University of Wisconsin Milwaukee

UWM Digital Commons

May 2021

An Optimization Analysis of the Subject Directory System on the Medlineplus Portal - An Investigation of Mental Health, Children, Teenagers, and Older Adults Related Health Topics

Yifan Zhu
University of Wisconsin-Milwaukee

Follow this and additional works at: https://dc.uwm.edu/etd

Part of the Library and Information Science Commons

Recommended Citation

AN OPTIMIZATION ANALYSIS OF THE SUBJECT DIRECTORY SYSTEM ON THE

MEDLINEPLUS PORTAL - AN INVESTIGATION OF MENTAL HEALTH, CHILDREN,

TEENAGERS, AND OLDER ADULTS RELATED HEALTH TOPICS

by

Yifan Zhu

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Information Studies

at

The University of Wisconsin-Milwaukee

May 2021

ABSTRACT

AN OPTIMIZATION ANALYSIS OF THE SUBJECT DIRECTORY SYSTEM ON THE
MEDLINEPLUS PORTAL - AN INVESTIGATION OF MENTAL HEALTH, CHILDREN,
TEENAGERS, AND OLDER ADULTS RELATED HEALTH TOPICS

by

Yifan Zhu

The University of Wisconsin-Milwaukee, 2021
Under the Supervision of Professor Jin Zhang

The Internet is a common means for people to search for health information. The subject

directory of MedlinePlus offers Internet searchers a browsing environment so that those seekers

could start from a broad term and refine their search terms to meet their real information needs,

thus resulting in a better information search. For those novice users who are not familiar with

relevant domain knowledge, MedlinePlus's directory can be of great assistance and enable the

portal to adopt to a more general population. Such a subject directory system and its involved

health topics in the MedlinePlus portal formed a network where a specific research methodology,

social network analysis, is applicable.

In this study, four health topic groups – mental health, children, teenagers, and older

adults - were selected as the focus for the investigation toward the subject directory on the

MedlinePlus portal. This study applied social network analysis to explore the health topic

directories and connection patterns among the health topics that comprised the subject directory

of the MedlinePlus portal, and identified the influential topics (i.e., those health topics which

play more important roles than others in connecting different topics) among the topic networks.

As a result, different recommendations were made toward mental health, children, teenagers, and

older adults related health topics, respectively. New optimized structural networks were suggested to be built for each of the four health topic subcategories according to the similarity values calculated through the cosine similarity measure in terms of the textual information contained in health topics' Web pages, as well as the key nodes identified in the networks of health topics. Evaluations were later conducted to compare the original and optimized structural networks of the four health topic groups regarding their topics' new similarity values. Newly identified influential health topics were verified to have improved the overall semantic connections among the whole networks. Last but not least, the recommendation results were evaluated by two health field experts and the evaluation outcomes proved that the recommendations suggested in this study were consistent with the opinions generated by health professionals.

The findings of this research will provide suggestions to optimize and enhance the current navigation guidance system in MedlinePlus, improve the information searching effectiveness among the portal users, offer insights to public health portal creators, and support other researchers focusing on subject directory systems.

iv

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

xiii

xiv

ACKNOWLEDGEMENTS

It has been quite a long journey and I have been fortunate to explore what I am truly passionate about as well as capable of during this impressive adventure.

I am forever grateful to my advisor and mentor, Dr. Jin Zhang, for his unreserved support and instruction throughout the past six years. It would not have been possible for me to complete this dissertation without my advisor's continuing guidance, encouragement, as well as support. He has modelled what an academic researcher should be, and has tailored my career plan of pursuing a research position in the academic environment.

I am also grateful to have exceptional committee members: Dr. Iris Xie, Dr. Xiangming Mu, Dr. MinSook Park, and Dr. Hengyi Fu. Thank you very much for reading my drafts and offering me enlightening feedback and suggestions. Thank you for your time and input.

I would like to thank Dr. Iris Xie for admitting me into the doctoral program, as well as the courses you taught about various qualitative research methods – they assisted me in building a solid foundation. Also, thank you and Dr. Xiangming Mu, for your continuing support and guidance throughout every stage of my doctoral program. I am grateful for the instruction received from Dr. MinSook Park, which prepared me to become a qualified course instructor. I would like to thank Dr. Fengyi Fu, for accepting my invitation of serving in my doctoral committee shortly after the birth of her child. In addition, I really appreciate the evaluation offered by the two health field experts from the University of California-Davis.

Last but not least, I would like to express my deepest gratitude to my parents, family, and friends who provided me endless love and support throughout these years.

Chapter 1. Introduction

1.1 Background

In the late 1980s, a report published by the National Academy of Sciences' Institute of Medicine claimed that "the public health of the nation was in disarray" (Heidenreich & Auflick, 2007). The report tried to invoke all the public health practitioners to rebuild their infrastructure as the ability of previous public health systems was identified as inadequate and the health of the public was considered unnecessarily threatened (Brodeur, 2005). As a result, Heidenreich and Auflick mentioned that proposals had been solicited from initiatives such as the "Turning Point: Collaborating for a New Century in Public Health" to "transform and strengthen the public health infrastructure" (p. 2). These proposals created opportunities for cooperation between public health agencies and communities at both state and local levels; thus embracing the concept of bridging the gap between public and non-public health entities at the very beginning of the 21st century.

As the Web keeps growing rapidly as a platform for information exchange, many major fields in people's daily lives have been heavily impacted - one of them is the accessibility to health information such as personal health-care and disease-associated questions (Ellis-Danquah, 2004; Yeo et al., 2010). This has been identified as an important change as people are able to find reliable health information online (Cline & Haynes, 2001). In the meantime, the health care system has gone through rapid development during the last decade which has led to higher medical cost as well as more advanced medical technology, and thus making it vital for health consumers to be well informed about their medical choices (C. Gray, 2006). As a result, some consumers use the Web to seek reliable sources of health information for quick and easy

response (Zhang, Zhai, Stevenson, & Xia, 2015), or for a place to share and communicate their experiences, and for anonymity and confidentiality (Bass et al., 2006). Many consumers also report that health information obtained from the Web is more satisfying compared with information collected from other media such as television, radio, newspaper, etc. (Eysenbach, 2003). Interestingly, consumers are not the only group that has benefited from online health information. Physicians are also involved. This makes patients today able to become partners in their own health care (C. Gray, 2000).

However, concerns also exist in the opinions held by some medical professionals regarding the fact that the quality of information on the Internet varies a lot (Susannah Fox & Fallows, 2003). Moreover, even for those most prevalent chronic diseases like cancer, many consumers are still not able to get access to the latest and most reliable information, thus requiring the cancer Web portals to be able to "translates evidence-based disease prevention and management information for text-based and multimedia delivery for diverse and low literacy citizens" (Schilling & McDaniel, 2010).

Among the Web resources that are available to the public for seeking health information, Web portals are playing a significant role. A Web portal refers to a site "that provides a number of services useful to a particular interest group or purpose" (Schleyer & Spallek, 2002). Such services might vary due to different sources or targeted user groups while many common features such as content aggregation, customization, and portal administration and maintenance function, have been shared among the majority of Web portals (Wege, 2002). However, some researchers believed that a Web portal with unique structural features can significantly impact user behaviors such as information searching and decision making (Baird, Furukawa, & Raghu,

2

2012).

Currently, the trend of searching health information online is indicated by the rapidly growing number of online public health portals (Benigeri, 2003; Evers, 2006). Among those comprehensive sites that are normally regarded as online health portals, MedlinePlus was launched in 1998 by the National Library of Medicine and was the first primary initiative for providing online health information to the public (Miller, Lacroix, & Backus, 2000). It provides a wide range of health-related information for both health consumers and health professionals that includes, but is not limited to, symptoms, diagnosis, prevention, and treatments about various diseases, introduction about medical terms, bodies, organs, and healthy daily living. Meanwhile, MedlinePlus provides a group of information retrieval tools including a search bar (engine), subject directory, sitemap, etc., to assist users in their information searching and site navigation. Among these tools, the health topic based subject directory system enables users to browse information from a general level to a specific level through its hierarchical structure, and vice versa. In addition, related subjects (health topics) are also listed in Web pages as individual topics so that users can "jump" to relevant topics when necessary.

Among hundreds of health conditions and diseases, mental disorders are generally characterized by the World Health Organization as "some combination of abnormal thoughts, emotions, behavior and relationships with others" ("WHO Mental Disorders Fact Sheet," 2018). According to the fact sheet published by WHO in 2018, the burden of mental disorders keeps growing "with significant impacts on health and major social, human rights and economic consequences in all countries of the world." Common mental diseases such as depression (300 million), bipolar affective disorder (60 million), and schizophrenia (23 million) have had a wide

3

range of patients globally. In the fact sheet, WHO also emphasized the importance of providing access to health care as well as social services that are able to offer treatment and support. Additionally, among the variety of health consumer groups, children and teenagers have long been the focus of previous studies (Kessler et al., 2005; Knitzer & Olson, 1982) since "adolescence and young adulthood" were regarded as critical timing for the appearance of mental health problems (Burns, Davenport, Durkin, Luscombe, & Hickie, 2010). Similarly, older adults were another major health consumer group where researchers worked to shed light on in terms of older adults' mental health information seeking behavior (Conner et al., 2010). Therefore, investigating child-related health topics, teenager-related health topics, and older adult-related health topics in the MedlinePlus portal is necessary for revealing the relationships between mental health and the three consumer groups.

This research focuses on the study of the subject directory system of MedlinePlus portal regarding mental health related health topics, child-related health topics, teenager-related health topics, and older adult-related health topics. The research objects are the health topics utilized as subject directories by the portal. Those health topics are collected and analyzed from four perspectives: the first is about the mental health related disorders, and the second is about the health topics related to children's health, the third perspective is about teenagers' heath related conditions, and the last is about health topics related to older adults' health. Social network analysis and both descriptive and inferential statistical analysis are presented to conduct a systematic analysis on mental health, children, teenagers, and older adults related subject directories on MedlinePlus.

1.2 Research problem, questions and hypotheses

Public health portals are playing a key role in people's daily life for obtaining health information. They provide basic knowledge and helpful resources to various sorts of health consumers. A Web portal like MedlinePlus contains dynamic, rich, and comprehensive information. The portal covers a variety of information. To better serve its users, the information has to be organized effectively and the portal has to provide its users with search mechanisms. An internal search engine and a subject directory are common means for this purpose. Query search allows users to find relevant information for a specific information need while browsing enables users to discover and explore relevant information related to a non-specific information need. A subject directory is a powerful tool to help users navigate the sophisticated Web portal and locate useful information to meet their information needs. Due to its hierarchy a subject directory helps users discover relevant information from a broad topic to a narrow topic, from a specific topic to a general topic, and from a topic of interest to more related topics. As a result, these unique characteristics make the subject directory a vital and useful tool in a Web portal.

Previous studies have shed light on both query search and browsing search when discussing the navigation system of public health portals. For browsing search and subject directory, those prior studies have mainly concentrated on the process and means of generating vocabulary. However, few researchers paid attention to the effectiveness of the existing topic-based subject directory among public health portals from the perspectives of different health themes or consumer groups.

5

1.2.1   *Research problem*

With the rapidly growing amount of people seeking health information online and large public health portals turning into vital health information resources, this research study focuses on the evaluation and optimization regarding the topic-based subject directory employed by the public health portal MedlinePlus. The primary research problem of this study is to explore, assess, and optimize the connections among the mental health topics, child-related health topics, teenager-related health topics, and older adult-related health topics in the subject directory of MedlinePlus. The research objects are the health topics related to mental health, children, teenagers, and older adults on MedlinePlus. These health topics refer to the corresponding topic pages created by MedlinePlus for introducing related information and resources. The connections among the health topics are important for the study. They can be used to construct networks. The connections among the health topics can be divided into two types: structural connections and semantic connections. To be more specific, structural connections represent the physical linkages set by the portal creators for connecting a specific health topic to other topics, while semantic connections refer to the linkages hidden behind the textual information of the topic pages among various health topics. In other words, the structural connection is determined by embedded links between the two topics while the semantic connection is defined by the similarity of the two topics' Web page content.

Based on the primary research problem, this study attempts to address the following six research questions as well as associated sub-questions. Moreover, corresponding null hypotheses are presented for the research questions in this study.

6

RQ1: How are health topics related to mental health, children, teenagers, and older adults connected in the subject directory on the MedlinePlus portal: are the structural and semantic connections of the health topics consistent?

To explore the overall research problem of this research, the first step is to figure out how the health topics are currently connected from both the structural and semantic perspectives. After various networks have been generated from both perspectives, it is then necessary to investigate the consistency between the structural and semantic connections among the selected health topics on the MedlinePlus portal. Theoretically speaking, in a well-designed subject directory, the semantic relationship should serve as the fundamental support for building a structural link network. The two networks should be consistent. If they are not consistent, end-users might lose vital related information or be guided to inappropriate pages during navigation. If that is the case, the subject directory needs optimization. This first research question can uncover the facts of connections built by the MedlinePlus portal creators and hidden behind each topic page's textual content.

MedlinePlus provides reliable and up-to-date health information to a wide range of health professionals and health consumers. It is sponsored by the world's largest medical library: The National Library of Medicine. The health topics-based subject directory of MedlinePlus is displayed in Figure 1. The MedlinePlus website uses five broad sections to divide its over 1000 health related issues: *Body Location/Systems, Disorders and Conditions, Diagnosis and Therapy, Demographic Groups, and Health and Wellness*.

Figure 1. The health topics-based subject directory of MedlinePlus

Each of these five sections includes a list of subcategories, and each subcategory has an introduction page that contains a group of health topics in an alphabetical order. A screenshot of an example page showing subcategories of *Blood, Heart and Circulation* is displayed in Figure 2.
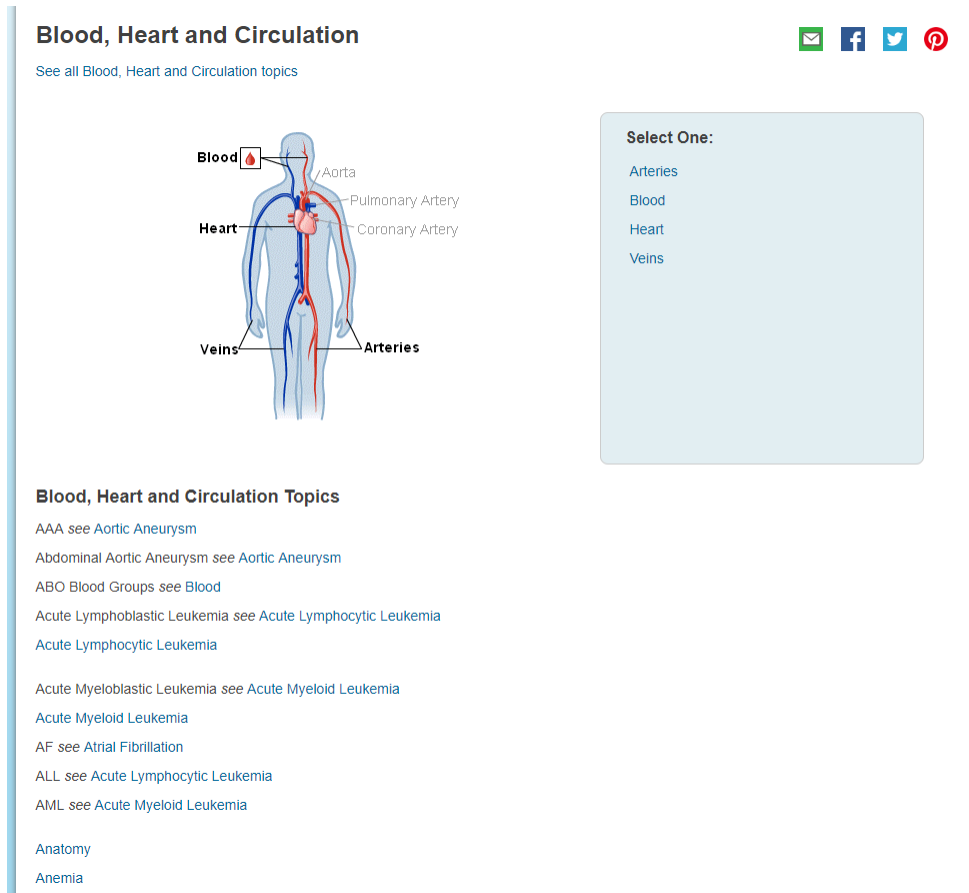
Figure 2. A page of a subcategory and its list of health topics on MedlinePlus

A health topic refers to a specific health-related issue and has an individual Web page starting from an overview table that includes "basics", "learn more", "see, play, and learn", "research", "resources", and "for you". Meanwhile, the Web page also has a "related health topics" column located on the right side to help end-users navigate other relevant health topics. For instance, the topic *Mental Disorders* has 19 related health topics: *Anxiety*, *Bipolar Disorder*, *Child Mental Health*, *Compulsive Gambling*, *Delirium*, *Depression*, *Eating Disorders*, etc. A screenshot of the health topic page of *Mental Disorders* is displayed in Figure 3 and the "related health topics" column is marked.

Figure 3. The topic page of *Mental Disorders* on MedlinePlus

Health topics and the relationships among them form a social network that reflects the

patterns and characteristics of mental health, children, teenagers, and older adults related topic

clusters. In the social network, actors refer to the health topics, while relationships among the health topics are represented as connections among actors/topics. The constructed social networks of mental health, children, teenagers, and older adults related health topics on MedlinePlus reveal patterns and characteristics of the relationships among those health topics. Interestingly, such relationships between a topic and one of its related topics are directional, which provides an ideal condition to apply the social network analysis method because in social network analysis, the interactions between two actors are not required to be mutual.

Social network analysis consists of measurements from two aspects: node-level (micro-level) and network-level (macro-level). Node-level measurements assess characteristics of a single topic's position within the network while network-level measurements examine features of the whole network. In this study, node-level measurements are applied to investigate the characteristics of each health topic's position within the corresponding networks, and network-level measurements are employed to explore the general patterns of the four health topic networks. The utilized measurements of social network analysis are displayed in Figure 4.
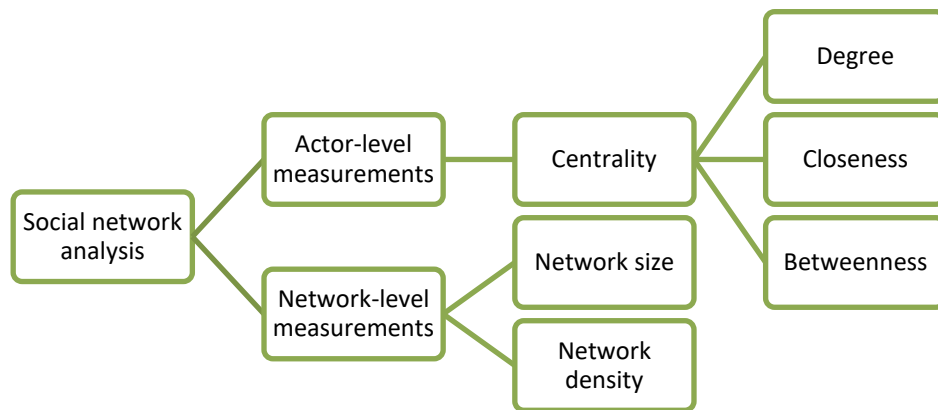
Figure 4. Social network analysis measurements

11

As shown in Figure 4, from the aspect of node-level measurements, centrality is selected as the actor feature for comparison. Centrality includes degree, betweenness and closeness. It can indicate the importance of an actor within a network from various perspectives. To be more specific, the degree refers to the amount of connections an actor possesses to other actors within the network. The betweenness represents the number of an actor sitting between pairs of other actors on their shortest paths (Freeman, 1978). The closeness measures the degree to which an actor is distant from other actors within the network (Freeman, 1978).

From the aspect of network-level, two measurements are applied to identify the patterns and characteristics of the networks regarding the mental health-related health topics, child-related health topics, teenager-related health topics, and older adult-related health topics. Various measurements can describe a social network from different perspectives. Network size represents the number of actors contained in a social network, therefore, it indicates the amount of health topics in the subject directory networks. Network density can be measured by the ratio of the number of all existing connections (or edges) among the actors to the number of whole connections a network could possibly have. In other words, network density refers to the proportion of actual connections among all possible connections (Borgatti, Everett, & Johnson, 2013).

1.2.2.1   RQ1.1 & RQ 1.2

Regarding the connections, two types of social networks can be generated among the health topics. The structural connections refer to the fact that one topic is listed as a "related health topic" to another topic. The structural connections have formed a structural network. On the MedlinePlus portal the structural network was literally set up by the portal creators to enable

end-users to navigate the website. The semantic connections refer to the semantic strength between two topics' Web page content. The semantic connections have formed a semantic network. The semantic network reflects how closely two health topics are related based on the similarity of their textual information. The varieties and frequencies of textual information (words) provide the quantitative evidence that is necessary for supporting the examination of relationships among health topics.

The consistency between the structural and semantic connection is examined from two general aspects: first, the structural and semantic connections are compared based on the similarity value among the health topics; second, node-level measurements (degree, closeness, and betweenness) are applied to identify influential nodes (health topics) within the two connection networks.

Following that thought, two sub-questions can be generated and applied to investigate the structural and semantic networks among the health topics. Figure 5 displays the structure of RQ1 as well as the two sub-questions RQ1.1, and RQ1.2.

RQ1.1: Are the similarity values between the structural and semantic connections among the health topics related to mental health, children, teenagers and older adults consistent on MedlinePlus?

RQ1.2: Are the node-level centrality measures between the structural and semantic connections among the health topics related to mental health, children, teenagers and older adults consistent on MedlinePlus?

13
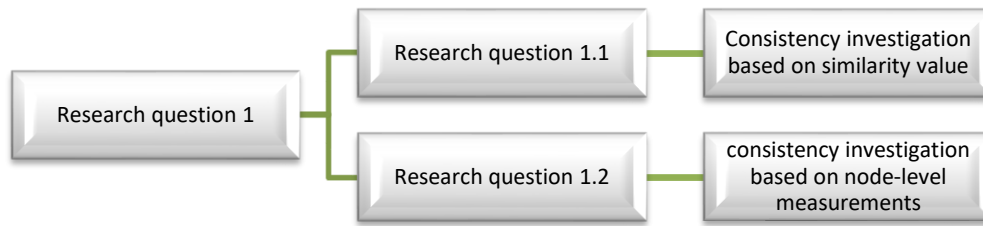
Figure 5. RQ1 structure

RQ1.1 is the fundamental part of the first question and provides evidence for RQ1.2. To answer RQ1.1, the textual information contained in each Web page of the selected health topics is transferred to similarity values according to a similarity measure in social network analysis. Descriptive statistical results including the average value and standard deviation are generated from the structural link network to serve as the threshold. If health topics without structural connections are found to have a similarity value larger than the threshold, it indicates that the structural connections do not match the semantic connections. Otherwise the structural and semantic connections are considered to be consistent.

RQ1.2 aims to identify and compare the influential health topics in the structural and semantic network among the health topics related to mental health, children, teenagers, and older adults on MedlinePlus. Influential actors/nodes (health topics) are also known as major players in the network. Compared with other actors, influential actors can have more impact on other actors. They play a more important role in controlling the information flow among the whole

14

network through locating at a comparatively central position. The influential health topics on the portal can be recognized when a topic is found to have many "related health topics" on its Web page and being listed as one of the "related health topics" on many other topics' Web pages. These characteristics quantify a health topic's impact or contribution to the subject directory and provide quantitative data to investigate the relationships among the health topics in the network.

In this study, the influential health topics are identified through social network analysis since one of the major purposes of social network analysis is to identify the major players in a network. To be more specific, node-level measurements including degree centrality, betweenness centrality, and closeness centrality are applied to discover influential health topics. Health topics which possess higher centrality measures are considered to play impactful roles in the network. Generally speaking, influential health topics should be roughly the same between the structural and semantic network. In addition, the ranking among those influential health topics should be consistent as well so that the most important health topics can play equal roles from both perspectives. If an influential health topic in the structural network does not have a high impact in the semantic network, or the most impactful health topic in the structural network is ranked low in the semantic network, it indicates that the two networks are inconsistent and optimizations are needed. Otherwise the structural network is proved to be well developed.

1.2.3   Research question 2 (RQ2)

RQ2: Are there significant differences between the original mental health subject directory and the optimized mental health subject directory in terms of its link structure on the MedlinePlus portal?

15

Suggestions for optimizing and enhancing the structural link networks among the selected mental health related health topics on the MedlinePlus portal are generated based on: 1) the similarity values of the structural and semantic networks; and 2) the features and rankings of important health topics in terms of connecting other topics in the mental health related networks. Those important health topics are identified through the analysis according to the node-level centrality measurements. Therefore, after the optimization recommendations are proposed, the optimized structural network should be evaluated through comparison to the corresponding original structural network.

1.2.3.1   RQ2.1 & RQ2.2

To identify the differences between the original and the optimized mental health related structural link networks, two sub-questions are listed to focus on various aspects:

RQ2.1: Are there any significant differences between the original and optimized structural networks of mental health related topics in terms of similarity values?

RQ2.2: Are there any significant differences between the original and optimized structural networks of mental health related topics in terms of node-level centrality measurements?

1.2.3.2   Hypothesis group 1

The following hypothesis is generated to answer RQ2.1:

H01: There are no significant differences between the original and the optimized structural networks in terms of similarity values among the health topics related to

mental health on the MedlinePlus portal.

H01 forms hypothesis group 1. The independent variable of hypothesis group 1 is the specific type of the structural link network (original/optimized). The dependent variable of hypothesis group 1 is the similarity measure results among the health topics in the original and optimized networks. The dependent variables can be measured through the similarity values.

1.2.3.3　Hypothesis group 2

The following three hypotheses are tested to answer RQ2.2:

H02$_{(a)}$: There are no significant differences between the original and the optimized structural networks in terms of node-level centrality measurements among the health topics related to mental health on the MedlinePlus portal according to degree centrality.

H02$_{(b)}$: There are no significant differences between the original and the optimized structural networks in terms of node-level centrality measurements among the health topics related to mental health on the MedlinePlus portal according to closeness centrality.

H02$_{(c)}$: There are no significant differences between the original and the optimized structural networks in terms of node-level centrality measurements among the health topics related to mental health on the MedlinePlus portal according to betweenness centrality.

H02$_{(a)}$, H02$_{(b)}$, and H02$_{(c)}$ compose hypothesis group 2. The independent variable of hypothesis group 2 is the specific type of the structural link network (original/optimized). The

dependent variable of hypothesis group 2 is the node-level centrality measurements among the health topics related to mental health in MedlinePlus. The dependent variables can be measured through degree centrality of each health topic, closeness centrality of each health topic, and betweenness centrality of each health topic.

1.2.4 Research question 3 (RQ3)

RQ3: Are there significant differences between the original children subject directory and the optimized children subject directory in terms of its link structure on the MedlinePlus portal?

Suggestions for optimizing and enhancing the structural link networks among the selected children related health topics on MedlinePlus are generated based on: 1) the similarity values of the structural and semantic network; and 2) the features and rankings of important health topics in terms of connecting other topics in the children related networks. Those important health topics are identified through the analysis according to the node-level centrality measurements. Therefore, after the optimization recommendations are proposed, the optimized structural network should be evaluated through comparison to the corresponding original structural network.

1.2.4.1 RQ3.1 & RQ3.2

To identify the differences between the original and the optimized children related structural link network, two sub-questions are listed to focus on various aspects:

RQ3.1: Are there any significant differences between the original and optimized structural networks of children related topics in terms of similarity values?

RQ3.2: Are there any significant differences between the original and optimized structural networks of children related topics in terms of node-level centrality measurements?

1.2.4.2 Hypothesis group 3

The following hypothesis is generated to answer RQ3.1:

H03: There are no significant differences between the original and the optimized structural networks in terms of similarity values among the health topics related to children on the MedlinePlus portal.

H03 forms hypothesis group 3. The independent variable of hypothesis group 3 is the specific type of the structural link network (original/optimized). The dependent variable of hypothesis group 4 is the similarity measure results among the health topics in the original and optimized networks. The dependent variables can be measured through the similarity values.

1.2.4.3 Hypothesis group 4

The following three hypotheses are tested to answer RQ3.2:

H04$_{(a)}$: There are no significant differences between the original and the optimized structural networks in terms of node-level centrality measurements among the health topics related to children on the MedlinePlus portal according to degree centrality.

H04$_{(b)}$: There are no significant differences between the original and the optimized structural networks in terms of node-level centrality measurements among the health topics related to children on the MedlinePlus portal according to closeness centrality.

19

H04(c): There are no significant differences between the original and the optimized structural networks in terms of node-level centrality measurements among the health topics related to children on the MedlinePlus portal according to betweenness centrality.

H04(a), H04(b), and H04(c) compose hypothesis group 4. The independent variable of hypothesis group 4 is the specific type of the structural link network (original/optimized). The dependent variable of hypothesis group 4 is the node-level centrality measurements among the health topics related to children in MedlinePlus. The dependent variables can be measured through degree centrality of each health topic, closeness centrality of each health topic, and betweenness centrality of each health topic.

1.2.5   Research question 4 (RQ4)

RQ4: Are there significant differences between the original teenagers subject directory and the optimized teenagers subject directory in terms of its link structure on the MedlinePlus portal?

Suggestions for optimizing and enhancing the structural link networks among the selected teenager-related health topics on MedlinePlus are generated based on: 1) the similarity values of the structural and semantic network; and 2) the features and rankings of important health topics in terms of connecting other topics in the teenagers related networks. Those important health topics are identified through the analysis according to the node-level centrality measurements. Therefore, after the optimization recommendations are proposed, the optimized structural network should be evaluated through comparison to the corresponding original structural network.

20

1.2.5.1   RQ4.1 & RQ4.2

To identify the differences between the original and the optimized teenagers related structural link network, two sub-questions are listed to focus on various aspects:

RQ4.1: Are there any significant differences between the original and optimized structural networks of teenagers related topics in terms of similarity values?

RQ4.2: Are there any significant differences between the original and optimized structural networks of teenagers related topics in terms of node-level centrality measurements?

1.2.5.2   Hypothesis group 5

The following hypothesis is generated to answer RQ4.1:

H05: There are no significant differences between the original and the optimized structural networks in terms of similarity values among the health topics related to teenagers on the MedlinePlus portal.

H05 forms hypothesis group 5. The independent variable of hypothesis group 5 is the specific type of the structural link network (original/optimized). The dependent variable of hypothesis group 5 is the similarity measure results among the health topics in the original and optimized networks. The dependent variables can be measured through the similarity values.

1.2.5.3   Hypothesis group 6

The following three hypotheses are tested to answer RQ4.2:

H06(a): There are no significant differences between the original and the optimized

structural networks in terms of node-level centrality measurements among the health

topics related to teenagers on the MedlinePlus portal according to degree centrality.

H06(b): There are no significant differences between the original and the optimized

structural networks in terms of node-level centrality measurements among the health

topics related to teenagers on the MedlinePlus portal according to closeness centrality.

H06(c): There are no significant differences between the original and the optimized

structural networks in terms of node-level centrality measurements among the health

topics related to teenagers on the MedlinePlus portal according to betweenness

centrality.

H06(a), H06(b), and H06(c) compose hypothesis group 6. The independent variable of

hypothesis group 6 is the specific type of the structural link network (original/optimized). The

dependent variable of hypothesis group 6 is the node-level centrality measurements among the

health topics related to teenagers on MedlinePlus. The dependent variables can be measured

through degree centrality of each health topic, closeness centrality of each health topic, and

betweenness centrality of each health topic.

1.2.6   Research question 5 (RQ5)

RQ5: Are there significant differences between the original older adults subject directory and

the optimized older adults subject directory in terms of its link structure on the MedlinePlus

portal?

Suggestions for optimizing and enhancing the structural link networks among the

selected older adult-related health topics on MedlinePlus are generated based on: 1) the

similarity values of the structural and semantic network; and 2) the features and rankings of important health topics in terms of connecting other topics in the older adults related networks. Those important health topics are identified through the analysis according to the node-level centrality measurements. Therefore, after the optimization recommendations are proposed, the optimized structural network should be evaluated through comparison to the corresponding original structural network.

1.2.6.1 RQ5.1 & RQ5.2

To identify the differences between the original and the optimized older adults related structural link network, two sub-questions are listed to focus on various aspects:

RQ5.1: Are there any significant differences between the original and optimized structural networks of older adults related topics in terms of similarity values?

RQ5.2: Are there any significant differences between the original and optimized structural networks of older adults related topics in terms of node-level centrality measurements?

1.2.6.2 Hypothesis group 7

The following hypothesis is generated to answer RQ5.1:

H07: There are no significant differences between the original and the optimized structural networks in terms of similarity values among the health topics related to older adults on the MedlinePlus portal.

H07 forms hypothesis group 7. The independent variable of hypothesis group 7 is the

23

specific type of the structural link network (original/optimized). The dependent variable of hypothesis group 7 is the similarity measure results among the health topics in the original and optimized networks. The dependent variables can be measured through the similarity values.

1.2.6.3   Hypothesis group 8

The following three hypotheses are tested to answer RQ5.2:

H08$_{(a)}$: There are no significant differences between the original and the optimized structural networks in terms of node-level centrality measurements among the health topics related to older adults on the MedlinePlus portal according to degree centrality.

H08$_{(b)}$: There are no significant differences between the original and the optimized structural networks in terms of node-level centrality measurements among the health topics related to older adults on the MedlinePlus portal according to closeness centrality.

H08$_{(c)}$: There are no significant differences between the original and the optimized structural networks in terms of node-level centrality measurements among the health topics related to older adults on the MedlinePlus portal according to betweenness centrality.

H08$_{(a)}$, H08$_{(b)}$, and H08$_{(c)}$ compose hypothesis group 8. The independent variable of hypothesis group 8 is the specific type of the structural link network (original/optimized). The dependent variable of hypothesis group 8 is the node-level centrality measurements among the health topics related to older adults on MedlinePlus. The dependent variables can be measured through degree centrality of each health topic, closeness centrality of each health topic, and betweenness centrality of each health topic.

24

1.2.7 Research question 6 (RQ6)

RQ6: Are there any significant differences between the optimized structural network from this study and the evaluation results from the health field experts?

Besides testing the effectiveness of the recommended connections within the structural networks of each health topic group themselves, it is also important to seek for confirmation from experts in the medical field. Therefore, the optimized structural networks are also compared to the evaluation results generated by two health field professionals.

These two health field experts were recruited from a different research institution and have a solid background in the medical/health area. They have either gone through the preliminary examination of a M.D. program or hold at least a master's degree in a medical or health related field and they possess at least five years of experience in the field. They were invited to offer manual reviews of the connections derived from the four health topic groups on the MedlinePlus portal. Considering the fact that most subject directories are derived from human review processes, it is necessary to make sure that the optimized structural networks proposed by this study fit with the opinions of health experts.

1.2.7.1 Hypothesis group 9

The following four hypotheses are generated to answer RQ6:

H09(a): There are no significant differences between the optimized structural network from this study and the evaluation results from the health field experts regarding the mental health subject directory.

25

H09(b): There are no significant differences between the optimized structural network

from this study and the evaluation results from the health field experts regarding the

children subject directory.

H09(c): There are no significant differences between the optimized structural network

from this study and the evaluation results from the health field experts regarding the

teenagers subject directory.

H09(d): There are no significant differences between the optimized structural network

from this study and the evaluation results from the health field experts regarding the older

adults subject directory.

H09(a), H09(b), H09(c), and H09(d) form hypothesis group 9. The independent variable of hypothesis group 9 is the recommendation decisions toward selected health topic connections proposed by this study and the health field experts. The dependent variable of hypothesis group 9 is the structural link connections. The dependent variables can be measured through the numbers of suggested and not suggested connections.

1.3 Definitions of terms

Before investigation and analysis of the subject directories related to mental health, children, teenagers, and older adults can begin, key concepts must be defined and justified. This section contains definitions and justifications of the key terms involved in this study.

1.3.1   Mental health & mental disorders

According to the definition of the National Library of Medicine,

26

Mental health … includes our emotional, psychological, and social well-being. It affects how we think, feel, and act as we cope with life. It also helps determine how we handle stress, relate to others, and make choices. Mental health is important at every stage of life, from childhood and adolescence through adulthood. … Mental illnesses … are serious disorders which can affect your thinking, mood, and behavior. They may be occasional or long-lasting. They can affect your ability to relate to others and function each day. Mental disorders are common; more than half of all Americans will be diagnosed with a mental disorder at some time in their life. ("Mental Health," n.d.)

1.3.2 Public portal

A Web "portal" was defined by Schleyer and Spallek (2002) as a "site that provides a number of services useful to a particular interest group or purpose" (Schleyer & Spallek, 2002, p. 204). In this study, only those public portals focusing on providing services and information useful to a community were discussed. Examples of websites include MedlinePlus, the World Health Organization, the National Institute of Health, etc. Patient portals as well as other websites that offer personal medical records related services were excluded.

1.3.3 MedlinePlus

Based on its own introductory page, MedlinePlus is a service of the world's largest medical library – the National Library of Medicine (NLM), and a part of the National Institute of Health (NIH) ("Learn about MedlinePlus," n.d.). It is constructed as an online health information resource and meant to assist patients, as well as their families and friends. The portal aims at presenting "high-quality and relevant health and wellness information that is trusted and easy to understand." MedlinePlus consists of a variety of features that include health topics, medical tests, medical encyclopedia, drugs & supplements, healthy recipes, and special collections. Both health consumers and professionals are expected to benefit from the portal resources.

27

1.3.4   Health topic

In this study using the MedlinePlus portal, a health topic can refer to a disease, an illness, a health condition, a wellness issue, or a specific health consumer group related health theme, such as *Child Development*. Each health topic has an independent Web page which introduces symptoms, causes, treatment, preventions and other related health information regarding the corresponding health topic. Currently, MedlinePlus has built Web pages for over 1000 health topics.

1.3.5   Navigation system

Based on guidelines provided by usability.gov, the word "navigation" refers to the means utilized for seeking information within a website. A navigation system should function in two ways: 1) assisting users to be aware of where they are and what content they are facing (Krug, 2006; Rosenfeld & Morville, 2002); and (2) helping users to locate specific information and link them to the destination page (Raju N & Harinarayana, 2008).

1.3.6   Subject directory

According to the ICT course demonstration of University of South California,

Subject directories, unlike search engines, are created and maintained by human editors, not electronic spiders or robots. The editors review and select sites for inclusion in their directories on the basis of previously determined selection criteria. The resources they list are usually annotated. Directories tend to be smaller than search engine databases, typically indexing only the home page or top level pages of a site. They may include a search engine for searching their own directory (or the Web, if a directory search yields unsatisfactory or no results). (Johnston, 2006)

Subject directories are important because users often begin their search in a too general way (Taylor, 1967) due to lack of prior knowledge about their search tasks (Pang et al., 2016),

28

hence a subject directory may suggest narrower terms that can better fit the information needs of the users. This process is normally regarded as a process of refinement, and indicates that users find real information related to their need via association - they start from a wrong term, then go through browsing in directories, and finally identify relevant information (Bates, 1989; Demelo et al., 2017).

1.3.7   Node

The term "node" refers to the health topics contained in the subject directory system of the MedlinePlus portal. Each health topic is represented by a node in the generated social networks.

1.3.8   Edge

The term "edge" refers to the physical relationship among health topics in this study. To be more specific, it refers to the tie shared between a pair of health topic in the social network. Therefore, the edges present the structural connections among the selected health topics related to mental health, children, teenagers, and older adults.

1.3.9   Health information

According to the definition provided by the American Health Information Management Association (AHIMA), health information refers to the data related to a person's medical records. Those records could include patient histories, such as symptoms, diagnoses, outcomes, etc. and lab results, as well as other clinical information. From a broader perspective, health information also contains general information about health conditions, services, healthcare resources, clinical trials, and so on, according to the "Health Information" Web page of the

29

National Institutes of Health ("Health Information," n.d.).

1.3.10 Health consumer

In Lewis et al.'s book section (2005), the authors quoted the definition for a "health information consumer" from The American Medical Informatics Association, Consumer Health Informatics Working Group, and the International Medical Informatics Association, Nursing Informatics Interest Group, as "a person who seeks information about health promotion, disease prevention, treatment of specific conditions, and management of various health conditions and chronic illnesses" (p.1). Meanwhile, Lewis et al. also pointed out that health information consumers not only contain individuals suffering from specific health conditions and their families and friends, but also the public "concerned about promoting optimal health" (Lewis, Chang, & Friedman, 2005).

1.3.11 Consumer health information

A widely used definition for consumer health information was from Patrick and Koss's working draft paper (1995) where the authors defined the concept as,

> Any information that enables individuals to understand their health and make health-related decision for themselves and their families. This includes information supporting individual and community-based health promotion and enhancement, self-care, shared (professional-patient) decision-making, patient education, patient information and rehabilitation, health education, using the healthcare systems and selecting insurance or healthcare provider. (Patrick & Koss, 1995)

1.3.12 Social network analysis

Social network analysis was first theorized as a new concept by (Barnes, 1954), with its roots lying in theories of social action (Coleman, 1986). It had become an established field

within the social science area by the 1980s (Borgatti, Mehra, Brass, & Labianca, 2009). This concept of networks has been applied to investigate social phenomena in a broad range of disciplines such as economics and psychology. Social network analysis was regarded as a gold mine by Borgatti et al. since it "provides an answer to a question that has preoccupied social philosophy since the time of Plato, namely, the problem of social order: how autonomous individuals can combine to create enduring, functioning societies" (p.892).

Social network analysis method examines, measures, and evaluates the characteristics of a network and connections among its actors (or nodes) on a network. Social network analysis focuses on the interactions between the actors. The uniqueness of the social network analysis method is that it can quantify the connections of a node to other nodes on a network from multiple perspectives. In addition, it measures not only the relationships between the node and its directly linked nodes but also the relationships between the node and indirectly linked nodes on the network.

1.3.13 *Centrality*

Centrality refers to a series of indicators generated through various measurements in order to identify key actors/nodes in a network. A group of centrality measures have been developed during the past decades to identify the core actors and investigate the structural characteristics of a network. In this study, three centrality measures are included. Degree centrality refers to the amount of connections an actor possesses to other actors within the network. Betweenness centrality represents the number of an actor falling between pairs of other actors on their shortest paths (Freeman, 1978). Closeness centrality measures the degree to which an actor is distant from other actors within the network (Freeman, 1978).

31

The purpose of applying similarity measures in this study is to determine the strength

between the nodes in the semantic connection network. According to Polamuri's article

published on Dataaspirant portal,

> The similarity measure is the measure of how much alike two data objects are. Similarity
> measure in a data mining context is a distance with dimensions representing features of
> the objects. If this distance is small, it will be the high degree of similarity where large
> distance will be the low degree of similarity. The similarity is subjective and is highly
> dependent on the domain and application. For example, two fruits are similar because of
> color or size or taste. Care should be taken when calculating distance across
> dimensions/features that are unrelated. The relative values of each element must be
> normalized, or one feature could end up dominating the distance calculation. Similarity
> are measured in the range 0 to 1 [0,1] (Polamuri, 2015).

Three similarity measures were applied in this study. Among them, the cosine-similarity

measure was utilized first with all research data. Based on the definition from the "ScienceDirect

Topics", "cosine similarity measures the similarity between two vectors of an inner product

space. It is measured by the cosine of the angle between two vectors and determines whether two

vectors are pointing in roughly the same direction. It is often used to measure document

similarity in text analysis" ("Cosine Similarity - an overview | ScienceDirect Topics," n.d.).

The cosine-similarity measure has been widely applied to the information retrieval area

(Baeza-Yates & Ribeiro-Neto, 1999; Zhang & Rasmussen, 2001) due to the fact that for

documents containing the same distributed or proportional weighted keywords, the cosine-

similarity measure can be an effective tool in identifying their similarities (Korfhage, 1997).

After results have been generated through the cosine-similarity measure, two other

similarity measures were employed as supplemental methods toward the same datasets for

comparing purposes. These additional two similarity measures are the Pearson correlation

similarity measure and the Euclidean distance similarity measure. In his book chapter, Segaran (2007) defines the Pearson correlation similarity measure as a way to measure how highly correlated two variables are. In other words, it "quantifies how well two data objects fit a line". Benefits of the Pearson correlation similarity measure include that it "can be used when quantities (i.e. scores) varies", also, "objects that describe the same data but use different values can still be used." The Euclidean distance similarity measure describes the "square root of the sum of squared differences between corresponding elements of the two vectors", and it always serve as the "basis of many measures of similarity and dissimilarity" (Borgatti, 2013).

Chapter 2. Literature review

2.1 Introduction

The Internet has arisen as one of the major health information resources. Among nearly six million Americans, up to 75% were found to have used the Web for seeking health information (Hewitt, Greenfield, & Stovall, 2006; Jibaja-Weiss & Volk, 2007). Moreover, about 70% of Web users regard the Internet as a tool that has helped improve their capability of finding health information. How they retrieved the information and later make use of the collected health information remains a mysterious (Eysenbach, 2003; Fogel, 2003). Health consumers have transferred from passively accepting orders from their physicians to actively searching for online information that relates to their health conditions (Nie, Zhao, Akbari, Shen, & Chua, 2015). As a result, due to this change of providing healthcare services, in recent years an increasingly growing number of community based healthcare services occurred to spread personalized health information and link consumers with physicians in a worldwide range through question answering (Nie, Li, Akbari, Shen, & Chua, 2014).

A more interesting finding is that among a group of selected Internet users that participated in a previous study. Although 80% of them did have confidence in seeking health information in the pretest interview, the fact is that almost half of them were not able to find quality information, which clearly indicated a need for skills improvement (Ivanitskaya, O'Boyle, & Casey, 2006). As Ellis-Danquah (2004) pointed out, the comprehensiveness of online health information has the potential to enable consumers to make important health-related

34

decisions. However, they have to overcome the challenges brought by the myriad of Web information resources and find specific, reliable, and timely health information first. "The availability of authoritative, easy-to-use, consumer health information facilitates both health education and information literacy, especially among underrepresented groups" (p. 63).

This section addresses previous studies covering public health portals regarding their navigation systems and subject directory, health information related to children, teenagers, and older adults, and their online health information seeking, and literature discussing social network analysis and its application in information studies and health informatics fields. The investigated research papers are from areas of health informatics, biomedical, social network, information studies, and their corresponding inter-disciplines. This paper aims to find trends, methods, technologies, challenges, and applications within the included topics.

2.2 Public health portals

According to Schleyer & Spallek (2002), a Web "portal" refers to a "site that provides a number of services useful to a particular interest group or purpose" (p. 204). Prior research found that 80% of the subset of online users seeking Web information started their search more from a general portal site (46%) than from a search engine like Google (33%) (Horrigan, Rainie, & Cornfield, 2003). Therefore, Web portals could be seen as the front door for the information needed by online users (Zhang, Zhai, Liu, & Stevenson, 2016). More than that, as users always desire easy access to and valid information from Web portals, the interaction between portals and users has been enhanced under the integration of Web 2.0 (Postigo, 2011). As a result, some researchers are convinced that a successfully built Web portal could significantly affect users' information searching and decision making through its innovative structure (Baird et al., 2012).

Given different sources and targeted user groups, Web portals vary on their services; however, there are some common features such as the portal administration and maintenance function, content aggregation and customization, multi-device support, user login page, etc., based on the nature and audience groups that are shared among many Web portals (Wege, 2002).

The rising number of online public health portals is considered to demonstrate the trend of seeking health information through the Internet (Benigeri, 2003; Evers, 2006). These health portals usually refer to comprehensive sites such as MedlinePlus where consumers can find information like symptoms, diagnosis, treatment, preventions, etc. regarding specific diseases or other health conditions; but a broader definition of public health portal may also involve Web directory sites, government sites, association sites, drug information, alternative medicine, nutrition and food safety, and other websites containing health issues (C. Gray, 2006). Except for these daily information seeking activities, other scholars also point out that there could be a pressing need for public health portals when bioterrorism and pandemics events containing public health threats, such as H1N1 flu, happen so that people can remain appropriately informed with objective and data-driven health information (Gilbert, 2009). For public health portals, besides the requirement of being professional and high quality, some consumers with special health conditions might also expect them to be warm and friendly (Schilling & McDaniel, 2010).

The popularity of online health portals created researchers' interest in evaluating the qualities like accessibility, content, credibility, scope, etc., and the research findings indicated that it is necessary to involve health professionals in the design and review of these health websites (Benigeri & Pluye, 2003). Effort has long been devoted into guiding consumers to reliable information resources as the American College of Physicians (ACP) Foundation and the

National Library of Medicine (NLM) "joined forces to encourage physicians to direct their patients to MedlinePlus" (Gray, 2006, p. 396) which is a nationwide public health portal created and maintained by the NLM. Gray also pointed out that despite the fact that most public health portals provide information including diagnosis and treatment, many of them still highly recommend that their consumers consult with family physicians and health care providers regarding specific medical conditions.

2.2.1   *Public health portals related studies*

Previous studies conducted focused on a variety of topics about public health portals. From the perspective of the service range, some local public health portals such as the AZHealthInfo located in Arizona (Heidenreich & Auflick, 2007) and Healthy Texas serving Texas consumers (Moore & Kaercher, 2008) have been researched in their history, development, mission, and management. Moreover, similar studies were also carried out in European public health portals like Spanish hospital websites by local scholars (García-Lacalle, Pina, & Royo, 2011).

From the aspect of users, studies focusing on minority groups and health disparities (Ellis-Danquah, 2004) were collected and challenges such as the lack of health literacy of health portal users by normal consumers (Liechty, 2011) and disable users (Kushalnagar et al., 2015) were also discussed. From the aspect of health portals, evaluations about the websites' interface design, usability, and information credibility of public health portals based on the feedback from users (Schilling & McDaniel, 2010) and health professionals (Pullen, Jones, & Timm, 2011) were also involved in this review. Similarly, evaluation toward the information quality contained in public health portals was found in Gray's (2006) research where he emphasized the

importance of critically evaluating websites including medical information and he introduced several rules created earlier by professionals and organizations. Well-known standards include the four criteria (authority, currency, objectivity, and audience) listed by the Medical Library Association, and the PILOT method (Purpose, Information, Links, Originator, Timeliness) proposed by Joan Price and Shannon Entin (Price, 2000). These famous rules as well as other evaluation means that could be applied in the assessment regarding health portals are part of the popular issues relating to public health portal research. Other types of evaluation, for instance, mainly focusing on a broad introduction toward usability and content of health portals were also employed regarding regional health websites in Kentucky (Gilbert, 2009).

2.2.2   Concerns and challenges on public health portals

A great number of Internet users have troubles finding authoritative information due to the lack of skills using Web search functions. As a result, it becomes quite challenging for many of them to differentiate the effective health information from misleading ones ("Clear & Simple | National Institutes of Health (NIH)," n.d.; Eysenbach, 2003). Moreover, for health information providers, previous research has revealed the fact that simply offering huge amount of information does not necessarily mean it is helpful (Friedman, Hoffman-Goetz, & Arocha, 2006); hence the mission of providing information is not only just displaying correct information but also focusing on how to present and make such information actually useful in real decision making scenarios (Fogel, 2003). Therefore, as stated in their research paper, Schilling and McDaniel (2010) listed a few challenges that portal designers face during the creation of the portal. These challenges contain the readability, reading skills, health literacy, information literacy, computing literacy, learning style, etc. According to the collected earlier literatures,

38

health literacy and health disparities are the two most commonly discussed topics. In addition, many scholars have indicated linkage existing between the two issues (Berkman et al., 2004; Kushalnagar et al., 2015).

2.2.2.1 Health literacy

Health literacy is an important issue that is worth further discussion. Health literacy is defined as to what extent an individual possesses the ability to collect, handle, and understand fundamental health information and services required for appropriate health decision making (*Healthy people 2010*, 2000; Ratzan, Parker, Selden, & Zorn, 2000). Based on the Integrated Conceptual Model for Health Literacy (Figure. 6) created by Sorensen et al. (2012), being health literate means more than just successfully accessing the health information, but furthermore, understanding and utilizing the collected information to achieve keeping and improving one's own health.
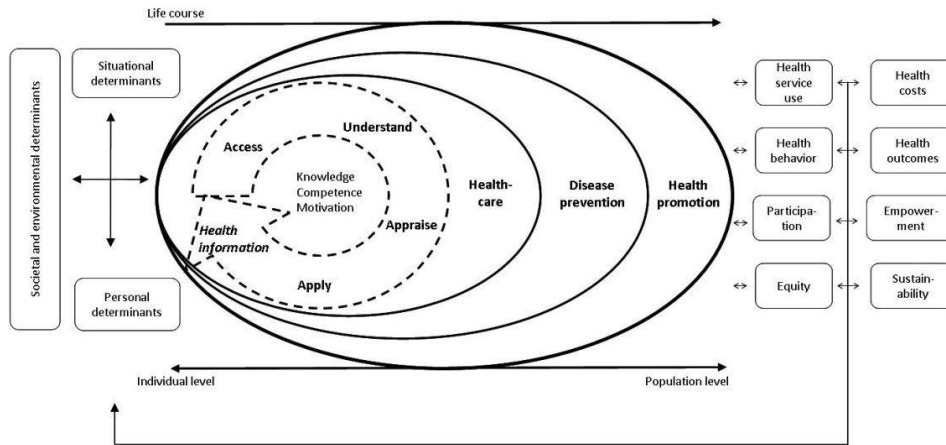
Figure 6. Integrated model of health literacy (Sørensen et al., 2012).

39

Prior studies have not only shed light on the fact that low health literacy should be regarded as a primary concern as it usually relates to disappointing health consequence, higher health care cost, difficulties in health communication, and could even lead to health disparities (Berkman et al., 2004; Committee on Health Literacy, 2004), but also have pointed out that more than one-third of American adults do not possess sufficient health literacy in managing their health care demand (Kutner, Greenberg, Jin, & Paulsen, 2006). This point was echoed by other researchers as they claimed that a lot of health care providers, as well as patients are still unaware of the lack of health literacy despite the fact that such a problem costs an estimate of fifty-eight billion dollars in the health care system of the United States every year (Committee on Health Literacy, 2004; Hewitt et al., 2006). Therefore, the development of health literacy has become a major target of the new Healthy People 2020 report (*Healthy people 2020*, 2010) as it has been proved that higher health literacy and access to targeted information could enhance patient compliance as well as their satisfaction with the treatment consequences (Helft, 2008).

An additional interesting finding about the health literacy issue stated in previous research was that health literacy relates to listening, speaking, problem-solving, and critical thinking ability, as well as conceptual knowledge (Schilling & McDaniel, 2010). Although health literacy was frequently linked to demographic elements like age, race, ethnicity, coverage of Internet access and health insurance, etc., no correlation with educational level was reported by Schilling and McDaniel. Ways of easing the problem of low health literacy include patient education and support, health care system improvement, health care provider accommodations, etc. (Kandula et al., 2009; Volandes et al., 2008), with extra emphasis on the assistance offered from social workers (Liechty, 2011).

40

2.2.2.2   Health disparities

Ellis-Danquah (2004) stated that unlike seeking health information in the real world, the seeking behavior itself for online health information does not create the so-called "health disparities", which is identified by the Institute of Medicine as,

> Differences in the quality of care received by minorities and non-minorities who have equal access to care–this is, when these groups have similar health insurance and the same access to a doctor–and when there are no differences between these groups in their preferences and needs for treatment. (Smedley, Stith, & Nelson, 2003)

Instead, there is no significant difference among consumers of various racial and ethnic backgrounds while searching for health information through the Internet. Therefore, Ellis-Danquah (2004) claimed that the Web has created a relatively equal way and can serve as a fairly influential means for health promotion and awareness. However, from the website content perspective, Ellis-Danquah also found that there are very few public health portals providing information on specific racial and ethnic health issues, compared with the websites sponsored by the government and other professional health organizations. Moreover, the health disparities regarding website content could also be related to health education and information literacy.

On the contrary, Schilling and McDaniel (2010) did identify that health disparities exist among public cancer information portals as many of them are not very well accessible to users who "have poor literacy skills, are socioeconomically disadvantaged, non-English speaking, or are otherwise on the wrong side of health disparities and the digital divide" (p. 244). Such point was echoed by Yi (2012) as he claimed that the global Web users always lack access to online information resources which were mainly written in English due to the language gap. Yi also mentioned that "word to word translation among languages is not sufficient to accommodate

41

language differences" (p. 250). On the other hand, when discussing about implementing cross language information retrieval, some previous studies stated several existing challenges such as translation ambiguity and vocabulary mismatch. Researchers claimed that those challenges were extremely common in the health field (Ballesteros & Croft, 1998; Gao et al., 2001). Moreover, recent researchers have also shed light on consumers with physical disabilities such as deaf users (Kushalnagar et al., 2015). Members from Deaf American Sign Language (ASL) reported challenges with searching and understanding health information, accessing health care services, and comprehending treatment procedures (McKee & Paasche-Orlow, 2012). Kushalnagar et al. considered the phenomenon as a sort of health disparity and pointed out that such fact might leave negative impact on disabled consumers while understanding and applying newly-learnt health information to their health care related decision making.

2.2.2.3   Regional public health portals and other challenges

For regional communities, before local public health portals were widely set up, researchers from some areas had pointed out that there was not even a single Web portal that was available for consumers to find information about local health care in a state like Texas, especially for residents with low health and computer literacy (Moore & Kaercher, 2008). Hence, regional public health portals were built "in response to fragmentation of local health services, limited access to local health information, literacy and language barriers to health information, and frustration with communications between provider and patient" (p. 96). On the other hand, for regional public health portals created during the first round of health information network development such as the AZHealthInfo (created for health information within the state of Arizona), researchers also found that they had uneven geographic and subject coverage

(Heidenreich & Auflick, 2007). Specific problems claimed by Heidenreich and Auflick include:

1) Despite the fact that most small towns are linked, not every place on the map could be able to find links

2) Links and pages created by rural communities and counties are often found to be short-lived or unstable

3) Some health topics on local portals have great amount of content and related resources while many other subjects lack sufficient information

4) Websites built and maintained by rural areas are always turned out of date

As a result, some scholars like Moore and Kaercher (2008) reported that regional health portals like HealthyTexas have not been utilized as frequent as was expected.

Another challenge faced by health portal creators is how to balance users' stay between portal content and external content (Schilling & McDaniel, 2010) as users might leave forever if they jumped to an external link before viewing at least four individual pages of the health portal itself. Hence, Schilling and McDaniel claimed that the ultimate idea of providing external health information resources is to keep consumers in the relevant pages within the portal, and then guide them to external links. Other means of attracting consumers to travel back to the portal include offering news feeds, FDA alerts, or task and content management features such as personal calendars, reminders, etc., and even some social network functions.

2.2.3   *Public health portals' development*

According to previous research, national and local public health portals have now

43

achieved reaching some common rules. For instance, Heidenreich and Auflick (2007) stated that the local public health portal of Arizona - the AZHealthInfo website, shares the same collection development guidelines and quality controls utilized by MedlinePlus. The main concerns toward the selected resources for public health portals today include the content and maintenance quality, the limitation for usage (for example, commercial content, advertisement, required registration, etc.), and accessibility toward the open public.

For the state level, Heidenreich and Auflick (2007) suggested that the committee of AZHealthInfo give up trying to cover every single health topic. Instead, they started to narrow down the scope and focused only on subjects that attract the most care among people who live in Arizona. Moreover, the local portal stopped creating duplicate pages with national resources like MedlinePlus and Go Local, but attempted to refer these links from appropriate locations on its own pages, thus connecting the proper information for the state residents. The same idea has been applied in Texas while building the Healthy Texas portal because the designers, at the beginning of construction, regarded the website as a bridge to unite national and local resources, as well as local services (Moore & Kaercher, 2008). Later, Moore and Kaercher emphasized that the organizers also believed that the local public health portal could have done more - that was, to meet specific and unique regional needs, just like what AZHealthInfo did. Other than that, another special aspect of regional health portals like Healthy Texas is the attention paid on patients' emotional support - Healthy Texas was trying to enable users to communicate or publish articles and other art works through the website so that consumers might get comfort from each other.

On the other hand, the user-centered approach has been more frequently applied in public

health portal design so that "users' points-of-view were at the forefront at every stage" (Schilling & McDaniel, 2010, p. 247). In addition, Schilling and McDaniel have also discussed several key points regarding the development of public health portals. For instance, a portal design focusing on cancer information might include:

1) Providing users with quality-filtered, user-driven evidence-based information on cancer care and management, risk and prevention, patient-provider communication, and lifestyle issues

2) Providing information in engaging and easily readable, graphical, multimedia, and interactive formats

3) Providing users with opportunities to personalize content and output, and encourage patient-provider communication and interaction. (p. 243)

Moreover, health portal designers have been trying to build content in non-textual formats over time in order to improve the effectiveness of information usage so that the portal could get closer to consumers with multiple learning styles and skill levels. The application of other content formats like audios, videos, games etc., is expected to strengthen the interaction between users and portals (Eichner & Dullabh, 2007; Houts, Doak, Doak, & Loscalzo, 2006; Kools, Van De Wiel, Ruiter, Crüts, & Kok, 2006). Echoing with the emphasis pointed on health portals' ability of interacting, other researchers have also conducted evaluations toward the interaction functions among a list of most popular public health portals (C. Gray, 2006), thus proving this part of development as a current trend in the health field. Following the trend, some scholars have started to pay attention to specific minority consumer groups such as deaf Web users and emphasized the importance of providing content in multimedia format through interactive design (Kushalnagar et al., 2015).

The previous studies introduced above have provided a big picture of the exploration conducted among public health portals and their users. To better discuss the research focusing on

the usability of these health portals, prior literatures paying attention to navigation systems and subject directory systems will be addressed in the following sections.

2.3 Navigation systems

Based on the definition given by the usability.gov guideline, navigation refers to the means utilized for seeking information within a website. It aims at assisting users to be aware of where they are and what content they are facing (Krug, 2006; Rosenfeld & Morville, 2002), as well as helping users to locate specific information and link them to the destination page (Raju N & Harinarayana, 2008). A broader range of navigation systems might also cover text size and contrast, ability to access information in non-textual formats, and directions for finding aids like an A-Z index. Every means for easing the use of a website could be regarded as part of a navigation system (Pullen et al., 2011).

2.3.1   Functions and roles of navigation systems

Navigation systems play a crucial role for Web portals in meeting seeking needs of users (Zhang et al., 2016) as users were considered to always need context to comprehend their search activities (Webster & Ahuja, 2006). As a result, navigation systems offer possibility and flexibility for users who have limited experience and knowledge regarding the portal, and help them to have the search process kept in control (Minetou, Chen, & Liu, 2008). The importance of a functional navigation system was extremely underlined in studies aiming at improving the ability of seeking health information among disabled health consumers (Kushalnagar et al., 2015).

Navigation systems have been studied a lot since the appearance of the Internet. Serving

46

as an important section of a Web portal's usability, functions and designs of navigation system have caused most researchers' interests.

2.3.2   Components and designs of navigation systems

2.3.2.1   The home page

Generally speaking, the design of the interface of Web portals is very important as it provides the connection between users and website information (Zhang et al., 2015). The home page has been regarded as the main entry and starting point of a navigation system for a website as it introduces the overall website structure and provides links to the lower-level pages (Raju N & Harinarayana, 2008). In addition, Raju N and Harinarayana emphasized that the link feature for returning to the home page should be offered from all the consecutive pages of the website for user to "quickly jump from lower level pages to home for new tasks or explore further" (p. 114). In their study focusing on university library websites which share similar functionalities with public health portals, Raju N and Harinarayana introduced three main types of return home links that were provided by most of the websites (93.35%) - the organization logo, an explicit home feature link, and a "going back home" banner, which was considered a clear improvement for the navigability feature of the website.

2.3.2.2   Global, embedded, local, and breadcrumb navigation

The global navigation, also known as the site-wide navigation system, refers to the feature that is designed to appear on every destination page throughout the whole website providing access to each of the key sections (Rosenfeld & Morville, 2002). In most cases for university library websites that contain the global navigation, the feature is usually presented as

47

the navigation bar located at the top of the page (Raju N & Harinarayana, 2008).

The embedded navigation is defined simply as "the collection of hyperlinks which supports global, local and contextual navigation on the websites" (Raju N & Harinarayana, 2008, p. 116). Raju N and Harinarayana pointed out that the importance of this feature grows as the amount of other features of websites increases, and the pop-up menus have been identified as the major means of achieving the embedded navigation features among university library websites. Earlier research argued that the embedded navigation sometimes looks less obvious to the users (Adkisson, 2002), but Raju N and Harinarayana stated that now the feature is being much more emphasized in order to save the "screen real estate" (p. 120).

Local navigation refers to the specific navigation information within a particular section of a website (Raju N & Harinarayana, 2008). The authors also mentioned that among university library websites, most of the local navigation features are figured out to be designed in the main content area. Last but not least, the breadcrumb navigation is the feature that tells user about where or in which part they are currently looking about the website (Raju N & Harinarayana, 2008). In other words, the feature shows the hierarchical structure of the site to the user, sometimes with a "you are here" label, like a map. In their study toward university library websites, Raju N and Harinarayana claim that most sites that apply a breadcrumb navigation feature use the right arrow mark while other sites decide to use the pipe symbol.

Similar classification for the types of navigation systems were also applied in articles focusing on portals providing hospital and medical related resources. In those articles, the authors' introduction included global, local, and contextual navigation systems given the definitions according to the range of functioning (Krug, 2006; Rosenfeld & Morville, 2002).

48

Another interesting finding was that prior scholars have claimed that many public health portals were coordinating with or under construction and maintenance by libraries, especially for regional health websites (Heidenreich & Auflick, 2007; Moore & Kaercher, 2008). Therefore, navigation systems of public health portals usually share a lot of same characteristics and functions with those of library websites mentioned above.

2.4 Subject directories

Before public health portals appeared and subject directories were applied, traditional means to organize online health-related resources were mainly based on normal information organization methods such as standard terms, classification, and taxonomy (Cimino, Octo Barnett, Hassan, Blewett, & Piggins, 1991; Kuśnierczyk, 2008; Lu, Lin, Chan, & Chen, 2008). Later, researchers started to shed light on ontology-based retrieval tools such as topic maps (Hajdukiewicz, Vicente, Doyle, Milgram, & Burns, 2001). According to Gruber (1993) and Pepper (2002), ontology was defined as a unique knowledge domain produced from one's conceptualization toward a specific object while topic maps were defined as an ISO standard for explaining various knowledge structures and linking them with appropriate information resources. In 2001, Beier and Tesche (2001) built a searching system based on a topic map transformed from the medical subject heading (MESH) and found the system was capable of providing fast access to relevant resources. Similarly, Stanescu et al. (2008) utilized a topic map to navigate medical resources and identified it as the most appropriate model to visualize health information. On the other hand, in the daily life health information seeking scenarios, the search vocabulary used by health professionals and normal health consumers has been widely recognized to be different (Ratzan et al., 2000; Zeng & Tse, 2006). Therefore, a need was

identified where physicians and health care providers could better communicate with their patients; and health consumers could better express their conditions and requirements (Zhang & Wolfram, 2009).

Subject directories, as well as the search engines, are the most common information retrieval tools for seeking Web information (Chung & Noh, 2003). Normal subject directory systems (such as Yahoo! were identified by Chung and Noh) "do not apply standard classification schemes like DDC or UDC but use specific schemes designed for Internet information resources" (p. 118) whereas some other specialized directory systems intended for online scientific information were found to follow standard library schemes. For instance, among the twenty-six websites offering scientific information investigated by McKiernan (2001), twenty-three of them classify their resources by either DDC or UDC. For public health portals, subject directories could also be generated from targeted audience group interviews, reviews of relevant literature and other existing health portals, and communication with clinical institutions or other coordinators (Schilling & McDaniel, 2010). However, although sharing common sources for classification, the organizational structures of subject directories among public health portals vary a lot (C. Gray, 2006); some use an alphabetical list while others apply a topic-based interface to organize their health information about specific diseases and conditions.

Compared with search engines, some researchers considered the subject directory services as a better retrieval tool since it provides more satisfying results that have gone through manual review and classification (Chung & Noh, 2003). However, Chung and Noh also pointed out that such manual classification work cannot persist as Web documents tremendously increase. When it comes to the health area, as Yeo et al. (2010) pointed out in their study,

50

"existing disease information systems support the classification of disease data and provide users with the data through the web" (p. 60) while the data formats applied among different databases vary. In addition, echoing with Chung and Noh regarding manual classification, some studies stated that the information organized by some Web portals are sponsored by questionable experts, display information without a powerful research base, and have set reading levels that apparently exceed the average level of Web users (Basch et al., 2005; Friedman & Hoffman-Goetz, 2006; Friedman et al., 2006).

2.4.1   *Effectiveness and quality improvement of subject directories*

Before automatic means for producing directory systems occurred, most of the subject directories on the Web for either general or specialized sites, were classified or categorized by the editors or surfers until some researchers like Chung and Noh (2003) and Yang and Lee (2004) proposed directory systems based on automatic classification, text-mining technology based on Self-Organizing Map, and text categorization techniques to automatically assign Web content to relevant subject categories, thus reducing the manual working hours. This idea originated from "Cora", a specialized retrieval engine that utilized an extended Naive Bayes classifier to achieve automatic classification of Web documents and group them into 75 predetermined subject categories. "Cora" was published in 1998 and its classification results were reported to reach 66% (McCallum, Nigam, Rennie, & Seymore, 1999). Following the idea, similar designs also appeared in Yeo et al.'s study (2010) where they applied the database technology and built a linked subject directory system in a consistent content format. Their system was formed of attributes such as diagnosis, symptoms, treatment, etc., in order to organize information collected from different professional databases. As a result, the challenges

51

faced by the public health consumers were eased.

Moreover, some other prior studies attempted to better collect and understand the vocabulary used by specific groups of consumers from a creative way. They obtained users' searching query terms through transaction logs of a public health portal and applied a clustering analysis to explore related terms about the five pre-selected keywords regarding obesity. Their findings revealed the fact that health professionals might be able to build a new thesaurus about obesity based on these users' vocabulary in a more effective way (Zhang & Wolfram, 2009). Later from another perspective, Zhang and An (2009) chose to focus on a subject directory that already existed on a public health portal - weight control - and then employed the transaction logs collected to explore other subjects that were most relevant to weight control through a combined approach containing component plane analysis and Self-Organizing Map. Their research findings indicated several subjects, such as nutrition and herbs, and women's health from various relevance levels to weight control, and provided an improved suggestion for a better related subject directory for the specific topic.

Recently, as social network analysis arose with greatly increased online searching, researchers have also proposed to optimize subject directory through a social network analysis of the structural and semantic relationships among various subject terms so that the interconnectedness among categories and subcategories of subject directory could be improved (Zhang et al., 2016). Similarly in the health field, researchers have also applied the same idea in order to check the relationships among the great number of health topics. The authors employed a creative research design combining social network analysis method, clustering analysis, and inferential statistical means to a public health portal with a topic-based navigation guidance

52

system - the World Health Organization portal (Zhang et al., 2015).

In conclusion, among the prior literatures, the exploration of subject directory systems can be divided into two parts: one is from the perspective of users since Web portal design is now shifting from system-oriented to user-oriented; the other is from the system's aspect which contains early studies focusing on the medical thesaurus, as well as newly-occurred means in detecting structural and semantic network relationships such as social network analysis.

2.5 The integration of public health portals, navigation system, and subject directory

Today, the increasing amount of websites as well as the information they contain has grown beyond one's imaginative power whereas the ability of organizing such great amount of information for easier retrieval falls behind (Raju N & Harinarayana, 2008). Information search mechanisms are crucial for Web portals in ensuring their accessibility and success (Zhang et al., 2015). The navigation system and other directional aids serve as part of the interactive design of the website (Schilling & McDaniel, 2010), hence should not only present information to consumers, but also reflect how these consumers interact with the information provided by the portal during their decision making ("Clear & Simple | National Institutes of Health (NIH)," n.d.; Hsee, 1996, 1998).

For the evaluation toward health portal navigation systems, Schilling and McDaniel (2010) explored a list of questions investigating: 1) if the connection built among menu categories, content options, and external links normally function; 2) if directional prompts and location markers of the portal pages are clearly displayed; and 3) if the Web users perceive comfort while navigating the portal without feeling lost or frustrated and can keep clear on where

53

they are located within the website at any time, etc. A typical qualified instance for public health portals' navigation systems is the one belonging to MedlinePlus as it was identified as "easy to navigate from the home page as well as from secondary pages since the navigation tabs for the primary topic areas appear throughout the site" (Gray, 2006, p. 400).

For most of the public health portals investigated, there are two main search functions provided - the subject directories for browsing search and the search engine capability for query search (Zhang et al., 2015). Browsing search had raised much attention from researchers since the occurrence of Web-based information seeking and a lot of models were developed in previous literature such as Choo et al.'s (2000). Some prior studies regarded browsing search as a different way of getting informed in an increasingly fragmental digital environment and argued that browsing search may provide a deeper engagement to information seekers compared with search engines and social media platforms (Möller et al., 2019).

Similar portal transitions of browsing search were observed in an Eastern European public library's website as well (Shevchenko, 2020). The researchers mentioned that the improvements brought by setting up related topic pages from certain patent subjects had led the average amount of portal page transitions to a 40% level. In addition to public health and library domains, subject directories were also widely utilized in other domains. For instance, a taxonomy containing 12 categories was developed for wayfinding behaviours in the UK (Barker, 2019).

The subject directory is usually designed to have a hierarchical or tree structure as it categorizes objects sharing similar attributes and characteristics into groups to enable users to browse the subject framework in a holistic and intuitive way (Zhang et al., 2016). Zhang et al.

54

also pointed out that unlike search engine function, a subject directory provides to users information from general to specific and related content; hence could be regarded as a classification scheme. For such topic-based navigation guidance system, previous studies have provided a great perspective where the authors paid attention on the linkage among the pre-selected topics from both structural aspect (like manually set hyperlinks) and semantic aspect (like the most frequently occurred words shared by two topics) through a comparison based on social network analysis means (Zhang et al., 2015). Recommendations for optimization and enhancement of the navigation guidance system would be offered in cases such as two topics sharing common keywords in their webpages which did not possess hyperlink to connect to each other. Another interesting perspective in public health portal related research was obtaining and analyzing data from transaction logs collected from health websites in order to improve both understanding of consumers' vocabulary (Zhang & Wolfram, 2009) and the structure of the subject directories applied by the portal (Zhang & An, 2009).

A navigation system without subject directory guidance might cause problems like navigation confusion, lack of interactivity, unimpressive interface design, etc. (Zimmerman & Paschal, 2009). As claimed in Heidenreich and Auflick's study (2007), the subject directories of some public health portals could be integrated from opinions of health professionals, terms and phrases used by library patrons, and vocabulary that is already applied by some authoritative organizations like the National Library of Medicine. On the other hand, the search engine capability of public health portals is intended to offer deeper indexing for not only the portal itself, but also the sites it links to so that "terms and phrases that are buried two or three levels deep within a linked web site" (p. 11) can be found.

Besides the discussion regarding the two means of information searching from the theoretical perspective, other researchers also investigated how health consumers seek information online in real life. Authors such as Pang et al. (2016) reported that users' health information seeking behaviour might be affected by the various nature of search tasks, and browsing search was found to be connected to serendipity as well as curiosity since browsing search could often bring health information that seekers "do not know much about" or "have not thought of" before the search task. This conclusion was further related to the fact that seekers might apply browsing search when they have little prior knowledge about a specific health topic. Hence, such searching could provide hints for further search directions. This finding was echoed in a later study conducted by Demelo et al. (2017) where they had identified the lack of domain-specific knowledge and vocabulary as a major barrier for health information seeking. As a result, Demelo et al. developed an ontology-driven search interface and found that the interface was helpful for addressing knowledge/vocabulary difficulty. Moreover, evidences were observed in another Pang et al.'s (2016) paper in which they obtained the navigation data of visitors to one of the largest public health portal in Australia. Their findings revealed that "a number of users continue to look for additional information after the first read, by navigating to other pages, browsing the home page, and using the search functions". To be more specific, significant navigation flows were investigated among categories like "conditions and treatments" and "A-Z health content", which implied the importance of subject directories in assisting health consumers for seeking information. However, these prior studies mostly focused on general health information seeking, few studies were found to shed light on the mental health related area.

2.5.1   *Subject directories of regional portals*

For regional and local public health portals, take for instance, the Arizona public health website (AZHealthInfo was introduced in Heidenreich and Auflick's article (2007) as an instance). The initial subject directory of this website came from consulting with public health professionals and local public librarians. Heidenreich and Auflick (2007) pointed out that in order to build a "home- grown", "user driven" subject directory, the indexing vocabulary absorbed terms and phrases used among the clients of the local public libraries. Most importantly, all the terms were required to be able to prepare for further edit and to include new words as needed. In the long term plan, Heidenreich and Auflick stated that the subject directory of AZHealthInfo should attempt to keep consistent with other health-related websites such as MedlinePlus, MeSH, and local public health service organizations. However, regional public health portals are also encouraged to retain those terms that cannot be found in national health portals as they might be able to reflect special informational desire for that specific place. The six broad subject directories identified by AZHealthInfo are:

1)   Demographic Groups

2)   Disorders and Conditions - Body Location Systems

3)   Environment

4)   Health and Wellness

5)   Medical Specialties

6)   Public Safety/Safety Issues

57

Slightly different from AZHealthInfo, the subject directory employed in Healthy Texas was nearly all developed from local opinions (Moore & Kaercher, 2008). According to Moore and Kaercher's article, the initial subject list was obtained through consumer drafts and local reports reflecting feedbacks of residents living along the Texas - Mexico border. Then the list was reviewed and edited by physicians, healthcare providers, and language experts.

2.6 Health information seeking

According to the definition provided by Niederdeppe et al. (2007), health information seeking refers to,

> Any nonroutine media use of interpersonal conversation about a specific health topic and thus includes behaviors such as viewing a special program about a health-related treatment, using a search engine to find information about a particular health topic on the Internet, and/or posing specific health-related questions to a friend, family member, or medical practitioner outside the normal flow of conversation. (p. 155)

Unlike information scanning, health information seeking emphasizes the intention and active effort devoted in the process of collecting specific health information besides the normal characteristics of information exposures (Griffin, Dunwoody, & Neuwirth, 1999). Meanwhile, Griffin et al. also pointed out that interpersonal resources should be regarded as an important part of health information seeking.

In Rains' book (2018), three principles were introduced as highly relevant to health information seeking: 1) the process of health information seeking is a consultation process that combines several various sources. Besides the Internet, health providers, friends and families, magazines, books, newspapers, brochures, and other types of sources may all play a role. Such preference for combining both online and offline sources among health information seekers were proved in several previous studies (Balka, Krueger, Holmes, & Stephen, 2010; Ruppel & Rains,

58

2012). 2) During the process of health information seeking, the selection and utilization of a specific information source can be affected by a variety of factors. Among those factors, three major groups of variables were identified: demographic factors such as age, gender, or education level; health-related factors including previous experiences with a health condition; and the characteristics of the information sources themselves (Johnson & Case, 2012; Johnson & Meischke, 1993). All three major groups of factors can impact source selection directly or indirectly. Moreover, besides the three major groups mentioned above, other factors were reported as well (Dutta-Bergman, 2013). For instance, the Internet might be abandoned during the health information seeking process due to uncertainty (Brashers, Goldsmith, & Hsieh, 2002). 3) Reactions to received health information may contain both positive and negative sides. Information collected does not necessarily lead to a decreased uncertainty but in some cases, information seekers may feel more uncertain about their health conditions or prognosis (Brashers, 2001; Hogan & Brashers, 2009). Health consumers obtaining information can result in acquiring additional information or simply being ignored. Hence, one cannot assume that positive outcomes are guaranteed after receiving information.

2.6.1   *Health information seeking online*

2.6.1.1   Demographic patterns of online health consumers

Starting from 2000, a group of surveys have been conducted by researchers at the Pew Internet and American Life Project, as well as the National Cancer Institute, to investigate the health information seeking behavior among U.S. adults. According to their findings, Internet use for collecting health information is a "pervasive phenomenon" (Rains, 2018). Researchers found that the proportion of all adult U.S. Internet users who sought health information online was a bit

more than 50% during 2000 (Rice, 2006). That number was reported to have increased to 72% by 2012 (S. Fox & Duggan, 2013). Those findings were echoed by the National Cancer Institute's surveys through which they concluded that the percentage of U.S. adult Internet users seeking health information online was 51% in 2003 and 79% during 2013 ("Health Information National Trends Survey Data," n.d.).

For patients that are curious about both present and long-term effect of their or their family's health conditions, information seeking is a "critical mechanism" for exploring answers toward questions raised by or related to the illness (Johnson & Case, 2012). Among various health consumer groups, some were reported to have a high percentage of population seeking health information online, such as patients with irritable bowel syndrome (Halpert et al., 2007) and patients eligible for bariatric surgery (Paolino et al., 2015), while other groups were found to perform less online seeking, such as hematology outpatients in Belgium (Laurent, Cremers, Verhoef, & Dierickx, 2012) and patients diagnosed with HIV/AIDS (Samal et al., 2011). In addition, among the general public, online information seeking was claimed to be distinguished by specific demographic characteristics (including gender, health condition level, educational background) and Internet-related factors (such as Web skills) despite the existence of exceptions and inconsistent conclusions in some studies (Rains, 2018).

2.6.1.2   Search topics of online health information seeking

Generally, health information sought by health consumers online could be classified into two topics: technical information regarding a health condition from a medical perspective; and experiential information that can provide personal experience as well as emotional support among people who are dealing with a specific illness (Fergie, Hilton, & Hunt, 2016; S. C. Kim,

Shah, Namkoong, Mctavish, & Gustafson, 2013).

Experiential information can normally be found in online communities and social network sites because in these platforms user generated content is more privileged. Technical information appears more frequently in health portals that are operated by authoritative organizations like the National Institute of Health since they cover information related to specific conditions and treatments (Rains, 2018). With the integration of Web 2.0, the connection between online Web portals and users has been strengthened (Postigo, 2011). As a result, these portals have become the front door to get access to the information needed by online health consumers (Zhang et al., 2016).

Moreover, from the aspect of constructing subject directories regarding online health information, various health conditions were found to be linked to specific health consumer groups. For instance, according to the WHO website and MedlinePlus portal, several health issues such as diabetes, development issues, and school health are shared among children and teenagers. However, it is clear that teenagers are faced with more mental illness related conditions since they are more connected to underage drinking, underage driving, sexual health, pregnancy, drug abuse, and depression. On the other hand, older adults are normally linked to health conditions such as elder abuse, memory issues like Alzheimer's Disease, stroke, heart problems, and so on. These facts indicate some relationships exist among specific health topics in the subject directory systems of many online public health portals.

2.6.1.3 Evaluation of online health information

Although a paradox has been reported to lie between a health consumer's perceptions and

61

use of the Internet as a source for health information (Rains, 2018), significant concerns regarding the quality of health information available online have been consistently identified among numerous previous studies covering a variety of health conditions. Among patients facing bariatric surgery, nearly 20% of them regarded surgery-related information posted online as untrustworthy sources (Paolino et al., 2015). Similar evidences displaying skepticism toward online health information were also reported among people suffering from multiple sclerosis (Colombo et al., 2014), deaf adults (Karras & Rintamaki, 2012), and women undergoing breast cancer (Balka et al., 2010). Nonetheless, health consumers continue to seek health information online.

Among various factors that may impact how health consumers evaluate the quality of information obtained online, the context in which the information is presented may be of great importance since it can trigger "a heuristic about the presence of gatekeepers monitoring the quality of information online" (Rains, 2018). Such heuristic can significantly affect health consumers' perceptions regarding the health information they receive. For instance, some health information posted on a formal portal was found to be more impactful than that posted on a personal blog or online community. This heuristic has been proposed in the MAIN model in Sundar's book section (2008) and proved through an experiment in Hu and Sundar's study (2010). Other factors that may influence the evaluation toward online health information include the seeker's prior experience and knowledge, name recognition, and good reviews from other health consumers (Prybutok & Ryan, 2015).

2.6.1.4 Motivations for health information seeking online

In his book, Rains (2018) also addressed and discussed the motivations for seeking health

information online. Five specific motivations were summarized by Rains based on the prior study conducted by Boot and Mejiman (2010) and displayed in Table 1.

Motivation	Definition
Acquiring knowledge	Desire to find and verify information related to one's health condition.
Managing uncertainty	Desire to cope with uncertainty related to one's health condition and increase one's sense of control.
Achieving solidarity	Desire to connect with others facing similar challenges.
Entertainment	Desire to pass the time or appease one's curiosity.
Accessibility	Potential to easily access large volumes of health information.

Table 1. Motivations for seeking health information online (Rains, 2018)

The first motivation identified for seeking health information online was to acquire knowledge. To be more specific, Internet users tend to explore and verify factual information through their seeking behavior (Boot & Meijman, 2010). Such motivation of acquiring knowledge includes: confirming one's existing knowledge, for instance, among young women suffering from breast cancer (Balka et al., 2010); attempting to become better informed about one's current situation, for example, among men coping with prostate cancer and women undergoing breast cancer in German (Kirschning & Von Kardorff, 2008); and preparing for or following up after the consultation with a health provider (Colombo et al., 2014), or even trying to question a health provider (Balka et al., 2010).

The second motivation could be related to the first one since it indicates that health consumers seek health information online to manage their uncertainty (Boot & Meijman, 2010). Previous researchers claimed that health consumers may obtain a better sense of control regarding their health conditions through the process of seeking information (Balka et al., 2010;

63

Kirschning & Von Kardorff, 2008).

The third motivation was represented from the social perspective (Boot & Meijman, 2010) where "information seeking can be a means to achieve solidarity with others facing similar situations and may be more prevalent among patients who possess more meaningful social relationships" (Rains, 2018, p. 88). This motivation has been proved by a prior study in which the researchers claimed that health consumers with more close confidants were more likely to seek health information online compared with those with few close confidants (Askelson, Campo, & Carter, 2011).

The fourth motivation indicates that seeking health information online may contain no specific purposes other than entertainment (Boot & Meijman, 2010). In a previous study conducted by Karras and Rintamaki (2012), the authors stated that information seeking was reported to be applied because of curiosity instead of meeting an immediate objective after they interviewed a group of deaf adults.

The last motivation relates to the characteristics of Internet resources. Previous researchers pointed out that people seek health information online due to the various types of content contained online, the flexibility of time limitation, and the ease of usability (De Choudhury, Morris, & White, 2014; Jones et al., 2014). Such convenience and availability of information may be of great importance for health consumers affected by any health care barriers (Rains, 2018) and was proved among oncology patients in Australia (Newnham et al., 2006). In addition, the convenience of online health information may also include information presented in one's native language (J. Yi & Zahn, 2010) and the possibility to obtain technical information that has been interpreted for a lay audience (Kirschning & Von Kardorff, 2008).

2.6.2 *Mental health information seeking on the Internet*

Earlier studies have revealed the fact that health consumers with poorer health conditions tended to rely on the Internet when collecting information related to both physical and mental health issues (Susannah Fox & Rainie, 2000). Following that thought, the standards for sites containing physical and mental health related content were proposed by researchers (Morahan-Martin, 2004). Some scholars compared common mental health disorders (CMHDs) with diabetes and claimed that health consumers suffering from these two types of health conditions shared several similarities such as the requirement of self-management (Sterling, von Esenwein, Tucker, Fricks, & Druss, 2010), hence they both considered seeking information online as a crucial element (Fergie et al., 2016).

Similarly, in a survey sent out to British Internet users, around 1/5 of them were reported to have used the Internet for obtaining mental health related information (Powell & Clarke, 2006). The proportion was even higher among those ones who had a past history of mental health issues or were suffering from psychological conditions at the moment. However, Powell and Clarke also pointed out that "the Internet is used more than it is trusted" since the Internet was regarded by only half of the users as an "accurate source" among all health consumers utilizing the Internet for mental health information seeking.

Moreover, prior studies claimed to find that information seeking behaviors regarding all primary mental illnesses and/or problems through Google had seasonal patterns (Ayers, Althouse, Allem, Rosenquist, & Ford, 2013). To be more specific, "All mental health queries followed seasonal patterns with winter peaks and summer troughs amounting to a 14% difference in volume for the U.S. and 11% for Australia." Interestingly, these seasonal patterns were also

65

evident for specific mental illnesses and problems such as anxiety, bipolar, depression, suicide, etc.

2.6.2.1 Mental health related information seeking among children

Children's health has been identified as a common topic among the experiences of seeking health information online since 1998 (Cyber Dialogue, 1998). However, the seeking behavior of children has been rarely discussed in the previous literature. Considering the lack of required online searching literacy and other qualifications among children, their needed health information, for instance, a cough condition (Pandolfini, Impicciatore, & Bonati, 2000), would normally be sought by their parents (Khoo, Bolt, Babl, Jury, & Goldman, 2008). Among all searched children conditions, mental illnesses such as ADHD (attention deficit/hyperactivity disorder) was also included (Sage et al., 2018; Tandi Lwoga & Florence Mosha, 2013).

2.6.2.2 Mental health related information seeking among teenagers

Unlike children, teenagers were identified to be able to search health information online on their own and the seeking behavior of them had been studied from a variety of perspectives. For mental health, Although 30.8% of young college students aged between 18 and 24 were found to use the Internet for mental health information and support, questions remained on how specifically they utilized the Internet (Horgan & Sweeney, 2010). Primary search topics among these young college students included depression, course work, general mental health information, etc.

Different from the findings mentioned above, in a qualitative research conducted in Oregon among three focus groups of young adults, five major topics related to mental health

66

information seeking online were medications, diagnosis, treatment options, access to health care, and supports and resources (Gowen, 2013). Gowen also classified the motivations for seeking mental health related information online among those young adults into five categories: "seeking out additional information"; "seeking out community"; "nowhere else to turn"; "preparing for a mental health visit"; and "anonymity". Moreover, two major barriers usually encountered during the information seeking process were identified – information overload and concern over information accuracy.

For young people aged from 12 to 25 in Australia, the Internet was used by the majority to connect with other young people on mental health issues (Burns et al., 2010). This phenomenon was echoed by other studies, and young adults were found to have contributed a great amount of mental health related content to various Internet platforms; thus receiving great opportunities for intervention from health professionals (Yonker, Zan, Scirica, Jethwani, & Kinane, 2015). Meanwhile, nearly half of the young adults in Australia were reported to search for information about specific mental health problems even if they were not having any. The authors also reported that such information seeking activities, including search patterns and the selection of online resource types regarding mental health, varied according to age and gender.

2.6.2.3 Mental health related information seeking among older adults

Besides children and teenagers, another important health consumer group that has raised great attention from researchers is the group of older adults. Compared with children and teenagers, older adults were found to encounter more barriers in seeking for mental health related information online. According to Conner et al.'s article (2010), "Stigma associated with having a mental illness has a negative influence on attitudes and intentions toward seeking mental health

67

services among older adults with depression, particularly African American elders."

2.7 Children, teenagers, and older adults related health information

According to the Dictionary by Merriam-Webster, child refers to "a young person especially between infancy and puberty", while teen refers to a person aged between 13 to 19. Meanwhile, based on the explanation provided by the WHO's article, older adult refers to people aged over 65 in most development countries, or aged over 60 agreed by the United Nations, or briefly "the age at which one can begin to receive pension benefits" ("Proposed working definition of an older person in Africa for the MDS Project," 2002). On the other hand, mental disorders are generally characterized by the World Health Organization as "some combination of abnormal thoughts, emotions, behavior and relationships with others" ("WHO Mental Disorders Fact Sheet," 2018).

2.7.1   Child-related health information

According to the WHO website and MedlinePlus portal, children are closely linked to a group of health issues including diabetes, violence, nutrition, growth disorders, genetic problems, school health, and mental health. As for the resources containing child-related health information, a list of websites specifically designed for children to get access to health information had been identified by Izenberg and Lieberman back to 1998 (Izenberg & Lieberman, 1998). However, Prior studies had questioned the reliability of health information related to children, such as children with high fevers (Impicciatore, Pandolfini, Casella, & Bonati, 1997), and childhood diarrhea (McClung, Murray, & Heitlinger, 1998), offered by some online websites. As a result, the researchers pointed out that even from those major academic

68

medical centers' online portals, health information might be inaccurate and of low quality.

For information about mental illness related to children, not surprisingly, the Internet had been recognized as one of the primary sources (Bouche & Migeot, 2008; Sage et al., 2018), or even the most important resources in Tanzania (Tandi Lwoga & Florence Mosha, 2013). Famous resources for obtaining child-related mental health information include not only those general public health portals such as WHO, CDC, MedlinePlus, etc., but also those ones that specifically focus on children group, like Child Mind Institute.

2.7.2   *Teenager-related health information*

According to the WHO website and MedlinePlus portal, several health issues such as diabetes, development issues, and school health are shared among children and teenagers. However, it is clear that teenagers are facing with more mental illness related conditions since they are more connected to underage drinking, underage driving, sexual health, pregnancy, drug abuse, and depression.

Among various health consumer groups, adolescents and young adults are unique when discussing mental health because almost 50% of lifetime prevalence of DSM-IV disorders appear by age 14, with three quarters of symptoms occurring by age 24 (Kessler et al., 2005). Compared with other consumer groups, teenagers have long raised attention from previous studies (Kessler et al., 2005; Knitzer & Olson, 1982) since "adolescence and young adulthood" were regarded as critical timing for the appearance of mental health problems (Burns, Davenport, Durkin, Luscombe, & Hickie, 2010).

There are a lot of resources for teenagers and their mental issues. For instance, Society

69

for Adolescent Health and Medicine, Center for Young Women's Health and Young Men's Health, Go Ask Alice!, Girls Health.Gov, etc.

2.7.3   *Older adult-related health information*

According to the WHO website and MedlinePlus portal, older adults are normally linked to health conditions such as elder abuse, memory issues like Alzheimer's Disease, stroke, heart problems, and so on. Besides children and teenagers, older adult is another health consumer group that has been closely connected to mental health. For the "homebound" elderly population, which refers to those older adults who need to make substantial effort or require assistance due to an illness or injury to leave their home, their mental health conditions were of great importance to be taken care of (Qiu et al., 2010).

On the other side, prior studies had also shed light on the mental health conditions of some minority groups among older adults. For instance, some researchers reported that for lesbian, gay, and bisexual older adults, "better mental health was correlated with higher self-esteem, less loneliness, and lower internalized homophobia. (D'Augelli, Grossman, Hershberger, & O'Connell, 2001)" Besides that, D'Augelli et al. also pointed out that men were more likely to catch internalized homophobia, alcohol abuse, and suicidality related to their sexual orientation than women. Another instance of studies focusing on the minority groups among older adults in terms of their mental health conditions was conducted toward the transgender elderly (Fredriksen-Goldsen et al., 2014). In their research, Fredriksen-Goldsen et al. claimed that "transgender older adults were at significantly higher risk of poor physical health, disability, depressive symptomatology, and perceived stress compared with nontransgender participants".

70

Resources for older adult-related mental health information are usually available on these general public health portals like WHO and Medlineplus. However, there are also institutions like the National Council on Aging giving specific careness to elderly people.

2.8 Social network analysis

The concept "network" was described by some researchers as a "collection of nodes and links" (E. Kim, Hou, Han, & Himelboim, 2016, p. 480) where nodes "refer to social actors which could be individuals, organizations, content, physical or virtual locations, or events" (p. 480). And the links refer to the connections and relationships existing among the nodes (Hansen, Shneiderman, & Smith, 2011). There were two important elements in the construction of a social network mentioned in Kim et al.'s article - betweenness centrality and closeness centrality. The betweenness centrality of a node "measures the number of shortest paths between two nodes that pass through that user and how often the user lies on the shortest path between two other nodes" (p. 481) while the closeness centrality assesses the average distance between two nodes in a social network (Barash & Golder, 2011).

Within the field of social science, the theory of networks has been applied to a wide range of disciplines covering psychology to economics and regarding a great variety of social phenomena (Borgatti et al., 2009). Scholars like Borgatti realized the great potential possessed by social network theory as he described such method as providing "an answer to a question that has preoccupied social philosophy since the time of Plato, namely, the problem of social order: how autonomous individuals can combine to create enduring, functioning societies" (p.892). The original history of social network analysis could be traced back to 1930s, but social network analysis was first theorized as a new concept by Barnes (1954), with its roots lying in theories of

71

social action (Coleman, 1986). As a result, it had become an established field within the social science area by the 1980s (Borgatti et al., 2009).

Social network analysis has been defined as a methodology for exploring social structures through the application of network and graph theories (Otte & Rousseau, 2002). It provides a research framework to measure structural relationships between members within a network and intends to reveal reality occurring among the interactions and progress behind the scenes (Borgatti et al., 2009). The goal of social network analysis is to discover and measure structural relationships among entities and nodes within a given network (Zhang et al., 2016, 2015). The relationships resulting from interactions among actors could be generally concluded as information exchanged between them. Such connections might be hyperlinks in a webpage or friendships between two people, etc. Due to its capability of reflecting the relationships existing in the network in a strong, intuitive, and vivid visualized way, social network analysis method has been widely applied in many domains. Within them, for Web portals, Zhang et al. (2016) stated that social network analysis could be employed to "gauge and compare the connection network structure and semantic network structure of the subject directory" (p. 2168).

2.8.1   Basic characteristics of social network analysis

The primary components of social network analysis are actors and relations where actors are also called nodes or vertices and relations might also be named as ties or edges. According to previous literature, actors refer to individual entities that are not limited to people, but could also include items such as Web pages, keywords, etc. (Hansen et al., 2011), and relations refer to links in the concepts of a social network - that is, the linkage between the actors. Borgatti et al. (2009) classified four types of relations containing similarities, social relations, interactions, and flows.

72

Their influences among each other are the primary focus of social network research.

According to Zhang et al. (2015), the relationships and connections that arise from the application of social network analysis can be revealing. To be more specific, the relationships explored through social network analysis are regarding nodes and ties. As a result, given the nodes and ties being abstract or physical, studies could be employed in various ways.

For physical objects like Web pages and abstract objects like a group of keywords or documents, researchers might be able to focus on these nodes and conduct studies like analyzing user-generated content applied among various media (Salah, Manovich, Salah, & Chow, 2013). On the other hand, while exploring relationships, prior researchers have found that such research could be able to pay attention to some quite specific fields, such as a sport social network (Hambrick, 2012). Furthermore regarding research focusing on the relationships of social network analysis, some scholars also pointed out the importance to differentiate strong ties from weak ties, as well as identify various kinds of connections within the network (Hoppe & Reinelt, 2010). In addition, Zhao et al. (2012) revealed the fact that when social network analysis is coordinated with community associations, not only the structural connection, but also semantic associations would be figured out.

2.8.2   *Application of social network analysis in Information Studies and Health Informatics*

According to Otte and Rousseau (2002), social network analysis has been employed to explore many fields within the information science area. Instances include subject classification, citation network, bibliometrics, etc. Within the setting of healthcare, prior scholars have attempted to figure out kinds of relationships existing between health care coworkers and finally

identified several thematic ties like social influence, describing actions, motivations for spread, etc., thus helping to improve the effectiveness of the health decision making process (Chambers, Wilson, Thompson, & Harden, 2012). Moreover, health care settings were considered to be possibly understood better if social network analysis was combined with public health communication methods (Luke & Harris, 2007) after Luke and Harris studied the history of the utilization of such coordinated research means toward the transmission in HIV and other sexually transmitted diseases. However, the application of social network analysis has still been considered insufficient in health care settings overall by Chambers et al. Previous exploration in health informatics also included discovering information exchanges among health consumers (Mertens, Saint-Charles, & Mergler, 2012), the impact left by culture on consumers' health information searching (Smith & Christakis, 2008), the feedback of usage from adolescents toward health information (N. J. Gray, Klein, Noyce, Sesselberg, & Cantrill, 2005), and hospital nursing (Pow, Gayen, Elliott, & Raeside, 2012).

Social network analysis was also applied in a recent research conducted by Kim et al. (2016) with content analysis of the breast cancer related health information on Twitter. In their article, Kim et al. selected the action of "retweet" and regarded it as the symbol of receiving health information through one's social network activities. After that, the scholars divided the influencing factors into two parts: characteristics of social network and content. The former one refers to social network relationships such as the number of followers, betweenness centrality, and closeness centrality; while the latter refers to the words utilized when posting a tweet containing health information about breast cancer. Their findings indicated that a user who possesses a higher number of followers, a greater level of personal influence, and closer

74

relationship and similarity with other users would get more retweets. Tweets containing positive feeling could also have more retweets compared with those including negative words. All the impacts were statistically significant. This study revealed the potential for other researchers focusing on the health information field to integrate social network analysis with other proper research methods, especially when the research target shares common characteristics with social network analysis and other specific approaches. Kim et al.'s application of social network analysis regarding social network objects was a great practice.

On the other hand, some researchers have claimed an interesting idea of employing social network analysis to the evaluation of the topic/subject based navigation systems possessed by public health portals like World Health Organization (Zhang et al., 2015) and the government agriculture portal (Zhang et al., 2016). In their studies, the scholars collected the link network relationship, the semantic network relationship, and an adjusted semantic network relationship through social network analysis and identified the differences existing among various relationships via the utilization of clustering analysis and inferential statistics. Their findings suggested that some subjects connected with hyperlinks might not have significant semantic relationship while some other topics sharing a significant number of keywords did not have the structural link relationship built. The study results were echoed by health expert as well. Therefore, recommendations for future optimization have been proposed.

Social network analysis has been widely utilized in both information studies and health informatics areas in recent years. It is extremely strong in exploring network relationships among a variety of actors and can be quite simply integrated if the research involves social network-related objects such as a social network website.

2.9 Summary

This chapter has reviewed previous research regarding public online portals, navigation systems and subject directories, health information seeking online among health consumers, health information related to children, teenagers, and older adults, as well as social network analysis with its applications in social science area. During the past decades, with the rapid growth of Internet technology, the trend of seeking health information online has been observed by a large group of scholars. As a result, recent years have witnessed a burst of research studying how information is organized and made accessible to health consumers, how health consumers are attempting to obtain health-related information in the online environment, and how social network analysis could be applied into information fields.

Compared with online information sources primarily consisting of user-generated content such as social network sites, public health portals are considered to contain the most reliable and up-to-date health-related information. Unlike social media sources through which health consumers seek health information based on the motivation of social and emotional support, the major motivation for utilizing public health portals to get access to health information is acquiring knowledge. Therefore, the effectiveness of organized information and the ease of retrieving targeted information in public health portals have raised attention among researchers.

Social network analysis has been widely applied in the social sciences for identifying key individuals as well as relationships among participants within specific networks. Besides typical networks formed by human beings, researchers have also recognized the potential for employing social network analysis in the information area, for instance, topics within a subject directory. Such creative applications have proposed a new aspect for examining and organizing information

in an information system.

The previous literature has shed more light on social network sites and search-based navigation means. Prior studies focused more on health consumers actively seeking specific health information or emotional support while the importance, benefits, and potential development of organized health information for general browsing purposes provided by health professional institutions remained not fully researched.

In addition, mental health related subject directory systems have seldom been addressed in previous research. Investigating the mental health section on a representative health Web portal's subject directory might provide opportunities for improving the effectiveness of the directory system, and offer a new perspective for both health consumers and professionals to understand relevant health information.

Chapter 3. Research Methodology

3.1 Introduction

The aim of this study was to investigate and evaluate the subject directory systems for mental health, children, teenagers, and older adults used in the MedlinePlus portal. It is an empirical study. The research objects are the health topics utilized as subject directories by the portal creators. Those health topics are connected in two ways: structural and semantic.

In this study, the research population was all health topics used in the subject directory system in MedlinePlus. Considering the fact that MedlinePlus has 44 subcategories and over 1000 health topics subject directories, a sampling strategy was conducted to select several subcategories and health topics from the whole population. The examined health topics were selected from four subcategories: Mental Health and Behavior under the Disorders and Conditions section, and Children, Teenagers, and Older Adults under the Demographic Groups section. All selected subcategories as well as their related health topics formed the sample.

Research data were collected from health topics of each sampled subcategory. To be more specific, since each health topic has an introductory Web page on MedlinePlus, two types of data were gathered through those Web pages of each selected health topic: the first data type was the structural connection which was reflected by the information displayed under the "related health topics" column; the other data type was the semantic connection information which was hidden behind all the textual content of the page.

This research study applied a mixed quantitative research method combining social

network analysis method, descriptive statistical analysis method, and inferential statistical analysis method to analyze the data collected from health topic pages about the four selected subcategories on the MedlinePlus portal. The reason for applying a mixed research method in this study was because when considering the subject directory system of the MedlinePlus portal as a social network, social network analysis method can be extremely effective in exploring network relationships among a variety of actors and can be simply integrated if the research involves social network-related objects. The outcome of social network analysis, visualized social networks, can be generated to identify various characteristics as well as patterns among different health topics. Along with that, descriptive and inferential statistical methods can be applied to evaluate various networks. To build the social network inherent in the subject directory, all the involved health topics served as the nodes in the network while their structural connections served as the relationships in the network. All social network data were imported to UciNet software to process necessary social network analysis such as similarities, node features, etc., so that key nodes, characteristics of nodes and relationships, and underestimated structural connections were identified within the subject directory system of MedlinePlus.

Inferential statistical analysis methods played a vital role in this study. It was utilized on a group of social network features generated from the social network analysis. Those features included similarity value, degree centrality, betweenness centrality, closeness centrality, etc. Inferential tests were also applied to evaluate the optimization suggestions proposed at a later stage. Applying inferential analysis methods helped in uncovering the consistency of results and relationship patterns among health topics related to mental health, children, teenagers, and older adults. For instance, the Mann-Whitney test was employed to assess the potential differences

79

existing between the original and optimized structural link network regarding the similarity values of the involved health topics.

The social network analysis was processed using UciNet (Version No. 6.669) and the inferential statistical analysis was processed using SPSS (Version No. 25.0).

The data collection and analysis methods utilized in this study are displayed in Figure 7. Detailed information regarding data collection and analysis methods are introduced and discussed in the following sections.
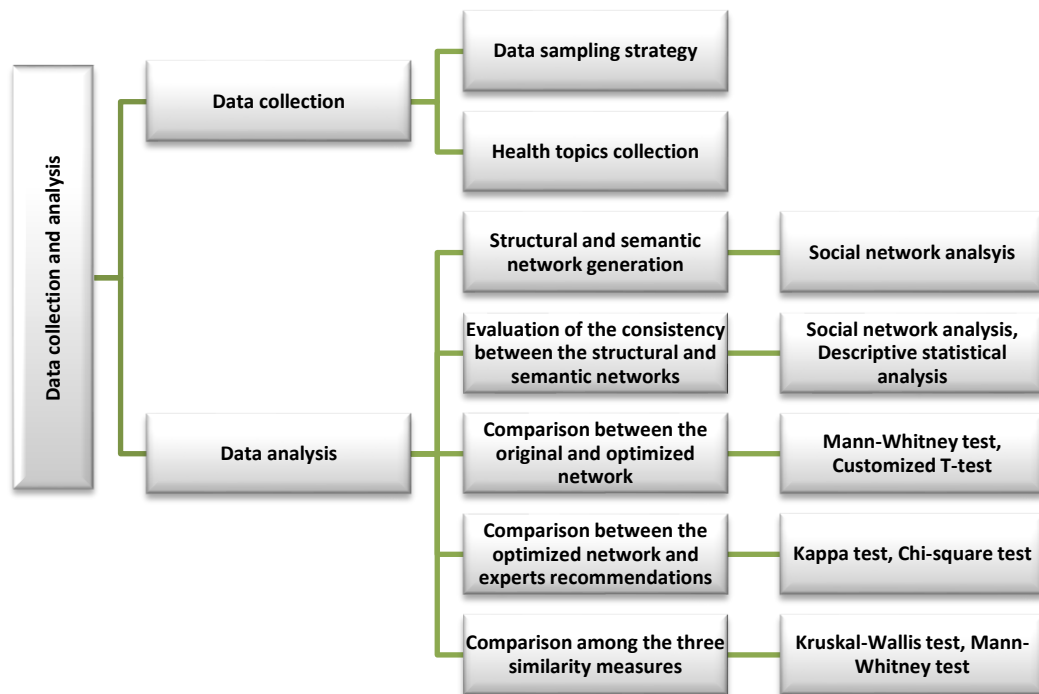
Figure 7. Data collection & analysis methodologies

3.2 Data collection

Three steps were followed to collect the research data. The first step was to determine the

health topics that could serve as the starting point for each subcategory group. It was decided the initial numbers as well as specific related health topics to start to construct the structural link network. The second step was to expand the health topic groups into a reasonable size through involving more health topics listed under the "related health topics" column from the previously selected health topics' Web pages. A proper group size enabled this study to generate a meaningful and strong social network so that in-depth explorations could be applied to an appropriate range of the subject directory system. The last step was to collect data related to each selected health topic through its individual Web page. The data gathered contained two parts: all the health topics listed as the "related health topics" were obtained as structural connection data and all the textual information in the introductory section was collected as semantic connection data.

Data collection also involved preparation work related to the expert evaluation at the final stage of the research. This preparation work involved two activities: a recruitment process of health field experts that was placed in the sampling strategy, and a dataset generation process for evaluation in the data collection plan.

3.2.1 *Sampling strategy*

The sampling strategy included the first and second steps mentioned above. For these two steps, the general process was that once a specific health topic was selected to serve as the starting point, this topic and its related topics were then collected to form the first level of health topics as Level 1. Subsequently, related health topics were collected from each of the health topic at Level 1 to form the second level as Level 2. Such data collection process was repeated until a third or fourth level was reached. An ideal health topic group should contain between 100 and

81

150 topics to generate a corresponding social network so that the network could possess enough information for later analysis. To ensure a proper network size and amount of selected health topics, two conditions were met during the first and second step:

1) The starting health topic was able to represent the corresponding subcategory group. In addition, it was preferred that this starting health topic be a general term like *Mental Disorders* rather than a specific disease or health condition term such as *Obsessive-compulsive Disorder*. The reason was that the MedlinePlus health topic-based subject directory system follows a general hierarchical structure. This means that for general terms like *Mental Disorders* and *Children's Health*, MedlinePlus tends to include many related general terms as well as specific health condition terms so that users can be directed from general health information to specific health information. However, for those health topics of specific health conditions, MedlinePlus has set only a few or no related health topics on their Web pages, thus often leading those pages to become "dead ends". Therefore, choosing a general health topic term ensured a proper number of involved health topics in Level 1.

2) The number of selected health topics within each subcategory group should reach 100 to 150 within four levels. Since the purpose of this study was to evaluate the subject directory system of the MedlinePlus portal, a preliminary investigation was performed in September 2019 to investigate the network size among a variety of health topic groups before the final four topic groups – mental health, children, teenagers, and older adults - were identified and selected.

Section	Subcategory	Starting health topic	Network size
Disorders and	Mental Health and Behavior	*Mental Disorders*	99 (Level 4)

www.manaraa.com

Conditions	Diabetes Mellitus	*Diabetes*	92 (Level 3)
Body Location/Systems	Brian and Nerves	*Brain Diseases*	109 (Level 3)
	Digestive System	*Liver Disease*	17 (Level 3)
	Lungs and Breathing	*Lung Diseases*	91 (Level 3)
	Kidneys and Urinary System	*Kidney Diseases*	75 (Level 5)
	Blood, Heart and Circulation	*Heart Diseases*	144 (Level 4)
Demographic Groups	Children and Teenagers	*Teen Health*	140 (Level 4)
	Children and Teenagers	*Children's Health*	159 (Level 3)
	Older Adults	*Older Adult Health*	150 (Level 4)

Table 2. Preliminary investigation results among a variety of health topic groups

The health topic groups were under 10 subcategories in three different sections of MedlinePlus's subject directory – *Body Location/Systems*, *Disorders and Conditions*, *and Demographic Groups*. Each health topic group contained a health topic serving as the starting point and the corresponding network size numbers were calculated and are displayed in Table 2 shown above. It is clear to find that those health topic groups with *Mental Disorders*, *Heart Diseases*, *Teen Health*, and *Older Adult Health* serving as the starting point fit best into the two conditions required by this study. The rest of the groups were excluded either because they had reached a too large network size within three levels (e.g., *Diabetes* and *Brain Diseases*) or they were not able to include enough health topics within four levels (e.g., *Liver Diseases*).

Another important finding examined during the preliminary investigation of the sampling process was that the health topics under the *Body Location/Systems* sections had been closely

83

connected in the original structural link network through the term "transplantation". Since the majority of the subcategories as well as their related health topics had involved an organ or body part term, the corresponding transplantation information regarding that specific organ or body part were found to appear under the "related health topics" columns for nearly every organ/body part health topic (Figure 8).
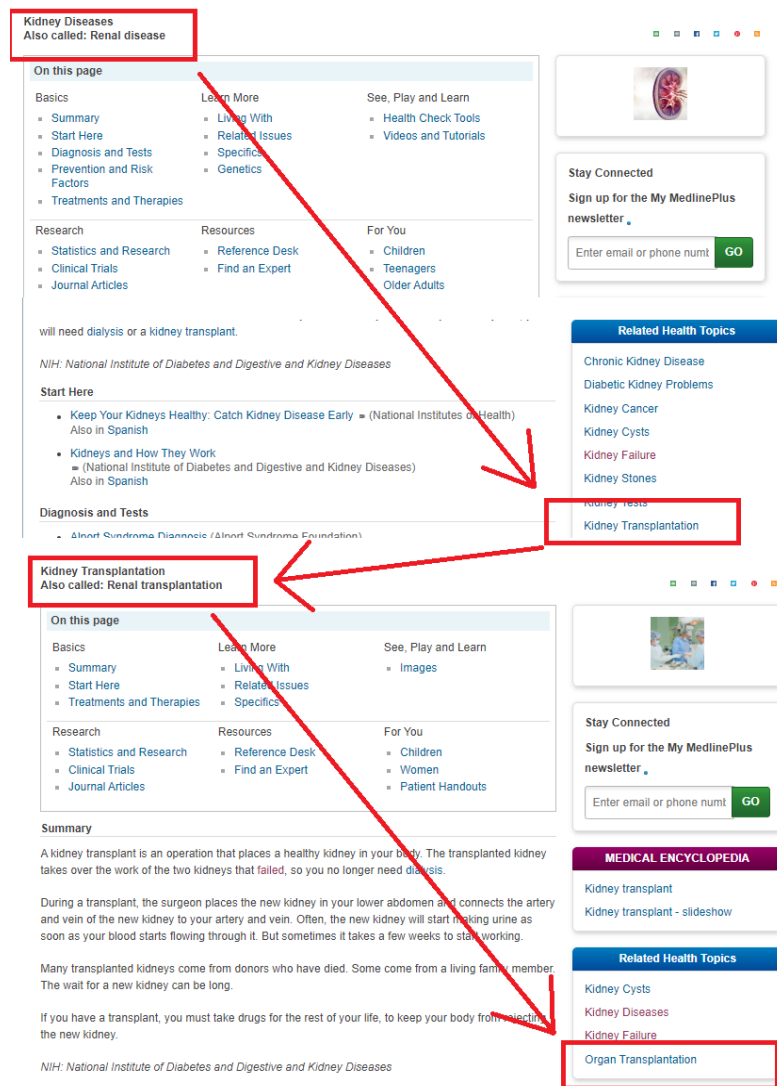
Figure 8. Transplantation related health topics linked from a specific organ's page

Following that, the page of a single organ's transplantation information would navigate to a health topic titled *Organ Transplantation*, whose Web page included transplantation related terms for other organs (Figure 9). These structural connections could be identified before Level 3 or 4 were reached. This indicated that the health topics under the *Body Location/Systems* sections had already been linked through the bridging topic *Organ Transplantation*. Such connections built by the portal creators may greatly impact the outcomes generated through social network analysis. Therefore, health topics under this section were excluded from this study.

Figure 9. Transplantation related health topics linked from the bridging topic *Organ Transplantation* to other specific organ's pages

Compared with those health topics located under other sections such as *Diagnosis and*

Therapy, and *Health and Wellness*, health topics under *Demographic Groups* have unique strengths. One strength is that the subcategories as well as their related health topics are closely linked among each other. For instance, development related issues have been commonly shared between children and teenagers. Also, these subcategories as well as their related health topics can be simply distinguished due to the fact that various health consumer groups are linked to different health conditions. Therefore, each demographic group has its specific concerned health conditions so that the relationships among health topics between various health consumer groups and their corresponding health conditions can be differentiated between each other. This provides an ideal condition for processing social network analysis through network features. This also explains why children was still selected as a subject in this study despite the fact that it had reached 159 health topics at level 3. This health consumer group not only has strong linkages toward the other two health consumer groups, but also connects closely to various types of mental disorders. Hence, the children related health topics were included for comparing purposes.

For the evaluation work that was performed after the data analysis process, two experts were invited to evaluate the study's results. Before the recruitment process started, an IRB approval was applied to ensure the following processes. The recruitment process began with a brief scan on qualified professionals – these professionals were to have gone through the preliminary examination of a M.D. program or hold at least a master's degree in a medical or health related field as well as having at least five years of experience in the field. An invitation list was generated and a recruitment email was sent out to each expert on the list with a short introduction to the research. For those experts that were willing to take part in the study, a

detailed information sheet and a consent form were sent and a signed confirmation was collected from them. The participating experts were able to know that they were free to drop out of the study at any point in the research.

3.2.2   Health topic data collection process

The data collection process presented the third step discussed above. The research data were from health topics' Web pages on MedlinePlus. Four subcategories identified after the sampling process – Mental Health and Behavior, Children, Teenagers, as well as Older Adults, became the data sources. All the information to be collected for this study are publicly accessible online. Although MedlinePlus has been observed to make minor adjustments to its subject directories regularly, no time patterns were found to exist among health topics. In other words, a health topic page created earlier would not lead it to be different from a health topic page created later in terms of their structural connections with other topics or page organization.

The data collection process was performed from two perspectives – structural and semantic. The structural data were gathered through the collection of information listed under the "related health topics" column of each topic. The general procedures have been described above.

Going forward, Web pages of all the involved three/four levels of health topics were gathered and the text on these Web pages was extracted through a coded Python program to form a word list. Textual information from the overview table, as well as side menu text and navigating hyperlinks was ignored because this textual information made no difference among various health topics' pages – this textual information such as "basics", "learn more", and "other language" has been standardized by the portal creators to keep a consistent format among

87

different Web pages of health topics. Hence, only words appearing as plain text from "basics" to "for you" were gathered. The automatically collected textual information was further filtered through manual review. Two types of content were removed: 1) some association names such as American Psychiatric Association and National Institute of Mental Health; 2) other data source information such as Medical Encyclopedia and MedlinePlus Health Topic. These two types of textual information were removed because MedlinePlus tends to add data source information after every word term included in a health topic's page. Hence, association names and other data sources were found to occur multiple times on every health topic's page. That textual information weighed greatly in the semantic dataset and might skew the analysis result. Hence, it was removed through manual review. The rest of the collected textual information was used in the semantic data in this study. This word list was then further cleansed. First, a stop-word list was applied to filter the word list to remove useless words. These stop-words mainly included those which only function from the grammar aspect, such as "a", "an", "the", "with", "of", "to" etc. Second, synonyms were combined; for instance, anorexia nervosa, binge eating, and bulimia were combined into eating disorders. Third, all the words on the word list were kept as their regular form. Different forms were normalized. For example, "psychotherapy" was converted to "psychotherapies", and "achieves" was converted to "achieve" etc.

3.2.3   *Collection of the data for expert evaluation*

Four lists of health topics were selected from the dataset collected in this study to serve as the evaluation dataset. Each of these lists contained 100 to 200 paired health topics based on the optimization results identified for the corresponding health topic group. Among them, about two thirds of the paired health topics were health topics that were suggested to add to the structural

network in the subject directory of MedlinePlus. These recommended health topics were identified through the analysis results of the cosine similarity measure. To be more specific, these suggested health topics were formed according to their cosine similarity measurement or other node feature test result value – the value was determined to be larger than the set threshold, for instance, the average number of all structurally connected health topics. Besides those recommended topics, the remaining one third of the paired health topics were also evaluated. Those health topics were not linked on the structural link network in the MedlinePlus portal and were found to have low semantic connections according to the similarity outcomes generated through the cosine similarity measure. In other words, these connections were collected from those topics that were possessing weak semantic relationships and they served as distracters in the evaluation dataset. The two groups of health topics were mixed in the evaluation dataset and evaluated together for comparison purposes. The original evaluation sheets for the four health topic subcategories are attached in Appendices 1-4.

3.3 Ethical consideration

Ethical issues should be taken into consideration when any human subject is included in research. It is vital to protect participants' rights during the whole research process. In this study, two domain experts from health or medicine field were invited to evaluate the optimization suggestions proposed by this study toward the mental health, children, teenagers, and older adults related subject directory applied by MedlinePlus. These domain experts were the human subjects involved in this study.

An application was sent to the Institutional Review Board (IRB) of the university and was approved. A consent form was sent to the participants with clear clarification regarding the

research objectives, objects, and procedures. Participants had to read and sign the consent form before they participated in the evaluation process. They were informed that they were able to withdraw from the research at any point if they wanted. Moreover, their evaluation responses were de-identified and safely stored to ensure anonymity and confidentiality.

3.4 Social network analysis

As discussed above, social networks representing the subject directory of mental health, children, teenagers, and older adults related health topics on MedlinePlus were built based on the data collected from the selected health topics' Web pages. Data collected from MedlinePlus were grouped into two types: nodes and edges, and they formed various networks.

3.4.1   Nodes

Nodes in a social network refer to actors or participants. In this study, each health topic selected from the subject directory of MedlinePlus served as a node in the social network analysis. Figure 10 displays an example of how mental health related health topics are presented by nodes in a network.
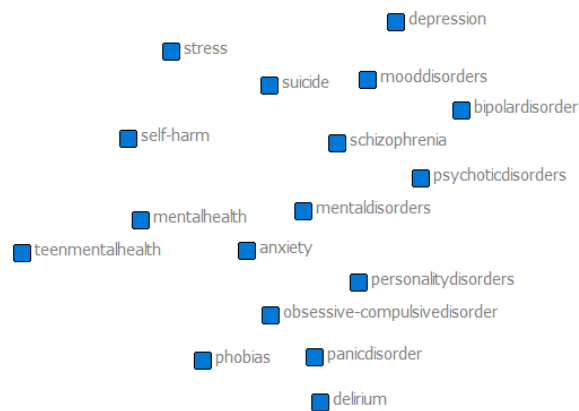
Figure 10. Nodes of a group of mental health related topics on MedlinePlus

3.4.2   Edges

Edges represent the relationships and connections among nodes in a social network. In this study, the structural connections (i.e., the hyperlink information set under the "related health topics" column) among health topics related to mental health, children, teenagers, and older adults in the subject directory of MedlinePlus served as the edges in the social network analysis. The semantic connections were not directly reflected through the edges in the network. Instead, they served as the fundamental bases when creating the structural links. Semantic connections were also applied as a baseline when evaluating the structural links. Figure 11 shows an example of edges among the same health topics related to mental health from the subject directory of MedlinePlus displayed in Figure 10.
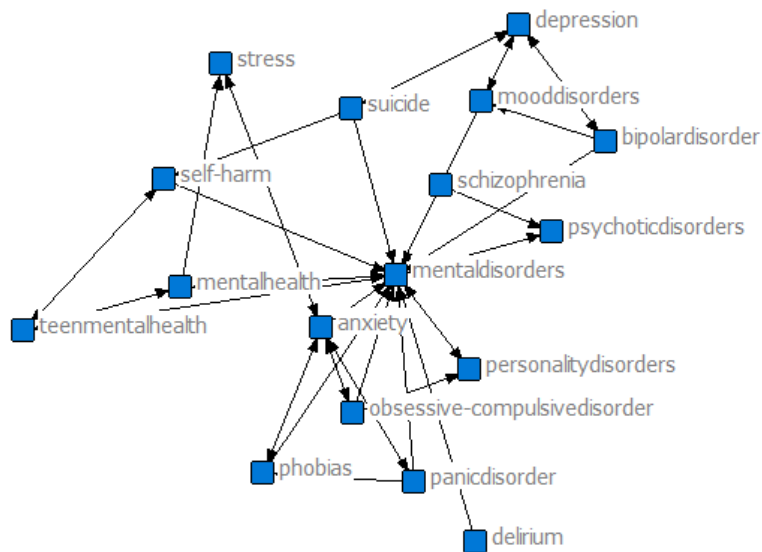
Figure 11. Edges among a group of mental health related topics on MedlinePlus

91

In Figure 11, an interesting fact worth noticing is that some of the edges between two health topics are unidirectional, such as the edge between *Mental Disorders* and *Delirium*, while others are bidirectional, such as the edge between *Anxiety* and *Stress*. Such unidirectional edges indicate the fact that those health topics have already been linked from one direction on the structural link network. For instance, users can be directed to the page of *Delirium* through the hyperlink provided in the "related health topics" column from the page of *Mental Disorders*. However, when users have reached *Delirium'*s page, they are not able to find the hyperlink to navigate to *Mental Disorders*. As a result, users would not be able to return to the previous topic. These unidirectional edges seem to indicate a hierarchy relationship that goes from general to specific or vice versa, such as *Mood Disorder* and *Bipolar Disorder*. Moreover, the standards according to which a unidirectional or bidirectional edge should be set between two health topics are not clear. For a similar hierarchical relationship shared between *Mood Disorder* and *Bipolar Disorder*, and between *Mood Disorder* and *Depression*, the former pair of health topics has been connected through a unidirectional edge while the latter has been linked with a bidirectional edge. In addition, it is reasonable to assume that other unidirectional edges may also represent other types of relationships such as relevance among other health topic groups.

Therefore, during the following procedures in the later data analysis phase in this study, it was important to differentiate unidirectional edges from those bidirectional ones. Patterns, characteristics, standards, and other related findings were processed and generated regarding both types of the edges.

3.4.3 Networks

The networks reflecting the mental health, children, teenagers, and older adults related

part of the subject directory applied by MedlinePlus were constructed based on the nodes and edges identified in the previous stage. Through the creation of the networks, key health topics as well as patterns and characteristics of the relationships among the selected health topics were explored and discussed.

Figure 11 could be regarded as a small network about mental health related health topics. On the MedlinePlus portal, health topics are linked through structural connections by the portal creators to indicate relationships existing among various health topics as well as provide additional related health information to the portal users to fulfill different health information seeking motivations. The structural connections are reflected by the hyperlinks in the "related health topics" column, however, those connection are not exactly the same in terms of the strength of the connections shared between different pairs of health topics. To further discover and evaluate the health topics and their relationships in the subject directory networks, involved nodes and edges were investigated through comparison of similarity measures, node features, and network features.

3.4.4 *Social network matrices*

After the data collection process, a group of matrices were built to represent the structural link network and the semantic network among the selected health topics related to mental health, children, teenagers, and older adults. It was of great importance to generate these matrices because they serve as the fundamental support for social network analysis.

The first matrix is a topic-topic link matrix (TTLM). It refers to the structural relationship among the collected health topics. In this matrix, t_{ij} refers to whether a health topic i is listed as a

93

"related health topic" of another health topic j, n represents the number of health topics selected from the subject directory and the matrix itself is a $n \times n$ asymmetrical matrix. The cell value of t_{ij} within the matrix indicates the relationship between two health topics – if topic i is listed as one of the related health topics of topic j, the cell value t_{ij} is assigned as 1, otherwise it is assigned as 0. Here, the matrix is asymmetrical because that topic i is a related health topic of topic j does not ensure that topic j will be a related health topic of topic i (Equation (1)).

$$\text{TTLM} = \begin{pmatrix} t_{11} & \cdots & & t_{1n} \\ & \cdots & t_{ij} & \\ & \cdots & & \\ t_{n1} & \cdots & & t_{nn} \end{pmatrix} \tag{1}$$

An equation that can be generated for the TTLM is displayed below. It indicates that a health topic cannot include itself as a related health topic (Equation (2)).

$$t_{ii} = 0, \qquad 1 \le i \le n \tag{2}$$

The second matrix created is a topic-keyword matrix (TKWM). In this matrix, each row refers to a health topic while each column refers to a keyword from the word list generated previously. The cell values, unlike those in the TTLM, represent the degree to which keyword j relates to topic i. This degree is determined by the weight, i.e., the term frequency (tf), of the keyword in the corresponding topic's Web page. Equation (3) is built for the TKWM and is shown below. h_{ij} is the frequency number of a keyword i that appears in a health topic j's Web page. n is the number of the selected health topics while m is the number of the keywords contained in the word list that are extracted from the health topic's Web pages.

$$\text{TKWM} = \begin{pmatrix} h_{11} & \dots & & h_{1m} \\ & \dots & h_{ij} & \\ & \dots & & \\ h_{n1} & \dots & & h_{nm} \end{pmatrix} \tag{3}$$

When calculating the term frequency in the topic-keyword matrix, a cutoff point of 2 was applied to further filter the dataset. For those words holding a cell value of 1 in the matrix, it indicates that they appear only in a single health topic's Web page. That means they are not able to provide sufficient semantic connection between two or more health topics. Hence, a word should at least have a cell value that is equal to or larger than 2 to be kept in the matrix. Furthermore, among the remaining words, some were eliminated if they only occur in a single health topic's Web page no matter how large the cell value is since there will be no semantic connections from other topics.

Following the creation of TTLM and TKWM, a topic-semantic matrix (TSM) was built to represent the semantic network among the selected health topics through similarity measures based on the term frequency data in the topic-keyword matrix (TKWM). The topic-semantic matrix is displayed in Equation (4). In this Equation, s_{ij} refers to the similarity value between two health topics i and j and n refers to the number of selected health topics. For similarity measure, the cosine-similarity measure was used as the major similarity measure in this study along with the Pearson correlation similarity measure and the Euclidean distance similarity measure for comparing purposes. the Pearson correlation similarity measure "can be used when quantities (i.e. scores) varies" and when "objects that describe the same data but use different values" Segaran (2007). The Euclidean distance similarity measure always serve as the "basis of many measures of similarity and dissimilarity" (Borgatti, 2013). Therefore, these two similarity

measures were applied as supplemental methods to compare and verify the results generated

according to the cosine similarity measure. The cosine-similarity is displayed in Equation (5).

This Equation aims to find how similar two health topics' Web pages are based on their textual

information. In this Equation, h_{ij} represents the frequency number i of a keyword j while n refers

to the number of selected health topics. Each cell value s_{ij} in Equation (4) is calculated based on

Equation (5). All the s_{ij} values will form the TSM. The cell value is between 0 to 1, in which 0

indicates that there is no similarity between two health topics' Web page content while 1

indicates that the two Web pages are the same. The cosine-similarity measure has been widely

applied to the information retrieval area (Baeza-Yates & Ribeiro-Neto, 1999; Zhang &

Rasmussen, 2001) due to the fact that for documents containing the same distributed or

proportional weighted keywords, the cosine-similarity measure can be an effective tool in

identifying their similarities (Korfhage, 1997). Equations for the Pearson correlation similarity

measure and the Euclidean distance similarity measure are displayed in Equations (6) and (7). In

Equation (6), μ refers to the average value.

$$\text{TSM} = \begin{pmatrix} s_{11} & \cdots & & s_{1n} \\ & \cdots & s_{ij} & \\ & \cdots & & \\ s_{n1} & \cdots & & s_{nn} \end{pmatrix} \tag{4}$$

$$CosineSimilarity: s_{ij} = \frac{\sum_{k=1}^{n} h_{ik} \times h_{jk}}{\left(\sum_{k=1}^{n} h_{ik}^2 \times \sum_{k=1}^{n} h_{jk}^2\right)^{\frac{1}{2}}} \tag{5}$$

$$PearsonCorrelation: s_{ij} = \frac{\sum_{k=1}^{n}(h_{ik} - \mu_{hi}) \times (h_{jk} - \mu_{hj})}{\sqrt{\sum_{k=1}^{n}(h_{ik} - \mu_{hi})^2} \sqrt{\sum_{k=1}^{n}(h_{jk} - \mu_{hj})^2}} \tag{6}$$

96

$$EuclideanDistance: s_{ij} = \sqrt{\sum_{k=1}^{n} (h_{ik} - h_{jk})^2} \qquad (7)$$

Based on Equation (5), the cell value (similarity) between topic i and topic j should stay the same as that between topic j and topic i. That means the TSM is a symmetrical matrix. Moreover, the TSM is not directional since two topics share the same similarity in both directions from the semantic aspect.

In the last stage, in order to compare the node features between the structural and semantic connections, the last matrix was generated to filter the cell value in TSM. This matrix is derived from TSM, hence it is titled as Cleaned Topic Semantic Matrix (CTSM). The CTSM is shown in Equation (8). In this Equation, every s_{ij} in TSM is checked and filtered based on a threshold. This threshold should be generated from the similarity measure outcomes - for instance, the average similarity value among the structurally connected health topics of all selected topics. If s_{ij} is equal or larger than the threshold, c_{ij} is assigned to 1; otherwise, c_{ij} is assigned to 0. All c_{ij} compose the CTSM. In other words, CTSM is another structural link network generated solely according to the semantic connection among health topics. After reflecting all those identified strong semantic connections in the CTSM, the original structural connections set by the portal creators were also added back to the matrix. These original structural connections are expected to set up a foundation for the later comparison between the original structural network and the optimized structural network. Containing both the original structural connections and the newly suggested connections have led CTSM to become a combined network. In conclusion, TSM and CTSM both play a vital role in representing the

97

semantic relationships among the involved health topics. They shall serve as the comparing baseline for later analysis.

$$\text{CTSM} = \begin{pmatrix} c_{11} & \cdots & & c_{1n} \\ & \cdots & c_{ij} & \\ & \cdots & & \\ c_{n1} & \cdots & & c_{nn} \end{pmatrix} \tag{8}$$

3.4.5 Network measurements

As introduced in the previous chapter, a group of network measurements were applied in this study to reveal the facts toward the subject directory system on the MedlinePlus portal. To be more specific, three network measurements were utilized: 1) the similarity measurements served as the fundamental support for investigating the consistency between the structural and semantic connections among the selected health topics; 2) node-level measurements were of great help when attempting to investigate the consistency between the structural and semantic connections among the selected health topics according to the discovery of key health topics and their characteristics; 3) network-level measurements assisted the exploration of the overall conditions regarding the four health topic groups through the findings toward the network characteristics. A summary of network measurements employed in this study along with their corresponding research questions and hypotheses, is generated and displayed in Table 3.

Level of measurement	Measurement	Research question	Hypothesis
Similarity	Similarity	RQ1.1, RQ2.1, RQ3.1, RQ4.1, RQ5.1	Hypothesis group 1, 3, 5, 7
Node-level	Degree centrality	RQ1.2, RQ2.2, RQ3.2, RQ4.2, RQ5.2	Hypothesis group 2, 4, 6, 8

98

Node-level	Betweenness centrality	RQ1.2, RQ2.2, RQ3.2, RQ4.2, RQ5.2	Hypothesis group 2, 4, 6, 8
Node-level	Closeness centrality	RQ1.2, RQ2.2, RQ3.2, RQ4.2, RQ5.2	Hypothesis group 2, 4, 6, 8

Table 3. Network measurements and corresponding RQs & hypotheses

3.4.5.1   Optimization

After the similarities among all the health topics on the semantic network were calculated, each edge on the semantic network had a similarity value. All edges were partitioned into one of 3 edge sets: Edge Set A, Edge Set B, and Edge Set C. Edge Set A contains the edges in which a topic is linked to itself. The similarity value between a topic and itself is always equal to 1. Since these edges make no contribution to later analysis, they were excluded. If an edge could find that its corresponding link exists on the structural link network, it was put in Edge Set B. If an edge could not find that its corresponding link exists on the structural link network, it was put in Edge Set C.

The average similarity among those edges in Edge Set B was calculated and it was used as the threshold to choose recommended topic edges for the structural link network. This process is called the optimization of the structural link network. If the similarity value of an edge in Edge Set C is larger than the threshold, the corresponding edge/link is recommended to add to the structural link network. Moreover, among the recommended edges in Edge Set C, there are two scenarios: 1) there is not any connection between two topics, T1 and T2, on the structural link network; 2) there is only a single connection from topic T1 (T2) to topic T2 (T1) on the structural link network. Both cases were considered. In conclusion, the similarity value of an edge in Edge Set C being larger than the threshold indicates that the recommended edge did not

99

exist on the structural link network, but its similarity is larger than the average similarity value of the edges on the structural link network. Hence, the recommended edge should be added to the structural link network.

As a result, the recommended edges formed a new set Edge Set D. This Edge Set D is a subset of Edge Set C. It is clear that after the process the optimized or finalized structural link network consists of both Edge Set B and Edge Set D.

In this study, the average similarity of the edges on the original structural link network was used as the threshold to determine or control the recommended edge set to add to the original structural link network. The rationale is that the similarities of the recommended links should have more close semantic relationships than the average links on the original structural link network. The threshold is adjustable. As the threshold increases, more edges are filtered out and the number of the recommended edges decreases, and vice versa. Notice that as the threshold increases, less relevant edges are removed from the recommended edge set. As a result, the optimization results get better. On the other hand, if the size of the recommended edge set is too small, reducing the threshold can result in an increase of the recommended edge set.

3.4.5.2   Node-level measurements

According to the Table 3, node-level measurements were applied to answer a series of research questions. These node features were employed to explore the differences between the structural and semantic connections regarding nodes that are playing more important roles than others in terms of connecting other topics contained in the networks. Among the node features, degree centrality was used to investigate how a specific node is linked with other nodes within a

network. In a directional network such as TTLM, the degree centrality (D) can be divided into in-degree and out-degree so that the inbound and outbound edges a given node is possessing can be measured. To be more specific, in-degree of a node refers to the inbound connections of the node and can be generated according to Equation (9) while out-degree of a node represents the outbound connections of the node and can be calculated through Equation (10). In a bidirectional network, degree centrality can be defined and measured based on Equation (11). In conclusion, a high degree centrality value refers to a high possibility of sending and receiving information flow within the network, hence can indicate the impact and importance a node possesses over other nodes in the network (Abraham, Hassanien, & Snásel, 2010). Here *n* refers to the number of nodes in the network. *i* and *j* refer to two node numbers within the network

$$D_{in}(\text{i}) = \frac{\sum_j t(i,j)}{n-1} \tag{9}$$

$$D_{out}(\text{j}) = \frac{\sum_i t(i,j)}{n-1} \tag{10}$$

$$D(\text{i}) = \frac{\sum_j c(i,j)}{n-1} \tag{11}$$

Betweenness centrality (B) refers to the times that a node serves as the bridge between two other nodes through the shortest path in the network. It reflects the capability of control a node possesses regarding the information flow in a network. According to Abraham et al.'s book (2010), a node with high betweenness centrality indicates that the node is a "broker" in the network. In this study, a health topic with high betweenness centrality reveals that the health topic is more likely to "control" the structural connections within the group's subject directory system. Take TTLM as an example, the betweenness centrality can be defined and generated

101

through Equation (12). Here n refers to the number of nodes in the network; $i, j,$ and k refer to various node numbers within the network. σ_{jk} refers to the total number of shortest paths from node j to node k and $\sigma_{jk}(i)$ indicates the total number of shortest paths from node j to node k through node i.

$$B(i) = \sum_{i \neq j \neq k} \frac{\sigma_{jk}(i)}{\sigma_{jk}} \tag{12}$$

Closeness centrality (C) refers to the combined distance a node has when relating to every other node within the network. It is normally used to measure how close a node is to other nodes in the network through shortest paths (Abraham et al., 2010). As a result, a node sitting in the middle of a network may possess a higher closeness centrality compared with another node sitting at the corner in a network where the majority of nodes are gathering toward the center of that network. In a directional network such as TTLM, the closeness centrality can be divided into in-closeness and out-closeness and they are defined and calculated based on Equation (13) and (14). In a bidirectional network, the closeness centrality is defined and measured through Equation (15). Here n refers to the number of nodes in the network, d refers to the distance between two nodes (node i and node j).

$$C_{in}(i) = \frac{n-1}{\sum_j d(i,j)}, \tag{13}$$

$$C_{out}(j) = \frac{n-1}{\sum_i d(i,j)}, \tag{14}$$

$$C(i) = \frac{n-1}{\sum_j d(i,j)}, \tag{15}$$

102

3.4.5.3   Network-level measurements

As described in Table 3, network-level measurements were applied to address a few research questions as well. To be more specific, two features were included during the process of comparing the four networks generated from mental health, children, teenagers, and older adults related health topics on the MedlinePlus portal: 1) network size is used to measure the number of nodes contained in each network; 2) network density (ND) refers to the proportion of the edges shared among the nodes in the network compared with the total number of edges that the network can possibly have. It may indicate the speed at which information spread among the nodes and can be defined and calculated through Equation (16). Here n refers to the number of nodes in a network and n' refers to the number of actual connections possessed by a network.

$$\text{ND} = \frac{n'}{n(n-1)} \tag{16}$$

3.5 Inferential statistical analysis

In this study, inferential statistical analysis is of importance in order to investigate the similarities and differences among the selected health topics related to mental health, children, teenagers, and older adults in the subject directory of MedlinePlus. Inferential analysis outcomes are vital in uncovering the patterns and characteristics of health topics and their connections as well as providing evidence for further optimizing the subject directory system. Every inferential analysis performed along with their corresponding research questions and hypotheses are discussed in detail in this section.

When studying social networks, inferential statistical analysis is performed on similarity

103

values and the network measurement results generated through the matrices and equations built previously, such as node features (degree centrality, betweenness centrality, and closeness centrality). One concern brought by those network measurements was that since the relationships among the nodes in the network depended on each other, it did not fulfill the requirement of "independent observation" for performing inferential analysis because, theoretically, the data should be extracted from a population that fits into a normal distribution (Borgatti et al., 2013). However, the issue was properly solved through the unique exponential random graph models (ERGMs) proposed in Borgatti et al.'s book.

The significant level (α) set for this study was 0.05. In other words, if a *p*-value generated from a hypothesis was smaller than 0.05, the null hypothesis was rejected. Otherwise it was not rejected. If a null hypothesis was rejected, an investigation was followed up to explore the reason of the rejection.

3.5.1   *Hypothesis groups 1, 3, 5, 7*

The hypothesis H01 in RQ2.1 is listed in the Section 1.2.3.2. The hypothesis H03 in RQ3.1 is listed in the Section 1.2.4.2. The hypothesis H05 in RQ4.1 is listed in the Section 1.2.5.2. The hypothesis H07 in RQ5.1 is listed in the Section 1.2.6.2. For H01, H03, H05 and H07, these four hypotheses were applied to the four different subcategories of health topics (mental, children, teenagers, and older adults) regarding the same objective – similarity values. Hence, these four hypotheses are listed in Table 4 shown below.

Subcategory	Mental Health and Behavior	Children	Teenagers	Older Adults
Hypotheses for	H01: There are no significant	H03: There are no significant	H05: There are no significant	H07: There are no significant

comparing similarity values between the original and the optimized networks	differences between the original and the optimized structural networks in terms of similarity values among the health topics related to mental health on the MedlinePlus portal.	differences between the original and the optimized structural networks in terms of similarity values among the health topics related to children on the MedlinePlus portal.	differences between the original and the optimized structural networks in terms of similarity values among the health topics related to teenagers on the MedlinePlus portal.	differences between the original and the optimized structural networks in terms of similarity values among the health topics related to older adults on the MedlinePlus portal.

Table 4. Four subcategories and their corresponding hypotheses

H01, H03, H05 and H07 form the first type of hypotheses groups, which focuses on the calculation of similarity values between the original and the optimized networks regarding the four health topic groups. The independent variable of this hypothesis type is the specific type of the structural link networks (original/optimized) of mental health, children, teenagers, and older adults related health topics. The dependent variable of this hypothesis type is the similarity measure results among the health topics in the original and the optimized networks. The dependent variables can be measured through the similarity values. A Mann-Whitney test was applied to test each hypothesis of this type.

3.5.2   *Hypothesis groups 2, 4, 6, 8*

The hypotheses $H02_{(a)}$, $H02_{(b)}$, and $H02_{(c)}$ in RQ2.2 is listed in the Section 1.2.3.3. The hypotheses $H04_{(a)}$, $H04_{(b)}$, and $H04_{(c)}$ in RQ3.2 is listed in the Section 1.2.4.3. The hypotheses $H06_{(a)}$, $H06_{(b)}$, and $H06_{(c)}$ in RQ4.2 is listed in the Section 1.2.5.3. The hypotheses $H08_{(a)}$, $H08_{(b)}$, and $H08_{(c)}$ in RQ5.2 is listed in the Section 1.2.6.3. For H02, H04, H06, and H08, there are four health topic groups selected from four various subcategories from the subject directory on the

105

MedlinePlus portal. Hence, each of the hypotheses in this hypothesis group is divided into 4 associated sub-hypotheses based on different subcategory of the subject directory (Mental Health and Behavior, Children, Teenagers, and Older Adults). Those 12 sub-hypotheses are listed in Table 5 shown below.

Subcategory	Mental Health and Behavior	Children	Teenagers	Older Adults
Hypotheses(a)	H02(a): There are no significant differences between the original and the optimized structural networks in terms of node-level centrality measurements among the health topics related to mental health on the MedlinePlus portal according to the degree centrality.	H04(a): There are no significant differences between the original and the optimized structural networks in terms of node-level centrality measurements among the health topics related to children on the MedlinePlus portal according to the degree centrality.	H06(a): There are no significant differences between the original and the optimized structural networks in terms of node-level centrality measurements among the health topics related to teenagers on the MedlinePlus portal according to the degree centrality.	H08(a): There are no significant differences between the original and the optimized structural networks in terms of node-level centrality measurements among the health topics related to older adults on the MedlinePlus portal according to the degree centrality.
Hypotheses(b)	H02(b): There are no significant differences between the original and the optimized structural networks in terms of node-level centrality measurements among the health topics related to mental health on the MedlinePlus portal according	H04(b): There are no significant differences between the original and the optimized structural networks in terms of node-level centrality measurements among the health topics related to children on the MedlinePlus portal according	H06(b): There are no significant differences between the original and the optimized structural networks in terms of node-level centrality measurements among the health topics related to teenagers on the MedlinePlus portal according	H08(b): There are no significant differences between the original and the optimized structural networks in terms of node-level centrality measurements among the health topics related to older adults on the MedlinePlus portal according

106

	to the closeness centrality.	to the closeness centrality.	to the closeness centrality.	to the closeness centrality.
Hypotheses(c)	H02(c): There are no significant differences between the original and the optimized structural networks in terms of node-level centrality measurements among the health topics related to mental health on the MedlinePlus portal according to the betweenness centrality.	H04(c): There are no significant differences between the original and the optimized structural networks in terms of node-level centrality measurements among the health topics related to children on the MedlinePlus portal according to the betweenness centrality.	H06(c): There are no significant differences between the original and the optimized structural networks in terms of node-level centrality measurements among the health topics related to teenagers on the MedlinePlus portal according to the betweenness centrality.	H08(c): There are no significant differences between the original and the optimized structural networks in terms of node-level centrality measurements among the health topics related to older adults on the MedlinePlus portal according to the betweenness centrality.

Table 5. Divided Hypotheses (a), (b), (c) and corresponding sub-hypotheses

H02, H04, H06 and H08 form the second type of hypotheses groups, which focuses on the calculation of the node-level centrality measurements between the original and the optimized networks regarding the four health topic groups. The independent variable of this hypothesis type is the specific type of the structural link networks (original/optimized) of mental health, children, teenagers, and older adults related health topics. The dependent variable of this hypothesis type is the node-level centrality measurements among the health topics in the original and the optimized networks. The dependent variables can be measured through degree centrality of each health topic, closeness centrality of each health topic, and betweenness centrality of each health topic. A customized T-test was applied to test each hypothesis of this type.

107

3.5.3   Hypothesis group 9

The hypotheses H09$_{(a)}$, H09$_{(b)}$, H09$_{(c)}$, and H09$_{(d)}$ in RQ6 is listed in the Section 1.2.7.1.

H09$_{(a)}$, H09$_{(b)}$, H09$_{(c)}$, and H09$_{(d)}$ form hypothesis group 9. The independent variable of hypothesis group 9 is the recommendation decisions of selected health topic connections generated by this study and the health field experts. The dependent variable of hypothesis group 9 is the structural link connections. The dependent variables can be measured through the numbers of suggested and not suggested connections. A Kappa test as well as a Chi-Square test were applied to test this hypothesis group.

3.6 Evaluation

After each hypothesis group was tested, two evaluators were invited to assess the optimization results. Both evaluators were recruited from a formal research institute in the United States. They have either gone through the preliminary examination of a M.D. program or obtained at least a master's degree in a medical or health related field. They have had at least five years of experience in the field.

Four lists that contained 200-300 paired health topics were generated for the corresponding health topic groups. There were two reasons for applying these paired health topics instead of networks for evaluation: 1) these pairs of health topics formed the whole network, hence they were key evaluation elements of a network; 2) compared with investigating a complicated network, evaluating a list of paired health topics was more straightforward and simple. Among them, about two thirds of the paired health topics were health topics which were suggested to be added to the structural network in the subject directory of MedlinePlus. Besides

108

those recommended topics, the remaining one third of paired health topics were also evaluated. Those health topics were not linked on the structural link network in the MedlinePlus portal and were found to have low semantic connections according to the similarity outcome generated in this study. Hence, they were used for comparative purposes. Such a mixed list of paired health topics could avoid potential bias from the experts.

A screenshot of a health topic's page including its "related health topics" list, along with a brief introduction about how the MedlinePlus portal creates and displays structural linkages among related health topics were provided to the evaluators. The evaluators were then asked to identify and mark the pairs of health topics that they considered to be relevant. A Kappa test was performed on the evaluation results generated between the two health filed experts. Then another Kappa test was employed to examine the consistency between the combined evaluation list from the two evaluators and the corresponding recommended results proposed by this study. Finally, a Chi-square test was employed to compare the combined evaluation list from the two evaluators and the corresponding recommended results proposed by this study.

3.7 Validity and reliability

Validity aims at evaluating how well the research questions have been addressed through the research study (Gravetter & Forzano, 2011) while reliability concerns whether the research outcomes can be repeatedly achieved under the same circumstances as the research study (Bryman, 2012). Validity and reliability are considered to be "tools of an essentially positivist epistemology" (Winter, 2000). They relate closely to the whole research process and need to be assured through the design of sampling and data collection methods as well as data analysis.

3.7.1 Internal validity

Validity assesses "the correctness or credibility of a description, conclusion, explanation, interpretation, or other sort of account" (Maxwell, 2012). To be more specific, internal validity attempts to verify that the outcomes of the study are caused by the variables that are within the consideration of the research investigation instead of any other unaccounted factors (Winter, 2000).

Internal validity was ensured through two means in this study. In terms of sampling and data collection, several health topics from multiple subcategories in the MedlinePlus portal were selected to serve as the starting point when constructing the subject directory social network. Meanwhile, for those common concerns like omission errors, commission errors, edge/node attribution errors, etc. mentioned in Borgatti et al.'s book (2013) in terms of the social network data designed and collected, several processes were followed. First, the social network data were gathered through the MedlinePlus portal instead of any individual creator or third-party platform. Second, the social network data were collected from the hyperlinks and textual information contained on the Web pages of selected health topics so that all nodes and edges were clearly identified from both structural and semantic perspectives. Third, since the data collection process involved multiple levels of health topics in the subject directory system, it helped avoid missing possible related topics if those topics were of interest. The node/edge attribution errors were addressed since all the interactions among health topics in the subject directory system were reflected by their structural and semantic connections. Relating to that, it is reasonable to assume that the relationships investigated were able to include all the relationships among the health topics in the subject directory system. Therefore, edge/node attribution errors, as well as

110

omission and commission errors were avoided.

In terms of data analysis, the internal validity was ensured through the process of conducting inferential statistical analysis in five parts: 1) the social network features were compared through inferential analysis among health topic groups from various subcategories and sections on MedlinePlus. Therefore, the findings of the study were determined by the relationships among health topics instead of the differences of subcategories or sections. In other words, the research results were not impacted by the condition of whether a health topic was from Mental Health and Behavior subcategory, Children subcategory, Teenagers subcategory, or Older Adults subcategory. 2) Calculation results regarding centrality measurements generated through social network analysis were normalized during the comparing process when necessary in order to eliminate any potential impact resulting from different network sizes. 3) Multiple features were utilized for optimizing the subject directory of MedlinePlus. Similarity values, node-level centrality features, as well as the evaluation results from the two health professionals were applied together during the process of optimization, which resulted in the conclusion of this study to become more convincing. 4) Besides applying the cosine similarity measure as the primary similarity measure for identifying recommended health topic connections, the Pearson correlation and the Euclidean distance similarity measures were also employed to compare and verify the recommendation results. Hence, the final suggestions presented in this study were analyzed through three different similarity measures to further ensure the internal validity. 5) Regarding the evaluation process, the consistency between the recommendation results from this study and the evaluation results from the two health field experts were twice confirmed through both the Kappa test and the Chi-square test.

111

3.7.2 *External validity*

External validity concerns whether the research findings can be generalized to a wider range than the study. It is defined as "the extent to which the results obtained in a research study hold true outside that specific study" (Gravetter & Forzano, 2011, p. 166). It indicates that the sample and data selected should be representative.

External validity was ensured in this study through two means. First, the four health topic groups derived from the four subcategories on the MedlinePlus portal (Mental Health and Behavior, Children, Teenagers, Older Adults) included a large number of health topics that consisted of not only clear structures formed by 3 to 4 levels of hyperlinks but also plentiful textual information. It ensured that there were sufficient nodes and edges among the selected health topics to be observed, collected and analyzed in later phases. In addition, according to the investigation results generated from the pilot study, health topics related to mental health, children, teenagers, and older adults were found to be associated frequently with health topics under other sections such as Health and Wellness. Such wide coverage as well as plentiful interactions improved the possibility that the research outcomes of this study can be generalized to the whole subject directory system of the MedlinePlus portal. Moreover, the findings of this study can be applied to other online portals that also utilize subject directory systems for further optimizing purposes as well.

The evaluation results and corresponding optimization suggestions for mental health, children, teenagers, and older adults related health topics utilized in the subject directory of MedlinePlus may also be applied to other public health portals on the Internet such as the website of the World Health Organization (WHO). Also, the patterns and characteristics explored

within the structural and semantic connections among health topic groups of mental health, children, teenagers, and older adults may be generalized to other subcategories and sections in health information related subject directories as well as other similar thesaurus such as NLM's Medical Subject Headings (MeSH). And last, the research methodologies employed in this study - including social network analysis regarding similarity measures, as well as node and network features, along with inferential statistical analysis - may be applied to investigate a health topic-based subject directory from other subcategories such as cancer, diabetes, etc. or various public health portals. The analysis conducted in this study was based on a hierarchical structure performed by MedlinePlus, hence, the method of this study can be applied to other online portals with hierarchical structures.

3.7.3   *Reliability*

Reliability refers to the consistency and stability throughout the process of data collection and analysis (Gravetter & Forzano, 2011). In this study, the research objects were the health topics applied in the subject directory of MedlinePlus. The structural and semantic connection data were considered to be stable and consistent in a long-term period despite slight changes to detailed information. For textual information, all data collected through the coded Python program were saved in an independent text file for further review.

During the sampling and data collection process, all structural and semantic data (hyperlinks and textual information) were gathered and kept in their original format. Few biases were identified during those procedures. When cleansing the word list generated from collected textual information, the stop-word list was built based on universally accepted rules within the field of social science. The adjustments made in word format, for instance, transferring the word

113

"achieves" to "achieve", followed the rules published by English language authorities.

However, in the data analysis, subjective observation as well as suggestions might have been made by the researcher based on individual experience, knowledge base, and judgement. To eliminate potential instability and inconsistency during the data analysis phase, two domain experts from health and medicine related areas were invited to perform an evaluation analysis regarding the research outcomes. A Kappa test and a Chi-Square test were applied to estimate the agreement level as well as potential differences between the study and domain experts in terms of the relationships among selected health topics.

3.7.4   Summary

This section has discussed the validity and reliability issues that were faced by the study. Solutions for addressing the potential validity and reliability concerns have been introduced that cover multiple research phases including sampling, data collection, and data analysis. A detailed research design as well as the results generated from the pilot study have been introduced in this section in order to support the proposed research procedures for this study. Proper considerations and approaches were applied to ensure that the research outcomes are sound. In addition, the methodology employed in this study that combines social network analysis and inferential statistical analysis can be expected to provide a positive model to other studies focusing on evaluating and optimizing health information related subject directories on online public health portals.

3.8 Research design

Figure 12 demonstrates the overall organization of the research problem, research

114

questions, and corresponding hypotheses raised in this study.
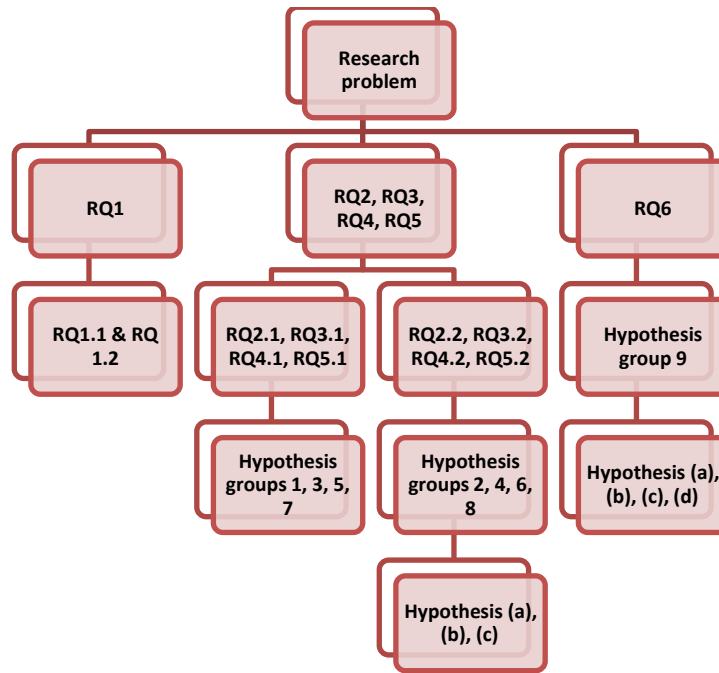
Figure 12. Research Design Organization

3.9 Summary

Table 6 summarizes the research questions with their corresponding hypothesis, independent variables (IV) with their valid values, dependent variables (DV) with their measurement, and the associated method used for testing.

Research questions	Hypothesis	IV	Valid values for IV	DV	Measurement for DV	Test
RQ2.1, RQ3.1, RQ4.1, RQ5.1	Hypothesis groups 1, 3, 5, 7	Specific type of the structural link networks	Original, optimized	Similarity measure results	Similarity values	Mann-Whitney test
RQ2.2,	Hypothes	Specific type	Original,	Node-level	Degree	Customiz

115

RQ3.2, RQ4.2, RQ5.2	is groups 2, 4, 6, 8	of the structural link networks	optimized	centrality measureme nts	centrality, closeness centrality, and betweennes s centrality	ed T-test
RQ6	Hypothes is group 9	Recommendat ion decisions	Recommend ed, not recommende d	Structural link connections	Numbers of suggested and not suggested connections	Kappa test and Chi- Square test

Table 6. Display of inferential analysis methods applied in this study

116

Chapter 4. Results and Analysis

In this chapter, all the results generated through social network analysis, descriptive and inferential statistical analysis are presented. These results revealed facts about selected health topics relating to mental health, children, teenagers, and older adults regarding their structural and semantic relationships. Optimizations were presented for each of the four health topic groups in terms of improving the similarity values and identifying influential health topics in the structural networks. The optimized structural networks were then compared with the original ones to ensure that the optimizations had generated significant improvements in MedlinePlus's subject directory. In the end, the recommendation results presented by this study were verified to be consistent with the evaluation outcomes reached by the two health field experts.

4.1 General description of the subject directories of the four health topic groups

4.1.1   Description of the collected data

The data were collected and processed from September 2019 to April 2020. Each of the four selected health topic groups was assigned a starting health topic to begin the construction of the first level in the original structural network. As described in the data collection section, these health topics included in Level 1 were generated from the "related health topics" column of the starting health topic. The starting health topic assigned to the mental health subcategory was *Mental Disorders*, and 99 health topics distributed at 4 levels were included in total. The starting health topic assigned to the children subcategory was *Children's Health*, and a total of 159 health topics distributed at 3 levels were included. The starting health topic assigned to the teenagers subcategory was *Teen Health*, and a total of 140 health topics distributed at 4 levels were

117

included. Last, the starting health topic assigned to the older adults subcategory was *Older Adult Health*, and a total of 150 health topics distributed at 4 levels were included.

For the three health consumer groups, the three starting health topics selected were of the same type: they fit into the requirements stated in the data collection section and had appropriate numbers of related health topics to form an appropriate first level in the network. The mental health subcategory is different from the three health consumer groups so the topic *Mental Disorders* was chosen to serve as the starting point for two reasons: 1) compared with the other candidate topic (*Mental Health* whose page only contains three related health topics (*Child Mental Health, Mental Disorders, and Teen Mental Health*)), the topic page of *Mental Disorders* includes a wider range of related health topics (19 in total). Those topics covered not only the same topics of *Mental Health*, but also a group of additional common mental health related topics such as *Anxiety*, *Bipolar Disorder*, etc; 2) it is the official term used by the MedlinePlus portal to represent mental related diseases - other related terms such as *Mental Illness* have been combined into the same topic page.

The detailed involvement of the selected health topics located in various levels regarding each of the four health topic groups is displayed in Table 7 shown below.

	Level 1	Level 2	Level 3	Level 4	Total
Mental health	20	11	20	48	99
Children	14	54	91		159
Teenagers	7	16	59	58	140

| Older adults | 13 | 16 | 43 | 78 | 150 |

Table 7. Numbers of health topics involved in each level of each subcategory group

For mental health related health topics, text was extracted from each of the 99 topics' introductory pages. A validation process was conducted and all stop-words were removed. Consequently, 2441 keywords were left. Two filtering processes followed: those keywords that only appeared once were eliminated, leaving 1413 keywords (cut-off point =1). For those keywords that appeared more than once, they were ensured to have appeared in more than one topic's page, otherwise they were excluded as well. This was the binary check process. As a result, 1211 keywords were found to appear in at least two different health topics' pages. Finally, when the keywords collection was adjusted further in cut-off point 2 to pursue more meaningful similarity measurement results, the keyword count was 934. The analytical results showed that the similarity data between cut-off points 2 and 3 were extremely close; hence the outcome of cut-off point 2 was selected to include a wider keyword range. This method was also applied to the other three health topic groups.

For children related health topics, text was extracted from each of the 159 topics' introductory pages. Through the same validation and filtering procedures, 3880 keywords were left after all stop-words were removed, and 2224 keywords were found to appear in more than just one health topic's Web page. Following that, 1919 keywords were kept after the binary check, and 1457 keywords formed the final list (cut-off point =2).

For teenagers related health topics, text was extracted from each of the 140 topics' introductory pages. Through the same validation and filtering procedures, 3329 keywords were

119

left after all stop-words were removed, and 1982 keywords were found to appear in more than just one health topic's Web page. Following that, 1742 keywords were kept after the binary check, and 1361 keywords formed the final list (cut-off point =2).

For older adults related health topics, text was extracted from each of the 150 topics' introductory pages. Through the same validation and filtering procedures, 3535 keywords were left after all stop-words were removed, and 2090 keywords were found to appear in more than just one health topic's Web page. Following that, 1872 keywords were kept after the binary check, and 1490 keywords formed the final list (cut-off point =2).

A summary regarding the collected textual information of the four health topic groups is displayed in Figure 13.
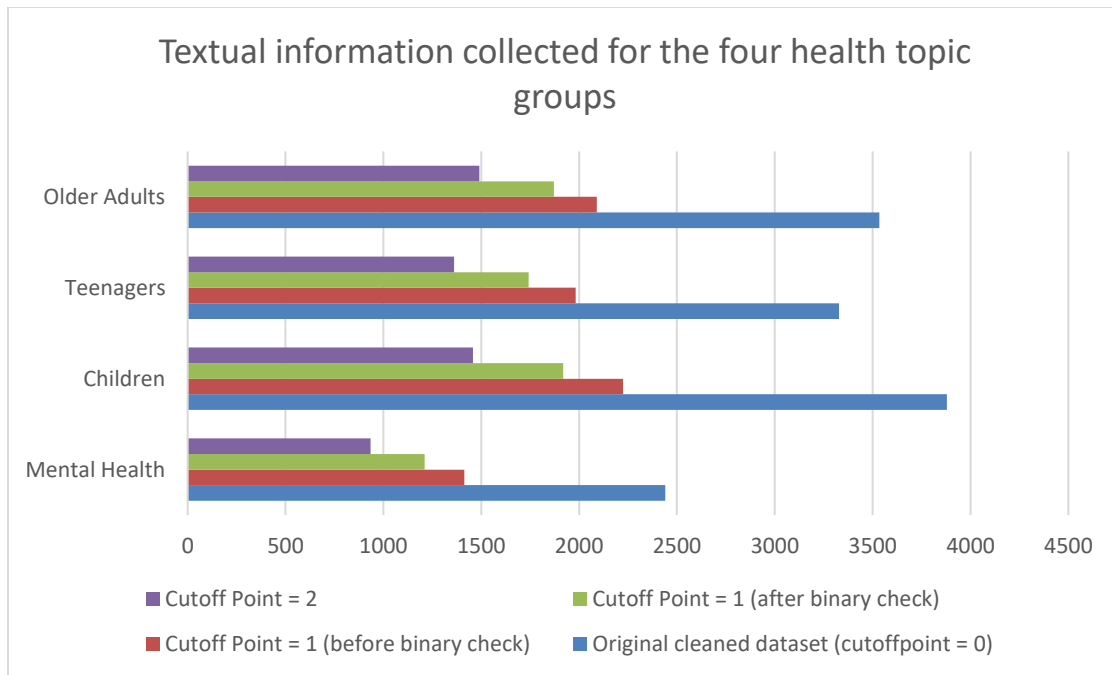
Figure 13. Textual information collected for the four health topic groups

4.1.2   Description of the structural link networks of the four health topic groups

This section describes how the chosen health topics of the four subcategory groups are connected structurally in the current subject directory of MedlinePlus. The structural link networks of the four health topic groups are described separately.

4.1.2.1   Structural link network of mental health related health topics

Two hundred sixty structural connections were observed in the original network created by the MedlinePlus portal. A visualized figure of the structural link network of mental health related topics is displayed in Figure 14.

In Figure 14, some of the edges between two health topics were unidirectional while others were bidirectional. Moreover, the health topics were not evenly distributed on the visualized network. Therefore, the selected 99 health topics could be classified into 3 groups: the first group was titled as "specific mental related disease group" and it included topics about specific mental related diseases/disorders such as *Mental Disorders*, *Post-traumatic Stress Disorder (PTSD)*. The second group was defined as "health consumer related group" since it included topics that were associated with specific health consumer groups such as children and teenagers. It is interesting that older adults related topics were not included in this group. The last group was labelled as "daily health element group". This group consisted of several topics that represented components for daily health, for instance, *Nutrition*, *Diets*, *Calcium*. Part of the "daily health element group" was isolated. There was a sub-group containing the topic *Nutrition* located at the right side of the figure. Note that a health topic might belong to multiple groups -

121

some topics like *Child Nutrition* and *Teen Mental Health* were considered to be included in two

groups and such overlapped topics were serving as the bridging topic to link two groups.
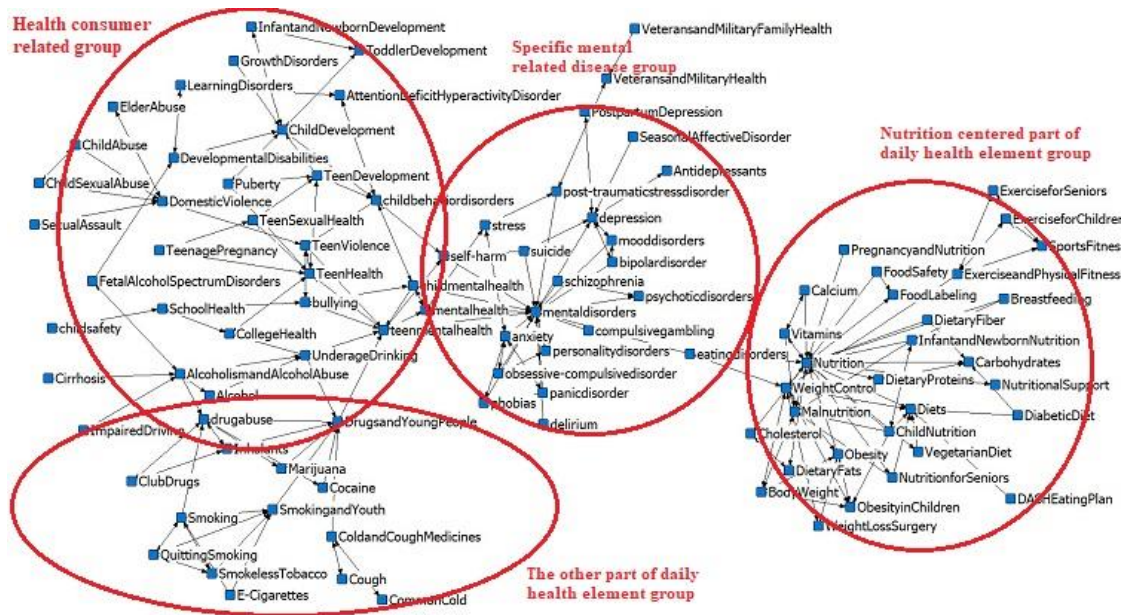
Figure 14. Display of the structural link network of mental health related topics

4.1.2.2   Structural link network of children related health topics

Four hundred seventy-two structural connections were observed in the original network

created by the MedlinePlus portal. A visualized figure of the structural link network of children

related topics is displayed in Figure 15.

In Figure 15, similar to the structural network of mental health related topics, both

unidirectional and bidirectional connections were identified in the structural network of the

children subcategory. However, compared with the mental health topics' network, the structural

network of children related health topics was more evenly distributed – no sections were found

to be apparently clustered within this network. The results showed that *Teen Health* occurred in

122

the network at a quite close location to *Children's Health*, which might have indicated a strong relationship between children related health topics and teenagers related health topics. Topics *Mental Health* as well as *Mental Disorders* were found to appear at the top edge of the network, with few specific mental related disease topics included. Last, *Older Adult Health* was not found in this network and the topic *Exercise for Older Adults* seemed to be the only health topic in this network that is directly related to the older adults subcategory. *Older Adult Health* might be reached at the fourth level through the path built by *Exercise for Older Adults*, but considering the fact that the network had already generated 159 health topics within three levels, it is reasonable to regard the relationship between children related health topics and older adults related health topics as not strong.
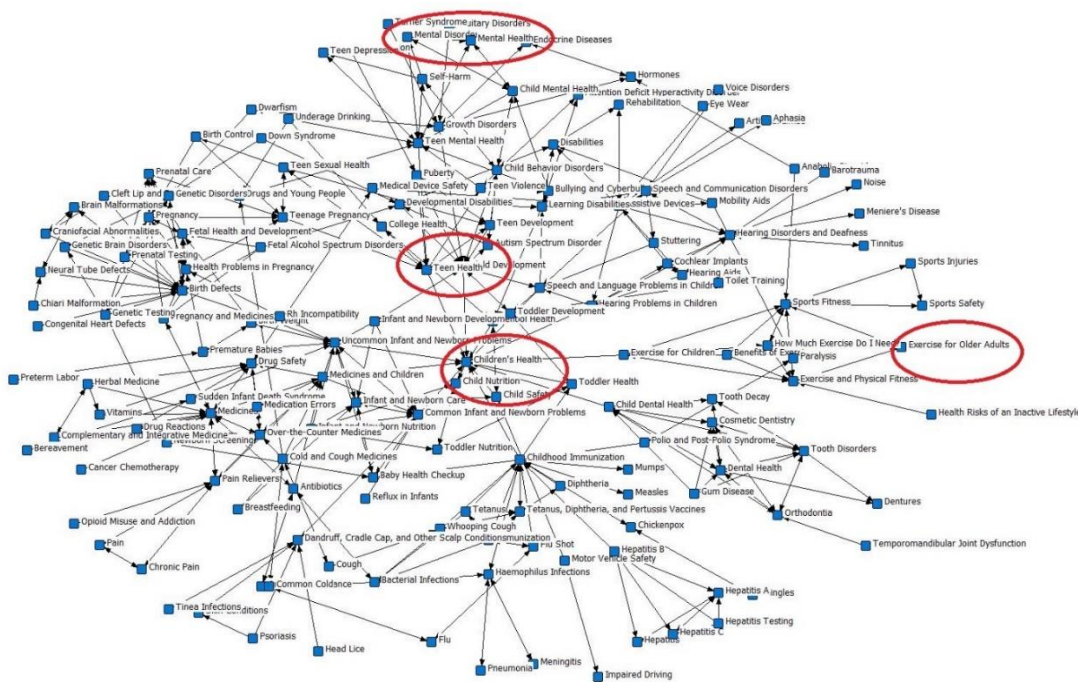
Figure 15. Display of the structural link network of children related topics

4.1.2.3   Structural link network of teenagers related health topics

Three hundred eighty-nine structural connections were observed in the original network created by the MedlinePlus portal. A visualized figure of the structural link network of teenagers related topics is displayed in Figure 16.

In Figure 16, again, both unidirectional and bidirectional connections were found in the network. Unlike the networks of the mental health subcategory and the children subcategory, the starting health topic of the teenagers subcategory, *Teen Health*, was a bit isolated in the network. It was located at the center of the network just like the other two starting health topics, however, no clustering patterns were found around *Teen Health*. Instead, two clustering groups of health topics were detected at the corners of the network. One of the clustered topic group was about pregnancy and sexual health and this group of health topics was located at the left bottom corner of the network. The two key bridging health topics that were linking this group and *Teen Health* were *Teen Sexual Health* and *Teen Pregnancy*. The other clustered topic group was the same one in the mental health related health topics' network about specific mental diseases. This group was located at the right bottom corner of the network and had indicated a strong linkage between teenagers and mental health. The key bridging health topic was *Teen Mental Health*.

In addition to the two clustering groups of health topics, some other interesting facts were also identified. First, *Children's Health* was not found in this network, meaning that it was not linked back to *Teen Health*, therefore the connection between these two starting health topics was a unidirectional connection. However, there were still a few children-related health topics located at the top of the network, which could prove the strong relationship between the two health consumer groups children and teenagers. Second, a group of health topics relating to daily health

124

elements were identified in the bottom middle area, with most of them being drug and alcohol

related conditions. This group of health topics were found to be connected to both the pregnancy

and sexual related topic group, as well as the mental health related topic group, which fit into the

common conceptual model of people. Last, *Older Adult Health* was also involved through the

path built by *Older Adult Mental Health* and was located at the edge in this network. Similarly

with the network of children related health topics, the relationship between the teenagers

subcategory and the older adults subcategory seemed to be not strong.
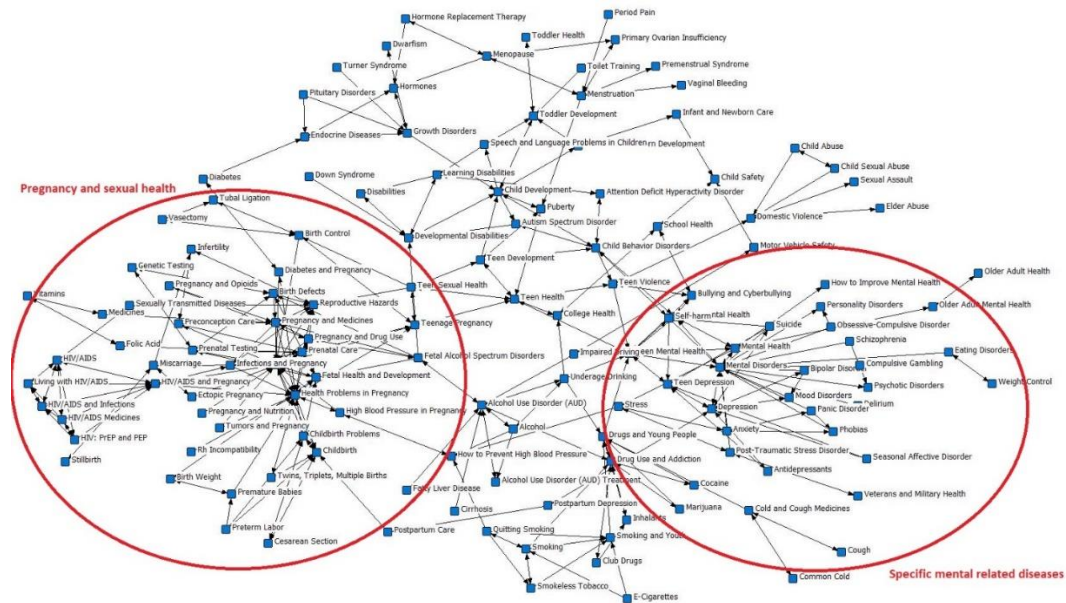
Figure 16. Display of the structural link network of teenagers related topics

4.1.2.4   Structural link network of older adults related health topics

Four hundred forty-six structural connections were observed in the original network

created in the MedlinePlus portal. A visualized figure of the structural link network of older

adults related topics is displayed in Figure 17.

In Figure 17, as other previously described networks, both unidirectional and bidirectional connections were identified. Among all the connections, two clustered health topic groups were found: the first group was the same one about specific mental diseases in the networks relating to mental health related topics and teenagers related topics. The other group was about daily health elements, with a specific focus on nutrition. This nutrition centered daily health elements group covered a wider range of health topics than the one included in the mental health's network. Regarding older adults related health topics, they were located at the left corner of the network and seemed not to be closely clustered. These health topics were found to be related to caregiving and exercising related topics. The older adults subcategory was connected to the children subcategory through two bridging topics *Child Nutrition* and *Nutrition for Older Adults*. *Teen Health* was not included in this network despite the fact that *Teen Mental Health* was connected through the specific mental disease group. This finding echoed the one generated in teenagers' section, and proved that the relationship between the teenagers subcategory and the older adults subcategory was not strong.
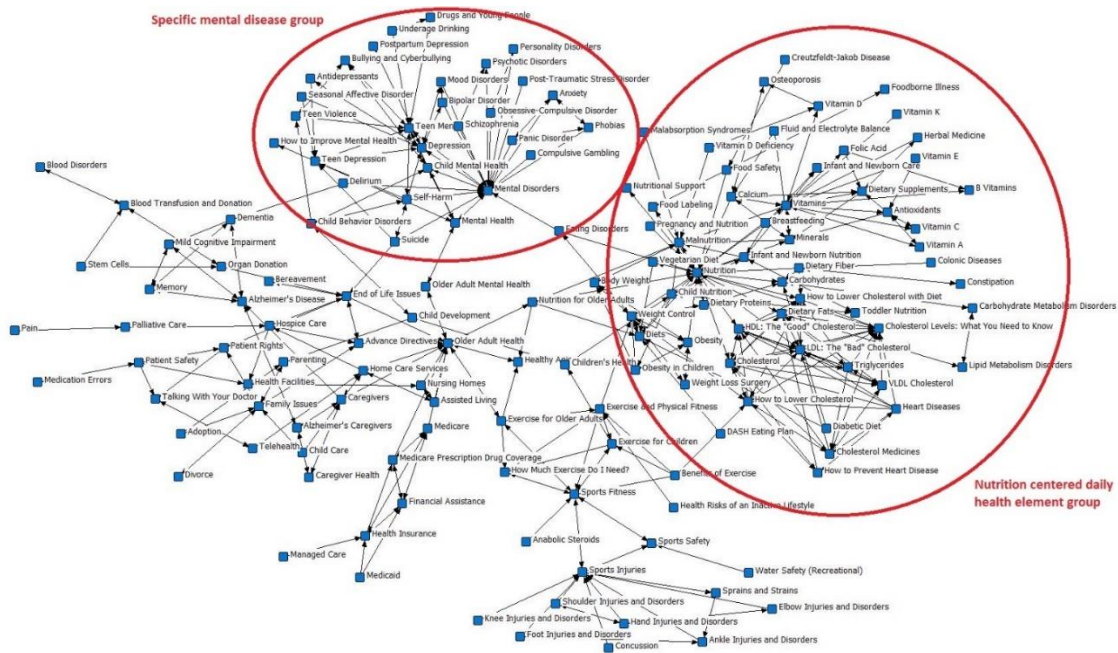
Figure 17. Display of the structural link network of older adults related topics

4.1.3   Description of the semantic networks of the four health topic groups

This section aims at finding the semantic relationships possessed by the four selected health topic groups on the MedlinePlus portal. The semantic networks of the four health topic groups are described separately.

4.1.3.1   Semantic network of mental health related health topics

For mental health related topics, the semantic relationship could be derived from every topic in terms of similarity value between its Web page's textual information and other selected health topics'. Hence, there could be 9,801 semantic connections in total among the selected 99 health topics. Because of the fact that a health topic cannot be listed as related to itself, 99 connections were excluded from the original 9,801 connections, leaving 9,702 connections. Among the 9,702 connections, 260 connections were already built in the structural network by

127

the portal creators to link hundreds of pairs of related health topics. To generate the semantic network, the cosine similarity measure was performed to check how similar two health topics' pages were based on the word frequency contained in the text. As a result, there were 260 edges built in the category of MedlinePlus (Edge Set B) and the average similarity value was 0.383677 while the average similarity among the rest of the 9442 pairs of topics (Edge Set C) was 0.098230. The overall average similarity was 0.105879.

4.1.3.2   Semantic network of children related health topics

For children related topics, there were 25,281 semantic connections initially among the 159 health topics and 25,122 left after the self-relating 159 connections were excluded. There were 472 edges built in the category of MedlinePlus (Edge Set B) and the average similarity value was 0.444744 while the average similarity among the rest of the 24,650 pairs of topics (Edge Set C) was 0.086589. The overall average similarity was 0.093318.

4.1.3.3   Semantic network of teenagers related health topics

For teenagers related topics, there were 19,600 semantic connections initially among the 140 health topics and 19,460 left after the self-relating 140 connections were excluded. There were 389 edges built in the category of MedlinePlus (Edge Set B) and the average similarity value was 0.461491 while the average similarity among the rest of the 19,071 pairs of topics (Edge Set C) was 0.104384. The overall average similarity was 0.111522.

4.1.3.4   Semantic network of older adults related health topics

For older adults related topics, there were 22,500 semantic connections initially among the 150 health topics and 22,350 left after the self-relating 150 connections were excluded. There

128

were 446 edges built in the category of MedlinePlus (Edge Set B) and the average similarity value was 0.440717 while the average similarity among the rest of the 21,904 pairs of topics (Edge Set C) was 0.097377. The overall average similarity was 0.104228.

4.1.3.5 Summary

Table 8 summarizes the average similarity values of structurally connected linkages, unconnected linkages, and all linkages regarding mental health, children, teenagers, and older adults related health topics.

	Average All	*Standard Deviation of similarity*	*Average Connected*	*Standard Deviation of similarity*	*Average Unconnected*	*Standard Deviation of similarity*
Mental health	0.105879	0.100205	0.383677	0.195807	0.09823	0.084155
Children	0.093318	0.095590	0.444744	0.210343	0.086589	0.077825
Teenagers	0.111522	0.112286	0.461491	0.209862	0.104384	0.097056
Older adults	0.104228	0.108716	0.440717	0.218706	0.097377	0.093464

Table 8. Similarity values of various topic groups' semantic connections

4.1.4 *Description of the network-level measurements of the four health topic groups*

Network-level measurements can uncover general characteristics and overall patterns of all nodes as well as their interactions within a network. Table 9 summarizes the network-level measurements of the four health topic groups. The network size measures the total number of nodes within a network and as noted in Table 9, the children related health topic group had the largest network size, followed by the older adults related health topic group, the teenagers related health topic group, and finally the mental health related topic group.

129

Network density measures the ratio of the number of all existing connections (or edges) among the nodes (or actors) to the number of whole connections a network could possibly have. As shown in Table 9, the mental health related topic group had the highest network density value, which indicated that the health topics in this group had a higher potential of connecting with others. The network of children related health topics had the lowest network density value.

	Mental health	Children	Teenagers	Older adults
Network size	99	159	140	150
Network density	0.027	0.019	0.020	0.020

Table 9. Network-level measurements of the four health topic groups.

4.2 Findings for research question 1 (RQ1)

RQ1: How are health topics related to mental health, children, teenagers, and older adults connected in the subject directory on the MedlinePlus portal: are the structural and semantic connections of the health topics consistent?

The first research question investigates if the current connections possessed by the selected health topics on the MedlinePlus portal are consistent from both the structural and the semantic perspectives.

4.2.1   RQ1.1

RQ1.1 is articulated as "Are the similarity values between the structural and semantic connections among the health topics related to mental health, children, teenagers and older adults consistent on MedlinePlus?" This question explores the consistency between the structural and semantic connections on the MedlinePlus portal. The investigation results of the four health topic

130

subcategories are discussed separately.

4.2.1.1   Mental health related health topics

Among the 9,442 comparisons that did not appear on the structural link network, 133 pairs of topics were found to have a similarity value larger than the threshold generated from the structurally connected topics. This indicated that the similarity values of the structural connections were not consistent with the similarity values of the semantic connections because a group of health topics possessing strong semantic relationships were not linked structurally. To be more specific, among these 133 pairs of topics, 110 pairs were recommended to set up bidirectional connections and these pairs of edges are listed in Table 10. Due to the word limitation, Table 10 only contains 20 out of the 110 pairs of the edges. For a complete list of the 110 bidirectional connections, please check the complete version of Table 10 in Appendix 5. Meanwhile, since these comparisons included the same health topics in both directions, these related topics are displayed only once in the table to save space.

Pairs of Topic A & Topic B		Pairs of Topic A & Topic B	
Topic A	**Topic B**	**Topic A**	**Topic B**
Child development	*Child mental health*	*Child abuse*	*Child development*
Child safety	*Child mental health*	*Child sexual abuse*	*Child development*
Learning disorders	*Child mental health*	*Teenage pregnancy*	*Teen development*
Toddler development	*Child mental health*	*Teen health*	*Body weight*
Obesity in children	*Child mental health*	*Weight loss surgery*	*Diets*

Table 10. 55 bidirectional pairs of health topics that require structural linkages in the mental health related topic group

In addition to the 110 pairs of bidirectional edges discussed above, there were another 23

131

pairs of health topics sharing a similarity value that was higher than the threshold. The difference was that these 23 pairs of the topics had already been linked from one direction on the structural link network. Hence, users would not be able to return to the previous topic. These health topics as well as their similarity measurement results are shown in Table 11. Table 11 only contains 6 out of the 23 pairs of the unidirectional edges. For a complete list of the 23 connections, please check the complete version of Table 11 in Appendix 6.

Topics and their related topic		Topics and their related topic	
Topic	Related topic	Topic	Related topic
Child mental health	Teen mental health	Quitting smoking	Smoking and youth
Bipolar disorder	Mood disorders	Smokeless tobacco	Smoking and youth
Veterans and military family health	Veterans and military health	Child abuse	Domestic violence

Table 11. 23 unidirectional edges recommended for the mental health topic group

The 23 pairs of unidirectional edges could be divided into two types: one type indicated a hierarchical relationship that was from general to specific or vice versa, such as *Diets* and *Vegetarian Diet*, *Nutrition* and *Child Nutrition*; the other type referred to a semantic relevance relationship, for example, *Child Mental Health* and *Teen Mental Health*, *Food Labeling* and *Food Safety*. Both types of unidirectional edges were supposed to be adjusted to bidirectional edges for navigating purposes.

4.2.1.2   Children related health topics

Regarding children related health topics, among the 24,650 semantic connections that did not appear on the structural link network, 184 pairs of topics were found to have a similarity

value larger than the threshold generated from the structurally connected topics. Again, the structural connections were found to be not consistent with the semantic connections. Among these 184 pairs of topics, 140 pairs were recommended to set up bidirectional connections and these pairs of edges are listed in Table 12. For a complete list of the 140 bidirectional connections, please check the complete version of Table 12 in Appendix 7.

Pairs of Topic A & Topic B		Pairs of Topic A & Topic B	
Topic A	**Topic B**	**Topic A**	**Topic B**
Children's Health	*Child Behavior Disorders*	*Child Dental Health*	*Dentures*
Children's Health	*Toddler Development*	*Child Dental Health*	*Child Mental Health*
Children's Health	*Baby Health Checkup*	*Child Dental Health*	*Child Nutrition*
Children's Health	*Child Mental Health*	*Child Development*	*Medicines and Children*
Child Dental Health	*Toddler Health*	*Child Development*	*Toddler Health*

Table 12. 70 bidirectional pairs of health topics that require structural linkages in the children related topic group

In addition to the 140 pairs of bidirectional edges discussed above, there were another 44 unidirectional pairs of health topics sharing a similarity value that was higher than the threshold. These health topics are shown in Table 13. For a complete list of the 44 connections, please check the complete version of Table 13 in Appendix 8.

Topics and their related topic		Topics and their related topic	
Topic	**Related topic**	**Topic**	**Related topic**
Children's Health	*Child Nutrition*	*Teen Mental Health*	*Teen Health*
Child Dental Health	*Children's Health*	*Teenage Pregnancy*	*Teen Health*
Child Dental Health	*Cosmetic Dentistry*	*Toddler Nutrition*	*Toddler Health*

133

Table 13. 44 unidirectional edges recommended for the children topic group

Similar to the mental health topic subcategory, both hierarchical and relevant relationships were identified in the recommended unidirectional connections of children related health topics. For instance, *Pregnancy* and *Pregnancy and Medicines* were presenting a hierarchical relationship while *Children's Health* and *Toddler's Health* were presenting a relevance relationship.

4.2.1.3   Teenagers related health topics

In terms of teenagers related health topics, among the 19,071 semantic connections that did not appear on the structural link network, 241 pairs of topics were found to have a similarity value larger than the threshold generated from the structurally connected topics. This finding fit into the prior results generated from the topic groups related to mental health and children. The structural connections were not consistent with the semantic connections in this health topic subcategory as well. Among these 241 pairs of topics, 190 pairs were recommended to set up bidirectional connections and these pairs of edges are listed in Table 14. For a complete list of the 190 bidirectional connections, please check the complete version of Table 14 in Appendix 9.

Pairs of Topic A & Topic B		Pairs of Topic A & Topic B	
Topic A	**Topic B**	**Topic A**	**Topic B**
Teen Health	*Teen Depression*	*Eating Disorders*	*Obsessive-Compulsive Disorder*
College Health	*Mental Health*	*Eating Disorders*	*Autism Spectrum Disorder*
Teen Development	*Teen Mental Health*	*Mood Disorders*	*Obsessive-Compulsive Disorder*
Teen Development	*Teenage Pregnancy*	*Mood Disorders*	*Personality Disorders*

134

| Teen Development | Teen Depression | Mood Disorders | Autism Spectrum Disorder |

Table 14. 95 bidirectional pairs of health topics that require structural linkages in the teenagers related topic group

Besides the 190 pairs of bidirectional edges discussed above, there were another 51 unidirectional pairs of health topics sharing a similarity value that was higher than the threshold. These health topics are shown in Table 15. For a complete list of the 51 connections, please check the complete version of Table 15 in Appendix 10.

Topics and their related topic		Topics and their related topic	
Topic	**Related topic**	**Topic**	**Related topic**
Teen Mental Health	*Teen Health*	*Pregnancy and Medicines*	*Health Problems in Pregnancy*
Teenage Pregnancy	*Teen Health*	*Pregnancy and Medicines*	*Pregnancy*
Child Mental Health	*Teen Mental Health*	*Pregnancy and Medicines*	*Infections and Pregnancy*

Table 15. 51 unidirectional edges recommended for the teenagers topic group

Instances of both hierarchical and relevance relationships were also identified among the recommended connections for teenagers related health topics. A hierarchical relationship could be reflected by paired health topics such as *Teen Mental Health* and *Teen Health*, and a relevance relationship could be reflected by pairs of health topics like *Child Mental Health* and *Teen Mental Health*.

4.2.1.4   Older adults related health topics

There were 21,904 semantic connections that were not included in the structural link network of the older adults topic subcategory. Within them, 244 pairs of topics were found to have a similarity value larger than the threshold generated from the structurally connected topics.

135

Similar to the other three health topic groups, the structural and semantic connections within the older adults topic subcategory were not consistent. Among these 244 pairs of topics, 196 pairs were recommended to set up bidirectional connections and these pairs of edges are listed in Table 16. For a complete list of the 196 bidirectional connections, please check the complete version of Table 16 in Appendix 11.

Pairs of Topic A & Topic B		Pairs of Topic A & Topic B	
Topic A	**Topic B**	**Topic A**	**Topic B**
Older Adult Health	*Mental Health*	*Child Nutrition*	*Child Behavior Disorders*
End of Life Issues	*Palliative Care*	*Dietary Fiber*	*Diets*
End of Life Issues	*Managed Care*	*Diets*	*Body Weight*
Exercise for Older Adults	*Benefits of Exercise*	*Diets*	*Dietary Supplements*
Exercise for Older Adults	*Exercise for Children*	*Eating Disorders*	*Mood Disorders*

Table 16. 98 bidirectional pairs of health topics that require structural linkages in the older adults related topic group

Other than the 196 pairs of bidirectional edges discussed above, there were another 48 unidirectional pairs of health topics sharing a similarity value that was higher than the threshold. These health topics are shown in Table 17. For a complete list of the 48 connections, please check the complete version of Table 17 in Appendix 12.

Topics and their related topic		Topics and their related topic	
Topic	**Related topic**	**Topic**	**Related topic**
Exercise for Older Adults	*Healthy Aging*	*Sprains and Strains*	*Ankle Injuries and Disorders*
Exercise for Older Adults	*How Much Exercise Do I Need?*	*Children's Health*	*Child Nutrition*
Home Care Services	*Hospice Care*	*Cholesterol Medicines*	*HDL: The "Good" Cholesterol*

136

Table 17. 48 unidirectional edges recommended for the older adults topic group

Similar to the other three health topic groups, both hierarchical and relevance relationships were investigated among the recommended unidirectional connections within the older adults topic subcategory. To be more specific, *Vitamin K* and *Vitamins* served as an example of a hierarchical relationship. *Sports Safety* and *Sports Injuries* served as a nice instance of a relevance relationship.

4.2.2   RQ1.2

RQ 1.2 is stated as "Are the node-level centrality measures between the structural and semantic connections among the health topics related to mental health, children, teenagers and older adults consistent on MedlinePlus?" This research question attempts to reveal the node-level features of the selected health topics among the four health topic groups. The outcomes of the four health topic subcategories are discussed separately.

4.2.2.1   Mental health related health topics

Following the procedures described in the methodology section, the node-level features including the three centrality measurements were analyzed regarding the mental health related health topics. Degree centrality, closeness centrality, and betweenness centrality were calculated for every selected health topic according to the original structural link network. Those health topics that were playing the most influential roles were identified and listed. Subsequently, the recommended connections generated from the earlier stages were added to the original structural link network to form a new connection network. This network contained both the original structural connections set up by the portal creators and the new connections that were found to

possess strong semantic relationships. In other words, this new network could be regarded as a revised semantic network based on the semantic relationships. Same node centrality features were then investigated from the revised semantic-based network and key health topics were again identified.

As a result, two ranking lists were built in terms of each of the three node centrality measurements – one was based on the original structural link network, and the other was based on the revised semantic-based network. The ranked health topics located on each ranking list were compared regarding their consistency. If a health topic was ranked high in the revised semantic-based network but low in the original structural link network, it indicated that the structural and semantic connections were not consistent in terms of possessing key health topics. Those impactful health topics derived from the revised semantic-based network should be regarded as the core topics among the recommended connections because they carried more important responsibilities within the whole network.

Table 18 displays the ranking lists of the degree centrality features between the original structural and revised semantic-based networks of the mental health topic subcategory. The three health topics that had the largest differences in terms of their rankings on the ranking lists are highlighted for both out-degree and in-degree centralities:

Out_Degree	Revised_Out_Degree	In_Degree	Revised_In_Degree
Nutrition	Child Nutrition	Nutrition	Nutrition
Weight Control	Nutrition	Mental Disorders	Mental Disorders
Mental Disorders	Child Mental Health	Teen Mental Health	**Child Nutrition**
Depression	Child Behavior Disorders	Weight Control	Child Development

Anxiety	Child Development	Depression	Child Behavior Disorders
Child Mental Health	**Obesity in Children**	Child Development	Child Mental Health
Teen Mental Health	Teen Mental Health	Drug Abuse	Teen Mental Health
Child Behavior Disorders	Weight Control	Diets	**Obesity in Children**
Underage Drinking	Teen Health	Teen Health	Weight Control
Teen Development	**Diets**	Drugs and Young People	Teen Health
Obesity	**Nutrition for Seniors**	Malnutrition	Diets
Child Nutrition	Teen Development	Child Behavior Disorders	Drugs and Young People
Mental Health	Child Abuse	Alcoholism and Alcohol Abuse	**Nutrition for Seniors**
Bullying	Child Sexual Abuse	Domestic Violence	Child Abuse
Child Development	Mental Disorders	Anxiety	Depression
Body Weight	Depression	Child Mental Health	Drug Abuse
Exercise and Physical Fitness	Underage Drinking	Obesity	Child Sexual Abuse
Teen Health	School Health	Bullying	School Health
Alcoholism and Alcohol Abuse	Exercise for Children	Body Weight	Exercise for Children
Drug Abuse	Anxiety	Smoking	College Health

Table 18. Degree centrality of the mental health subcategory

From Table 18, four different health topics were identified as key nodes in the network in terms of their in-degree centrality and out-degree centrality. The range for qualifying as "key nodes" in the network was determined to be the top 20 topics. This was not only because 20 was an appropriate size compared with the total number of selected health topics (99), but also because these 20 topics of each of the three centrality measures could provide an appropriate range for combining key nodes of different centrality ranking lists at later analysis stages. For the out-degree feature, *Obesity in Children*, *Diets*, and *Nutrition for Seniors* were not included in the

139

top 20 rankings of the original structural network. For the in-degree feature, *Child Nutrition*, *Obesity in Children*, and *Nutrition for Seniors* were not included in the top 20 rankings of the original structural network. However, according to the revised semantic-based network, these four health topics were linked to a few other topics through strong semantic relationships, hence they were serving as key nodes in the revised network. In other words, the importance of these four health topics, especially the two topics that occurred in both the in-degree and out-degree ranking lists (*Obesity in Children* and *Nutrition for Seniors*), was underestimated in the original subject directory of MedlinePlus portal.

Table 19 shows the ranking lists of the closeness centrality features between the original structural and revised semantic-based network of the mental health topic subcategory. The three health topics with the largest differences in terms of their rankings on the ranking lists are highlighted for both out-closeness and in-closeness centralities:

Out_Closeness	Revised_Out_Closeness	In_Closeness	Revised_In_Closeness
Eating Disorders	Underage Drinking	Mental Disorders	Child Mental Health
Weight Control	Child Nutrition	Teen Mental Health	Teen Mental Health
Exercise and Physical Fitness	Child Mental Health	Mental Health	**Child Nutrition**
Underage Drinking	**Child Development**	Child Mental Health	Child Behavior Disorders
Nutrition	**Obesity in Children**	Teen Health	Mental Disorders
Obesity	**Child Behavior Disorders**	Self-Harm	Child Development
Diets	Child Sexual Abuse	Bullying	Teen Health
Alcoholism and Alcohol Abuse	Teen Health	Depression	**Obesity in Children**
Body Weight	Child Abuse	Personality Disorders	Mental Health

140

Weight Loss Surgery	Teen Mental Health	Psychotic Disorders	**Child Sexual Abuse**
Drugs and Young People	Nutrition	Child Behavior Disorders	Child Abuse
Alcohol	Weight Control	Stress	College Health
Child Nutrition	Mental Disorders	Teen Violence	Teenage Pregnancy
Breastfeeding	Mental Health	College Health	Nutrition
Child Mental Health	Diets	School Health	Self-Harm
Teen Mental Health	Child Safety	Teen Development	Bullying
Smoking and Youth	College Health	Child Development	Teen Development
Infant and Newborn Nutrition	Nutrition for Seniors	Teen Sexual Health	Learning Disorders
Fetal Alcohol Spectrum Disorder	Fetal Alcohol Spectrum Disorder	Suicide	Child Safety
Mental Disorders	Eating Disorders	Bipolar Disorder	School Health

Table 19. Closeness centrality of the mental health subcategory

From Table 19, five different health topics were identified as key nodes in the network in terms of their in-closeness centrality and out-closeness centrality. Similarly, for the out-closeness feature, *Children Development*, *Obesity in Children*, and *Child Behavior Disorders* were not included in the top 20 rankings of the original structural network. For the in-closeness feature, *Child Nutrition*, *Obesity in Children*, and *Child Sexual Abuse* were not included in the top 20 rankings of the original structural network. However, according to the revised semantic-based network, these five health topics were linked to a few other topics through strong semantic relationships, hence they served as key nodes in the revised network. In other words, the importance of these five health topics, especially the topic *Obesity in Children* that occurred in both the in-closeness and out-closeness ranking lists, was underestimated in the original subject directory of MedlinePlus portal.

141

Table 20 shows the ranking list of the betweenness centrality feature between the original structural and revised semantic-based network of the mental health topic subcategory. The three health topics that had the largest differences in terms of their rankings on the ranking lists are highlighted for the betweenness centrality:

Betweenness	Revised_Betweenness
Mental Disorders	**Child Nutrition**
Teen Mental Health	Mental Disorders
Weight Control	Child Mental Health
Eating Disorders	Teen Mental Health
Nutrition	Nutrition
Child Behavior Disorders	Drugs and Young People
Child Mental Health	Child Development
Mental Health	Mental Health
Teen Health	Depression
Depression	Stress
Stress	Child Behavior Disorders
Drugs and Young People	Teen Health
Teen Violence	**Obesity in Children**
Bullying	**Diets**
Child Development	Exercise for Children
Domestic Violence	Weight Control
Anxiety	Post-Traumatic Stress Disorder
Attention Deficit Hyperactivity Disorder	Anxiety
Self-Harm	Underage Drinking

Table 20. Betweenness centrality of the mental health subcategory

According to the results shown in Table 20, three health topics were identified as key nodes in the network in terms of their betweenness centrality. To be more specific, *Child Nutrition*, *Obesity in Children*, and *Diets* were not included in the top 20 rankings of the original structural network. However, according to the revised semantic-based network, these three health topics were linked to a few other topics through strong semantic relationships, hence they served as key nodes in the revised network. In other words, the importance of these three health topics was underestimated in the original subject directory of MedlinePlus portal.

In conclusion, the node-level centrality measures among the health topics related to mental health were not consistent between the structural and semantic connections. A group of health topics that possessed strong semantic relationships were ignored in the original structural network. To better reflect the semantic connections in the mental health related subject directory, these key topics' connections, as well as the rest of the suggested connections, should be added to the structural link network. The key nodes identified included: *Obesity in Children*, *Diets*, *Nutrition for Seniors*, *Child Nutrition*, *Children Development*, *Child Behavior Disorders*, and *Child Sexual Abuse*.

4.2.2.2   Children related health topics

Following the same processes applied to the mental health subcategory, two ranking lists were built in terms of each of the three node centrality measurements – one was based on the original structural link network, and the other was based on the revised semantic-based network.

143

The ranked health topics located in each ranking list were compared regarding their consistency.

Table 21 displays the ranking lists of the degree centrality features between the original structural and revised semantic-based network of the children topic subcategory. The three health topics that had the largest differences in terms of their rankings on the ranking lists are highlighted for both out-degree and in-degree centralities:

Out_Degree	Revised_Out_Degree	In_Degree	Revised_In_Degree
Childhood Immunization	Child Development	Children's Health	Children's Health
Medicines	Medicines and Children	Childhood Immunization	**Medicines and Children**
Child Development	**Children's Health**	Medicines	**Toddler Health**
Medicines and Children	**Toddler Health**	Birth Defects	Medicines
Hearing Disorders and Deafness	**Child Mental Health**	Uncommon Infant and Newborn Problem	Uncommon Infant and Newborn Problems
Birth Defects	Child Dental Health	Hearing Disorders and Deafness	Child Development
Child Dental Health	Childhood Immunization	Teen Mental Health	Childhood Immunization
Dental Health	Teen Health	Sports Fitness	Teen Mental Health
Child Behavior Disorders	Child Behavior Disorders	Assistive Devices	Birth Defects
Sports Fitness	Infant and Newborn Care	Common Infant and Newborn Problems	**Child Mental Health**
Assistive Devices	Teen Mental Health	Child Development	Child Dental Health
Teen Mental Health	Child Nutrition	Dental Health	Infant and Newborn Care
Fetal Health and Development	Medicines	Over-the-Counter Medicines	Child Nutrition
Common Infant and Newborn Problems	Fetal Health and Development	Drug Safety	Common Infant and Newborn Problems
Teen Health	Uncommon Infant and Newborn Problems	Pregnancy	Teen Health
Uncommon Infant and Newborn Problems	Teen Development	Genetic Disorders	Child Behavior Disorders
Infant and Newborn Development	Birth Defects	Teen Health	Baby Health Checkup

144

Teen Development	Common Infant and Newborn Problems	Infant and Newborn Care	Hearing Disorders an Deafness
Infant and Newborn Care	Exercise for Children	Speech and Communication Disorders	Teenage Pregnancy
Immunization	Dental Health	Tooth Disorders	Fetal Health and Development

Table 21. Degree centrality of the children subcategory

From Table 21, four different health topics were identified as key nodes in the network in terms of their in-degree centrality and out-degree centrality. For the out-degree feature, *Children's Health*, *Toddler Health*, and *Child Mental Health* were not included in the top 20 rankings of the original structural network. For the in-degree feature, *Medicines and Children*, *Toddler Health*, and *Child Mental Health* were not included in the top 20 rankings of the original structural network. However, according to the revised semantic-based network, these four health topics were linked to a few other topics through strong semantic relationships, hence they served as key nodes in the revised network. In other words, the importance of these four health topics, especially the two topics that occurred in both the in-degree and out-degree ranking lists (*Toddler Health* and *Child Mental Health*), was underestimated in the original subject directory of MedlinePlus portal.

Table 22 shows the ranking lists of the closeness centrality features between the original structural and revised semantic-based network of the children topic subcategory. The three health topics with the largest differences in terms of their rankings on the ranking lists are highlighted for both out-closeness and in-closeness centralities:

Out_Closeness	Revised_Out_Closeness	In_Closeness	Revised_In_Closeness
Fetal Alcohol Spectrum Disorders	Childhood Immunization	Children's Health	Children's Health

145

Cold and Cough Medicines	Medicines and Children	Child Nutrition	Toddler Health
Childhood Immunization	**Child Development**	Uncommon Infant and Newborn Problems	Uncommon Infant and Newborn Problems
Medicines	**Children's Health**	Common Infant and Newborn Problems	**Medicines and Children**
Herbal Medicine	**Toddler Health**	Infant and Newborn Care	Child Development
Medicines and Children	Medicines	Baby Health Checkup	Child Nutrition
Growth Disorders	Infant and Newborn Development	Infant and Newborn Nutrition	**Child Mental Health**
Hepatitis B	Child Nutrition	Toddler Health	Baby Health Checkup
Whooping Cough	Fetal Health and Development	Toddler Nutrition	Common Infant and Newborn Problems
Down Syndrome	Child Mental Health	Birth Weight	Child Behavior Disorders
Dwarfism	Fetal Alcohol Spectrum Disorders	Child Safety	Child Safety
Cough	Child Behavior Disorders	Toddler Development	Toddler Development
Pregnancy and Medicines	Antibiotics	Dandruff, Cradle Cap, and Other Scalp Conditions	**Child Dental Health**
Immunization	Cold and Cough Medicines	Child Development	Infant and Newborn Development
Tetanus	Infant and Newborn Care	Fetal Health and Development	Medicines
Antibiotics	Baby Health Checkup	Teen Health	Fetal Health and Development
Common Cold	Birth Defects	School Health	Infant and Newborn Care
Complementary and Integrative Medicine	Teen Health	Infant and Newborn Development	Infant and Newborn Nutrition
Drug Safety	Toddler Development	Child Behavior Disorders	Teen Health
Flu Shot	Hearing Problems in Children	Speech and Language Problems	Teen Mental Health

Table 22. Closeness centrality of the children subcategory

From Table 22, six health topics were identified as key nodes in the network in terms of their in-closeness centrality and out-closeness centrality. For the out-closeness feature, *Children Development*, *Children's Health*, and *Toddler Health* were not included in the top 20 rankings of

the original structural network. For the in-closeness feature, *Medicines and Children*, *Child Mental Health*, and *Child Dental Health* were not included in the top 20 rankings of the original structural network. However, according to the revised semantic-based network, these six health topics were linked to a few other topics through strong semantic relationships, hence they served as key nodes in the revised network. In other words, the importance of these six health topics was underestimated in the original subject directory of MedlinePlus portal.

Table 23 shows the ranking list of the betweenness centrality feature between the original structural and revised semantic-based network of the children topic subcategory. The three health topics that had the largest differences in terms of their rankings on the ranking lists are highlighted for the betweenness centrality:

Betweenness	Revised_Betweenness
Birth Defects	Children's Health
Childhood Immunization	Childhood Immunization
Medicines	**Medicines and Children**
Fetal Health and Development	Child Development
Uncommon Infant and Newborn Problems	Medicines
Child Development	Fetal Health and Development
Pregnancy and Medicines	**Prenatal Testing**
Birth Weight	**Child Safety**
Assistive Devices	Birth Defects
Children's Health	Assistive Devices
Child Nutrition	Toddler Health
Toddler Development	Baby Health Checkup

147

Antibiotics	Antibiotics
Bacterial Infections	Hepatitis Testing
Haemophilus Infections	Child Dental Health
Speech and Language Problems in Children	Haemophilus Infections
Infant and Newborn Nutrition	Speech and Language Problems in Children
Toddler Health	Hearing Disorders and Deafness
Child Behavior Disorders	Sports Safety
Hearing Disorders and Deafness	Bacterial Infections

Table 23. Betweenness centrality of the children subcategory

According to the results shown in Table 23, three health topics were identified as key nodes in the network in terms of their betweenness centrality. To be more specific, *Medicines and Children*, *Prenatal Testing*, and *Child Safety* were not included in the top 20 rankings of the original structural network. However, according to the revised semantic-based network, these three health topics were linked to a few other topics through strong semantic relationships, hence they served as key nodes in the revised network. In other words, the importance of these three health topics was underestimated in the original subject directory of MedlinePlus portal.

In conclusion, the node-level centrality measures among the health topics related to children were not consistent between the structural and semantic connections. A group of health topics that possessed strong semantic relationships were ignored in the original structural network. To better reflect the semantic connections in the children related subject directory, these key topics' connections, as well as the rest of the suggested connections, should be added to the structural link network. The key nodes identified included: *Children's Health*, *Toddler Health*, *Child Mental Health*, *Medicines and Children*, *Children Development*, *Child Dental Health*,

148

Prenatal Testing, and *Child Safety*.

4.2.2.3   Teenagers related health topics

Similar to the prior two topic groups, two ranking lists were built in terms of each of the three node centrality measurements – one was based on the original structural link network, and the other was based on the revised semantic-based network. The ranked health topics located in each ranking list were compared regarding their consistency.

Table 24 displays the ranking lists of the degree centrality features between the original structural and revised semantic-based network of the teenagers topic subcategory. The three health topics that had the largest differences in terms of their rankings on the ranking lists are highlighted for both out-degree and in-degree centralities:

Out_Degree	Revised_Out_Degree	In_Degree	Revised_In_Degree
Health Problems in Pregnancy	Health Problems in Pregnancy	Pregnancy	Pregnancy
Pregnancy	Pregnancy	Mental Disorders	Mental Disorders
Prenatal Care	Pregnancy and Drug Use	Health Problems in Pregnancy	Health Problems in Pregnancy
Depression	Mental Disorders	Teen Mental Health	Pregnancy and Medicines
Pregnancy and Drug Use	Pregnancy and Medicines	Reproductive Hazards	**Pregnancy and Drug Use**
Teen Mental Health	**Teenage Pregnancy**	Depression	**Teenage Pregnancy**
Child Development	Teen Mental Health	Child Development	Teen Mental Health
Mental Disorders	**Infections and Pregnancy**	Pregnancy and Medicines	**Infections and Pregnancy**
Mental Health	Child Development	Alcohol Use Disorder (AUD)	Preconception Care
Child Behavior Disorders	**Miscarriage**	Drug Use and Addiction	Pregnancy and Nutrition
Fetal Health and Development	Pregnancy and Nutrition	Prenatal Care	Child Development

149

Pregnancy and Medicines	Preconception Care	Fetal Health and Development	Miscarriage
Fetal Alcohol Spectrum Disorder	Mental Health	Teen Health	HIV/AIDS and Pregnancy
HIV/AIDS	Childbirth Problems	Childbirth	Reproductive Hazards
Teen Development	HIV/AIDS and Pregnancy	Drugs and Young People	Ectopic Pregnancy
Teen Sexual Health	Ectopic Pregnancy	Teen Depression	Teen Depression
Underage Drinking	Teen Development	Mental Health	Mental Health
Menstruation	Child Mental Health	Child Behavior Disorders	Child Mental Health
Child Mental Health	Child Behavior Disorders	HIV/AIDS	Child Behavior Disorders
Alcohol Use Disorder (AUD)	Obsessive-Compulsive Disorder	Menstruation	Obsessive-Compulsive Disorder

Table 24. Degree centrality of the teenagers subcategory

From Table 24, four different health topics were identified as key nodes in the network in terms of their in-degree centrality and out-degree centrality. For the out-degree feature, *Teenage Pregnancy*, *Infections and Pregnancy*, and *Miscarriage* were not included in the top 20 rankings of the original structural network. For the in-degree feature, *Pregnancy and Drug Use*, *Teenage Pregnancy*, and *Infections and Pregnancy* were not included in the top 20 rankings of the original structural network. However, according to the revised semantic-based network, these four health topics were linked to a few other topics through strong semantic relationships, hence they served as key nodes in the revised network. In other words, the importance of these four health topics, especially the two topics that occurred in both the in-degree and out-degree ranking lists (*Teenage Pregnancy* and *Infections and Pregnancy*), was underestimated in the original subject directory of MedlinePlus portal.

Table 25 shows the ranking lists of the closeness centrality features between the original structural and revised semantic-based network of the teenagers topic subcategory. The three

150

health topics with the largest differences in terms of their rankings on the ranking lists are

highlighted for both out-closeness and in-closeness centralities:

Out_Closeness	Revised_Out_Closeness	In_Closeness	Revised_In_Closeness
Alcohol Use Disorder (AUD)	Teenage Pregnancy	Teen Health	Teen Mental Health
Fetal Alcohol Spectrum Disorder	Teen Mental Health	Teen Mental Health	Teenage Pregnancy
Underage Drinking	Teen Development	Teen Violence	Teen Health
HIV/AIDS and Pregnancy	Pregnancy and Drug Use	Teen Sexual Health	Teen Development
Health Problems in Pregnancy	Health Problems in Pregnancy	Child Behavior Disorders	Pregnancy
Alcohol	Teen Health	Teen Development	Child Mental Health
Pregnancy and Drug Use	Pregnancy	Reproductive Hazards	Child Development
Teenage Pregnancy	Child Mental Health	College Health	**Health Problems in Pregnancy**
Teen Sexual Health	**Pregnancy and Medicines**	Bullying and Cyberbullying	Mental Disorders
Pregnancy	Fetal Alcohol Spectrum Disorder	Teenage Pregnancy	**Pregnancy and Medicines**
Prenatal Care	**Infections and Pregnancy**	Child Development	**Pregnancy and Drug Use**
Teen Development	**Preconception Care**	Pregnancy	Mental Health
Mental Health	Miscarriage	Mental Disorders	Infections and Pregnancy
Impaired Driving	Pregnancy and Nutrition	Child Mental Health	Child Behavior Disorders
Teen Health	Child Development	Self-harm	Self-harm
Teen Mental Health	Mental Disorders	Mental Health	Preconception Care
Cirrhosis	Ectopic Pregnancy	Teen Depression	Pregnancy and Nutrition
Fatty Liver Disease	Teen Sexual Health	How to Prevent High Blood Pressure	Developmental Disabilities
Child Mental Health	Underage Drinking	School Health	Miscarriage
Mental Disorders	Prenatal Care	Puberty	Ectopic Pregnancy

Table 25. Closeness centrality of the teenagers subcategory

151

From Table 25, five different health topics were identified as key nodes in the network in terms of their in-closeness centrality and out-closeness centrality. For the out-closeness feature, *Pregnancy and Medicines*, *Infections and Pregnancy*, and *Preconception Care* were not included in the top 20 rankings of the original structural network. For the in-closeness feature, *Health Problems in Pregnancy*, *Pregnancy and Medicines*, and *Pregnancy and Drug Use* were not included in the top 20 rankings of the original structural network. However, according to the revised semantic-based network, these five health topics were linked to a few other topics through strong semantic relationships, hence they served as key nodes in the revised network. In other words, the importance of these five health topics, especially the topic *Pregnancy and Medicines* that occurred in both the in-closeness and out-closeness ranking lists, was underestimated in the original subject directory of MedlinePlus portal.

Table 26 shows the ranking list of the betweenness centrality feature between the original structural and revised semantic-based network of the teenagers topic subcategory. The three health topics that had the largest differences in terms of their rankings on the ranking lists are highlighted for the betweenness centrality:

Betweenness	Revised_Betweenness
Teen Health	Teen Mental Health
Teen Mental Health	Teenage Pregnancy
Health Problems in Pregnancy	Mental Disorders
Pregnancy	Child Development
Teen Sexual Health	Health Problems in Pregnancy
Teenage Pregnancy	Pregnancy and Drug Use

152

Child Development	Fetal Alcohol Spectrum Disorders
Mental Disorders	Pregnancy
Teen Violence	**Drugs and Young People**
Mental Health	**Fetal Health and Development**
Fetal Alcohol Spectrum Disorders	**HIV/AIDS and Pregnancy**
Pregnancy and Drug Use	Teen Development
Alcohol Use Disorder (AUD)	Child Mental Health
Child Behavior Disorders	Alcohol Use Disorder (AUD)
Stress	Infant and Newborn Development
Puberty	Teen Sexual Health
Reproductive Hazards	Teen Health
Depression	Growth Disorders
Teen Development	Mental Health
High Blood Pressure in Pregnancy	Developmental Disabilities

Table 26. Betweenness centrality of the teenagers subcategory

According to the results shown in Table 26, three health topics were identified as key nodes in the network in terms of their betweenness centrality. To be more specific, *Drugs and Young People*, *Fetal Health and Development*, and *HIV/AIDS and Pregnancy* were not included in the top 20 rankings of the original structural network. However, according to the revised semantic-based network, these three health topics were linked to a few other topics through strong semantic relationships, hence they served as key nodes in the revised network. In other words, the importance of these three health topics was underestimated in the original subject directory of MedlinePlus portal.

153

In conclusion, the node-level centrality measures among the health topics related to teenagers were not consistent between the structural and semantic connections. A group of health topics that possessed strong semantic relationships were ignored in the original structural network. To better reflect the semantic connections in the teenagers related subject directory, these key topics' connections, as well as the rest of the suggested connections, should be added to the structural link network. The key nodes identified included: *Teenage Pregnancy*, *Infections and Pregnancy*, *Miscarriage*, *Pregnancy and Drug Use*, *Pregnancy and Medicines*, *Preconception Care*, *Health Problems in Pregnancy*, *Drugs and Young People*, *Fetal Health and Development*, and *HIV/AIDS and Pregnancy*.

4.2.2.4 Older adults related health topics

Following the same procedures applied to the other three subcategories, two ranking lists were built in terms of each of the three node centrality measurements – one was based on the original structural link network, and the other was based on the revised semantic-based network. The ranked health topics located in each ranking list were compared regarding their consistency.

Table 27 displays the ranking lists of the degree centrality features between the original structural and revised semantic-based network of the older adults topic subcategory. The three health topics that had the largest differences in terms of their rankings on the ranking lists are highlighted for both out-degree and in-degree centralities:

Out_Degree	Revised_Out_Degree	In_Degree	Revised_In_Degree
Nutrition	Nutrition	Nutrition	Nutrition
Cholesterol	Mental Disorders	Mental Disorders	Mental Disorders

154

Vitamins	Vitamins	Vitamins	Vitamins
Dietary Fats	Child Nutrition	Older Adult Health	Teen Mental Health
Weight Control	**Teen Mental Health**	Cholesterol	Older Adult Health
Triglycerides	Cholesterol	Cholesterol Levels: What You Need to Know	How to Lower Cholesterol with Diets
Depression	**Child Mental Health**	LDL: The "Bad" Cholesterol	**Child Nutrition**
Heart Diseases	**Healthy Aging**	Teen Mental Health	Cholesterol
How to Lower Cholesterol	Mental Health	HDL: The "Good" Cholesterol	**Child Mental Health**
Sports Fitness	Nutrition for Older Adults	Sports Injuries	Cholesterol Levels: What You Need to Know
Child Nutrition	Dietary Fats	Malnutrition	HDL: The "Good" Cholesterol
Mental Disorders	Weight Control	Depression	LDL: The "Bad" Cholesterol
Cholesterol Medicines	How to Lower Cholesterol with Diets	Sports Fitness	**Healthy Aging**
Caregivers	Vitamin D	How to Lower Cholesterol with Diets	Mental Health
End of Life Issues	Exercise for Older Adults	Weight Control	How to Lower Cholesterol
Exercise for Older Adults	Older Adult Mental Health	How to Lower Cholesterol	Diets
Nutrition for Older Adults	Managed Care	Diets	Sports Injuries
Suicide	Heart Diseases	Exercise and Physical Fitness	VLDL Cholesterol
Mental Health	How to Lower Cholesterol	VLDL Cholesterol	Cholesterol Medicines
Calcium	Triglycerides	Minerals	Malnutrition

Table 27. Degree centrality of the older adults subcategory

From Table 27, four different health topics were identified as key nodes in the network in terms of their in-degree centrality and out-degree centrality. For the out-degree feature, *Teen Mental Health*, *Child Mental Health*, and *Healthy Aging* were not included in the top 20 rankings of the original structural network. For the in-degree feature, *Child Nutrition*, *Child Mental Health*, and *Healthy Aging* were not included in the top 20 rankings of the original structural

155

network. However, according to the revised semantic-based network, these four health topics were linked to a few other topics through strong semantic relationships, hence they served as key nodes in the revised network. In other words, the importance of these four health topics, especially the two topics that occurred in both the in-degree and out-degree ranking lists (*Child Mental Health* and *Healthy Aging*), was underestimated in the original subject directory of MedlinePlus portal.

Table 28 shows the ranking lists of the closeness centrality features between the original structural and revised semantic-based network of the older adults topic subcategory. The three health topics with the largest differences in terms of their rankings on the ranking lists are highlighted for both out-closeness and in-closeness centralities:

Out_Closeness	Revised_Out_Closeness	In_Closeness	Revised_In_Closeness
Eating Disorders	Nutrition	Older Adult Health	Nutrition
Caregivers	**Healthy Aging**	Nutrition for Older Adults	Healthy Aging
Home Care Services	**Child Nutrition**	Nutrition	Nutrition for Older Adults
Assisted Living	Nutrition for Older Adults	Healthy Aging	Older Adult Health
End of Life Issues	Mental Health	Diets	**Child Nutrition**
Suicide	**Child Mental Health**	Malnutrition	**Mental Health**
Advance Directives	Child Care	Older Adult Mental Health	Older Adult Mental Health
Nursing Homes	Older Adult Mental Health	Exercise for Older Adults	Diets
Weight Control	Mental Disorders	Weight Control	**Child Mental Health**
Child Care	Teen Mental Health	Mental Disorders	Weight Control
Nutrition	Eating Disorders	Vitamins	Teen Mental Health
Child Development	Health Insurance	Dietary Proteins	Malnutrition

Exercise and Physical Fitness	Weight Control	Cholesterol	Exercise for Older Adults
Older Adult Health	How to Improve Mental Health	Cholesterol Levels: What You Need to Know	Health Insurance
Older Adult Mental Health	Infant and Newborn Nutrition	How to Lower Cholesterol with Diets	How to Improve Mental Health
Organ Donation	Children's Health	LDL: The "Bad" Cholesterol	Children's Health
Nutrition for Older Adults	Older Adult Health	HDL: The "Good" Cholesterol	Cholesterol
Family Issues	Toddler Nutrition	Dietary Fats	Cholesterol Levels: What You Need to Know
Telehealth	Child Behavior Disorders	Minerals	HDL: The "Good" Cholesterol
Mental Health	Exercise for Older Adults	Carbohydrates	LDL: The "Bad" Cholesterol

Table 28. Closeness centrality of the older adults subcategory

From Table 28, four different health topics were identified as key nodes in the network in terms of their in-closeness centrality and out-closeness centrality. For the out-closeness feature, *Healthy Aging*, *Child Nutrition*, and *Child Mental Health* were not included in the top 20 rankings of the original structural network. For the in-closeness feature, *Child Nutrition*, *Mental Health*, and *Child Mental Health* were not included in the top 20 rankings of the original structural network. However, according to the revised semantic-based network, these four health topics were linked to a few other topics through strong semantic relationships, hence they served as key nodes in the revised network. In other words, the importance of these four health topics, especially the two topics that occurred in both the in-closeness and out-closeness ranking lists (*Child Nutrition* and *Child Mental Health*), was underestimated in the original subject directory of MedlinePlus portal.

Table 29 shows the ranking lists of the betweenness centrality feature between the original structural and revised semantic-based network of the older adults topic subcategory. The

157

three health topics that had the largest differences in terms of their rankings on the ranking lists were highlighted for the betweenness centrality:

Betweenness	Revised_Betweenness
Nutrition	Nutrition
Older Adult Health	**Healthy Aging**
Mental Disorders	Mental Disorders
Weight Control	Vitamins
Nutrition for Older Adults	Exercise for Older Adults
Depression	**Child Nutrition**
End of Life Issues	Sports Fitness
Eating Disorders	Older Adult Health
Suicide	Caregivers
Mental Health	**Child Care**
Older Adult Mental Health	Teen Mental Health
Vitamins	Weight Control
Sports Fitness	Sports Injuries
Exercise for Older Adults	Eating Disorders
Exercise and Physical Fitness	Health Insurance
Patient Rights	Cholesterol
Caregivers	End of Life Issues
Home Care Services	Child Mental Health
Talking with Your Doctor	Managed Care
Sports Injuries	Nutrition for Older Adults

Table 29. Betweenness centrality of the older adults subcategory

158

According to the results shown in Table 29, three health topics were identified as key nodes in the network in terms of their betweenness centrality. To be more specific, *Healthy Aging*, *Child Nutrition*, and *Child Care* were not included in the top 20 rankings of the original structural network. However, according to the revised semantic-based network, these three health topics were linked to a few other topics through strong semantic relationships, hence they served as key nodes in the revised network. In other words, the importance of these three health topics was underestimated in the original subject directory of MedlinePlus portal.

In conclusion, the node-level centrality measures among the health topics related to older adults were not consistent between the structural and semantic connections. A group of health topics that possessed strong semantic relationships were ignored in the original structural network. To better reflect the semantic connections in the older adults related subject directory, these key topics' connections, as well as the rest of the suggested connections, should be added to the structural link network. The key nodes identified included: *Teen Mental Health*, *Child Mental Health*, *Healthy Aging*, *Child Nutrition*, *Mental Health*, and *Child Care*.

4.2.3   Summary

The first research question concerned the investigation of the current subject directory applied by the MedlinePlus portal. This research question covered all the four health topic subcategories and revealed facts about inconsistencies hidden behind the structural and semantic connections. The results generated for this question suggested additional research questions for further optimization and evaluation.

Specifically, RQ1.1 compared the connections from both the structural and semantic

159

perspectives and identified a list of recommended connections for each of the four health topic subcategories. Following that, RQ1.2 identified the core topics that were previously underestimated in the structural link network.

Through a series of descriptive statistical analysis, along with various comparisons and observations, the current structural link networks relating to mental health, children, teenagers, and older adults applied on the MedlinePlus portal were found to be not consistent with their semantic relationships among the involved health topics. Portal users were not able to navigate between hundreds of pairs of health topics possessing similar Web page content. Furthermore, among these missed connections, a few health topics were found to have great impact within the whole network in terms of linking other topics through semantic connections. The control and responsibilities taken by these hidden core health topics were underestimated.

These hidden core health topics within each topic subcategory might not be the most influential ones in the network, but they indicated a weak structural connection and the portal creators should be made aware of them. Moreover, some hidden core health topics were found to be commonly shared among multiple health topic groups. For instance, children related health topics were found to be underestimated in not only the children subcategory, but also the mental health subcategory and older adults subcategory. This indicated that these children related health topics were playing a more important role than the portal creators thought in the whole subject directory system on the MedlinePlus portal. Adding connections to these health topics would not only improve the structural network of the children subcategory, but also other related subcategories like mental health and older adults.

160

4.3 Findings for research question 2 (RQ2)

RQ2: Are there significant differences between the original mental health subject directory and the optimized mental health subject directory in terms of its link structure on the MedlinePlus portal?

As discussed in the above section, the structural and semantic connections were not consistent regarding mental health related health topics in the subject directory of MedlinePlus. Recommendations were made for a group of health topics in order to add more relevant connections that had high similarity values in terms of Web page textual information. Moreover, some hidden core health topics were detected through the comparison of the ranking lists between the structural and semantic network. These health topics were identified through the analysis of node-level centrality features. As a result, an optimized structural link network was developed. It is important to ensure that this optimized subject directory works better than the original one in terms of organizing related health topics from both the structural and semantic perspectives.

4.3.1   Introduction of the optimized structural network

4.3.1.1   Description of the structural connections in the optimized network

This section aims at providing an overview of the optimized subject directory regarding the revised structural connections. As introduced in the above section, there were 260 connections in the original structural link network. After conducting a calculation toward the average similarity values held among all the selected health topics, the threshold was set according to the average number of structurally connected topics. The threshold was 0.383677.

Following that, all the unconnected topics were investigated to find pairs of health topics whose similarity values were higher than the threshold. As a result, 133 connections were found and among them, 110 were bidirectional connections while 23 were unidirectional connections. All these 133 connections were then added to the original structural link network to form an optimized subject directory. This optimized subject directory is shown in Figure 18.
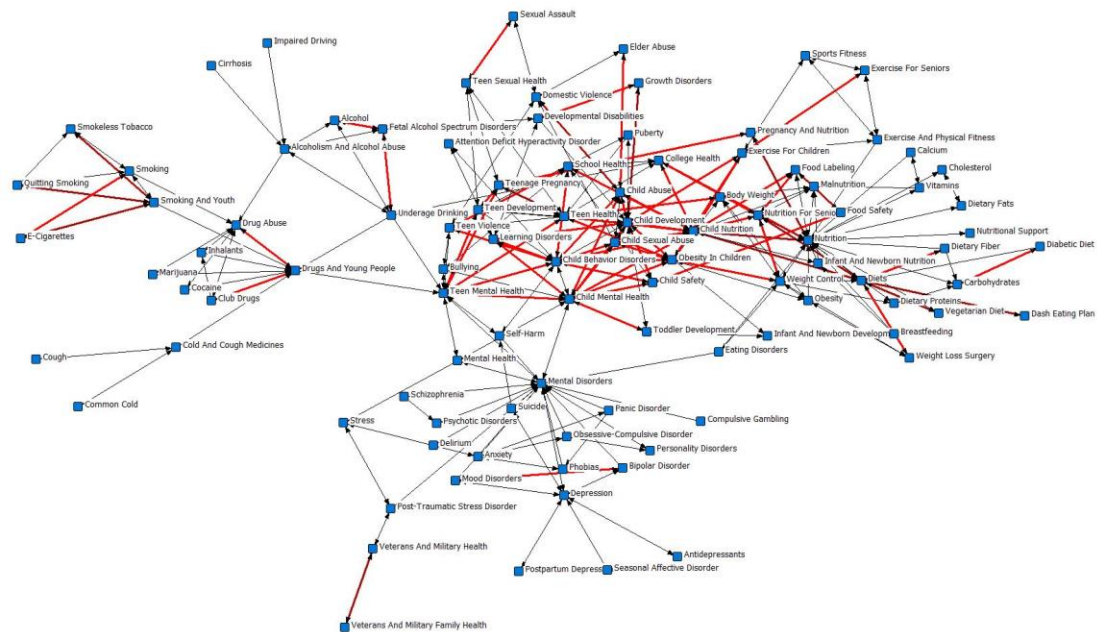
Figure 18. Optimized mental health structural link network

In Figure 18, the original structural connections are indicated with black lines while the recommended connections are in red lines. According to the figure, it is clear to find that most of the recommended connections were added to the health consumer groups (i.e., children and teenagers) related health topics. Newly suggested connections were also found in the daily health element group. Compared with those two groups of health topics, specific mental disease related topics only had two new connections in the optimized subject directory, which indicated that this

162

section of health topics had already been well-developed.

4.3.1.2   Description of the semantic connections of the optimized network

This section identifies the new semantic relationships of the optimized mental health subject directory. For mental health related topics, the overall average similarity value among all the 9,702 potential connections stayed the same at 0.105879 after the optimization process. However, since the structured connections in the optimized subject directory combined both Edge Set B and Edge Set D, there were 393 connections in total as a result, thus leading the average similarity value to increase to 0.421036. The average similarity among the rest of the 9,309 pairs of topics (new Edge Set C) was decreased to 0.092574.

4.3.1.3   Identification of the key nodes

This section tends to reveal specific hidden core health topics in the original structural link network. In the previous stage, several core health topics were identified based on the analysis results of node-level centrality measures. To further filter these hidden core health topics, each topic was assigned with a weighted score if it appeared to be the health topic that had the largest difference in terms of the rankings between the original structural link network and the revised semantic-based network. Also, these health topics were required to be ranked in the top 20 of the corresponding node centrality features in the revised semantic-based network.

For instance, Topic A was listed as the first topic in the betweenness centrality list of the revised semantic-based network but was listed as the thirtieth topic in the same node centrality measure list of the original structural link network; Topic B was listed as the fifth topic in the betweenness centrality list of the revised semantic-based network but was listed as the twenty-

163

fifth topic on the same node centrality measure list of the original structural link network; Topic C was listed as the twenty-fifth topic in the betweenness centrality list of the revised semantic-based network but was listed as the sixty-fifth topic in the same node centrality measure list of the original structural link network. As a result, Topic A and Topic B would be selected as more impactful topics within the subject directory, and Topic A would be given a weighted score of 1 while Topic B would be given a weighted score of 2. For all the key nodes identified in the previous stage, these topics were gathered together and the five topics possessing the lowest weighted scores were kept as the final group of core health topics.

For the mental health subcategory, the five core health topics identified were: *Obesity in Children*, *Child Nutrition*, *Diets*, *Child Development*, and *Nutrition for Seniors*. Among them, *Obesity in Children* had 14 recommended connections (12 bidirectional and 2 unidirectional); *Child Nutrition* had 25 recommended connections (22 bidirectional and 3 unidirectional); *Diets* had 8 recommended connections (4 bidirectional and 4 unidirectional); *Child Development* had 17 recommended connections (16 bidirectional and 1 unidirectional); and *Nutrition for Seniors* had 11 recommended connections (10 bidirectional and 1 unidirectional). Combining all connections relating to these five core topics, there were 66 suggested links in total, excluding the overlapped links among the five core topics. These 66 suggested links weighted 49.62% of all the recommended links, hence could prove the important roles played by the five core topics. The full list of suggested links related to the five core health topics could be found in Appendix 13.

The connections of the five core health topics were later added to the original structural network to form an adjusted structural network. The number of total connections included in this

164

adjusted network was 326. Node-level centrality measures were then analyzed from the adjusted structural network in order to investigate how the five core health topics might affect the node centrality measures within the network. The ranking lists of the three node centrality features are displayed in Tables 30-34. The five core topics are shown in red in the revised semantic-based network's ranking lists. In the adjusted structural network, the five core topics are shown in red bold font.

Out_Degree	Revised_Out_Degree	Adjusted_Out_Degree
Nutrition	*Child Nutrition*	***Child Nutrition***
Weight Control	*Nutrition*	***Child Development***
Mental Disorders	*Child Mental Health*	*Nutrition*
Depression	*Child Behavior Disorders*	***Obesity in Children***
Anxiety	*Child Development*	*Weight Control*
Child Mental Health	*Obesity in Children*	*Child Mental Health*
Teen Mental Health	*Teen Mental Health*	***Diets***
Child Behavior Disorders	*Weight Control*	***Nutrition for Seniors***
Underage Drinking	*Teen Health*	*Child Behavior Disorders*
Teen Development	*Diets*	*Mental Disorders*
Obesity	*Nutrition for Seniors*	*Depression*
Child Nutrition	*Teen Development*	*Teen Health*
Mental Health	*Child Abuse*	*Anxiety*
Bullying	*Child Sexual Abuse*	*Teen Mental Health*
Child Development	*Mental Disorders*	*Underage Drinking*
Body Weight	*Depression*	*Teen Development*

165

Exercise and Physical Fitness	Underage Drinking	Obesity
Teen Health	School Health	Exercise for Children
Alcoholism and Alcohol Abuse	Exercise for Children	Child Sexual Abuse
Drug Abuse	Anxiety	Mental Health

Table 30. Out-degree rankings among the three networks

In_Degree	Revised_In_Degree	Adjusted_In_Degree
Nutrition	Nutrition	Nutrition
Mental Disorders	Mental Disorders	Mental Disorders
Teen Mental Health	*Child Nutrition*	***Child Nutrition***
Weight Control	*Child Development*	***Child Development***
Depression	Child Behavior Disorders	***Obesity in Children***
Child Development	Child Mental Health	Weight Control
Drug Abuse	Teen Mental Health	***Diets***
Diets	*Obesity in Children*	Teen Health
Teen Health	Weight Control	Teen Mental Health
Drugs and Young People	Teen Health	Child Mental Health
Malnutrition	*Diets*	***Nutrition for Seniors***
Child Behavior Disorders	Drugs and Young People	Child Behavior Disorders
Alcoholism and Alcohol Abuse	*Nutrition for Seniors*	Depression
Domestic Violence	Child Abuse	Drug Abuse
Anxiety	Depression	Drugs and Young People
Child Mental Health	Drug Abuse	Malnutrition

Obesity	*Child Sexual Abuse*	*Exercise for Children*
Bullying	*School Health*	*Alcoholism and Alcohol Abuse*
Body Weight	*Exercise for Children*	*College Health*
Smoking	*College Health*	*Domestic Violence*

Table 31. In-degree rankings among the three networks

Out_Closeness	**Revised_Out_Closeness**	**Adjusted_Out_Closeness**
Eating Disorders	*Underage Drinking*	***Child Nutrition***
Weight Control	*Child Nutrition*	*Child Mental Health*
Exercise and Physical Fitness	*Child Mental Health*	*Underage Drinking*
Underage Drinking	*Child Development*	***Obesity in Children***
Nutrition	*Obesity in Children*	***Child Development***
Obesity	*Child Behavior Disorders*	*Child Behavior Disorders*
Diets	*Child Sexual Abuse*	*Weight Control*
Alcoholism and Alcohol Abuse	*Teen Health*	*Nutrition*
Body Weight	*Child Abuse*	***Diets***
Weight Loss Surgery	*Teen Mental Health*	***Nutrition for Seniors***
Drugs and Young People	*Nutrition*	*Child Sexual Abuse*
Alcohol	*Weight Control*	*Teen Health*
Child Nutrition	*Mental Disorders*	*Exercise for Children*
Breastfeeding	*Mental Health*	*Mental Disorders*
Child Mental Health	*Diets*	*Eating Disorders*
Teen Mental Health	*Child Safety*	*Mental Health*

167

Smoking and Youth	*College Health*	*Child Abuse*
Infant and Newborn Nutrition	*Nutrition for Seniors*	*Teen Mental Health*
Fetal Alcohol Spectrum Disorder	*Fetal Alcohol Spectrum Disorder*	*Obesity*
Mental Disorders	*Eating Disorders*	*College Health*

Table 32. Out-closeness rankings among the three networks

In_Closeness	**Revised_In_Closeness**	**Adjusted_In_Closeness**
Mental Disorders	*Child Mental Health*	***Child Nutrition***
Teen Mental Health	*Teen Mental Health*	*Child Mental Health*
Mental Health	*Child Nutrition*	*Teen Mental Health*
Child Mental Health	*Child Behavior Disorders*	*Mental Disorders*
Teen Health	*Mental Disorders*	*Teen Health*
Self-Harm	*Child Development*	***Obesity in Children***
Bullying	*Teen Health*	*Child Behavior Disorders*
Depression	*Obesity in Children*	***Child Development***
Personality Disorders	*Mental Health*	*Mental Health*
Psychotic Disorders	*Child Sexual Abuse*	*Bullying*
Child Behavior Disorders	*Child Abuse*	*Self-Harm*
Stress	*College Health*	*Nutrition*
Teen Violence	*Teenage Pregnancy*	*College Health*
College Health	*Nutrition*	***Nutrition for Seniors***
School Health	*Self-Harm*	*School Health*
Teen Development	*Bullying*	*Weight Control*

168

Child Development	Teen Development	*Diets*
Teen Sexual Health	Learning Disorders	Malnutrition
Suicide	Child Safety	Exercise for Children
Bipolar Disorder	School Health	Food Labeling

Table 33. In-closeness rankings among the three networks

Betweeness	Revised_Betweeness	Adjusted_Betweeness
Mental Disorders	Child Nutrition	**Child Nutrition**
Teen Mental Health	Mental Disorders	Mental Disorders
Weight Control	Child Mental Health	Child Mental Health
Eating Disorders	Teen Mental Health	Teen Mental Health
Nutrition	Nutrition	Nutrition
Child Behavior Disorders	Drugs and Young People	**Child Development**
Child Mental Health	Child Development	Teen Health
Mental Health	Mental Health	Drugs and Young People
Teen Health	Depression	**Obesity in Children**
Depression	Stress	Depression
Stress	Child Behavior Disorders	Mental Health
Drugs and Young People	Teen Health	Stress
Teen Violence	Obesity in Children	**Diets**
Bullying	Diets	Weight Control
Child Development	Exercise for Children	Exercise for Children
Domestic Violence	Weight Control	Child Behavior Disorders

169

Anxiety	Post-Traumatic Stress Disorder	Teen Violence
Attention Deficit Hyperactivity Disorder	Anxiety	Domestic Violence
Self-Harm	Underage Drinking	Anxiety
Post-Traumatic Stress Disorder	College Health	Smoking and Youth

Table 34. Betweenness rankings among the three networks

From the ranking lists displayed in Tables 30-34 regarding the three node-level centrality measures among the original structural network, the revised semantic-based network, and the adjusted core topic-based network, it is clear to see that the five core topics improved to higher rankings and these changes led the node centrality ranking lists of the adjusted network to become more consistent with the ranking lists of the revised semantic-based network. In other words, identifying core health topics might greatly assist a subject directory in better reflecting its topics' semantic relationships.

4.3.2   RQ2.1 & Hypothesis group 1

RQ2.1 is "Are there any significant differences between the original and optimized structural networks of mental health related topics in terms of similarity values?" It examines if similarity value differences appeared after the optimization process.

The similarity values were not able to be analyzed through the standard T-test since they did not have a normal distribution. Therefore, the Mann-Whitney test was applied to investigate the differences of similarity values between the original and optimized subject directories. SPSS was employed to conduct the inferential statistical analysis for the mental health related topics.

RQ2.1 was answered by hypothesis group 1. This hypothesis group contained only one

170

hypothesis. The Mann-Whitney test was applied to test the hypothesis in hypothesis group 1. Table 35 summarizes the mean rank as well as the sum of ranks of the similarity values from the original structural network (n=260) and the optimized structural network (n=393). The statistical analysis revealed that there was a systematic difference of similarity values between the original and optimized structural networks of mental health related topics (p=0.002<0.05). Therefore, the hypothesis was rejected. In other words, the similarity value of the connections was significantly improved after the optimization process.

	Similarity value	
	Original structural network	*Optimized structural network*
Mean rank	*298.26*	*346.01*
Sum of ranks	*77548.50*	*135982.50*
z-statistic	*-3.166*	
p-value	*0.002*	

Table 35. Statistical analysis result for hypothesis group 1

4.3.3   *RQ2.2 & Hypothesis group 2*

RQ2.2 is "Are there any significant differences between the original and optimized structural networks of mental health related topics in terms of node-level centrality measurements?" The aim of this question is to determine if node-level centrality measures differences occurred between the original and optimized structural network after the optimization process.

Node-level centrality measures were not able to be analyzed through standard tests since they did not have a normal distribution. However, UciNet provides a series of permutation tests (also called randomization tests) to modify standard methods (Borgatti, Everett, & Johnson, 2013). A series of customized T-tests were applied to investigate the differences of three node

171

centrality features between the original and optimized subject directories.

RQ2.2 was answered by hypothesis group 2, and this hypothesis group contained three sub-hypotheses. Each sub-hypothesis under hypothesis group 2 was aimed at comparing one of the three node-level centrality measures (degree centrality, closeness centrality, and betweenness centrality) between the original and optimized structures.

A series of customized T-tests were applied to test the three sub-hypotheses under hypothesis group 2. Tables 36-38 include the means and standard deviations (SD) of degree centrality, closeness centrality, and betweenness centrality of all original structural connections (n=99) and all optimized structural connections (n=99). In Table 36, the p-values of both the out-degree centrality and in-degree centrality were smaller than the significant level (0.05), which indicated that there was a significant difference of degree centrality measures between the original structural network and the optimized structural network. That is, the degree centrality measures in the optimized structural network had generated a significantly higher average value than the original structural network.

	Out-Degree		In-Degree	
	Original structural network	*Optimized structural network*	*Original structural network*	*Optimized structural network*
Mean	*0.027*	*0.040*	*0.027*	*0.040*
Standard deviation (SD)	*0.017*	*0.032*	*0.032*	*0.041*
t-statistic	*-0.014*		*-0.014*	
p-value	*0.0002*		*0.0077*	

Table 36. Degree centrality rankings between the original and optimized networks

In Table 37, the p-values of both the out-closeness centrality and in-closeness centrality were smaller than the significant level (0.05), which indicated that there was a significant

172

difference of closeness centrality measures between the original structural network and the optimized structural network. That is, the closeness centrality measures in the optimized structural network had generated a significantly higher average value than the original structural network.

	Out-Closeness		In-Closeness	
	Original structural network	Optimized structural network	Original structural network	Optimized structural network
Mean	0.136	0.203	0.152	0.238
Standard deviation (SD)	0.007	0.020	0.057	0.086
t-statistic	-0.067		-0.085	
p-value	0.0001		0.0001	

Table 37. Closeness centrality rankings between the original and optimized networks

In Table 38, unlike the degree and closeness centrality measures, the p-value of the betweenness centrality was larger than the significant level (0.05), which indicated that there was no significant difference of the betweenness centrality measure between the original structural network and the optimized structural network. That is, the betweenness centrality measure in the optimized structural network had not generated a significantly higher average value than the original structural network. A few causes could have caused this, with one of them being that the betweenness centrality measure only tests how a node is connecting two other nodes within the network. For a node that serving as a starting point or ending point in the network, since the connections are all gathered on that node, the betweenness centrality measure might not recognize this node as an impactful one since it is not connecting other nodes on both sides. Another reason might be that since more connections were added to the structural network, more pairs of health topics had been linked directly. Hence, the former bridging topics were not needed any more, thus leading the betweenness centrality to decrease.

173

	Betweenness	
	Original structural network	*Optimized structural network*
Mean	*1.743*	*2.146*
Standard deviation (SD)	*3.795*	*4.469*
t-statistic	*-0.403*	
p-value	*0.4975*	

Table 38. Betweenness centrality rankings between the original and optimized networks

As a result, for hypothesis group 2, H02$_{(a)}$ and H02$_{(b)}$ were rejected while H02$_{(c)}$ failed to be rejected.

4.3.4   Summary

The second research question concentrated on comparing the original structural network and the optimized structural network of mental health related health topics in the subject directory on MedlinePlus. The introductory section examined the optimized structural network in terms of structural connections, semantic connections, and key nodes in the network, while RQ2.1 and RQ2.2 investigated the differences between the original and optimized structural networks regarding their connections' similarity values as well as node-level centrality measures.

It was found that the recommended new structural connections were mostly added to the children and teenagers related health topics, and besides the health consumer groups, connections were also found in daily health element related topics. The specific mental diseases related health topics were well developed. Moreover, the average similarity value of the structural connections in the optimized network was improved after new connections were added. This difference was proved to be significant through the Mann-Whitney test. Interestingly, core health topics were identified to have a huge impact on the whole network. Through the creation of key nodes' connections, the node-level centrality ranking lists in the adjusted structural network moved

174

closer to the semantic-based network. In other words, key influential nodes were more capable of reflecting the semantic relationships in the network compared with normal nodes. In addition, the node centrality measures were tested through a customized T-test, the results showed that the degree centrality and closeness centrality measures were significantly increased in the optimized structural network while the betweenness centrality measure was not significantly different from the original structural link network.

4.4 Findings for research question 3 (RQ3)

RQ3: Are there significant differences between the original children subject directory and the optimized children subject directory in terms of its link structure on the MedlinePlus portal?

As discussed in section 4.2, the structural and semantic connections were not consistent regarding children related health topics in the subject directory of MedlinePlus. Recommendations were made for a group of health topics in order to add more relevant connections that had high similarity values in terms of Web page textual information. Moreover, some hidden core health topics were detected through the comparison of the ranking lists between the structural and semantic network. These health topics were identified through the analysis of node-level centrality features. As a result, an optimized structural link network was developed. It is important to ensure that this optimized subject directory works better than the original one in terms of organizing related health topics from both the structural and semantic perspectives.

4.4.1 *Introduction of the optimized structural network*

4.4.1.1 Description of the structural connections in the optimized network

This section aims at providing an overview of the optimized subject directory regarding the revised structural connections. As introduced in section 4.2, there were 472 connections in the original structural link network. After conducting a calculation toward the average similarity values held among all the selected health topics, the threshold was set according to the average number of structurally connected topics. The threshold was 0.444744. Following that, all the unconnected topics were investigated to find pairs of health topics whose similarity values were higher than the threshold. As a result, 184 connections were found and among them, 140 were bidirectional connections while 44 were unidirectional connections. All these 184 connections were then added to the original structural link network to form an optimized subject directory. This optimized subject directory is shown in Figure 19.
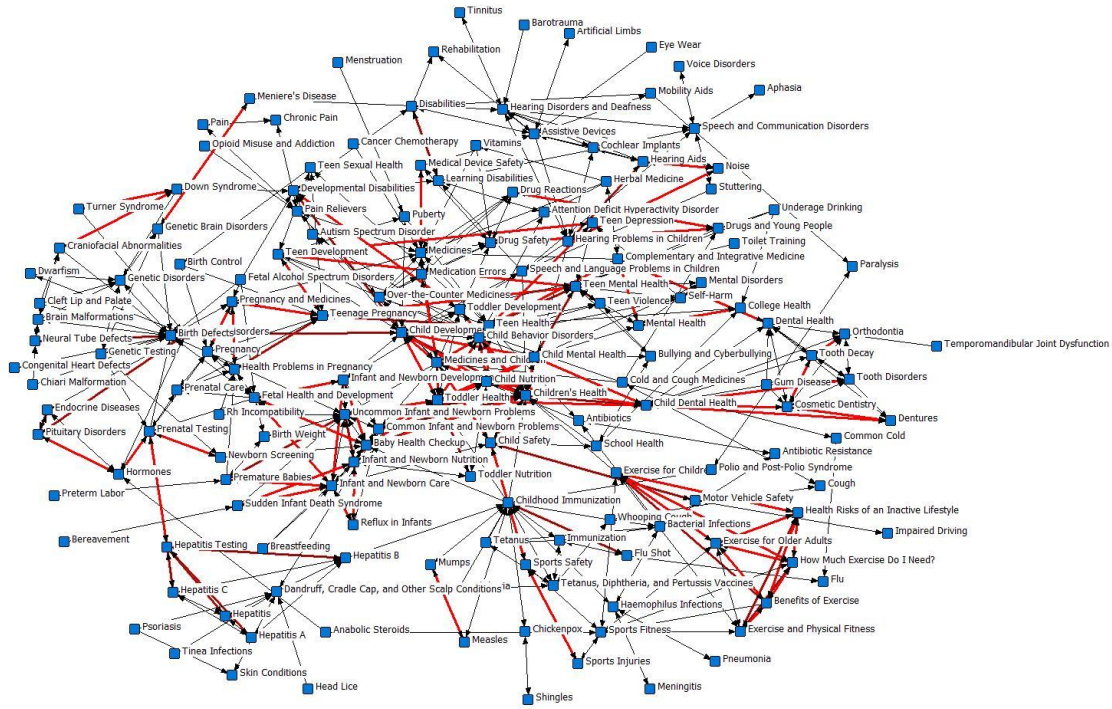
Figure 19. Optimized children structural link network

In Figure 19, the original structural connections are shown in black lines while the recommended connections are in red lines. Compared with the optimized structural network of mental health related health topics, recommended connections were more evenly distributed in children's network. Similar to the mental health's optimized network, it is clear to find that a lot of recommended connections were added to the health consumer groups (i.e., children, teenagers, toddlers, infants, older adults etc.) related health topics. Moreover, newly suggested connections were also found in the health topics relating to pregnancy, hepatitis and daily exercises.

4.4.1.2   Description of the semantic connections of the optimized network

This section identifies the new semantic relationships of the optimized children subject

177

directory. For children related topics, the overall average similarity value among all the 25,122 potential connections stayed the same at 0.093318 after the optimization process. However, since the structured connections in the optimized subject directory combined both Edge Set B and Edge Set D, there were 656 connections in total as a result, thus leading the average similarity value to increase to 0.476088. The average similarity among the rest of the 24,466 pairs of topics (new Edge Set C) was decreased to 0.083055.

4.4.1.3   Identification of the key nodes

This section tends to reveal specific hidden core health topics in the original structural link network. In the previous stage, several core health topics were identified based on the analysis results of node-level centrality measures. To further filter these hidden core health topics, each topic was assigned with a weighted score if it appeared to be the health topic that had the largest difference in terms of the rankings between the original structural link network and the revised semantic-based network. Also, these health topics were required to be ranked in the top 20 of the corresponding node centrality features in the revised semantic-based network.

For the children subcategory, the five core health topics identified were: *Medicines and Children*, *Children's Health*, *Toddler Health*, *Child Development*, and *Child Mental Health*. Among them, *Medicines and Children* had 14 recommended connections (12 bidirectional and 2 unidirectional); *Children's Health* had 16 recommended connections (8 bidirectional and 8 unidirectional); *Toddler Health* had 20 recommended connections (18 bidirectional and 2 unidirectional); *Child Development* had 11 recommended connections (8 bidirectional and 3 unidirectional); and *Child Mental Health* had 13 recommended connections (12 bidirectional and 1 unidirectional). Combining all connections relating to these five core topics, there were 57

178

suggested links in total, excluding the overlapped links among the five core topics. These 57 suggested links weighted 30.98% of all the recommended links, hence could prove the important roles played by the five core topics. The full list of suggested links related to the five core health topics could be found in Appendix 14.

The connections of the five core health topics were later added to the original structural network to form an adjusted structural network. The number of total connections included in this adjusted network was 529. Node-level centrality measures were then analyzed from the adjusted structural network in order to investigate how the five core health topics might affect the node centrality measures within the network. The ranking lists of the three node centrality features are displayed in Tables 39-43. The five core topics are displayed in red in the revised semantic-based network's ranking lists. In the adjusted structural network, the five core topics are shown in red bold font.

Out_Degree	Revised_Out_Degree	Adjusted_Out_Degree
Childhood Immunization	*Child Development*	***Child Development***
Medicines	*Medicines and Children*	***Medicines and Children***
Child Development	*Children's Health*	***Children's Health***
Medicines and Children	*Toddler Health*	***Toddler Health***
Hearing Disorders and Deafness	*Child Mental Health*	***Child Mental Health***
Birth Defects	*Child Dental Health*	*Childhood Immunization*
Child Dental Health	*Childhood Immunization*	*Child Behavior Disorders*
Dental Health	*Teen Health*	*Child Dental Health*
Child Behavior Disorders	*Child Behavior Disorders*	*Medicines*

179

Sports Fitness	Infant and Newborn Care	Hearing Disorders and Deafness
Assistive Devices	Teen Mental Health	Teen Mental Health
Teen Mental Health	Child Nutrition	Birth Defects
Fetal Health and Development	Medicines	Child Nutrition
Common Infant and Newborn Problems	Fetal Health and Development	Common Infant and Newborn Problems
Teen Health	Uncommon Infant and Newborn Problems	Uncommon Infant and Newborn Problems
Uncommon Infant and Newborn Problems	Teen Development	Dental Health
Infant and Newborn Development	Birth Defects	Sports Fitness
Teen Development	Common Infant and Newborn Problems	Assistive Devices
Infant and Newborn Care	Exercise for Children	Fetal Health and Development
Immunization	Dental Health	Teen Health

Table 39. Out-degree rankings among the three networks

In_Degree	Revised_In_Degree	Adjusted_In_Degree
Children's Health	*Children's Health*	***Children's Health***
Childhood Immunization	*Medicines and Children*	***Medicines and Children***
Medicines	*Toddler Health*	***Toddler Health***
Birth Defects	Medicines	Uncommon Infant and Newborn Problems
Uncommon Infant and Newborn Problems	Uncommon Infant and Newborn Problems	***Child Development***
Hearing Disorders and Deafness	*Child Development*	Childhood Immunization
Teen Mental Health	Childhood Immunization	Medicines
Sports Fitness	Teen Mental Health	Birth Defects
Assistive Devices	Birth Defects	***Child Mental Health***

180

Common Infant and Newborn Problems	*Child Mental Health*	*Common Infant and Newborn Problems*
Child Development	*Child Dental Health*	*Hearing Disorders and Deafness*
Dental Health	*Infant and Newborn Care*	*Teen Mental Health*
Over-the-Counter Medicines	*Child Nutrition*	*Child Behavior Disorders*
Drug Safety	*Common Infant and Newborn Problems*	*Sports Fitness*
Pregnancy	*Teen Health*	*Assistive Devices*
Genetic Disorders	*Child Behavior Disorders*	*Child Dental Health*
Teen Health	*Baby Health Checkup*	*Child Nutrition*
Infant and Newborn Care	*Hearing Disorders and Deafness*	*Dental Health*
Speech and Communication Disorders	*Teenage Pregnancy*	*Baby Health Checkup*
Tooth Disorders	*Fetal Health and Development*	*Over-the-Counter Medicines*

Table 40. In-degree rankings among the three networks

Out_Closeness	**Revised_Out_Closeness**	**Adjusted_Out_Closeness**
Fetal Alcohol Spectrum Disorders	*Childhood Immunization*	***Medicines and Children***
Cold and Cough Medicines	*Medicines and Children*	***Child Development***
Childhood Immunization	*Child Development*	*Childhood Immunization*
Medicines	*Children's Health*	*Cold and Cough Medicines*
Herbal Medicine	*Toddler Health*	*Medicines*
Medicines and Children	*Medicines*	***Children's Health***
Growth Disorders	*Infant and Newborn Development*	***Toddler Health***
Hepatitis B	*Child Nutrition*	*Antibiotics*
Whooping Cough	*Fetal Health and Development*	***Child Mental Health***

181

Down Syndrome	*Child Mental Health*	*Exercise for Children*
Dwarfism	*Fetal Alcohol Spectrum Disorders*	*Child Behavior Disorders*
Cough	*Child Behavior Disorders*	*Fetal Alcohol Spectrum Disorders*
Pregnancy and Medicines	*Antibiotics*	*Child Nutrition*
Immunization	*Cold and Cough Medicines*	*Hearing Problems in Children*
Tetanus	*Infant and Newborn Care*	*Hepatitis B*
Antibiotics	*Baby Health Checkup*	*Medication Errors*
Common Cold	*Birth Defects*	*Over-the-Counter Medicines*
Complementary and Integrative Medicine	*Teen Health*	*Pain Relievers*
Drug Safety	*Toddler Development*	*Herbal Medicine*
Flu Shot	*Hearing Problems in Children*	*Developmental Disabilities*

Table 41. Out-closeness rankings among the three networks

In_Closeness	Revised_In_Closeness	Adjusted_In_Closeness
Children's Health	*Children's Health*	***Children's Health***
Child Nutrition	*Toddler Health*	***Toddler Health***
Uncommon Infant and Newborn Problems	*Uncommon Infant and Newborn Problems*	*Uncommon Infant and Newborn Problems*
Common Infant and Newborn Problems	*Medicines and Children*	***Medicines and Children***
Infant and Newborn Care	*Child Development*	***Child Development***
Baby Health Checkup	*Child Nutrition*	***Child Mental Health***
Infant and Newborn Nutrition	*Child Mental Health*	*Common Infant and Newborn Problems*
Toddler Health	*Baby Health Checkup*	*Baby Health Checkup*
Toddler Nutrition	*Common Infant and Newborn Problems*	*Child Behavior Disorders*

Birth Weight	Child Behavior Disorders	Child Nutrition
Child Safety	Child Safety	Child Safety
Toddler Development	Toddler Development	Toddler Development
Dandruff, Cradle Cap, and Other Scalp Conditions	Child Dental Health	Child Dental Health
Child Development	Infant and Newborn Development	Medicines
Fetal Health and Development	Medicines	Infant and Newborn Care
Teen Health	Fetal Health and Development	Toddler Nutrition
School Health	Infant and Newborn Care	Over-the-Counter Medicines
Infant and Newborn Development	Infant and Newborn Nutrition	Speech and Language Problems
Child Behavior Disorders	Teen Health	Drug Safety
Speech and Language Problems	Teen Mental Health	Medication Errors

Table 42. In-closeness rankings among the three networks

Betweeness	Revised_Betweeness	Adjusted_Betweeness
Birth Defects	*Children's Health*	***Children's Health***
Childhood Immunization	Childhood Immunization	***Child Development***
Medicines	*Medicines and Children*	***Medicines and Children***
Fetal Health and Development	*Child Development*	Childhood Immunization
Uncommon Infant and Newborn Problems	Medicines	Medicines
Child Development	Fetal Health and Development	Antibiotics
Pregnancy and Medicines	Prenatal Testing	Uncommon Infant and Newborn Problems
Birth Weight	Child Safety	***Toddler Health***
Assistive Devices	Birth Defects	Haemophilus Infections

183

Children's Health	Assistive Devices	Bacterial Infections
Child Nutrition	*Toddler Health*	Birth Defects
Toddler Development	Baby Health Checkup	**Child Mental Health**
Antibiotics	Antibiotics	Child Dental Health
Bacterial Infections	Hepatitis Testing	Assistive Devices
Haemophilus Infections	Child Dental Health	Speech and Language Problems in Children
Speech and Language Problems in Children	Haemophilus Infections	Speech and Communication Disorders
Infant and Newborn Nutrition	Speech and Language Problems in Children	Growth Disorders
Toddler Health	Hearing Disorders and Deafness	Hearing Disorders and Deafness
Child Behavior Disorders	Sports Safety	Fetal Health and Development
Hearing Disorders and Deafness	Bacterial Infections	Birth Weight

Table 43. Betweenness rankings among the three networks

From the ranking lists displayed in Tables 39-43 regarding the three node-level centrality measures among the original structural network, the revised semantic-based network, and the adjusted core topic-based network, it is clear to see that the five core topics improved to higher rankings and these changes led the node centrality ranking lists of the adjusted network to become more consistent with the ranking lists of the revised semantic-based network. In other words, identifying core health topics might greatly assist a subject directory in better reflecting its topics' semantic relationships.

4.4.2   RQ3.1 & Hypothesis group 3

RQ3.1 is "Are there any significant differences between the original and optimized structural networks of children related topics in terms of similarity values?" It examines if

similarity value differences appeared after the optimization process.

The similarity values of the selected children related health topics were not able to be analyzed through the standard T-test since they did not have a normal distribution. Therefore, the Mann-Whitney test was applied to investigate the differences of similarity values between the original and optimized subject directories. SPSS was employed to conduct the inferential statistical analysis for the children related topics.

RQ3.1 was answered by hypothesis group 3. This hypothesis group contained only one hypothesis. The Mann-Whitney test was applied to test the hypothesis in hypothesis group 3. Table 44 summarizes the mean rank and sum of ranks of the similarity values from the original structural network (n=472) and the optimized structural network (n=656). The statistical analysis revealed that there was a systematic difference of similarity values between the original and optimized structural networks of children related topics (p=0.002<0.05). Therefore, the hypothesis was rejected. In other words, the similarity value of the connections was significantly improved after the optimization process.

	Similarity value	
	Original structural network	*Optimized structural network*
Mean rank	*528.76*	*590.21*
Sum of ranks	*249575.00*	*387181.00*
z-statistic	*-3.125*	
p-value	*0.002*	

Table 44. Statistical analysis result for hypothesis group 3

4.4.3   *RQ3.2 & Hypothesis group 4*

RQ3.2 is stated as "Are there any significant differences between the original and

185

optimized structural networks of children related topics in terms of node-level centrality measurements?" This question aims at investigating if any differences of node-level centrality measures occurred between the original and optimized structural network after the optimization process.

Similar with the mental health topic subcategory, a series of customized T-tests were applied to investigate the differences of three node centrality features between the original and optimized subject directories. UciNet was employed to conduct the inferential statistical analysis for the children related topics.

RQ3.2 was answered by hypothesis group 4, and this hypothesis group contained three sub-hypotheses. Each sub-hypothesis under hypothesis group 4 was aimed at comparing one of the three node-level centrality measures (degree centrality, closeness centrality, and betweenness centrality) between the original and optimized structures.

A series of customized T-tests were applied to test the three sub-hypotheses under hypothesis group 4. Tables 45-47 include the means and standard deviations (SD) of degree centrality, closeness centrality, and betweenness centrality of all original structural connections (n=159) and all optimized structural connections (n=159). In Table 45, the p-values of both the out-degree centrality and in-degree centrality were smaller than the significant level (0.05), which indicated that there was a significant difference of degree centrality measures between the original structural network and the optimized structural network. That is, the degree centrality measures in the optimized structural network had generated a significantly higher average value than the original structural network.

186

	Out-Degree		In-Degree	
	Original structural network	Optimized structural network	Original structural network	Optimized structural network
Mean	0.019	0.026	0.019	0.026
Standard deviation (SD)	0.011	0.017	0.017	0.022
t-statistic	-0.007		-0.007	
p-value	0.0001		0.001	

Table 45. Degree centrality rankings between the original and optimized networks

In Table 46, the p-values of both the out-closeness centrality and in-closeness centrality were smaller than the significant level (0.05) as well, which indicated that there was a significant difference of closeness centrality measures between the original structural network and the optimized structural network. That is, the closeness centrality measures in the optimized structural network had generated a significantly higher average value than the original structural network.

	Out-Closeness		In-Closeness	
	Original structural network	Optimized structural network	Original structural network	Optimized structural network
Mean	0.087	0.175	0.121	0.209
Standard deviation (SD)	0.007	0.015	0.066	0.077
t-statistic	-0.088		-0.088	
p-value	0.0001		0.0001	

Table 46. Closeness centrality rankings between the original and optimized networks

In Table 47, unlike the degree and closeness centrality measures, the p-value of the betweenness centrality was larger than the significant level (0.05), which indicated that there was no significant difference of betweenness centrality measure between the original structural network and the optimized structural network. That is, the betweenness centrality measure in the optimized structural network had not generated a significantly higher average value than the original structural network. This result echoed the findings investigated in the optimized network

187

of mental health related health topics.

	Betweenness	
	Original structural network	Optimized structural network
Mean	2.405	2.078
Standard deviation (SD)	3.868	3.463
t-statistic	0.327	
p-value	0.4255	

Table 47. Betweenness centrality rankings between the original and optimized networks

As a result, for hypothesis group 4, H04$_{(a)}$ and H04$_{(b)}$ were rejected while H04$_{(c)}$ failed to be rejected.

4.4.4   Summary

The third research question concentrated on comparing the original structural network and the optimized structural network of children related health topics in the subject directory on MedlinePlus. The introductory section explored the optimized structural network in terms of structural connections, semantic connections, and key nodes in the network, while RQ3.1 and RQ3.2 investigated the differences between the original and optimized structural networks regarding their connections' similarity values as well as node-level centrality measures.

It was found that the recommended new structural connections were mostly added to the health consumer related health topics. These consumer groups included children, teenagers, older adults, infants, toddlers, etc. Besides the health consumer groups, connections were also found in pregnancy, hepatitis, and daily exercises related topics. Moreover, the average similarity value of the structural connections in the optimized network was improved after new connections were added. This difference was proved to be significant through the Mann-Whitney test. Key health topics were identified to have a huge impact on the whole network again in the children topic

188

subcategory. It echoed the conclusion that through the creation of key nodes' connections, the node-level centrality ranking lists in the adjusted structural network moved closer to the semantic-based network. In other words, key influential nodes were more capable of reflecting the semantic relationships in the network compared with normal nodes. In addition, the node centrality measures were tested through the customized T-test, the results showed that the degree centrality and closeness centrality measures were significantly increased in the optimized structural network while the betweenness centrality measure was not significantly different from the original structural link network. This conclusion also fit into the earlier results generated in the network of the mental health topic subcategory.

4.5 Findings for research question 4 (RQ4)

RQ4: Are there significant differences between the original teenagers subject directory and the optimized teenagers subject directory in terms of its link structure on the MedlinePlus portal?

As discussed in section 4.2, the structural and semantic connections were not consistent regarding teenagers related health topics in the subject directory of MedlinePlus. Recommendations were made for a group of health topics in order to add more relevant connections that had high similarity values in terms of Web page textual information. Moreover, some hidden core health topics were detected through the comparison of the ranking lists between the structural and semantic network. These health topics were identified through the analysis of node-level centrality features. As a result, an optimized structural link network was developed. It is important to ensure that this optimized subject directory works better than the original one in terms of organizing related health topics from both the structural and semantic perspectives.

189

4.5.1 Introduction of the optimized structural network

4.5.1.1 Description of the structural connections in the optimized network

This section aims at providing an overview of the optimized subject directory regarding the revised structural connections. As introduced in section 4.2, there were 389 connections in the original structural link network. After conducting a calculation toward the average similarity values held among all the selected health topics, the threshold was set according to the average number of structurally connected topics. The threshold was 0.461491. Following that, all the unconnected topics were investigated to find pairs of health topics that were sharing a similarity value higher than the threshold. As a result, 241 connections were found and among them, 190 were bidirectional connections while 51 were unidirectional connections. All these 241 connections were then added to the original structural link network to form an optimized subject directory. This optimized subject directory is shown in Figure 20.

190
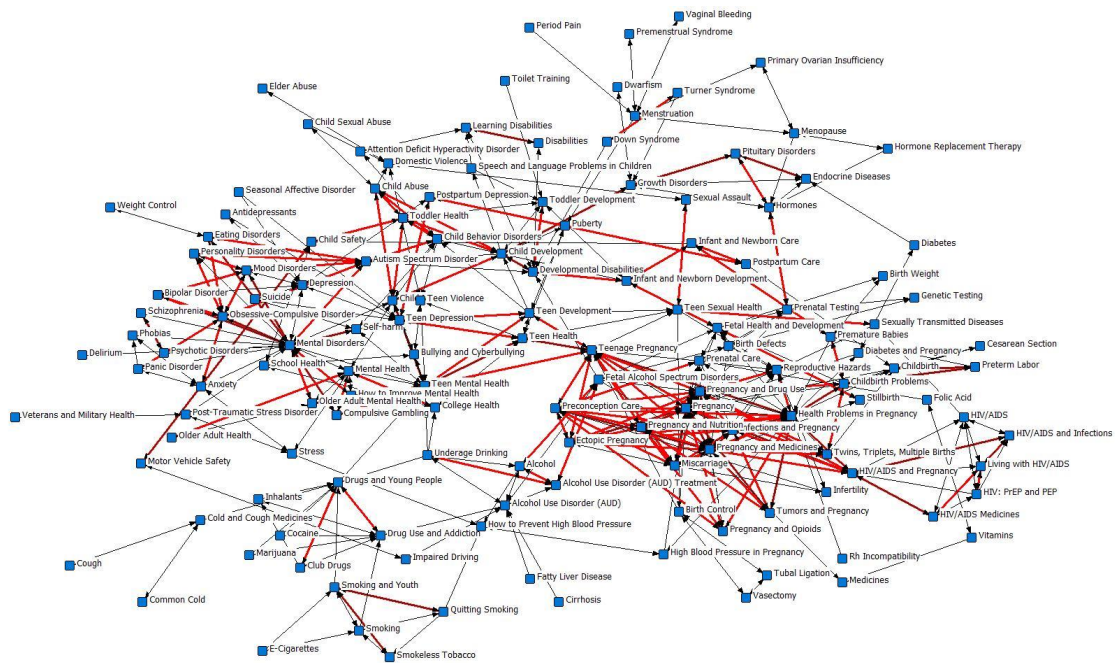
Figure 20. Optimized teenagers structural link network

In Figure 20, the original structural connections are indicated in black lines while the recommended connections are in red lines. Compared with the optimized structural networks of mental health and children related health topics, recommended connections were quite clustered in teenagers related health topics' network. A lot of recommended connections were found in pregnancy and specific mental diseases related health topics.

4.5.1.2   Description of the semantic connections of the optimized network

This section identifies the new semantic relationships of the optimized teenagers subject directory. For teenagers related topics, the overall average similarity value among all the 19,460 potential connections stayed the same at 0.111522 after the optimization process. However, since the structured connections in the optimized subject directory combined both Edge Set B and Edge Set D, there were 630 connections in total as a result, thus leading the average similarity value to increase to 0.504422. The average similarity among the rest of the 18,830 pairs of topics (new Edge Set C) was decreased to 0.098377.

4.5.1.3   Identification of the key nodes

This section tends to reveal specific hidden core health topics in the original structural link network. In the previous stage, several core health topics were identified based on the analysis results of node-level centrality measures. To further filter these hidden core health topics, each topic was assigned with a weighted score if it appeared to be the health topic that had the largest difference in terms of the rankings between the original structural link network and the revised semantic-based network. Also, these health topics were required to be ranked in

191

the top 20 of the corresponding node centrality features in the revised semantic-based network.

For the teenagers subcategory, the five core health topics identified were: *Teenage Pregnancy*, *Infections and Pregnancy*, *Pregnancy and Drug Use*, *Pregnancy and Medicines*, and *Health Problems in Pregnancy*. Among them, *Teenage Pregnancy* had 20 recommended connections (18 bidirectional and 2 unidirectional); *Infections and Pregnancy* had 17 recommended connections (14 bidirectional and 3 unidirectional); *Pregnancy and Drug Use* had 21 recommended connections (18 bidirectional and 3 unidirectional); *Pregnancy and Medicines* had 19 recommended connections (14 bidirectional and 5 unidirectional); and *Health Problems in Pregnancy* had 18 recommended connections (12 bidirectional and 6 unidirectional). Combining all connections relating to these five core topics, there were 82 suggested links in total, excluding the overlapped links among the five core topics. These 82 suggested links weighted 34.02% of all the recommended links, hence could prove the important roles played by the five core topics. The full list of suggested links related to the five core health topics could be found in Appendix 15.

The connections of the five core health topics were later added to the original structural network to form an adjusted structural network. The number of total connections included in this adjusted network was 471. Node-level centrality measures were then analyzed from the adjusted structural network in order to investigate how the five core health topics might affect the node centrality measures within the network. The ranking lists of the three node centrality measures are displayed in Tables 48-52. The five core topics are marked in red in the revised semantic-based network's ranking lists. In the adjusted structural network, the five core topics are marked in red bold font.

192

Out_Degree	Revised_Out_Degree	Adjusted_Out_Degree
Health Problems in Pregnancy	*Health Problems in Pregnancy*	***Health Problems in Pregnancy***
Pregnancy	*Pregnancy*	***Pregnancy and Drug Use***
Prenatal Care	*Pregnancy and Drug Use*	***Pregnancy and Medicines***
Depression	*Mental Disorders*	***Teenage Pregnancy***
Pregnancy and Drug Use	*Pregnancy and Medicines*	***Infections and Pregnancy***
Teen Mental Health	*Teenage Pregnancy*	*Pregnancy*
Child Development	*Teen Mental Health*	*Miscarriage*
Mental Disorders	*Infections and Pregnancy*	*Teen Mental Health*
Mental Health	*Child Development*	*Prenatal Care*
Child Behavior Disorders	*Miscarriage*	*Depression*
Fetal Health and Development	*Pregnancy and Nutrition*	*HIV/AIDS and Pregnancy*
Pregnancy and Medicines	*Preconception Care*	*Ectopic Pregnancy*
Fetal Alcohol Spectrum Disorders	*Mental Health*	*Teen Development*
HIV/AIDS	*Childbirth Problems*	*Child Development*
Teen Development	*HIV/AIDS and Pregnancy*	*Mental Disorders*
Teen Sexual Health	*Ectopic Pregnancy*	*Mental Health*
Underage Drinking	*Teen Development*	*Child Behavior Disorders*
Menstruation	*Child Mental Health*	*Childbirth Problems*
Child Mental Health	*Child Behavior Disorders*	*Fetal Health and Development*
Alcohol Use Disorder (AUD)	*Obsessive-Compulsive Disorders*	*Pregnancy and Nutrition*

Table 48. Out-degree rankings among the three networks

In_Degree	Revised_In_Degree	Adjusted_In_Degree

193

Pregnancy	*Pregnancy*	***Health Problems in Pregnancy***
Mental Disorders	*Mental Disorders*	*Pregnancy*
Health Problems in Pregnancy	*Health Problems in Pregnancy*	*Mental Disorders*
Teen Mental Health	*Pregnancy and Medicines*	***Pregnancy and Medicines***
Reproductive Hazards	*Pregnancy and Drug Use*	***Pregnancy and Drug Use***
Depression	*Teenage Pregnancy*	***Teenage Pregnancy***
Child Development	*Teen Mental Health*	***Infections and Pregnancy***
Pregnancy and Medicines	*Infections and Pregnancy*	*Teen Mental Health*
Alcohol Use Disorder (AUD)	*Preconception Care*	*Reproductive Hazards*
Drug Use and Addiction	*Pregnancy and Nutrition*	*HIV/AIDS and Pregnancy*
Prenatal Care	*Child Development*	*Depression*
Fetal Health and Development	*Miscarriage*	*Preconception Care*
Teen Health	*HIV/AIDS and Pregnancy*	*Miscarriage*
Childbirth	*Reproductive Hazards*	*Child Development*
Drugs and Young People	*Ectopic Pregnancy*	*Alcohol Use Disorder (AUD)*
Teen Depression	*Teen Depression*	*Drug Use and Addiction*
Mental Health	*Mental Health*	*Prenatal Care*
Child Behavior Disorders	*Child Mental Health*	*Ectopic Pregnancy*
HIV/AIDS	*Child Behavior Disorders*	*Fetal Health and Development*
Menstruation	*Obsessive-Compulsive Disorders*	*Teen Health*

Table 49. In-degree rankings among the three networks

Out_Closeness	Revised_Out_Closeness	Adjusted_Out_Closeness
Alcohol Use Disorder (AUD)	*Teenage Pregnancy*	***Teenage Pregnancy***

194

Fetal Alcohol Spectrum Disorders	*Teen Mental Health*	***Health Problems in Pregnancy***
Underage Drinking	*Teen Development*	***Pregnancy and Drug Use***
HIV/AIDS and Pregnancy	*Pregnancy and Drug Use*	***Pregnancy and Medicines***
Health Problems in Pregnancy	*Health Problems in Pregnancy*	*Teen Mental Health*
Alcohol	*Teen Health*	***Infections and Pregnancy***
Pregnancy and Drug Use	*Pregnancy*	*Pregnancy*
Teenage Pregnancy	*Child Mental Health*	*Ectopic Pregnancy*
Teen Sexual Health	*Pregnancy and Medicines*	*Miscarriage*
Pregnancy	*Fetal Alcohol Spectrum Disorders*	*Pregnancy and Nutrition*
Prenatal Care	*Infections and Pregnancy*	*Preconception Care*
Teen Development	*Preconception Care*	*Fetal Alcohol Spectrum Disorders*
Mental Health	*Miscarriage*	*Teen Development*
Impaired Driving	*Pregnancy and Nutrition*	*Prenatal Care*
Teen Health	*Child Development*	*Teen Health*
Teen Mental Health	*Mental Disorders*	*Teen Sexual Health*
Cirrhosis	*Ectopic Pregnancy*	*Underage Drinking*
Fatty Liver Disease	*Teen Sexual Health*	*Alcohol Use Disorder (AUD)*
Child Mental Health	*Underage Drinking*	*Child Mental Health*
Mental Disorders	*Prenatal Care*	*Birth Control*

Table 50. Out-closeness rankings among the three networks

In_Closeness	Revised_In_Closeness	Adjusted_In_Closeness
Teen Health	*Teen Mental Health*	*Teen Mental Health*
Teen Mental Health	*Teenage Pregnancy*	***Teenage Pregnancy***

Teen Violence	*Teen Health*	*Teen Health*
Teen Sexual Health	*Teen Development*	*Pregnancy*
Child Behavior Disorders	*Pregnancy*	***Health Problems in Pregnancy***
Teen Development	*Child Mental Health*	***Pregnancy and Medicines***
Reproductive Hazards	*Child Development*	*Mental Disorders*
College Health	*Health Problems in Pregnancy*	***Pregnancy and Drug Use***
Bullying and Cyberbullying	*Mental Disorders*	***Infections and Pregnancy***
Teenage Pregnancy	*Pregnancy and Medicines*	*Teen Development*
Child Development	*Pregnancy and Drug Use*	*Self-harm*
Pregnancy	*Mental Health*	*Child Behavior Disorders*
Mental Disorders	*Infections and Pregnancy*	*Mental Health*
Child Mental Health	*Child Behavior Disorders*	*Bullying and Cyberbullying*
Self-harm	*Self-harm*	*Preconception Care*
Mental Health	*Preconception Care*	*Miscarriage*
Teen Depression	*Pregnancy and Nutrition*	*Ectopic Pregnancy*
How to Prevent High Blood Pressure	*Developmental Disabilities*	*Pregnancy and Nutrition*
School Health	*Miscarriage*	*Teen Depression*
Puberty	*Ectopic Pregnancy*	*Teen Sexual Health*

Table 51. In-closeness rankings among the three networks

Betweeness	Revised_Betweeness	Adjusted_Betweeness
Teen Health	*Teen Mental Health*	***Teenage Pregnancy***
Teen Mental Health	*Teenage Pregnancy*	*Teen Mental Health*
Health Problems in Pregnancy	*Mental Disorders*	***Health Problems in Pregnancy***

Pregnancy	Child Development	Mental Disorders
Teen Sexual Health	*Health Problems in Pregnancy*	Child Development
Teenage Pregnancy	*Pregnancy and Drug Use*	**Pregnancy and Drug Use**
Child Development	Fetal Alcohol Spectrum Disorders	Fetal Alcohol Spectrum Disorders
Mental Disorders	Pregnancy	Alcohol Use Disorder (AUD)
Teen Violence	Drugs and Young People	Teen Health
Mental Health	Fetal Health and Development	Puberty
Fetal Alcohol Spectrum Disorders	HIV/AIDS and Pregnancy	Mental Health
Pregnancy and Drug Use	Teen Development	Child Behavior Disorders
Alcohol Use Disorder (AUD)	Child Mental Health	Teen Development
Child Behavior Disorders	Alcohol Use Disorder (AUD)	HIV/AIDS and Pregnancy
Stress	Infant and Newborn Development	Drugs and Young People
Puberty	Teen Sexual Health	Pregnancy
Reproductive Hazards	Teen Health	Teen Violence
Depression	Growth Disorders	**Pregnancy and Medicines**
Teen Development	Mental Health	Depression
High Blood Pressure in Pregnancy	Developmental Disabilities	Child Mental Health

Table 52. Betweenness rankings among the three networks

From the ranking lists displayed in Tables 48-52 regarding the three node-level centrality measures among the original structural network, the revised semantic-based network, and the adjusted core topic-based network, it is clear to see that the five core topics improved to higher rankings and these changes led the node centrality ranking lists of the adjusted network to become more consistent with the ranking lists of the revised semantic-based network. In other

197

words, identifying core health topics might greatly assist a subject directory in better reflecting its topics' semantic relationships.

4.5.2   RQ4.1 & Hypothesis group 5

RQ4.1 is described as "Are there any significant differences between the original and optimized structural networks of teenagers related topics in terms of similarity values?" It examines if similarity value differences appeared after the optimization process.

The similarity values were not able to be analyzed through the standard T-test since they did not have a normal distribution. Therefore, the Mann-Whitney test was applied to investigate the differences of similarity values between the original and optimized subject directories. SPSS was employed to conduct the inferential statistical analysis for the teenager related topics.

RQ4.1 was answered by hypothesis group 5. This hypothesis group contained only one hypothesis. The Mann-Whitney test was applied to test the hypothesis in hypothesis group 5. Table 53 summarizes the mean rank and sum of ranks of the similarity values from the original structural network (n=389) and the optimized structural network (n=630). The statistical analysis revealed that there was a systematic difference of similarity values between the original and optimized structural networks of teenagers related topics (p=0.000<0.05). Therefore, the hypothesis was rejected. In other words, the similarity value of the connections was significantly improved after the optimization process.

	Similarity value	
	Original structural network	*Optimized structural network*
Mean rank	*467.08*	*536.50*
Sum of ranks	*181695.00*	*337995.00*

198

z-statistic	-3.658
p-value	0.000

Table 53. Statistical analysis result for hypothesis group 5

4.5.3   *RQ4.2 & Hypothesis group 6*

RQ4.2 is stated as "Are there any significant differences between the original and optimized structural networks of teenagers related topics in terms of node-level centrality measurements?" This question aims at investigating if any differences of node-level centrality measures occurred between the original and optimized structural network after the optimization process.

Similar with the mental health topic subcategory as well as the children topic subcategory, a series of customized T-tests were applied to investigate the differences of three node centrality features between the original and optimized subject directories. UciNet was employed to conduct the inferential statistical analysis for the teenagers related topics.

RQ4.2 was answered by hypothesis group 6, and this hypothesis group contained three sub-hypotheses. Each sub-hypothesis under hypothesis group 6 was aimed at comparing one of the three node-level centrality measures (degree centrality, closeness centrality, and betweenness centrality) between the original and optimized subject structures.

A series of customized T-tests were applied to test the three sub-hypotheses under hypothesis group 6. Tables 54-56 include the means and standard deviations (SD) of degree centrality, closeness centrality, and betweenness centrality of all original structural connections (n=140) and all optimized structural connections (n=140). In Table 54, the p-values of both the out-degree centrality and in-degree centrality were smaller than the significant level (0.05),

199

which indicated that there was a significant difference of degree centrality measures between the original structural network and the optimized structural network. That is, the degree centrality measures in the optimized structural network had generated a significantly higher average value than the original structural network.

	Out-Degree		In-Degree	
	Original structural network	Optimized structural network	Original structural network	Optimized structural network
Mean	0.020	0.032	0.020	0.032
Standard deviation (SD)	0.012	0.025	0.021	0.030
t-statistic	-0.012		-0.012	
p-value	0.0001		0.0001	

Table 54. Degree centrality rankings between the original and optimized networks

In Table 55, the p-values of both the out-closeness centrality and in-closeness centrality were smaller than the significant level (0.05) as well, which indicated that there was a significant difference of closeness centrality measures between the original structural network and the optimized structural network. That is, the closeness centrality measures in the optimized structural network had generated a significantly higher average value than the original structural network.

	Out-Closeness		In-Closeness	
	Original structural network	Optimized structural network	Original structural network	Optimized structural network
Mean	0.119	0.193	0.146	0.221
Standard deviation (SD)	0.008	0.019	0.063	0.071
t-statistic	-0.075		-0.075	
p-value	0.0001		0.0001	

Table 55. Closeness centrality rankings between the original and optimized networks

In Table 56, unlike the degree and closeness centrality measures, the p-value of the betweenness centrality was larger than the significant level (0.05), which indicated that there was

200

no significant difference of betweenness centrality measure between the original structural network and the optimized structural network. That is, the betweenness centrality measure in the optimized structural network had not generated a significantly higher average value than the original structural network. This result echoed the findings investigated in the optimized networks of mental health and children related health topics.

	Betweenness	
	Original structural network	*Optimized structural network*
Mean	*2.238*	*1.912*
Standard deviation (SD)	*4.000*	*3.647*
t-statistic	*0.326*	
p-value	*0.4691*	

Table 56. Betweenness centrality rankings between the original and optimized networks

As a result, for hypothesis group 6, H06$_{(a)}$ and H06$_{(b)}$ were rejected while H06$_{(c)}$ failed to be rejected.

4.5.4   Summary

The fourth research question concentrated on comparing the original structural network and the optimized structural network of teenagers related health topics in the subject directory on MedlinePlus. The introductory section examined the optimized structural network in terms of structural connections, semantic connections, and key nodes in the network, while RQ4.1 and 4.2 explored the differences between the original and optimized structural networks regarding their connections' similarity values as well as node-level centrality measures.

It was found that the recommended new structural connections were mostly added to the pregnancy related health topics. These recommendations were echoed by the five core topics identified through the process of identifying key nodes in the revised semantic-based network –

201

the five core topics recognized were all about pregnancy related issues. This finding had proved the great impact possessed by the pregnancy related health topics in the network of the teenagers subcategory. Besides the pregnancy related health topics, connections were also found among health topics relating to specific mental diseases. The close relationship between mental health and teenagers was verified. Moreover, the average similarity value of the structural connections in the optimized network was improved after new connections were added. This difference was found to be significant through the Mann-Whitney test. The five pregnancy related key health topics were identified to have a huge impact in the whole network again in the teenagers topic subcategory. It proved the conclusion that through the creation of key nodes' connections, the node-level centrality ranking lists in the adjusted structural network moved closer to the semantic-based network. In other words, key influential nodes were more capable of reflecting the semantic relationships in the network compared with normal nodes. In addition, the node centrality measures were tested through the customized T-test, the results showed that the degree and closeness centrality measures were significantly increased in the optimized structural network while the betweenness centrality measure was not significantly different from the original structural link network. This conclusion also fit into the earlier results generated in the network of the mental health topic subcategory and children topic subcategory.

4.6 Findings for research question 5 (RQ5)

RQ5: Are there significant differences between the original older adults subject directory and the optimized older adults subject directory in terms of its link structure on the MedlinePlus portal?

As discussed in section 4.2, the structural and semantic connections were not consistent

regarding older adults related health topics in the subject directory of MedlinePlus. Recommendations were made for a group of health topics in order to add more relevant connections that had high similarity values in terms of Web page textual information. Moreover, some hidden core health topics were detected through the comparison of the ranking lists between the structural and semantic network. These health topics were identified through the analysis of node-level centrality features. As a result, an optimized structural link network was developed. It is important to ensure that this optimized subject directory works better than the original one in terms of organizing related health topics from both the structural and semantic perspectives.

4.6.1 *Introduction of the optimized structural network*

4.6.1.1 Description of the structural connections in the optimized network

This section aims at providing an overview of the optimized subject directory regarding the revised structural connections. As introduced in section 4.2, there were 446 connections in the original structural link network. After conducting a calculation for the average similarity values held among all the selected health topics, the threshold was set according to the average number of structurally connected topics. The threshold was 0.440717. Following that, all the unconnected topics were investigated to find pairs of health topics that were sharing a similarity value higher than the threshold. As a result, 244 connections were found and among them, 196 were bidirectional connections while 48 were unidirectional connections. All these 244 connections were then added to the original structural link network to form an optimized subject directory. This optimized subject directory is shown in Figure 21.

203
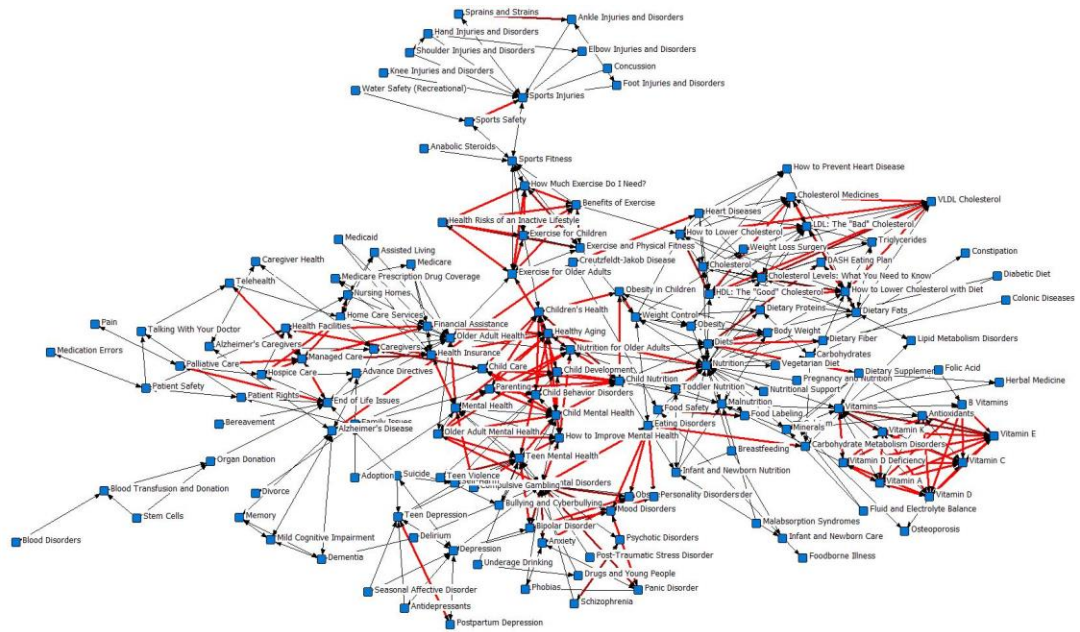
Figure 21. Optimized older adults structural link network

In Figure 21, the original structural connections are indicated in black lines while the recommended connections are in red lines. Compared with the optimized structural networks of mental health, teenagers, and children related health topics, recommended connections for the older adults subcategory were gathered in a few groups of health topics. These groups covered several health consumers groups, specific mental diseases, and some daily health elements. To be more specific, a lot of recommended connections were found in the interactions among mental health and the three consumers groups (children, teenagers, and older adults). Meanwhile, for daily health related issues, vitamins and cholesterol related health topics possessed a great number of recommended connections as well. Compared with vitamins and cholesterol, sports injuries related topics were well developed.

204

4.6.1.2 Description of the semantic connections of the optimized network

This section identifies the new semantic relationships of the optimized older adults subject directory. For older adults related topics, the overall average similarity value among all the 22,350 potential connections stayed the same at 0.104228 after the optimization process. However, since the structured connections in the optimized subject directory combined both Edge Set B and Edge Set D, there were 690 connections in total as a result, thus leading the average similarity value to increase to 0.493671. Following that, the average similarity among the rest of the 21,660 pairs of topics (new Edge Set C) was decreased to 0.091822.

4.6.1.3 Identification of the key nodes

This section tends to reveal specific hidden core health topics in the original structural link network. In the previous stage, several core health topics were identified based on the analysis results of node-level centrality measures. To further filter these hidden core health topics, each topic was assigned with a weighted score if it appeared to be the health topic that had the largest difference in terms of the rankings between the original structural link network and the revised semantic-based network. Also, these health topics were required to be ranked in the top 20 of the corresponding node centrality features in the revised semantic-based network.

For the older adults subcategory, the five core health topics identified were: *Child Nutrition*, *Healthy Aging*, *Teen Mental Health*, *Child Mental Health*, and *Mental Health*. Among them, *Child Nutrition* had 15 recommended connections (12 bidirectional and 3 unidirectional); *Healthy Aging* had 16 recommended connections (14 bidirectional and 2 unidirectional); *Teen Mental Health* had 12 recommended connections (10 bidirectional and 2 unidirectional); *Child*

205

Mental Health had 13 recommended connections (12 bidirectional and 1 unidirectional); and *Mental Health* had 10 recommended connections (10 bidirectional only). Combining all connections relating to these five core topics, there were 59 suggested links in total, excluding the overlapped links among the five core topics. These 59 suggested links weighted 24.18% of all the recommended links, hence could prove the important roles played by the five core topics. The full list of suggested links related to the five core health topics could be found in Appendix 16.

The connections of the five core health topics were later added to the original structural network to form an adjusted structural network. The number of total connections included in this adjusted network was 505. Node-level centrality measures were then analyzed from the adjusted structural network in order to investigate how the five core health topics might affect the node centrality measures within the network. The ranking lists of the three node centrality measures are displayed in Tables 57-61. The five core topics are marked in red in the revised semantic-based network's ranking lists. In the adjusted structural network, the five core topics are marked in red bold font.

Out_Degree	Revised_Out_Degree	Adjusted_Out_Degree
Nutrition	*Nutrition*	*Nutrition*
Cholesterol	*Mental Disorders*	***Child Nutrition***
Vitamins	*Vitamins*	***Teen Mental Health***
Dietary Fats	*Child Nutrition*	*Cholesterol*
Weight Control	*Teen Mental Health*	*Vitamins*
Triglycerides	*Cholesterol*	***Child Mental Health***

206

Depression	Child Mental Health	Healthy Aging
Heart Diseases	Healthy Aging	Mental Health
How to Lower Cholesterol	Mental Health	Dietary Fats
Sports Fitness	Nutrition for Older Adults	Weight Control
Child Nutrition	Dietary Fats	Triglycerides
Mental Disorders	Weight Control	Nutrition for Older Adults
Cholesterol Medicines	How to Lower Cholesterol with Diet	Depression
Caregivers	Vitamin D	Heart Diseases
End of Life Issues	Exercise for Older Adults	How to Lower Cholesterol
Exercise for Older Adults	Older Adult Mental Health	Sports Fitness
Nutrition for Older Adults	Managed Care	Mental Disorders
Suicide	Heart Diseases	Cholesterol Medicines
Mental Health	How to Lower Cholesterol	Older Adult Health
Calcium	Triglycerides	Caregivers

Table 57. Out-degree rankings among the three networks

In_Degree	Revised_In_Degree	Adjusted_In_Degree
Nutrition	Nutrition	Nutrition
Mental Disorders	Mental Disorders	Mental Disorders
Vitamins	Vitamins	Teen Mental Health
Older Adult Health	Teen Mental Health	Vitamins
Cholesterol	Older Adult Health	Older Adult Health
Cholesterol Levels: What You Need to Know	How to Lower Cholesterol with Diet	Child Nutrition
LDL: The "Bad" Cholesterol	Child Nutrition	Cholesterol

Teen Mental Health	Cholesterol	*Child Mental Health*
HDL: The "Good" Cholesterol	*Child Mental Health*	Cholesterol Levels: What You Need to Know
Sports Injuries	Cholesterol Levels: What You Need to Know	LDL: The "Bad" Cholesterol
Malnutrition	HDL: The "Good" Cholesterol	*Healthy Aging*
Depression	LDL: The "Bad" Cholesterol	*Mental Health*
Sports Fitness	*Healthy Aging*	HDL: The "Good" Cholesterol
How to Lower Cholesterol with Diet	*Mental Health*	Sports Injuries
Weight Control	How to Lower Cholesterol	Malnutrition
How to Lower Cholesterol	Diets	Depression
Diets	Sports Injuries	Sports Fitness
Exercise and Physical Fitness	VLDL Cholesterol	How to Lower Cholesterol with Diet
VLDL Cholesterol	Cholesterol Medicines	Weight Control
Minerals	Malnutrition	How to Lower Cholesterol

Table 58. In-degree rankings among the three networks

Out_Closeness	Revised_Out_Closeness	Adjusted_Out_Closeness
Eating Disorders	Nutrition	Nutrition
Caregivers	*Healthy Aging*	*Child Nutrition*
Home Care Services	*Child Nutrition*	Nutrition for Older Adults
Assisted Living	Nutrition for Older Adults	Child Care
End of Life Issues	*Mental Health*	*Child Mental Health*
Suicide	*Child Mental Health*	*Healthy Aging*
Advance Directives	Child Care	*Mental Health*
Nursing Homes	Older Adult Mental Health	*Teen Mental Health*

208

Weight Control	Mental Disorders	Infant and Newborn Nutrition
Child Care	*Teen Mental Health*	Older Adult Mental Health
Nutrition	Eating Disorders	Eating Disorders
Child Development	Health Insurance	Parenting
Exercise and Physical Fitness	Weight Control	Weight Control
Older Adult Health	How to Improve Mental Health	Child Development
Older Adult Mental Health	Infant and Newborn Nutrition	Older Adult Health
Organ Donation	Children's Health	Dietary Fiber
Nutrition for Older Adults	Older Adult Health	Food Safety
Family Issues	Toddler Nutrition	Vitamins
Telehealth	Child Behavior Disorders	Breastfeeding
Mental Health	Exercise for Older Adults	Calcium

Table 59. Out-closeness rankings among the three networks

In_Closeness	Revised_In_Closeness	Adjusted_In_Closeness
Older Adult Health	Nutrition	***Healthy Aging***
Nutrition for Older Adults	*Healthy Aging*	Nutrition
Nutrition	Nutrition for Older Adults	Nutrition for Older Adults
Healthy Aging	Older Adult Health	Older Adult Health
Diets	*Child Nutrition*	***Child Nutrition***
Malnutrition	*Mental Health*	***Mental Health***
Older Adult Mental Health	Older Adult Mental Health	***Child Mental Health***
Exercise for Older Adults	Diets	Older Adult Mental Health
Weight Control	*Child Mental Health*	Diets

Mental Disorders	Weight Control	*Teen Mental Health*
Vitamins	*Teen Mental Health*	Malnutrition
Dietary Proteins	Malnutrition	Exercise for Older Adults
Cholesterol	Exercise for Older Adults	Weight Control
Cholesterol Levels: What You Need to Know	Health Insurance	How to Improve Mental Health
How to Lower Cholesterol with Diet	How to Improve Mental Health	Health Insurance
LDL: The "Bad" Cholesterol	Children's Health	Children's Health
HDL: The "Good" Cholesterol	Cholesterol	Mental Disorders
Dietary Fats	Cholesterol Levels: What You Need to Know	Vitamins
Minerals	HDL: The "Good" Cholesterol	Dietary Proteins
Carbohydrates	LDL: The "Bad" Cholesterol	Cholesterol

Table 60. In-closeness rankings among the three networks

Betweeness	Revised_Betweeness	Adjusted_Betweeness
Nutrition	Nutrition	Nutrition
Older Adult Health	*Healthy Aging*	*Healthy Aging*
Mental Disorders	Mental Disorders	*Child Nutrition*
Weight Control	Vitamins	Older Adult Health
Nutrition for Older Adults	Exercise for Older Adults	Vitamins
Depression	*Child Nutrition*	Mental Disorders
End of Life Issues	Sports Fitness	Caregivers
Eating Disorders	Older Adult Health	Child Care
Suicide	Caregivers	Sports Fitness

Mental Health	Child Care	*Child Mental Health*
Older Adult Mental Health	*Teen Mental Health*	*Teen Mental Health*
Vitamins	Weight Control	Exercise for Older Adults
Sports Fitness	Sports Injuries	*Mental Health*
Exercise for Older Adults	Eating Disorders	Sports Injuries
Exercise and Physical Fitness	Health Insurance	Cholesterol
Patient Rights	Cholesterol	Nutrition for Older Adults
Caregivers	End of Life Issues	Weight Control
Home Care Services	*Child Mental Health*	End of Life Issues
Talking With Your Doctor	Managed Care	Family Issues
Sports Injuries	Nutrition for Older Adults	Home Care Services

Table 61. Betweenness rankings among the three networks

From the ranking lists displayed in Tables 57-61 regarding the three node-level centrality measures among the original structural network, the revised semantic-based network, and the adjusted core topic-based network, it is clear to see that the five core topics improved to higher rankings and these changes led the node centrality ranking lists of the adjusted network to become more consistent with the ranking lists of the revised semantic-based network. In other words, identifying core health topics might greatly assist a subject directory in better reflecting its topics' semantic relationships.

4.6.2   RQ5.1 & Hypothesis group 7

RQ5.1 is described as "Are there any significant differences between the original and optimized structural networks of older adults related topics in terms of similarity values?" It

211

examines if similarity value differences appeared after the optimization process.

The similarity values of the selected older adults related health topics were not able to be analyzed through the standard T-test since they did not have a normal distribution. Therefore, the Mann-Whitney test was applied to investigate the differences of similarity values between the original and optimized subject directories. SPSS was employed to conduct the inferential statistical analysis toward the older adults related topics.

RQ5.1 was answered by hypothesis group 7. This hypothesis group contained only one hypothesis. The Mann-Whitney test was applied to test the hypothesis in hypothesis group 7. Table 62 summarizes the mean rank and sum of ranks of the similarity values from the original structural network (n=446) and the optimized structural network (n=690). The statistical analysis revealed that there was a systematic difference of similarity values between the original and optimized structural networks of older adults related topics (p=0.000<0.05). Therefore, the hypothesis was rejected. In other words, the similarity value of the connections was significantly improved after the optimization process.

	Similarity value	
	Original structural network	Optimized structural network
Mean rank	516.43	602.15
Sum of ranks	230330.00	415486.00
z-statistic	-4.300	
p-value	0.000	

Table 62. Statistical analysis result for hypothesis group 7

4.6.3   RQ5.2 & Hypothesis group 8

RQ5.2 is stated as "Are there any significant differences between the original and

212

optimized structural networks of older adults related topics in terms of node-level centrality measurements?" This question aims at investigating if any differences of node-level centrality measures occurred between the original and optimized structural network after the optimization process.

Similar with the other three topic subcategories, a series of customized T-tests were applied to investigate the differences of three node centrality features between the original and optimized subject directories. UciNet was employed to conduct the inferential statistical analysis for the older adults related topics.

RQ5.2 was answered by hypothesis group 8, and this hypothesis group contained three sub-hypotheses. Each sub-hypothesis under hypothesis group 8 was aimed at comparing one of the three node-level centrality measures (degree centrality, closeness centrality, and betweenness centrality) between the original and optimized subject structures.

A series of customized T-tests were applied to test the three sub-hypotheses under hypothesis group 8. Tables 63-65 include the means and standard deviations (SD) of degree centrality, closeness centrality, and betweenness centrality of all original structural connections (n=150) and all optimized structural connections (n=150). In Table 63, the p-values of both the out-degree centrality and in-degree centrality were smaller than the significant level (0.05), which indicated that there was a significant difference of the degree centrality measures between the original structural network and the optimized structural network. That is, the degree centrality measures in the optimized structural network had generated a significantly higher average value than the original structural network.

213

	Out-Degree		In-Degree	
	Original structural network	*Optimized structural network*	*Original structural network*	*Optimized structural network*
Mean	*0.020*	*0.031*	*0.020*	*0.031*
Standard deviation (SD)	*0.015*	*0.022*	*0.023*	*0.027*
t-statistic	*-0.011*		*-0.011*	
p-value	***0.0001***		***0.0001***	

Table 63. Degree centrality rankings between the original and optimized networks

In Table 64, the p-values of both the out-closeness centrality and in-closeness centrality were smaller than the significant level (0.05) as well, which indicated that there was a significant difference of the closeness centrality measures between the original structural network and the optimized structural network. That is, the closeness centrality measures in the optimized structural network had generated a significantly higher average value than the original structural network.

	Out-Closeness		In-Closeness	
	Original structural network	*Optimized structural network*	*Original structural network*	*Optimized structural network*
Mean	*0.106*	*0.189*	*0.142*	*0.218*
Standard deviation (SD)	*0.006*	*0.021*	*0.070*	*0.070*
t-statistic	*-0.083*		*-0.076*	
p-value	***0.0001***		***0.0001***	

Table 64. Closeness centrality rankings between the original and optimized networks

In Table 65, unlike the degree and closeness centrality measures, the p-value of the betweenness centrality was larger than the significant level (0.05), which indicated that there was no significant difference of the betweenness centrality measure between the original structural network and the optimized structural network. That is, the betweenness centrality measure in the optimized structural network had not generated a significantly higher average value than the original structural network. This result echoed the findings investigated in the optimized

214

networks of the other three topic groups.

	Betweenness	
	Original structural network	Optimized structural network
Mean	2.367	1.868
Standard deviation (SD)	5.261	4.210
t-statistic	0.498	
p-value	0.3739	

Table 65. Betweenness centrality rankings between the original and optimized networks

As a result, for hypothesis group 8, H08$_{(a)}$ and H08$_{(b)}$ were rejected while H08$_{(c)}$ failed to be rejected.

4.6.4   Summary

The fifth research question concentrated on comparing the original structural network and the optimized structural network of older adults related health topics in the subject directory on MedlinePlus. The introductory section examined the optimized structural network in terms of structural connections, semantic connections, and key nodes in the network, while RQ5.1 and 5.2 investigated the differences between the original and optimized structural networks regarding their connections' similarity values as well as node-level centrality measures.

It was found that the recommended new structural connections were mostly added to the mental health related topics, the three health consumer groups related topics (children, teenagers, and older adults), and two specific daily health elements related topics (vitamins and cholesterol). These recommendations were partially echoed by the five core topics identified through the process of identifying key nodes in the revised semantic-based network – the five core topics recognized were covering health topics relating to children, teenagers, and mental health. This finding had also proved the strong relationships existing among mental health,

215

children, teenagers, and older adults. Moreover, the average similarity value of the structural connections in the optimized network was improved after new connections were added. This difference was found to be significant through the Mann-Whitney test. The five core health topics were identified to have a huge impact in the whole network again in the older adults topic subcategory. It proved the conclusion that through the creation of key nodes' connections, the node-level centrality ranking lists in the adjusted structural network moved closer to the semantic-based network. In other words, key influential nodes were more capable of reflecting the semantic relationships in the network compared with normal nodes. In addition, the node centrality measures were tested through the customized T-test, the results showed that the degree and closeness centrality measures were significantly increased in the optimized structural network while the betweenness centrality measure was not significantly different from the original structural link network. This conclusion also fit into the earlier results generated in the network of the mental health topic subcategory, the children topic subcategory, and the teenagers topic subcategory.

4.7 Findings for research question 6 (RQ6)

RQ6: Are there any significant differences between the optimized structural network from this study and the evaluation results from the health field experts?

After optimization results were generated in this study, two evaluators were invited to assess the optimization results. Both of the two evaluators were recruited from a formal research institute in the United States; they are research fellows in the Department of Dermatology of the University of California-Davis. One of them holds a Doctoral degree in the Predictive Medicine field while the other had finished taking all the coursework of a M.D. program. Both evaluators

216

have 5 - 10 years of experience in the field.

Four lists that contained 100 - 200 paired health topics were generated for each of the four health topic subcategories. Each list consisted of two sub-groups of connections: 1) the recommended connections generated according to the cosine similarity measure in this study; 2) about 60 pairs of health topics that had the lowest similarity values and were not structurally connected by the portal creators. The second sub-group of connections was added for comparing purposes. A screenshot of a health topic's page including its "related health topics" list, along with a brief introduction about how the MedlinePlus portal creates and displays structural linkages among related health topics were provided to the evaluators. The evaluators were then asked to identify and mark the pairs of health topics that they considered as related.

RQ6 was answered by hypothesis group 9, and this hypothesis group contained four sub-hypotheses. Each sub-hypothesis was applied to investigate the evaluation result for a specific health topic subcategory.

4.7.1   Mental health related health topics

For recommendation results regarding mental health related health topics, one evaluator confirmed 80 relevant pairs out of the 149 connections listed (53.69%) while the other evaluator confirmed relevant 82 pairs (55.03%).

Firstly, the consistency of the evaluation outcomes between the two evaluators were investigated. A Kappa test was performed between the two evaluation results and the "Measure of Agreement" value was 0.838 ($p < 0.001$), which indicated that a substantial agreement had been reached between the two evaluators.

217

Following that, two statistical methods were then performed to ensure that the recommendation results from this study and the evaluation outcomes from the two health field experts were consistent as well. Therefore, the two evaluation results from the two health professionals were combined and another Kappa test was employed to examine the consistency between the combined evaluation list and the corresponding recommended results from this study. The "Measure of Agreement" value was 0.743 ($p < 0.001$), which also achieved a moderate agreement.

Along with the Kappa test, a Chi-square test was employed to compare the combined evaluation list from the two evaluators and the corresponding recommended results from this study to further check the agreement reached between the results from this study and the results from the two health professionals. The Pearson Chi-square value, *df*, and *p*-value were 1.100, 1, and 0.294, respectively. The test results showed there was no significant difference between the results from this study and the results from the two evaluators. It suggested that the recommended results in this study were consistent with those from the expert evaluators.

4.7.2   *Children related health topics*

For recommendation results regarding children related health topics, one evaluator confirmed 115 relevant pairs out of the 179 connections listed (64.25%) while the other evaluator confirmed 113 relevant pairs (63.13%).

Similar procedures applied for the mental health subcategory were followed again toward the children related health topics. First, the consistency of the evaluation outcomes between the two evaluators were investigated. A Kappa test was performed between the two evaluation

results and the "Measure of Agreement" value was 0.952 ($p < 0.001$), which indicated that a substantial agreement had been reached between the two evaluators.

Following that, two statistical methods were then performed to ensure that the recommendation results from this study and the evaluation outcomes from the two health field experts were consistent as well. Therefore, the two evaluation results from the two health professionals were combined and another Kappa test was employed to examine the consistency between the combined evaluation list from the two evaluators and the corresponding recommended results from this study. The "Measure of Agreement" value was 0.951 ($p < 0.001$), which also achieved a perfectly strong agreement.

Along with the Kappa test, a Chi-square test was employed to compare the combined evaluation list from the two evaluators and the corresponding recommended results from this study to further check the agreement reached between the results from this study and the results from the two health professionals. The Pearson Chi-square value, *df*, and *p*-value were 0.049, 1, and 0.825, respectively. The test results showed there was no significant difference between the results from this study and the results from the two evaluators. It suggested that the recommended results in this study were consistent with those from the expert evaluators.

4.7.3   *Teenagers related health topics*

For recommendation results regarding teenagers related health topics, one evaluator confirmed 148 relevant pairs out of the 214 connections listed (69.16%) while the other evaluator confirmed 147 relevant pairs (68.69%).

Similar procedures applied for the mental health and children subcategories were

followed again toward the teenagers related health topics. First, the consistency of the evaluation outcomes between the two evaluators were investigated. A Kappa test was performed between the two evaluation results and the "Measure of Agreement" value was 0.989 ($p < 0.001$), which indicated that a substantial agreement had been reached between the two evaluators.

Following that, two statistical methods were then performed to ensure that the recommendation results from this study and the evaluation outcomes from the two health field experts were consistent as well. Therefore, the two evaluation results from the two health professionals were combined and another Kappa test was employed to examine the consistency between the combined evaluation list from the two evaluators and the corresponding recommended results from this study. The "Measure of Agreement" value was 0.826 ($p < 0.001$), which also achieved a strong agreement.

Along with the Kappa test, a Chi-square test was employed to compare the combined evaluation list from the two evaluators and the corresponding recommended results from this study to further check the agreement reached between the results from this study and the results from the two health professionals. The Pearson Chi-square value, *df*, and *p*-value were 0.043, 1, and 0.835, respectively. The test results showed there was no significant difference between the results from this study and the results from the two evaluators. It suggested that the recommended results in this study were consistent with those from the expert evaluators.

4.7.4 *Older adults related health topics*

For recommendation results regarding older adults related health topics, one evaluator confirmed 142 relevant pairs out of the 219 connections listed (64.84%) while the other

220

evaluator confirmed 143 relevant pairs (65.30%).

Similar procedures applied for the other three subcategories were followed again toward the older adults related health topics. First, the consistency of the evaluation outcomes between the two evaluators were investigated. A Kappa test was performed between the two evaluation results and the "Measure of Agreement" value was 0.930 ($p < 0.001$), which indicated that a substantial agreement had been reached between the two evaluators.

Following that, two statistical methods were then performed to ensure that the recommendation results from this study and the evaluation outcomes from the two health field experts were consistent as well. Therefore, the two evaluation results from the two health professionals were combined and another Kappa test was employed to examine the consistency between the combined evaluation list from the two evaluators and the corresponding recommended results from this study. The "Measure of Agreement" value was 0.836 ($p < 0.001$) for the recommendation results, which achieved a strong agreement.

Along with the Kappa test, a Chi-square test was employed to compare the combined evaluation list from the two evaluators and the corresponding recommended results from this study to further check the agreement reached between the results from this study and the results from the two health professionals. The Pearson Chi-square value, *df*, and *p*-value were 0.000, 1, and 1.000, respectively. The test results showed there was no significant difference between the results from this study and the results from the two evaluators. It suggested that the recommended results in this study were consistent with those from the expert evaluators.

221

H09$_{(a)}$, H09$_{(b)}$, H09$_{(c)}$, and H09$_{(d)}$ were all failed to be rejected. It indicated that the recommended structural connections generated for the four health topic subcategories by this study were consistent with the evaluation results from the two health field experts.

4.8 Results summary

Through the investigation toward the similarity values contained in the original structural network, inconsistencies were uncovered from all the four health topic groups in RQ1 – there were hundreds of pairs of health topics that were sharing strong semantic relationships but being ignored by the MedlinePlus portal creators when considering building structural connections. Therefore, recommendations were generated based on the threshold values derived from the structurally linked health topics.

Following that, RQ2–5 revealed the differences between the original and optimized structural networks of mental health, children, teenagers, and older adults related health topics. The optimized networks were formed of new structural connections, stronger semantic relationships, as well as newly recognized influential health topics. The similarity values and node centrality measures of the optimized structural network were significantly improved compared with the original structural networks.

Last but not least, the optimization recommendations proposed by this study were evaluated by two health field experts. Their evaluation outcomes confirmed that the suggested connections generated by this study fit into professional assessments. In other words, the recommendation results were supported by both the semantic relationships and professional

judgements.

Chapter 5. Discussion and implications

This section includes a discussion of the impact of various thresholds applied for optimization, a comparison among the three similarity measures, and a discussion regarding the intersections as well as key nodes identified among the four health topic groups. In addition, the revised semantic network is compared with the prior studies. Last but not least, both the theoretical and practical implications of the findings from this research study are presented.

5.1 The impact of different thresholds on recommendation results

During the process of optimization, the suggested connections were recognized through the comparison between the semantic relationships among those health topics that were not connected structurally and a threshold. This threshold served as the foundation when identifying unconnected health topics with strong semantic relationships since it reflected the facts about how the original structural link network was built. When applying the cosine similarity measure, the threshold was the average similarity value of the structurally connected health topics.

For comparing purposes, the semantic networks of the four health topic groups were processed through UciNet using the Pearson correlation similarity measure and the Euclidean distance similarity measure. Since the original similarity values generated by the Euclidean distance similarity measure were larger than 1, these original similarity values were than divided by 1 to get normalized values. The thresholds were then determined by the similarity values and recommendations were identified within the same health topic networks of mental health, children, teenagers, and older adults subcategories.

224

5.1.1 Mental health related topics

Table 66 displays the average similarity values of the original structural link network of mental health related health topics generated through the three similarity measures.

	Average All	Average Connected	Average Unconnected
Cosine similarity measure	0.105879406	0.383676923	0.098229824
Pearson correlation measure	0.083203463	0.366607692	0.075399492
Euclidean distance measure	0.024334813	0.028425028	0.024222183

Table 66. Similarity values regarding mental health related topics generated by the three similarity measures

According to the similarity value calculated toward the structurally connected health topics, two groups of pairs of health topics were identified to possess strong semantic relationships with no structural connections created. These two groups of pairs of semantically related health topics were investigated by the Pearson correlation and the Euclidean distance similarity measures, respectively.

To be more specific, the threshold set for the health topics to be added into the optimized network based on the Pearson correlation similarity analysis was the average similarity value of structurally linked health topics. As a result, 135 connections were discovered and suggested to be added into the optimized structural network of mental health related health topics. Among these 135 connections, 112 connections were bidirectional while 23 connections were unidirectional.

Unlike the recommendations concluded through the cosine and the Pearson correlation similarity measures, the average similarity value did not work for optimizing the original

225

structural network through the Euclidean distance similarity measure. The reason was that after the normalization process, the similarity values were too close to each other. If using the average similarity value as the threshold, 2,264 connections would be recommended to the optimized structural link network. Compared with the numbers of all the connections in the network (9,702) and the original structural connections created in the network (260), the number of the suggested connections was too large. Therefore, the threshold was then adjusted to further filter the health topics that possessed similarity values larger than the average number. Figure 22 shows the number of left recommended connections with their corresponding thresholds calculated based on the Euclidean distance similarity measure. It indicated that when applying the average similarity value of all the structurally connected health topics, there were 2,264 connections left; when applying the similarity value of the connection that was ranked in the 100th of all the structurally connected health topics, the suggested connections decreased to 1,696; when applying the similarity value of the 50th connection, there were 440 recommendations left; when applying the similarity value of the connection which was ranked in the 25th of all the structurally connection health topics, the suggested connections were lowered to 223 (214 bidirectional connections and 9 unidirectional connections). In order to maintain a similar recommendation number with the other two similarity measures, the 25th similarity value was finally selected to serve as the threshold for the optimized structural network built through the Euclidean distance similarity measure.
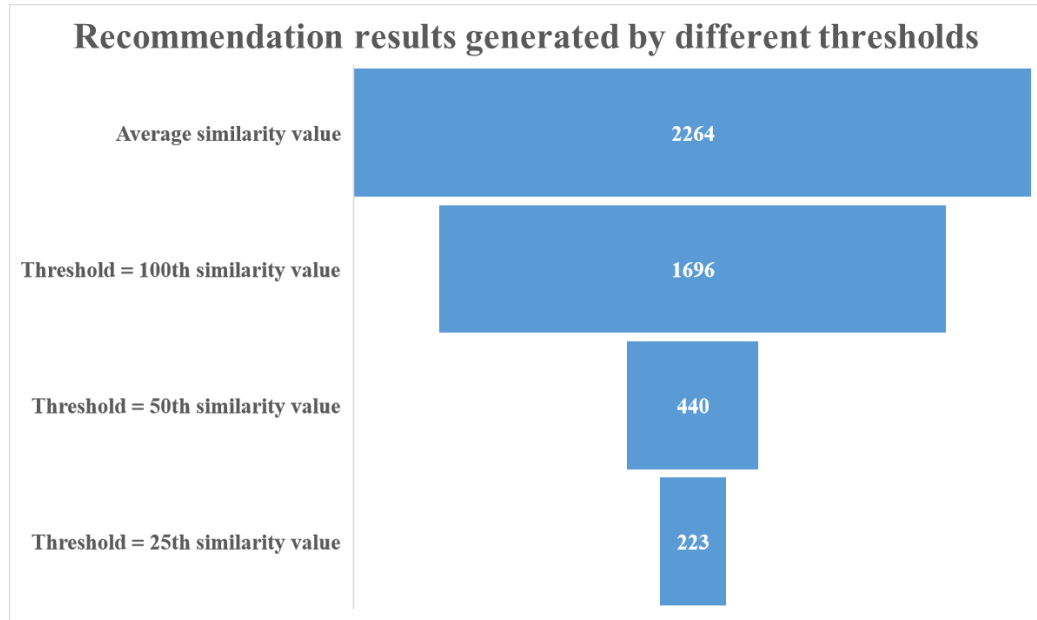
Figure 22. Recommendation results generated by different thresholds

In conclusion, regarding the optimized structural network of mental health related health topics, 133 new connections were suggested by the cosine similarity measure, 135 new connections were recommended by the Pearson correlation similarity measure, and 223 new connections were proposed by the Euclidean distance similarity measure. Interestingly enough, the recommended connections identified through the cosine and the Pearson correlation similarity measures were found to possess a great overlap – 131 connections were recommended by both two similarity measures. The only pair of bidirectional connection that was uniquely discovered by the cosine similarity measure was between topics *Nutrition for Seniors* and *Weight Control*, while the two unique pairs of bidirectional connections investigated through the Pearson correlation similarity measure being between *Child Development* and *School Health*, and between *Infant and Newborn Development* and *Infant and Newborn Nutrition*. Compared with the suggestions proposed by the Pearson correlation similarity measure, the results concluded

227

through the Euclidean distance measure only contained 7 common connections with the results of the cosine similarity measure. The rest of the 216 connections uncovered by the Euclidean distance similarity measure were unique.

According to the connections recommended by both the Pearson correlation and the Euclidean distance similarity measures, the same earlier procedures were followed again. Weighted scores were assigned to health topics that were found to be ranked differently between the original structural network and the revised semantic-based network. As a result, five core health topics were identified again for each of the two similarity measures. Table 67 shows the five core health topics generated by the three similarity measures among their recommended connections.

	Cosine similarity measure	Pearson correlation similarity measure	Euclidean distance similarity measure
Core health topic 1	*Obesity in Children*	*Obesity in Children*	*Compulsive Gambling*
Core health topic 2	*Child Nutrition*	*Child Nutrition*	*Inhalants*
Core health topic 3	*Diets*	*Diets*	*Personality Disorders*
Core health topic 4	*Child Development*	*Child Development*	*Malnutrition*
Core health topic 5	*Nutrition for Seniors*	*School Health*	*College Health*

Table 67. five core health topics identified through the three similarity measures

According to this table, it is clear that the 5 key nodes in the mental health topics' network identified by the cosine and the Pearson correlation similarity measures were similar. The core topics explored were mostly focusing on specific health consumer groups including children and older adults. Besides specific health consumer groups, daily health elements such as *Diets* were also revealed. The only difference was about the last core health topic – the cosine similarity measure recognized the topic *Nutrition for Seniors* while the Pearson correlation measure investigated the topic *School Health*. This difference might be caused by the unique suggested connections discovered by these two similarity measures. Different from the other two

228

similarity measures, the five core health topics identified through the Euclidean distance similarity measure were mostly concentrating on specific mental diseases and daily health elements. The complete list of recommendations proposed by the Euclidean distance similarity measure could be found in Appendix 17.

Moreover, regarding the connections of the five core topics identified through the Pearson correlation and the Euclidean distance similarity measures, the results were quite different. The Pearson correlation similarity measure explored 131 common recommendations and only 1 different core topic from the cosine similarity measure's findings. Hence, it was not surprised to find that the number of the connections relating to the five core health topics generated by the Pearson correlation similarity measure was very close to the results of the cosine similarity measure – 64 connections were suggested in total, excluding the overlapping connections among the five core topics. However, for the results concluded by the Euclidean distance similarity measure, it was found that the number of connections relating to the five core health topics reached to 162, excluding the overlapping connections among the five core topics. These connections weighted 72.65% of all the recommended connections. In other words, the suggested connections generated through the Euclidean distance similarity measure were highly clustered among the 5 key health topics in the mental health related network. The health topic *Compulsive Gambling* was suggested to be linked to other 34 health topics, including topics like *Eating Disorders* and *Malnutrition*. These recommended connections were not very convincing.

5.1.2   *Children related topics*

Table 68 displays the average similarity values of the original structural link network of children related health topics generated through the three similarity measures.

229

	Average All	*Average Connected*	*Average Unconnected*
Cosine similarity measure	0.093318287	0.444743644	0.086589168
Pearson correlation measure	0.07667017	0.434135593	0.069825396
Euclidean distance measure	0.02093257	0.026847407	0.020819312

Table 68. Similarity values regarding children related topics generated by the three similarity measures

According to the similarity value calculated toward the structurally connected health topics, two groups of pairs of health topics were identified to possess strong semantic relationships with no structural connections created. These two groups of pairs of semantically related health topics were identified by the Pearson correlation and the Euclidean distance similarity measures, respectively.

To be more specific, the threshold set for the health topics to be added into the optimized network based on the Pearson correlation similarity analysis was the average similarity value of structurally linked health topics. As a result, 177 connections were discovered and suggested to be added into the optimized structural network of children related health topics. Among these 177 connections, 134 connections were bidirectional while 43 connections were unidirectional.

Unlike the recommendations concluded through the cosine and the Pearson correlation similarity measures, the average similarity value did not work for optimizing the original structural network through the Euclidean distance similarity measure toward children related health topics either. The similarity values again, were too close to each other. If using the average similarity value as the threshold, 3640 connections would be recommended to the optimized structural link network. Compared with the numbers of all the connections in the network (25,122) and the original structural connections created in the network (472), the number of the

230

suggested connections was too large. Therefore, the threshold was then adjusted to further filter the health topics that were possessing similarity values larger than the average number. Figure 23 shows the number of left recommended connections with their corresponding thresholds calculated based on the Euclidean distance similarity measure. It indicated that when applying the average similarity value of all the structurally connected health topics, there were 3,640 connections left; when applying the similarity value of the connection that was ranked in the 100th of all the structurally connected health topics, the suggested connections decreased to 983; when applying the similarity value of the 50th connection, there were 320 recommendations left; when applying the similarity value of the connection which was ranked in the 25th of all the structurally connection health topics, the suggested connections were lowered to 119 (116 bidirectional connections and 3 unidirectional connections). In order to maintain a similar recommendation number with the other two similarity measures, the 25th similarity value was finally selected to serve as the threshold for the optimized structural network built through the Euclidean distance similarity measure.
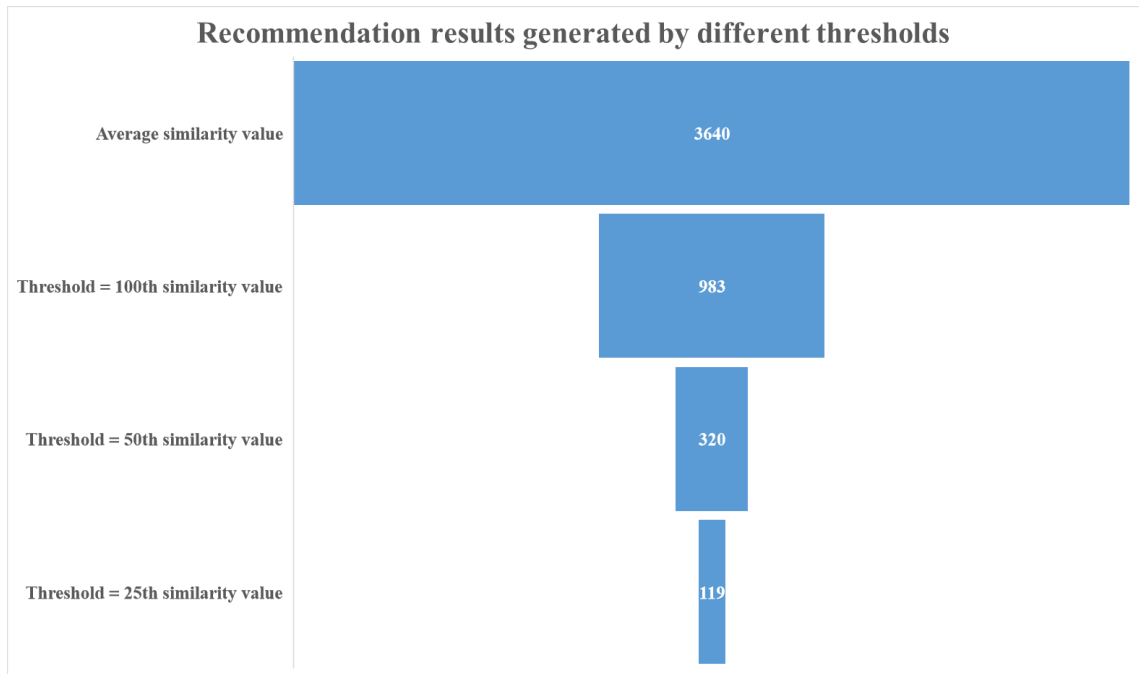
Figure 23. Recommendation results generated by different thresholds

In conclusion, regarding the optimized structural network of children related health topics, 184 new connections were suggested by the cosine similarity measure, 177 new connections were recommended by the Pearson correlation similarity measure, and 119 new connections were proposed by the Euclidean distance similarity measure. Within the children related health topics network, the overlap condition was uncovered again between the recommendation results generated by the cosine and the Pearson correlation similarity measures. Different from what was found in mental health related health topics where each of the two similarity measures still possessed their unique recommendations, for children related health topics, the suggested connections proposed by the Pearson correlation similarity measure were completely involved in the suggested connections investigated by the cosine similarity measure. In other words, the recommendations of the Pearson correlation similarity measure were a subset

232

www.manaraa.com

of the recommendations of the cosine similarity measure. The cosine similarity measure explored 6 more bidirectional connections and 1 additional unidirectional connection than the Pearson correlation similarity measure. Moreover, these 7 additional connections were not related to any of the five core health topics identified through the cosine similarity measure in the earlier stage. Compared with the suggestions proposed by the Pearson correlation similarity measure, the results concluded through the Euclidean distance measure only contained 4 common connections with the results of the cosine similarity measure. The rest of the 115 connections uncovered by the Euclidean distance similarity measure were unique.

According to the connections recommended by both the Pearson correlation and the Euclidean distance similarity measures, the same earlier procedures were followed again. Weighted scores were assigned to health topics that were found to be ranked differently between the original structural network and the revised semantic-based network. As a result, five core health topics were identified again for each of the two similarity measures. Table 69 shows the five core health topics generated by the three similarity measures among their recommended connections.

	Cosine similarity measure	Pearson correlation similarity measure	Euclidean distance similarity measure
Core health topic 1	*Medicines and Children*	*Medicines and Children*	*Dwarfism*
Core health topic 2	*Children's Health*	*Children's Health*	*Mobility Aids*
Core health topic 3	*Toddler Health*	*Toddler Health*	*Artificial Limbs*
Core health topic 4	*Child Development*	*Child Development*	*Barotrauma*
Core health topic 5	*Child Mental Health*	*Child Mental Health*	*Baby Health Checkup*

Table 69. five core health topics identified through the three similarity measures

According to this table, it is not surprising to find that the 5 key nodes in the children topics' network identified by the cosine and the Pearson correlation similarity measures were the same. The core topics explored were mostly focusing on specific health consumer groups

233

including children and toddlers. These five key nodes were closely related to the children subcategory. Different from the other two similarity measures, the five core health topics identified through the Euclidean distance similarity measure were mostly concentrating on specific diseases, devices, and medical checking procedures. The complete list of recommendations proposed by the Euclidean distance similarity measure could be found in Appendix 18.

Moreover, regarding the connections of the five core topics identified through the Pearson correlation and the Euclidean distance similarity measures, the results were quite different. The Pearson correlation similarity measure explored 177 common recommendations from the cosine similarity measure's findings. Hence, it was not surprised to find that the number of the connections relating to the five core health topics generated by the Pearson correlation similarity measure was the same with the results of the cosine similarity measure. However, for the results concluded by the Euclidean distance similarity measure, it was found that the number of connections relating to the five core health topics reached to 100, excluding the overlapping connections among the five core topics. These connections weighted 84.03% of all the recommended connections. In other words, the suggested connections generated through the Euclidean distance similarity measure were highly clustered among the 5 key health topics in the children related network. This finding echoed the conclusion generated toward the mental health related network. The health topic *Dwarfism* was suggested to be linked to other 16 topics, including topics like *Dentures* and *Cosmetic Dentistry*. These recommended connections were not very convincing.

Table 70 displays the average similarity values of the original structural link network of teenagers related health topics generated through the three similarity measures.

	Average All	*Average Connected*	*Average Unconnected*
Cosine similarity measure	*0.111522199*	*0.461491003*	*0.104383724*
Pearson correlation measure	*0.094826105*	*0.45129563*	*0.087555031*
Euclidean distance measure	*0.01926639*	*0.022770394*	*0.019194918*

Table 70. Similarity values regarding teenagers related topics generated by the three similarity measures

According to the similarity value calculated toward the structurally connected health topics, two groups of pairs of health topics were identified to possess strong semantic relationships with no structural connections built. These two groups of pairs of semantically related health topics were identified by the Pearson correlation and the Euclidean distance similarity measures, respectively.

To be more specific, the threshold set for the health topics to be added into the optimized network based on the Pearson correlation similarity analysis was the average similarity value of structurally linked health topics. As a result, 248 connections were discovered and suggested to be added into the optimized structural network of teenagers related health topics. Among these 248 connections, 196 connections were bidirectional while 52 connections were unidirectional.

Unlike the recommendations concluded through the cosine and the Pearson correlation similarity measures, the average similarity value did not work for optimizing the original structural network through the Euclidean distance similarity measure toward teenagers related

235

health topics either. The similarity values were too close to each other similarly with mental health and children related health topics. If using the average similarity value as the threshold, 5,225 connections would be recommended to the optimized structural link network. Compared with the numbers of all the connections in the network (19,460) and the original structural connections created in the network (389), the number of the suggested connections was too large. Therefore, the threshold was then adjusted to further filter the health topics that were possessing similarity values larger than the average number. Figure 24 shows the number of left recommended connections with their corresponding thresholds calculated based on the Euclidean distance similarity measure. It indicated that when applying the average similarity value of all the structurally connected health topics, there were 5,225 connections left; when applying the similarity value of the connection that was ranked in the 100th of all the structurally connected health topics, the suggested connections decreased to 2,094; when applying the similarity value of the 50th connection, there were 376 recommendations left; when applying the similarity value of the connection which was ranked in the 25th of all the structurally connection health topics, the suggested connections were lowered to 168 (162 bidirectional connections and 6 unidirectional connections). In order to maintain a similar recommendation number with the other two similarity measures, the 25th similarity value was finally selected to serve as the threshold for the optimized structural network built through the Euclidean distance similarity measure.
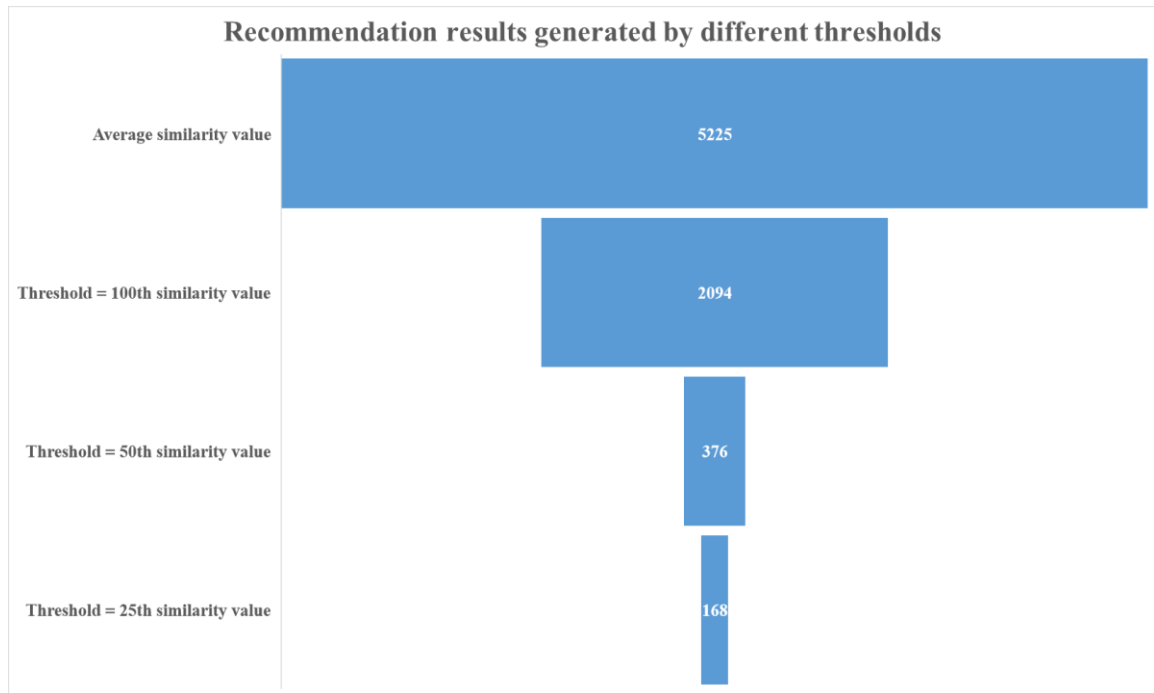
Figure 24. Recommendation results generated by different thresholds

In conclusion, regarding the optimized structural network of teenagers related health topics, 241 new connections were suggested by the cosine similarity measure, 248 new connections were recommended by the Pearson correlation similarity measure, and 168 new connections were proposed by the Euclidean distance similarity measure. Within the teenagers related health topics network, the overlap condition was uncovered again between the recommendation results generated by the cosine and the Pearson correlation similarity measures. Different from what was found in children related health topics where the recommendation results identified by the Pearson correlation similarity measure became a subset of the recommendation results identified by the cosine similarity measure, for teenagers related health topics, the suggested connections proposed by the cosine similarity measure were completely involved in the suggested connections investigated by the Pearson correlation similarity measure.

237

In other words, the recommendations of the cosine similarity measure were a subset of the recommendations of the Pearson correlation similarity measure. The Pearson correlation similarity measure explored 6 more bidirectional connections and 1 additional unidirectional connection than the cosine similarity measure. Moreover, these 7 additional connections were not related to any of the five core health topics identified through the cosine similarity measure in the earlier stage. Compared with the suggestions proposed by the Pearson correlation similarity measure, the results concluded through the Euclidean distance measure only contained 9 common connections with the results of the Pearson correlation similarity measure. The rest of the 159 connections uncovered by the Euclidean distance similarity measure were unique.

According to the connections recommended by both the Pearson correlation and the Euclidean distance similarity measures, the same earlier procedures were followed again. Weighted scores were assigned to health topics that were found to be ranked differently between the original structural network and the revised semantic-based network. As a result, five core health topics were identified again for each of the two similarity measures. Table 71 shows the five core health topics generated by the three similarity measures among their recommended connections.

	Cosine similarity measure	Pearson correlation similarity measure	Euclidean distance similarity measure
Core health topic 1	*Teenage Pregnancy*	*Teenage Pregnancy*	*Compulsive Gambling*
Core health topic 2	*Infections and Pregnancy*	*Infections and Pregnancy*	*Dwarfism*
Core health topic 3	*Pregnancy and Drug Use*	*Pregnancy and Drug Use*	*Postpartum Care*
Core health topic 4	*Pregnancy and Medicines*	*Pregnancy and Medicines*	*Fetal Health and Development*
Core health topic 5	*Health Problems in Pregnancy*	*Health Problems in Pregnancy*	*Motor Vehicle Safety*

Table 71. five core health topics identified through the three similarity measures

238

According to this table, it is not surprising to find that the 5 key nodes in the teenagers topics' network identified by the cosine and the Pearson correlation similarity measures were the same. The core topics explored were all focusing on pregnancy related health topics. These five key nodes were closely related to the teenagers subcategory. Different from the other two similarity measures, the five core health topics identified through the Euclidean distance similarity measure were distributed in various fields. These fields include specific mental diseases, children related health conditions, pregnancy, and daily health elements. The complete list of recommendations proposed by the Euclidean distance similarity measure could be found in Appendix 19.

Moreover, regarding the connections of the five core topics identified through the Pearson correlation and the Euclidean distance similarity measures, the results were quite different. The Pearson correlation similarity measure explored 241 common recommendations from the cosine similarity measure's findings. Hence, it was not surprised to find that the number of the connections relating to the five core health topics generated by the Pearson correlation similarity measure was the same with the results of the cosine similarity measure. However, for the results concluded by the Euclidean distance similarity measure, it was found that the number of connections relating to the five core health topics reached to 149, excluding the overlapping connections among the five core topics. These connections weighted 88.69% of all the recommended connections. In other words, the suggested connections generated through the Euclidean distance similarity measure were highly clustered among the 5 key health topics in the teenagers related network. This finding echoed the conclusion generated toward the mental health and children related networks. Moreover, the topics *Compulsive Gambling* and *Dwarfism*

239

were identified as the first core topic of the mental health subcategory and the children

subcategory, respectively. In teenagers related health topics' network, these two topics were

regarded as core health topics again, with a combined 53 topics linked to them in the suggestion

list. Some suggested connections such as *Compulsive Gambling* and *Fetal Health and*

Development, as well as *Dwarfism* and *Elder Abuse*, were not very convincing.

5.1.4   *Older adults related topics*

Table 72 displays the average similarity values of the original structural link network of

older adults related health topics generated through the three similarity measures.

|	*Average All*	*Average Connected*	*Average Unconnected*
Cosine similarity measure	*0.104228*	*0.440717*	*0.097377*
Pearson correlation measure	*0.087436779*	*0.42994843*	*0.080462701*
Euclidean distance measure	*0.020158054*	*0.023983732*	*0.020080157*

Table 72. Similarity values regarding older adults related topics generated by the three similarity measures

According to the similarity value calculated toward the structurally connected health

topics, two groups of pairs of health topics were identified to possess strong semantic

relationships with no structural connections created. These two groups of pairs of semantically

related health topics were identified by the Pearson correlation and the Euclidean distance

similarity measures, respectively.

To be more specific, the threshold set for the health topics to be added into the optimized

network based on the Pearson correlation similarity analysis was the average similarity value of

structurally linked health topics. As a result, 244 connections were discovered and suggested to

240

be added into the optimized structural network of older adults related health topics. Among these 244 connections, 196 connections were bidirectional while 48 connections were unidirectional.

Unlike the recommendations concluded through the cosine and the Pearson correlation similarity measures, the average similarity value did not work for optimizing the original structural network through the Euclidean distance similarity measure toward older adults related health topics either. The similarity values were too close to each other similarly with mental health, children, and teenagers related health topics. If using the average similarity value as the threshold, 5,376 connections would be recommended to the optimized structural link network. Compared with the numbers of all the connections in the network (22,350) and the original structural connections created in the network (446), the number of the suggested connections was too large. Therefore, the threshold was then adjusted to further filter the health topics that were possessing similarity values larger than the average number. Figure 25 shows the number of left recommended connections with their corresponding thresholds calculated based on the Euclidean distance similarity measure. It indicated that when applying the average similarity value of all the structurally connected health topics, there were 5,376 connections left; when applying the similarity value of the connection that was ranked in the 100th of all the structurally connected health topics, the suggested connections decreased to 1,963; when applying the similarity value of the 50th connection, there were 596 recommendations left; when applying the similarity value of the connection which was ranked in the 25th of all the structurally connection health topics, the suggested connections were lowered to 248 (242 bidirectional connections and 6 unidirectional connections). In order to maintain a similar recommendation number with the other two similarity measures, the 25th similarity value was finally selected to serve as the threshold for the

optimized structural network built through the Euclidean distance similarity measure.
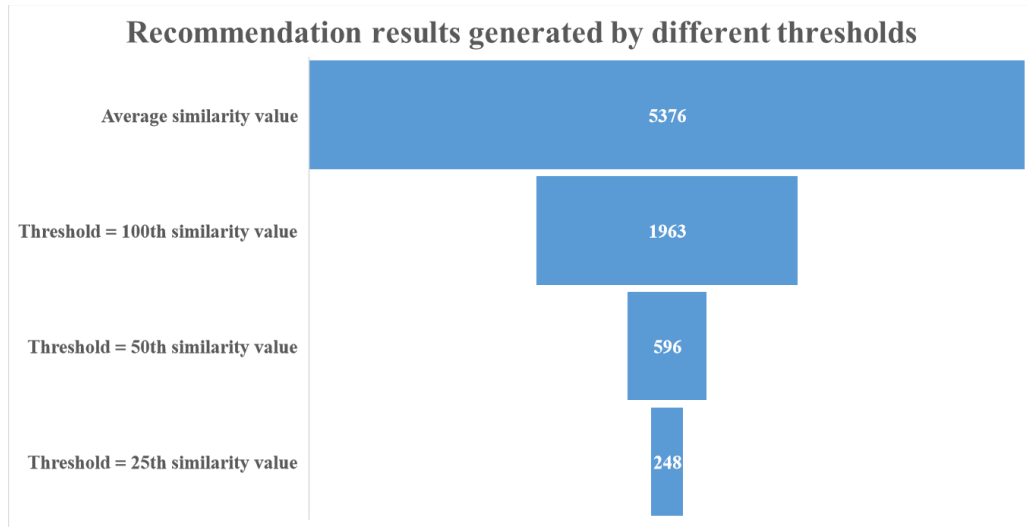
Figure 25. Recommendation results generated by different thresholds

In conclusion, regarding the optimized structural network of older adults related health topics, 244 new connections were suggested by the cosine similarity measure, 244 new connections were recommended by the Pearson correlation similarity measure, and 248 new connections were proposed by the Euclidean distance similarity measure. Within the older adults related health topics network, the overlap condition was uncovered again between the recommendation results generated by the cosine and the Pearson correlation similarity measures. Different from what was found in other health topic groups where the recommendation results identified between the Pearson correlation similarity and the cosine similarity measure being slightly different, for older adults related health topics, the suggested connections proposed by the cosine similarity measure were completely the same with the suggested connections investigated by the Pearson correlation similarity measure. This was a unique circumstance and indicated that the two similarity measures generated same analysis outcomes through different

242

mathematical methods. Compared with the suggestions proposed by the cosine and the Pearson correlation similarity measures, the results concluded through the Euclidean distance measure only contained 37 common connections with the results of the cosine and the Pearson correlation similarity measures. The rest of the 211 connections uncovered by the Euclidean distance similarity measure were unique.

According to the connections recommended by both the Pearson correlation and the Euclidean distance similarity measures, the same earlier procedures were followed again. Weighted scores were assigned to health topics that were found to be ranked differently between the original structural network and the revised semantic-based network. As a result, five core health topics were identified again for each of the two similarity measures. Table 73 shows the five core health topics generated by the three similarity measures among their recommended connections.

	Cosine similarity measure	Pearson correlation similarity measure	Euclidean distance similarity measure
Core health topic 1	*Child Nutrition*	*Child Nutrition*	*Compulsive Gambling*
Core health topic 2	*Healthy Aging*	*Healthy Aging*	*Financial Assistance*
Core health topic 3	*Teen Mental Health*	*Teen Mental Health*	*Malnutrition*
Core health topic 4	*Child Mental Health*	*Child Mental Health*	*Medicaid*
Core health topic 5	*Mental Health*	*Mental Health*	*Patient Rights*

Table 73. five core health topics identified through the three similarity measures

According to this table, it is expected to find that the 5 key nodes in the older adults topics' network identified by the cosine and the Pearson correlation similarity measures were the same. The core topics explored were primarily focusing on specific health consumer groups, mental health, and daily health elements. These five key nodes, excepting for *Healthy Aging*, were not closely related to the older adults subcategory. However, they indicated strong relationships among older adults, children, teenagers, and mental health since each of the topic

243

group had at least a related health topic involved in the list of core health topics for older adults related health topics. Different from the other two similarity measures, the five core health topics identified through the Euclidean distance similarity measure were distributed in various daily health elements and specific mental diseases. The complete list of recommendations proposed by the Euclidean distance similarity measure could be found in Appendix 20.

Moreover, regarding the connections of the five core topics identified through the Pearson correlation and the Euclidean distance similarity measures, the results were quite different. The Pearson correlation similarity measure explored all 244 common recommendations from the cosine similarity measure's findings. Hence, it was expected to find that the number of the connections relating to the five core health topics generated by the Pearson correlation similarity measure was the same with the results of the cosine similarity measure. However, for the results concluded by the Euclidean distance similarity measure, it was found that the number of connections relating to the five core health topics reached to 169, excluding the overlapping connections among the five core topics. These connections weighted 68.15% of all the recommended connections. In other words, the suggested connections generated through the Euclidean distance similarity measure were highly clustered among the 5 key health topics in the older adults related network. This finding echoed the conclusion generated toward the mental health, children, and teenagers related networks. Moreover, the topic *Compulsive Gambling* was identified as the first core topic of the mental health subcategory and the teenagers subcategory. In older adults related health topics' network, this topics was regarded as the first core health topic again, with suggested connections linking to 33 other health topics. Some suggested connections such as *Compulsive Gambling* and *Sports Fitness*, as well as *Compulsive Gambling*

244

and *Health Facilities*, were not very convincing.

5.2 The impact of different similarity methods on the optimization

In order to reveal the differences among the three selected similarity measures (the cosine similarity measure, the Pearson correlation measure, and the Euclidean distance measure), the same optimization procedures toward the four health topic groups (mental health, children, teenagers, and older adults) were performed again with the Pearson correlation similarity measure and the Euclidean distance similarity measure. The recommendation results for further improving the original structural link networks of the four health topic subcategories were generated again. These results were compared with the earlier optimization recommendations identified with the cosine similarity measure from two perspectives: an inferential statistical analysis regarding the similarity values of the suggested health topics among the three similarity measures, and another inferential statistical analysis toward the recommendation results concluded by each of the three similarity measurements.

5.2.1   Differences among the similarity values of the recommendation results generated with the three similarity measures

For each of the four health topic subcategories, the similarity values generated through the cosine, the Pearson correlation, and the Euclidean distance similarity measures were imported into SPSS. After that, a Kruskal-Wallis test was performed to check if there were any significant differences among the similarity values of the suggested connections calculated by the three similarity measures. Furthermore, a series of Mann-Whitney tests were applied to investigate if there were any significant differences of the similarity values between the cosine

245

similarity measure and the Pearson correlation similarity measure since their recommendation results had great overlap.

The results showed that there were significant differences among the similarity values calculated by the three similarity measures toward all the four health topic groups (p-value=0.000<0.05). Moreover, for mental health related health topics, there was no significant difference in terms of similarity values generated between the cosine and the Pearson correlation similarity measures (p-value=0.053>0.05). For children related health topics, there was no significant difference in terms of similarity values generated between the cosine and the Pearson correlation similarity measures either (p-value=0.429>0.05). For teenagers related health topics, there was significant difference in terms of similarity values generated between the cosine and the Pearson correlation similarity measures (p-value=0.045<0.05). Last but not least, for older adults related health topics, no significant difference was discovered in terms of similarity values generated between the cosine and the Pearson correlation similarity measures again (p-value=0.134>0.05).

In conclusion, the similarity values of the optimized structural network generated through the three similarity measures were significantly different among each of the four health topic subcategories. However, no significant differences were observed regarding the similarity values of the suggested connections calculated between the cosine and the Pearson correlation similarity measure in terms of the health topics relating to mental health, children, and older adults. For teenagers related health topics, the similarity values of the suggested connections generated through the cosine and the Pearson correlation similarity measures were significantly different.

246

5.2.2 Differences within the recommendation results in each of the three similarity measures in terms of the four health topic groups

After exploring the differences of the similarity values generated among the three similarity measures, a new series of Mann-Whitney tests were performed to further check if there were any significant differences among the similarity values of the suggested connections within each similarity measure's recommendation results. To be more specific, the comparisons were conducted toward the similarity values between the original and optimized structural networks generated within each similarity measure.

Table 74 summarizes the Mann-Whitney test results regarding the similarity values within the recommendation results of the cosine similarity measure. Table 75 summarizes the Mann-Whitney test results regarding the similarity values within the recommendation results of the Pearson correlation similarity measure. Table 76 summarizes the Mann-Whitney test results regarding the similarity values within the recommendation results of the Euclidean distance similarity measure.

	Mental health		Children		Teenagers		Older adults	
	Original	*Optimized*	*Original*	*Optimized*	*Original*	*Optimized*	*Original*	*Optimized*
Mean rank	298.26	346.01	528.76	590.21	467.08	536.50	516.43	602.15
Sum of ranks	77548.50	135982.50	249575	387181	181695	337995	230330	415486
z-statistic	-3.166		-3.125		-3.658		-4.300	
p-value	0.002		0.002		0.000		0.000	

Table 74. Summary of the Mann-Whitney test results in terms of similarity values between the original and optimized networks of the cosine similarity measure

	Mental health		Children		Teenagers		Older adults	
	Original	*Optimized*	*Original*	*Optimized*	*Original*	*Optimized*	*Origina l*	*Optimize d*

Mean rank	298.85	347.19	525.77	586.62	470.89	539.52	516.56	602.07
Sum of ranks	77701.50	137138.50	248164.50	380716.50	183174.50	343676.50	230385	415431
z-statistic	-3.199		-3.107		-3.600		-4.290	
p-value	0.001		0.002		0.000		0.000	

Table 75. Summary of the Mann-Whitney test results in terms of similarity values between the original and optimized networks of the Pearson correlation similarity measure

	Mental health		*Children*		*Teenagers*		*Older adults*	
	Original	*Optimized*	*Original*	*Optimized*	*Original*	*Optimized*	*Original*	*Optimized*
Mean rank	272.91	425.34	476.38	576.42	396.53	527.25	454.81	644.85
Sum of ranks	70957.50	205438.50	224851.50	340664.50	154251.50	293679.50	202845	447525
z-statistic	-9.233		-5.279		-7.240		-9.511	
p-value	0.000		0.000		0.000		0.000	

Table 76. Summary of the Mann-Whitney test results in terms of similarity values between the original and optimized networks of the Euclidean distance similarity measure

The results showed that there were significant differences in the recommendation results toward all the four health topic groups with the three similarity measures. In other words, the similarity values of the optimized structural link networks of health topics relating to mental health, children, teenagers and older adults had all been significantly improved after the optimization process through the cosine, the Pearson correlation and the Euclidean distance similarity measures.

5.2.3   Expert evaluation results toward the recommendation results from the Pearson correlation similarity measure

After recommendation results were identified through the Pearson correlation and the Euclidean distance similarity measures separately, the suggested connections were evaluated by the two health field experts following the same procedures, respectively. For mental health

248

related health topics, the "Measure of Agreement" value of the Kappa test was 0.732 ($p < 0.001$) for the recommendation results of the Pearson correlation similarity measure, which achieved a moderate agreement. The Pearson Chi-square value, *df*, and *p*-value were 0.859, 1, and 0.354, respectively for the recommendation results generated through the Pearson correlation similarity measure. The test results showed there was no significant difference between the results generated by this study and the two evaluators.

In terms of children related health topics, the "Measure of Agreement" value of the Kappa test was 0.952 ($p < 0.001$) for the recommendation results of the Pearson correlation similarity measure, which achieved a perfectly strong agreement. The Pearson Chi-square value, *df*, and *p*-value were 0.048, 1, and 0.826, respectively for the recommendation results generated through the Pearson correlation similarity measure. The test results showed there was no significant difference between the results generated by this study and the two evaluators either.

For teenagers related health topics, the "Measure of Agreement" value of the Kappa test was 0.828 ($p < 0.001$) for the recommendation results of the Pearson correlation similarity measure, which also achieved a strong agreement. The Pearson Chi-square value, *df*, and *p*-value were 0.043, 1, and 0.836, respectively for the recommendation results generated through the Pearson correlation similarity measure. The test results showed there was no significant difference between the results generated by this study and the two evaluators again.

Regarding older adults related health topics, the "Measure of Agreement" value of the Kappa test was 0.836 ($p < 0.001$) for the recommendation results of the Pearson correlation similarity measure, which achieved a strong agreement. Also, the Pearson Chi-square value, *df*, and *p*-value were 0.000, 1, and 1.000, respectively for the recommendation results generated

249

through the Pearson correlation similarity measure. The test results showed there was no significant difference between the results generated by this study and the two evaluators.

In conclusion, it suggested that the recommended results in this study were consistent with those from the expert evaluators.

5.2.4   Expert evaluation results toward the recommendation results from the Euclidean distance similarity measure

On the other hand, differences were investigated in terms of the optimization results of the Euclidean distance similarity measure. For mental health related health topics, the "Measure of Agreement" value was only 0.270 ($p < 0.001$) for the recommendation results of the Euclidean distance similarity measure through the Kappa test, which did not achieve a strong agreement. In addition, the Pearson Chi-square value, *df*, and *p*-value were 19.387, 1, and 0.000, respectively for the recommendation results generated through the Euclidean distance similarity measure. The test results showed there was significant difference between the results generated by this study and the two evaluators.

For children related health topics, the "Measure of Agreement" value of the Kappa test was only 0.396 ($p < 0.001$) for the recommendation results of the Euclidean distance similarity measure, which only reached a minimal agreement. The Pearson Chi-square value, *df*, and *p*-value were 16.090, 1, and 0.000, respectively for the recommendation results generated through the Euclidean distance similarity measure. The test results showed there was significant difference between the results generated by this study and the two evaluators.

Regarding teenagers related health topics, the "Measure of Agreement" value of the

Kappa test was only 0.123 ($p < 0.001$) for the recommendation results of the Euclidean distance similarity measure, which only reached a minimal agreement. The Pearson Chi-square value, *df*, and *p*-value were 36.168, 1, and 0.000, respectively for the recommendation results generated through the Euclidean distance similarity measure. The test results showed there was significant difference between the results generated by this study and the two evaluators.

In terms of older adults related health topics, the "Measure of Agreement" value of the Kappa test was only 0.252 ($p < 0.001$) for the recommendation results of the Euclidean distance similarity measure, which only reached a minimal agreement. Moreover, the Pearson Chi-square value, *df*, and *p*-value were 36.699, 1, and 0.000, respectively for the recommendation results generated through the Euclidean distance similarity measure. The test results showed there was significant difference between the results generated by this study and the two evaluators.

In conclusion, it suggested that the recommended results in this study were not consistent with those from the expert evaluators in terms of the suggested connections identified by the Euclidean distance similarity measure.

5.2.5   Summary

According to the findings observed during the optimization process, the Pearson correlation similarity measure had generated groups of recommendations that were very similar to the recommendation results identified through the cosine similarity measure toward the four health topic groups. This consistency between the two similarity measures was also proved by the two health professionals through their evaluation results. Hence, these two similarity measures might be applied together into optimizations toward more subject directories. When

employing one of them for similarity calculation, the other could serve as the supplemental methodology for verification purposes. This could ensure a more convincing and effective optimization result.

On the other side, the Euclidean distance similarity measure might not be an ideal methodology when calculating the semantic relationships of textual information. Both the suggestions analyzed through this study and the evaluation results generated by the health field experts had shed light on some inappropriate suggested connections toward the selected health topics relating to the four health topic subcategories. Compared with the cosine and the Pearson correlation similarity measures, the Euclidean distance similarity measure was less effective in optimizing the health topic subject directory system on the MedlinePlus portal.

5.3 The connection of recommendation results among the four health topic groups

5.3.1 *Strong semantic relationships among several health topics*

During the comparisons conducted between the recommendation results proposed by this study and the evaluation results generated by health field experts, a small group of suggested connections were marked as "not related" by health professionals despite of the fact that the two linked health topics shared an extremely high similarity value in terms of their Web pages. One example would be *Exercise for Seniors* and *Exercise for Children* – these two topics had a similarity value of 0.669, which might be caused by their common sections of text introducing about "exercise," while they were actually targeting different health consumer groups. These health topics were mostly regarded as irrelevant health topics in a hierarchical structural subject directory.

252

However, there could be another circumstance worth noticing when considering the fact that users checking on information from MedlinePlus might be the family member that is responsible to take care of the whole family (e.g. a mother). If that is the case, such structural linkages set for the same daily health topic among various health consumer groups might be of great help.

5.3.2   *The intersections among the four health topic groups*

In order to further explore the relationships among the four selected health topic groups, the original structural link networks as well as the five core health topics identified in the recommendation results were observed and compared toward each health topic subcategory.

For mental health related health topics, the children and teenagers subcategories were involved at the first level through the health topics *Child Mental Health* and *Teen Mental Health*. After the connections to these two bridging topics were built, a few more children and teenagers related health topics were found in the mental health subject directory network. On the contrary, only two health topics relating to the older adults subcategory were included and they were not involved until the last level. Surprisingly enough, in the optimization results, the topic *Nutrition for Seniors* was one of the five core health topics that were underestimated. This might prove the strong relationship between the older adult subcategory and mental health related health topics.

Regarding children related health topics, the only closely linked health topic group was teenagers. The topic *Teen Health* occurred at the first level in the original structural network. However, mental health and older adults related health topics only appeared at the third level. Despite of the fact that *Child Mental Health* was serving as one of the five core health topics in

253

the recommendation results, the five core health topics were all targeting on children and toddlers.

Compared with mental health and children related health topics, the network of teenagers related health topics was found to have stronger relationships with the other three subcategories. The children subcategory was involved at the second level while the mental health and older adults subcategories were included at the third level. It might indicate that the health topics relating to teenagers were playing more important roles in connecting health topics of other subcategories in the whole subject directory system on the MedlinePlus portal. Regarding the five core health topics in the recommendation results, similar to the children's network, optimization suggestions for teenagers related health topics were only focusing on the teenagers group.

Last but not least, older adults related health topics were also found to be connected to the other three health topic groups. To be more specific, the children subcategory was involved at the second level, while the mental health and teenagers related health topics appeared at the third and fourth level, respectively. Interestingly enough, although older adults related health topics were not closely related to the other three health topic subcategories in others' networks, the five core health topics suggested for older adults related health topics included topics relating to all of the other three groups – *Child Nutrition*, *Teen Mental Health*, *Child Mental Health*, and *Mental Health*. These four core health topics listed in the suggested connections for older adults related health topics had indicated two facts: 1) the original structural connections of the older adults subcategory was poorly designed in the whole subject directory system on MedlinePlus; 2) the hidden semantic relationships between health topics relating to older adults and health topics

254

relating to mental health, children, and teenagers were strong.

5.3.3   Key health topics identified during the optimization process

During the process of optimization, a few unique health topics were recognized among the selected four health topic groups. Among these unique health topics, some of them were already playing influential roles in multiple original structural link networks. For instance, *Nutrition* was found to be a key node in the networks of both mental health and older adults related health topics. Similarity, *Eating Disorder* was found to serve as an important bridging topic in the networks of both mental health and older adults related health topics as well.

On the other side, the importance of some unique health topics was not recognized in the original subject directory system on the MedlinePlus portal. Therefore, they were identified through the optimization process in this study. For instance, *Children's Health* was the starting point when collecting health topics relating to the children subcategory. It had a few related health topics listed on its Web page, including the topic *Teen Health*, which had indicated the strong relationship between the subcategories of children and teenagers. However, among all the selected health topics, it was found that *Children's Health* was listed as a related health topic for only one health topic – *Child Nutrition*. In other words, all the connections possessed by *Children's Health* was unidirectional, and once a user had left the page of *Children's Health*, it would be nearly impossible for the user to get back to this page. Hence, adding connections to such unique health topics should improve the effectiveness of users' browsing search.

5.4 The revised semantic-based network

In prior studies, former researchers tended to create a semantic network that was built

purely according to the semantic relationships possessed by the selected health topics (Zhu & Zhang, 2020; Zhang et al., 2016). This semantic network had differences with the original structural network from two aspects: 1) new health topics were added into the semantic network because they had a high similarity in terms of their textual information; 2) several structurally connected health topics in the original structural network were removed in semantic network due to having low similarity values.

Compared with those prior studies, the revised semantic-based network built in this study decided to keep the structurally connected health topics that were included in the original structural network even if they possessed a low similarity value in terms of their textual information. The reason was that these connections were derived from official subject headings such as MeSH and had been manually reviewed by health professionals. In other words, these connections were created by the MedlinePlus portal for different reasons. The health topics connected might be generated from other relationships such as an ontology-based system instead of semantic relationships. Therefore, keeping these original structural connections along with the newly recommended health topics in the revised semantic-based network could better reflect the overall subject directory system of MedlinePlus from a broader view – both semantic relationships and other types of relationships could be included and serve as the foundation for building a theoretically more effective subject directory. This is an innovation toward the prior research studies.

5.5 Implications

The major goal of this health information study is to provide assistance to both health consumers and professionals through improving the navigation system of MedlinePlus. Public

health portals serve as a vital resource for obtaining health information. Therefore, understanding and evaluating the structure, mechanism, and effectiveness and then proposing recommendations for possible optimizations accordingly can enable health consumers to search and obtain health information more effectively and efficiently. Meanwhile, health professionals, including the public health portal creators, may be able to optimize subject directory-based navigation systems. The implications of this study come from the theoretical, practical, and methodological aspects.

5.5.1   *Theoretical implications*

The theoretical implications lie in the discovery of characteristics and patterns among the health topics in the subject directory of MedlinePlus. Previous studies about a subject directory were mostly from the traditional perspectives of knowledge organization and other library and information studies fields. They attempted to assess and improve a subject directory through ideas such as ontology-based development, controlled vocabulary, or comparisons between computer generation and manual review. This study regarded the selected topics in the subject directory as independent individual objects and explored the relationships among these topics from both structural and semantic aspects. Visualized social networks were generated to find clustering patterns among different health topics and optimize structural networks of the four health topic groups. Inferential statistical methods were applied to evaluate various networks after the optimization process.

To be more specific, multiple procedures and similarity measures were employed in this study to ensure the effectiveness of the optimization results. Three procedures were followed to identify the suggested connections in the recommendation results toward the four health topic groups. The original structural link networks of mental health, children, teenagers, and older

257

adults related health topics were firstly examined through the social network analysis and descriptive statistical analysis in terms of similarity value. As a result, recommendations were presented based on the comparison of the semantic relationship between the structurally connected and unconnected health topics. Following that, node-level centrality features were calculated and investigated to further identify core health topics that were playing more important roles in the corresponding networks. Last but not least, the optimization suggestions, including the recognized core health topics, were evaluated by two health professionals. The evaluation outcomes supported the recommendation results proposed in this study. Such combined procedures from multiple perspectives could be further applied to other research studies toward subject directory optimization.

In addition, three similarity measures were utilized in this study for comparing purposes. The major findings about the optimization suggestions were determined by the analysis results derived from the cosine similarity measure. After that, the Pearson correlation similarity measure as well as the Euclidean distance similarity measure were employed to verify the recommendations. As a result, the three similarity measures generated both common and unique suggestions, which indicated that some similarity measures might work more effectively than others in terms of optimizing a health topic related subject directory system. Moreover, some similarity measures might be applied together to explore more convincing suggestions.

An instance that could prove the significance of applying multiple similarity measures was *Veterans and Military Family Health*. This topic was a "dead end" in the original subject directory on MedlinePlus since it had no related health topics. However, the recommendation results generated from all the three similarity measures were suggesting that this topic should be

258

linked back to its parent topic - *Veterans and Military Health*. Such a suggestion should be regarded as a strong recommendation since it was supported by multiple similarity measure analysis results.

5.5.2   *Practical implications*

This study explores and evaluates the connections among the mental health related topics, child-related health topics, teenager-related health topics, and older adult-related health topics, in the subject directory of MedlinePlus. A subject directory has been recognized as a vital and effective mechanism for browsing and navigating a Web portal. Therefore, identifying the connection patterns can reveal the fact about how those health topics are related and how effective the subject directory is on assisting users in obtaining useful health information. The relationships identified and assessed can affect whether users are able to get directed to proper topics while browsing the portal. It can also provide suggestions for the creators of the MedlinePlus portal as well as other portal creators for improving the structure of the subject directory so that users can navigate the system effectively and receive better information searching experiences.

Investigating the inconsistency between the structural and semantic relationships among health topics in the subject directory may assist the creators of the MedlinePlus portal as well as other portal creators in recognizing isolated health topics. Moreover, for those influential topics that are playing vital roles in the network in terms of connecting other topics, if any of these influential topics is not being located at the appropriate position in the network, it can be identified and adjusted as well. Evaluating and comparing each health topic within the network can also provide clues and ideas for portal creators as well as other health professionals to

259

understand the topics investigated, and develop a more effective subject directory accordingly.

The findings of this study contribute to optimizing the subject directory of MedlinePlus for mental health, children, teenagers, and older adults related health information. The results help both health consumers and professionals to understand how the health topics are built and connected within the subject directory, and enable people to identify the inconsistency existing between the structural and semantic relationships among those health topics. The recommended optimizations proposed by this study will help to develop public health portals' navigation systems through building more consistent and effective subject directories. These suggestions will better support users for browsing when seeking health information.

5.5.3   *Methodological implications*

Social network analysis has been widely employed in the field of social science, especially when faced with social media related content. It has also raised attention from scholars in the LIS field. Social network analysis provides a unique way to uncover the facts and characteristics existing in the communications and connections among members within a community. This study applied social network analysis to explore the communication and connection patterns among the health topics that have formed into the subject directory of the MedlinePlus portal, and identified the influential topics, which referred to those health topics playing more important roles than others in connecting different topics, among the topic networks. The application of social network analysis in this study could be utilized to investigate a subject directory on other public health portals.

A mixed research method was employed to compare various networks generated among

health topics from the subject directory system of MedlinePlus. This study applied social network analysis, descriptive statistical methods and inferential statistical methods to identify the clustering patterns and evaluate the relationship characteristics among mental health related health topics, children related health topics, teenagers related health topics, and older adults related health topics. This combined research method could be applied to explore health topics of other health conditions/diseases or health consumer groups.

Chapter 6. Conclusions

This last chapter provides an overview of the significance of the study, and summarizes the research problem as well as specific research questions with their corresponding findings. In addition, this chapter also includes the limitations of this study and several future research directions.

6.1 The implications of the study

This study intended to optimize the subject directory system on the MedlinePlus portal through an investigation toward the consistency between the structural and semantic relationships among the selected health topics relating to the four health topic subcategories. A mixed research method was employed to compare various networks generated among health topics from the subject directory system of MedlinePlus. This study applied social network analysis, descriptive statistical methods and inferential statistical methods to identify the clustering patterns and evaluate the relationship characteristics among mental health related health topics, children related health topics, teenagers related health topics, and older adults related health topics. This combined research method could be applied to explore health topics of other health conditions/diseases or health consumer groups.

Three procedures were followed to identify the suggested connections in the recommendation results toward the four health topic groups. Recommendations were presented based on the comparison of the semantic relationship between the structurally connected and unconnected health topics. Following that, node-level centrality features were calculated and investigated to further identify core health topics that were playing more important roles in the

corresponding networks. Last but not least, the optimization suggestions, including the recognized core health topics, were evaluated by two health professionals. The evaluation outcomes supported the recommendation results proposed in this study. Such combined procedures from multiple perspectives could be further applied to other research studies toward subject directory optimization.

In addition, three similarity measures were utilized in this study for comparing purposes. The major findings about the optimization suggestions were determined by the analysis results derived from the cosine similarity measure. After that, the Pearson correlation similarity measure as well as the Euclidean distance similarity measure were employed to verify the recommendations. As a result, the three similarity measures generated both common and unique suggestions, which indicated that some similarity measures might work more effectively than others in terms of optimizing a health topic related subject directory system. Moreover, some similarity measures might be applied together to explore more convincing suggestions.

The findings of this study contributed to optimizing the subject directory of MedlinePlus for mental health, children, teenagers, and older adults related health information. The results would help both health consumers and professionals to understand how the health topics were built and connected within the subject directory, and enable people to identify the inconsistency existing between the structural and semantic relationships among those health topics. The recommended optimizations proposed by this study would help to develop public health portals' navigation systems through building more consistent and effective subject directories. These suggestions would better support users for browsing when seeking health information.

6.2 Summary of the research problem, questions and major findings

This research study concentrates on the investigation and optimization regarding the topic-based subject directory applied by the representative public health portal MedlinePlus. The primary research problem of this study is to explore, assess, and optimize the connections among the mental health related health topics, children related health topics, teenagers related health topics, and older adults related health topics in the subject directory of MedlinePlus.

The time spent for the data collection stage was 7 months. After collecting around 600 health topics in total for the four health topic groups, both the structural and semantic relationships among the selected health topics were extracted, explored, and compared. Social network analysis, descriptive statistical analysis, as well as inferential analysis methods were applied to analyze the collected data.

RQ1: How are health topics related to mental health, children, teenagers, and older adults connected in the subject directory on the MedlinePlus portal: are the structural and semantic connections of the health topics consistent?

The first research question investigated if the current connections possessed by the selected health topics on the MedlinePlus portal were consistent from both the structural and the semantic perspectives. RQ1.1 examined the fact about the consistency problem between the structural and semantic connections on the MedlinePlus portal. RQ1.2 revealed the node-level features of the selected health topics among the four health topic groups. As a result, the current structural link networks relating to mental health, children, teenagers, and older adults applied on the MedlinePlus portal were found to be not consistent with their semantic relationships among

264

the involved health topics. Hundreds of pairs of health topics possessing similar Web page content were not able to be navigated to each other by the portal users. Furthermore, among these missed connections, a few health topics were found to have great impact within the whole network in terms of linking other topics through semantic connections. The control and responsibilities taken by these hidden core health topics were underestimated.

RQ2: Are there significant differences between the original mental health subject directory and the optimized mental health subject directory in terms of its link structure on the MedlinePlus portal?

The second research question focused on comparing the original structural network and the optimized structural network of mental health related health topics in the subject directory on MedlinePlus. The introduction section examined the optimized structural network in terms of the structural connections, semantic connections, and key nodes in the network, while RQ2.1 and RQ2.2 investigated the differences between the original and optimized structural networks regarding their connections' similarity values as well as node-level centrality measures.

It was found that for the mental health subcategory, the five core health topics identified were: *Obesity in Children*, *Child Nutrition*, *Diets*, *Child Development*, and *Nutrition for Seniors*. The recommended new structural connections were mostly added to the children and teenagers related health topics, and besides the health consumer groups, connections were also found in daily health element related topics. The specific mental diseases related health topics were well developed. Moreover, the average similarity value of the structural connections in the optimized network was improved after new connections were added. This difference was proved to be significant through the Mann-Whitney test. Interestingly enough, core health topics were

265

identified to have huge impact on the whole network. It indicated that through the creation of key nodes' connections, the node-level centrality ranking lists in the adjusted structural network were moving closer to the semantic-based network. In other words, key influential nodes were more capable of reflecting the semantic relationships in the network compared with normal nodes. In addition, the node centrality measures were tested through a customized T-test, the results showed that the degree centrality and closeness centrality measures were significantly increased in the optimized structural network while the betweenness centrality measure was not significantly increased in the optimized structural network.

RQ3: Are there significant differences between the original children subject directory and the optimized children subject directory in terms of its link structure on the MedlinePlus portal?

The third research question was concentrating on comparing the original structural network and the optimized structural network of children related health topics in the subject directory on MedlinePlus. The introduction section explored the optimized structural network in terms of the structural connections, semantic connections, and key nodes in the network, while RQ3.1 and RQ3.2 investigated the differences between the original and optimized structural networks regarding their connections' similarity values as well as node-level centrality measures.

It was found that for the children subcategory, the five core health topics identified were: *Medicines and Children*, *Children's Health*, *Toddler Health*, *Child Development*, and *Child Mental Health*. The recommended new structural connections were mostly added to the health consumer related health topics. These consumer groups included children, teenagers, older adults, infants, toddlers, etc. Besides the health consumer groups, connections were also found in pregnancy, hepatitis, and daily exercises related topics. Moreover, the average similarity value of

the structural connections in the optimized network was improved after new connections were added. This difference was proved to be significant through the Mann-Whitney test. Key health topics were identified to have huge impact on the whole network again in the children topic subcategory. It echoed the conclusion that through the creation of key nodes' connections, the node-level centrality ranking lists in the adjusted structural network were moving closer to the semantic-based network. In other words, key influential nodes were more capable of reflecting the semantic relationships in the network compared with normal nodes. In addition, the node centrality measures were tested through the customized T-test, the results showed that the degree centrality and closeness centrality measures were significantly increased in the optimized structural network while the betweenness centrality measure was not significantly different from the original structural link network. This conclusion also fit into the earlier results generated in the network of the mental health topic subcategory.

RQ4: Are there significant differences between the original teenagers subject directory and the optimized teenagers subject directory in terms of its link structure on the MedlinePlus portal?

The fourth research question was concentrating on comparing the original structural network and the optimized structural network of teenagers related health topics in the subject directory on MedlinePlus. The introduction section examined the optimized structural network in terms of the structural connections, semantic connections, and key nodes in the network, while RQ4.1 and 4.2 explored the differences between the original and optimized structural networks regarding their connections' similarity values as well as node-level centrality measures.

It was found that for the teenagers subcategory, the five core health topics identified were: *Teenage Pregnancy*, *Infections and Pregnancy*, *Pregnancy and Drug Use*, *Pregnancy and*

Medicines, and *Health Problems in Pregnancy*. The recommended new structural connections were mostly added to the pregnancy related health topics. These recommendations were echoed by the five core topics identified through the process of identifying key nodes in the revised semantic-based network –the five core topics recognized were all about pregnancy related issues. This finding had proved the great impact possessed by the pregnancy related health topics in the network of the teenagers subcategory. Besides the pregnancy related health topics, connections were also found among health topics relating to specific mental diseases. The close relationship between mental health and teenagers was verified. Moreover, the average similarity value of the structural connections in the optimized network was improved after new connections were added. This difference was found to be significant through the Mann-Whitney test. The 5 pregnancy related key health topics were identified to have huge impact in the whole network again in the teenagers topic subcategory. It proved the conclusion that through the creation of key nodes' connections, the node-level centrality ranking lists in the adjusted structural network were moving closer to the semantic-based network. In other words, key influential nodes were more capable of reflecting the semantic relationships in the network compared with normal nodes. In addition, the node centrality measures were tested through the customized T-test, the results showed that the degree and closeness centrality measures were significantly increased in the optimized structural network while the betweenness centrality measure was not significantly different from the original structural link network. This conclusion also fit into the earlier results generated in the network of the mental health topic subcategory and children topic subcategory.

RQ5: Are there significant differences between the original older adults subject directory and the optimized older adults subject directory in terms of its link structure on the MedlinePlus

portal?

The fifth research question was concentrating on comparing the original structural network and the optimized structural network of older adults related health topics in the subject directory on MedlinePlus. The introduction section examined the optimized structural network in terms of the structural connections, semantic connections, and key nodes in the network, while RQ5.1 and 5.2 investigated the differences between the original and optimized structural networks regarding their connections' similarity values as well as node-level centrality measures.

It was found that for the older adults subcategory, the five core health topics identified were: *Child Nutrition*, *Healthy Aging*, *Teen Mental Health*, *Child Mental Health*, and *Mental Health*. The recommended new structural connections were mostly added to the mental health related topics, the three health consumer groups related topics (children, teenagers, and older adults), and two specific daily health elements related topics (vitamins and cholesterol). These recommendations were partially echoed by the five core topics identified through the process of identifying key nodes in the revised semantic-based network – the five core topics recognized were covering health topics relating to children, teenagers, and mental health. This finding had also proved the strong relationships existing among mental health, children, teenagers, and older adults. Moreover, the average similarity value of the structural connections in the optimized network was improved after new connections were added. This difference was found to be significant through the Mann-Whitney test. The five core health topics were identified to have huge impact in the whole network again in the older adults topic subcategory. It proved the conclusion that through the creation of key nodes' connections, the node-level centrality ranking lists in the adjusted structural network were moving closer to the semantic-based network. In

269

other words, key influential nodes were more capable of reflecting the semantic relationships in the network compared with normal nodes. In addition, the node centrality measures were tested through the customized T-test, the results showed that the degree and closeness centrality measures were significantly increased in the optimized structural network while the betweenness centrality measure was not significantly different from the original structural link network. This conclusion also fit into the earlier results generated in the network of the mental health topic subcategory, the children topic subcategory, and the teenagers topic subcategory.

RQ6: Are there any significant differences between the optimized structural network from this study and the evaluation results from the health field experts?

After optimization results were generated in this study, two evaluators were invited to assess the optimization results. The sixth research question verified the consistency between the recommendation results from this study and the evaluation results from the health professionals.

As a result, the optimization recommendations proposed by this study were evaluated by two health field experts. Their evaluation outcomes confirmed that the suggested connections generated by this study fit into professional assessments. In other words, the recommendation results were supported by both the semantic relationships and professional judgements.

6.3 Limitations of the study

One limitation of this study was not recommending the removal of the currently linked edges with low semantic relationships. Even though, based on the semantic relationships identified and calculated, new relevant edges were recommended to be added to the original structural link network, those edges from similarities of the health topic edges on the original

structural link network that were very low were not recommended to be removed.

Another limitation of this study was that the optimization process conducted in this study had only focused on the system's side. All the recommendations were identified and presented according to the structural link networks built in the subject directory system on the MedlinePlus portal. No investigation was performed from the users' perspective.

The third limitation of this study was that the optimization results were presented based on only part of the subject directory system on the MedlinePlus portal. To be more specific, suggestions were only made in terms of health topics among mental health, children, teenagers, and older adults related health topics. Recommendations regarding topics from other subcategories were not included.

The last limitation of this study was that the methodologies applied for optimization in this study were limited. Only the similarity values and the three node-level centrality features were employed for proposing suggested connections. Other features of social network analysis such as clustering coefficient were not involved. In addition, when calculating the similarity values of the selected health topics, only three similarity measures were utilized. The calculation results from other similarity measures remained unknown.

6.4 Future research directions

The future research directions include, but are not limited to, applying the social network analysis method to a wider range of subject directory systems in health portals or health information systems. More health topic subcategories such as body/organ diseases, traditional disorders and conditions, daily health elements, as well as other health consumers groups can be

271

included in future studies. Meanwhile, more elements of the social network analysis methodology, especially from the network level, such as clustering coefficient can be used in similar studies. More similarity measures can also be employed to set up various thresholds in terms of exploring new structural connections. As for the evaluation performed by the health field experts, additional qualitative methods can be applied for quantitative analysis. For instance, interviews can be conducted toward the experts regarding their evaluation outcomes so that more in-depth interpretations about their decisions can be received to better justify their evaluation results.

Moreover, besides evaluating and optimizing subject directories from the system's side, users could also be involved. Different from inviting health field experts for evaluation in this study, users could also serve as the evaluators toward the optimization suggestions. Furthermore, users' perspectives could be taken into considerations through usability tests so that different recommendations could be identified according to users' feedbacks and thoughts through their interactions toward the system. Last but not least, more methodologies might be applied to discover potential improvements regarding health topic related subject directories.

References

Abraham, A., Hassanien, A.-E., & Snásel, V. (2010). *Computational Social Network Analysis: Trends, Tools and Research Advances*. Retrieved from https://books.google.com/books?hl=en&lr=&id=-S1KiURSfRAC&oi=fnd&pg=PR5&dq=Computational+Social+Network+Analysis+A+Abraham,+AE+Hassanien,+V+Snasel&ots=nJgJHTb1cI&sig=se3-tcDwW8eDZX6U_jHA_qeOgi8

Adkisson, H. P. (2002). Identifying de-facto standards for e-commerce web sites. *IEEE International Professional Communication Conference*, 22–45. https://doi.org/10.1109/ipcc.2002.1049085

Askelson, N. M., Campo, S., & Carter, K. D. (2011). Completely isolated? health information seeking among social isolates. *Health Education and Behavior*, *38*(2), 116–122. https://doi.org/10.1177/1090198110369055

Ayers, J. W., Althouse, B. M., Allem, J.-P., Rosenquist, J. N., & Ford, D. E. (2013). Seasonality in Seeking Mental Health Information on Google. *American Journal of Preventive Medicine*, *44*(5), 520–525. https://doi.org/10.1016/j.amepre.2013.01.012

Baeza-Yates, R., & Ribeiro-Neto, B. (1999). Modern Information Retrieval. In *Information Retrieval*. https://doi.org/10.1080/14735789709366603

Baird, A., Furukawa, M. F., & Raghu, T. S. (2012). Understanding Contingencies Associated with the Early Adoption of Customer-Facing Web Portals. *Journal of Management Information Systems*, *29*(2), 293–324. https://doi.org/10.2753/MIS0742-1222290210

Balka, E., Krueger, G., Holmes, B. J., & Stephen, J. E. (2010). Situating Internet Use:

273

Information-Seeking Among Young Women with Breast Cancer. *Journal of Computer-Mediated Communication*, *15*(3), 389–411. https://doi.org/10.1111/j.1083-6101.2010.01506.x

Ballesteros, L., & Croft, W. B. (1998). Resolving ambiguity for cross-language retrieval. *SIGIR Forum (ACM Special Interest Group on Information Retrieval)*, 64–71. https://doi.org/10.1145/290941.290958

Barash, V., & Golder, S. (2011). Twitter: Conversation, Entertainment, and Information, All in One Network! In *Analyzing Social Media Networks with NodeXL: Insights from a Connected World* (pp. 143–164). https://doi.org/10.1016/B978-0-12-382229-1.00010-2

Barker, A. (2019). Navigating Life: A Taxonomy of Wayfinding Behaviours. *Journal of Navigation*, *72*(3), 539–554. https://doi.org/10.1017/S0373463319000043

Barnes, J. A. (1954). Class and Committees in a Norwegian Island Parish. *Human Relations*, *7*(1), 39–58. https://doi.org/10.1177/001872675400700102

Basch, E., Artz, D., Dulko, D., Scher, K., Sabbatini, P., Hensley, M., … Schrag, D. (2005). Patient Online Self-Reporting of Toxicity Symptoms During Chemotherapy. *Journal of Clinical Oncology*, *23*(15), 3552–3561. https://doi.org/10.1200/JCO.2005.04.275

Bass, S. B., Ruzek, S. B., Gordon, T. F., Fleisher, L., McKeown-Conn, N., & Moore, D. (2006). Relationship of Internet Health Information Use With Patient Behavior and Self-Efficacy: Experiences of Newly Diagnosed Cancer Patients Who Contact the National Cancer Institute's Cancer Information Service. *Journal of Health Communication*, *11*(2), 219–236. https://doi.org/10.1080/10810730500526794

Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search

274

interface. *Online Review*, *13*(5), 407–424. https://doi.org/10.1108/eb024320

Beier, J., & Tesche, T. (2001). Navigation and interaction in medical knowledge spaces using topic maps. *International Congress Series*, *1230*(C), 384–388. https://doi.org/10.1016/S0531-5131(01)00087-5

Benigeri, M. (2003). Shortcomings of health information on the Internet. *Health Promotion International*, *18*(4), 381–386. https://doi.org/10.1093/heapro/dag409

Benigeri, M., & Pluye, P. (2003). Shortcomings of health information on the Internet. *Health Promotion International*, *18*(4), 381–386. https://doi.org/10.1093/heapro/dag409

Berkman, N. D., DeWalt, D. A., Pignone, M., Sheridan, S. L., Lohr, K. N., Lux, L., … Bonito, A. J. (2004). *Literacy and Health Outcomes. Evidence Report/Technology Assessment No. 87 (Prepared by RTI International–University of North Carolina Evidence-based Practice Center under Contract No. 290-02-0016)*. Rockville, MD: Agency for Healthcare Research and Quality.

Boot, C. R. L., & Meijman, F. J. (2010). The public and the Internet: Multifaceted drives for seeking health information. *Health Informatics Journal*, *16*(2), 145–156. https://doi.org/10.1177/1460458210364786

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2013). *Analyzing Social Networks*. Los Angeles: SAGE Publications Ltd.

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network Analysis in the Social Sciences. *Science*, *323*(5916), 892–895. https://doi.org/10.1126/science.1165821

Bouche, G., & Migeot, V. (2008). Parental use of the Internet to seek health information and primary care utilisation for their child: A cross-sectional study. *BMC Public Health*, *8*(1),

275

300. https://doi.org/10.1186/1471-2458-8-300

Brashers, D. E. (2001). Communication and Uncertainty Management. *Journal of Communication*, *51*(3), 477–497. https://doi.org/10.1111/j.1460-2466.2001.tb02892.x

Brashers, D. E., Goldsmith, D. J., & Hsieh, E. (2002). Information Seeking and Avoiding in Health Contexts. *Human Communication Research*, *28*(2), 258–271. https://doi.org/10.1111/j.1468-2958.2002.tb00807.x

Brodeur, P. (2005). The Turning Point Initiative. In S. L. Issacs & J. R. Knickman (Eds.), *To Improve Health and Health Care, Volume VIII The Robert Wood Johnson Foundation Anthology* (pp. 1–16). Retrieved from https://www.rwjf.org/en/library/research/2005/01/to-improve-health-and-health-care-volume-viii.html

Bryman, A. (2012). *Social Research Methods*. https://doi.org/10.4324/9780429463013-34

Burns, J. M., Davenport, T. A., Durkin, L. A., Luscombe, G. M., & Hickie, I. B. (2010). The internet as a setting for mental health service utilisation by young people. *Medical Journal of Australia*, *192*, S22–S26. https://doi.org/10.5694/J.1326-5377.2010.TB03688.X

Chambers, D., Wilson, P., Thompson, C., & Harden, M. (2012). Social network analysis in healthcare settings: A systematic scoping review. *PLoS ONE*, *7*(8), e41911. https://doi.org/10.1371/journal.pone.0041911

Choo, C. W., Detlor, B., & Turnbull, D. (2000). Information seeking on the Web: An integrated model of browsing and searching. *First Monday*, *5*(2), 3–16. https://doi.org/10.5210/fm.v5i2.729

Chung, Y. M., & Noh, Y. H. (2003). Developing a specialized directory system by automatically classifying Web documents. *Journal of Information Science*.

https://doi.org/10.1177/016555150302900204

Cimino, C., Octo Barnett, G., Hassan, L., Blewett, D. R., & Piggins, J. L. (1991). Interactive

query workstation: Standardizing access to computer-based medical resources. *Computer*

Methods and Programs in Biomedicine, *35*(4), 293–299. https://doi.org/10.1016/0169-

2607(91)90007-G

Clear & Simple | National Institutes of Health (NIH). (n.d.). Retrieved November 25, 2019, from

https://www.nih.gov/institutes-nih/nih-office-director/office-communications-public-

liaison/clear-communication/clear-simple

Cline, R. J. W., & Haynes, K. M. (2001). Consumer health information seeking on the Internet:

the state of the art. *Health Education Research*, *16*(6), 671–692.

https://doi.org/10.1093/her/16.6.671

Coleman, J. S. (1986). Social Theory, Social Research, and a Theory of Action. *American*

Journal of Sociology, *91*(6), 1309–1335. https://doi.org/10.1086/228423

Colombo, C., Mosconi, P., Confalonieri, P., Baroni, I., Traversa, S., Hill, S. J., … Filippini, G.

(2014). Web search behavior and information needs of people with multiple sclerosis:

Focus group study and analysis of online postings. *Journal of Medical Internet Research*,

16(7). https://doi.org/10.2196/ijmr.3034

Committee on Health Literacy, I. of M. (US). (2004). *Health Literacy: A Prescription to End*

Confusion (L. Nielsen-Bohlman, A. Panzer, & D. Kindig, Eds.). Retrieved from

https://www.ncbi.nlm.nih.gov/books/NBK216035/

Conner, K. O., Copeland, V. C., Grote, N. K., Koeske, G., Rosen, D., Reynolds, C. F., & Brown,

C. (2010). Mental Health Treatment Seeking Among Older Adults With Depression: The

277

Impact of Stigma and Race. *The American Journal of Geriatric Psychiatry*, *18*(6), 531–543. https://doi.org/10.1097/JGP.0b013e3181cc0366

Cosine Similarity - an overview | ScienceDirect Topics. (n.d.). Retrieved November 24, 2019, from https://www.sciencedirect.com/topics/computer-science/cosine-similarity

Cyber Dialogue. (1998). *Online health reaches critical mass. Nua Internet Surveys*.

D'Augelli, A. R., Grossman, A. H., Hershberger, S. L., & O'Connell, T. S. (2001). Aspects of mental health among older lesbian, gay, and bisexual adults. *Aging & Mental Health*, *5*(2), 149–158. https://doi.org/10.1080/13607860120038366

De Choudhury, M., Morris, M. R., & White, R. W. (2014). Seeking and sharing health information online: Comparing search engines and social media. *Conference on Human Factors in Computing Systems - Proceedings*, 1365–1375. https://doi.org/10.1145/2556288.2557214

Demelo, J., Parsons, P. & Sedig, K. (2017). Ontology-Driven Search and Triage: Design of a Web-Based Visual Interface for MEDLINE. *JMIR Medical Informatics*, *5*(1), e4. https://doi.org/10.2196/medinform.6918

Dutta-Bergman, M. J. (2013). Media use theory and internet use for health care. In M. Murero & R. E. Rice (Eds.), *The Internet and Health Care: Theory, Research, and Practice* (pp. 83–104). https://doi.org/10.4324/9780203810675

Eichner, J., & Dullabh, P. (2007). Accessible Health Information Technology (IT) for Populations with Limited Literacy: A Guide for Developers and Purchasers of Health IT. *National Opinion Research Center for the National Resource Center for Health IT*, (08). Retrieved from www.ahrq.gov

Ellis-Danquah, L. V. (2004). Addressing health disparities: African American consumer information resources on the web. *Medical Reference Services Quarterly*, *23*(4), 61–73. https://doi.org/10.1300/J115v23n04_06

Evers, K. E. (2006). eHealth Promotion: The Use of the Internet for Health Promotion. *American Journal of Health Promotion*, *20*(4), 1–14. https://doi.org/10.4278/0890-1171-20.4.1

Eysenbach, G. (2003). The Impact of the Internet on Cancer Outcomes. *CA: A Cancer Journal for Clinicians*, *53*(6), 356–371. https://doi.org/10.3322/canjclin.53.6.356

Fergie, G., Hilton, S., & Hunt, K. (2016). Young adults' experiences of seeking online information about diabetes and mental health in the age of social media. *Health Expectations*, *19*(6), 1324–1335. https://doi.org/10.1111/hex.12430

Fogel, J. (2003). Internet use for Cancer Information among Racial/Ethnic Populations and Low Literacy Groups. *Cancer Control*, *10*(5_suppl), 45–51. https://doi.org/10.1177/107327480301005s07

Fox, S., & Duggan, M. (2013). Health online. In *Pew Internet & American Life Project*. https://doi.org/10.1007/978-1-4419-1005-9_1474

Fox, Susannah, & Fallows, D. (2003). Internet Health Resources. *Pew Internet & American Life Project*, 42. Retrieved from http://www.pewinternet.org

Fox, Susannah, & Rainie, L. (2000). *The online health care revolution: How the Web helps Americans take better care of themselves. Pew Internet & American Life Project: Online Report*. Retrieved from http://www.medicalmarketingmatters.com/wp-content/uploads/PIP_Health_Report.pdf.pdf

Fredriksen-Goldsen, K. I., Cook-Daniels, L., Kim, H.-J., Erosheva, E. A., Emlet, C. A., Hoy-

279

Ellis, C. P., … Muraco, A. (2014). Physical and Mental Health of Transgender Older Adults: An At-Risk and Underserved Population. *The Gerontologist*, *54*(3), 488–500. https://doi.org/10.1093/geront/gnt021

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239. https://doi.org/10.1016/0378-8733(78)90021-7

Friedman, D. B., & Hoffman-Goetz, L. (2006). A systematic review of readability and comprehension instruments used for print and web-based cancer information. *Health Education & Behavior*, *33*(3), 352–373.

Friedman, D. B., Hoffman-Goetz, L., & Arocha, J. F. (2006). Health literacy and the World Wide Web: comparing the readability of leading incident cancers on the Internet. *Medical Informatics and the Internet in Medicine*, *31*(1), 67–87.

Gao, J., Nie, J.-Y., Zhang, J., Xun, E., Su, Y., Zhou, M., & Huang, C. (2001). TREC-9 CLIR Experiments at MSRCN. *The Nineth Text REtrieval Conference (TREC-9)*.

García-Lacalle, J., Pina, V., & Royo, S. (2011). The unpromising quality and evolution of Spanish public hospital web sites. *Online Information Review*, *35*(1), 86–112. https://doi.org/10.1108/14684521111113605

Gilbert, K. (2009). Internet Reviews: Kentucky Public Health Web Resources. *Kentucky Libraries*, *73*(4), 24–27. Retrieved from http://web.b.ebscohost.com/ehost/detail/detail?sid=08bf359c-cd2f-4cc6-9388-e71b1a766d14%40sessionmgr101&vid=0&hid=107&bdata=JkF1dGhUeXBlPWlwLHVpZCZzaXRlPWVob3N0LWxpdmUmc2NvcGU9c2l0ZQ%3D%3D#AN=502988423&db=llf

Gowen, L. K. (2013). Online Mental Health Information Seeking in Young Adults with Mental

280

Health Challenges. *Journal of Technology in Human Services*, *31*(2), 97–111.

https://doi.org/10.1080/15228835.2013.765533

Gravetter, F., & Forzano, L. (2011). *Research methods for the behavioral sciences*.

https://doi.org/10.5860/choice.197263

Gray, C. (2000). INTERNET RESOURCES: Medical resources information for the consumer.

College & Research Libraries News, *61*(11), 995–1002.

https://doi.org/10.5860/crln.61.11.995

Gray, C. (2006). Health and medical resources: information for the consumer. *Journal of Library

Administration*, *44*(1/2), 395–428. https://doi.org/10.1300/J111v44n01_05

Gray, N. J., Klein, J. D., Noyce, P. R., Sesselberg, T. S., & Cantrill, J. A. (2005). Health

information-seeking behaviour in adolescence: The place of the internet. *Social Science and

Medicine*, *60*(7), 1467–1478. https://doi.org/10.1016/j.socscimed.2004.08.010

Griffin, R. J., Dunwoody, S., & Neuwirth, K. (1999). Proposed model of the relationship of risk

information seeking and processing to the development of preventive behaviors.

Environmental Research, *80*(2 II). https://doi.org/10.1006/enrs.1998.3940

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge

Acquisition*, *5*(2), 199–220. https://doi.org/10.1006/knac.1993.1008

Hajdukiewicz, J. R., Vicente, K. J., Doyle, D. J., Milgram, P., & Burns, C. M. (2001). Modeling

a medical environment: An ontology for integrated medical informatics design.

International Journal of Medical Informatics, *62*(1), 79–99. https://doi.org/10.1016/S1386-

5056(01)00128-9

Halpert, A., Dalton, C. B., Palsson, O., Morris, C., Hu, Y., Bangdiwala, S., … Drossman, D.

281

(2007). What Patients Know About Irritable Bowel Syndrome (IBS) and What They Would Like to Know. National Survey on Patient Educational Needs in IBS and Development and Validation of the Patient Educational Needs Questionnaire (PEQ). *The American Journal of Gastroenterology*, *102*(9), 1972–1982. https://doi.org/10.1111/j.1572-0241.2007.01254.x

Hambrick, M. E. (2012). Six Degrees of Information: Using Social Network Analysis to Explore the Spread of Information Within Sport Social Networks. *International Journal of Sport Communication*, *5*(1), 16–34. https://doi.org/10.1123/ijsc.5.1.16

Hansen, D. L., Shneiderman, B., & Smith, M. A. (2011). Analyzing Social Media Networks With NodeXL. In *Analyzing Social Media Networks with NodeXL*. https://doi.org/10.1016/C2009-0-64028-9

Health Information. (n.d.). Retrieved from https://www.nih.gov/health-information

Health Information National Trends Survey Data. (n.d.). Retrieved from National Cancer Institute website: https://hints.cancer.gov/data/default.aspx

Healthy people 2010. (2000).

Healthy people 2020. (2010).

Heidenreich, F. L., & Auflick, P. A. (2007). AZHealthInfo: a database that complements Arizona's new "Go Local" directory of services and enhances online access to local health information. *Journal of Consumer Health on the Internet*, *11*(1), 1–13. https://doi.org/10.1300/J381v11n01_01

Helft, P. R. (2008, March). A new age for cancer information seeking: Are we better off now? *Journal of General Internal Medicine*, Vol. 23, pp. 350–352. https://doi.org/10.1007/s11606-007-0496-6

Hewitt, M., Greenfield, S., & Stovall, E. (2006). From Cancer Patient to Cancer Survivor. In *From Cancer Patient to Cancer Survivor*. https://doi.org/10.17226/11468

Hogan, T. P., & Brashers, D. E. (2009). The theory of communication and uncertainty management: Implications from the wider realm of information behavior. In *Uncertainty, information management, and disclosure decisions: Theories and applications.* (pp. 61–82). https://doi.org/10.4324/9780203933046-11

Hoppe, B., & Reinelt, C. (2010). Social network analysis and the evaluation of leadership networks. *Leadership Quarterly*, *21*(4), 600–619. https://doi.org/10.1016/j.leaqua.2010.06.004

Horgan, Á., & Sweeney, J. (2010). Young students' use of the Internet for mental health information and support. *Journal of Psychiatric and Mental Health Nursing*, *17*(2), 117–123. https://doi.org/10.1111/j.1365-2850.2009.01497.x

Horrigan, J. B., Rainie, L., & Cornfield, M. (2003). *Part Three: The Portals | Pew Research Center*. Retrieved from https://www.pewresearch.org/internet/2003/03/20/part-three-the-portals/

Houts, P. S., Doak, C. C., Doak, L. G., & Loscalzo, M. J. (2006). The role of pictures in improving health communication: A review of research on attention, comprehension, recall, and adherence. *Patient Education and Counseling*, Vol. 61, pp. 173–190. https://doi.org/10.1016/j.pec.2005.05.004

Hsee, C. K. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes*, *67*(3), 247–257. https://doi.org/10.1006/obhd.1996.0077

Hsee, C. K. (1998). Less is better: when low-value options are valued more highly than high-value options. *Journal of Behavioral Decision Making*, *11*(2), 107–121. https://doi.org/10.1002/(sici)1099-0771(199806)11:2<107::aid-bdm292>3.3.co;2-p

Hu, Y., & Sundar, S. S. (2010). Effects of online health sources on credibility and behavioral intentions. *Communication Research*, *37*(1), 105–132. https://doi.org/10.1177/0093650209351512

Impicciatore, P., Pandolfini, C., Casella, N., & Bonati, M. (1997). Reliability of health information for the public on the world wide web: systematic survey of advice on managing fever in children at home. *BMJ*, *314*(7098), 1875–1875. https://doi.org/10.1136/bmj.314.7098.1875

Ivanitskaya, L., O'Boyle, I., & Casey, A. M. (2006). Health Information Literacy and Competencies of Information Age Students: Results From the Interactive Online Research Readiness Self-Assessment (RRSA). *Journal of Medical Internet Research*, *8*(2), e6. https://doi.org/10.2196/jmir.8.2.e6

Izenberg, N., & Lieberman, D. A. (1998). The Web, Communication Trends, and Children's Health Part 4: How Children Use the Web. *Clinical Pediatrics*, *37*(6), 335–340. https://doi.org/10.1177/000992289803700601

Jibaja-Weiss, M. L., & Volk, R. J. (2007). Utilizing Computerized Entertainment Education in the Development of Decision Aids for Lower Literate and Naïve Computer Users. *Journal of Health Communication*, *12*(7), 681–697. https://doi.org/10.1080/10810730701624356

Johnson, J. D., & Case, D. O. (2012). *Health information seeking*. New York: New York : Peter Lang.

Johnson, J. D., & Meischke, H. (1993). A Comprehensive Model of Cancer-Related Information

  Seeking Applied to Magazines. *Human Communication Research*, *19*(3), 343–367.

  https://doi.org/10.1111/j.1468-2958.1993.tb00305.x

Johnston, K. (2006). What are subject directories? Retrieved from Board of Trustees of the

  University of South Carolina website:

  https://www.tcd.ie/Education/ICT/unit02/demonstration03.htm

Jones, L. M., Veinot, T. C. E., Pressler, S. J., Seng, J. S., Mccall, A. M., Fernandez, D., &

  Coleman-Burns, P. W. (2014). Internet Health Information Seeking (IHIS):An Integrative

  Review of the Literature. *Western Journal of Nursing Research*, *36*(10), 1376–1377.

  https://doi.org/10.1177/0193945914540100

Kandula, N. R., Nsiah-Kumi, P. A., Makoul, G., Sager, J., Zei, C. P., Glass, S., … Baker, D. W.

  (2009). The relationship between health literacy and knowledge improvement after a

  multimedia type 2 diabetes education program. *Patient Education and Counseling*, *75*(3),

  321–327. https://doi.org/10.1016/j.pec.2009.04.001

Karras, E., & Rintamaki, L. S. (2012). An Examination of Online Health Information Seeking by

  Deaf People. *Health Communication*, *27*(2), 194–204.

  https://doi.org/10.1080/10410236.2011.575539

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005).

  Lifetime Prevalence and Age-of-Onset Distributions of DSM-IV Disorders in the National

  Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*(6), 593.

  https://doi.org/10.1001/archpsyc.62.6.593

Khoo, K., Bolt, P., Babl, F. E., Jury, S., & Goldman, R. D. (2008). Health information seeking

285

by parents in the Internet age. *Journal of Paediatrics and Child Health*, *44*(7–8), 419–423. https://doi.org/10.1111/j.1440-1754.2008.01322.x

Kim, E., Hou, J., Han, J. Y., & Himelboim, I. (2016). Predicting Retweeting Behavior on Breast Cancer Social Networks: Network and Content Characteristics. *Journal of Health Communication*, *21*(4), 479–486. https://doi.org/10.1080/10810730.2015.1103326

Kim, S. C., Shah, D. V., Namkoong, K., Mctavish, F. M., & Gustafson, D. H. (2013). Predictors of Online Health Information Seeking Among Women with Breast Cancer: The Role of Social Support Perception and Emotional Well-Being. *Journal of Computer-Mediated Communication*, *18*(2), 98–118. https://doi.org/10.1111/jcc4.12002

Kirschning, S., & Von Kardorff, E. (2008). The use of the Internet by women with breast cancer and men with prostate cancer-results of online research. *Journal of Public Health*, *16*(2), 133–143. https://doi.org/10.1007/s10389-007-0134-0

Knitzer, J., & Olson, L. (1982). Unclaimed children: The failure of public responsibility to children and adolescents in need of mental health services. In *Children's Defense Fund*. Retrieved from https://eric.ed.gov/?id=ED223346

Kools, M., Van De Wiel, M. W. J., Ruiter, R. A. C., Crüts, A., & Kok, G. (2006). The effect of graphic organizers on subjective and objective comprehension of a health education text. *Health Education and Behavior*, *33*(6), 760–772. https://doi.org/10.1177/1090198106288950

Korfhage, R. R. (1997). Information storage and retrieval / Robert R. Korfhage. In *Information storage and retrieval*. New York: Wiley Computer Pub.

Krug, S. (2006). *Don't make me think!: a common sense approach to Web usability*. Pearson

Education India.

Kushalnagar, P., Naturale, J., Paludneviciene, R., Smith, S. R., Werfel, E., Doolittle, R., …
DeCaro, J. (2015). Health Websites: Accessibility and Usability for American Sign
Language Users. *Health Communication*, *30*(8), 830–837.
https://doi.org/10.1080/10410236.2013.853226

Kuśnierczyk, W. (2008). Taxonomy-based partitioning of the Gene Ontology. *Journal of
Biomedical Informatics*, *41*(2), 282–292. https://doi.org/10.1016/j.jbi.2007.07.007

Kutner, M., Greenberg, E., Jin, Y., & Paulsen, C. (2006). The health literacy of America's
adults: results from the 2003 National Assessment of Adult Literacy. *Education*, *6*, 1–59.
https://doi.org/10.1592/phco.22.5.282.33191

Laurent, M. R., Cremers, S., Verhoef, G., & Dierickx, D. (2012). Internet use for health
information among haematology outpatients: A cross-sectional survey*. *Informatics for
Health and Social Care*, *37*(2), 62–73. https://doi.org/10.3109/17538157.2011.606481

Learn about MedlinePlus. (n.d.). Retrieved from
https://medlineplus.gov/about/general/aboutmedlineplus/

Lewis, D., Chang, B. L., & Friedman, C. P. (2005). Consumer Health Informatics. In D. Lewis,
G. Eysenbach, R. Kukafka, P. Z. Stavri, & H. B. Jimison (Eds.), *Consumer Health
Informatics* (pp. 1–7). https://doi.org/10.1007/0-387-27652-1_1

Liechty, J. M. (2011, May 1). Health literacy: Critical opportunities for social work leadership in
health care and research. *Health and Social Work*, Vol. 36, pp. 99–107.
https://doi.org/10.1093/hsw/36.2.99

Lu, W. H., Lin, R. S., Chan, Y. C., & Chen, K. H. (2008). Using Web resources to construct

multilingual medical thesaurus for cross-language medical information retrieval. *Decision Support Systems*, *45*(3), 585–595. https://doi.org/10.1016/j.dss.2007.07.004

Luke, D. A., & Harris, J. K. (2007). Network Analysis in Public Health: History, Methods, and Applications. *Annual Review of Public Health*, *28*(1), 69–93. https://doi.org/10.1146/annurev.publhealth.28.021406.144132

Maxwell, J. A. (2012). *Qualitative Research Design: An Interactive Approach: An Interactive Approach*. Retrieved from https://books.google.com/books?hl=en&lr=&id=xAHCOmtAZd0C&oi=fnd&pg=PR5&dq=Maxwell,+J.+A.+(2005b).+Qualitative+Research+Design:+An+Interactive+Approach.+SAGE.&ots=Y1HRnnucd1&sig=UHdq3YjnJbIdXwUJ6dfkr9Oqndc

McCallum, A., Nigam, K., Rennie, J., & Seymore, K. (1999). A machine learning approach to building domain-specific search engines. *IJCAI International Joint Conference on Artificial Intelligence*, *2*, 662–667. Retrieved from www.campsearch.com

McClung, H. J., Murray, R. D., & Heitlinger, L. A. (1998). The Internet as a Source for Current Patient Information. *PEDIATRICS*, *101*(6), e2–e2. https://doi.org/10.1542/peds.101.6.e2

McKee, M. M., & Paasche-Orlow, M. K. (2012). Health literacy and the disenfranchised: The importance of collaboration between limited english proficiency and health literacy researchers. *Journal of Health Communication*, *17*(SUPPL. 3), 7–12. https://doi.org/10.1080/10810730.2012.712627

McKiernan, G. (2001). *Beyond bookmarks: schemes for organizing the Web* (Vol. 2007). Vol. 2007.

Mental Health. (n.d.). Retrieved from https://medlineplus.gov/mentalhealth.html

288

Mertens, F., Saint-Charles, J., & Mergler, D. (2012). Social communication network analysis of the role of participatory research in the adoption of new fish consumption behaviors. *Social Science and Medicine*, *75*(4), 643–650. https://doi.org/10.1016/j.socscimed.2011.10.016

Miller, N., Lacroix, E. M., & Backus, J. E. (2000). MEDLINEplus: building and maintaining the National Library of Medicine's consumer health Web service. *Bulletin of the Medical Library Association*, *88*(1), 11–17. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10658959

Minetou, C. G., Chen, S. Y., & Liu, X. (2008). Investigation of the use of navigation tools in web-based learning: A data mining approach. *International Journal of Human-Computer Interaction*, *24*(1), 48–67. https://doi.org/10.1080/10447310701771480

Möller, J., van de Velde, R. N., Merten, L., & Puschmann, C. (2019). Explaining Online News Engagement Based on Browsing Behavior: Creatures of Habit? *Social Science Computer Review*, 089443931982801. https://doi.org/10.1177/0894439319828012

Moore, M., & Kaercher, D. (2008). Healthy Texas: The Eight-Year Odyssey of a Community Health Web Portal. *Journal of Consumer Health On the Internet*, *12*(2), 95–104. https://doi.org/10.1080/15398280802121331

Morahan-Martin, J. M. (2004). How Internet Users Find, Evaluate, and Use Online Health Information: A Cross-Cultural Review. *CyberPsychology & Behavior*, *7*(5), 497–510. https://doi.org/10.1089/cpb.2004.7.497

Newnham, G. M., Burns, W. I., Snyder, R. D., Dowling, A. J., Ranieri, N. F., Gray, E. L., & McLachlan, S. A. (2006). Information from the internet: Attitudes of Australian oncology patients. *Internal Medicine Journal*, *36*(11), 718–723. https://doi.org/10.1111/j.1445-

5994.2006.01212.x

Nie, L., Li, T., Akbari, M., Shen, J., & Chua, T.-S. (2014). WenZher. *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval - SIGIR '14*, 1245–1246. https://doi.org/10.1145/2600428.2611176

Nie, L., Zhao, Y.-L., Akbari, M., Shen, J., & Chua, T.-S. (2015). Bridging the Vocabulary Gap between Health Seekers and Healthcare Knowledge. *IEEE Transactions on Knowledge and Data Engineering*, *27*(2), 396–409. https://doi.org/10.1109/TKDE.2014.2330813

Niederdeppe, J., Hornik, R. C., Kelly, B. J., Frosch, D. L., Romantan, A., Stevens, R. S., … Schwartz, J. S. (2007). Examining the dimensions of cancer-related information seeking and scanning behavior. *Health Communication*, *22*(2), 153–167. https://doi.org/10.1080/10410230701454189

Otte, E., & Rousseau, R. (2002). Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science*, *28*(6), 441–453. https://doi.org/10.1177/016555150202800601

Pandolfini, C., Impicciatore, P., & Bonati, M. (2000). Parents on the Web: Risks for Quality Management of Cough in Children. *Pediatrics*, *105*(1), e1–e1. https://doi.org/10.1542/peds.105.1.e1

Pang, P. C. I., Chang, S., Verspoor, K. & Pearce, J. (2016). Designing health websites based on users' web-based information-seeking behaviors: A mixed-method observational study. *Journal of Medical Internet Research*, *18*(6), e145. https://doi.org/10.2196/jmir.5661

Pang, P. C. I., Harrop, M., Verspoor, K., Pearce, J., & Chang, S. (2016). What are health website visitors doing? Insights from visualisations toward exploratory search. *Proceedings of the*

28th Australian Computer-Human Interaction Conference, OzCHI 2016, 631–633.

https://doi.org/10.1145/3010915.3011843

Paolino, L., Genser, L., Fritsch, S., de' Angelis, N., Azoulay, D., & Lazzati, A. (2015). The

Web-Surfing Bariatic Patient: the Role of the Internet in the Decision-Making Process.

Obesity Surgery, *25*(4), 738–743. https://doi.org/10.1007/s11695-015-1578-x

Patrick, K., & Koss, S. (1995). *Consumer health information: white paper (Working draft 15*

May 1995.). Retrieved from http://nii.nist.gov/pubs/chi.html

Pepper, S. (2002). *The TAO of Topic Maps*.

Polamuri, S. (2015). Five Most Popular Similarity Measures Implementation in Python.

Retrieved from https://dataaspirant.com/2015/04/11/five-most-popular-similarity-measures-

implementation-in-python/

Postigo, H. (2011). Questioning the Web 2.0 discourse: Social roles, production, values, and the

case of the Human Rights Portal. *Information Society*, *27*(3), 181–193.

https://doi.org/10.1080/01972243.2011.566759

Pow, J., Gayen, K., Elliott, L., & Raeside, R. (2012). Understanding complex interactions using

social network analysis. *Journal of Clinical Nursing*, *21*(19–20), 2772–2779.

https://doi.org/10.1111/j.1365-2702.2011.04036.x

Powell, J., & Clarke, A. (2006). Internet information-seeking in mental health. *British Journal of*

Psychiatry, *189*(3), 273–277. https://doi.org/10.1192/bjp.bp.105.017319

Price, J. (2000). *The complete idiot's guide to online medical resources* (1st ed.). Alpha.

Proposed working definition of an older person in Africa for the MDS Project. (2002). Retrieved

from https://www.who.int/healthinfo/survey/ageingdefnolder/en/

291

Prybutok, G., & Ryan, S. (2015). Social Media: The Key to Health Information Access for 18- to 30-Year-Old College Students. *CIN - Computers Informatics Nursing*, *33*(4), 132–141. https://doi.org/10.1097/CIN.0000000000000147

Pullen, K., Jones, D., & Timm, D. F. (2011). Healthy connections for consumers: National library of medicine web sites. *Journal of Hospital Librarianship*, *11*(1), 23–38. https://doi.org/10.1080/15323269.2011.537979

Qiu, W. Q., Dean, M., Liu, T., George, L., Gann, M., Cohen, J., & Bruce, M. L. (2010). Physical and Mental Health of Homebound Older Adults: An Overlooked Population. *Journal of the American Geriatrics Society*, *58*(12), 2423–2428. https://doi.org/10.1111/j.1532-5415.2010.03161.x

Rains, S. A. (2018). Coping with Illness Digitally. In *Coping with Illness Digitally*. https://doi.org/10.7551/mitpress/11219.001.0001

Raju N, V., & Harinarayana, N. (2008). An analysis of usability features of library web sites. *Annals of Library and Information Studies*, *55*(2), 111–122.

Ratzan, S. C., Parker, R. M., Selden, C. R., & Zorn, M. (2000). *Health literacy*. National library of medicine current bibliographies in medicine: health literacy. Bethesda, MD: National Institutes of Health, US Department of Health and Human Services.

Rice, R. E. (2006). Influences, usage, and outcomes of Internet health information searching: Multivariate results from the Pew surveys. *International Journal of Medical Informatics*, *75*(1), 8–28. https://doi.org/10.1016/j.ijmedinf.2005.07.032

Rosenfeld, L., & Morville, P. (2002). *Information architecture for the world wide web*. " O'Reilly Media, Inc."

Ruppel, E. K., & Rains, S. A. (2012). Information Sources and the Health Information-Seeking Process: An Application and Extension of Channel Complementarity Theory. *Communication Monographs*, *79*(3), 385–405. https://doi.org/10.1080/03637751.2012.697627

Sage, A., Carpenter, D., Sayner, R., Thomas, K., Mann, L., Sulzer, S., … Sleath, B. (2018). Online Information-Seeking Behaviors of Parents of Children With ADHD. *Clinical Pediatrics*, *57*(1), 52–56. https://doi.org/10.1177/0009922817691821

Salah, A. A., Manovich, L., Salah, A. A., & Chow, J. (2013). Combining Cultural Analytics and Networks Analysis: Studying a Social Network Site with User-Generated Content. *Journal of Broadcasting and Electronic Media*, *57*(3), 409–426. https://doi.org/10.1080/08838151.2013.816710

Samal, L., Saha, S., Chander, G., Korthuis, P. T., Sharma, R. K., Sharp, V., … Beach, M. C. (2011). Internet Health information seeking behavior and antiretroviral adherence in persons living with HIV/AIDS. *AIDS Patient Care and STDs*, *25*(7), 445–449. https://doi.org/10.1089/apc.2011.0027

Schilling, K., & McDaniel, A. M. (2010). Development and Evaluation of a Cancer Information Web Portal: The Impact of Design and Presentation on User Engagement. *Journal of Consumer Health On the Internet*, *14*(3), 242–262. https://doi.org/10.1080/15398285.2010.501736

Schleyer, T., & Spallek, H. (2002). An evaluation of five dental Internet portals. *The Journal of the American Dental Association*, *133*(2), 204–212. https://doi.org/10.14219/jada.archive.2002.0145

293

Segaran, T. (2007). *Programming Collective Intelligence: Building Smart Web 2.0 Applications*. Retrieved from https://books.google.com/books?hl=en&lr=&id=fEsZ3Ey-Hq4C&oi=fnd&pg=PR13&dq=Segaran,+Toby.+Programming+Collective+Intelligence:+Building+Smart+Web+2.0+Applications.+Sebastopol,+CA:+O%27Reilly+Media,+2007.&ots=Sf-7TQJTDg&sig=OIWjGTQklOmWlWWVBnMRsaoVGAs

Shevchenko, L. (2020). Analysis of library website users' behavior to optimize virtual information and library services. *Journal of Information Science Theory and Practice*, *8*(1), 45–55. https://doi.org/10.1633/JISTaP.2020.8.1.4

Smedley, B., Stith, A., & Nelson, A. (2003). *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care.*

Smith, K. P., & Christakis, N. A. (2008). Social Networks and Health. *Annual Review of Sociology*, *34*(1), 405–429. https://doi.org/10.1146/annurev.soc.34.040507.134601

Sørensen, K., Van Den Broucke, S., Fullam, J., Doyle, G., Pelikan, J., Slonska, Z., & Brand, H. (2012, December 25). Health literacy and public health: A systematic review and integration of definitions and models. *BMC Public Health*, Vol. 12, p. 80. https://doi.org/10.1186/1471-2458-12-80

Stănescu, L., Burdescu, D., Mihai, G., Ion, A., & Stoica, C. (2008). Topic map for medical E-learning. *Studies in Computational Intelligence*, *162*, 305–310. https://doi.org/10.1007/978-3-540-85257-5_35

Sterling, E. W., von Esenwein, S. A., Tucker, S., Fricks, L., & Druss, B. G. (2010). Integrating Wellness, Recovery, and Self-management for Mental Health Consumers. *Community Mental Health Journal*, *46*(2), 130–138. https://doi.org/10.1007/s10597-009-9276-6

Sundar, S. S. (2008). The MAIN Model : A Heuristic Approach to Understanding Technology

Effects on Credibility. In A. J. Flanagin & M. Metzger (Eds.), *Digital media, youth, and*

credibility (pp. 73–100). https://doi.org/10.1162/dmal.9780262562324.073

Tandi Lwoga, E., & Florence Mosha, N. (2013). Information seeking behaviour of parents and

caregivers of children with mental illness in Tanzania. *Library Review*, *62*(8/9), 567–584.

https://doi.org/10.1108/LR-10-2012-0116

Taylor, R. S. (1967). *Question-Negotiation An Information-Seeking in Libraries*.

https://apps.dtic.mil/sti/citations/AD0659468 (Archived by the Internet Archive at

https://web.archive.org/web/20200623153637/https://apps.dtic.mil/sti/citations/AD0659468

)

Volandes, A. E., Paasche-Orlow, M., Gillick, M. R., Cook, E. F., Shaykevich, S., Abbo, E. D., &

Lehmann, L. (2008). Health literacy not race predicts end-of-life care preferences. *Journal*

of Palliative Medicine, *11*(5), 754–762. https://doi.org/10.1089/jpm.2007.0224

Webster, J., & Ahuja, J. S. (2006). Enhancing the design of Web navigation systems: The

influence of user disorientation on engagement and performance. *MIS Quarterly:*

Management Information Systems, *30*(3), 661–678. https://doi.org/10.2307/25148744

Wege, C. (2002). Portal server technology. *IEEE Internet Computing*, *6*(3), 73–77.

https://doi.org/10.1109/MIC.2002.1003134

WHO Mental Disorders Fact Sheet. (2018). Retrieved September 20, 2019, from

https://www.who.int/en/news-room/fact-sheets/detail/mental-disorders

Winter, G. (2000). A Comparative Discussion of the Notion of "Validity" in Qualitative and

Quantitative Research. *The Qualitative Report*, *Volume 4*(3 & 4), 1–14. Retrieved from

https://nsuworks.nova.edu/tqr

Yang, H.-C., & Lee, C.-H. (2004). A text mining approach on automatic generation of web directories and hierarchies. *Expert Systems with Applications*, *27*(4), 645–663.

Yeo, M.-H., Lee, Y.-K., Roh, K.-J., Park, H.-S., Kim, H.-S., Park, J.-H., … Yoo, J.-S. (2010). Design and Implementation of a Directory System for Disease Services. *International Journal of Contents*, *6*(1), 59–64. https://doi.org/10.5392/IJoC.2010.6.1.059

Yi, J., & Zahn, M. (2010). Internet Use among Chinese-American breast Cancer survivors. *International Quarterly of Community Health Education*, *31*(4), 359–368. https://doi.org/10.2190/IQ.31.4.d

Yi, M. (2012). Effective medical resources searching using an ontology-driven medical information retrieval system h1n1 case study. *Electronic Library*, *30*(2), 248–263. https://doi.org/10.1108/02640471211221368

Yonker, L. M., Zan, S., Scirica, C. V, Jethwani, K., & Kinane, T. B. (2015). "Friending" Teens: Systematic Review of Social Media in Adolescent and Young Adult Health Care. *Journal of Medical Internet Research*, *17*(1), e4. https://doi.org/10.2196/jmir.3692

Zeng, Q. T., & Tse, T. (2006). Exploring and developing consumer health vocabularies. *Journal of the American Medical Informatics Association*, *13*(1), 24–29. https://doi.org/10.1197/jamia.M1761

Zhang, J., & An, L. (2009). Visualization relevance analysis for weight control-related subjects. *Proceedings of the ASIST Annual Meeting*, *46*(1), 1–5. https://doi.org/10.1002/meet.2009.145046032

Zhang, J., & Rasmussen, E. M. (2001). Developing a new similarity measure from two different

296

perspectives. *Information Processing and Management*, *37*(2), 279–294.

https://doi.org/10.1016/S0306-4573(00)00027-3

Zhang, J., & Wolfram, D. (2009). Visual analysis of obesity-related query terms on HealthLink.

Online Information Review, *33*(1), 43–57. https://doi.org/10.1108/14684520910944382

Zhang, J., Zhai, S., Liu, H., & Stevenson, J. A. (2016). Social network analysis on a topic-based

navigation guidance system in a public health portal. *Journal of the Association for*

Information Science and Technology, *67*(5), 1068–1088. https://doi.org/10.1002/asi.23468

Zhang, J., Zhai, S., Stevenson, J. A., & Xia, L. (2015). Optimization of the subject directory in a

government agriculture department web portal. *Journal of the Association for Information*

Science and Technology, *67*(9), 2166–2180. https://doi.org/10.1002/asi.23550

Zhao, Z., Feng, S., Wang, Q., Huang, J. Z., Williams, G. J., & Fan, J. (2012). Topic oriented

community detection through social objects and link analysis in social networks.

Knowledge-Based Systems, *26*, 164–173. https://doi.org/10.1016/j.knosys.2011.07.017

Zhu, Y., & Zhang, J. (2020). Social network analysis on the Mental Health sub-topic in the

MedlinePlus subject directory. *Information Research: An International Electronic Journal,*

25(4), paper 876. https://doi.org/10.47989/irpaper876

Zimmerman, D., & Paschal, D. B. (2009). An exploratory usability evaluation of Colorado State

University Libraries' digital collections and the Western Waters Digital Library Web sites.

Journal of Academic Librarianship, *35*(3), 227–240.

https://doi.org/10.1016/j.acalib.2009.03.011

APPENDICES

Appendix A. Evaluation sheet for mental health related topics

Topic A	Topic B	Strongly Relevant	Relevant	Little Relevant	Not Relevant
Bipolar Disorder	Mood Disorders				
Child Abuse	Domestic Violence				
Child Behavior Disorders	Teen Violence				
Child Nutrition	Diets				
Child Nutrition	Nutrition				
Child Nutrition	Obesity In Children				
Dash Diet	Diets				
Diabetic Diet	Carbohydrates				
Drugs And Young People	Drug Abuse				
E-Cigarettes	Smoking And Youth				
E-Cigarettes	Smoking				
Food Safety	Food Labeling				
Growth Disorders	Child Development				
Nutrition For Seniors	Diets				
Pregnancy And Nutrition	Nutrition				
Quitting Smoking	Smoking And Youth				
Smokeless Tobacco	Smoking And Youth				
Child Mental Health	Teen Mental Health				
Teen Mental Health	Teen Health				
Teenage Pregnancy	Teen Health				
Vegetarian Diet	Diets				
Veterans And Military Family Health	Veterans And Military Health				
Weight Control	Obesity In Children				
Alcohol	Fetal Alcohol Spectrum Disorders				
Body Weight	Teen Health				
Child Mental Health	Child Abuse				
Child Behavior Disorders	Child Abuse				
Child Development	Child Abuse				
Child Nutrition	Child Abuse				
Child Abuse	Elder Abuse				
Teen Mental Health	Child Behavior Disorders				
Child Behavior Disorders	School Health				
Child Behavior Disorders	Child Safety				
Child Behavior Disorders	Learning Disorders				
Child Behavior Disorders	Obesity In Children				
Child Behavior Disorders	Child Nutrition				
Child Behavior Disorders	Child Sexual Abuse				
Child Mental Health	Child Development				
Child Development	Child Safety				
Child Development	Learning Disorders				
Child Development	Obesity In Children				
Child Development	Child Nutrition				
Child Development	Exercise For Children				
Child Development	Child Sexual Abuse				

Child Mental Health	Child Safety				
Child Mental Health	Learning Disorders				
Child Mental Health	Toddler Development				
Child Mental Health	Obesity In Children				
Child Mental Health	Child Nutrition				
Child Mental Health	Child Sexual Abuse				
School Health	Child Nutrition				
Teen Health	Child Nutrition				
College Health	Child Nutrition				
Child Nutrition	Nutrition For Seniors				
Child Nutrition	Exercise For Children				
Child Nutrition	Food Labeling				
Child Nutrition	Child Sexual Abuse				
Child Safety	Food Safety				
Obesity In Children	Child Sexual Abuse				
Drugs And Young People	Club Drugs				
Nutrition	College Health				
College Health	Nutrition For Seniors				
Developmental Disabilities	Growth Disorders				
Diets	Dietary Fiber				
Diets	Weight Loss Surgery				
Obesity In Children	Exercise For Children				
Exercise For Children	Exercise For Seniors				
Underage Drinking	Fetal Alcohol Spectrum Disorders				
Nutrition For Seniors	Food Labeling				
Nutrition For Seniors	Food Safety				
Nutrition For Seniors	Weight Control				
Teen Health	Obesity In Children				
Pregnancy And Nutrition	Teenage Pregnancy				
Teen Violence	School Health				
Teen Sexual Health	Sexual Assault				
Teen Mental Health	Teen Development				
Teen Development	Teenage Pregnancy				
Teen Mental Health	Teenage Pregnancy				
Child Safety	Nutrition				
Dash Diet	Toddler Development				
Infant And Newborn Development	Dash Eating Plan				
Psychotic Disorders	Dash Eating Plan				
Drugs And Young People	Dash Eating Plan				
Dash Diet	Drugs And Young People				
Schizophrenia	Dash Eating Plan				
Dash Diet	Schizophrenia				
Bipolar Disorder	Dash Eating Plan				
Veterans And Military Family Health	Dash Eating Plan				
Toddler Development	Dash Eating Plan				
Dash Diet	Veterans And Military Family Health				
Child Safety	Compulsive Gambling				
Child Mental Health	Nutrition For Seniors				
Child Mental Health	Exercise For Seniors				
Child Safety	Dash Diet				
Child Safety	Personality Disorders				
Cocaine	Exercise For Seniors				

299

Developmental Disabilities	Food Safety				
Child Safety	Delirium				
Dash Diet	Psychotic Disorders				
Dash Diet	Developmental Disabilities				
Dash Diet	Infant And Newborn Development				
Delirium	Dietary Fiber				
Food Safety	Personality Disorders				
Child Safety	Seasonal Affective Disorder				
Child Safety	Diabetic Diet				
Developmental Disabilities	Nutrition For Seniors				
Developmental Disabilities	Exercise For Seniors				
Developmental Disabilities	Dietary Proteins				
Food Labeling	Personality Disorders				
Psychotic Disorders	Teen Sexual Health				
Seasonal Affective Disorder	Toddler Development				
Bipolar Disorder	Dash Diet				
Child Mental Health	Nutrition				
Developmental Disabilities	Dietary Fats				
Food Labeling	Psychotic Disorders				
Psychotic Disorders	Nutrition For Seniors				
Psychotic Disorders	Nutrition				
Psychotic Disorders	Exercise For Seniors				
Infant And Newborn Development	Cirrhosis				
Schizophrenia	Nutrition				
Toddler Development	Cirrhosis				
Delirium	Nutrition				
Self-Harm	Nutrition				
Seasonal Affective Disorder	Child Development				
Psychotic Disorders	Exercise And Physical Fitness				
Veterans And Military Health	Food Labeling				
Developmental Disabilities	Cholesterol				
Stress	Developmental Disabilities				
Suicide	Developmental Disabilities				
Exercise And Physical Fitness	Developmental Disabilities				
Developmental Disabilities	Breastfeeding				
Alcohol	Developmental Disabilities				
Developmental Disabilities	Drug Abuse				
Developmental Disabilities	Smoking				
Nutrition	Developmental Disabilities				
Developmental Disabilities	Vegetarian Diet				
Developmental Disabilities	Child Nutrition				
Marijuana	Developmental Disabilities				
Depression	Developmental Disabilities				

300

Weight Control	Developmental Disabilities				
Developmental Disabilities	Food Labeling				
Child Safety	Developmental Disabilities				
Developmental Disabilities	Vitamins				
Attention Deficit Hyperactivity Disorder	Developmental Disabilities				
Alcoholism And Alcohol Abuse	Developmental Disabilities				
Developmental Disabilities	Diabetic Diet				
Developmental Disabilities	Quitting Smoking				
Developmental Disabilities	Teenage Pregnancy				
Developmental Disabilities	Calcium				

Appendix B. Evaluation sheet for children related topics

Topic A	Topic B	Strongly Relevant	Relevant	Little Relevant	Not Relevant
Children's Health	Child Nutrition				
Child Dental Health	Children's Health				
Child Dental Health	Cosmetic Dentistry				
Child Dental Health	Tooth Disorders				
Child Development	Children's Health				
Child Safety	Children's Health				
Common Infant and Newborn Problems	Children's Health				
Exercise for Children	How Much Exercise Do I Need?				
Medicines and Children	Children's Health				
Medicines and Children	Medication Errors				
Toddler Health	Children's Health				
Uncommon Infant and Newborn Problems	Children's Health				
Developmental Disabilities	Child Development				
Growth Disorders	Child Development				
Infant and Newborn Development	Infant and Newborn Care				
Motor Vehicle Safety	Child Safety				
Flu Shot	Childhood Immunization				
Infant and Newborn Nutrition	Common Infant and Newborn Problems				
Infant and Newborn Nutrition	Uncommon Infant and Newborn Problems				
Benefits of Exercise	Exercise for Children				
Benefits of Exercise	Exercise and Physical Fitness				
Hearing Aids	Hearing Problems in Children				
Teen Mental Health	Teen Health				
Teenage Pregnancy	Teen Health				
Toddler Nutrition	Toddler Health				
Premature Babies	Uncommon Infant and Newborn Problems				
Child Mental Health	Teen Mental Health				
Self-Harm	Mental Disorders				
Learning Disabilities	Disabilities				
Pituitary Disorders	Growth Disorders				
Pituitary Disorders	Endocrine Diseases				
Hepatitis Testing	Hepatitis B				
Hepatitis Testing	Hepatitis				
Hepatitis Testing	Hepatitis A				
Hepatitis Testing	Hepatitis C				
Exercise for Older Adults	How Much Exercise Do I Need?				
Health Risks of an Inactive Lifestyle	Exercise and Physical Fitness				
How Much Exercise Do I Need?	Exercise and Physical Fitness				

302

Sports Injuries	Sports Safety				
Pregnancy and Medicines	Health Problems in Pregnancy				
Pregnancy and Medicines	Pregnancy				
Teen Violence	Teen Mental Health				
Health Problems in Pregnancy	Teenage Pregnancy				
Neural Tube Defects	Birth Defects				
Children's Health	Child Behavior Disorders				
Children's Health	Toddler Development				
Children's Health	Baby Health Checkup				
Children's Health	Child Mental Health				
Child Dental Health	Toddler Health				
Child Dental Health	Dentures				
Child Dental Health	Child Mental Health				
Child Dental Health	Child Nutrition				
Child Development	Medicines and Children				
Child Development	Toddler Health				
Child Development	Child Mental Health				
Child Development	Child Nutrition				
Child Safety	Medicines and Children				
Child Safety	Sports Safety				
Common Infant and Newborn Problems	Toddler Health				
Exercise for Children	Exercise for Older Adults				
Exercise for Children	Health Risks of an Inactive Lifestyle				
Hearing Problems in Children	Noise				
Medicines and Children	Toddler Health				
Medicines and Children	Child Behavior Disorders				
Medicines and Children	Child Mental Health				
Medicines and Children	Child Nutrition				
Teen Health	Child Nutrition				
Teen Health	Complementary and Integrative Medicine				
Teen Health	Teen Depression				
Toddler Health	Uncommon Infant and Newborn Problems				
Toddler Health	Child Behavior Disorders				
Toddler Health	Baby Health Checkup				
Toddler Health	Child Mental Health				
Toddler Health	Child Nutrition				
Dental Health	Complementary and Integrative Medicine				
Orthodontia	Tooth Decay				
Tooth Decay	Cosmetic Dentistry				
Child Behavior Disorders	Child Nutrition				
Developmental Disabilities	Toddler Development				
Infant and Newborn Development	Baby Health Checkup				

303

Infant and Newborn Development	Fetal Health and Development				
Teen Development	Teen Mental Health				
Teen Development	Teenage Pregnancy				
Teen Development	Teen Depression				
Infant and Newborn Care	Reflux in Infants				
Infant and Newborn Care	Premature Babies				
Infant and Newborn Care	Sudden Infant Death Syndrome				
Infant and Newborn Care	Fetal Health and Development				
Measles	Mumps				
Baby Health Checkup	Fetal Health and Development				
Infant and Newborn Nutrition	Reflux in Infants				
Infant and Newborn Nutrition	Sudden Infant Death Syndrome				
Benefits of Exercise	Exercise for Older Adults				
Benefits of Exercise	Health Risks of an Inactive Lifestyle				
Benefits of Exercise	How Much Exercise Do I Need?				
Hearing Aids	Noise				
Medicines	Medical Device Safety				
College Health	Mental Health				
Teen Mental Health	Teenage Pregnancy				
Teenage Pregnancy	Pregnancy and Medicines				
Cosmetic Dentistry	Dentures				
Child Mental Health	Child Nutrition				
Down Syndrome	Turner Syndrome				
Down Syndrome	Craniofacial Abnormalities				
Hormones	Pituitary Disorders				
Hormones	Prenatal Testing				
Hepatitis Testing	Prenatal Testing				
Health Risks of an Inactive Lifestyle	Exercise for Older Adults				
Health Risks of an Inactive Lifestyle	How Much Exercise Do I Need?				
Meniere's Disease	Genetic Brain Disorders				
Complementary and Integrative Medicine	Mental Health				
Drug Reactions	Drugs and Young People				
Drug Safety	Drugs and Young People				
Prenatal Testing	Newborn Screening				
Newborn Screening	Prenatal Testing				
Exercise for Older Adults	Health Risks of an Inactive Lifestyle				
Medical Device Safety	Assistive Devices				
Artificial Limbs	Mobility Aids				
Dwarfism	Artificial Limbs				

304

Dwarfism	Mobility Aids				
Mobility Aids	Barotrauma				
Artificial Limbs	Barotrauma				
Dwarfism	Barotrauma				
Assistive Devices	Dwarfism				
Dwarfism	Craniofacial Abnormalities				
Temporomandibular Joint Dysfunction	Dwarfism				
Dwarfism	Brain Malformations				
Temporomandibular Joint Dysfunction	Mobility Aids				
Baby Health Checkup	Dwarfism				
Temporomandibular Joint Dysfunction	Artificial Limbs				
Assistive Devices	Barotrauma				
Dentures	Mobility Aids				
Dentures	Artificial Limbs				
Orthodontia	Dentures				
Temporomandibular Joint Dysfunction	Barotrauma				
Baby Health Checkup	Mobility Aids				
Dentures	Dwarfism				
Artificial Limbs	Craniofacial Abnormalities				
Mobility Aids	Craniofacial Abnormalities				
Barotrauma	Tinnitus				
Mobility Aids	Brain Malformations				
Dentures	Barotrauma				
Artificial Limbs	Brain Malformations				
Temporomandibular Joint Dysfunction	Brain Malformations				
Cosmetic Dentistry	Mobility Aids				
Baby Health Checkup	Barotrauma				
Baby Health Checkup	Artificial Limbs				
Barotrauma	Craniofacial Abnormalities				
Mobility Aids	Drug Safety				
Assistive Devices	Temporomandibular Joint Dysfunction				
Medical Device Safety	Mobility Aids				
Barotrauma	Brain Malformations				
Artificial Limbs	Drug Safety				
Dentures	Temporomandibular Joint Dysfunction				
Cosmetic Dentistry	Artificial Limbs				
Dwarfism	Tinnitus				
Mobility Aids	Tinnitus				
Brain Malformations	Chiari Malformation				
Dwarfism	Chiari Malformation				
Orthodontia	Mobility Aids				
Cosmetic Dentistry	Dwarfism				
Dwarfism	Drug Safety				
Barotrauma	Drug Safety				
Barotrauma	Meniere's Disease				

305

Artificial Limbs	Tinnitus				
Mobility Aids	Chiari Malformation				
Artificial Limbs	Chiari Malformation				
Orthodontia	Dwarfism				
Uncommon Infant and Newborn Problems	Dwarfism				
Assistive Devices	Dentures				
Cosmetic Dentistry	Barotrauma				
Orthodontia	Artificial Limbs				
Toddler Nutrition	Dwarfism				
Cochlear Implants	Dentures				
Hearing Aids	Preterm Labor				
Stuttering	Preterm Labor				
Pituitary Disorders	Preterm Labor				
Sports Safety	Genetic Brain Disorders				
Impaired Driving	Preterm Labor				
Mobility Aids	Herbal Medicine				
Assistive Devices	Psoriasis				
Stuttering	Drug Safety				
Cough	Preterm Labor				
Drug Safety	Cleft Lip and Palate				
Drug Safety	Bereavement				
Child Safety	Cosmetic Dentistry				
Child Safety	Meniere's Disease				
Child Safety	Genetic Brain Disorders				
Tetanus	Toilet Training				
Whooping Cough	Drug Safety				
Toilet Training	Hepatitis Testing				
Toilet Training	Meniere's Disease				
Toilet Training	Preterm Labor				
Medical Device Safety	Cleft Lip and Palate				
Toddler Development	Medical Device Safety				
Hearing Disorders and Deafness	Toddler Nutrition				
Sudden Infant Death Syndrome	Cosmetic Dentistry				
Cosmetic Dentistry	Toilet Training				
Toilet Training	Drug Safety				
Medical Device Safety	Brain Malformations				
Gum Disease	Assistive Devices				
Immunization	Temporomandibular Joint Dysfunction				
Artificial Limbs	Herbal Medicine				
Sports Safety	Meniere's Disease				
Dandruff, Cradle Cap, and Other Scalp Conditions	Drug Safety				
Reflux in Infants	Teen Violence				
Assistive Devices	Sudden Infant Death Syndrome				
Birth Weight	Meniere's Disease				
Child Dental Health	Genetic Brain Disorders				
Infant and Newborn Care	Meniere's Disease				
Immunization	Chiari Malformation				
Tetanus, Diphtheria, and Pertussis Vaccines	Assistive Devices				

306

Cosmetic Dentistry	Drug Reactions				
Mobility Aids	Noise				
Tooth Decay	Hepatitis C				
Flu	Hepatitis C				
Tooth Decay	Flu				
Bullying and Cyberbullying	Hepatitis C				
Bullying and Cyberbullying	Flu				
Tooth Decay	Bullying and Cyberbullying				
Hepatitis C	Congenital Heart Defects				
Birth Defects	Hepatitis C				
Birth Defects	Flu				
Flu	Congenital Heart Defects				
Hepatitis C	Pregnancy and Medicines				
Tooth Decay	Birth Defects				
Tooth Decay	Congenital Heart Defects				
Reflux in Infants	Hepatitis C				
Reflux in Infants	Flu				
Flu	Pregnancy and Medicines				
Flu	Mental Disorders				
Hepatitis C	Mental Disorders				
Bullying and Cyberbullying	Birth Defects				
Tooth Decay	Pregnancy and Medicines				
Hepatitis C	Cleft Lip and Palate				
Tooth Disorders	Hepatitis C				
Tooth Decay	Mental Disorders				
Hepatitis C	Health Problems in Pregnancy				
Hepatitis C	Mental Health				

307

Appendix C. Evaluation sheet for teenagers related topics

Topic A	Topic B	Strongly Relevant	Relevant	Little Relevant	Not Relevant
Teen Mental Health	Teen Health				
Teenage Pregnancy	Teen Health				
Child Mental Health	Teen Mental Health				
Drugs and Young People	Drug Use and Addiction				
Self-harm	Mental Disorders				
Health Problems in Pregnancy	Teenage Pregnancy				
Pregnancy	Preconception Care				
Prenatal Care	Preconception Care				
Alcohol	Alcohol Use Disorder (AUD) Treatment				
Child Behavior Disorders	Autism Spectrum Disorder				
Developmental Disabilities	Child Development				
Growth Disorders	Child Development				
Infant and Newborn Development	Infant and Newborn Care				
Bipolar Disorder	Mental Disorders				
Bipolar Disorder	Mood Disorders				
Compulsive Gambling	Mental Disorders				
Eating Disorders	Mental Disorders				
Mood Disorders	Mental Disorders				
Obsessive-Compulsive Disorder	Mental Disorders				
Obsessive-Compulsive Disorder	Personality Disorders				
Schizophrenia	Psychotic Disorders				
Childbirth Problems	Health Problems in Pregnancy				
HIV/AIDS and Pregnancy	Health Problems in Pregnancy				
HIV/AIDS and Pregnancy	Pregnancy				
HIV/AIDS and Pregnancy	Infections and Pregnancy				
Infections and Pregnancy	Health Problems in Pregnancy				
Pregnancy and Medicines	Health Problems in Pregnancy				
Pregnancy and Medicines	Pregnancy				
Pregnancy and Medicines	Infections and Pregnancy				
Pregnancy and Drug Use	Pregnancy				
Pregnancy and Drug Use	Pregnancy and Medicines				
Pregnancy and Drug Use	Reproductive Hazards				
Tumors and Pregnancy	Health Problems in Pregnancy				
Tumors and Pregnancy	Pregnancy				
Ectopic Pregnancy	Pregnancy				

308

Miscarriage	Pregnancy				
Pregnancy and Nutrition	Pregnancy				
Motor Vehicle Safety	Child Safety				
Learning Disabilities	Disabilities				
Pituitary Disorders	Growth Disorders				
Pituitary Disorders	Endocrine Diseases				
Quitting Smoking	Smoking and Youth				
Smokeless Tobacco	Smoking and Youth				
Postpartum Depression	Postpartum Care				
Preterm Labor	Childbirth Problems				
Preterm Labor	Childbirth				
HIV/AIDS and Infections	HIV/AIDS and Pregnancy				
HIV/AIDS Medicines	HIV/AIDS and Pregnancy				
HIV/AIDS Medicines	HIV/AIDS and Infections				
HIV/AIDS Medicines	Living with HIV/AIDS				
Preconception Care	Pregnancy and Medicines				
Teen Health	Teen Depression				
College Health	Mental Health				
Teen Development	Teen Mental Health				
Teen Development	Teenage Pregnancy				
Teen Development	Teen Depression				
Teen Mental Health	Teenage Pregnancy				
Teen Mental Health	Compulsive Gambling				
Teen Mental Health	How to Improve Mental Health				
Teen Mental Health	Older Adult Mental Health				
Teen Sexual Health	Sexual Assault				
Teen Sexual Health	Sexually Transmitted Diseases				
Teenage Pregnancy	Infections and Pregnancy				
Teenage Pregnancy	Pregnancy and Medicines				
Teenage Pregnancy	Pregnancy and Drug Use				
Teenage Pregnancy	Ectopic Pregnancy				
Teenage Pregnancy	Miscarriage				
Teenage Pregnancy	Pregnancy and Nutrition				
Teenage Pregnancy	Preconception Care				
Underage Drinking	Alcohol Use Disorder (AUD) Treatment				
Child Development	Child Mental Health				
Child Development	Toddler Health				
Child Development	Child Abuse				
Child Mental Health	Toddler Health				
Child Mental Health	Child Abuse				
Drugs and Young People	Club Drugs				
Mental Disorders	How to Improve Mental Health				

Mental Disorders	Older Adult Mental Health				
Mental Disorders	Autism Spectrum Disorder				
Mental Health	Compulsive Gambling				
Mental Health	Older Adult Health				
Teen Depression	Postpartum Depression				
Health Problems in Pregnancy	Ectopic Pregnancy				
Health Problems in Pregnancy	Miscarriage				
Health Problems in Pregnancy	Pregnancy and Nutrition				
Health Problems in Pregnancy	Reproductive Hazards				
Health Problems in Pregnancy	Preconception Care				
Health Problems in Pregnancy	Pregnancy and Opioids				
Pregnancy	Twins, Triplets, Multiple Births				
Pregnancy	Pregnancy and Opioids				
Child Safety	Toddler Health				
Alcohol	Fetal Alcohol Spectrum Disorders				
Child Behavior Disorders	Toddler Health				
Child Behavior Disorders	Child Abuse				
Developmental Disabilities	Infant and Newborn Development				
Developmental Disabilities	Toddler Development				
Infant and Newborn Development	Fetal Health and Development				
Anxiety	Obsessive-Compulsive Disorder				
Bipolar Disorder	Obsessive-Compulsive Disorder				
Eating Disorders	Obsessive-Compulsive Disorder				
Eating Disorders	Autism Spectrum Disorder				
Mood Disorders	Obsessive-Compulsive Disorder				
Mood Disorders	Personality Disorders				
Mood Disorders	Autism Spectrum Disorder				
Obsessive-Compulsive Disorder	Panic Disorder				
Obsessive-Compulsive Disorder	Autism Spectrum Disorder				
Personality Disorders	Autism Spectrum Disorder				
How to Improve Mental Health	Older Adult Mental Health				

Childbirth Problems	Pregnancy and Drug Use				
Childbirth Problems	Premature Babies				
Childbirth Problems	Twins, Triplets, Multiple Births				
Fetal Health and Development	Stillbirth				
HIV/AIDS and Pregnancy	Pregnancy and Medicines				
HIV/AIDS and Pregnancy	Pregnancy and Drug Use				
Infections and Pregnancy	Pregnancy and Drug Use				
Infections and Pregnancy	Tumors and Pregnancy				
Infections and Pregnancy	Ectopic Pregnancy				
Infections and Pregnancy	Miscarriage				
Infections and Pregnancy	Pregnancy and Nutrition				
Infections and Pregnancy	Preconception Care				
Pregnancy and Medicines	Tumors and Pregnancy				
Pregnancy and Medicines	Twins, Triplets, Multiple Births				
Pregnancy and Medicines	Ectopic Pregnancy				
Pregnancy and Medicines	Miscarriage				
Pregnancy and Medicines	Pregnancy and Nutrition				
Pregnancy and Drug Use	Twins, Triplets, Multiple Births				
Pregnancy and Drug Use	Ectopic Pregnancy				
Pregnancy and Drug Use	Miscarriage				
Pregnancy and Drug Use	Pregnancy and Nutrition				
Pregnancy and Drug Use	Preconception Care				
Premature Babies	Infant and Newborn Care				
Tumors and Pregnancy	Pregnancy and Nutrition				
Ectopic Pregnancy	Pregnancy and Nutrition				
Ectopic Pregnancy	Preconception Care				
Miscarriage	Pregnancy and Nutrition				
Miscarriage	Preconception Care				
Pregnancy and Nutrition	Preconception Care				
Pregnancy and Nutrition	Pregnancy and Opioids				
Prenatal Testing	Hormones				
Infant and Newborn Care	Postpartum Care				
Alcohol Use Disorder (AUD) Treatment	Fetal Alcohol Spectrum Disorders				
Down Syndrome	Turner Syndrome				
Hormones	Pituitary Disorders				
Toddler Health	Child Abuse				
HIV: PrEP and PEP	Living with HIV/AIDS				

311

Preconception Care	Pregnancy and Opioids				
HIV: PrEP and PEP	HIV/AIDS and Pregnancy				
Child Mental Health	Child Safety				
Eating Disorders	Personality Disorders				
Toddler Health	Child Sexual Abuse				
E-Cigarettes	Smoking and Youth				
Compulsive Gambling	Dwarfism				
Compulsive Gambling	Postpartum Care				
Dwarfism	Postpartum Care				
Compulsive Gambling	Vasectomy				
Compulsive Gambling	Schizophrenia				
Compulsive Gambling	Motor Vehicle Safety				
Compulsive Gambling	Stillbirth				
Vasectomy	Dwarfism				
Motor Vehicle Safety	Dwarfism				
Dwarfism	Stillbirth				
Compulsive Gambling	Tubal Ligation				
Schizophrenia	Dwarfism				
Compulsive Gambling	Elder Abuse				
Compulsive Gambling	Disabilities				
Compulsive Gambling	Psychotic Disorders				
Postpartum Care	Stillbirth				
Compulsive Gambling	E-Cigarettes				
Dwarfism	Pituitary Disorders				
Compulsive Gambling	Smokeless Tobacco				
Tubal Ligation	Dwarfism				
Compulsive Gambling	Pituitary Disorders				
Disabilities	Dwarfism				
Dwarfism	Elder Abuse				
Vasectomy	Postpartum Care				
Compulsive Gambling	Preterm Labor				
Dwarfism	E-Cigarettes				
Compulsive Gambling	Mood Disorders				
Compulsive Gambling	Folic Acid				
Dwarfism	Smokeless Tobacco				
Psychotic Disorders	Dwarfism				
Motor Vehicle Safety	Postpartum Care				
Teen Violence	Compulsive Gambling				
Compulsive Gambling	Birth Weight				
Compulsive Gambling	Seasonal Affective Disorder				
Compulsive Gambling	Phobias				
Folic Acid	Dwarfism				
Birth Weight	Postpartum Care				
Twins, Triplets, Multiple Births	Postpartum Care				
Vasectomy	Motor Vehicle Safety				
Dwarfism	Preterm Labor				
Vaginal Bleeding	Compulsive Gambling				
Schizophrenia	Postpartum Care				
Compulsive Gambling	Twins, Triplets, Multiple Births				
Dwarfism	Birth Weight				
Tubal Ligation	Postpartum Care				

312

Teen Violence	Dwarfism				
Smoking and Youth	Compulsive Gambling				
Vasectomy	Stillbirth				
Compulsive Gambling	Speech and Language Problems in Children				
Preterm Labor	Postpartum Care				
Vaginal Bleeding	Dwarfism				
Motor Vehicle Safety	Stillbirth				
Compulsive Gambling	Fetal Health and Development				
Disabilities	Postpartum Care				
Fetal Health and Development	Postpartum Care				
Twins, Triplets, Multiple Births	Dwarfism				
Twins, Triplets, Multiple Births	Birth Weight				
Seasonal Affective Disorder	Dwarfism				
Twins, Triplets, Multiple Births	Stillbirth				
Premature Babies	Postpartum Care				
Vaginal Bleeding	Postpartum Care				
Dwarfism	Speech and Language Problems in Children				
Elder Abuse	Postpartum Care				
Preterm Labor	Stillbirth				
Smoking and Youth	Dwarfism				
Compulsive Gambling	Premature Babies				
Schizophrenia	Vasectomy				
Motor Vehicle Safety	Elder Abuse				
Teen Development	Compulsive Gambling				
E-Cigarettes	Smokeless Tobacco				
Tubal Ligation	Motor Vehicle Safety				
Pituitary Disorders	Postpartum Care				
Mood Disorders	Dwarfism				
Phobias	Dwarfism				
Preconception Care	Postpartum Care				
Psychotic Disorders	Postpartum Care				
Underage Drinking	Compulsive Gambling				
Fetal Health and Development	Dwarfism				
Schizophrenia	Motor Vehicle Safety				
Health Problems in Pregnancy	Elder Abuse				
Infant and Newborn Development	Cocaine				
Toddler Development	Cirrhosis				
Primary Ovarian Insufficiency	Elder Abuse				
Premenstrual Syndrome	Smokeless Tobacco				
Seasonal Affective Disorder	Toddler Health				
Alcohol	How to Prevent High Blood Pressure				

313

Down Syndrome	How to Prevent High Blood Pressure				
Child Safety	Premenstrual Syndrome				
Seasonal Affective Disorder	Childbirth				
Seasonal Affective Disorder	Toilet Training				
Cough	Preterm Labor				
Smoking and Youth	Childbirth				
Tumors and Pregnancy	Elder Abuse				
Impaired Driving	Preterm Labor				
Toilet Training	E-Cigarettes				
Prenatal Care	Elder Abuse				
Child Safety	Personality Disorders				
Infant and Newborn Development	Cirrhosis				
Seasonal Affective Disorder	Prenatal Testing				
Seasonal Affective Disorder	Smokeless Tobacco				
Tumors and Pregnancy	Child Sexual Abuse				
Childbirth	Impaired Driving				
Toilet Training	HIV/AIDS and Infections				
Cough	Cesarean Section				
Prenatal Care	Seasonal Affective Disorder				
Premenstrual Syndrome	Toilet Training				
Seasonal Affective Disorder	Fetal Health and Development				
Rh Incompatibility	Elder Abuse				
Cirrhosis	Toilet Training				
Toilet Training	Preterm Labor				
Elder Abuse	Cesarean Section				
Impaired Driving	Postpartum Care				
Growth Disorders	Childbirth				
Premenstrual Syndrome	Motor Vehicle Safety				
Childbirth Problems	Impaired Driving				
Elder Abuse	Birth Weight				
Menstruation	Elder Abuse				
Seasonal Affective Disorder	Cesarean Section				
Fetal Health and Development	Elder Abuse				
Childbirth	Alcohol Use Disorder (AUD) Treatment				
Prenatal Testing	Elder Abuse				
Impaired Driving	Cesarean Section				
Pituitary Disorders	Preterm Labor				
Fatty Liver Disease	How to Prevent High Blood Pressure				
Drug Use and Addiction	How to Prevent High Blood Pressure				
How to Prevent High Blood Pressure	HIV/AIDS Medicines				

314

Anxiety	How to Prevent High Blood Pressure				
Bullying and Cyberbullying	How to Prevent High Blood Pressure				
Suicide	How to Prevent High Blood Pressure				
Eating Disorders	How to Prevent High Blood Pressure				
How to Prevent High Blood Pressure	HIV/AIDS				
Birth Defects	How to Prevent High Blood Pressure				
Post-Traumatic Stress Disorder	How to Prevent High Blood Pressure				
Pregnancy and Medicines	How to Prevent High Blood Pressure				
Marijuana	How to Prevent High Blood Pressure				
Mental Disorders	How to Prevent High Blood Pressure				
How to Prevent High Blood Pressure	HIV: PrEP and PEP				
How to Prevent High Blood Pressure	Genetic Testing				
Depression	How to Prevent High Blood Pressure				
Infections and Pregnancy	How to Prevent High Blood Pressure				
Fatty Liver Disease	Drug Use and Addiction				
Pregnancy and Nutrition	How to Prevent High Blood Pressure				
Attention Deficit Hyperactivity Disorder	How to Prevent High Blood Pressure				
Health Problems in Pregnancy	How to Prevent High Blood Pressure				
Pregnancy	How to Prevent High Blood Pressure				
Domestic Violence	How to Prevent High Blood Pressure				
Alcohol Use Disorder (AUD)	How to Prevent High Blood Pressure				

315

Appendix D. Evaluation sheet for older adults related topics

Topic A	Topic B	Strongly Relevant	Relevant	Little Relevant	Not Relevant
Exercise for Older Adults	Healthy Aging				
Exercise for Older Adults	How Much Exercise Do I Need?				
Home Care Services	Hospice Care				
Nutrition for Older Adults	Healthy Aging				
Child Care	Caregivers				
Benefits of Exercise	Exercise and Physical Fitness				
Benefits of Exercise	Exercise for Children				
Exercise for Children	How Much Exercise Do I Need?				
Health Risks of an Inactive Lifestyle	Exercise and Physical Fitness				
How Much Exercise Do I Need?	Exercise and Physical Fitness				
Sports Injuries	Sports Safety				
Managed Care	Health Insurance				
Child Nutrition	Nutrition				
Child Nutrition	Obesity in Children				
Dietary Fats	How to Lower Cholesterol with Diet				
Eating Disorders	Mental Disorders				
Food Safety	Food Labeling				
Weight Control	Obesity in Children				
Child Mental Health	Teen Mental Health				
Child Development	Parenting				
Child Development	Children's Health				
Bipolar Disorder	Mental Disorders				
Bipolar Disorder	Mood Disorders				
Mood Disorders	Mental Disorders				
Sprains and Strains	Ankle Injuries and Disorders				
Children's Health	Child Nutrition				
Cholesterol Medicines	HDL: The "Good" Cholesterol				
Cholesterol Medicines	LDL: The "Bad" Cholesterol				
Cholesterol Medicines	VLDL Cholesterol				
HDL: The "Good" Cholesterol	How to Lower Cholesterol with Diet				
HDL: The "Good" Cholesterol	VLDL Cholesterol				
How to Lower Cholesterol	HDL: The "Good" Cholesterol				
How to Lower Cholesterol	LDL: The "Bad" Cholesterol				
How to Lower Cholesterol with Diet	Cholesterol Levels: What You Need to Know				

316

VLDL Cholesterol	Cholesterol Levels: What You Need to Know				
DASH Eating Plan	Diets				
Vitamin D Deficiency	Vitamin D				
Vitamin E	Vitamins				
Vitamin E	Antioxidants				
Vitamin K	Vitamins				
Anxiety	Mental Disorders				
Compulsive Gambling	Mental Disorders				
Obsessive-Compulsive Disorder	Mental Disorders				
Obsessive-Compulsive Disorder	Personality Disorders				
Panic Disorder	Mental Disorders				
Schizophrenia	Psychotic Disorders				
Self-Harm	Mental Disorders				
Teen Violence	Teen Mental Health				
Older Adult Health	Mental Health				
End of Life Issues	Palliative Care				
End of Life Issues	Managed Care				
Exercise for Older Adults	Benefits of Exercise				
Exercise for Older Adults	Exercise for Children				
Exercise for Older Adults	Health Risks of an Inactive Lifestyle				
Healthy Aging	Older Adult Mental Health				
Healthy Aging	Financial Assistance				
Healthy Aging	Health Insurance				
Healthy Aging	Nutrition				
Healthy Aging	Mental Health				
Healthy Aging	How to Improve Mental Health				
Healthy Aging	Teen Mental Health				
Home Care Services	Nursing Homes				
Home Care Services	Managed Care				
Nutrition for Older Adults	Older Adult Mental Health				
Nutrition for Older Adults	Mental Health				
Nutrition for Older Adults	Child Nutrition				
Nutrition for Older Adults	How to Improve Mental Health				
Older Adult Mental Health	Child Mental Health				
Older Adult Mental Health	How to Improve Mental Health				
Older Adult Mental Health	Mental Disorders				
Older Adult Mental Health	Teen Mental Health				
Hospice Care	Managed Care				
Child Care	Child Nutrition				
Child Care	Child Mental Health				

317

Child Care	Child Development				
Child Care	Children's Health				
Child Care	Child Behavior Disorders				
Telehealth	Managed Care				
Financial Assistance	Health Facilities				
Financial Assistance	Managed Care				
Health Insurance	Health Facilities				
Health Insurance	Mental Health				
Health Facilities	Managed Care				
Nutrition	Toddler Nutrition				
Mental Health	Compulsive Gambling				
Parenting	Child Nutrition				
Parenting	Child Mental Health				
Parenting	Children's Health				
Parenting	Child Behavior Disorders				
Palliative Care	Managed Care				
Alzheimer's Disease	Creutzfeldt-Jakob Disease				
Benefits of Exercise	Health Risks of an Inactive Lifestyle				
Benefits of Exercise	How Much Exercise Do I Need?				
Exercise for Children	Health Risks of an Inactive Lifestyle				
Health Risks of an Inactive Lifestyle	How Much Exercise Do I Need?				
Child Nutrition	Child Mental Health				
Child Nutrition	Child Development				
Child Nutrition	Child Behavior Disorders				
Dietary Fiber	Diets				
Diets	Body Weight				
Diets	Dietary Supplements				
Eating Disorders	Mood Disorders				
Eating Disorders	Carbohydrate Metabolism Disorders				
Eating Disorders	Obsessive-Compulsive Disorder				
Eating Disorders	Personality Disorders				
Vitamins	Vitamin D Deficiency				
Child Mental Health	Child Development				
Child Mental Health	Children's Health				
How to Improve Mental Health	Mental Disorders				
How to Improve Mental Health	Teen Mental Health				
Teen Mental Health	Bipolar Disorder				
Teen Mental Health	Compulsive Gambling				
Bipolar Disorder	Obsessive-Compulsive Disorder				
Mood Disorders	Obsessive-Compulsive Disorder				
Mood Disorders	Panic Disorder				
Mood Disorders	Personality Disorders				

318

Postpartum Depression	Teen Depression				
Children's Health	Obesity in Children				
Children's Health	Child Behavior Disorders				
Cholesterol Medicines	How to Lower Cholesterol with Diet				
Heart Diseases	Creutzfeldt-Jakob Disease				
How to Lower Cholesterol	VLDL Cholesterol				
How to Lower Cholesterol with Diet	VLDL Cholesterol				
How to Lower Cholesterol with Diet	DASH Eating Plan				
Vitamin D Deficiency	B Vitamins				
Vitamin D Deficiency	Vitamin A				
Vitamin D Deficiency	Vitamin C				
Vitamin D Deficiency	Vitamin E				
Vitamin D Deficiency	Vitamin K				
Antioxidants	Vitamin D				
B Vitamins	Vitamin A				
B Vitamins	Vitamin C				
B Vitamins	Vitamin D				
B Vitamins	Vitamin E				
B Vitamins	Vitamin K				
Vitamin A	Vitamin C				
Vitamin A	Vitamin D				
Vitamin A	Vitamin E				
Vitamin A	Vitamin K				
Vitamin C	Vitamin D				
Vitamin C	Vitamin E				
Vitamin C	Vitamin K				
Vitamin D	Vitamin E				
Vitamin D	Vitamin K				
Vitamin E	Vitamin K				
Anxiety	Obsessive-Compulsive Disorder				
Medicaid	Financial Assistance				
Water Safety (Recreational)	Sports Safety				
Folic Acid	B Vitamins				
Assisted Living	Financial Assistance				
Assisted Living	Patient Rights				
Assisted Living	Compulsive Gambling				
Medicare Prescription Drug Coverage	Medicaid				
Medicare Prescription Drug Coverage	Compulsive Gambling				
Child Care	Sports Fitness				
Child Care	Financial Assistance				
Child Care	Health Facilities				
Child Care	Divorce				
Child Care	Medicaid				
Child Care	Malnutrition				
Child Care	Water Safety (Recreational)				

319

Child Care	Patient Rights				
Child Care	Carbohydrate Metabolism Disorders				
Child Care	Toddler Nutrition				
Child Care	Vitamin A				
Child Care	Vitamin C				
Child Care	Vitamin E				
Child Care	Compulsive Gambling				
Sports Fitness	Medicaid				
Sports Fitness	Vitamin C				
Sports Fitness	Compulsive Gambling				
Financial Assistance	Divorce				
Financial Assistance	Malnutrition				
Financial Assistance	Water Safety (Recreational)				
Financial Assistance	Patient Rights				
Financial Assistance	Carbohydrate Metabolism Disorders				
Financial Assistance	Toddler Nutrition				
Financial Assistance	Vitamin A				
Financial Assistance	Vitamin C				
Financial Assistance	Vitamin E				
Financial Assistance	Compulsive Gambling				
Financial Assistance	Schizophrenia				
Health Facilities	Medicaid				
Health Facilities	Patient Rights				
Health Facilities	Compulsive Gambling				
Divorce	Medicaid				
Divorce	Malnutrition				
Divorce	Patient Rights				
Divorce	Toddler Nutrition				
Divorce	Vitamin C				
Divorce	Compulsive Gambling				
Sports Safety	Compulsive Gambling				
Medicaid	Malnutrition				
Medicaid	Water Safety (Recreational)				
Medicaid	Patient Rights				
Medicaid	Carbohydrate Metabolism Disorders				
Medicaid	Toddler Nutrition				
Medicaid	Vitamin A				
Medicaid	Vitamin C				
Medicaid	Vitamin E				
Medicaid	Compulsive Gambling				
Medicaid	Schizophrenia				
Dietary Fiber	Compulsive Gambling				
Malnutrition	Water Safety (Recreational)				
Malnutrition	Patient Rights				
Malnutrition	Fluid and Electrolyte Balance				
Malnutrition	Carbohydrate Metabolism Disorders				
Malnutrition	Toddler Nutrition				
Malnutrition	Antioxidants				

320

Malnutrition	B Vitamins				
Malnutrition	Vitamin A				
Malnutrition	Vitamin C				
Malnutrition	Vitamin E				
Malnutrition	Vitamin K				
Malnutrition	Compulsive Gambling				
Malnutrition	Schizophrenia				
Nutritional Support	Toddler Nutrition				
Nutritional Support	Compulsive Gambling				
Mood Disorders	Carbohydrate Metabolism Disorders				
Mood Disorders	Compulsive Gambling				
Water Safety (Recreational)	Vitamin C				
Water Safety (Recreational)	Compulsive Gambling				
Patient Rights	Carbohydrate Metabolism Disorders				
Patient Rights	Toddler Nutrition				
Patient Rights	Vitamin A				
Patient Rights	Vitamin C				
Patient Rights	Vitamin E				
Patient Rights	Compulsive Gambling				
Fluid and Electrolyte Balance	Compulsive Gambling				
Carbohydrate Metabolism Disorders	Vitamin C				
Carbohydrate Metabolism Disorders	Vitamin E				
Carbohydrate Metabolism Disorders	Compulsive Gambling				
Toddler Nutrition	Vitamin A				
Toddler Nutrition	Vitamin C				
Toddler Nutrition	Vitamin E				
Toddler Nutrition	Compulsive Gambling				
Lipid Metabolism Disorders	Compulsive Gambling				
Colonic Diseases	Compulsive Gambling				
DASH Eating Plan	Compulsive Gambling				
Creutzfeldt-Jakob Disease	Compulsive Gambling				
Malabsorption Syndromes	Compulsive Gambling				
Antioxidants	Compulsive Gambling				
B Vitamins	Compulsive Gambling				
Folic Acid	Vitamin C				
Folic Acid	Compulsive Gambling				
Vitamin A	Compulsive Gambling				
Vitamin C	Compulsive Gambling				
Vitamin C	Schizophrenia				
Vitamin E	Compulsive Gambling				
Vitamin K	Compulsive Gambling				
Compulsive Gambling	Psychotic Disorders				
Compulsive Gambling	Schizophrenia				
Compulsive Gambling	Teen Violence				

321

Home Care Services	Ankle Injuries and Disorders				
Medicare Prescription Drug Coverage	Elbow Injuries and Disorders				
Medicare Prescription Drug Coverage	Hand Injuries and Disorders				
Adoption	Hand Injuries and Disorders				
Divorce	Dietary Proteins				
Palliative Care	Folic Acid				
Blood Transfusion and Donation	Food Labeling				
Blood Transfusion and Donation	Vegetarian Diet				
Assisted Living	Elbow Injuries and Disorders				
Assisted Living	Folic Acid				
Medicare Prescription Drug Coverage	Exercise for Children				
Heart Diseases	Cholesterol Levels: What You Need to Know				
Heart Diseases	Post-Traumatic Stress Disorder				
Food Labeling	Hand Injuries and Disorders				
Food Labeling	Shoulder Injuries and Disorders				
Medicare Prescription Drug Coverage	Shoulder Injuries and Disorders				
Health Insurance	Ankle Injuries and Disorders				
Adoption	How to Lower Cholesterol with Diet				
Adoption	VLDL Cholesterol				
Medicaid	Ankle Injuries and Disorders				
Food Labeling	Ankle Injuries and Disorders				
Vegetarian Diet	Ankle Injuries and Disorders				
Ankle Injuries and Disorders	Folic Acid				
Assisted Living	Ankle Injuries and Disorders				
Medicare Prescription Drug Coverage	Adoption				
Medicare Prescription Drug Coverage	Sports Injuries				
Nursing Homes	Vegetarian Diet				
Nursing Homes	Folic Acid				
Adoption	Blood Transfusion and Donation				
Blood Transfusion and Donation	Managed Care				
Managed Care	Ankle Injuries and Disorders				

Vegetarian Diet	Shoulder Injuries and Disorders				
Medicare Prescription Drug Coverage	Malabsorption Syndromes				
Bereavement	Antioxidants				
Nutrition	Blood Transfusion and Donation				
How to Improve Mental Health	Malabsorption Syndromes				
Assisted Living	Carbohydrates				
Sports Injuries	How to Lower Cholesterol with Diet				
Nutrition	Ankle Injuries and Disorders				
Sprains and Strains	How to Lower Cholesterol with Diet				
Health Risks of an Inactive Lifestyle	Nutritional Support				
Dietary Proteins	Delirium				
Family Issues	B Vitamins				
Benefits of Exercise	Nutritional Support				
Sports Injuries	Food Labeling				
Managed Care	Lipid Metabolism Disorders				
Seasonal Affective Disorder	Toddler Nutrition				
Seasonal Affective Disorder	How to Lower Cholesterol with Diet				
Ankle Injuries and Disorders	How to Lower Cholesterol with Diet				
Eating Disorders	Cholesterol Levels: What You Need to Know				
Eating Disorders	How to Prevent Heart Disease				
Eating Disorders	Heart Diseases				
Cholesterol Levels: What You Need to Know	Anxiety				
Cholesterol Levels: What You Need to Know	Bullying and Cyberbullying				
Suicide	Cholesterol Levels: What You Need to Know				
How to Prevent Heart Disease	Anxiety				
How to Prevent Heart Disease	Bullying and Cyberbullying				
Suicide	How to Prevent Heart Disease				
Dementia	Cholesterol Levels: What You Need to Know				
Cholesterol Levels: What You Need to Know	Post-Traumatic Stress Disorder				
Heart Diseases	Anxiety				
Suicide	Heart Diseases				
Eating Disorders	Bullying and Cyberbullying				
Heart Diseases	Bullying and Cyberbullying				

Suicide	Eating Disorders				
Cholesterol	Eating Disorders				
Cholesterol Levels: What You Need to Know	How to Prevent Heart Disease				
Alzheimer's Disease	Cholesterol Levels: What You Need to Know				
Cholesterol Levels: What You Need to Know	Vitamin D Deficiency				
How to Prevent Heart Disease	Post-Traumatic Stress Disorder				
Cholesterol	Anxiety				
Eating Disorders	Vitamin D Deficiency				
Dementia	How to Prevent Heart Disease				

324

Appendix E. Full Table 9

Topic A	Topic B	Topic A	Topic B
Child development	Child mental health	Child abuse	Child development
Child safety	Child mental health	Child sexual abuse	Child development
Learning disorders	Child mental health	Teenage pregnancy	Teen development
Toddler development	Child mental health	Teen health	Body weight
Obesity in children	Child mental health	Weight loss surgery	Diets
Child nutrition	Child mental health	Dietary fiber	Diets
Child abuse	Child mental health	College health	Nutrition
Child sexual abuse	Child mental health	Obesity in children	Teen health
Child behavior disorders	Teen mental health	Child nutrition	Teen health
Teen development	Teen mental health	Fetal alcohol spectrum disorders	Alcohol
Teenage pregnancy	Teen mental health	Food safety	Child safety
School health	Child behavior disorders	Child nutrition	College health
Child safety	Child behavior disorders	Nutrition for seniors	College health
Learning disorders	Child behavior disorders	Growth disorders	Developmental disabilities
Obesity in children	Child behavior disorders	Sexual assault	Teen sexual health
Child nutrition	Child behavior disorders	Exercise for children	Obesity in children
Child abuse	Child behavior disorders	Child sexual abuse	Obesity in children
Child sexual abuse	Child behavior disorders	Nutrition for seniors	Child nutrition
Nutrition for seniors	Weight control	Exercise for children	Child nutrition
Club drugs	Drugs and young people	Food labeling	Child nutrition
School health	Teen violence	Child abuse	Child nutrition
Fetal alcohol spectrum disorders	Underage drinking	Child sexual abuse	Child nutrition
Child nutrition	School health	Food labeling	Nutrition for seniors
Child safety	Child development	Food safety	Nutrition for seniors
Learning disorders	Child development	Exercise for seniors	Exercise for children
Obesity in children	Child development	Teenage pregnancy	Pregnancy and nutrition
Child nutrition	Child development	Elder abuse	Child abuse
Exercise for children	Child development		

325

Appendix F. Full Table 10

Topic	Related topic	Topic	Related topic
Child mental health	Teen mental health	Quitting smoking	Smoking and youth
Bipolar disorder	Mood disorders	Smokeless tobacco	Smoking and youth
Veterans and military family health	Veterans and military health	Child abuse	Domestic violence
Child behavior disorders	Teen violence	Teen mental health	Teen health
Growth disorders	Child development	Teenage pregnancy	Teen health
Child nutrition	Diets	Weight control	Obesity in children
Dash diet	Diets	Child nutrition	Obesity in children
Nutrition for seniors	Diets	Diabetic diet	Carbohydrates
Vegetarian diet	Diets	Food safety	Food labeling
Child nutrition	Nutrition	Drugs and young people	Drug abuse
Pregnancy and nutrition	Nutrition	E-cigarettes	Smoking
E-cigarettes	Smoking and youth		

Appendix G. Full Table 11

Topic A	Topic B	Topic A	Topic B
Children's Health	Child Behavior Disorders	Infant and Newborn Development	Baby Health Checkup
Children's Health	Toddler Development	Infant and Newborn Development	Fetal Health and Development
Children's Health	Baby Health Checkup	Teen Development	Teen Mental Health
Children's Health	Child Mental Health	Teen Development	Teenage Pregnancy
Child Dental Health	Toddler Health	Teen Development	Teen Depression
Child Dental Health	Dentures	Infant and Newborn Care	Reflux in Infants
Child Dental Health	Child Mental Health	Infant and Newborn Care	Premature Babies
Child Dental Health	Child Nutrition	Infant and Newborn Care	Sudden Infant Death Syndrome
Child Development	Medicines and Children	Infant and Newborn Care	Fetal Health and Development
Child Development	Toddler Health	Measles	Mumps
Child Development	Child Mental Health	Baby Health Checkup	Fetal Health and Development
Child Development	Child Nutrition	Infant and Newborn Nutrition	Reflux in Infants
Child Safety	Medicines and Children	Infant and Newborn Nutrition	Sudden Infant Death Syndrome
Child Safety	Sports Safety	Benefits of Exercise	Exercise for Older Adults
Common Infant and Newborn Problems	Toddler Health	Benefits of Exercise	Health Risks of an Inactive Lifestyle
Exercise for Children	Exercise for Older Adults	Benefits of Exercise	How Much Exercise Do I Need?
Exercise for Children	Health Risks of an Inactive Lifestyle	Hearing Aids	Noise
Hearing Problems in Children	Noise	Medicines	Medical Device Safety
Medicines and Children	Toddler Health	College Health	Mental Health
Medicines and Children	Child Behavior Disorders	Teen Mental Health	Teenage Pregnancy
Medicines and Children	Child Mental Health	Teenage Pregnancy	Pregnancy and Medicines
Medicines and Children	Child Nutrition	Cosmetic Dentistry	Dentures
Teen Health	Child Nutrition	Child Mental Health	Child Nutrition
Teen Health	Complementary and Integrative Medicine	Down Syndrome	Turner Syndrome
Teen Health	Teen Depression	Down Syndrome	Craniofacial Abnormalities
Toddler Health	Uncommon Infant and Newborn Problems	Hormones	Pituitary Disorders
Toddler Health	Child Behavior Disorders	Hormones	Prenatal Testing
Toddler Health	Baby Health Checkup	Hepatitis Testing	Prenatal Testing
Toddler Health	Child Mental Health	Health Risks of an Inactive Lifestyle	Exercise for Older Adults
Toddler Health	Child Nutrition	Health Risks of an Inactive Lifestyle	How Much Exercise Do I Need?

Dental Health	Complementary and Integrative Medicine	Meniere's Disease	Genetic Brain Disorders
Orthodontia	Tooth Decay	Complementary and Integrative Medicine	Mental Health
Tooth Decay	Cosmetic Dentistry	Drug Reactions	Drugs and Young People
Child Behavior Disorders	Child Nutrition	Drug Safety	Drugs and Young People
Developmental Disabilities	Toddler Development	Prenatal Testing	Newborn Screening

Appendix H. Full Table 12

Topic	Related topic	Topic	Related topic
Children's Health	Child Nutrition	Teen Mental Health	Teen Health
Child Dental Health	Children's Health	Teenage Pregnancy	Teen Health
Child Dental Health	Cosmetic Dentistry	Toddler Nutrition	Toddler Health
Child Dental Health	Tooth Disorders	Premature Babies	Uncommon Infant and Newborn Problems
Child Development	Children's Health	Child Mental Health	Teen Mental Health
Child Safety	Children's Health	Self-Harm	Mental Disorders
Common Infant and Newborn Problems	Children's Health	Learning Disabilities	Disabilities
Exercise for Children	How Much Exercise Do I Need?	Pituitary Disorders	Growth Disorders
Medicines and Children	Children's Health	Pituitary Disorders	Endocrine Diseases
Medicines and Children	Medication Errors	Hepatitis Testing	Hepatitis B
Toddler Health	Children's Health	Hepatitis Testing	Hepatitis
Uncommon Infant and Newborn Problems	Children's Health	Hepatitis Testing	Hepatitis A
Developmental Disabilities	Child Development	Hepatitis Testing	Hepatitis C
Growth Disorders	Child Development	Exercise for Older Adults	How Much Exercise Do I Need?
Infant and Newborn Development	Infant and Newborn Care	Health Risks of an Inactive Lifestyle	Exercise and Physical Fitness
Motor Vehicle Safety	Child Safety	How Much Exercise Do I Need?	Exercise and Physical Fitness
Flu Shot	Childhood Immunization	Sports Injuries	Sports Safety
Infant and Newborn Nutrition	Common Infant and Newborn Problems	Pregnancy and Medicines	Health Problems in Pregnancy
Infant and Newborn Nutrition	Uncommon Infant and Newborn Problems	Pregnancy and Medicines	Pregnancy
Benefits of Exercise	Exercise for Children	Teen Violence	Teen Mental Health
Benefits of Exercise	Exercise and Physical Fitness	Health Problems in Pregnancy	Teenage Pregnancy
Hearing Aids	Hearing Problems in Children	Neural Tube Defects	Birth Defects

329

Appendix I. Full Table 13

Topic A	Topic B	Topic A	Topic B
Teen Health	Teen Depression	Eating Disorders	Obsessive-Compulsive Disorder
College Health	Mental Health	Eating Disorders	Autism Spectrum Disorder
Teen Development	Teen Mental Health	Mood Disorders	Obsessive-Compulsive Disorder
Teen Development	Teenage Pregnancy	Mood Disorders	Personality Disorders
Teen Development	Teen Depression	Mood Disorders	Autism Spectrum Disorder
Teen Mental Health	Teenage Pregnancy	Obsessive-Compulsive Disorder	Panic Disorder
Teen Mental Health	Compulsive Gambling	Obsessive-Compulsive Disorder	Autism Spectrum Disorder
Teen Mental Health	How to Improve Mental Health	Personality Disorders	Autism Spectrum Disorder
Teen Mental Health	Older Adult Mental Health	How to Improve Mental Health	Older Adult Mental Health
Teen Sexual Health	Sexual Assault	Childbirth Problems	Pregnancy and Drug Use
Teen Sexual Health	Sexually Transmitted Diseases	Childbirth Problems	Premature Babies
Teenage Pregnancy	Infections and Pregnancy	Childbirth Problems	Twins, Triplets, Multiple Births
Teenage Pregnancy	Pregnancy and Medicines	Fetal Health and Development	Stillbirth
Teenage Pregnancy	Pregnancy and Drug Use	HIV/AIDS and Pregnancy	Pregnancy and Medicines
Teenage Pregnancy	Ectopic Pregnancy	HIV/AIDS and Pregnancy	Pregnancy and Drug Use
Teenage Pregnancy	Miscarriage	Infections and Pregnancy	Pregnancy and Drug Use
Teenage Pregnancy	Pregnancy and Nutrition	Infections and Pregnancy	Tumors and Pregnancy
Teenage Pregnancy	Preconception Care	Infections and Pregnancy	Ectopic Pregnancy
Underage Drinking	Alcohol Use Disorder (AUD) Treatment	Infections and Pregnancy	Miscarriage
Child Development	Child Mental Health	Infections and Pregnancy	Pregnancy and Nutrition
Child Development	Toddler Health	Infections and Pregnancy	Preconception Care
Child Development	Child Abuse	Pregnancy and Medicines	Tumors and Pregnancy
Child Mental Health	Toddler Health	Pregnancy and Medicines	Twins, Triplets, Multiple Births
Child Mental Health	Child Abuse	Pregnancy and Medicines	Ectopic Pregnancy
Drugs and Young People	Club Drugs	Pregnancy and Medicines	Miscarriage
Mental Disorders	How to Improve Mental Health	Pregnancy and Medicines	Pregnancy and Nutrition
Mental Disorders	Older Adult Mental Health	Pregnancy and Drug Use	Twins, Triplets, Multiple Births
Mental Disorders	Autism Spectrum Disorder	Pregnancy and Drug Use	Ectopic Pregnancy
Mental Health	Compulsive Gambling	Pregnancy and Drug Use	Miscarriage
Mental Health	Older Adult Health	Pregnancy and Drug Use	Pregnancy and Nutrition
Teen Depression	Postpartum Depression	Pregnancy and Drug Use	Preconception Care
Health Problems in Pregnancy	Ectopic Pregnancy	Premature Babies	Infant and Newborn Care

330

Health Problems in Pregnancy	Miscarriage	Tumors and Pregnancy	Pregnancy and Nutrition
Health Problems in Pregnancy	Pregnancy and Nutrition	Ectopic Pregnancy	Pregnancy and Nutrition
Health Problems in Pregnancy	Reproductive Hazards	Ectopic Pregnancy	Preconception Care
Health Problems in Pregnancy	Preconception Care	Miscarriage	Pregnancy and Nutrition
Health Problems in Pregnancy	Pregnancy and Opioids	Miscarriage	Preconception Care
Pregnancy	Twins, Triplets, Multiple Births	Pregnancy and Nutrition	Preconception Care
Pregnancy	Pregnancy and Opioids	Pregnancy and Nutrition	Pregnancy and Opioids
Child Safety	Toddler Health	Prenatal Testing	Hormones
Alcohol	Fetal Alcohol Spectrum Disorders	Infant and Newborn Care	Postpartum Care
Child Behavior Disorders	Toddler Health	Alcohol Use Disorder (AUD) Treatment	Fetal Alcohol Spectrum Disorders
Child Behavior Disorders	Child Abuse	Down Syndrome	Turner Syndrome
Developmental Disabilities	Infant and Newborn Development	Hormones	Pituitary Disorders
Developmental Disabilities	Toddler Development	Toddler Health	Child Abuse
Infant and Newborn Development	Fetal Health and Development	HIV: PrEP and PEP	Living with HIV/AIDS
Anxiety	Obsessive-Compulsive Disorder	Preconception Care	Pregnancy and Opioids
Bipolar Disorder	Obsessive-Compulsive Disorder		

Appendix J. Full Table 14.

Topic	Related topic	Topic	Related topic
Teen Mental Health	Teen Health	Pregnancy and Medicines	Health Problems in Pregnancy
Teenage Pregnancy	Teen Health	Pregnancy and Medicines	Pregnancy
Child Mental Health	Teen Mental Health	Pregnancy and Medicines	Infections and Pregnancy
Drugs and Young People	Drug Use and Addiction	Pregnancy and Drug Use	Pregnancy
Self-harm	Mental Disorders	Pregnancy and Drug Use	Pregnancy and Medicines
Health Problems in Pregnancy	Teenage Pregnancy	Pregnancy and Drug Use	Reproductive Hazards
Pregnancy	Preconception Care	Tumors and Pregnancy	Health Problems in Pregnancy
Prenatal Care	Preconception Care	Tumors and Pregnancy	Pregnancy
Alcohol	Alcohol Use Disorder (AUD) Treatment	Ectopic Pregnancy	Pregnancy
Child Behavior Disorders	Autism Spectrum Disorder	Miscarriage	Pregnancy
Developmental Disabilities	Child Development	Pregnancy and Nutrition	Pregnancy
Growth Disorders	Child Development	Motor Vehicle Safety	Child Safety
Infant and Newborn Development	Infant and Newborn Care	Learning Disabilities	Disabilities
Bipolar Disorder	Mental Disorders	Pituitary Disorders	Growth Disorders
Bipolar Disorder	Mood Disorders	Pituitary Disorders	Endocrine Diseases
Compulsive Gambling	Mental Disorders	Quitting Smoking	Smoking and Youth
Eating Disorders	Mental Disorders	Smokeless Tobacco	Smoking and Youth
Mood Disorders	Mental Disorders	Postpartum Depression	Postpartum Care
Obsessive-Compulsive Disorder	Mental Disorders	Preterm Labor	Childbirth Problems
Obsessive-Compulsive Disorder	Personality Disorders	Preterm Labor	Childbirth
Schizophrenia	Psychotic Disorders	HIV/AIDS and Infections	HIV/AIDS and Pregnancy
Childbirth Problems	Health Problems in Pregnancy	HIV/AIDS Medicines	HIV/AIDS and Pregnancy
HIV/AIDS and Pregnancy	Health Problems in Pregnancy	HIV/AIDS Medicines	HIV/AIDS and Infections
HIV/AIDS and Pregnancy	Pregnancy	HIV/AIDS Medicines	Living with HIV/AIDS
HIV/AIDS and Pregnancy	Infections and Pregnancy	Preconception Care	Pregnancy and Medicines
Infections and Pregnancy	Health Problems in Pregnancy		

Appendix K. Full Table 15

Topic A	Topic B	Topic A	Topic B
Older Adult Health	Mental Health	Child Nutrition	Child Behavior Disorders
End of Life Issues	Palliative Care	Dietary Fiber	Diets
End of Life Issues	Managed Care	Diets	Body Weight
Exercise for Older Adults	Benefits of Exercise	Diets	Dietary Supplements
Exercise for Older Adults	Exercise for Children	Eating Disorders	Mood Disorders
Exercise for Older Adults	Health Risks of an Inactive Lifestyle	Eating Disorders	Carbohydrate Metabolism Disorders
Healthy Aging	Older Adult Mental Health	Eating Disorders	Obsessive-Compulsive Disorder
Healthy Aging	Financial Assistance	Eating Disorders	Personality Disorders
Healthy Aging	Health Insurance	Vitamins	Vitamin D Deficiency
Healthy Aging	Nutrition	Child Mental Health	Child Development
Healthy Aging	Mental Health	Child Mental Health	Children's Health
Healthy Aging	How to Improve Mental Health	How to Improve Mental Health	Mental Disorders
Healthy Aging	Teen Mental Health	How to Improve Mental Health	Teen Mental Health
Home Care Services	Nursing Homes	Teen Mental Health	Bipolar Disorder
Home Care Services	Managed Care	Teen Mental Health	Compulsive Gambling
Nutrition for Older Adults	Older Adult Mental Health	Bipolar Disorder	Obsessive-Compulsive Disorder
Nutrition for Older Adults	Mental Health	Mood Disorders	Obsessive-Compulsive Disorder
Nutrition for Older Adults	Child Nutrition	Mood Disorders	Panic Disorder
Nutrition for Older Adults	How to Improve Mental Health	Mood Disorders	Personality Disorders
Older Adult Mental Health	Child Mental Health	Postpartum Depression	Teen Depression
Older Adult Mental Health	How to Improve Mental Health	Children's Health	Obesity in Children
Older Adult Mental Health	Mental Disorders	Children's Health	Child Behavior Disorders
Older Adult Mental Health	Teen Mental Health	Cholesterol Medicines	How to Lower Cholesterol with Diet
Hospice Care	Managed Care	Heart Diseases	Creutzfeldt-Jakob Disease
Child Care	Child Nutrition	How to Lower Cholesterol	VLDL Cholesterol
Child Care	Child Mental Health	How to Lower Cholesterol with Diet	VLDL Cholesterol
Child Care	Child Development	How to Lower Cholesterol with Diet	DASH Eating Plan
Child Care	Children's Health	Vitamin D Deficiency	B Vitamins
Child Care	Child Behavior Disorders	Vitamin D Deficiency	Vitamin A
Telehealth	Managed Care	Vitamin D Deficiency	Vitamin C
Financial Assistance	Health Facilities	Vitamin D Deficiency	Vitamin E
Financial Assistance	Managed Care	Vitamin D Deficiency	Vitamin K
Health Insurance	Health Facilities	Antioxidants	Vitamin D

333

Health Insurance	Mental Health	B Vitamins	Vitamin A
Health Facilities	Managed Care	B Vitamins	Vitamin C
Nutrition	Toddler Nutrition	B Vitamins	Vitamin D
Mental Health	Compulsive Gambling	B Vitamins	Vitamin E
Parenting	Child Nutrition	B Vitamins	Vitamin K
Parenting	Child Mental Health	Vitamin A	Vitamin C
Parenting	Children's Health	Vitamin A	Vitamin D
Parenting	Child Behavior Disorders	Vitamin A	Vitamin E
Palliative Care	Managed Care	Vitamin A	Vitamin K
Alzheimer's Disease	Creutzfeldt-Jakob Disease	Vitamin C	Vitamin D
Benefits of Exercise	Health Risks of an Inactive Lifestyle	Vitamin C	Vitamin E
Benefits of Exercise	How Much Exercise Do I Need?	Vitamin C	Vitamin K
Exercise for Children	Health Risks of an Inactive Lifestyle	Vitamin D	Vitamin E
Health Risks of an Inactive Lifestyle	How Much Exercise Do I Need?	Vitamin D	Vitamin K
Child Nutrition	Child Mental Health	Vitamin E	Vitamin K
Child Nutrition	Child Development	Anxiety	Obsessive-Compulsive Disorder

334

Appendix L. Full Table 16

Topic	Related topic	Topic	Related topic
Exercise for Older Adults	Healthy Aging	Sprains and Strains	Ankle Injuries and Disorders
Exercise for Older Adults	How Much Exercise Do I Need?	Children's Health	Child Nutrition
Home Care Services	Hospice Care	Cholesterol Medicines	HDL: The "Good" Cholesterol
Nutrition for Older Adults	Healthy Aging	Cholesterol Medicines	LDL: The "Bad" Cholesterol
Child Care	Caregivers	Cholesterol Medicines	VLDL Cholesterol
Benefits of Exercise	Exercise and Physical Fitness	HDL: The "Good" Cholesterol	How to Lower Cholesterol with Diet
Benefits of Exercise	Exercise for Children	HDL: The "Good" Cholesterol	VLDL Cholesterol
Exercise for Children	How Much Exercise Do I Need?	How to Lower Cholesterol	HDL: The "Good" Cholesterol
Health Risks of an Inactive Lifestyle	Exercise and Physical Fitness	How to Lower Cholesterol	LDL: The "Bad" Cholesterol
How Much Exercise Do I Need?	Exercise and Physical Fitness	How to Lower Cholesterol with Diet	Cholesterol Levels: What You Need to Know
Sports Injuries	Sports Safety	VLDL Cholesterol	Cholesterol Levels: What You Need to Know
Managed Care	Health Insurance	DASH Eating Plan	Diets
Child Nutrition	Nutrition	Vitamin D Deficiency	Vitamin D
Child Nutrition	Obesity in Children	Vitamin E	Vitamins
Dietary Fats	How to Lower Cholesterol with Diet	Vitamin E	Antioxidants
Eating Disorders	Mental Disorders	Vitamin K	Vitamins
Food Safety	Food Labeling	Anxiety	Mental Disorders
Weight Control	Obesity in Children	Compulsive Gambling	Mental Disorders
Child Mental Health	Teen Mental Health	Obsessive-Compulsive Disorder	Mental Disorders
Child Development	Parenting	Obsessive-Compulsive Disorder	Personality Disorders
Child Development	Children's Health	Panic Disorder	Mental Disorders
Bipolar Disorder	Mental Disorders	Schizophrenia	Psychotic Disorders
Bipolar Disorder	Mood Disorders	Self-Harm	Mental Disorders
Mood Disorders	Mental Disorders	Teen Violence	Teen Mental Health

Appendix M. Full list of recommended connections about the five core topics relating to the mental health subcategory (format: topic, related topic, similarity value)

Obesity In Children		
Obesity In Children	*Child Mental Health*	*0.451*
Obesity In Children	*Child Behavior Disorders*	*0.422*
Obesity In Children	*Child Development*	*0.504*
Obesity In Children	*Teen Health*	*0.389*
Obesity In Children	*Exercise For Children*	*0.397*
Obesity In Children	*Child Sexual Abuse*	*0.39*
Child Behavior Disorders	*Obesity In Children*	*0.422*
Child Development	*Obesity In Children*	*0.504*
Child Mental Health	*Obesity In Children*	*0.451*
Child Nutrition	*Obesity In Children*	*0.568*
Child Sexual Abuse	*Obesity In Children*	*0.39*
Exercise For Children	*Obesity In Children*	*0.397*
Teen Health	*Obesity In Children*	*0.389*
Weight Control	*Obesity In Children*	*0.43*
NutritionForSeniors		
Nutrition For Seniors	*Weight Control*	*0.39*
Nutrition For Seniors	*Diets*	*0.444*
Nutrition For Seniors	*College Health*	*0.456*
Nutrition For Seniors	*Child Nutrition*	*0.545*
Nutrition For Seniors	*Food Labeling*	*0.491*
Nutrition For Seniors	*Food Safety*	*0.455*
Child Nutrition	*Nutrition For Seniors*	*0.545*
College Health	*Nutrition For Seniors*	*0.456*
Food Labeling	*Nutrition For Seniors*	*0.491*
Food Safety	*Nutrition For Seniors*	*0.455*
Weight Control	*Nutrition For Seniors*	*0.39*
ChildNutrition		
Child Nutrition	*Child Mental Health*	*0.59*
Child Nutrition	*Child Behavior Disorders*	*0.565*
Child Nutrition	*School Health*	*0.463*
Child Nutrition	*Child Development*	*0.649*
Child Nutrition	*Diets*	*0.394*
Child Nutrition	*Nutrition*	*0.6*
Child Nutrition	*Teen Health*	*0.455*
Child Nutrition	*College Health*	*0.446*
Child Nutrition	*Obesity In Children*	*0.568*
Child Nutrition	*Nutrition For Seniors*	*0.545*
Child Nutrition	*Exercise For Children*	*0.465*
Child Nutrition	*Food Labeling*	*0.407*
Child Nutrition	*Child Abuse*	*0.434*
Child Nutrition	*Child Sexual Abuse*	*0.471*

336

Child Abuse	*Child Nutrition*	*0.434*
Child Behavior Disorders	*Child Nutrition*	*0.565*
Child Development	*Child Nutrition*	*0.649*
Child Mental Health	*Child Nutrition*	*0.59*
Child Sexual Abuse	*Child Nutrition*	*0.471*
College Health	*Child Nutrition*	*0.446*
Exercise For Children	*Child Nutrition*	*0.465*
Food Labeling	*Child Nutrition*	*0.407*
Nutrition For Seniors	*Child Nutrition*	*0.545*
School Health	*Child Nutrition*	*0.463*
Teen Health	*Child Nutrition*	*0.455*
ChildDevelopment		
Child Development	*Child Mental Health*	*0.735*
Child Development	*Child Safety*	*0.431*
Child Development	*Learning Disorders*	*0.436*
Child Development	*Obesity In Children*	*0.504*
Child Development	*Child Nutrition*	*0.649*
Child Development	*Exercise For Children*	*0.43*
Child Development	*Child Abuse*	*0.531*
Child Development	*Child Sexual Abuse*	*0.575*
Child Abuse	*Child Development*	*0.531*
Child Mental Health	*Child Development*	*0.735*
Child Nutrition	*Child Development*	*0.649*
Child Safety	*Child Development*	*0.431*
Child Sexual Abuse	*Child Development*	*0.575*
Exercise For Children	*Child Development*	*0.43*
Growth Disorders	*Child Development*	*0.442*
Learning Disorders	*Child Development*	*0.436*
Obesity In Children	*Child Development*	*0.504*
Diets		
Diets	*Weight Loss Surgery*	*0.506*
Diets	*Dietary Fiber*	*0.497*
Child Nutrition	*Diets*	*0.394*
Dash Diet	*Diets*	*0.503*
Dietary Fiber	*Diets*	*0.497*
Nutrition For Seniors	*Diets*	*0.444*
Vegetarian Diet	*Diets*	*0.404*
Weight Loss Surgery	*Diets*	*0.506*

337

Appendix N. Full list of recommended connections about the five core topics relating to the children subcategory (format: topic, related topic, similarity value)

Medicines and Children		
Medicines and Children	*Children's Health*	*0.628*
Medicines and Children	*Child Development*	*0.458*
Medicines and Children	*Child Safety*	*0.548*
Medicines and Children	*Toddler Health*	*0.591*
Medicines and Children	*Child Behavior Disorders*	*0.476*
Medicines and Children	*Child Mental Health*	*0.526*
Medicines and Children	*Child Nutrition*	*0.529*
Medicines and Children	*Medication Errors*	*0.566*
Child Development	*Medicines and Children*	*0.458*
Child Safety	*Medicines and Children*	*0.548*
Toddler Health	*Medicines and Children*	*0.591*
Child Behavior Disorders	*Medicines and Children*	*0.476*
Child Mental Health	*Medicines and Children*	*0.526*
Child Nutrition	*Medicines and Children*	*0.529*
Children's Health		
Children's Health	*Child Behavior Disorders*	*0.689*
Children's Health	*Toddler Development*	*0.489*
Children's Health	*Baby Health Checkup*	*0.469*
Children's Health	*Child Mental Health*	*0.747*
Children's Health	*Child Nutrition*	*0.735*
Child Dental Health	*Children's Health*	*0.58*
Child Development	*Children's Health*	*0.714*
Child Safety	*Children's Health*	*0.487*
Common Infant and Newborn Problems	*Children's Health*	*0.544*
Medicines and Children	*Children's Health*	*0.628*
Toddler Health	*Children's Health*	*0.889*
Uncommon Infant and Newborn Problems	*Children's Health*	*0.514*
Child Behavior Disorders	*Children's Health*	*0.689*
Toddler Development	*Children's Health*	*0.489*
Baby Health Checkup	*Children's Health*	*0.469*
Child Mental Health	*Children's Health*	*0.747*
Toddler Health		
Toddler Health	*Children's Health*	*0.889*
Toddler Health	*Child Dental Health*	*0.527*
Toddler Health	*Child Development*	*0.661*
Toddler Health	*Common Infant and Newborn Problems*	*0.593*
Toddler Health	*Medicines and Children*	*0.591*

338

Toddler Health	Uncommon Infant and Newborn Problems	0.596
Toddler Health	Child Behavior Disorders	0.599
Toddler Health	Baby Health Checkup	0.512
Toddler Health	Child Mental Health	0.659
Toddler Health	Child Nutrition	0.642
Child Dental Health	Toddler Health	0.527
Child Development	Toddler Health	0.661
Common Infant and Newborn Problems	Toddler Health	0.593
Medicines and Children	Toddler Health	0.591
Uncommon Infant and Newborn Problems	Toddler Health	0.596
Child Behavior Disorders	Toddler Health	0.599
Baby Health Checkup	Toddler Health	0.512
Toddler Nutrition	Toddler Health	0.489
Child Mental Health	Toddler Health	0.659
Child Nutrition	Toddler Health	0.642
Child Development		
Child Development	Children's Health	0.714
Child Development	Medicines and Children	0.458
Child Development	Toddler Health	0.661
Child Development	Child Mental Health	0.576
Child Development	Child Nutrition	0.555
Medicines and Children	Child Development	0.458
Toddler Health	Child Development	0.661
Developmental Disabilities	Child Development	0.52
Growth Disorders	Child Development	0.472
Child Mental Health	Child Development	0.576
Child Nutrition	Child Development	0.555
Child Mental Health		
Child Mental Health	Children's Health	0.747
Child Mental Health	Child Dental Health	0.478
Child Mental Health	Child Development	0.576
Child Mental Health	Medicines and Children	0.526
Child Mental Health	Toddler Health	0.659
Child Mental Health	Teen Mental Health	0.637
Child Mental Health	Child Nutrition	0.623
Children's Health	Child Mental Health	0.747
Child Dental Health	Child Mental Health	0.478
Child Development	Child Mental Health	0.576
Medicines and Children	Child Mental Health	0.526
Toddler Health	Child Mental Health	0.659
Child Nutrition	Child Mental Health	0.623

339

Appendix O. Full list of recommended connections about the five core topics relating to the teenagers subcategory (format: topic, related topic, similarity value)

TeenagePregnancy		
Teenage Pregnancy	*Teen Health*	*0.519*
Teenage Pregnancy	*Teen Development*	*0.514*
Teenage Pregnancy	*Teen Mental Health*	*0.462*
Teenage Pregnancy	*Infections and Pregnancy*	*0.648*
Teenage Pregnancy	*Pregnancy and Medicines*	*0.696*
Teenage Pregnancy	*Pregnancy and Drug Use*	*0.638*
Teenage Pregnancy	*Ectopic Pregnancy*	*0.578*
Teenage Pregnancy	*Miscarriage*	*0.484*
Teenage Pregnancy	*Pregnancy and Nutrition*	*0.639*
Teenage Pregnancy	*Preconception Care*	*0.563*
Teen Development	*Teenage Pregnancy*	*0.514*
Teen Mental Health	*Teenage Pregnancy*	*0.462*
Health Problems in Pregnancy	*Teenage Pregnancy*	*0.696*
Infections and Pregnancy	*Teenage Pregnancy*	*0.648*
Pregnancy and Medicines	*Teenage Pregnancy*	*0.696*
Pregnancy and Drug Use	*Teenage Pregnancy*	*0.638*
Ectopic Pregnancy	*Teenage Pregnancy*	*0.578*
Miscarriage	*Teenage Pregnancy*	*0.484*
Pregnancy and Nutrition	*Teenage Pregnancy*	*0.639*
Preconception Care	*Teenage Pregnancy*	*0.563*
InfectionsandPregnancy		
Infections and Pregnancy	*Teenage Pregnancy*	*0.648*
Infections and Pregnancy	*Health Problems in Pregnancy*	*0.814*
Infections and Pregnancy	*Pregnancy and Drug Use*	*0.66*
Infections and Pregnancy	*Tumors and Pregnancy*	*0.49*
Infections and Pregnancy	*Ectopic Pregnancy*	*0.657*
Infections and Pregnancy	*Miscarriage*	*0.574*
Infections and Pregnancy	*Pregnancy and Nutrition*	*0.73*
Infections and Pregnancy	*Preconception Care*	*0.613*
Teenage Pregnancy	*Infections and Pregnancy*	*0.648*
HIV/AIDS and Pregnancy	*Infections and Pregnancy*	*0.496*
Pregnancy and Medicines	*Infections and Pregnancy*	*0.82*
Pregnancy and Drug Use	*Infections and Pregnancy*	*0.66*
Tumors and Pregnancy	*Infections and Pregnancy*	*0.49*
Ectopic Pregnancy	*Infections and Pregnancy*	*0.657*
Miscarriage	*Infections and Pregnancy*	*0.574*
Pregnancy and Nutrition	*Infections and Pregnancy*	*0.73*
Preconception Care	*Infections and Pregnancy*	*0.613*
PregnancyandDrugUse		
Pregnancy and Drug Use	*Teenage Pregnancy*	*0.638*
Pregnancy and Drug Use	*Pregnancy*	*0.737*

340

Pregnancy and Drug Use	Childbirth Problems	0.487
Pregnancy and Drug Use	HIV/AIDS and Pregnancy	0.484
Pregnancy and Drug Use	Infections and Pregnancy	0.66
Pregnancy and Drug Use	Pregnancy and Medicines	0.737
Pregnancy and Drug Use	Twins, Triplets, Multiple Births	0.497
Pregnancy and Drug Use	Ectopic Pregnancy	0.565
Pregnancy and Drug Use	Miscarriage	0.534
Pregnancy and Drug Use	Pregnancy and Nutrition	0.675
Pregnancy and Drug Use	Reproductive Hazards	0.494
Pregnancy and Drug Use	Preconception Care	0.618
Teenage Pregnancy	Pregnancy and Drug Use	0.638
Childbirth Problems	Pregnancy and Drug Use	0.487
HIV/AIDS and Pregnancy	Pregnancy and Drug Use	0.484
Infections and Pregnancy	Pregnancy and Drug Use	0.66
Twins, Triplets, Multiple Births	Pregnancy and Drug Use	0.497
Ectopic Pregnancy	Pregnancy and Drug Use	0.565
Miscarriage	Pregnancy and Drug Use	0.534
Pregnancy and Nutrition	Pregnancy and Drug Use	0.675
Preconception Care	Pregnancy and Drug Use	0.618
PregnancyandMedicines		
Pregnancy and Medicines	Teenage Pregnancy	0.696
Pregnancy and Medicines	Health Problems in Pregnancy	0.899
Pregnancy and Medicines	Pregnancy	0.903
Pregnancy and Medicines	HIV/AIDS and Pregnancy	0.52
Pregnancy and Medicines	Infections and Pregnancy	0.82
Pregnancy and Medicines	Tumors and Pregnancy	0.547
Pregnancy and Medicines	Twins, Triplets, Multiple Births	0.469
Pregnancy and Medicines	Ectopic Pregnancy	0.74
Pregnancy and Medicines	Miscarriage	0.619
Pregnancy and Medicines	Pregnancy and Nutrition	0.831
Teenage Pregnancy	Pregnancy and Medicines	0.696
HIV/AIDS and Pregnancy	Pregnancy and Medicines	0.52
Pregnancy and Drug Use	Pregnancy and Medicines	0.737
Tumors and Pregnancy	Pregnancy and Medicines	0.547
Twins, Triplets, Multiple Births	Pregnancy and Medicines	0.469
Ectopic Pregnancy	Pregnancy and Medicines	0.74
Miscarriage	Pregnancy and Medicines	0.619
Pregnancy and Nutrition	Pregnancy and Medicines	0.831
Preconception Care	Pregnancy and Medicines	0.666
HealthProblemsinPregnancy		
Health Problems in Pregnancy	Teenage Pregnancy	0.696
Health Problems in Pregnancy	Ectopic Pregnancy	0.728
Health Problems in Pregnancy	Miscarriage	0.663
Health Problems in Pregnancy	Pregnancy and Nutrition	0.807
Health Problems in Pregnancy	Reproductive Hazards	0.502
Health Problems in Pregnancy	Preconception Care	0.701
Health Problems in Pregnancy	Pregnancy and Opioids	0.531

341

Childbirth Problems	Health Problems in Pregnancy	0.484
HIV/AIDS and Pregnancy	Health Problems in Pregnancy	0.497
Infections and Pregnancy	Health Problems in Pregnancy	0.814
Pregnancy and Medicines	Health Problems in Pregnancy	0.899
Tumors and Pregnancy	Health Problems in Pregnancy	0.565
Ectopic Pregnancy	Health Problems in Pregnancy	0.728
Miscarriage	Health Problems in Pregnancy	0.663
Pregnancy and Nutrition	Health Problems in Pregnancy	0.807
Reproductive Hazards	Health Problems in Pregnancy	0.502
Preconception Care	Health Problems in Pregnancy	0.701
Pregnancy and Opioids	Health Problems in Pregnancy	0.531

Appendix P. Full list of recommended connections about the five core topics relating to the older adults subcategory (format: topic, related topic, similarity value)

ChildNutrition		
Child Nutrition	Nutrition for Older Adults	0.49
Child Nutrition	Child Care	0.562
Child Nutrition	Nutrition	0.57
Child Nutrition	Parenting	0.522
Child Nutrition	Child Mental Health	0.626
Child Nutrition	Child Development	0.561
Child Nutrition	Obesity in Children	0.559
Child Nutrition	Child Behavior Disorders	0.573
Nutrition for Older Adults	Child Nutrition	0.49
Child Care	Child Nutrition	0.562
Parenting	Child Nutrition	0.522
Child Mental Health	Child Nutrition	0.626
Child Development	Child Nutrition	0.561
Children's Health	Child Nutrition	0.743
Child Behavior Disorders	Child Nutrition	0.573
HealthyAging		
Healthy Aging	Older Adult Mental Health	0.559
Healthy Aging	Financial Assistance	0.489
Healthy Aging	Health Insurance	0.564
Healthy Aging	Nutrition	0.494
Healthy Aging	Mental Health	0.64
Healthy Aging	How to Improve Mental Health	0.594
Healthy Aging	Teen Mental Health	0.47
Exercise for Older Adults	Healthy Aging	0.445
Nutrition for Older Adults	Healthy Aging	0.781
Older Adult Mental Health	Healthy Aging	0.559
Financial Assistance	Healthy Aging	0.489
Health Insurance	Healthy Aging	0.564
Nutrition	Healthy Aging	0.494
Mental Health	Healthy Aging	0.64
How to Improve Mental Health	Healthy Aging	0.594
Teen Mental Health	Healthy Aging	0.47
TeenMentalHealth		
Teen Mental Health	Healthy Aging	0.47
Teen Mental Health	Older Adult Mental Health	0.539
Teen Mental Health	How to Improve Mental Health	0.593
Teen Mental Health	Bipolar Disorder	0.451
Teen Mental Health	Compulsive Gambling	0.505

343

Healthy Aging	*Teen Mental Health*	*0.47*
Older Adult Mental Health	*Teen Mental Health*	*0.539*
Child Mental Health	*Teen Mental Health*	*0.634*
How to Improve Mental Health	*Teen Mental Health*	*0.593*
Bipolar Disorder	*Teen Mental Health*	*0.451*
Compulsive Gambling	*Teen Mental Health*	*0.505*
Teen Violence	*Teen Mental Health*	*0.444*
ChildMentalHealth		
Child Mental Health	*Older Adult Mental Health*	*0.445*
Child Mental Health	*Child Care*	*0.554*
Child Mental Health	*Parenting*	*0.487*
Child Mental Health	*Child Nutrition*	*0.626*
Child Mental Health	*Teen Mental Health*	*0.634*
Child Mental Health	*Child Development*	*0.577*
Child Mental Health	*Children's Health*	*0.751*
Older Adult Mental Health	*Child Mental Health*	*0.445*
Child Care	*Child Mental Health*	*0.554*
Parenting	*Child Mental Health*	*0.487*
Child Nutrition	*Child Mental Health*	*0.626*
Child Development	*Child Mental Health*	*0.577*
Children's Health	*Child Mental Health*	*0.751*
MentalHealth		
Mental Health	*Older Adult Health*	*0.485*
Mental Health	*Healthy Aging*	*0.64*
Mental Health	*Nutrition for Older Adults*	*0.482*
Mental Health	*Health Insurance*	*0.46*
Mental Health	*Compulsive Gambling*	*0.494*
Older Adult Health	*Mental Health*	*0.485*
Healthy Aging	*Mental Health*	*0.64*
Nutrition for Older Adults	*Mental Health*	*0.482*
Health Insurance	*Mental Health*	*0.46*
Compulsive Gambling	*Mental Health*	*0.494*

344

Appendix Q. Complete lists of recommended connections generated through the Euclidean distance similarity value toward mental health related topics (format: topic, related topic, similarity value).

Topic	Related topic	Similarity Value
Mental Disorders	Veterans And Military Family Health	0.041884817
Mental Disorders	Inhalants	0.042144302
Bipolar Disorder	Mood Disorders	0.042448425
Compulsive Gambling	Mental Disorders	0.051232133
Compulsive Gambling	Delirium	0.058024835
Compulsive Gambling	Eating Disorders	0.054392168
Compulsive Gambling	Mood Disorders	0.057352604
Compulsive Gambling	Obsessive-Compulsive Disorder	0.04494584
Compulsive Gambling	Panic Disorder	0.050251256
Compulsive Gambling	Personality Disorders	0.074537865
Compulsive Gambling	Phobias	0.045407074
Compulsive Gambling	Psychotic Disorders	0.046127589
Compulsive Gambling	Schizophrenia	0.045930553
Compulsive Gambling	Self-Harm	0.047890427
Compulsive Gambling	Child Behavior Disorders	0.041523066
Compulsive Gambling	Postpartum Depression	0.057448153
Compulsive Gambling	Seasonal Affective Disorder	0.043979242
Compulsive Gambling	Drugs And Young People	0.041169205
Compulsive Gambling	Underage Drinking	0.041135335
Compulsive Gambling	Teen Development	0.043936731
Compulsive Gambling	Weight Loss Surgery	0.062138818
Compulsive Gambling	Veterans And Military Family Health	0.064151912
Compulsive Gambling	Cold And Cough Medicines	0.044367541
Compulsive Gambling	Inhalants	0.065372295
Compulsive Gambling	Smoking And Youth	0.041995632
Compulsive Gambling	Domestic Violence	0.051639556
Compulsive Gambling	College Health	0.052342319
Compulsive Gambling	Malnutrition	0.058421452
Compulsive Gambling	Nutrition For Seniors	0.041593877
Compulsive Gambling	Sports Fitness	0.047619048
Compulsive Gambling	Nutritional Support	0.045314483
Compulsive Gambling	Common Cold	0.044281096
Compulsive Gambling	Club Drugs	0.042718612
Compulsive Gambling	Child Sexual Abuse	0.042220815
Compulsive Gambling	Elder Abuse	0.05220024
Compulsive Gambling	Sexual Assault	0.04381161
Compulsive Gambling	Cirrhosis	0.053225463
Delirium	Compulsive Gambling	0.058024835
Delirium	Mood Disorders	0.042914771
Delirium	Personality Disorders	0.04805613
Delirium	Postpartum Depression	0.042562247
Delirium	Weight Loss Surgery	0.04454343

345

Delirium	Veterans And Military Family Health	0.043601482
Delirium	Inhalants	0.045787546
Delirium	Malnutrition	0.044367541
Delirium	Cirrhosis	0.041450777
Eating Disorders	Compulsive Gambling	0.054392168
Eating Disorders	Mood Disorders	0.041739711
Eating Disorders	Personality Disorders	0.0464231
Eating Disorders	Postpartum Depression	0.041064389
Eating Disorders	Weight Loss Surgery	0.04415011
Eating Disorders	Veterans And Military Family Health	0.042295817
Eating Disorders	Inhalants	0.043851956
Eating Disorders	College Health	0.042914771
Eating Disorders	Malnutrition	0.042371086
Mood Disorders	Mental Disorders	0.042756969
Mood Disorders	Compulsive Gambling	0.057352604
Mood Disorders	Delirium	0.042914771
Mood Disorders	Eating Disorders	0.041739711
Mood Disorders	Panic Disorder	0.041169205
Mood Disorders	Personality Disorders	0.0483372
Mood Disorders	Self-Harm	0.040893105
Mood Disorders	Postpartum Depression	0.044676764
Mood Disorders	Weight Loss Surgery	0.04512839
Mood Disorders	Veterans And Military Family Health	0.044324276
Mood Disorders	Inhalants	0.045549786
Mood Disorders	College Health	0.040859688
Mood Disorders	Malnutrition	0.04299411
Mood Disorders	Elder Abuse	0.042295817
Mood Disorders	Cirrhosis	0.041703157
Obsessive-Compulsive Disorder	Compulsive Gambling	0.04494584
Obsessive-Compulsive Disorder	Personality Disorders	0.040791352
Panic Disorder	Compulsive Gambling	0.050251256
Panic Disorder	Mood Disorders	0.041169205
Panic Disorder	Personality Disorders	0.043192813
Panic Disorder	Weight Loss Surgery	0.040859688
Panic Disorder	Inhalants	0.041958629
Personality Disorders	Compulsive Gambling	0.074537865
Personality Disorders	Delirium	0.04805613
Personality Disorders	Eating Disorders	0.0464231
Personality Disorders	Mood Disorders	0.0483372
Personality Disorders	Panic Disorder	0.043192813
Personality Disorders	Psychotic Disorders	0.041238814
Personality Disorders	Schizophrenia	0.041450777
Personality Disorders	Self-Harm	0.042107036
Personality Disorders	Postpartum Depression	0.04708763
Personality Disorders	Weight Loss Surgery	0.049446203
Personality Disorders	Veterans And Military Family Health	0.048393341
Personality Disorders	Inhalants	0.051435038
Personality Disorders	Domestic Violence	0.043233895
Personality Disorders	College Health	0.04512839
Personality Disorders	Malnutrition	0.04805613
Personality Disorders	Elder Abuse	0.04406451
Personality Disorders	Cirrhosis	0.045979125

346

Phobias	Compulsive Gambling	0.045407074
Psychotic Disorders	Compulsive Gambling	0.046127589
Psychotic Disorders	Personality Disorders	0.041238814
Schizophrenia	Compulsive Gambling	0.045930553
Schizophrenia	Personality Disorders	0.041450777
Self-Harm	Compulsive Gambling	0.047890427
Self-Harm	Mood Disorders	0.040893105
Self-Harm	Personality Disorders	0.042107036
Child Behavior Disorders	Compulsive Gambling	0.041523066
Child Behavior Disorders	Child Sexual Abuse	0.04503693
Postpartum Depression	Compulsive Gambling	0.057448153
Postpartum Depression	Delirium	0.042562247
Postpartum Depression	Eating Disorders	0.041064389
Postpartum Depression	Mood Disorders	0.044676764
Postpartum Depression	Personality Disorders	0.04708763
Postpartum Depression	Weight Loss Surgery	0.043519889
Postpartum Depression	Veterans And Military Family Health	0.043355734
Postpartum Depression	Inhalants	0.04503693
Postpartum Depression	College Health	0.040893105
Postpartum Depression	Malnutrition	0.042874293
Seasonal Affective Disorder	Compulsive Gambling	0.043979242
Drugs And Young People	Compulsive Gambling	0.041169205
Underage Drinking	Compulsive Gambling	0.041135335
Teen Development	Compulsive Gambling	0.043936731
Weight Loss Surgery	Compulsive Gambling	0.062138818
Weight Loss Surgery	Delirium	0.04454343
Weight Loss Surgery	Eating Disorders	0.04415011
Weight Loss Surgery	Mood Disorders	0.04512839
Weight Loss Surgery	Panic Disorder	0.040859688
Weight Loss Surgery	Personality Disorders	0.049446203
Weight Loss Surgery	Postpartum Depression	0.043519889
Weight Loss Surgery	Veterans And Military Family Health	0.045930553
Weight Loss Surgery	Inhalants	0.04828119
Weight Loss Surgery	Domestic Violence	0.042107036
Weight Loss Surgery	College Health	0.044107269
Weight Loss Surgery	Malnutrition	0.047782875
Weight Loss Surgery	Elder Abuse	0.042333418
Weight Loss Surgery	Cirrhosis	0.045549786
Veterans And Military Family Health	Mental Disorders	0.041884817
Veterans And Military Family Health	Compulsive Gambling	0.064151912
Veterans And Military Family Health	Delirium	0.043601482
Veterans And Military Family Health	Eating Disorders	0.042295817
Veterans And Military Family Health	Mood Disorders	0.044324276
Veterans And Military Family Health	Personality Disorders	0.048393341
Veterans And Military Family Health	Postpartum Depression	0.043355734
Veterans And Military Family Health	Veterans And Military Health	0.040995368

347

Veterans And Military Family Health	Weight Loss Surgery	0.045930553
Veterans And Military Family Health	Inhalants	0.049086982
Veterans And Military Family Health	Domestic Violence	0.04311274
Veterans And Military Family Health	College Health	0.044367541
Veterans And Military Family Health	Malnutrition	0.044992351
Veterans And Military Family Health	Elder Abuse	0.043519889
Veterans And Military Family Health	Cirrhosis	0.041958629
Cold And Cough Medicines	Compulsive Gambling	0.044367541
Inhalants	Mental Disorders	0.042144302
Inhalants	Compulsive Gambling	0.065372295
Inhalants	Delirium	0.045787546
Inhalants	Eating Disorders	0.043851956
Inhalants	Mood Disorders	0.045549786
Inhalants	Panic Disorder	0.041958629
Inhalants	Personality Disorders	0.051435038
Inhalants	Postpartum Depression	0.04503693
Inhalants	Weight Loss Surgery	0.04828119
Inhalants	Veterans And Military Family Health	0.049086982
Inhalants	Domestic Violence	0.043979242
Inhalants	College Health	0.045407074
Inhalants	Malnutrition	0.04698365
Inhalants	Sports Fitness	0.041559305
Inhalants	Elder Abuse	0.04415011
Inhalants	Cirrhosis	0.04485713
Smoking And Youth	Compulsive Gambling	0.041995632
Domestic Violence	Compulsive Gambling	0.051639556
Domestic Violence	Personality Disorders	0.043233895
Domestic Violence	Weight Loss Surgery	0.042107036
Domestic Violence	Veterans And Military Family Health	0.04311274
Domestic Violence	Inhalants	0.043979242
Domestic Violence	College Health	0.042562247
College Health	Compulsive Gambling	0.052342319
College Health	Eating Disorders	0.042914771
College Health	Mood Disorders	0.040859688
College Health	Personality Disorders	0.04512839
College Health	Postpartum Depression	0.040893105
College Health	Weight Loss Surgery	0.044107269
College Health	Veterans And Military Family Health	0.044367541
College Health	Inhalants	0.045407074
College Health	Domestic Violence	0.042562247
College Health	Malnutrition	0.042874293
College Health	Nutrition For Seniors	0.044676764
College Health	Sports Fitness	0.041238814
Malnutrition	Compulsive Gambling	0.058421452
Malnutrition	Delirium	0.044367541
Malnutrition	Eating Disorders	0.042371086

348

Malnutrition	Mood Disorders	0.04299411
Malnutrition	Personality Disorders	0.04805613
Malnutrition	Postpartum Depression	0.042874293
Malnutrition	Weight Loss Surgery	0.047782875
Malnutrition	Veterans And Military Family Health	0.044992351
Malnutrition	Inhalants	0.04698365
Malnutrition	College Health	0.042874293
Malnutrition	Cirrhosis	0.044367541
Nutrition For Seniors	Compulsive Gambling	0.041593877
Nutrition For Seniors	College Health	0.044676764
Nutrition For Seniors	Malnutrition	0.041559305
Sports Fitness	Compulsive Gambling	0.047619048
Sports Fitness	Inhalants	0.041559305
Sports Fitness	College Health	0.041238814
Nutritional Support	Compulsive Gambling	0.045314483
Common Cold	Compulsive Gambling	0.044281096
Club Drugs	Compulsive Gambling	0.042718612
Club Drugs	Inhalants	0.041169205
Smokeless Tobacco	Smoking And Youth	0.041848008
Child Sexual Abuse	Compulsive Gambling	0.042220815
Child Sexual Abuse	Child Behavior Disorders	0.04503693
Child Sexual Abuse	Domestic Violence	0.040893105
Elder Abuse	Compulsive Gambling	0.05220024
Elder Abuse	Mood Disorders	0.042295817
Elder Abuse	Personality Disorders	0.04406451
Elder Abuse	Weight Loss Surgery	0.042333418
Elder Abuse	Veterans And Military Family Health	0.043519889
Elder Abuse	Inhalants	0.04415011
Sexual Assault	Compulsive Gambling	0.04381161
Cirrhosis	Compulsive Gambling	0.053225463
Cirrhosis	Delirium	0.041450777
Cirrhosis	Mood Disorders	0.041703157
Cirrhosis	Personality Disorders	0.045979125
Cirrhosis	Weight Loss Surgery	0.045549786
Cirrhosis	Veterans And Military Family Health	0.041958629
Cirrhosis	Inhalants	0.04485713
Cirrhosis	Malnutrition	0.044367541

349

Appendix R. Complete lists of recommended connections generated through the Euclidean distance similarity value toward children related topics (format: topic, related topic, similarity value).

Topic	Related topic	Similarity Value
Toddler Health	Children's Health	0.049630255
Uncommon Infant and Newborn Problems	Dwarfism	0.045739377
Orthodontia	Dentures	0.050443906
Orthodontia	Dwarfism	0.045787546
Orthodontia	Artificial Limbs	0.04536176
Orthodontia	Mobility Aids	0.0464231
Baby Health Checkup	Dwarfism	0.051988563
Baby Health Checkup	Artificial Limbs	0.04828119
Baby Health Checkup	Mobility Aids	0.049568752
Baby Health Checkup	Barotrauma	0.0483372
Assistive Devices	Dentures	0.045549786
Assistive Devices	Temporomandibular Joint Dysfunction	0.047673532
Assistive Devices	Dwarfism	0.054232876
Assistive Devices	Barotrauma	0.051570316
Cochlear Implants	Dentures	0.045267303
Toddler Nutrition	Dwarfism	0.04536176
Cosmetic Dentistry	Dentures	0.061428835
Cosmetic Dentistry	Dwarfism	0.046373586
Cosmetic Dentistry	Artificial Limbs	0.047036689
Cosmetic Dentistry	Mobility Aids	0.048449612
Cosmetic Dentistry	Barotrauma	0.045502116
Dentures	Orthodontia	0.050443906
Dentures	Assistive Devices	0.045549786
Dentures	Cochlear Implants	0.045267303
Dentures	Cosmetic Dentistry	0.061428835
Dentures	Temporomandibular Joint Dysfunction	0.047245583
Dentures	Dwarfism	0.049507401
Dentures	Artificial Limbs	0.051366345
Dentures	Mobility Aids	0.051503914
Dentures	Barotrauma	0.049029221
Temporomandibular Joint Dysfunction	Assistive Devices	0.047673532
Temporomandibular Joint Dysfunction	Dentures	0.047245583
Temporomandibular Joint Dysfunction	Dwarfism	0.053149083
Temporomandibular Joint Dysfunction	Artificial Limbs	0.051639556
Temporomandibular Joint Dysfunction	Mobility Aids	0.052058931
Temporomandibular Joint Dysfunction	Barotrauma	0.049751244
Temporomandibular Joint Dysfunction	Brain Malformations	0.048678382

350

Dwarfism	Uncommon Infant and Newborn Problems	0.045739377
Dwarfism	Orthodontia	0.045787546
Dwarfism	Baby Health Checkup	0.051988563
Dwarfism	Assistive Devices	0.054232876
Dwarfism	Toddler Nutrition	0.04536176
Dwarfism	Cosmetic Dentistry	0.046373586
Dwarfism	Dentures	0.049507401
Dwarfism	Temporomandibular Joint Dysfunction	0.053149083
Dwarfism	Artificial Limbs	0.062869357
Dwarfism	Mobility Aids	0.06074596
Dwarfism	Barotrauma	0.057352604
Dwarfism	Tinnitus	0.04698365
Dwarfism	Drug Safety	0.046373586
Dwarfism	Brain Malformations	0.052851329
Dwarfism	Chiari Malformation	0.046524612
Dwarfism	Craniofacial Abnormalities	0.053604932
Hepatitis Testing	Hepatitis A	0.046324177
Artificial Limbs	Orthodontia	0.04536176
Artificial Limbs	Baby Health Checkup	0.04828119
Artificial Limbs	Cosmetic Dentistry	0.047036689
Artificial Limbs	Dentures	0.051366345
Artificial Limbs	Temporomandibular Joint Dysfunction	0.051639556
Artificial Limbs	Dwarfism	0.062869357
Artificial Limbs	Mobility Aids	0.069671846
Artificial Limbs	Barotrauma	0.057448153
Artificial Limbs	Tinnitus	0.046127589
Artificial Limbs	Drug Safety	0.047350727
Artificial Limbs	Brain Malformations	0.048854365
Artificial Limbs	Chiari Malformation	0.045884188
Artificial Limbs	Craniofacial Abnormalities	0.049207755
Medical Device Safety	Assistive Devices	0.048000768
Medical Device Safety	Mobility Aids	0.047510452
Mobility Aids	Orthodontia	0.0464231
Mobility Aids	Baby Health Checkup	0.049568752
Mobility Aids	Cosmetic Dentistry	0.048449612
Mobility Aids	Dentures	0.051503914
Mobility Aids	Temporomandibular Joint Dysfunction	0.052058931
Mobility Aids	Dwarfism	0.06074596
Mobility Aids	Artificial Limbs	0.069671846
Mobility Aids	Medical Device Safety	0.047510452
Mobility Aids	Barotrauma	0.058823529
Mobility Aids	Tinnitus	0.046829634
Mobility Aids	Drug Safety	0.048000768
Mobility Aids	Brain Malformations	0.049086982
Mobility Aids	Chiari Malformation	0.046076579
Mobility Aids	Craniofacial Abnormalities	0.049207755
Barotrauma	Baby Health Checkup	0.0483372
Barotrauma	Assistive Devices	0.051570316
Barotrauma	Cosmetic Dentistry	0.045502116
Barotrauma	Dentures	0.049029221
Barotrauma	Temporomandibular Joint Dysfunction	0.049751244

351

Barotrauma	Dwarfism	0.057352604
Barotrauma	Artificial Limbs	0.057448153
Barotrauma	Mobility Aids	0.058823529
Barotrauma	Meniere's Disease	0.046176579
Barotrauma	Tinnitus	0.049207755
Barotrauma	Drug Safety	0.046373586
Barotrauma	Brain Malformations	0.047458592
Barotrauma	Craniofacial Abnormalities	0.048111619
Meniere's Disease	Barotrauma	0.046176579
Tinnitus	Dwarfism	0.04698365
Tinnitus	Artificial Limbs	0.046127589
Tinnitus	Mobility Aids	0.046829634
Tinnitus	Barotrauma	0.049207755
Drug Safety	Dwarfism	0.046373586
Drug Safety	Artificial Limbs	0.047350727
Drug Safety	Mobility Aids	0.048000768
Drug Safety	Barotrauma	0.046373586
Brain Malformations	Temporomandibular Joint Dysfunction	0.048678382
Brain Malformations	Dwarfism	0.052851329
Brain Malformations	Artificial Limbs	0.048854365
Brain Malformations	Mobility Aids	0.049086982
Brain Malformations	Barotrauma	0.047458592
Brain Malformations	Chiari Malformation	0.046726788
Chiari Malformation	Dwarfism	0.046524612
Chiari Malformation	Artificial Limbs	0.045884188
Chiari Malformation	Mobility Aids	0.046076579
Chiari Malformation	Brain Malformations	0.046726788
Craniofacial Abnormalities	Dwarfism	0.053604932
Craniofacial Abnormalities	Artificial Limbs	0.049207755
Craniofacial Abnormalities	Mobility Aids	0.049207755
Craniofacial Abnormalities	Barotrauma	0.048111619

Appendix S. Complete lists of recommended connections generated through the Euclidean distance similarity value toward teenagers related topics (format: topic, related topic, similarity value).

Topic	Related topic	Similarity Value
Teen Development	Compulsive Gambling	0.037138825
Teen Sexual Health	Sexual Assault	0.036736343
Underage Drinking	Compulsive Gambling	0.036760651
Teen Violence	Compulsive Gambling	0.040690104
Teen Violence	Dwarfism	0.038690706
Prenatal Care	Preconception Care	0.038983315
Vaginal Bleeding	Compulsive Gambling	0.039073184
Vaginal Bleeding	Dwarfism	0.037822913
Vaginal Bleeding	Postpartum Care	0.037477045
Smoking and Youth	Compulsive Gambling	0.038662285
Smoking and Youth	Dwarfism	0.037345483
Compulsive Gambling	Teen Development	0.037138825
Compulsive Gambling	Underage Drinking	0.036760651
Compulsive Gambling	Teen Violence	0.040690104
Compulsive Gambling	Vaginal Bleeding	0.039073184
Compulsive Gambling	Smoking and Youth	0.038662285
Compulsive Gambling	Mood Disorders	0.041739711
Compulsive Gambling	Phobias	0.040096231
Compulsive Gambling	Psychotic Disorders	0.044676764
Compulsive Gambling	Schizophrenia	0.050125313
Compulsive Gambling	Seasonal Affective Disorder	0.040160643
Compulsive Gambling	Tubal Ligation	0.045643343
Compulsive Gambling	Vasectomy	0.050766575
Compulsive Gambling	Fetal Health and Development	0.037795752
Compulsive Gambling	Premature Babies	0.037345483
Compulsive Gambling	Twins, Triplets, Multiple Births	0.038954462
Compulsive Gambling	Folic Acid	0.04148689
Compulsive Gambling	Motor Vehicle Safety	0.049937578
Compulsive Gambling	Disabilities	0.044810898
Compulsive Gambling	Dwarfism	0.074951282
Compulsive Gambling	Pituitary Disorders	0.043192813
Compulsive Gambling	Speech and Language Problems in Children	0.038179597
Compulsive Gambling	E-Cigarettes	0.044194988
Compulsive Gambling	Smokeless Tobacco	0.043685291
Compulsive Gambling	Elder Abuse	0.04503693
Compulsive Gambling	Preterm Labor	0.041811264
Compulsive Gambling	Birth Weight	0.040225261
Compulsive Gambling	Postpartum Care	0.05581603
Compulsive Gambling	Stillbirth	0.048449612
Mood Disorders	Compulsive Gambling	0.041739711
Mood Disorders	Dwarfism	0.037063118
Phobias	Compulsive Gambling	0.040096231
Phobias	Dwarfism	0.037063118
Psychotic Disorders	Compulsive Gambling	0.044676764

353

Psychotic Disorders	Dwarfism	0.040859688
Psychotic Disorders	Postpartum Care	0.036810719
Schizophrenia	Compulsive Gambling	0.050125313
Schizophrenia	Psychotic Disorders	0.044587123
Schizophrenia	Vasectomy	0.037319003
Schizophrenia	Motor Vehicle Safety	0.036736343
Schizophrenia	Dwarfism	0.04536176
Schizophrenia	Postpartum Care	0.039013733
Seasonal Affective Disorder	Compulsive Gambling	0.040160643
Seasonal Affective Disorder	Dwarfism	0.037582682
Tubal Ligation	Compulsive Gambling	0.045643343
Tubal Ligation	Motor Vehicle Safety	0.037087861
Tubal Ligation	Dwarfism	0.043275056
Tubal Ligation	Postpartum Care	0.03883646
Vasectomy	Compulsive Gambling	0.050766575
Vasectomy	Schizophrenia	0.037319003
Vasectomy	Motor Vehicle Safety	0.039747208
Vasectomy	Dwarfism	0.048111619
Vasectomy	Postpartum Care	0.041995632
Vasectomy	Stillbirth	0.038236531
Fetal Health and Development	Compulsive Gambling	0.037795752
Fetal Health and Development	Dwarfism	0.036760651
Fetal Health and Development	Postpartum Care	0.037715924
Fetal Health and Development	Stillbirth	0.040160643
Premature Babies	Compulsive Gambling	0.037345483
Premature Babies	Postpartum Care	0.037529085
Twins, Triplets, Multiple Births	Compulsive Gambling	0.038954462
Twins, Triplets, Multiple Births	Dwarfism	0.037609538
Twins, Triplets, Multiple Births	Birth Weight	0.037609538
Twins, Triplets, Multiple Births	Postpartum Care	0.039777247
Twins, Triplets, Multiple Births	Stillbirth	0.037555864
Folic Acid	Compulsive Gambling	0.04148689
Folic Acid	Dwarfism	0.04
Motor Vehicle Safety	Compulsive Gambling	0.049937578
Motor Vehicle Safety	Schizophrenia	0.036736343
Motor Vehicle Safety	Tubal Ligation	0.037087861
Motor Vehicle Safety	Vasectomy	0.039747208
Motor Vehicle Safety	Dwarfism	0.047835446
Motor Vehicle Safety	Elder Abuse	0.037216226
Motor Vehicle Safety	Postpartum Care	0.04075644
Motor Vehicle Safety	Stillbirth	0.037822913
Disabilities	Compulsive Gambling	0.044810898
Disabilities	Dwarfism	0.04311274
Disabilities	Postpartum Care	0.037770056
Dwarfism	Teen Violence	0.038690706
Dwarfism	Vaginal Bleeding	0.037822913
Dwarfism	Smoking and Youth	0.037345483
Dwarfism	Compulsive Gambling	0.074951282
Dwarfism	Mood Disorders	0.037063118
Dwarfism	Phobias	0.037063118
Dwarfism	Psychotic Disorders	0.040859688
Dwarfism	Schizophrenia	0.04536176
Dwarfism	Seasonal Affective Disorder	0.037582682
Dwarfism	Tubal Ligation	0.043275056
Dwarfism	Vasectomy	0.048111619
Dwarfism	Fetal Health and Development	0.036760651

354

Dwarfism	Twins, Triplets, Multiple Births	0.037609538
Dwarfism	Folic Acid	0.04
Dwarfism	Motor Vehicle Safety	0.047835446
Dwarfism	Disabilities	0.04311274
Dwarfism	Pituitary Disorders	0.044107269
Dwarfism	Speech and Language Problems in Children	0.037477045
Dwarfism	E-Cigarettes	0.041811264
Dwarfism	Smokeless Tobacco	0.041238814
Dwarfism	Elder Abuse	0.042371086
Dwarfism	Preterm Labor	0.039223377
Dwarfism	Birth Weight	0.038925652
Dwarfism	Postpartum Care	0.051778595
Dwarfism	Stillbirth	0.046027801
Pituitary Disorders	Compulsive Gambling	0.043192813
Pituitary Disorders	Dwarfism	0.044107269
Pituitary Disorders	Postpartum Care	0.037087861
Speech and Language Problems in Children	Compulsive Gambling	0.038179597
Speech and Language Problems in Children	Dwarfism	0.037477045
E-Cigarettes	Smoking and Youth	0.039497591
E-Cigarettes	Compulsive Gambling	0.044194988
E-Cigarettes	Dwarfism	0.041811264
E-Cigarettes	Smokeless Tobacco	0.037114014
Smokeless Tobacco	Smoking and Youth	0.042144302
Smokeless Tobacco	Compulsive Gambling	0.043685291
Smokeless Tobacco	Dwarfism	0.041238814
Smokeless Tobacco	E-Cigarettes	0.037114014
Postpartum Depression	Postpartum Care	0.038720669
Elder Abuse	Compulsive Gambling	0.04503693
Elder Abuse	Motor Vehicle Safety	0.037216226
Elder Abuse	Dwarfism	0.042371086
Elder Abuse	Postpartum Care	0.037477045
Sexual Assault	Teen Sexual Health	0.036736343
Preterm Labor	Compulsive Gambling	0.041811264
Preterm Labor	Dwarfism	0.039223377
Preterm Labor	Postpartum Care	0.038096689
Preterm Labor	Stillbirth	0.037423749
Birth Weight	Compulsive Gambling	0.040225261
Birth Weight	Twins, Triplets, Multiple Births	0.037609538
Birth Weight	Dwarfism	0.038925652
Birth Weight	Postpartum Care	0.039808917
HIV/AIDS and Infections	HIV/AIDS and Pregnancy	0.039044198
Preconception Care	Postpartum Care	0.036885397
Postpartum Care	Vaginal Bleeding	0.037477045
Postpartum Care	Compulsive Gambling	0.05581603
Postpartum Care	Psychotic Disorders	0.036810719
Postpartum Care	Schizophrenia	0.039013733
Postpartum Care	Tubal Ligation	0.03883646
Postpartum Care	Vasectomy	0.041995632
Postpartum Care	Fetal Health and Development	0.037715924
Postpartum Care	Premature Babies	0.037529085

355

Postpartum Care	Twins, Triplets, Multiple Births	0.039777247
Postpartum Care	Motor Vehicle Safety	0.04075644
Postpartum Care	Disabilities	0.037770056
Postpartum Care	Dwarfism	0.051778595
Postpartum Care	Pituitary Disorders	0.037087861
Postpartum Care	Elder Abuse	0.037477045
Postpartum Care	Preterm Labor	0.038096689
Postpartum Care	Birth Weight	0.039808917
Postpartum Care	Preconception Care	0.036885397
Postpartum Care	Stillbirth	0.04441089
Stillbirth	Compulsive Gambling	0.048449612
Stillbirth	Vasectomy	0.038236531
Stillbirth	Fetal Health and Development	0.040160643
Stillbirth	Twins, Triplets, Multiple Births	0.037555864
Stillbirth	Motor Vehicle Safety	0.037822913
Stillbirth	Dwarfism	0.046027801
Stillbirth	Preterm Labor	0.037423749
Stillbirth	Postpartum Care	0.04441089

Appendix T. Complete lists of recommended connections generated through the Euclidean distance similarity value toward older adults related topics (format: topic, related topic, similarity value).

Topic	Related topic	Similarity Value
Assisted Living	Financial Assistance	0.042448425
Assisted Living	Patient Rights	0.039103742
Assisted Living	Compulsive Gambling	0.040995368
Medicare Prescription Drug Coverage	Medicaid	0.040521922
Medicare Prescription Drug Coverage	Compulsive Gambling	0.04177633
Child Care	Sports Fitness	0.040389353
Child Care	Financial Assistance	0.047945534
Child Care	Health Facilities	0.040389353
Child Care	Divorce	0.044676764
Child Care	Medicaid	0.046076579
Child Care	Malnutrition	0.044324276
Child Care	Water Safety (Recreational)	0.044194988
Child Care	Patient Rights	0.044810898
Child Care	Carbohydrate Metabolism Disorders	0.039872408
Child Care	Toddler Nutrition	0.045739377
Child Care	Vitamin A	0.041310365
Child Care	Vitamin C	0.044499822
Child Care	Vitamin E	0.040096231
Child Care	Compulsive Gambling	0.055300559
Sports Fitness	Child Care	0.040389353
Sports Fitness	Medicaid	0.038690706
Sports Fitness	Vitamin C	0.038807824
Sports Fitness	Compulsive Gambling	0.043685291
Financial Assistance	Assisted Living	0.042448425
Financial Assistance	Child Care	0.047945534
Financial Assistance	Health Facilities	0.046324177
Financial Assistance	Divorce	0.041884817
Financial Assistance	Managed Care	0.039936102
Financial Assistance	Malnutrition	0.044992351
Financial Assistance	Water Safety (Recreational)	0.040723245
Financial Assistance	Patient Rights	0.051639556
Financial Assistance	Carbohydrate Metabolism Disorders	0.040291712
Financial Assistance	Toddler Nutrition	0.043233895
Financial Assistance	Vitamin A	0.041274558
Financial Assistance	Vitamin C	0.045739377
Financial Assistance	Vitamin E	0.042144302
Financial Assistance	Compulsive Gambling	0.060856865
Financial Assistance	Schizophrenia	0.041416442
Health Facilities	Child Care	0.040389353
Health Facilities	Financial Assistance	0.046324177
Health Facilities	Medicaid	0.039528817
Health Facilities	Patient Rights	0.04485713

357

Health Facilities	Compulsive Gambling	0.044810898
Divorce	Child Care	0.044676764
Divorce	Financial Assistance	0.041884817
Divorce	Medicaid	0.042333418
Divorce	Malnutrition	0.041238814
Divorce	Patient Rights	0.039747208
Divorce	Toddler Nutrition	0.039436842
Divorce	Vitamin C	0.041884817
Divorce	Compulsive Gambling	0.049630255
Sports Safety	Compulsive Gambling	0.042371086
Medicaid	Medicare Prescription Drug Coverage	0.040521922
Medicaid	Child Care	0.046076579
Medicaid	Sports Fitness	0.038690706
Medicaid	Financial Assistance	0.051297835
Medicaid	Health Facilities	0.039528817
Medicaid	Divorce	0.042333418
Medicaid	Malnutrition	0.045454545
Medicaid	Water Safety (Recreational)	0.040995368
Medicaid	Patient Rights	0.04647488
Medicaid	Carbohydrate Metabolism Disorders	0.040893105
Medicaid	Toddler Nutrition	0.041995632
Medicaid	Vitamin A	0.041416442
Medicaid	Vitamin C	0.047140904
Medicaid	Vitamin E	0.041848008
Medicaid	Compulsive Gambling	0.059548621
Medicaid	Schizophrenia	0.040389353
Managed Care	Financial Assistance	0.039936102
Dietary Fiber	Compulsive Gambling	0.038807824
Malnutrition	Child Care	0.044324276
Malnutrition	Financial Assistance	0.044992351
Malnutrition	Divorce	0.041238814
Malnutrition	Medicaid	0.045454545
Malnutrition	Water Safety (Recreational)	0.040456348
Malnutrition	Patient Rights	0.042914771
Malnutrition	Fluid and Electrolyte Balance	0.040859688
Malnutrition	Carbohydrate Metabolism Disorders	0.042874293
Malnutrition	Toddler Nutrition	0.04299411
Malnutrition	Antioxidants	0.039254171
Malnutrition	B Vitamins	0.044587123
Malnutrition	Vitamin A	0.049207755
Malnutrition	Vitamin C	0.0527037
Malnutrition	Vitamin E	0.048854365
Malnutrition	Vitamin K	0.041344524
Malnutrition	Compulsive Gambling	0.055555556
Malnutrition	Schizophrenia	0.039436842
Nutritional Support	Toddler Nutrition	0.039840637
Nutritional Support	Compulsive Gambling	0.041416442
Mood Disorders	Carbohydrate Metabolism Disorders	0.039073184
Mood Disorders	Compulsive Gambling	0.041135335
Water Safety (Recreational)	Child Care	0.044194988
Water Safety (Recreational)	Financial Assistance	0.040723245
Water Safety (Recreational)	Sports Safety	0.042640287

358

Water Safety (Recreational)	Medicaid	0.040995368
Water Safety (Recreational)	Malnutrition	0.040456348
Water Safety (Recreational)	Vitamin C	0.040389353
Water Safety (Recreational)	Compulsive Gambling	0.047510452
Patient Rights	Assisted Living	0.039103742
Patient Rights	Child Care	0.044810898
Patient Rights	Financial Assistance	0.051639556
Patient Rights	Health Facilities	0.04485713
Patient Rights	Divorce	0.039747208
Patient Rights	Medicaid	0.04647488
Patient Rights	Malnutrition	0.042914771
Patient Rights	Carbohydrate Metabolism Disorders	0.039132817
Patient Rights	Toddler Nutrition	0.03990423
Patient Rights	Vitamin A	0.039528817
Patient Rights	Vitamin C	0.04315367
Patient Rights	Vitamin E	0.039528817
Patient Rights	Compulsive Gambling	0.054800526
Fluid and Electrolyte Balance	Malnutrition	0.040859688
Fluid and Electrolyte Balance	Compulsive Gambling	0.042069836
Carbohydrate Metabolism Disorders	Child Care	0.039872408
Carbohydrate Metabolism Disorders	Financial Assistance	0.040291712
Carbohydrate Metabolism Disorders	Medicaid	0.040893105
Carbohydrate Metabolism Disorders	Malnutrition	0.042874293
Carbohydrate Metabolism Disorders	Mood Disorders	0.039073184
Carbohydrate Metabolism Disorders	Patient Rights	0.039132817
Carbohydrate Metabolism Disorders	Vitamin C	0.041666667
Carbohydrate Metabolism Disorders	Vitamin E	0.038954462
Carbohydrate Metabolism Disorders	Compulsive Gambling	0.050251256
Toddler Nutrition	Child Care	0.045739377
Toddler Nutrition	Financial Assistance	0.043233895
Toddler Nutrition	Divorce	0.039436842
Toddler Nutrition	Medicaid	0.041995632
Toddler Nutrition	Malnutrition	0.04299411
Toddler Nutrition	Nutritional Support	0.039840637
Toddler Nutrition	Patient Rights	0.03990423
Toddler Nutrition	Vitamin A	0.040893105
Toddler Nutrition	Vitamin C	0.043894303
Toddler Nutrition	Vitamin E	0.039715636
Toddler Nutrition	Compulsive Gambling	0.048621578
Lipid Metabolism Disorders	Compulsive Gambling	0.040356754
Colonic Diseases	Compulsive Gambling	0.042032701
DASH Eating Plan	Compulsive Gambling	0.03928347
Creutzfeldt-Jakob Disease	Compulsive Gambling	0.04
Malabsorption Syndromes	Compulsive Gambling	0.040521922
Antioxidants	Malnutrition	0.039254171
Antioxidants	Compulsive Gambling	0.040521922

359

B Vitamins	Malnutrition	0.044587123
B Vitamins	Vitamin A	0.068041097
B Vitamins	Vitamin C	0.066666667
B Vitamins	Vitamin D	0.042448425
B Vitamins	Vitamin E	0.06900835
B Vitamins	Vitamin K	0.0625
B Vitamins	Compulsive Gambling	0.041135335
Folic Acid	B Vitamins	0.041204829
Folic Acid	Vitamin C	0.041593877
Folic Acid	Compulsive Gambling	0.040961783
Vitamin A	Child Care	0.041310365
Vitamin A	Financial Assistance	0.041274558
Vitamin A	Medicaid	0.041416442
Vitamin A	Malnutrition	0.049207755
Vitamin A	Patient Rights	0.039528817
Vitamin A	Toddler Nutrition	0.040893105
Vitamin A	B Vitamins	0.068041097
Vitamin A	Vitamin C	0.098531875
Vitamin A	Vitamin D	0.046574449
Vitamin A	Vitamin E	0.107828337
Vitamin A	Vitamin K	0.077615647
Vitamin A	Compulsive Gambling	0.048393341
Vitamin C	Child Care	0.044499822
Vitamin C	Sports Fitness	0.038807824
Vitamin C	Financial Assistance	0.045739377
Vitamin C	Divorce	0.041884817
Vitamin C	Medicaid	0.047140904
Vitamin C	Malnutrition	0.0527037
Vitamin C	Water Safety (Recreational)	0.040389353
Vitamin C	Patient Rights	0.04315367
Vitamin C	Carbohydrate Metabolism Disorders	0.041666667
Vitamin C	Toddler Nutrition	0.043894303
Vitamin C	B Vitamins	0.066666667
Vitamin C	Folic Acid	0.041593877
Vitamin C	Vitamin A	0.098531875
Vitamin C	Vitamin D	0.042107036
Vitamin C	Vitamin E	0.103691414
Vitamin C	Vitamin K	0.073126143
Vitamin C	Compulsive Gambling	0.058122639
Vitamin C	Schizophrenia	0.039132817
Vitamin D	B Vitamins	0.042448425
Vitamin D	Vitamin A	0.046574449
Vitamin D	Vitamin C	0.042107036
Vitamin D	Vitamin E	0.048854365
Vitamin D	Vitamin K	0.052058931
Vitamin E	Child Care	0.040096231
Vitamin E	Financial Assistance	0.042144302
Vitamin E	Medicaid	0.041848008
Vitamin E	Malnutrition	0.048854365
Vitamin E	Patient Rights	0.039528817
Vitamin E	Carbohydrate Metabolism Disorders	0.038954462
Vitamin E	Toddler Nutrition	0.039715636
Vitamin E	Antioxidants	0.045930553
Vitamin E	B Vitamins	0.06900835

360

Vitamin E	Vitamin A	0.107828337
Vitamin E	Vitamin C	0.103691414
Vitamin E	Vitamin D	0.048854365
Vitamin E	Vitamin K	0.08838607
Vitamin E	Compulsive Gambling	0.049207755
Vitamin K	Malnutrition	0.041344524
Vitamin K	Vitamins	0.042795395
Vitamin K	B Vitamins	0.0625
Vitamin K	Vitamin A	0.077615647
Vitamin K	Vitamin C	0.073126143
Vitamin K	Vitamin D	0.052058931
Vitamin K	Vitamin E	0.08838607
Vitamin K	Compulsive Gambling	0.040064103
Compulsive Gambling	Assisted Living	0.040995368
Compulsive Gambling	Medicare Prescription Drug Coverage	0.04177633
Compulsive Gambling	Child Care	0.055300559
Compulsive Gambling	Sports Fitness	0.043685291
Compulsive Gambling	Financial Assistance	0.060856865
Compulsive Gambling	Health Facilities	0.044810898
Compulsive Gambling	Divorce	0.049630255
Compulsive Gambling	Sports Safety	0.042371086
Compulsive Gambling	Medicaid	0.059548621
Compulsive Gambling	Dietary Fiber	0.038807824
Compulsive Gambling	Malnutrition	0.055555556
Compulsive Gambling	Nutritional Support	0.041416442
Compulsive Gambling	Mood Disorders	0.041135335
Compulsive Gambling	Water Safety (Recreational)	0.047510452
Compulsive Gambling	Patient Rights	0.054800526
Compulsive Gambling	Fluid and Electrolyte Balance	0.042069836
Compulsive Gambling	Carbohydrate Metabolism Disorders	0.050251256
Compulsive Gambling	Toddler Nutrition	0.048621578
Compulsive Gambling	Lipid Metabolism Disorders	0.040356754
Compulsive Gambling	Colonic Diseases	0.042032701
Compulsive Gambling	DASH Eating Plan	0.03928347
Compulsive Gambling	Creutzfeldt-Jakob Disease	0.04
Compulsive Gambling	Malabsorption Syndromes	0.040521922
Compulsive Gambling	Antioxidants	0.040521922
Compulsive Gambling	B Vitamins	0.041135335
Compulsive Gambling	Folic Acid	0.040961783
Compulsive Gambling	Vitamin A	0.048393341
Compulsive Gambling	Vitamin C	0.058122639
Compulsive Gambling	Vitamin E	0.049207755
Compulsive Gambling	Vitamin K	0.040064103
Compulsive Gambling	Psychotic Disorders	0.044367541
Compulsive Gambling	Schizophrenia	0.049446203
Compulsive Gambling	Teen Violence	0.040456348
Psychotic Disorders	Compulsive Gambling	0.044367541
Schizophrenia	Financial Assistance	0.041416442
Schizophrenia	Medicaid	0.040389353
Schizophrenia	Malnutrition	0.039436842
Schizophrenia	Vitamin C	0.039132817
Schizophrenia	Compulsive Gambling	0.049446203
Schizophrenia	Psychotic Disorders	0.04441089
Teen Violence	Compulsive Gambling	0.040456348

361

Curriculum Vitae

Yifan Zhu

School of Information Studies
University of Wisconsin-Milwaukee, Milwaukee

EDUCATION

University of Wisconsin-Milwaukee, WI, USA
Ph.D., Information Studies Program, GPA: 4.0/4.0   Graduation: May. 2021 (expected)
Advisor: Dr. Jin Zhang

University of Wisconsin-Madison, WI, USA
M.L.I.S., Library and Information Science, GPA: 3.8/4.0 Graduation: May. 2015
Advisor: Dr. Kyung-Sun "Sunny" Kim

Sichuan University, Chengdu, China
B.M., Information Resources Management, GPA: 3.2/4.0 Graduation: May. 2013
Advisor: Dr. Guihua Li

RESEARCH INTERESTS

- Information Retrieval
- Health Information
- Information Seeking Behavior
- Information Privacy and Ethics

SKILLS

- Languages: Python, SQL, HTML, CSS
- Toolkits: UciNet, Octopus, SPSS, SAS

362

PUBLICATIONS

Peer-Review Journal:
1. **Zhu, Y.**, & Zhang, J. (2020). Social network analysis on the Mental Health sub-topic in the MedlinePlus subject directory. *Information Research: An International Electronic Journal, 25*(4), paper 876. (SSCI journal)
2. Reuter, K., **Zhu, Y.**, Angyan, P., Le, N., Merchant, A. A., & Zimmer, M. (2019). Public Concern About Monitoring Twitter Users and Their Conversations to Recruit for Clinical Trials: Survey Study. *Journal of medical Internet research*, *21*(10), e15455. (Co-first author. SCI journal. This article has been selected by the Editorial Board of the 2020 IMIA Yearbook of Medical Informatics for listing in the 2020 edition of the Yearbook as one of the best articles published in 2019 in the Consumer Health Informatics and Education.)
3. Zhang, J., Zheng, Z., Wang, Y., & **Zhu, Y.** (2020). A study on users' preference towards diabetes-related video clips on YouTube. *BMC Medical Informatics and Decision Making*, *20*(1), 1-16. (SCI journal.)
4. Li, G., Wang, Y., & **Zhu, Y.** (2014). The reception and response mechanism to online rumors and their risk management. *Journal of the China Society for Scientific and Technical Information, 33*(3), pp305-312. (In Chinese. Best journal under the discipline catalog of Information Studies in China.)

Book Chapter Co-Author:
1. Jiang, X. (2013). *The Development Report of the Social Management Innovation for Sichuan Province*. Beijing: China Renmin University Press Co. LTD

Conference Paper:
1. **Zhu, Y.**, & Zhang, J. (2020). An investigation study on the mental disorder related topics in the subject directory of MedlinePlus portal. *iConference 2020 Proceedings*.
2. Li, G., Wang, Y., & **Zhu, Y.** (2012). Influencing factors of public trust of network rumors. *Conference abstract of the 8th Cross-Strait Conference on Public Administration*, Chengdu, pp104. (In Chinese)

HONORS AND AWARDS

- 2011-2012   School of Public Administration Scholarship, Sichuan University
- 2013-2014   School of Library and Information Studies Scholarship, University of Wisconsin-Madison
- 2015-2021   Chancellor's Graduate Student Award, University of Wisconsin-Milwaukee

CERTIFICATES

- Certificate in Strategic Innovation, University of Wisconsin-Madison, 2015
- IRB-Social & Behavior Researchers, CITI Program

TEACHING EXPERIENCE

Instructor:
1. Introduction to System Analysis, Online & On-site Sessions, University of Wisconsin-Milwaukee. (Sep 2018 - May 2021)
2. Human Factors in Information Seeking and Use, Online & On-site Sessions, University of Wisconsin-Milwaukee. (Sep 2020 – May 2021)

Teaching Assistant:
1. Foundations of Library and Information Science, University of Wisconsin-Milwaukee. (Sep - Dec 2016)
2. Information Architecture, University of Wisconsin-Milwaukee. (Jan - May 2017)
3. Introduction to Data Science, University of Wisconsin-Milwaukee. (Jan - May 2017)
4. Introduction to System Analysis, University of Wisconsin-Milwaukee. (Sep - Dec 2017)
5. Information Technology Ethics, University of Wisconsin-Milwaukee. (Jan – May 2018)

Guest Speaker:
1. Weekly Chapter: "Information needs and user behavior", Foundations of Library and Information Science, University of Wisconsin-Milwaukee. (Oct 2016)