August 2018

Analysis of Family-Health-Related Topics on Wikipedia

Yanyan Wang
University of Wisconsin-Milwaukee

ANALYSIS OF FAMILY-HEALTH-RELATED TOPICS ON WIKIPEDIA

by

Yanyan Wang

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

August 2018

ABSTRACT

ANALYSIS OF FAMILY-HEALTH-RELATED TOPICS ON WIKIPEDIA

by

Yanyan Wang

The University of Wisconsin-Milwaukee, 2018
Under the Supervision of Professor Jin Zhang

New concepts, terms, and topics always emerge; and meanings of existing terms and topics keep changing all the time. These phenomena occur more frequently on social media than on conventional media because social media allows a huge number of users to generate information online. Retrieving relevant results in different time periods of a fast-changing topic becomes one of the most difficult challenges in the information retrieval field. Among numerous topics discussed on social media, health-related topics are a major category which attracts increasing attention from the general public.

This study investigated and explored the evolution patterns of family-health-related topics on Wikipedia. Three family-health-related topics (Child Maltreatment, Family Planning, and Women's Health) were selected from the World Health Organization Website and their associated entries were retrieved on Wikipedia. Historical numeric and text data of the entries from 2010 to 2017 were collected from a Wikipedia data dump and the Wikipedia Web pages. Four periods were defined: 2010 to 2011, 2012 to 2013, 2014 to 2015, and 2016 to 2017. Coding, subject analysis, descriptive statistical analysis, inferential statistical analysis, SOM

approach, and n-gram approach were employed to explore the internal characteristics and external popularity evolutions of the topics.

The findings illustrate that the external popularities of the family-health-related topics declined from 2010 to 2017, although their content on Wikipedia kept increasing. The emerged entries had three features: specialization, summarization, and internationalization. The subjects derived from the entries became increasingly diverse during the investigated periods. Meanwhile, the developing trajectories of the subjects varied from one to another. According to the developing trajectories, the subjects were grouped into three categories: growing subject, diminishing subject, and fluctuating subject. The popularities of the topics among the Wikipedia viewers were consistent, while among the editors were not. For each topic, its popularity trend among the editors and the viewers was inconsistent. Child Maltreatment was the most popular among the three topics, Women's Health was the second most popular, while Family Planning was the least popular among the three.

The implications of this study include: (1) helping health professionals and general users get a more comprehensive understanding of the investigated topics; (2) contributing to the developments of health ontologies and consumer health vocabularies; (3) assisting Website designers in organizing online health information and helping them identify popular family-health-related topics; (4) providing a new approach for query recommendation in information retrieval systems; (5) supporting temporal information retrieval by presenting the temporal changes of family-health-related topics; and (6) providing a new combination of data collection and analysis methods for researchers.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Dissertation writing has been a period of intense learning for me, not only in the academic research, but also on a personal level. Without this valuable experience, I will never know how far I can push myself.

I would like to express the deepest appreciation to my advisor, Dr. Jin Zhang, who has provided me with unconditional support and countless help during my entire graduate study. His inner passion for research and teaching affected me a lot. Without his guidance and persistent help, this dissertation would not have been possible.

I would like to thank my committee members, Dr. Iris Xie, Dr. Nadine Kozak, Dr. Xiangming Mu, and Dr. Timothy Patrick. Thank you so much for being my committee members, reading my proposal and dissertation drafts, and giving me insightful suggestions and comments.

I am so lucky to have the continuing and strong support from my family, my boyfriend (Dr. Zhi Chen), and my friends. When I was upset and stressed, they always encouraged me and cheered me up. I would also like to extend my gratitute to my best roommate, Dr. Zhe Kong.

1. INTRODUCTION

1.1. Background & Rationale

With the development of computer technology and Internet technology, the volume of information and data keeps increasing. Concepts and terms regarding certain topics are always changing. Not only the meanings of concepts and terms, but also the concepts and terms relevant to certain topics change over time. With the improvement of social media, the amount of user-generated content on social media grows and the changing of concepts, themes and topics are much quicker than before. These changes cause problems in information retrieval. For instance, the term “cloud” has existed for hundreds of years. Several decades ago there was no relation between the term cloud and the term computer. However, with the generation of cloud computing, cloud and computer often occur in one document together. Therefore, it is necessary to explore the temporal features of concepts, terms, and topics.

In the past decade, the use of social media kept increasing globally (Boyd & Ellison, 2007; Moorhead et al., 2013). It was reported by the eBizMBA Rank that popular social media platforms had millions of users. Facebook had 900 million unique monthly visitors and Twitter had attracted 310 million unique monthly visitors by October 2015. Meanwhile, the volume of user-generated content on social media grew rapidly. For example, Twitter users had broadcasted more than 500 million tweets per day by 2014 (Matta, Doiron, & Leveridge, 2014). Since the information on social media is generated by a great number of individuals, the user-generated content on these platforms reflect the general public’s perceptions and consensus to some extent. A good example is Wikipedia. Being the largest online knowledge collaboration,

Wikipedia allows users to create, revise, and delete entries. Statistics offered by Wikipedia indicate that more than 900 thousand content creators generated nearly 5 million articles. Given these numbers, it is reasonable to assume that the content entries on Wikipedia reflect a consensus of the general public. Examining content on social media is necessary for researchers to gain insight into the general public's perceptions. To explore how certain topics change on social media will show the evolution of the general public's understanding of these topics.

There are challenges for examining content on social media. One of the challenges is that the amount of data to be collected is tremendous and the another is that information on social media is dynamic. Most social media platforms do not record historical versions of information or the historical data are unavailable for researchers. Different from other social media platforms, Wikipedia stores all the historical versions and data of entries, and since Wikipedia is a non-profit organization and all its content is created by users, the data on Wikipedia are open to anyone and relatively easy to access and collect. Hence, to investigate the evolution of topics on social media, Wikipedia is a valid data source.

Containing such a huge volume of information, social media is not only regarded as the channel for information creation and diffusion, and connection building and maintaining, but also a platform and source for seeking information online. As the quality of life of the general public improves, people pay more attention to the health status of themselves and their family members. Seeking health information on social media is becoming more common than in previous years. It was reported that 23% of social media users followed their friends' health experiences and updates and 15% of them sought and retrieved health information from social media sites (Fox, 2011). In this situation, the content and quality of health information are

important for users, especially for patients and their families. Examining health-related topics on social media can assist users in seeking and using health information.

The purpose of this study is to explore how family-health-related topics change on social media. Wikipedia was selected for data collection because its historical data were comprehensive and accessible. Temporal analysis was used to examine the evolution of health-related topics. This method has been used to explore changes and patterns of certain objects and predict future trends of certain objects. In this study, the temporal analysis method was to compare the internal characteristics and external popularities of a specific topic in different time periods. Internal characteristics of a topic included the concepts relevant to the topic, the subjects and themes hidden in the text of the concepts, and the relations among them. Each Wikipedia entry contained information related to a specific concept. External popularity of a topic is represented by the number of views and edits of the topics. To reveal relationships among entries, open coding method and clustering approaches were employed to group them into categories and clusters. A Self-Organizing Map (SOM) was utilized as the clustering approach in this study. The SOM approach is a neural network method that measures similarities among items of input data so as to form similarity graphs. This approach was adopted in research in a variety of fields like biology, artificial intelligence, and finance, but it has not been widely used in the information science. In this study this approach was applied to the exploration of the selected topics' internal characteristics.

This study has six chapters: Introduction, Literature Review, Research Methodology, Results, Discussion and Implications, and Conclusion. The first chapter introduces the background and rationale, the research problem and research questions, and the research

design of this study. The second chapter reviews the primary relevant literature of this study. The third chapter presents the data collection procedures and analysis and provides a detailed description of the methods and approaches used in the study. The fourth chapter presents the results obtained for the research questions. The fifth chapter discuss the unique findings presented in the Results chapter and compares the findings with those of previous studies reported in the literature. The last chapter summarizes the previous chapters and presents the conclusions of this study.

1.2. Research Problem, Questions and Hypotheses

1.2.1. Research Problem Statement

Over the past several decades, the general public has focused increasing attention on their health status as well as that of their families. In recent years the number of users who seek health information online has continued to grow. These users seek health information not only from online databases (e.g. PubMed) and Websites for the general public (e.g. WebMD and MayoClinic) but also from social media platforms (e.g. Wikipedia, Facebook, and Yahoo!Answers). Social media allows the general public to create, revise, share, and seek information, and to communicate with each other. The user-generated content on social media contains a variety of topics and health is one of the primary topics. As the health-related content on social media increases, social media has become an important health information source for the general public.

Among the different types of social media platforms, Wikipedia can be considered a representative social media Website. It is a rapidly growing platform containing vast interlinked

information (Milne & Witten, 2013). Currently it is the largest online information collaboration consisting of user-generated content. Since the health-related information on Wikipedia is created by the general public, it reflects the general public's interests and understandings of health-related topics to a large extent. Therefore, the changes of this information over time show the changes of the general public's interests and understandings of health-related topics.

This study's research problem is to investigate and discover the evolution of three family-health-related topics derived from the social media website Wikipedia. The research problem can be divided into three layers: (1) the internal characteristics and external popularities of specific family-health-related topics on Wikipedia in certain time periods; (2) the evolution patterns of the topics over time; (3) the commonalities and differences among the evolution patterns of the topics. Accordingly, the research questions to be addressed are stated below.

1.2.2. Research Question One

The first layer of the research problem examined the family-health-related topics and their associated entries, themes, and subjects on Wikipedia. However, because of the time and text length limitations, it was impossible to examine all the related topics on Wikipedia. Therefore, several typical family-health-related topics and their associated entries on Wikipedia were investigated. The research question and sub-questions are:

RQ1: What are the associated entries, emerged themes and subjects, and relations among them in each of the selected family-health-related topics?

RQ1a: What are the associated entries and the main themes of the selected family-health-related topics discussed on Wikipedia?

RQ1b: What are the subjects of each theme of the selected family-health-related topics?

RQ1c: What are the relations among the themes, subjects, and entries?

Three family-health-related topics (Child Maltreatment, Family Planning, and Women's Health) were selected for this study. The criteria for choosing the qualified topics included: (1) the topics should be popular family-health-related topics widely discussed by the general public; (2) the topics should cover different aspects of family health issues; (3) in order to collect data from Wikipedia, there should be more than 100 relevant entries for each of the topics on Wikipedia and the entries should contain sufficient data; (4) the lengths of the selected topics' history should be longer than 8 years. These criteria are described in detail in the Methodology section.

Each family-health-related topic included one seed entry and a number of associated entries on Wikipedia. These entries draw a picture of a specific topic from different aspects. R1a intends to explore the different themes of every selected topic. Each theme stands for one aspect of the specific topic. R1b aims to extract the subjects from the content of the entries belonging to the specific themes, respectively. Based on the findings, the relations among the entries, subjects, and themes are constructed in R1c. In other words, the internal characteristics of every selected topic were examined.

1.2.3. Research Question Two

The second layer of the research problem explored the evolutions of the selected topics. The evolution of the specific topics can be observed from two aspects: internal characteristics and external popularity. Therefore, the second research question is divided into two sub-questions: the evolution of the internal characteristics for each topic and the evolution of the external popularity for each topic.

RQ2: What are the evolution patterns for each of the selected family-health-related topics in terms of the internal characteristics and external popularity?

RQ2a: What are the new entries created in each investigated time period for each theme of the selected topics? What subjects are emerging and disappearing in each period for each theme of the selected topics? For each topic, what are the evolution patterns of its internal characteristics during these time periods?

RQ2b: What are the evolution patterns for the selected topics in terms of their associated entries' number of views and number of edits? What are the evolution patterns for the themes of each selected topic in terms of the associated entries' number of views and number of edits?

H01: There were no significant differences among the investigated time periods in terms of the number of page views of the entries relevant to each of the topics.

H02: There were no significant differences among the investigated time periods in terms of the number of page edits of the entries relevant to each of the topics.

The second research question investigated the changes of the selected family-health-related topics over the past decade. The internal characteristics of a specific topic in different time periods show the emergences, growths, and disappearances of entries, subjects, and terms in each theme. Figure 1 presents the structure of entries, themes, and subjects for a topic in a specific time period.

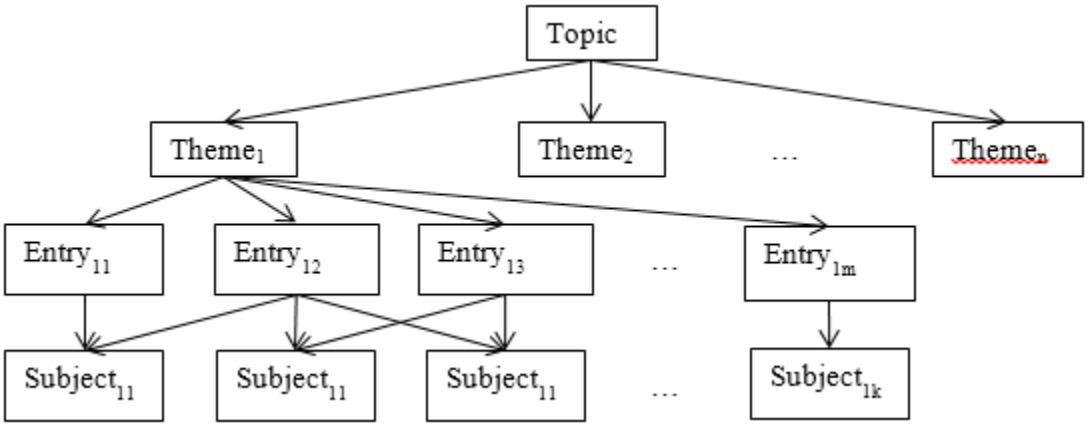


Figure 1. Concept Map

Figure 1 demonstrates that there could be multiple themes contained in one topic and every theme relates to one to many entries. Each entry can only belong to one theme. The content of the entries in different time periods were collected from the corresponding historical versions of entries on Wikipedia. For one period, the subjects of a theme were determined by the content of the entries within this theme. An entry may relate to one or more than one subject, while different entries could relate to the same subjects. Therefore, according to the entries included, a theme may have more than one subject.

It is possible that a subject is either broader or narrower than a theme, since a theme was confined by a subject in the schema. For instance, the Child prostitution entry belonged to

the Abuse, violence, harm, and subordination theme. Its content involved the definitions and causes of child prostitution. The definition part was associated with the abuse and violence subject and the cause part was associated with the social factor subject. The abuse and violence subject was relatively narrower than the theme, while the social factor subject was broader.

In this study, the external popularity of an entry was measured by its number of page edits and number of page views. The number of page edits reflects the popularity of an entry among the Wikipedia editors and the number of page views reflects the popularity of an entry among the Wikipedia viewers. Regarding one topic, the total number of page views of all the entries associated with it reflects its popularity, and so does the total number of page edits. The popularities of the themes in a topic can also be revealed by these two measures.

Since the revision data of Wikipedia entries are available from July 2006 and other data are available from when entries were created, this study explored the internal characteristics and external popularities of the selected topics from 2010 to 2017. Four periods were defined: 2010 to 2011, 2012 to 2013, 2014 to 2015, and 2016 to 2017. The detailed criteria for determining these time periods are described in the Methodology section.

1.2.4. Research Question Three

The third layer of the research problem tends to compare the evolution patterns of different topics and to explore the differences and commonalities among them. Similar to the second research question, the internal characteristic evolution patterns and the external popularity evolution patterns of the selected topics were compared.

RQ3: What are the differences and commonalities among the selected topics in terms of their evolution patterns?

H03: There were no significant differences among the selected topics in terms of the number of the page edits of the associated entries.

H04: There were no significant differences among the selected topics in terms of the number of the page views of the associated entries.

After RQ1 and RQ2 explored the evolution patterns of three family-health-related topics individually, this research question compared these evolution patterns to determine commonalities and differences among them from both qualitative and quantitative perspectives.

In order to clarify the connections among the research problem, research questions, and hypotheses, Figure 2 displays their structure. There are three research questions under the research problem. Research question one contains three sub-questions and research question two contains two sub-questions. The first two hypotheses are related to the second sub-question of research question two. The third and fourth hypotheses are associated to research question three.

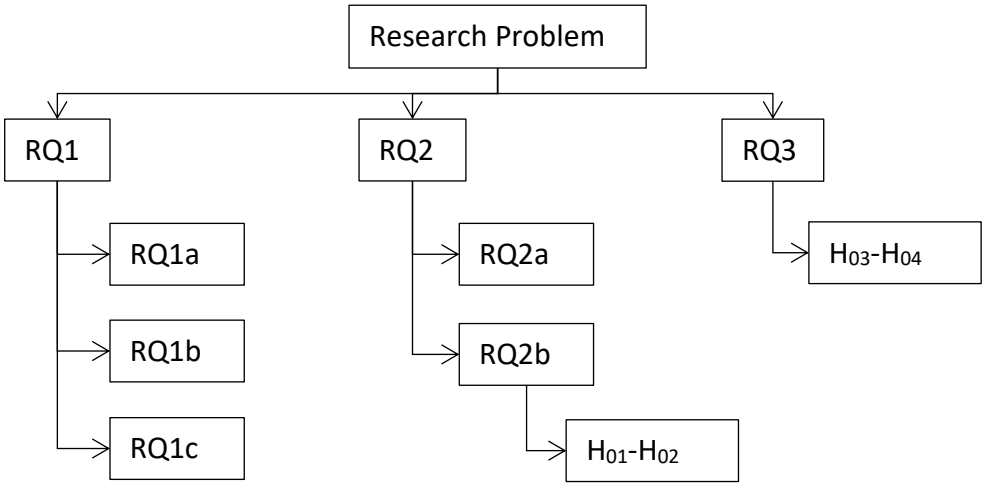


Figure 2. Structure of Research Problem, Research Questions, and Hypotheses

1.3. Research Design

To address the research problem and answer the research questions, a mixed method study was conducted. Three family-health-related topics were selected and the data of their related entries on Wikipedia were collected. Main themes and subjects in different time periods were identified by statistical methods, coding method, and subject analysis methods. How and when the entries and the subjects emerged, grew, and disappeared were explored. Important and popular themes and subjects in different time periods were observed. The external popularity evolution of the selected topics and their themes were demonstrated by temporal analysis methods. Additionally, the different topics' evolution patterns were compared and their commonalities and differences were investigated. To achieve the research goals, both qualitative and quantitative methods, like coding, subject analysis, statistical analysis, clustering analysis, text mining, and information visualization methods, were employed in this study.

1.4. Definitions of Terms

This section is an overview of the terms found in this study and the definitions of these terms. Some of the definitions were proposed in previous research studies, while the remaining ones were defined in this study. These definitions were created to understand how the corresponding terms were considered in the context of this study.

- *Social media*: A group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content (Kaplan & Haenlein, 2010). There are different types of social media, such as blogs, microblogs, RSS feeds, wikis, and so on (Xie & Stevenson, 2014).

- *Wikipedia*: "Wikipedia is a free online encyclopedia created and edited by volunteers around the world." (Wikipedia, 2017)

- *WHO Website*: The WHO Website is the official Website of the World Health Organization (WHO). The primary goal of WHO is to "build a better, healthier future for people all over the world" (WHO Website, 2017).

- *Temporal analysis*: Analysis where temporal data or temporal text play an important role. In this proposal, the concept of temporal analysis includes longitudinal analysis, time-series analysis and trend analysis, but excludes spatial-temporal analysis.

- *Temporal information retrieval*: An emerging research area which aims to bring temporal relevance into information retrieval. Temporal information retrieval considers temporal relevance as important as document relevance in information retrieval.

- *Health information*: Information that refers to general health, drugs and supplements, specific populations, genetics, environmental health and toxicology, clinical trials, and biomedical literature (National Library of Medicine, 2016).

- *Consumer*: People who "seek information about health promotion, disease prevention, treatment of specific conditions, and management of various health conditions and chronic illnesses" (Lewis, Eysenbach, Kukafka, Stavri, & Jimison, 2006, p. 1). In this proposal, consumer only stands for the consumer of health information.

- *Consumer health information*: "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." (as cited in Suess, 2001, p. 41-42)

- *Consumer health vocabulary*: The health terms and phrases used by consumers. Health scientists try to develop consumer health vocabularies to fill the language gap between health professionals and consumers.

- *Family health*: "A dynamic, changing state of well-being, which includes the biological, psychological, spiritual, sociological, and culture factors of individual members and the whole family system" (Kaakinen, Coehlo, Steele, Tabacco, & Hanson, 2014, p.5).

- *Concept*: An abstract or generic idea generalized from particular instances. In this study, the title of every Wikipedia entry represents a concept and the main body of the Wikipedia

entry introduces the corresponding concept. For example, the "Cluster analysis" entry on Wikipedia is a concept related to data mining topic.

- *Theme*: A specific and distinctive concern of a group of concepts. In this study, the concepts collected were assigned to several categories and every category had its own theme. For example, the theme of the data-mining-related concepts "cluster analysis", "odds algorithm", and "genetic algorithm" is "data mining methods."

- *Subject*: In this study, the subject of an article is the focus of it. Every Wikipedia entry could have one or more than one subjects. For example, after reviewing the content of the data-mining-related concept "stellar wind", the subjects of this concept are "security" and "information collection and analysis."

- *Number of page views*: The number of times a Wikipedia entry is viewed by users. These data reflect the popularity of an entry. Daily number of page views is the number of times a Wikipedia entry is viewed by users in one day. Monthly number of page views is the number of times a Wikipedia entry is viewed by users in one month.

- *Number of page edits*: The number of times a Wikipedia entry is edited by users. These data reflect the popularity of an entry. Monthly number of page edits is the number of times a Wikipedia entry is edited by users in one month.

- *Neural network*: "A system composed of many simple processing elements operating in parallel whose function is determined by network structure, connection strengths, and the processing performed at computing elements or nodes" (Widrow, 1988).

- *Self-Organizing Map*: Self-Organizing Map (SOM) is a popular unsupervised learning approach that projects high-dimensional data on to low-dimensional output (Kohonen, 1990). It is a widely used neural network method which measures similarities between items of input data so as to form similarity graphs.

1.5. Chapter One Summary

This chapter presented the background and rationale of the study. The research problem, research questions, hypotheses, and research design were described and addressed in detail. To make this study clear for the reader, the concepts appearing in this study were defined and the connections among them were illustrated in a concept map.

2. LITERATURE REVIEW

In this day and age, because of the development of Web 2.0 technology, the general public is able to create, revise, share, and delete information and communicate with each other on a variety of social media platforms. Being environments for generating user-generated content, social media platforms are not only regarded as the place for information creation and sharing, but also the source for seeking information. For example, YouTube is recognized as the largest video sharing platform for the general public to watch videos. Thousands of entries created on Wikipedia attract users to look for certain knowledge on this Website. Due to the importance of social media, research studies focusing on social media have been increasing in the past decade. Research topics tend to focus on user behaviors on social media, social media content, and social network analysis.

The development of information technology and Internet techniques enables the general public to search health information online. In the early years of Internet use, online health information was mostly posted by health specialists; this is different from the situation today. In recent years, social media platforms serve as a useful channel for general users, patients, and patients' relatives and friends to access information from each other, as well as create health information online. Therefore, social media platforms are becoming more and more important for online health information seeking and the studies which talk about health information on social media have increased. These research studies cover diverse research topics and consumer health and family health are two important topics.

A number of methods have been applied to social media studies and health-related studies. Among the various methods, temporal analysis has played an important role in those research studies. Both qualitative and quantitative research methods have been employed in temporal analysis. As the volume of data increased, data mining methods and text mining methods have been frequently used in temporal analysis. SOM is a widely used method applied to data mining and text mining. This chapter reviews the literature about temporal analysis, health information, social media, and SOM.

2.1. Temporal Analysis

For a long time, social science research relied on cross-sectional analysis which allowed researchers to measure variables and collect data on a single occasion (Fitzmaurice, Laird, & Ware, 2012). Cross-sectional analysis is one of the observational study methods applied to many research fields, especially in the heath and sociology fields (Rosenbaum, 2002). This method provides a way for researchers to observe, compare and analyze certain samples from different populations at one point in time with cross-sectional data. Also, various variables are observed at the same time. As a result, the differences between multiple variables and sample groups are revealed. However, since cross-sectional analysis only offers a snapshot of a single moment, it is not possible to provide definite information of cause-and-effect relationships or to reflect changes or trends of certain subjects. Therefore, in addition to the cross-sectional data, temporal data are utilized in order to take time dimension into account.

As with the cross-sectional analysis, social science's interest in temporal analysis has a long history (Hannan & Tuma, 1979). The needs of exploring changes and patterns of certain

objects, uncovering the cause-and-effect relationships between certain objects, and predicting future trends of certain objects has increased in the past decades. With the development of science and technology, the world today is changing rapidly in all aspects of life (Tran, Gaber, & Sattler, 2014). Accordingly, detecting these changes, describing the patterns, seeking the causes of changes, and predicting the corresponding future trends are key issues in academic and commercial research. In this sense, using temporal analysis in research studies is necessary.

2.1.1. Temporal Analysis

As it was mentioned before, temporal analysis enables researchers to collect data at different time points. In other words, the snapshots of certain objects in different time periods are recorded, analyzed and compared. One of the strengths of temporal analysis is that time interval in temporal analysis is flexible. It means in one study, researchers could observe objects across decades, years, or even seconds (Harzing, 2014; Naaman, Becker, & Gravano, 2011; St. Jean, 2014). Therefore, when the time interval is long, the history and development of certain objects in a long period will be demonstrated. When the time interval is short, changes and patterns in a short time period will be illustrated. Hence, the second strength of temporal analysis is that it reveals how certain objects change and develop over time. Based on detecting the trends or patterns in the past, it is possible to predict the future trends or patterns of the same objects, which is the third advantage. In addition, different from cross-sectional analysis, temporal analysis characterizes the factors which influence changes (Fitzmaurice et al., 2012). Thus, evidences of cause-and-effect relationships can be obtained by this method.

Every coin has two sides. Temporal analysis also has its weaknesses. First of all, to conduct temporal analysis, temporal data need to be collected. This is time-consuming in some cases. Secondly, since the length of the time interval is flexible, for the same research purpose, different timescales might be selected by different researchers (Naaman, 2010). As a result, even if the same data set is utilized, different findings could be obtained depending on difference timescales. It is one of the main challenges for temporal analysis. Last but not least, temporal analysis can only be applied to analyze the same objects with the same time interval. This means it is only capable of capturing within-subject difference rather than between-subject difference (Fitzmaurice et al., 2012). Alternately, using temporal analysis, researchers cannot explore the differences between several sample groups or populations. To make up this shortcoming, some researchers combine temporal analysis and cross-sectional analysis together so as to gather the whole picture of the investigated items (Deursen & Dijk, 2015; Efron, 2013).

Because of the strengths of temporal analysis, it is widely used in multiple research fields, such as business, health, medicine, education, sociology and so on (Ghapanchi, 2015; J. Lin, Wang, Wang, & Lu, 2013; Montiel-Overall & Grimes, 2013; Pivovarov, Albers, Hripcsak, Sepulveda, & Elhadad, 2014). For instance, in the business field, temporal analysis helps to investigate user behavior patterns and to detect product developing trends. In health and medicine, this method allows researchers to figure out factors that cause health issues. In the computer science field, the use of temporal analysis is widely accepted in one of its main branches – data mining.

The applications of temporal analysis in the library and information science field are also indispensable. Temporal analysis is employed in the studies of information retrieval, information visualization, user behavior, informetrics, library science, and so forth. Among the relevant studies of user behavior, temporal analysis provides ways to characterize user behavior patterns and how the patterns change over time (Jain, Rajyalakshmi, Tripathy, & Bagchi, 2013; Pelechrinis & Krishnamurthy, 2012). In informetrics studies that apply temporal analysis, the long-term developing trends of studies, collaboration networks, and citations in different fields are detected (Hossain, Karimi, Wigand, & Crawford, 2015; Peng, 2015; Taba, Hossain, Atkinson, & Lewis, 2015). Regarding the studies in the library science area, temporal analysis deals with both library services and use of library resources and services (Collins & Quan-Haase, 2014; Radford & Connaway, 2013).

2.1.2. Temporal Analysis Applied to Information Retrieval

Both quantitative and qualitative research methods are frequently employed in information retrieval studies. As a widely used research method, temporal analysis is frequently utilized in information retrieval studies. The studies employing temporal analysis primarily focus on search behavior and the information retrieval system. In addition, there is a newly emerging research area, temporal information retrieval, associated with temporal analysis.

2.1.2.1.Information retrieval

Information retrieval is the sum of all ways to answer users' information needs as accurately as possible (Cooper, 1971; Manning, Raghavan & Schütze, 2008). Studies in this area concentrate on either system-oriented or user-oriented aspects.

The research topics of the system-oriented studies usually focus on generating, improving and evaluating information retrieval systems; and the models, algorithms, languages and techniques which are related to them. Search engine, online database, digital library, and online public access catalog are four primary types of information retrieval systems. Three information retrieval models occupy important places among all the models. These are the Boolean Information Retrieval Model, the Vector Space Model, and the Probability Information Retrieval Model (Manning, Raghavan & Schütze, 2008). When the World Wide Web was created, the needs for searching information online emerged along with strong demands of improving search efficiency. New models and algorithms were proposed in the 1990s. PageRank proposed by Brin and Page plays an important role in online information retrieval (Brin & Page, 2012; Page, Brin, Motwani, & Winograd, 1999). Recently, apart from retrieving text information, studies of image retrieval, music retrieval and multimedia retrieval have increased. To provide more user-friendly search systems, new techniques such as information visualization techniques have been applied to optimize interface design. Meanwhile, the development of social media also raises new questions in information retrieval, such as social tagging, temporal information retrieval, and people search (Campos, Dias, Jorge, & Jatowt, 2014; Han, He, Yue, Jiang, & Jeng, 2012; Li, Shan, & Lin, 2011; Wang, Clements, Yang, de Vries, & Reinders, 2010).

Other research studies have been conducted from a user-oriented perspective. By analysis of users' search behavior, two main information retrieval paradigms have been proposed. They are query searching and browsing. User behavior studies have investigated both searching and browsing behavior in different types of information retrieval systems. These studies included users' motivations, preferences, eye movement patterns, information literacy,

and so on (Granka, Joachims, & Gay, 2004; Julien, Tan, & Merillat, 2013; Ross et al., 2009). In recent years, with the improvement of mobile devices, exploring user information seeking behavior on these devices has become a hot topic in the information retrieval area.

2.1.2.2.Information search behavior

Exploring users' information behavior is one of the research topics which most often apply temporal analysis. A part of the studies observe and compare user behavior across a number of years, while the remaining studies explore user behavior patterns within a relatively short time period.

Deursen and Dijk (2015) investigated search skills across different user groups in a longitudinal study. The participants' search skills were examined every year from 2010 to 2013. The results revealed that age and educational background made differences in users' search skills. From 2010 to 2013, search skills of the participants who were older than 65 years old increased. Researchers from health science also conducted a number of studies referring to search behavior. Consumer health information behavior is currently a hot topic in this area. For example, St. Jean argued that combining longitudinal analysis and card-sorting technique was a useful means for observing patients' information behavior, including search behavior (St. Jean, 2014).

In addition, think-aloud protocol, transaction log analysis and eye-tracking method are applied together with temporal analysis for user studies. These three approaches are capable of capturing temporal data, or orders of actions. Thus, they are employed for exploring user's searching and browsing behavior patterns, especially transition patterns. Hölscher and Strube

(2000) utilized the think-aloud protocol to record users' actions and orders of actions in Web-based information search. Based on the orders of actions, they built several transition models for information seeking behavior, and interaction between users and search engines. Goodrum, Bejune, and Siochi (2003) identified 18 states for online image search behavior. They gathered transaction log data and coded each record with the identified states. Depending on the coding results and the temporal data in transaction log, they detected the maximal repeating patterns of state transition. Liu and Kešelj (2007), and Hassan, Jones, and Klinkner (2010) also modeled navigation and search behavior by transaction log data and temporal analysis. Furthermore, they applied the models for prediction. Eye-tracking devices record users' eye movement data, including time and position of fixations, saccades, blinks and clicks. Similar to transaction log data, eye-tracking data are collected for investigating eye movement patterns in information seeking. For instance, Puolamäki et al. (2005) gathered the eye-tracking data of scientific article searching, and estimated relevance from eye tracking data. Then they performed the Hidden Markov Model to detect eye movement and relevance judgment patterns. Moreover, the "scan path" function embedded in eye-tracking software can present the entire process of eye movements (Räihä, Aula, Majaranta, Rantala, & Koivunen, 2005).

Search queries contained in transaction logs reflect interactions between users and information retrieval systems in online search. Search queries not only relate to user behavior, but also associate with information retrieval systems. Temporal analysis of search queries explores users' query formation patterns (Wang, Berry, & Yang, 2003). The way users modify the search queries reflects their mental models in information search. At the same time, the changes of the vocabulary utilized in search queries show the evolution of language use.

These examples show that the use of temporal analysis helps researchers gain insights into users' information behaviors. User behavior patterns and changes of user behavior patterns over time are both detected by this means.

2.1.2.3. Information retrieval system

To improve the effectiveness of an information retrieval system, research studies discussed the possibility of taking the time feature into account in information retrieval, and proposed models and frameworks for extracting temporal information from documents, metadata, social tags, and queries (Alonso, Gertz, & Baeza-Yates, 2007; Radinsky et al., 2013; Ruocco & Ramampiaro, 2015; Uricchio, Ballan, Bertini, & Bimbo, 2013). Not only publication time but also temporal information contained in the content was extracted. With a variety of temporal information, information retrieval systems will be able to retrieve documents on multiple time dimensions.

Evaluation of information retrieval systems is another main research topic. Harzing (2014) explored the coverage of Google Scholar by retrieving studies in chemistry, economics, medicine, and physics from 2012 to 2013. The findings reflected that the volume of documents in each research field retrieved by Google Scholar increased at different paces over time. The expansion of different information retrieval systems was also detected. Winter, Zadpoor and Dodou (2014) reported that Google Scholar expanded faster than Web of Science, but the quality of literature retrieved by Google Scholar was not as good as those retrieved by Web of Science.

Information retrieval studies also strengthen temporal analysis methodology. Since one of the shortcomings of temporal analysis is the variation of indices used in analysis and there is no criterion for selection of these indices, comparing and evaluating the indices are meaningful. Tseng et al. (2009) compared and evaluated a number of trend indices by information retrieval measures. They reported that the linear regression slope performed better than the other investigated indices. This study made a contribution to temporal analysis on both theoretical and practical sides.

The applications of temporal analysis in information retrieval enhance both user behavior studies and system-oriented studies. Conversely, information retrieval research contributes in the methodology of temporal analysis.

2.1.2.4. Temporal information retrieval

Since the volume of information increases quite rapidly in the digital age, new documents, topics and terms emerge every day. The spelling and meaning of terms also change over time. It means documents containing the same terms may refer to totally different themes, and different terms used in queries and documents in different time periods may have the same meaning. Therefore, vocabulary mismatch is a huge challenge for current information retrieval systems. Also, as knowledge quickly updates, the value of documents and information is not only content dependent, but also time dependent. Metzger (2007) considered that timeliness was a key factor, beyond relevance, accuracy, objectivity and coverage, which determines document quality. As a result, retrieving documents and information which match

users' information needs from a time perspective is necessary. Thus, a new area, temporal information retrieval, is generated to bring temporal relevance into information retrieval.

So far, some applications of temporal information retrieval have been created in industry. For instance, YAGO2 provides an interface for searching the temporal and spatial knowledge base from Wikipedia (Hoffart, Suchanek, Berberich, & Weikum, 2013). Time Explorer allows users to search news over time and analyzes how news changes over time so as to predict future news (Matthews et al., 2010). Google Ngram Viewer visualizes the historical trends of the use of particular keywords in books stored in Google Book from 1800 to 2000 (Michel et al., 2011). In addition, Twitter, one of the most popular social networks, lists recently popular hashtags and terms in the Trend panel in order to help users explore hot topics in time. These applications reflect the interest of the public and industry which support the idea that temporal information retrieval is a promising area.

Researchers in academia also show great interest in this area. Their focuses cover query understanding and presentation, temporal ranking, future-related information retrieval, and so on (Campos et al., 2014). Although Efron (2013) argued that cross-temporal search was similar to cross-language retrieval, other researchers have different ideas and have proposed new frameworks and algorithms. Information retrieval framework generally contains four steps, including document processing, indexing, query processing, and ranking documents (Campos et al., 2014). The studies in the temporal information retrieval area also pay attention to topics relevant to these steps. For the document processing step, most of the study in this area is in relation to Web crawling and Web archiving. Internet Archive, started in 1996, is one of the most famous projects (Kahle, 1997). Another project, the Internet Memory Foundation, which

was launched in 2004, also provides open memory of Web pages. Some other studies focus on the use of content of Web archives. These studies either propose methods to recover lost online information or maintain the accessibility of Web pages (McCown & Nelson, 2008; Van de Sompel et al., 2009). Regarding the indexing step, Arikan, Bedathur, and Berberich (2009) suggested that two types of indexes should be created: one index to store text documents, the other to store temporal data derived from the documents. Pasca (2008) proposed an index which contains both dates and text in a fact repository. Different from the previous studies, the two indexes proposed by Matthews et al. (2010) included one for each document and one for each sentence. The studies of query processing could be assigned into two categories, the recency-sensitive queries and the time-sensitive queries (Campos et al., 2014). The former category aims to retrieve the latest documents that are topically relevant, while the latter one attempts to search the documents from a particular time period. Similar to the query processing step, the works of temporal ranking also concentrate on two aspects: recency-sensitive ranking and time-sensitive ranking.

All these studies show that the utilization of temporal analysis provides ways to detect users' information seeking behavior overtime and enhances information retrieval systems by (1) extracting more temporal information from documents and queries, and (2) introducing temporal relevance into information retrieval and generating temporal information retrieval. The implications of using temporal analysis cover both theoretical and practical sides. For the theoretical implications, temporal analysis evokes the emergence of temporal information retrieval, explores the features and development of information retrieval systems, strengthens the studies of user behaviors, and inspires researchers to generate new research methods.

Most of the practical implications rely on the theoretical implications. Researchers propose suggestions for improving effectiveness and efficiency of information retrieval systems by taking temporal relevance into consideration, and modifying search functions and interface design. A unique practical implication of temporal analysis is the invention of the scan path function embedded in eye-tracking devices.

2.1.3. Temporal Analysis Applied to Data Mining

Hey (2012) asserted in his work *The Fourth Paradigm: Data-Intensive Scientific Discovery* that nowadays scientists are overwhelmed with data sets from various information sources. These data sets are obtained by different instruments from many fields, which contain rich information and knowledge. To extract knowledge from data and information is a vital topic in scientific research. However, it is difficult to deal with huge data sets with traditional research approaches. In such a situation, the data mining area emerged to extract hidden information and knowledge from vast amounts of data.

2.1.3.1. Data mining

Data mining is a method to reveal previously unknown and reliable insights from large data sets (Elkan, 2001). Turban, Sharda, Delen, and Efraim (2007) defined data mining as ‘‘the process that uses statistical, mathematical, artificial intelligence and machine-learning techniques to extract and identify useful information and subsequently gain knowledge from large databases” (p. 305). A similar definition proposed by Berson and Smith (2002), Lejeune (2001) and Ahmed (2004) regards data mining as the process of extracting or detecting hidden patterns or information from large databases. Since the massive volume of data from different

fields keeps growing, useful analysis methods and techniques are urgently needed. Therefore, data mining has become an increasingly important research field (Liao, Chu, & Hsiao, 2012).

With the development of data mining techniques, lots of methods from other fields have been introduced to the data mining area, such as generalization, characterization, classification, clustering, data visualization, and database technology (Aggarwal, Kumar, Khatter, & Aggarwal, 2012; Liao, Chu and Hsiao, 2012). At the same time, data mining has been applied to many research fields such as education, medicine, finance, product design, business intelligence, and so on (Aliev, Aliev, Guirimov, & Uyar, 2008; Borghini, Crotti, Pietra, Favero, & Bianucci, 2010; Gregori et al., 2011; Trafalis & White, 2003).

2.1.3.2. Data mining studies applying temporal analysis

Nowadays, the importance of temporal data is growing in various fields, like Web usage monitoring, biomedicine, geography, and finance (Yoo & Shekhar, 2009). Hence, several different types of databases are launched to store time-related data including temporal database, sequence database and time-series database. Mining these data sets helps to reveal the features of object evolution and illustrate the trend of their changes (Han & Kamber, 2006). A few studies aim to explore evolutions of research objects during several periods. For instance, Stevenson and Zhang (2015) studied the changes of institutional repositories from 1992 to 2013. They divided the data of institutional repositories into four periods, which were 1992 to 2001, 2002 to 2005, 2006 to 2009, and 2010 to 2013. For each period, the multidimensional scaling method was employed to obtain term clusters, and the content analysis method was applied to discover the theme of each term cluster. The themes of different time periods were

compared so as to reveal the evolution of institutional repositories from 1992 to 2013. More studies analyze large amounts of ordered data with temporal features, such as time-series data and sequence data. These studies are regarded as temporal data mining studies (Laxman & Sastry, 2006). There are several tasks of temporal data mining, which are prediction, classification, clustering, search and retrieval, and pattern discovery (Laxman & Sastry, 2006). The first four tasks have been conducted in the traditional time-series analysis, while the fifth task is of more recent origin. Since information retrieval was discussed in the previous section, this section focuses on the remaining four tasks.

(1) Temporal clustering

Liao (2005) defined clustering as the ability "to identify structure in an unlabeled data set by objectively organizing data into homogeneous groups where the within-group-object similarity is minimized and the between-group-object dissimilarity is maximized" (p. 1857). Accordingly, temporal clustering aims to cluster sequences or time series based on the similarity among them (Laxman & Sastry, 2006). A number of different methods have been proposed for sequence clustering such as the model-based sequence clustering methods (Law & Kwok, 2000; Sebastiani, Ramoni, Cohen, Warwick, & Davis, 1999).

At the same time, time series clustering became popular. Liao (2005) reported that there were three sets of time series clustering methods, including the raw-data-based methods, the feature-based methods, and the model-based methods. These methods were developed based on the original clustering methods like agglomerative hierarchical clustering, k-Means, fuzzy c-means, Self-Organizing Map, and so forth (Liao, 2005).

Recently, temporal clustering has been applied to a variety of research topics, particularly in event and trend detection and future prediction. Jatowt, Kanazawa, Oyama, and Tanaka (2009) proposed a clustering method to extract and summarize future-related information from text, including content and timestamps. By extracting information from a reference text corpus, Jatowt and Yeung (2011) generated a model-based clustering algorithm to predict future events. In addition, some researchers applied multiple clustering methods in one research study. For instance, Liu, Jiang and Ma (2013) employed both keyword analysis and citation analysis for exploring emerging trends in knowledge networks. Because citation analysis has an obvious time lag effect which may influence the results of temporal analysis, they applied keyword analysis to strengthen their study. The results showed that the approach utilized in this study could also be performed for predicting future emerging trends. Furthermore, topic trends and emerging events on social media are also detected by temporal analysis. Researchers detected trends and events by clustering blog posts, tweets, hashtags, email messages, photographs, and so on (Petrović, Osborne, & Lavrenko, 2010; Rattenbury, Good, & Naaman, 2007; Sayyadi, Hurst, & Maykov, 2009; Zhao, Mitra, & Chen, 2007).

(2) Temporal classification

As a traditional data mining task, classification is a primary task of temporal data mining as well. However, due to the nature of temporal data, special methods and approaches need to be employed for temporal classification (Fu, 2011). Two main topics of temporal classification are sequence classification and time series classification. Certain examples of sequence classification are speech recognition, demarcating gene and non-gene regions in a genome sequence, on-line signature verification, and so on (Laxman & Sastry, 2006). Several methods

and algorithms have been proposed and utilized in time series classification. For instance, Zhang, Ho, and Lin (2004) proposed a nearest neighbor classification algorithm to automatically select parameters from time-series data. Xi, Keogh, Shelton, Wei, and Ratanamahatana (2006) created semi-supervised time series classifiers for small data sets.

This paper mainly investigates the studies focusing on temporal text classification, which is an important and popular current research topic. The temporal information embedded in text can be divided into two categories: (1) timestamp, which is the time when the text was created, published or modified; and (2) focus time, which is the time that the content refers to (Campos et al., 2014). Toyoda and Kitsuregawa (2006) and Nunes, Ribeiro, and David (2007) first studied timestamps of documents. They estimated the Web page creation date by the scores of linking Web pages. Nunes et al. (2007) extracted three features from Web pages - incoming links, outgoing links, and HTML src attributes. They utilized link structure analysis to analyze these data to estimate the last-modified date of Web pages. However, it is unclear whether the results are reliable or not (Campos et al., 2014). To overcome this problem, Web archives, Last-Modified HTTP response header, time when last crawled by Google, and time when first tweets of the investigated documents were collected and utilized to further date these documents (Jatowt, Kawai, & Tanaka, 2007; SalahEldeen & Nelson, 2013).

A few studies aim to determine focus time of documents. Jatowt and Yeung (2011) first detected events from related news by clustering methods and estimated the time when these events took place. Then they compared these events with the events which occurred in certain Web pages in order to estimate focus time of these Web pages. Kawai, Jatowt, Tanaka, Kunieda, and Yamada (2010), and Strötgen, Alonso, and Gertz (2012) presented approaches to

extract the most relevant temporal expressions from text. After the timestamp and focus time of a document was determined, it could be assigned into one or several time periods because different parts of a document were probably associated with different time periods.

(3) Pattern discovery and prediction

Pattern discovery has its root in data mining (Laxman & Sastry, 2006). In temporal data mining, it tends to identify frequent patterns and surprising patterns by using time-series data (Fu, Chung, Ng, & Luk, 2001; Keogh, Lonardi, & Chiu, 2002). Clustering methods are usually employed for pattern discovery, such as distance-based clustering, SOM, and so on (Fu et al., 2001; H. Wang, Wang, Yang, & Yu, 2002). Apart from analyzing temporal sequences of raw data, there is a growing trend of analyzing the results of temporal data mining (Lingras, Hogo, Snorek, & West, 2005). In other words, these studies aim to mine the changes. Change mining monitors models and patterns overtime, compares them, and detect their changes (Böttcher, Höppner, & Spiliopoulou, 2008). Incremental mining methods were generated for investigating the updating of patterns or models.

The studies cited indicate that the five temporal data mining tasks are usually associated with each other in academic research. In Fu's (2011) work, pattern discovery and clustering were considered one task. Actually, clustering methods are frequently applied for pattern discovery. In some cases, temporal clustering supports temporal classification, while clustering and classification both support pattern discovery. In general, prediction is the purpose of pattern discovery. Meanwhile, classification, clustering, pattern discovery and prediction also contribute to information retrieval. Temporal analysis enhances data mining from different

perspectives. First of all, temporal data, such as sequence data and time-series data; are primary information resource for all five tasks. To analyze this type of data for different tasks, old methods are modified while new methods are created. By these methods, researchers are able to extract temporal data from text, cluster and classify items with temporal features, discover patterns, and predict future trends relying on temporal data. Temporal analysis strengthens all five tasks of data mining.

In data mining, temporal analysis has both theoretical and practical implications. Its contribution to research methodologies of data mining studies consists of two aspects: (1) introducing temporal analysis methods generated in other fields, and (2) generating new temporal data mining methods. It also reveals emerging events, detects current trends, predicts future trends, and explores changes of certain research objects over time. Regarding the practical implications, temporal analysis improves the effectiveness and efficiency of information retrieval systems. On one hand, it enhances temporal information search by extracting temporal information from documents. On the other hand, it detects and presents trending topics and events in social media applications, which strengthens information browsing functions embedded in these applications.

2.2. Social Media Studies

The use of social media is increasing globally (Boyd & Ellison, 2007; Moorhead, Hazlett, Harrison, Carroll, Irwin, & Hoving, 2013). The *eBizMBA Rank* reported that by October 2015, Facebook had 900 million unique monthly visitors, Twitter had 310 million unique monthly visitors, and the number of LinkedIn visitors had reached 255 million. The volume of content

increased as the number of users grew. It was reported that Twitter users had broadcasted more than 500 million tweets per day by 2014 (Matta et al., 2014). The statistics offered by Wikipedia presented that currently it has more than 900 thousand content creators generating nearly 5 million articles. In addition to content creation, diffusion, and utilization, social media is the platform for social connections. Social media users connect with each other by becoming friends, following, commenting, tagging and so forth.

The large number of users and the huge volume of content on social media attract attention from different research fields such as business, health, sociology, education, anthropology and so on (Bredl, Hünniger, & Jensen, 2012; Joseph, 2012; Kietzmann, Hermkens, McCarthy, & Silvestre, 2011; Moorhead et al., 2013). Studies focus on users, content, and relationships between users and content on social media.

2.2.1. Social Media

Academic and industry researchers have proposed many notions for the description of the emergence of Websites and applications of user-generated content. For instance, "commons-based peer production" and "produsage" represent the new content creation mode (Benkler & Nissenbaum, 2006; Bruns, 2008). Notions such as social software, social networking, crowd-sourcing, wisdom of crowds, Wikinomics, and collective intelligence also illustrate the phenomenon from different perspectives (Benkler, 2006; Boyd & Ellison, 2007; Coates, 2003; Hermida, 2010; Malone, Laubacher, & Dellarocas, 2009; Surowiecki, 2005; Tapscott & Williams, 2008). However, "social media" is one of the most widely accepted notions.

Ahlqvist, Bäck, Halonen, and Heinonen (2008) defined social media as the way people generate, share information and communicate in virtual communities and networks. Kaplan and Haenlein (2010) gave their definition, which defined social media as "a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content" (p. 61). Recently, Xie and Stevenson (2014) stated that "social media is a means of communication through the Internet that enables social interaction" (p. 502). All of these definitions indicate that (1) social media depends on the Internet and Web 2.0 technologies, and (2) social media is the platform for the general public to generate content and communicate spontaneously.

Under the big umbrella of social media, social media sites and applications vary a lot. For example, Twitter, which is regarded as a microblog, allows users to communicate and create posts less than 140 words (Kwak, Lee, Park, & Moon, 2010); Wikipedia provides opportunities for collaborative information and knowledge production (Bruns, 2006); and YouTube is a multimedia broadcasting Website (Burgess & Green, 2013). Although hundreds of social media sites and applications have been created, new social media applications emerge every day. With the development of mobile devices, geo-mapping tools (e.g. Google Maps) and self-tracking applications (e.g. Qualified Self) have been invented. Each kind of social media has its own features.

The diversity of social media causes difficulty in classifying them. Kaplan and Haenlein (2009) suggested that social media should be classified into six groups according to: (1) "the degree of self-disclosure it requires and the type of self-presentation it allows" (p. 62); and (2) "the richness of the medium and the degree of social presence it allows" (p. 61). The six groups

include blogs, collaborative projects (e.g. wikis), social networking sites, content communities (e.g. YouTube), virtual social worlds (e.g. Second Life) and virtual game worlds (e.g. World of Warcraft). However, Shi, Rui, and Whinston (2013) argued that social media like Twitter should be classified as social broadcasting technology. Two new categories, social bookmarking and social news, were introduced by Agarwal and Yiliyasi (2010). Later on, Xie and Stevenson (2014) reviewed the previous literature and proposed their classification of eight categories. This paper adopts Xie and Stevenson's classification and definitions of different social media applications. Table 1 displays these eight categories, their definitions, and the corresponding examples.

Types	Definitions	Example
Blogs	Allows a user to share thoughts and opinions on subjects in a diary like fashion in a series of posts. Creates discussions or an informational site published online and consisting of discrete entries or "posts."	Blog
Micro blogs	Allows users to communicate with a handle or username that the user creates, and can write short messages, typically 140 characters that are sent to the user's followers.	Twitter
Photo sharing	Online image and video hosting site that allows users to share, comment, and connect through posted images.	Facebook; Flickr; Pinterest; Twitter
Podcasts	Multimedia digital file that is stored on the Internet and is available to download, similar to a radio broadcast that is available freely online.	Podcast
RSS feeds	Rich Site Summary or Really Simple Syndication is frequently updated Web feed that indicates news, events, blog entries that a user can subscribe to and follow. RSS takes current headlines from different Websites and pushes those headlines down to your computer for quick scanning.	RSS feeds
Social networks	Online platform, for users to communicate and connect via interests, backgrounds and activities, which are part of a large social network.	Facebook, Twitter; Reddit
Video sharing	Content distribution of videos, typically available for free to the public.	YouTube

Wikis	Allow users to create and edit Web page content online. Hyperlinks and crosslinks connect between pages. Users are allowed and encouraged to edit wikis.	Wiki

Table 1. Definitions of Different Types of Social Media (Xie & Stevenson, 2014, p. 504)

The definition of each category reveals the specific functionalities of the social media within it. Social media applications in one category, like social networks, share a set of similar characteristics. In each kind of social media, users, their behaviors and their connections are complex. Numerous users play various roles on social media so that their behaviors vary a lot. The users can generate, edit, and delete content, interact with other users or content, and create connections with other users. These behaviors generate huge volumes of data, including numeric data, text data, multimedia data, and so on. The relationships between the data obtained from social media are also complicated, because an action may create several different types of data, a piece of data may relate to more than one object, and different types of data may associate to the same object. Due to complexity of social media users, users' behaviors, users' connections, and data, the research studies concentrating on social media are complicated. A variety of data collection and data analysis approaches are employed in the studies about social media.

2.2.2. Data Collection on Social Media Studies

Social media provides users with opportunities for generating, sharing, seeking, and receiving information in the context of multiuser communication (Kaplan & Haenlein, 2010; Maness, 2006; Moorhead et al., 2013). In the context of social media, user status and activities are complicated. Both quantitative and qualitative research methods have been applied to the social-media-related studies (Bolton et al., 2013). Generally speaking, in quantitative studies,

there are four different types of measurement scales including nominal, ordinal, interval and ratio, while in qualitative studies, verbal data, observation data, document data and visual data are collected (F. Gravetter & Forzano, 2015; Maxwell, 2012). In the social-media-related studies, specific types of data are collected and utilized, including interval data, ratio data, nominal data, ordinal data, text, and multimedia.

2.2.2.1. Data integration

In general, different types of data reflect characteristics of research objects from different aspects. As mentioned before, different types of data, including numeric data, non-numeric data, text and multimedia, are usually integrated in one study.

The nature of social media calls for the combination of different types of data, because in social media different types of data are created together regularly (e.g. online product reviews, videos on YouTube, and online maps). For instance, in order to detect the factors that impact the helpfulness of product reviews, Mudambi and Schuff (2010) selected six products on Amazon.com and captured all the reviews. They collected: (1) the star rating (1 to 5) the reviewer gave the product; (2) the total number of people that voted in response to the question, "Was this review helpful to you (yes/no)?" (p. 191) (3) the number of people who voted that the review was helpful; and (4) the word count of the review (Mudambi & Schuff, 2010). Different data were used as different variables. In addition, product type was another independent variable. Then all these variables were utilized in the Tobit regression. This study implies that different types of data can be analyzed by the same approach. Sometimes, the data obtained need to be transformed for data analysis. Gilbert and Karahalios (2009) extracted

users' education background from their Twitter profile and categorized those data into three categories. For some studies, one data collection approach or one data analysis method is not enough. For example, He, Zha, and Li (2013) manually saved the data (both textual and numeric data) from Facebook and Twitter about three pizza chains. Both SPSS Clementine text mining tool and NVivo 9 were adopted for content analysis in the study.

In addition to data obtained from social media applications, data collected from participants by survey, questionnaire or interview are also utilized. For instance, Gilbert and Karahalios (2009) collected survey data to supplement the findings concluded from data obtained from Twitter. They recruited 35 participants and collected the data of their friends on Facebook, such as the number of messages, intimacy words, mutual friends, groups in common, and so on. Meanwhile, the participants were asked to rate the strength of their friendships on Facebook. Based on those two data collection approaches, they identified 74 predictors of tie strength. The results showed that those predictors successfully predicted strong and weak ties over 85% of the time with a certain data set containing more than 2000 Facebook posts. Ross, Terras, Warwick, and Welsh (2011)investigated in what ways an academic community made use of Twitter. They analyzed more than 4000 tweets with open coding and text mining to detect user conventions. At the same time, they undertook a small qualitative survey so as to ascertain users' attitudes towards using a Twitter enabled backchannel in conference. In this study, the data for both quantitative and qualitative analyses were integrated. These studies show that although data collected directly from social media applications reveal some characteristics of certain research objects; usually they do not reflect participants' perceptions or motivations accurately. Traditional research methods such as questionnaire and interview

allow researchers to gain insights into participants' perceptions and motivations. Therefore, to ensure the validity and reliability of research findings, various data collection approaches are applied, and different types of data are integrated.

All these examples indicate that different types of data complement each other in research studies. There are a variety of approaches available for collecting different types of data from social media applications.

2.2.2.2. Data collection methods

(1) Publicly available datasets

So far, a good deal of datasets containing social media data is accessible for researchers (Paltoglou, 2014). These datasets contain various types of information, such as reviews, comments, tweets, and so on. A particular example is the ICWSM Spinn3r Dataset (Burton et al., 2009; Burton & Soboroff, 2011). This dataset contains several million blog posts scraped by Spinn3r. In addition, in Blitzer, Dredze, and Pereira's (2007) paper, they offered a new dataset including Amazon product reviews for four types of products. The four types were books, DVDs, electronics and kitchen appliances (Blitzer et al., 2007). Paltoglou, Thelwall, and Buckely (2010) provided two datasets for textual sentiment analysis in their paper. One dataset consisted of the information from the BBC Message Boards which contains users' opinions about "ethical, religious and news-related issues" (Paltoglou et al., 2010). The other dataset includes information from a social network site, Digg. Both datasets were used in research studies (Chmiel et al., 2011; Mitrović, Paltoglou, & Tadić, 2011; Thelwall, Buckley, & Paltoglou, 2011).

(2) Unique data collection approaches

A lot of unique tools have been created for collecting social media data. Some of the tools are browser extensions or plugins. Particular examples can be seen in the Firefox extension Greasemonkey and the Chrome plugin NCapture. Gilbert and Karahalios (2009) utilized the Firefox extension Greasemonkey to randomly select participants' Facebook friends. Furthermore, the Greasemonkey enabled the researchers to add survey questions on user's personal Facebook homepages. In this way, the researchers guided them to rate the tie strengths between themselves and their friends. The NCapture provides ways to capture Facebook wall posts and comments, Twitter content, and LinkedIn group discussions. However, this tool needs to be utilized with NVivo for content analysis and cannot work independently. In addition to extensions/plugins, there is a variety of software available. NodeXL, an add-on to Excel, offers another approach to obtain social media data. It is capable of accessing and gathering data from Outlook, Twitter, Facebook, Flickr and YouTube (Hansen, Shneiderman, & Smith, 2010). For Facebook, it is able to collect fan lists, group discussion content, and timeline data of certain users. For Twitter, it can capture the data of both user networks and tweets. Apart from user networks, video information on YouTube and tags on Flickr can be obtained by NodeXL as well. This tool is widely applied to social network analysis.

Facebook, Twitter, Yahoo! Answer, and some other social media allow researchers to access and obtain the data in their databases by using APIs (Jansen, Zhang, Sobel, & Chowdury, 2009; Li, Lei, Khadiwala, & Chang, 2012). In addition to these social media applications, some online services also provide accesses to data on social media. Jansen et al. (2009) used the Summize4 to collect tweets. The Summize4 was a service for searching tweets, and it also provided an API for data collection.

APIs are required to be employed with tools like Python or r (Meyer, Hornik, & Feinerer, 2008; M. A. Russell, 2013). Python and r are two of the most popular open source tools which are frequently utilized to create scripts for collecting data from social media. For example, Lipizzi, Iandoli, and Ramirez Marquez (2015) created a Python script to download tweets over time and store them in their own database.

Browser extensions/plugins and software such as NodeXL have more user-friendly interfaces, and are easier for researchers to use than Python and r. However, they are not as flexible and powerful as the latter. APIs also have limitations. For instance, Twitter only allows one hundred API calls per hour for one account. Therefore, all the approaches have their own strengths and weaknesses.

2.2.3. Data Mining Applied to Social Media Studies

Both qualitative and quantitative methods have been employed in social media studies. For instance, descriptive statistical methods and inferential statistical methods (e.g. t-test, ANOVA, Pearson's correlation) are frequently applied to these studies. Among the various quantitative methods used, data mining methods occupy an important position.

2.2.3.1. Web mining

Web mining as a branch of data mining is playing an increasingly important role in research. Web mining can be classified into three different types, Web content mining, Web structure mining and Web usages mining (Singh & Singh, 2010). Web content mining analyzes the content of Web resources and deals with knowledge and information discovery from Web pages. For example, Hosseini and Abolhassani (2007) used Query-ERL co-clustering to group

queries and URLs. Web structure mining aims at generating a structural summary about Websites and Web pages. Moussiades and Vakali (2009) extracted the implicit structures hidden in the Web hyperlink connectivity with graph clustering algorithm. Web usages mining explores user behaviors online, especially users' interactions with Websites (Singh & Singh, 2010).

2.2.3.2. Social Web mining

Social Web mining is one of the primary components in the studies related to social media. The methods applied to these studies are mostly based on information diffusion in social networks and semantic analysis of content published on social media sites (Lipizzi et al., 2015; Passmore, 2011; Saif, He, & Alani, 2012). The methods and approaches in the former category are created for social network analysis. The latter category includes applications like sentiment analysis and text mining (He et al., 2013; Saif et al., 2012). In the investigated research studies, sentiment analysis and text mining are interconnected and usually utilized together.

(1) Social network analysis

Social network analysis has its root in network science, and tries to reveal human relationships and connections (Hansen et al., 2010). With the wide spread of social media, the ways for creating and maintaining connections among family members, friends, classmates, colleagues and other people have changed in the past decades. Social media enables users to manage social connections easier than before. The use of social media generates large volumes of social network data. These data have high value in both academic research and practice

(Smith et al., 2009). In recent years, various tools have been invented to analyze and visualize social networks.

NodeXL UCINET, Pajek, and Gephi are popular social network analysis tools (Bastian, Heymann, & Jacomy, 2009; Borgatti, Everett, & Freeman, 2002; Hansen et al., 2010; Kolaczyk & Csárdi, 2014; Nooy, Mrvar, & Batagelj, 2011). All of these tools allow users to import data, visualize networks, change the layout of networks, and calculate metrics of networks. In this way, roles and importance of nodes in networks, interactions between nodes, and properties of whole networks are revealed. Meanwhile, the tools also enable users to cluster nodes in networks. These tools have user-friendly interfaces, and are easy to use. However, the algorithms embedded in these tools are pre-determined and cannot be changed by researchers. In addition, these tools can only visualize static social networks.

Other tools, such as NetworkX in Python and igraph in r, are embedded in programming environments (Kolaczyk & Csárdi, 2014). These tools provide opportunities for users to manipulate the visualization of social networks and the calculation of metrics. Moreover, they offer means to draw and manipulate dynamic social networks (Kolaczyk & Csárdi, 2014). Therefore, these tools are more flexible in social network analysis, but using them requires knowledge in programming. The more a user knows about programming, the better s/he can use these tools.

(2) Sentiment analysis and text mining

Text mining aims to solve the crisis of information overload (Feldman & Sanger, 2007). It is widely utilized in multiple fields like business, sociology, health sciences, and so on (Bansal &

Koudas, 2007; Eynon, Schroeder, & Fry, 2009; Jansen et al., 2009). Techniques in the areas of data mining, machine learning, natural language processing and information retrieval are combined and applied to text mining (Feldman & Sanger, 2007; Wiebe, Wilson, & Cardie, 2006). In general, text mining includes several steps: tokenizing, filtering, lemmatizing, and creating matrix (Lipizzi et al., 2015). Sentiment analysis is known as opinion mining, which is related to text mining (Thelwall et al., 2011). Therefore, methods for text mining are also employed in sentiment analysis.

Sentiment analysis aims to predict opinion or emotion from content by consistent, repeatable, algorithmic approaches (Hill, Dean, & Murphy, 2013). It focuses on both sentiment polarity and sentiment strength of content (Thelwall et al., 2011). Currently, several commercial platforms conduct sentiment analysis to assess customers' judgments by analyzing their tweets (Cognizant, 2014; Crimson Hexagon, 2014). Social media also provides means for users to seek content based on the emotion embedded. For instance, Twitter allows users to search tweets in positive or negative mode; Amazon.com asks consumers to rate their satisfaction by a five-point scale and enables users to filter results by customer reviews. Additionally, a number of tools are available for automatically extracting opinions and emotion. For instance, the Summize4 analyzes tweets and rates them by a five-point Likert scale, from wretched to great (Jansen et al., 2009). The SAS Text Miner can automatically uncover and place into categories hidden themes from large document collections and group documents (Miner, 2012).

Identifying opinion-expressing terms in texts is one of the major tasks in sentiment analysis. In the past, the General Inquirer lexicon was created by Hatzivassiloglou and McKeown (1997) to retrieve semantic orientation information. The Multi-perspective Question Answering

(MPQA) Opinion Corpus, which focused on questions and answers, was described by Wiebe et al. (2006). Based on previous studies, Wilson, Wiebe, and Hoffmann (2005) proposed the OpinionFinder lexicon to determine the emotion (neutral or polar) of an expression and disambiguate the polarity of the polar expressions. Recently, several dictionaries, for example, the Linguistic Inquiry and Word Count (LIWC) dictionary, have been compiled for sentiment analysis (Gilbert & Karahalios, 2009). The lexicon-based classifiers were established based on these lexicons and dictionaries (Paltoglou, 2014).

Moreover, machine learning algorithms and natural language processing approaches are applied to sentiment analysis. Tools like LingPipe, Mallet and ApacheOpenNLP are available for opinion mining (Alias-i, 2008; McCallum, 2002). Open source tools like Python, Weka and r not only offer opportunities for data collection, but also contain functionalities of machine learning and natural language processing (Paltoglou, 2014). To strengthen the power of these tools and to make them easier to use, thousands of researchers have created "packages" for these tools. For example, among the packages for r, the "tm" and "Textir" packages focus on text mining, while the "rattle" package provides a number of approaches for hypotheses testing, clustering and modeling (Meyer et al., 2008). The machine learning algorithm extensions of Weka also provide functions for text mining.

Although many algorithms have been proposed for detecting sentiment polarity, only a few studies focus on sentiment strength detection (Pang & Lee, 2005; Strapparava & Mihalcea, 2008; Wilson et al., 2005). Neviarouskaya, Prendinger, and Ishizuka (2007), and Strapparava and Mihalcea (2008) proposed their own algorithms for detecting sentiment strength in text. Thelwall, Buckley, Paltoglou, Cai, and Kappas (2010) presented the SentiStrength algorithm to

detect sentiment strength of short text. Since the content on the majority of social network sites is short, this algorithm is widely applied. The SentiStrength classifies text for both positive and negative sentiment on a scale of 1 to 5. However, the accurate detection of sentiment is domain-dependent (Thelwall et al., 2011). It means the same term may represent various sentiment polarities or sentiment strength in different domains or contexts. Thus, using the same lexicons or algorithms for data obtained from different domains or contexts will cause bias in findings.

Approaches invented in other areas have also been introduced to sentiment analysis. For instance, the action-object pair approach was utilized to analyze textual data such as tweets and posts (Jansen et al., 2009). This approach was proposed by Zhang and Jansen (2008), who originally aimed to extract the relationships between users and actions from the transaction log.

2.2.4. Coding Methods Applied to Social Media Studies

Coding is the main categorizing strategy for qualitative data analysis (Maxwell, 2005a). It intends to segment data, rearrange them into different categories, and compare those categories in order to develop theory. Several computer-assisted tools have been invented for coding, such as NVivo, ATLAS/ti, MAXQda, Cassandre and Transana (Bazeley & Jackson, 2013; Seale, Gobo, Gubrium, & Silverman, 2004). NVivo is one of the most popular tools. It allows researchers to analyze different types of data, including Web pages, images, videos, tweets on Twitter, posts on Facebook, and so on. Furthermore, users can import and export data, create

and organize nodes and categories, visualize coding results, and produce reports of results (Bazeley & Jackson, 2013).

The investigated social media studies reveal that different types of data have been collected from social media and manifold data analysis methods have been adopted for various research purposes. Coding and data mining methods are the representative data analysis method utilized in the social media studies.

2.3. Health Information Studies

According to the official Website of US National Library of Medicine, the range of health information covers seven categories, including general health information, drugs and supplements, specific populations, genetics, environmental health and toxicology, clinical trials, and biomedical literature. Each category contains plenty of health topics and the information associated to these topics is recognized as health information no matter if it is posted on traditional media or social media platforms. During the past decades, health information has held the general public's attention as the quality of people's life improves. A new concept emerged in response to this situation, which is consumer health information.

2.3.1. Consumer Health Information

Consumers of health information are defined as the people who "seek information about health promotion, disease prevention, treatment of specific conditions, and management of various health conditions and chronic illnesses" (Lewis, Eysenbach, Kukafka, Stavri, & Jimison, 2006, p. 1). This definition shows that not only the patients or their families and friends, but also other people who have interests in health information and seek health information are the

consumers of health information. Therefore, the consumer group of health information is quite huge. Deering and Harris (1996) proposed several dimensions, such as age, disability, race and ethnicity, and gender, for identifying different consumer groups.

On the other hand, Patrick and Koss proposed the definition of consumer health information as "any information that enables individuals to understand their health and make health-related decision for themselves and their families" (as cited in Suess, 2001, p. 41). In their opinion, consumer health information 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" (as cited in Suess, 2001, p. 41-42). Agreeing with Patrick and Koss's work, Deering and Harris (1996) further summarized that there are three broad purposes of consumer health information: personal health, medical treatment, and public health.

2.3.2. Health Information on Social Media

According to Tu's (2011) report, the proportion of people among all the consumers, who sought health information online increased from 2001 to 2010. This report presented that in 2001 the proportion was 15.9% and it rose to 31.1% in 2007 and finally reached 32.6% in 2010. To the contrary, the proportion seeking health information through book, magazines, and newspapers dropped from 32.9% to 18.2% from 2007 to 2010. Similarly, the proportion seeking health information through TV or radio reduced from 15.6% to 10.0%. A survey in 2013 reported that 87% of US adults use the Internet and among them 72% stated that they sought

health information online during the past year (Street, NW, Washington, & Inquiries, 2013). These statistics reveals that the use of Internet for seeking health information increased rapidly in the past sixteen years.

The emergence of Web 2.0 advocated the creation of social media, which changed the way of health care interaction between health organizations and individuals (Moorhead et al., 2013). Conceptions like Health 2.0, Medicine 2.0, and Science 2.0 were generated thereby (Eysenbach, 2008). Eysenbach (2008) identified five aspects of the Web 2.0 in health, health care, medicine, and science. These were social networking, participation, apomediation, collaboration, and openness.

With the broad adoption of social media applications, the general public, patients, and health professionals started to communicate about health issues via this new channel (McNab, 2009; Thackeray, Neiger, Hanson, & McKenzie, 2008). About 45% of the hospitals in Norway and Sweden use LinkedIn and 22% of the hospitals in Norway use Facebook (Heidelberger, 2011). The purposes that health professionals use social media include facilitating communication, increasing skills, and increasing knowledge (Hamm et al., 2013). The information sources used by patients for seeking health information are diverse but less authoritative than those used by health professionals. Dawson (2010) reported that according to his survey, Facebook was ranked the fourth health information source and YouTube and Twitter were also prevalent sources. Additionally, 81% of European consumers and 63% of US consumers trust the health information on social media applications. Wikipedia is the most popular health information source with Italians and Spanish consumers.

The health-related information on social media covers a wide range of topics, including diseases and treatments, nutrition, health care and insurance, healthy lifestyle, and so forth. Furthermore, social media allows users to develop their own health stories, interact with each other, and search for health topics online. Family-health-related topics are a main category among the various health-related topics online.

2.3.3. Family Health Studies

The health risks faced by the general public are numerous. In the latter age of the twentieth century, health-related research studies focused on the treatment of illnesses, injuries, and infectious diseases (Halfon, Larson, Lu, Tullis, & Russ, 2014). As the evidences of the influence of social factors and behavior on human health accrued, new determinants of the general public's health emerged, such as lifestyle, families, social services, and environment (Fuemmeler et al., 2017; Halfon et al., 2014). Among these determinants, the role of families is crucial because they are considered as a unit of health care (Kaakinen et al., 2014).

Kaakinen et al. (2014) defined family as "two or more individuals who depend on one another for emotional, physical, and economical support" (p. 6). According to the previous literature, the concept of family health is used interchangeably with the some terms such as family functioning, familial health, resilient families, and balanced families and therefore, the definition of family health various (Kaakinen et al., 2014; Kim, Kim-Godwin, & Koenig, 2016). Based on the definition of health given by the WHO, Craft-Rosenberg and Pehler (2011) defined family health as "the state of physical, mental, and emotional wellness of the family system and its individual members" (p. xxxii). Kaakinen et al. (2014) proposed a similar definition which

centers on some family-health-related factors. Their definition is that "family health is a dynamic, changing state of well-being, which includes the biological, psychological, spiritual, sociological, and culture factors of individual members and the whole family system" (p.5). From another perspective, Bomar (2004) defined this term as "a holistic state referring to the complex process of negotiating and solving day-to-day family life events and crises and providing for a quality life for its members" (p.10). In this study, the definition of family health references the definition from Rosenberg and Pehler's book.

Worldwide health agencies have conducted plenty of surveys to collect family health data and information. For example, the Ministry of Health and Family Welfare of India designated the International Institute for Population Sciences (IIPS) to conduct the National Family Health Survey for India from 1992. The purposes of the survey are (1) collecting health and family welfare data for policy formulation and (2) seeking information of emerging health and family welfare issues (IIPS, 2017). Other countries like Iraq, Philippine, and Syria conduct their own family health surveys for similar purposes. Additionally, there are specific organizations providing technical assistance for family health survey, such as the Demographic and Health Surveys Program, which has supported family-health-related surveys in more than 90 countries. In the United States, some statewide surveys were completed by state level departments. For instance, the Wisconsin Department of Health Services surveys the Wisconsin residents every year and reports key findings of the surveys.

The surveys mentioned before usually contain data and information of household characteristics, fertility, infant and child health, family planning, family health insurance, poverty status, and so forth; topics which refer to various health-related topics. The family

health research studies also cover a wide range of research topics. A number of studies explore the factors that influence family health. The Circumplex Model of Marital and Family Functioning proposed by Olson (2000) has three dimensions: marital and family cohesion, flexibility, and communication; whether or not a family is healthy can be assessed by the three dimensions. Olson argued that a family with balanced levels (neither too high nor too low) of the three dimensions was healthier than that with unbalanced levels. Moreover, the traits of a healthy family are proposed by other researchers, like communicating and listening, affirming and supporting each other, developing a sense of trust, sharing a religious core, and so forth (Kaakinen et al., 2014). Black and Lobo (2008) emphasized optimal growth, functioning, and well-being of families and the interactions among these factors.

Apart from the factors influencing family health, the family-health-related studies often center on particular groups like infants and children, adolescents, women, and so on. Fuemmeler et al. (2017) reviewed the literature related to child health and provided recommendations for ensuring good health of children. They found that childhood obesity and childhood trauma and violence, two outstanding issues, influence child health at the population level and potentially impact health over the life course. Childhood obesity is caused by pre-pregnancy obesity and gestational weight gain, tobacco use during pregnancy, and contextual factors such as home and school environment. Moreover, evidences show that breastfeeding protects against childhood obesity (Hansstein, 2016). Childhood trauma and violence is caused by prenatal maternal stress, adverse childhood events, and family and school context.

Regarding adolescent health, DiClemente, Hansen, and Ponton (2013) listed 12 health risks: tobacco use, disordered eating, alcohol use, drug use, suicide and suicide behavior,

unintentional injury, delinquency, adolescent violence, adolescent pregnancy, sexually transmitted diseases, runaway and homelessness, and academic underachievement and school refusal. They concluded that family is a crucial factor impacting those health risk behaviors. Most adolescents have close relations with their parents and are influenced by their parents when determining whether to engage in risk-taking behavior or not. For instance, adolescents who receive more emotional support and acceptance from parents are less likely to abuse substances and have sexual activities (Turner, Irwin, Tschann, & Millstein, 1993). Family structure also influences adolescents' behaviors. Flewelling and Bauman (1990) presented that adolescents from single-parent families tended to abuse substances more than those from intact families.

Women are another particular group which gains attentions from the family health researchers. According to Kotch's (2005) book, age at first marriage and age at first pregnancy for women both increase from 1970s. The proportion of single parent (mostly female) families grows because of high divorce rate and the increase of unmarried mothers. These phenomena lead to the family-health-related research on pregnant women and single mothers. Pregnant women face various health risks, including virus infections (e.g. dengue and ZIKA), physical issues (e.g. overweight and obesity), psychological issues (e.g. depressive symptoms), and so on (Brasil et al., 2016; Marcus, Flynn, Blow, & Barry, 2003; Oteng-Ntim, Tezcan, Seed, Poston, & Doyle, 2015). Compared with married mothers, single mothers face higher risks of physical and mental disorders (Subramaniam, Prasad, Abdin, Vaingankar, & Chong, 2017). Furthermore, since single mothers generally have less work experience and education than their peers, they are less competitive in the labor market (Damaske, Bratter, & Frech, 2017). This shortcoming

increases the single-mother household poverty rate. During 2006 to 2008 the single-mother household poverty rate was 45% in the United States (DeNavas-Walt, 2010). Lack of education and poverty both lead to poor access to health care resources, which will cause potential health risks for single mother families (Cohen et al., 2014).

Different groups of people face different health problems, including mental and physical, and internal and external problems. Therefore, plenty of research topics emerge for the family health researchers to explore the factors that cause the problems and suggest the solutions to the problems.

2.4. SOM Studies

2.4.1. SOM History, Theories, and Algorithms

The initial idea of self-organization was proposed by Willshaw and Malsburg (1976) for explaining the formation of neural connections during ontogenesis. Their theory was extended by Takeuchi and Amari (1979) and the convergence properties and dynamic stability of the previous model were improved. Then Kohonen (1982) revised the previous models and proposed a simpler and more practical SOM algorithm. A number of SOM models were developed based on the theories and original models; these include the Growing Neural Gas model and the Evolving SOM model (López-Rubio & Díaz Ramos, 2014).

The SOM introduced by Kohonen (1982) is an unsupervised learning approach which transforms a set of high-dimensional data to a low-dimensional feature map. It is a widely used neural network method that measures similarities among items of input data so as to forms similarity graphs. The whole procedure of this approach is a recursive regression process

(Kohonen et al., 2000). Zhang (2007) reviewed the history and improvement of the SOM approach in his book, as well as the SOM theories, models, and algorithms.

There are two basic learning algorithms for SOM, which are sequential learning and batch learning (J. Zhang, An, Tang, & Hong, 2009). After comparing these two algorithms, Fort, Letremy, and Cottrell (2002) concluded that advantages of the batch learning algorithm are "simplicity of the computation, quickness, better final distortion, no adaptation parameter to tune", and "deterministic reproducible results", while the disadvantages include "bad organization, bad visualization, too unbalanced classes", and "strong dependence of the initialization". Another advantage of the batch learning algorithm is that it is not impacted by the convergence problem (Ding & Patra, 2007).

Regarding the output of the SOM approach, the Unified Distance Matrix (U-matrix) display and the component plane display are two prevalent types of SOM displays (An, Zhang, & Yu, 2011). The U-matrix display proposed by Ultsch and Siemon (1990) aims to reveal the differences and similarities among the weight vectors in the output display. The values of a U-matrix represent the Euclidean distances among the weight vectors. Low values stand for clusters, while high values stand for cluster boundaries. The U-matrix is visualized in the final display to show the clusters. Component planes are significant visual representations in the SOM literature (Kohonen, 1995). A component plane reflects the contribution of a specific variable of the input matrix and reveals the features of the items in the matrix and the connections among the items in terms of this variable. The component plane can also be converted to various colors in the final display.

2.4.2. Applications of SOM

As it is stated in Chapter One, the SOM approach has been used in many fields, such as finance, industry, and biology. Apart from the fields mentioned before, health science is another research field that adopts this approach frequently. The health-related research studies usually apply the SOM approach to data clustering. For instance, Carboni and Russu (2015) collected the wellbeing and life quality data of different regions in Italy and clustered the regions in terms of the data obtained by the SOM approach. The findings show that regional differences in wellbeing exist in Italy. Chakraborty et al. (2006) analyzed life-style data and found that sleep, exercise, and mind are related to each other for people's health status.

In the field of library and information science, the SOM approach is mainly employed for user behavior exploration, information system optimization, and information extraction. An example of an information extraction tool is the framework SOFIE, which extracts new information from text documents (Suchanek et al., 2009).

For the user-oriented studies, Stenmark (2008) analyzed the search engine log data with a clustering method which was based on self-organizing maps in order to group search engine users into three groups: casual seeker, users applying more holistic approach, and information-seeking savvy employees. Ding and Patra (2007) proposed a user modeling method which revealed users' preferences by clustering their search queries. These methods and models provide ways to developing personalized Web searches for information retrieval systems.

For the system-oriented studies, the SOM approach is usually used for visualization and algorithm optimization. When used for document cluster analysis, the SOM approach is able to

demonstrate and visualize the distributions of documents in a collection (Zhang, 2007). For instance, Lagus, Honkela, Kaski, and Kohonen (1996) generated WEBSOM for visualizing similarity relations among documents. VisualLink, created by White, Lin, and Buzydlowski (2004), is an information visualizer for presenting bibliographic records. To optimize positioning systems, Takizawa et al.(2006) proposed a self-organizing location estimation method. Compare with other location estimation methods, this method avoids using a large number of sensors and is more accurate.

Some research studies are helpful for improving library service. An et al. (2011) employed the SOM approach to cluster journals in library and information science in terms of their indexed subjects; he obtained 19 clusters for the 60 investigated journals. Based on the findings, they suggest that both quality and subject coverage of journals should be considered when librarians are determining which journals to purchase.

Zhang, An, and their colleagues conducted a series of studies relevant to both health and information science. They focused on health subject directory and users' traversal activities and explored the association between health subject directory Web pages in terms of users' browsing activities (J. Zhang & An, 2010; J. Zhang et al., 2009). They collected subjects and path data among them from a health subject directory. The path data recorded how users browsed subjects one after another. The path data of all the investigated subjects were analyzed by the SOM approach and they revised the U-matrix algorithm so as to achieve the best results. The new U-matrix algorithm performed well in distinguishing irrelevant subjects. The findings of these studies demonstrated that the users' traversal activities can reveal the relations among the investigated health subjects. All these studies show that SOM is not only used as data

analysis methods but also is regarded as a useful algorithm for system optimization and model development.

2.5. Chapter Two Summary

This chapter reviewed the previous research studies relevant to temporal analysis, social media, health information, and the SOM approach. The literature review reveals that the needs of health information have grown during the past decades. With the development of Internet technology, more and more general users tend to seek health information online. Social media sites are popular health information sources for the general public. In the large volume of health information online, family health information is a major category.

Since social media sites contain huge user-generated content and users' interaction data, they are regarded as research objects and data sources for researchers. Different types of data (e.g. numerical data, textual data, and multimedia data) have been collected from social media sites (e.g. Facebook, Twitter, YouTube, and Wikipedia) and various data analysis methods have been employed to analyze the data. Both qualitative methods (e.g. coding) and quantitative methods (e.g. statistical methods and data mining methods) are frequently used in the social media studies. Among the quantitative methods, SOM is a unique unsupervised learning approach for reducing high-dimensional data to low-dimensional presentations. This approach has been widely used in information science, health science, biology, finance, and many other fields.

Temporal analysis, a type of data analysis methods handling temporal data and temporal text, is widely applied to research studies. This method can be used to analyze the

rich temporal data and text (e.g. data of Twitter trends and historical versions of entries on Wikipedia) contained on social media. However, there are few research studies that employed this approach to analyze Wikipedia temporal data and text. To fill the gap, this study explores the temporal changes and evolution patterns of entries on Wikipedia.

3. RESEARCH METHODOLOGY

3.1. Introduction

The volume of data in the world keeps growing (Manyika et al., 2011). With the development of Web 2.0 technique, information increased more rapidly and was shared much wider and faster than in previous decades. New terms emerged and meanings of existing terms changed. For instance, “Google” was initially created as the name of a search engine, but currently it is also used as “search”. Along with the change of a term, the concepts having relations to it also changed. For example, the term “cloud” has been used for hundreds of years to represent the weather phenomenon and its related concepts are mist, coalescence, weather lore, and so forth. In the past decades, with the advance of computer science, “cloud computing” became a popular topic in computer science. Meanwhile, terms such as information security and Web computing became related to “cloud”. The changes of relevant concepts cause problems in information retrieval. Returning relevant results in different time periods of a fast-changing topic in various information retrieval systems becomes one of the biggest challenges in the information retrieval field. To solve this problem, it is necessary to explore how topics and concepts change overtime.

Social media, generated and based on the Web 2.0 technologies, plays an important role in information creation and dissemination. Thousands of users from all over the world create, share, and seek information on social media. The information of certain topics on social media platforms reflects the general public’s interests and perceptions of those topics. The changes of specific topics and concepts on social media are much faster than on those in conventional

media. Therefore, changes of topics are more apparent on the former than the latter. It makes social media a proper place for observing and detecting how topics change. Moreover, since users who seek information on social media usually tend to satisfy their temporal information needs, it requires the information retrieval functions embedded in some social media applications to combine the time dimension into relevance judgment. Exploring temporal features and evolution patterns of terms, concepts, and topics will contribute to information retrieval. In addition, the evolution patterns of concepts and topics on social media reveal the changes of the general public's perceptions.

In recent years the public pays more attention to health issues than in the previous years. The improvement of Internet techniques leads to increasing use of online information by the public. The proportion of people who seek health information online has grown during the past sixteen years. As a platform for user-generated content, social media applications provide a new way for health professionals, patients, and general users to create, revise, share, and seek health information online, and communicate with each other. Currently social media is an important health information source for the public.

The health information on social media covers a variety of topics. Family health is a popular topic among the public because family is now considered as a crucial factor affecting human health. This study aims to examine and uncover the evolution patterns of family-health-related topics on social media Wikipedia. The research questions are:

RQ1: What are the associated entries, emerged themes and subjects, and relations among them in each of the selected family-health-related topics?

RQ1a: What are the associated entries and the main themes of the selected family-health-related topics discussed on Wikipedia?

RQ1b: What are the subjects of each theme of the selected family-health-related topics?

RQ1c: What are the relations among the themes, subjects, and entries?

RQ2: What are the evolution patterns for each of the selected family-health-related topics in terms of the internal characteristics and external popularity?

RQ2a: What are the new entries created in each investigated time period for each theme of the selected topics? What subjects are emerging and disappearing in each period for each theme of the selected topics? For each topic, what are the evolution patterns of its internal characteristics during these time periods?

RQ2b: What are the evolution patterns for the selected topics in terms of their associated entries' number of views and number of edits? What are the evolution patterns for the themes of each selected topic in terms of the associated entries' number of views and number of edits?

RQ3: What are the differences and commonalities among the selected topics in terms of their evolution patterns?

3.2. Assumptions

Since Wikipedia allows all the users to create and revise entries and the number of the Wikipedia users is vast, this study assumes that the entries on Wikipedia reflect the perceptions

and opinions of a large number of the general public. In other words, the information on Wikipedia stands for the consensus of a great number of people. Based on this assumption, using the information on Wikipedia enables researchers to detect the change of topics over a broad scope.

It is also assumed that the number of page edits and views of an entry stands for its popularity, and the popularity of an entry reflects the popularity of the corresponding concept. Although the click-through rate or the number of edits does not equal the number of readers or editors, it is in proportion to the number of these two groups of people. The number of page views reveals the interests in the specific concepts from the readers, while the number of edits reflects the number of people who have the knowledge of the specific concepts to some extent. The more readers viewing an entry, the more popular is the topic or concept discussed in the entry. Based on the monthly and yearly numbers of views and edits of an entry from when it was created, the evolution of its popularity can be revealed.

Text mining approaches and clustering approaches were applied to this study. The vector space model was utilized when using these approaches. There are assumptions about text mining approaches, clustering approaches, and the vector space model in previous research studies. One of the implicit assumptions of text mining is the bag-of-words assumption, which assumes that the order of words in a document does not matter (Miner, 2012). Based on this assumption, breaking text into tokens becomes reasonable. Tokenizing is the basis of lemmatizing and counting term frequency. Since texts need be tokenized, it is difficult to distinguish homographs. A homograph is a group of terms spelled the same but having different meanings. It is difficult to identify the meaning of a single term without its

context. In this circumstance, another assumption is that homographs do not affect the results of text mining (Miner, 2012). Hence, there is no need to distinguish the homographs and count their frequencies separately. This assumption makes the data processing step for the vector space model easier.

These assumptions lay the theoretical and practical foundations for both the data collection and data analysis processes of this study. Based on these assumptions the research problems and questions were solved.

3.3. Data Collection

This section presents the whole procedure of data collection, including the selections of a social media platform, topics, and entries; and the collection of the associated entries' text data, page views data, and page edits data.

3.3.1. Selection of a Social Media Platform

To explore the change of topics on social media, Wikipedia was selected as the data source in this study. Milne and Witten (2013) argued that Wikipedia is a rapidly growing platform containing vast interlinked information. It is the biggest encyclopedia consisting of user-generated articles and the semantic relations between them (Milne & Witten, 2008, 2013). With the contribution of editors from all over the world, Wikipedia exhibits today's knowledge unsurpassed in breadth and actuality (Iba et al., 2010). The previous literature declared the reliable accuracy and completeness of health-related information on Wikipedia (Kräenbring et al., 2014). The richness and accuracy of its content makes it an important resource for knowledge sharing and citation, and even for research.

In addition, the history of each entry on Wikipedia is accessible to users, which means all the historical versions of the entry are recorded by Wikipedia and could be viewed and collected by researchers. For each version of an entry, the time when it was created, the content that was edited, and the editor are all recorded. Wikipedia also allows users to compare any two versions of an entry. Furthermore, hourly and monthly page views data and revision statistics (e.g. number of total revisions, number of editors, first edit time, and last edit time) of every entry are provided by Wikipedia. Therefore, it is possible to track the temporal changes in content and popularities of the entries on Wikipedia. The details of the entries on Wikipedia are demonstrated in the Text Collection section.

Compared to other social media, Wikipedia has its own advantages in this research study. Although social media platforms, such as social network sites and blogs, allow the general public to create, edit, share, and delete information, they do not record every historical version of topics. Moreover, in recent years, social network sites, like Facebook and Twitter, block the way for researchers to collect historical data. For instance, Twitter only provides researchers the APIs to retrieve the most recent tweets instead of the tweets created in a specific time period. However, Wikipedia offers data dumps for researchers to collect data and as mentioned before, all the historical data of entries are recorded. Another strength of Wikipedia is that it provides large amounts of health information widely used by the lay public and health professionals (Heilman et al., 2011). Studies have found that junior physicians used Wikipedia as health information source more than other Websites except Google; and the English Wikipedia was utilized more frequently than other health resources (e.g. WebMD and NHS Direct) according to a search engine ranking in 2008 (Heilman et al., 2011; Hughes, Joshi,

Lemonde, & Wareham, 2009). Therefore, to the best of the researcher's knowledge, Wikipedia is the most proper social media application for data collection in this study.

3.3.2. Selection of Topics

In the Introduction chapter, three primary criteria were proposed for choosing the qualified health-related topics in this research: (1) the topics should be popular family-health-related topics widely discussed by the general public; (2) the topics should cover different aspects of family health issues; (3) in order to collect data from Wikipedia, there should be more than 100 relevant entries of these topics on Wikipedia and the entries should contain sufficient data; (4) the lengths of the selected topics' history should be longer than 8 years.

This study aims to explore the evolution patterns of family-health-related topics. However, because of the limitation of time and dissertation length, it is impossible to examine all the important family-health-related topics. Thus, the topics selected should be representative in the area. Because the data of the selected topics are supposed to be obtained from Wikipedia, it is required that these topics should be widely discussed and attractive to the general public; otherwise there won't be enough associated entries on Wikipedia.

Topics referring to different aspects of family health were selected because the topics about different aspects have different background, history, and characteristics and hence, it is possible that these topics have different evolution patterns. Moreover, exploring evolution patterns of these topics can reveal the developments of family health from different aspects.

Since the data were collected from Wikipedia, there should be associated entries of the selected topics on the platform. For a selected topic, if no relevant entries are created on

Wikipedia, no data about the topics can be obtained from Wikipedia. If the relevant entries are too few or correspond to specific themes, the data set collected cannot reflect the whole picture of the topic. If the relevant entries do not contain sufficient data (e.g. historical versions, page views data, and page edits data), the data set obtained would be incomplete and the results obtained would be biased. Therefore, the selected topics should have more than 100 relevant entries containing sufficient data on Wikipedia.

In addition, to explore the evolution pattern of a topic, it should have a relatively long history. If the history of a topic is too short, the potential drawbacks of it include: (1) if there are not enough associated entries or historical versions of the entries on Wikipedia then the data might be insufficient for this study; and (2) if the content of the associated entries do not change a lot from the corresponding original versions then the evolution of a topic is not obvious.

The family-health-related topics examined in this study were selected from the World Health Organization (WHO)'s official Website. The WHO is "the designated agency on worldwide health matters" and its tasks cover "setting norms and standards, articulating policy options, providing technical support to countries, monitoring and assessing health trends, and shaping the global health research agenda" (Chorev, 2012, p.2). As one of the largest health agencies, the WHO has more budget and responsibilities and establishes more programs than other agencies (Chorev, 2012, p.13). The WHO official Website posts health news, data, publications, and programs for both general public and experts. There is a visible "health topics" tab on the Website and the Web page of this tab contains some of the most important, popular, and influential health topics.

All the health topics on the WHO Website were examined and three topics that met the previous criteria were selected for this study, which were *Child Maltreatment*, *Family Planning*, and *Women's Health*. Table 2 displays the definitions and creation time of the corresponding entries of these topics on Wikipedia. The corresponding entry of *Child Maltreatment* was *Child abuse*.

Topics	Definition	Entry Creation Time
Child Maltreatment (Child abuse)	Physical, sexual, or psychological mistreatment or neglect of a child or children, especially by a parent or other caregiver	June 2nd, 2002
Family Planning	The practice of controlling the number of children in a family and the intervals between their births, particularly by means of artificial contraception or voluntary sterilization	May 21st, 2003
Women's Health	The health of women	May 15th, 2004

Table 2. Definitions and Attributes of the Three Selected Topics

3.3.3. Selection of Entries

According to the criteria for the topic selection, three family-health-related topics were investigated and the corresponding entries of these topics on Wikipedia were regarded as the "seeds" for seeking the associated entries of these topics. A seed was the starting point for data collection. The seed entry was the first level entry in data collection. In the "See also" section of this entry there were several links connecting to associated entries which were called the second level entries. Similarly, in the "See also" sections of the second level entries, there were links connecting to relevant entries as well. These relevant entries were named the third level entries. According to a pilot study which explored the internal characteristic of the data-mining-related concepts on Wikipedia, its data collection process revealed that most of the fourth level

entries were irrelevant to the seed entry. Therefore, the entries on the fourth level were not examined, only the associated entries on the first, second and third levels were collected in this study.

Another means for seeking related entries was query search offered by Wikipedia. The search results returned were ranked by relevance. The selected topics were used as search terms and the top 100 search results returned were examined by the researcher. The associated entries which were not the same as the entries obtained by the first means were collected. Whether the entries obtained were relevant to the selected topics or not was judged by the researcher based on her domain knowledge. For every selected topic, more than 100 qualified entries were obtained by the two means.

3.3.4. Selection of Time Periods

Since Wikipedia was created in 2001, the time periods to be investigated should be later than 2001. Moreover, the beginning of the time periods should be later than the corresponding entries of the selected topics created on Wikipedia. The historical versions and the page edits data for Wikipedia entries were accessible from the date the entries were created. However, the page views data of Wikipedia entries were available from December 2007. Since the data before December 2007 were incomplete, this study explores the internal characteristics and external popularities of the selected topics after this time point.

To determine the time periods, several criteria were proposed: (1) The time interval should be long enough, otherwise the analysis on changes of internal characteristics or external popularities might not be sufficient; (2) There should be several time periods that reflect

different developing stages of the selected topics; (3) The lengths of time periods determined for different topics should be equal so that the evolution patterns are comparable in a fair way; and (4) The total length of the defined time periods should be long enough to observe the evolutions of the selected topics from both internal and external perspectives. In this case, the total length of the time periods was from 2010 to 2017 because eight years was long enough to observe the evolutions for the selected topics. Four time periods were defined for the selected topics, which were 2010 to 2011, 2012 to 2013, 2014 to 2015, and 2016 to 2017. Table 3 presents the four periods and the corresponding time spans. The time interval in this study was two years, long enough to reveal the changes of topics; and the number of time periods was four, which was able to reveal the topics' internal characteristics and external popularities in different developing stages.

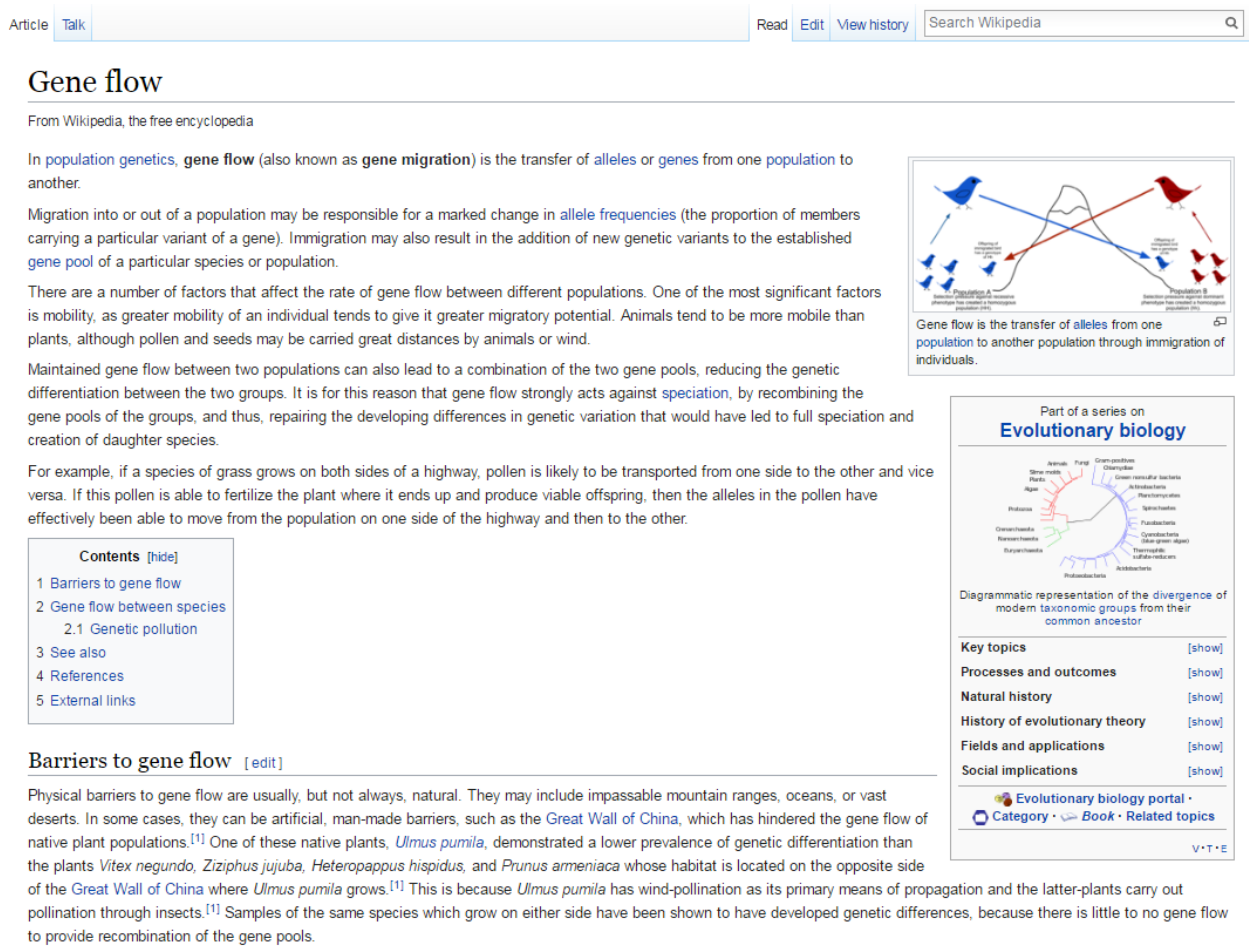
Time Period	**Period 1**	**Period 2**	**Period 3**	**Period 4**
Time Span	2010-2011	2012-2013	2014-2015	2016-2017

Table 3. Four Time Periods and the Corresponding Time Spans

3.3.5. Text Collection

Every entry on Wikipedia contains several sections, such as content, main text, reference, and so on. Although not all entries consist of the same sections, certain sections are included in almost all entries. They are title, other entry/entries associated to this entry, a short description of the entry, content, main body, "See also", and reference. The content section includes the content table of an entry; the main text or main body of an entry; and the reference section which consists of references and URLs of references of an entry.

Figure 3 and Figure 4 are the screenshots of the "Gene flow" entry on Wikipedia. Figure 3 shows the top half of the Gene flow's article page. There are five tabs on the top of the page, which are Article, Talk, Read, Edit, and View history. The Talk page is the place where the Wikipedia editors discuss the improvements of the entry's content and the Edit page is the place for editing the content. The View history page provides the links to the historical data and historical versions of the entry. The current version of the entry, which is presented in Figure 3, is displayed under the Article and Read tabs. The title "Gene flow" is on the top of the page and beneath the title is the text of the entry. The first several paragraphs are a brief introduction of gene flow and a table of contents is listed after them. Beneath the table of contents is the main body of the entry. On the right side of the page are two figures relevant to the entry.


Article Talk Read Edit View history Search Wikipedia

Gene flow

From Wikipedia, the free encyclopedia

In population genetics, **gene flow** (also known as **gene migration**) is the transfer of alleles or genes from one population to another.

Migration into or out of a population may be responsible for a marked change in allele frequencies (the proportion of members carrying a particular variant of a gene). Immigration may also result in the addition of new genetic variants to the established gene pool of a particular species or population.

There are a number of factors that affect the rate of gene flow between different populations. One of the most significant factors is mobility, as greater mobility of an individual tends to give it greater migratory potential. Animals tend to be more mobile than plants, although pollen and seeds may be carried great distances by animals or wind.

Maintained gene flow between two populations can also lead to a combination of the two gene pools, reducing the genetic differentiation between the two groups. It is for this reason that gene flow strongly acts against speciation, by recombining the gene pools of the groups, and thus, repairing the developing differences in genetic variation that would have led to full speciation and creation of daughter species.

For example, if a species of grass grows on both sides of a highway, pollen is likely to be transported from one side to the other and vice versa. If this pollen is able to fertilize the plant where it ends up and produce viable offspring, then the alleles in the pollen have effectively been able to move from the population on one side of the highway and then to the other.

Gene flow is the transfer of alleles from one population to another population through immigration of individuals.

Contents [hide]

1 Barriers to gene flow
2 Gene flow between species
2.1 Genetic pollution
3 See also
4 References
5 External links

Part of a series on
Evolutionary biology

Diagrammatic representation of the divergence of modern taxonomic groups from their common ancestor

Key topics [show]
Processes and outcomes [show]
Natural history [show]
History of evolutionary theory [show]
Fields and applications [show]
Social implications [show]

Evolutionary biology portal · Category · Book · Related topics

v · t · e

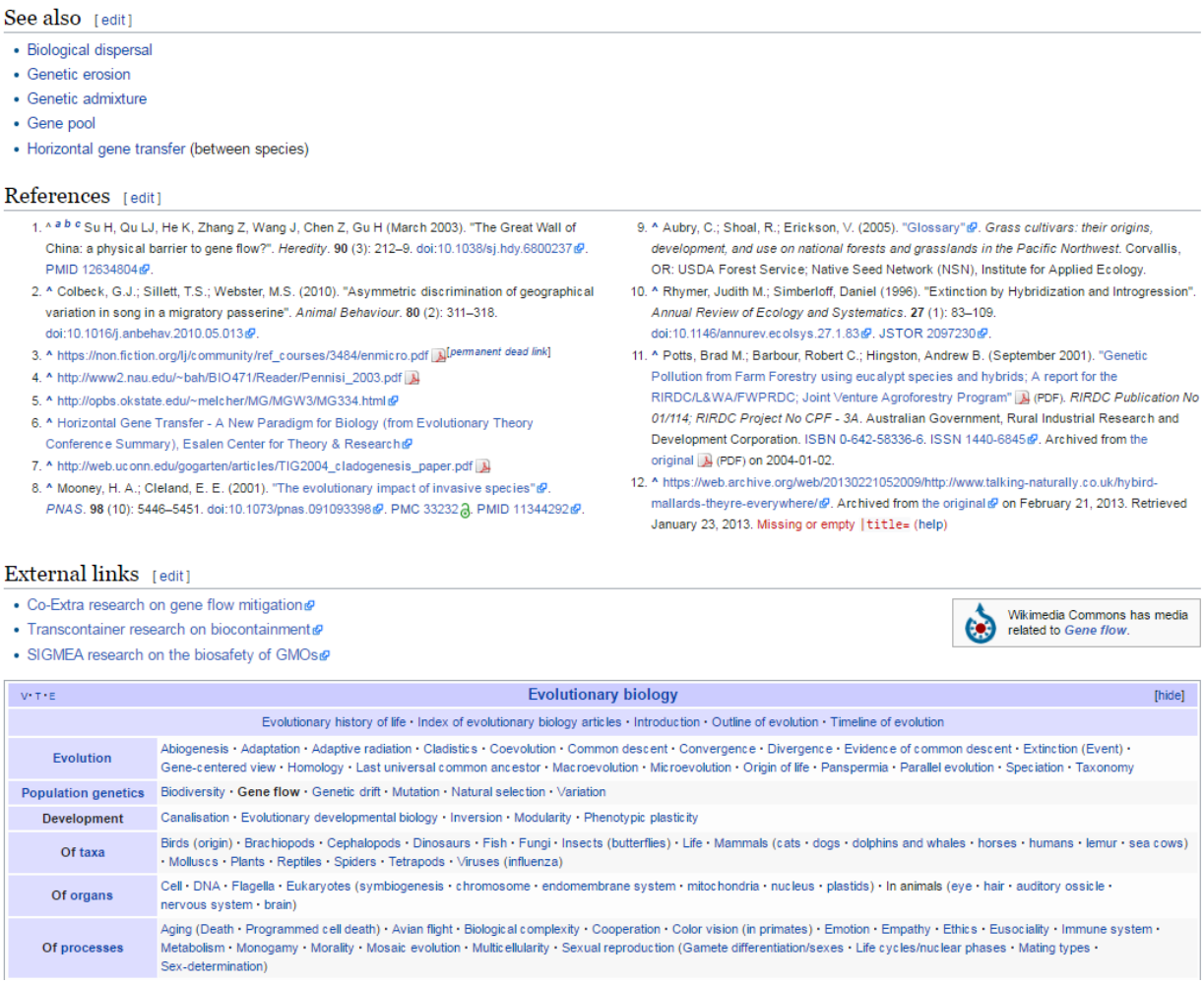
Barriers to gene flow [edit]

Physical barriers to gene flow are usually, but not always, natural. They may include impassable mountain ranges, oceans, or vast deserts. In some cases, they can be artificial, man-made barriers, such as the Great Wall of China, which has hindered the gene flow of native plant populations.[1] One of these native plants, *Ulmus pumila*, demonstrated a lower prevalence of genetic differentiation than the plants *Vitex negundo, Ziziphus jujuba, Heteropappus hispidus,* and *Prunus armeniaca* whose habitat is located on the opposite side of the Great Wall of China where *Ulmus pumila* grows.[1] This is because *Ulmus pumila* has wind-pollination as its primary means of propagation and the latter-plants carry out pollination through insects.[1] Samples of the same species which grow on either side have been shown to have developed genetic differences, because there is little to no gene flow to provide recombination of the gene pools.

Figure 3. Top half of the Gene flow Entry on Wikipedia

Figure 4 presents the bottom half of the Gene flow Article page. After the main body of the entry, there are three sections: See also, References, and External links. The See also section contains eight Wikipedia entries related to the Gene flow entry, while the External links section contains three relevant Web page links. The References section lists the literature cited in the content. The table under the External links section displays some evolutionary-biology-related concepts, including Gene flow. However, this kind of index table does not frequently occur in Wikipedia entries. Apart from these sections, there are some other common sections

which often occur in entries, like Bibliography sections and Further readings sections. In this study, the text of all the sections of each entry were collected.



See also [edit]

- Biological dispersal
- Genetic erosion
- Genetic admixture
- Gene pool
- Horizontal gene transfer (between species)

References [edit]

1. ^ a b c Su H, Qu LJ, He K, Zhang Z, Wang J, Chen Z, Gu H (March 2003). "The Great Wall of China: a physical barrier to gene flow?". *Heredity*. **90** (3): 212–9. doi:10.1038/sj.hdy.6800237. PMID 12634804.
2. ^ Colbeck, G.J.; Sillett, T.S.; Webster, M.S. (2010). "Asymmetric discrimination of geographical variation in song in a migratory passerine". *Animal Behaviour*. **80** (2): 311–318. doi:10.1016/j.anbehav.2010.05.013.
3. ^ https://non.fiction.org/lj/community/ref_courses/3484/enmicro.pdf [permanent dead link]
4. ^ http://www2.nau.edu/~bah/BIO471/Reader/Pennisi_2003.pdf
5. ^ http://opbs.okstate.edu/~melcher/MG/MGW3/MG334.html
6. ^ Horizontal Gene Transfer - A New Paradigm for Biology (from Evolutionary Theory Conference Summary), Esalen Center for Theory & Research
7. ^ http://web.uconn.edu/gogarten/articles/TIG2004_cladogenesis_paper.pdf
8. ^ Mooney, H. A.; Cleland, E. E. (2001). "The evolutionary impact of invasive species". *PNAS*. **98** (10): 5446–5451. doi:10.1073/pnas.091093398. PMC 33232. PMID 11344292.
9. ^ Aubry, C.; Shoal, R.; Erickson, V. (2005). "Glossary". *Grass cultivars: their origins, development, and use on national forests and grasslands in the Pacific Northwest*. Corvallis, OR: USDA Forest Service; Native Seed Network (NSN), Institute for Applied Ecology.
10. ^ Rhymer, Judith M.; Simberloff, Daniel (1996). "Extinction by Hybridization and Introgression". *Annual Review of Ecology and Systematics*. **27** (1): 83–109. doi:10.1146/annurev.ecolsys.27.1.83. JSTOR 2097230.
11. ^ Potts, Brad M.; Barbour, Robert C.; Hingston, Andrew B. (September 2001). "Genetic Pollution from Farm Forestry using eucalypt species and hybrids; A report for the RIRDC/L&WA/FWPRDC; Joint Venture Agroforestry Program" (PDF). *RIRDC Publication No 01/114; RIRDC Project No CPF - 3A*. Australian Government, Rural Industrial Research and Development Corporation. ISBN 0-642-58336-6. ISSN 1440-6845. Archived from the original (PDF) on 2004-01-02.
12. ^ https://web.archive.org/web/20130221052009/http://www.talking-naturally.co.uk/hybird-mallards-theyre-everywhere/. Archived from the original on February 21, 2013. Retrieved January 23, 2013. Missing or empty |title= (help)

External links [edit]

- Co-Extra research on gene flow mitigation
- Transcontainer research on biocontainment
- SIGMEA research on the biosafety of GMOs

Wikimedia Commons has media related to *Gene flow*.

v • t • e	Evolutionary biology	[hide]
Evolutionary history of life • Index of evolutionary biology articles • Introduction • Outline of evolution • Timeline of evolution		
Evolution	Abiogenesis • Adaptation • Adaptive radiation • Cladistics • Coevolution • Common descent • Convergence • Divergence • Evidence of common descent • Extinction (Event) • Gene-centered view • Homology • Last universal common ancestor • Macroevolution • Microevolution • Origin of life • Panspermia • Parallel evolution • Speciation • Taxonomy	
Population genetics	Biodiversity • **Gene flow** • Genetic drift • Mutation • Natural selection • Variation	
Development	Canalisation • Evolutionary developmental biology • Inversion • Modularity • Phenotypic plasticity	
Of taxa	Birds (origin) • Brachiopods • Cephalopods • Dinosaurs • Fish • Fungi • Insects (butterflies) • Life • Mammals (cats • dogs • dolphins and whales • horses • humans • lemur • sea cows) • Molluscs • Plants • Reptiles • Spiders • Tetrapods • Viruses (influenza)	
Of organs	Cell • DNA • Flagella • Eukaryotes (symbiogenesis • chromosome • endomembrane system • mitochondria • nucleus • plastids) • In animals (eye • hair • auditory ossicle • nervous system • brain)	
Of processes	Aging (Death • Programmed cell death) • Avian flight • Biological complexity • Cooperation • Color vision (in primates) • Emotion • Empathy • Ethics • Eusociality • Immune system • Metabolism • Monogamy • Morality • Mosaic evolution • Multicellularity • Sexual reproduction (Gamete differentiation/sexes • Life cycles/nuclear phases • Mating types • Sex-determination)	

Figure 4. Bottom half of the Gene flow Entry on Wikipedia

For each entry, the text data of the last version generated in 2011, 2013, 2015, and 2017 were collected based on the time periods determined. Figure 5 illustrates the top part of the View history page for the Gene flow entry. This page allows users to search historical versions of the entry by a search box and to browse all the historical versions. Six external tools displayed on this page provide revision statistics, editor information, number of watchers, and

page view statistics to researchers and the general public. The historical versions and their attributes (e.g. time when the revision was made and editor) are all listed in reverse chronological order under the external tools. Users are able to view any historical version and compare any two historical versions.

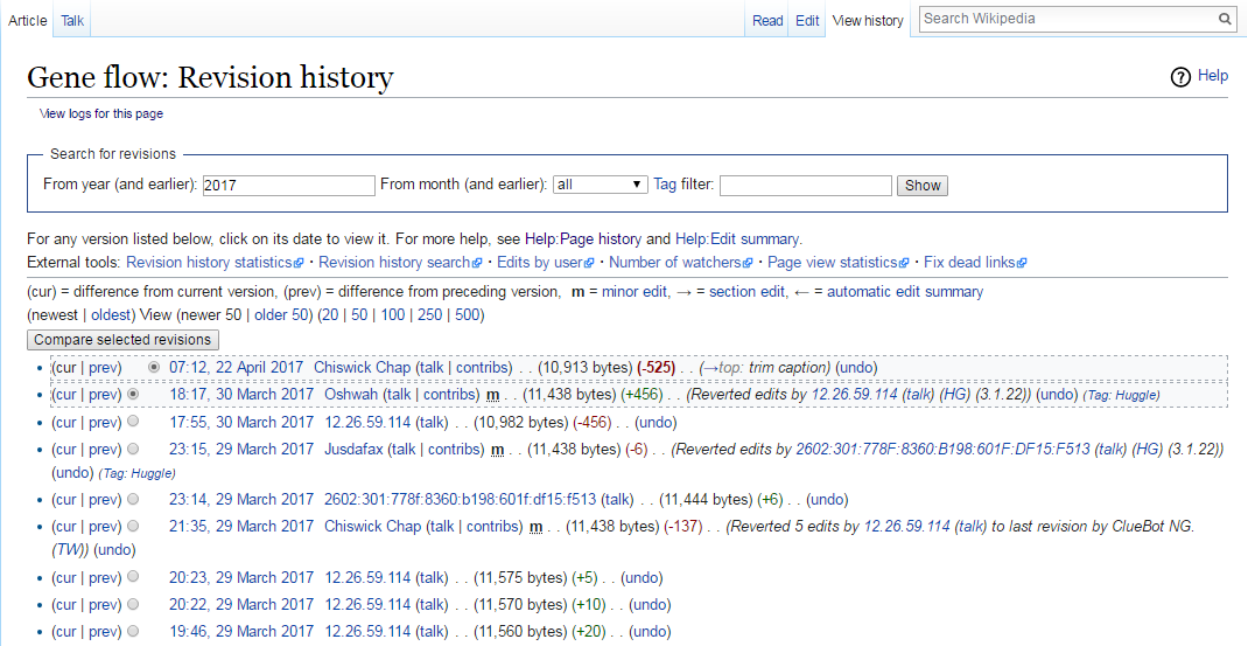


Figure 5. Top Part of the View History Page of the Gene flow Entry on Wikipedia

The WikipediR package run on r was adopted for text data collection (Keyes, 2017). It enables the researcher to retrieve and gather the text content of an entry's current and historical versions.

3.3.6. Page Views and Edits Data Collection

The page views data during 2010 to 2017 were collected from Wikimedia Downloads. This Website provides the Wikipedia data dumps that store all the historical page views data of all the Wikipedia entries. Wikipedia data dump stores the hourly page views data from January

2010 to November 2011 and the monthly page views data from December 2011 to December 2017 for each entry. These data sets were downloaded and r and RStudio were adopted to extract the page views data for the associated entries from the data sets. Then the monthly and yearly numbers of page views were calculated based on these data.

As it was shown before, the View history page of an entry displays its revision history. The data of the revisions (e.g. revision time, revision ID, and the page links of revisions) during 2010 to 2017 were collected by r and RStudio from the View history page of each associated entry. The monthly and yearly numbers of page edits were calculated based on the revision time data.

After collecting the numerical and textual data by the methods mentioned before, data cleansing and transformation were conducted. The r and RStudio tools are open source software for data analysis. RStudio is an integrated development environment for r. It offers a productive user interface for r and supports direct code execution, debugging, workspace management, plotting, and so on. A bunch of packages have been created for r and RStudio to enrich their data collection, analysis, presentation, and visualization functionalities. These two tools were applied to text data cleansing and transformation in this study.

3.3.7. Ethic Issue

There are always concerns about collecting data from people in academic research (Oliver, 2010). Guillemin and Gillam (2004) suggest that there are procedural and practical dimensions of ethics in research studies. The procedural ethics requires seeking approval from a research ethics committee and the practical ethics refers to important ethical moments when

conducting research studies, such as “when participants indicate discomfort with their answer, or reveal a vulnerability” (Guillemin & Gillam, 2004, p265). In general, research ethic issues usually occur in human subject research studies.

The data collected for this study include historical versions of specific Wikipedia entries and page edits and page views data of these entries. These data are all published online open to the public so they are not confidential. Secondly, since this study does not focus on Wikipedia users or editors, no information (e.g. names, emails, and addresses) about people were viewed or collected and the researcher did not interact with or observe people in person or online. Therefore, this study is not human subject research and the data collected are not about individual persons.

3.4. Data Analysis

This section describes the whole data analysis procedure of this study. Coding, subject analysis methods, statistical methods, text mining approaches, clustering approaches, and visualization tools were utilized for data analysis and presentation.

3.4.1. Categories and Themes

The entries obtained related to the topic on different aspects; in order to explore the relations among the entries, they were grouped into several categories in terms of their content. Since there are no existing categories of these entries, the content analysis method is not proper to utilize in this case. This study is not concerned with language usage so that the discourse analysis method is not suitable. The open coding method allows researchers to develop concepts and categories from data, which met the researcher’s research requirements

(Pandit, 1996). Therefore, this method was employed to analyze the associated entries and group them into several categories.

The first step of coding was to identify the concepts in the data. In this study because of the features of the text data collected, the concepts had already been extracted from the entries. For each selected topic, the related concepts were manually compared one by one and labeled according to their properties. With more and more concepts labeled, the labels were refined. The concepts with more similarities were grouped together, while those containing more different properties were separated. Finally, all the concepts obtained for a topic were categorized into several categories based on their content, and the concepts in the same category had common properties. According to the properties, the theme of each category was generated by the researcher. In this research every category had one theme. To ensure the reliability of the coding results, an expert with health background was invited to code a part of the entries and the Cohen's kappa coefficient was 0.629 which means a substantial agreement was obtained between the two coding results (Viera & Garrett, 2005).

3.4.2. Text Data Organization

The open coding method classified the entries into different themes in each selected topic. The historical versions generated in 2017 obtained for the entries were accordingly assigned into the corresponding themes. The subjects of every theme were extracted from the entries belonging to it by clustering approaches and text mining approaches. To apply these approaches, the text data obtained for the themes were cleansed and transformed first.

The open source software r and RStudio were adopted for text data cleansing and transformation. The tm package offered by r assisted to process the text data. This package removed the punctuations and stop words from the text data, stripped the white spaces, and stemmed the terms. Stop words were some of the most common, short function words, such as prepositions and articles. The obtained words were stemmed so that the words with same roots were combined into one term. Furthermore, some meaningless words were excluded from the terms obtained, such as numbers, dates, equations, formulas, special symbols, and so on. Then a matrix of the vector space model was presented for each theme.

Each theme formed a document-term matrix where its columns were the terms, and its rows were the entries. The frequencies of all the terms in each entry in the theme were counted. All the entries were ranked alphabetically and numbered from 1 to m. A matrix is presented in Eq.1. As this equation displays, the matrix has m rows and n columns. The value of the cell (a_{ij}) in the matrix represents the frequency of the term j in the entry i.

$$M = \begin{pmatrix} a_{11} & a_{12} & .. & .. & a_{1n} \\ a_{21} & a_{22} & .. & .. & a_{2n} \\ .. & .. & a_{ij} & .. & .. \\ & & ... & & \\ a_{m1} & a_{m2} & .. & .. & a_{mn} \end{pmatrix} \quad \text{(Eq. 1)}$$

Since the text data of several hundred entries were gathered for each theme, the matrices were huge. Terms whose frequencies were less than 4 were removed from the final document-term matrices.

The final document-term matrices were then transformed to Term Frequency × Inverse Document Frequency (TF-IDF) matrices. Each value of the TF-IDF matrices (w_{ij}) was calculated based on Eq.2. In this equation, m is the number of the entries of a matrix and e_j represents the number of the entries containing the term j in the matrix. The TF-IDF matrices obtained were the input matrices for the following clustering process.

$$w_{ij} = a_{ij} \times Log(m/e_j) \text{ (Eq. 2)}$$

3.4.3. SOM Approach

To cluster the entries in the themes, the Self-Organizing Map (SOM) approach was employed in this study. As first introduced by Kohonen (1990), the SOM approach is a popular unsupervised learning approach. It is a widely used neural network method that measures similarities among items of input data so as to form similarity graphs. The whole procedure of this approach is a recursive regression process (Kohonen et al., 2000). The SOM has an input layer and an output layer. As was mentioned before, the input matrix is described in Section 3.4.2 and each row is an entry, which is defined as:

$$D_i = [a_{i1}, a_{i2}, a_{i3}, \cdots, a_{in}], i = 1, \cdots, m$$

Many of the outputs of the SOM are two-dimensional and are in grid structure. A grid contains many nodes, which are also known as neurons. A neuron in the grid is related to a weight vector:

$$M_i = [m_{i1}, m_{i2}, m_{i3}, \cdots, m_{in}] \subset R^n, i = 1, \cdots, k \text{ (Eq.3)}$$

In Eq.3, R^n stands for an n-dimensional vector set and n represents the number of attributes of each vector; and k is the number of neurons in the corresponding grid. The weight vectors in the grid were initialized and small values (near zero) were randomly assigned to all of their elements. In the training and learning process, all neurons competed for a winning neuron of a randomly selected input vector. This step repeated many times until the training and learning process ended. If D_i was a selected vector, the corresponding winning node would be:

$$\mathrm{C}(D_i) = min\{(\sum_{r=1}^{n}(a_{ir} - m_{lr})^2)^{1/2}\}, l = 1,2,\cdots,k \text{ (Eq.4)}$$

When the winning nodes emerged, the whole grid converged. Learning rate is another mechanism that ensures the convergence. The learning rate decreased as time goes by. When this rate was zero, the training and learning process ended and the grid converged. After the grid converged, all the entries were projected to the final display map based on the same algorithm. The algorithm found their wining nodes and assigned them to the corresponding nodes. As a result, similar entries were assigned to the same node on the output display map.

Several toolboxes in Matlab contain the functions of SOM, such as the nnet toolbox. MatLab is a computing environment for solving engineering and scientific problems. It offers an interface and toolboxes for users to process, present, analyze, and visualize data.

In this study, the SOM toolbox created by Himberg, Alhoniemi and Parhankangas (2000) was utilized. This toolbox provides functions for data input, normalization and training, SOM creation, and visualization. Multiple algorithms and visualization functions have been proposed for SOM. In order to achieve sound visualization results, the batch learning algorithm and the U-matrix technique were applied to this case. The batch learning algorithm was applied in the

training process of SOM and U-matrix was used for projecting the clustering results to SOM displays. The algorithm and function intended to reveal the distribution of the entries of each matrix on the SOM display. Every entry was projected to the SOM display as a number. Numbers with shorter distances among them were more similar than those with longer distances. Moreover, the similarity among entries was indicated by color of a SOM output. The color projected to the SOM display background was determined by a U-matrix (Ultsch & Siemon, 1990). Higher values of the U-matrix stood for cluster borders, while lower values represented clusters.

According to the distances between numbers and the colors on background, the entries of each matrix were clustered. The criteria for clustering the numbers are: (1) the numbers locate in a same SOM node were grouped into one cluster; (2) if the numbers locate in two or more nodes, and the nodes are adjacent, or separated by only one empty node, and at the same time, the numbers locate in the same area where the U-matrix values were lower than half of the highest U-matrix value of the matrix, then these numbers will be grouped into one cluster. In this way, the entries of each theme were assigned into several clusters. Then the subjects of each cluster were extracted from the entries by subject analysis. Similar to the coding process, the subjects were reviewed by the health science expert.

3.4.4. Subject Analysis

To identify the subjects of the clusters and themes, the n-gram approach was employed. The n-gram package offered by r extracts the n-word phrases in unstructured text files. The n-

gram package was developed based on the n-gram model which predicts the next word of an existing word or string in a specific text file.

In this study, the historical revisions of the entries in one cluster were merged into one document. For each theme, the historical revisions in a specific period of its entries were also merged into one document. The most used 2-word, 3-word, and 4-word phrases in each document were extracted by the n-gram package. The set phrases (e.g. “as long as”) and meaningless phrases (e.g. “of the” and “the study is a”) were removed from the data set. If a phrase was a part of another one and the two phrases had the same meaning, then they were regarded as one phrase and their frequencies were added together (e.g. “child development index” and “the child development index”). After data processing, a list containing phrases and frequencies was obtained for each document.

The researcher manually reviewed the lists to summarize the subjects of each document. One phrase could relate to more than one subjects and different phrases could relate to the same subjects. In this way, the subjects of each cluster were generated and the subjects in each period of each theme were generated.

To find the increasing and decreasing phrases in each theme, the differences of the frequencies obtained from the adjunct periods for a phrase was calculated. For example, the frequency of the phrase “human trafficking” obtained from Period 2 minus the corresponding frequency obtained from Period 1 was the frequency difference between Periods 1 and 2. The frequency differences between Periods 2 and 3, and Periods 3 and 4 were also calculated for the “human trafficking” phrase. After all the frequency differences were obtained for each

theme, the researcher reviewed the most increasing and most decreasing phrases and generated the subjects from them.

3.4.5. Inferential Analysis

Inferential statistical tests allow the researcher to gain insights into the differences among the objects. In addition to descriptive statistical methods, inferential statistical analysis was applied to test the differences among the determined periods for each investigated topic, and the differences among the selected topics. The hypotheses are:

H01: There were no significant differences among the investigated time periods in terms of the number of page views of the entries relevant to each of the topics.

H02: There were no significant differences among the investigated time periods in terms of the number of page edits of the entries relevant to each of the topics.

H03: There were no significant differences among the selected topics in terms of the number of the page edits of the associated entries.

H04: There were no significant differences among the selected topics in terms of the number of the page views of the associated entries.

For the first two hypotheses, since the independent variable (time period) was categorical, the dependent variables (number of page views and number of page edits) were continuous with repeated measures, and the distributions of the dependent variables did not follow the normal distribution, the Friedman's Test was applied to test the differences among the periods. To explore the difference among every two periods, a series of pairwise

comparisons were conducted. Because the distribution of the differences among every two periods was not symmetrical, so the Sign Test was used.

To test the third and fourth hypotheses, because the independent variable (topic) was categorical and had more than two levels, and the dependent variables (number of page views and number of page edits) were numeric but did not follow the normal distribution, so the Kruskal-Wallis H Test and its post-hoc pairwise comparisons in SPSS were utilized.

These tests helped the researcher solve the RQ2b of the second research question and the third research question. SPSS, one of the most popular statistical analysis and reporting software in scientific research, was used for inferential analysis. In this study, the significant level of the inferential statistical test was 0.05.

3.4.6. Temporal Analysis

Temporal analysis is the analysis where temporal data or temporal text play an important role. In this study, the internal characteristics of each topic in the four defined periods were compared. The hypothesis testing examining the differences between the four periods was another kind of temporal analysis.

To explore the evolutions of external popularities for the selected topics, the yearly number of page views and the yearly number of page edits were analyzed. For each topic and theme, the data of their associated entries were added up. For example, to show the popularity of *Child Maltreatment* in 2017, the numbers of page views in 2017 of all the associated entries of *Child Maltreatment* were added up, and as well as the numbers of page edits of these entries. These data reflect the external popularities of the topics and themes. After the data of

the topics and themes acquired, descriptive statistical methods and line charts and bar charts were employed to demonstrate how the popularities changed over time.

3.5. Validity and Reliability

3.5.1. Validity

Gravetter and Forzano (2011) defined the validity of a study as "the degree to which the study accurately answers the question it was intended to answer" (p.167). Maxwell (2005b) explained more straightforwardly that validity refers to "the correctness or credibility of a description, conclusion, explanation, interpretation, or other sort of account" (p.106). Internal validity is concerned with the explanation of results. A valid study allows only one interpretation of the results. To ensure internal validity of the results, both quantitative and qualitative research methods were utilized in this study. The results obtained by qualitative research methods support those obtained by quantitative research methods, and vice versa. Using multiple research methods helped avoid bias in the results.

External validity concerns whether the results obtained could be generalized to other "populations, settings, times, measures, or characteristics" (Gravetter & Forzano, 2011, p.168). Although only three family-health-related topics were investigated in this study, the selected topics were representative topics for family health and several hundreds of their associated entries were examined. In this case, the evolution patterns obtained could be generalized to other family-health-related topics, and even other health-related topics.

This study collected data from Wikipedia, one of the largest collaborative projects. The content on Wikipedia is generated by the public, which is same as other social media platforms.

Accordingly, the evolutions of the topics, themes, subjects, and terms on Wikipedia can reveal the general public's understandings and perceptions and these evolutions could be generalized to other social media platforms.

The research methods used in this study could be applied to other similar situations. In this study, the textual and numeric data collected from Wikipedia were analyzed by coding, clustering approaches, text mining approaches, and statistical methods. It confirmed the applicability of the mixed methods. Therefore, the same methods could be applied to other similar situations, such as data mining and text mining studies on social media.

3.5.2. Reliability

Reliability means the stability and consistency of data collection and analysis processes (Gravetter & Forzano, 2011). To avoid the instability and inconsistency of data collection, automatic data collection methods were utilized instead of manual data collection. To ensure data integrity, all the historical data during 2010 to 2017 of the related entries were gathered by the approaches mentioned in the Data Collection section. Moreover, several rounds of data collection were conducted in order to avoid the lack of data.

Two means were applied for seeking relevant entries of the selected topics on Wikipedia. One means was to locate a seed concept and seek relevant entries through its "see also" section. As it was mentioned before, three levels of entries were collected. The other means was to retrieve relevant entries by query search on Wikipedia.

A relatively unstable step in this study was coding the entries. To ensure the reliability of this study, two independent coders participated in the open coding process. They were asked

to code a part of the entries with the same coding scheme. After that the Cohen's kappa inter-coder reliability was calculated to examine whether they agreed on the coding results.

3.6. Chapter Three Summary

This chapter expands the whole picture of this study. The research background, research problem and questions, assumptions, and research methodology are all demonstrated in this chapter. Particularly, the data collection and analysis procedure and the methods and approaches used in these two steps are illustrated in detail.

Table 4 presents a summary of this study, including the research questions and sub-questions, data collection methods and the data collected, and data analysis methods. For each of the sub-questions, their corresponding data collection and analysis methods and approaches are listed in the same row. This table shows that different types of data were collected in this study, such as text data and numerical data. These data can reveal the characteristics of the selected topics from different aspects. The data obtained were integrated and both qualitative and quantitative methods and approaches were adopted in data analysis. In Table 4 the qualitative methods are coding and subject analysis, while the quantitative methods include the SOM approach, the n-gram approach, the descriptive statistical methods, and the inferential statistical methods.

Research Questions	Sub-Questions	Data Collection	Data Analysis
RQ1: What are the associated entries, emerged	RQ1a: What are the associated entries and the main themes of the selected family-health-related topics discussed on Wikipedia?	Text of entries collected from Wikipedia	Coding

themes and subjects, and relations among them in each of the selected family-health-related topics?	RQ1b: What are the subjects of each theme of the selected family-health-related topics?	Text of entries collected from Wikipedia	SOM, n-gram, subject analysis
	RQ1c: What are the relations among the themes, subjects, and entries?	Entries collected from Wikipedia, themes obtained in RQ1a, clusters and subjects obtained in R1b	Coding, SOM, n-gram, subject analysis
RQ2: What are the evolution patterns for each of the selected family-health-related topics in terms of the internal characteristics and external popularity?	RQ2a: What are the new entries created in each investigated time period for each theme of the selected topics? What subjects are emerging and disappearing in each period for each theme of the selected topics? For each topic, what are the evolution patterns of its internal characteristics during these time periods?	Historical versions of the entries collected from Wikipedia	N-gram, subject analysis
	RQ2b: What are the evolution patterns for the selected topics in terms of their associated entries' number of views and number of edits? What are the evolution patterns for the themes of each selected topic in terms of the associated entries' number of views and number of edits?	Historical page views and edits data collected from Wikipedia	Descriptive statistical methods, Friedman's Test, Sign Test
RQ3: What are the differences and commonalities among the selected topics in terms of their evolution patterns?		Historical page views and edits data collected from Wikipedia, evolution patterns obtained from RQ2	Kruskal-Wallis H Test and post-hoc pairwise comparison

Table 4. Summary of Research Questions and Methodology

4. RESULTS

4.1. Descriptive Results

4.1.1. Topics and Themes

According to the topic selection strategy presented in Chapter 3, three health-related topics were selected in this study, which were *Child Maltreatment*, *Family Planning*, and *Women's Health*. For each topic, its corresponding "seed" entries on Wikipedia were used to collect related entries. The seed entries were same as the selected topics or were the synonyms of the selected topics. For instance, the "Child abuse" entry was the seed entry of the *Child Maltreatment* topic because child abuse was a synonym of child maltreatment; the "Family planning" entry was the *Family Planning* topic's seed entry.

As it was mentioned in the Data Collection section, query search was another approach for collecting the relevant entries. The selected topics were utilized as the search queries to retrieve the relevant entries on Wikipedia. After obtaining all the relevant entries for each topic, the researcher manually reviewed and labeled the entries, and assigned them into several categories in terms of their themes.

Table 5 lists the three selected topics, the themes of each topic, and the description of each theme. It shows the related entries of the *Child Maltreatment* topic were assigned to four categories, and the themes of the categories were *Abuse, violence, harm, and subordination* (AVHS); *Children, youth, families and friends* (CYFF); *Health problems and risks* (CM-HPR); and *Support and protection* (CM-SP), respectively. Three themes were generated for the *Family*

Planning topic, including *Family planning and reproductive health* (FPRH); *Human and environment* (HE); and *Population problems* (PP). The themes of *Women's Health* included *Discrimination, violence, harm, and subordination* (DVHS); *Health problems and risks* (WH-HPR); *Medical and interdisciplinary subjects* (MIS); and *Support and protection* (WH-SP).

Selected Topics	Themes	Descriptions
Child Maltreatment	Abuse, violence, harm, and subordination (AVHS)	Entries related to mental and physical abuse, violence, and harm, and subordination to children
	Children, youth, families and friends (CYFF)	Entries related to children and children's social relations
	Health problems and risks (CM-HPR)	Entries related to health problems and risks, including the causes of child abuse
	Support and protection (CM-SP)	Entries related to policies, laws, research studies, literary and artistic work, treatments, people, and organizations that support and protect children, and prevent children from being abused
Family Planning	Family planning and reproductive health (FPRH)	Entries related to policies, laws, research studies, methods, services, people, and organizations about family planning and reproductive health
	Human and environment (HE)	Entries related to human rights, human life, human society, and environment
	Population problems (PP)	Entries related to human population
Women's Health	Discrimination, violence, harm, and subordination (DVHS)	Entries related to mental and physical violence and harm, discrimination, and subordination to women
	Health problems and risks (WH-HPR)	Entries related to health problems and risks
	Medical and interdisciplinary subjects (MIS)	Entries related to medical subjects and health-related interdisciplinary subjects
	Support and protection (WH-SP)	Entries related to policies, laws, research studies, literary and artistic work, treatments, people, and organizations that support and protect women, and improve women's health

Table 5. Selected Topics and Themes of Each Topic

The total number of the relevant entries of the four selected topics was 578. Table 6 displays the numbers of entries of the topics and themes. *Child Maltreatment*, *Family Planning*, and *Food Safety* had 241, 150, and 207 entries, respectively. The results reflect that *Child Maltreatment* was the most popular among the three selected topics because more entries were created for it than the other three topics. Accordingly, *Women's Health* was less popular than *Child Maltreatment*, but received more attentions than *Family Planning*.

Selected Topics	No. of Entries	Themes	No. of Entries
Child Maltreatment	241	Abuse, violence, harm, and subordination	118
		Children, youth, families and friends	28
		Health problems and risks	33
		Support and protection	62
Family Planning	150	Family planning and reproductive health	95
		Human and environment	28
		Population problems	27
Women's Health	207	Discrimination, violence, harm, and subordination	37
		Health problems and risks	25
		Medical and interdisciplinary subjects	46
		Support and protection	99
Total Entries	578	-	-

Table 6. Numbers of Entries Per Topic and Theme

Since the three selected topics were all about family health, it is possible that some of their relevant entries were overlapped. The results of the entry collection process show that there were a few entries related to different topics, so the sum of the numbers of the three topics was larger than the number of the total relevant entries. For instance, the entry Abortion was relevant to *Child Maltreatment*, *Family Planning*, and *Women's Health* three topics. The entry Birth control was related to both *Family Planning* and *Women's Health*.

The AVHS theme (118 entries) had the most entries among the four themes of *Child Maltreatment*. The CM-SP theme, with 62 entries, achieved the second place, while CM-HPR (33 entries) and CYFF (28 entries) occupied the last two positions among the four. These findings reveal that the Wikipedia editors had more interests in the harms of child maltreatment and the support and protection of children.

For the *Family Planning* topic, the FPRH (95 entries) outnumbered the other two themes. The numbers of entries of HE (28 entries) and PP (27 entries) were much less than the FPRH theme, but close to each other. It shows that people paid more attention on the policies, methods, services, and organizations of family planning and reproductive health than the other related themes.

Among the four themes of the *Women's Health* topic, the WH-HPR theme (25 entries) had the smallest number of entries. The DVHS theme (37 entries) and the MIS theme (46 entries) reached the third and second places, while the WH-SP theme (99 entries) became the first, which reflects that the general public cared more about the protection of women's health than the other themes. Tables 5 and 6 provide the answers to RQ1a.

After retrieving and collecting all the relevant entries, the creation time, the historical versions, the page edits data, and the page views data of each entry were collected. Table 7 presents the numbers of the entries created from 2010 to 2017 of each topic and theme, which partially answers RQ2a. In general, the total number of the entries created each year increased from 2010 to 2012 and decreased after 2012. A similar trend was found for the *Family Planning* topic. In 2011 and 2012, new entries were generated more than the other investigated years.

This phenomenon was caused by the creation of new entries of the FPRH theme. Moreover, this theme mainly accounted for the increase of entries of the *Family Planning* topic.

Topics	Themes	2010	2011	2012	2013	2014	2015	2016	2017
Child Maltreatment	Abuse, violence, harm, and subordination	7	7	4	6	5	4	6	2
	Children, youth, families and friends	0	2	3	1	1	0	0	0
	Health problems and risks	3	1	0	1	0	1	0	1
	Support and protection	3	1	2	1	2	3	1	1
	Total	**13**	**11**	**9**	**9**	**8**	**8**	**7**	**4**
Family Planning	Family planning and reproductive health	2	9	13	2	2	2	2	5
	Human and environment	0	0	1	0	0	0	0	0
	Population problems	0	1	0	0	3	0	0	0
	Total	**2**	**10**	**14**	**2**	**5**	**2**	**2**	**5**
Women's Health	Discrimination, violence, harm, and subordination	1	1	1	1	2	1	0	1
	Health problems and risks	0	1	0	2	0	3	0	1
	Medical and interdisciplinary subjects	1	0	1	4	2	1	0	1
	Support and protection	7	4	5	4	3	8	3	3
	Total	**9**	**6**	**7**	**11**	**7**	**13**	**3**	**6**
Four Topics	**Total**	**24**	**27**	**29**	**22**	**20**	**23**	**12**	**15**

Table 7. Number of Entries Created during the Investigated Time Periods

The number of new entries of the *Child Maltreatment* topic kept decreasing from 2010 to 2017. The AVHS theme had much more new entries than the other three themes from 2010 to 2017. In other words, it caused the growth of the Child-Maltreatment-related entries to a large extent. The CM-SP theme also played a relatively important role in the growth of the Child-Maltreatment-related entries compared with the remaining two themes.

The *Women's Health* topic's number of entries had a steady rise. The WH-SP theme contributed to the entry increase the most among the four themes every year from 2010 to 2017. In 2010 and 2015, more new entries were created for this theme compared with the

other years. Another special case is that for the MIS theme, 4 new entries were generated in 2013, which was larger than the other years.

4.1.2. Descriptive Results of Page Edits

In addition to the creation time, the page edits data were collected for all the relevant entries. The numbers of yearly page edits for each topic and theme were calculated based on the page edits data obtained. The results presented in this section are associated with RQ2b and RQ3. Table 8 displays the mean, standard deviation (SD), and minimum and maximum numbers of yearly page edits for each topic and theme.

Topics	Themes	Mean	Std. Deviation	Minimum	Maximum
Child Maltreatment	Abuse, violence, harm, and subordination	6457.25	1426.682	5493	9615
	Children, youth, families and friends	1202.88	259.243	941	1690
	Health problems and risks	2037	613.105	1414	3125
	Support and protection	1637	210.288	1458	2043
	All themes	**11334.13**	**2298.441**	**9391**	**16473**
Family Planning	Family planning and reproductive health	4403	1730.62	2787	8078
	Human and environment	2164.13	557.573	1571	3043
	Population problems	2472.25	515.704	1506	3044
	All themes	**9039.38**	**2213.684**	**6882**	**13653**
Women's Health	Discrimination, violence, harm, and subordination	2801.38	389.715	2257	3350
	Health problems and risks	1911.25	373.024	1345	2362
	Medical and interdisciplinary subjects	2219.38	325.985	1760	2648
	Support and protection	3104	338.152	2655	3583
	All themes	**10036**	**725.487**	**9265**	**10994**

Table 8. Descriptive Statistical Analysis Results of Yearly Edits for Each Topic and Theme

Figure 6 to Figure 9 show the changes of the yearly page edits from 2010 to 2017 for each topic and theme. The *X*-axes of these figures represent the year and the *Y*-axes represent the number of yearly page edits (NYPE). Figure 6 illustrates the NYPEs of the three selected

topics. This figure demonstrates that the *Child Maltreatment* topic (Mean=11334.13, SD=2298.441) had higher NYPE than the other topics except 2011 and 2015, although its NYPE declined during the investigated eight years. The *Family Planning* topic (Mean=9039.38, SD=2213.684) occupied the second or third place among the three during the investigated eight years except 2011 when it reached the first place, which was its first peak. The number of this topic's yearly page edits fluctuated from 2010 to 2017 and reached its second peak in 2015. Similar to the *Family Planning* topic, the *Women's Health* topic (Mean=10036, SD=725.487) received the second or third place from 2010 to 2017 except 2015, but it did not fluctuate as widely as the *Family Planning* topic, and its standard deviation was much smaller than *Family Planning*'s.

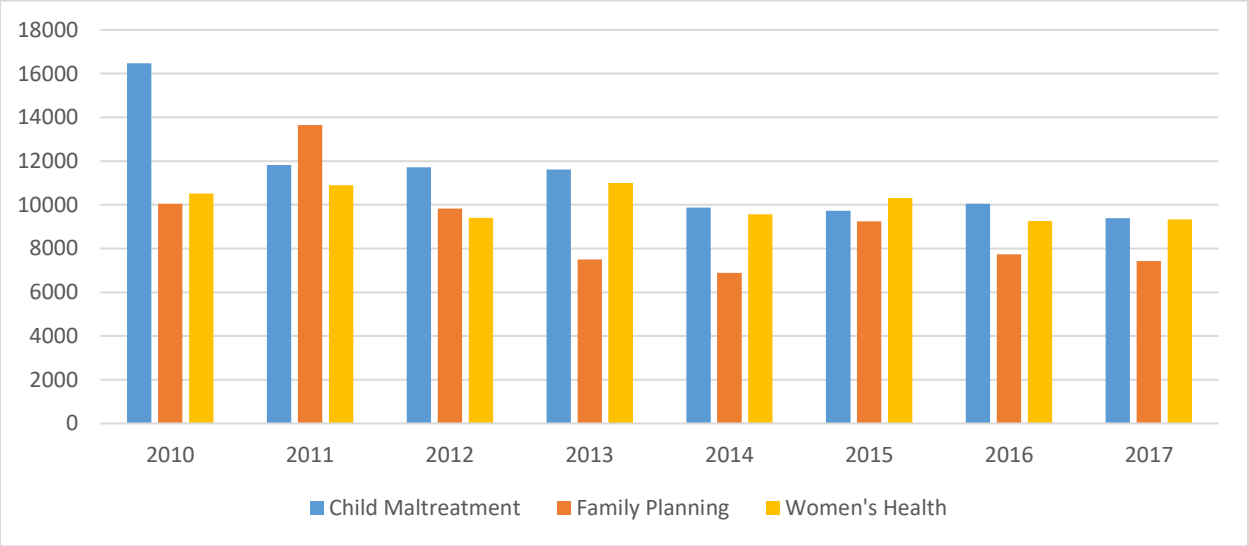


Figure 6. Numbers of Yearly Page Edits for Each Topic

4.1.2.1. The Child Maltreatment topic

Figure 7 illustrates the NYPEs of the four themes of the *Child Maltreatment* topic. This figure reveals that the *Abuse, violence, harm, and subordination* theme (Mean=6457.25, SD=1426.682) always had the highest NYPE among the four themes, while the other three themes each had less than half of the *Abuse, violence, harm, and subordination* theme's NYPE. It reveals that the *Abuse, violence, harm, and subordination* theme was the most popular among the four for Wikipedia editors, while the remaining three themes attracted much less attentions from the editors. The NYPE of the *Abuse, violence, harm, and subordination* theme decreased rapidly from 2010 to 2013 and dropped slightly after that. The NYPEs of the other three themes fluctuated from 2010 to 2017.

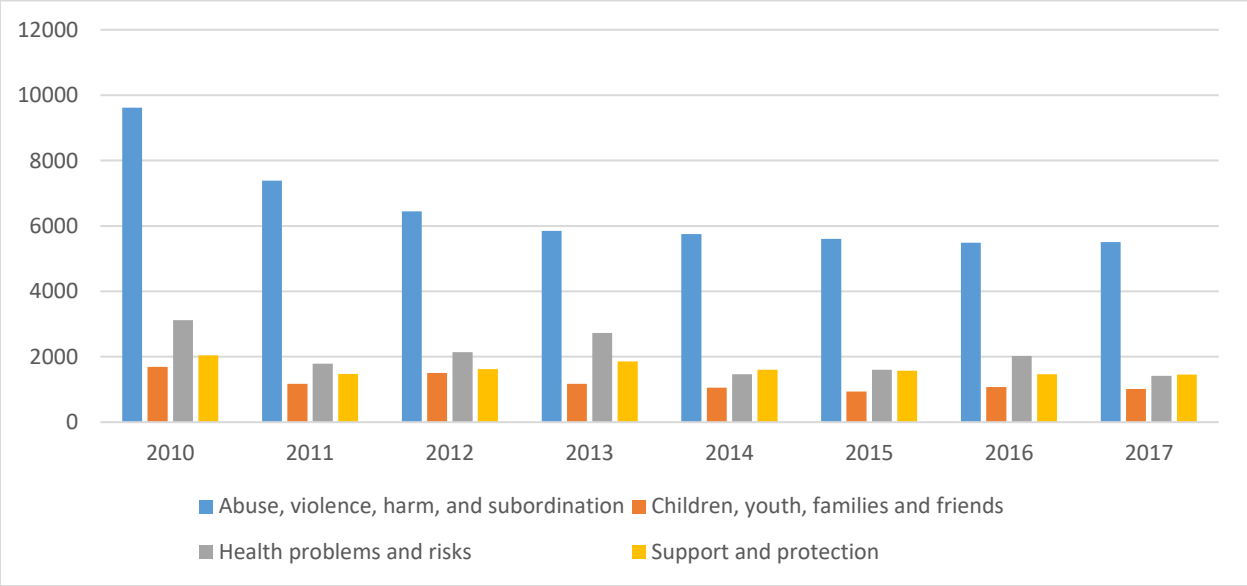


Figure 7. Numbers of Yearly Page Edits for Each Theme of Child Maltreatment

4.1.2.2. The Family Planning topic

Figure 8 demonstrates the NYPEs of the three themes of the *Family Planning* topic. The NYPE of the *Family planning and reproductive health* theme (Mean=4403, SD=1730.62) reached

its peak in 2011. The *Family planning and reproductive health* theme achieved the first position among the four themes from 2010 to 2017, although in the last two years its NYPEs were close to the NYPEs of the *Population problems* theme. The general trend of the NYPE of the *Human and environment* theme (Mean=2164.13, SD=557.573) declined during the investigated periods. The *Population problems* theme's NYPE (Mean=2472.25, SD=515.704) decreased from 2010 to 2013, but increased after that. Its NYPE was smaller than the *Human and environment* theme's NYPE from 2010 to 2013, but was larger than that from 2014 to 2017.

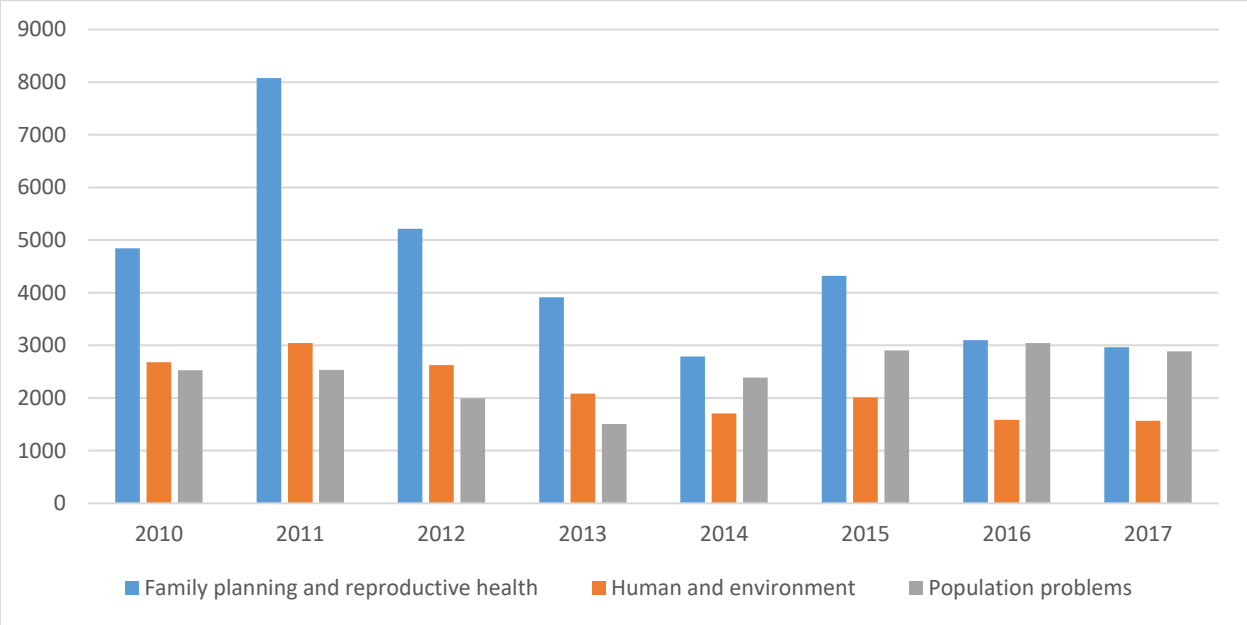


Figure 8. Numbers of Yearly Page Edits for Each Theme of Family Planning

4.1.2.3. The Women's Health topic

Figure 9 illustrates the NYPEs of the four themes of the *Women's Health* topic. This figure reveals that the *Support and protection* theme (Mean=3104, SD=338.152) received the largest NYPEs six years (2010 to 2012 and 2015 to 2017) among the investigated eight years. In

2013 and 2014 the *Discrimination, violence, harm, and subordination* theme (Mean=2801.38, SD=389.715) surpassed *Support and protection* and occupied the first position. The NYPEs of these two themes and *Medical and interdisciplinary subjects* (Mean=2219.38, SD=325.985) fluctuated from 2010 to 2017 and no obvious ascending or descending trend was found for them. The NYPEs of the *Health problems and risks* theme (Mean=1911.25, SD=373.024) rose from 2010 and reached peak in 2012, and then began to drop and reached its trough in 2017.

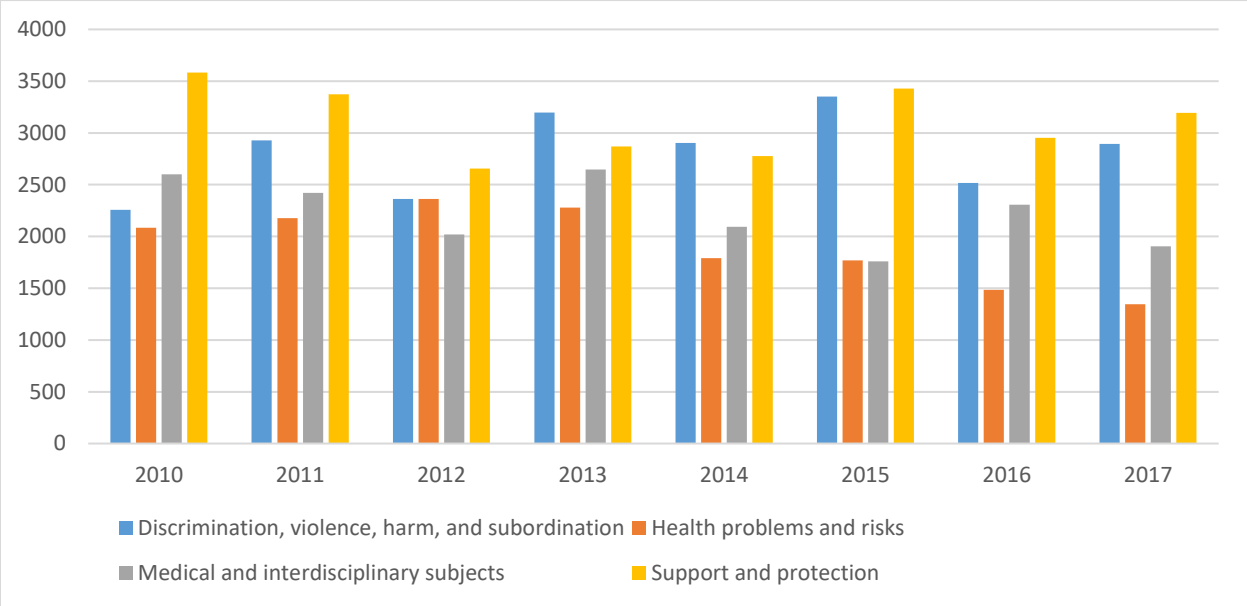


Figure 9. Numbers of Yearly Page Edits for Each Theme of Women's Health

4.1.3. Descriptive Results of Page Views

Apart from the page edits data, the page views data were collected for all the relevant entries. The numbers of yearly page views for each topic and theme were calculated based on the page views data obtained. The results presented in this section are associated with RQ2b and RQ3. Table 9 displays the mean, standard deviation (SD), and minimum and maximum numbers of yearly page views for each topic and theme.

Topics	Themes	Mean	Std. Deviation	Minimum	Maximum
Child Maltreatment	Abuse, violence, harm, and subordination	14344791	4322007	8756682	19658107
	Children, youth, families and friends	3027738	1035425	1670939	4419021
	Health problems and risks	6583360	2157364	3813014	9724321
	Support and protection	3839355	1160124	2131875	5203632
	All themes	**27795244**	**8449141**	**16372510**	**38037718**
Family Planning	Family planning and reproductive health	9219838	3274985	4686995	13168869
	Human and environment	4517944	1007057	3070288	5883305
	Population problems	6117393	985824.7	4886729	8113559
	All themes	**19855176**	**4744800**	**12644012**	**25124209**
Women's Health	Discrimination, violence, harm, and subordination	4817745	1599211	3193557	7613161
	Health problems and risks	4841570	1940105	2296627	7122197
	Medical and interdisciplinary subjects	6127769	1964491	3501912	8582655
	Support and protection	8539977	2922811	4622018	11876762
	All themes	**24327061**	**8041988**	**13614114**	**34816639**

Table 9. Descriptive Statistical Analysis Results of Yearly Views for Each Topic and Theme

Figure 10 to Figure 13 show the changes of the yearly page views from 2010 to 2017 for each topic and theme. The *X*-axes of these figures represent the year and the *Y*-axes represent the number of yearly page views (NYPV). Figure 10 illustrates the NYPVs of the three selected topics. This figure demonstrates that the *Child Maltreatment* topic (Mean=27795244.38, SD=8449140.605) had higher NYPV than the other topics during the investigated periods. The *Women's Health* topic (Mean=24327061, SD=8041987.722) and the *Family Planning* topic (Mean=19855175.5, SD=4744799.9) ranked the second and third places, respectively. The NYPVs of all the three topics reduced slightly from 2010 to 2011, increased from 2011 to 2013, and then decreased again from 2013. All the topics had their peaks in 2013.

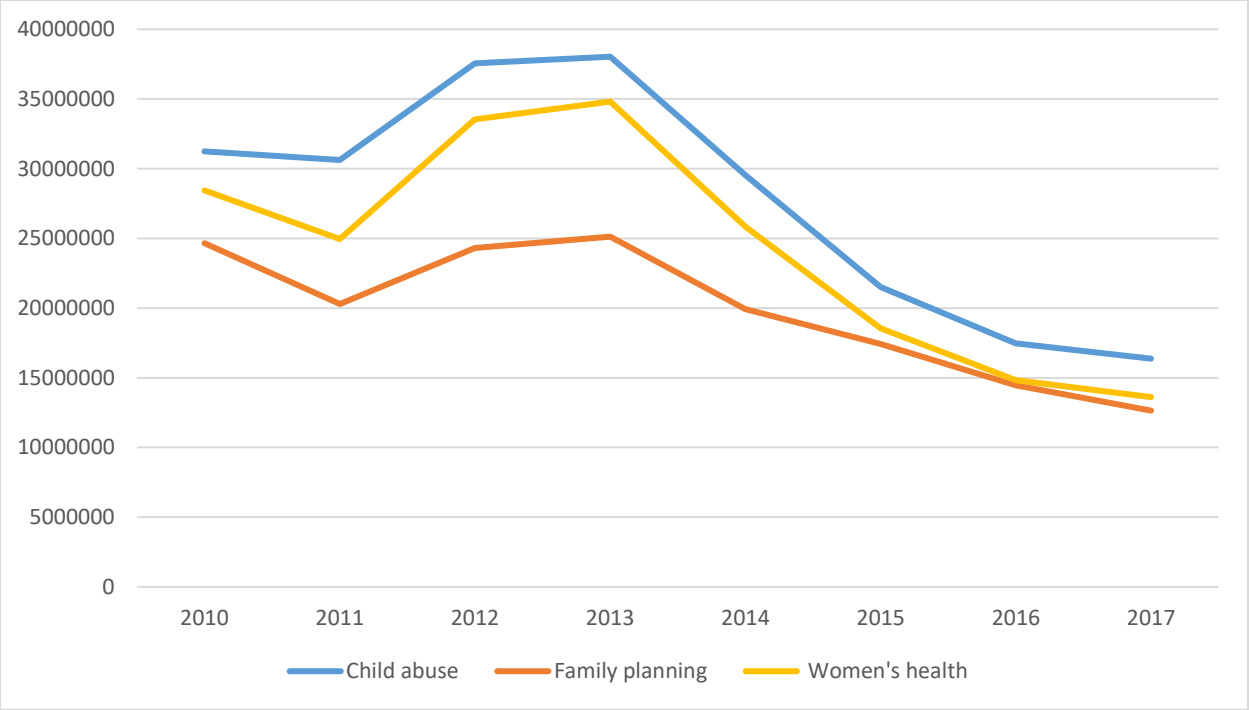


Figure 10. Numbers of Yearly Page Views for Each Topic

4.1.3.1. The Child Maltreatment topic

Figure 11 presents the NYPVs of the four themes of the *Child Maltreatment* topic. This figure reveals that the *Abuse, violence, harm, and subordination* theme (Mean=14344791.13, SD=4322007.268) and the *Health problems and risks* theme (Mean=6583360.25, SD=2157363.792) always had the highest and second highest NYPV among the four themes, respectively. The *Support and protection* theme (Mean=3839354.88, SD=1160124.038) ranked the third place most of the time except the first year (2010) when the *Children, youth, families and friends* theme (Mean=3027738.13, SD=1035424.697) achieved the third place. The NYPVs of the *Abuse, violence, harm, and subordination* theme, the *Health problems and risks* theme, and the *Children, youth, families and friends* theme increased from 2011 to 2012 and decreased

from 2013 to 2017. Different from the other three themes, the NYPV of *Support and protection* climbed to the peak in 2014 and declined since then.

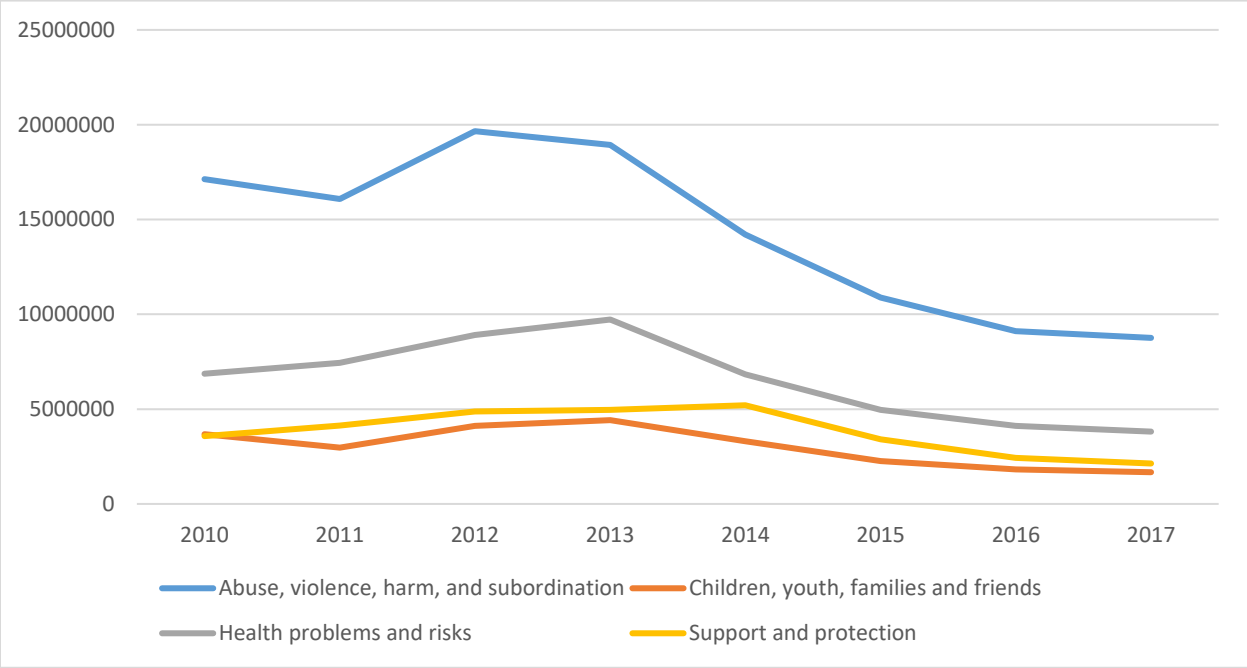


Figure 11. Numbers of Yearly Page Views for Each Theme of Child Maltreatment

4.1.3.2. The Family Planning topic

Figure 12 demonstrates the NYPVs of the three themes of the *Family Planning* topic. The trend of the *Family planning and reproductive health* theme (Mean=9219838.13, SD=3274984.835) was similar to the trend of the *Family Planning* topic and its NYPV was the highest among the three themes from 2010 to 2015. In 2016 and 2017, the *Population problems* theme (Mean=6117393.38, SD=985824.66) surpassed the *Family planning and reproductive health* theme and reached the first place. The *Human and environment* theme's NYPV (Mean=4517944, SD=1007057.393) was always the smallest among the three except 2014

when the *Population problems* theme's NYPV was even smaller. The overall trends of the three themes were all decreasing.

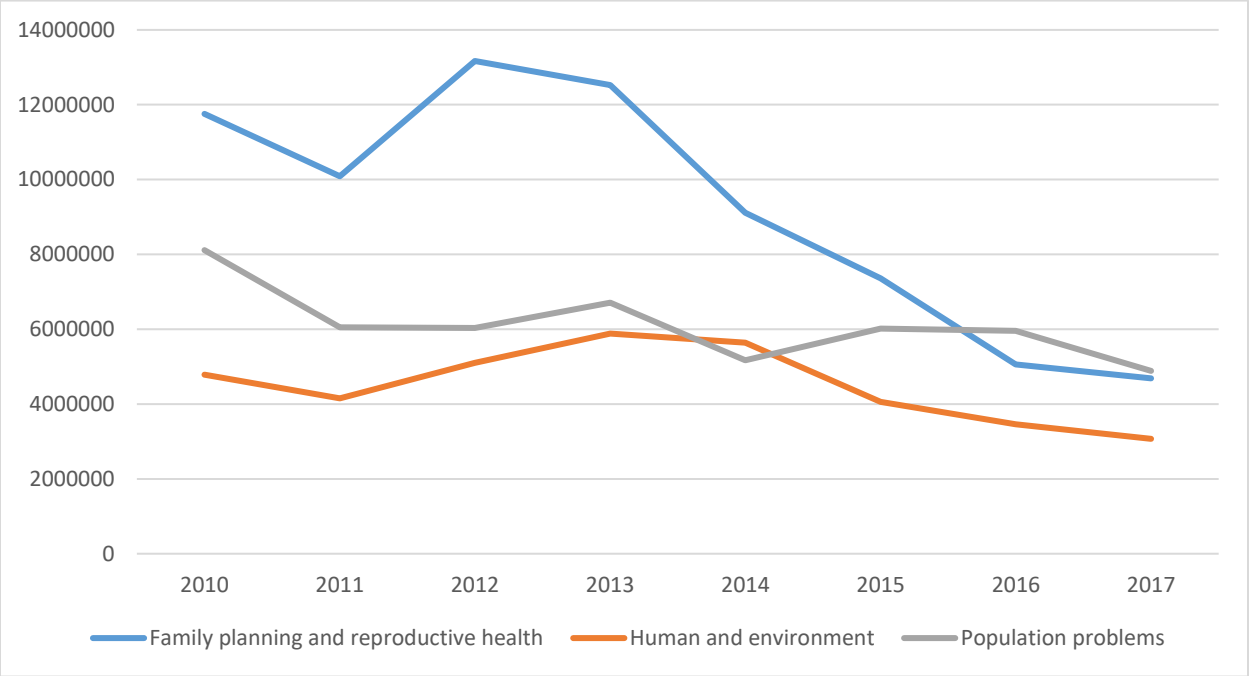


Figure 12. Numbers of Yearly Page Views for Each Theme of Family Planning

4.1.3.3. The Women's Health topic

Figure 13 displays the NYPVs of the four themes of the *Women's Health* topic. The trends of the *Support and protection* theme (Mean=8539976.88, SD=2922811.285), the *Medical and interdisciplinary subjects* theme (Mean=6127769.25, SD=1964490.902), and the *Health problems and risks* theme (Mean=4841570, SD=1940105.038) were similar to the trend of the entire *Women's Health* topic. The first two themes ranked the top two places among the four themes from 2010 to 2017. The third theme occupied the third place from 2010 to 2012, but fell to the last place from 2013. The trend of the *Discrimination, violence, harm, and subordination* theme (Mean=4817745, SD=1599210.84) was slightly different from the other

three themes because its NYPV increased rapidly from 2010 to 2013. However, the decreasing of its NYPV from 2013 to 2017 was similar to the other themes of the *Women's Health* topic.

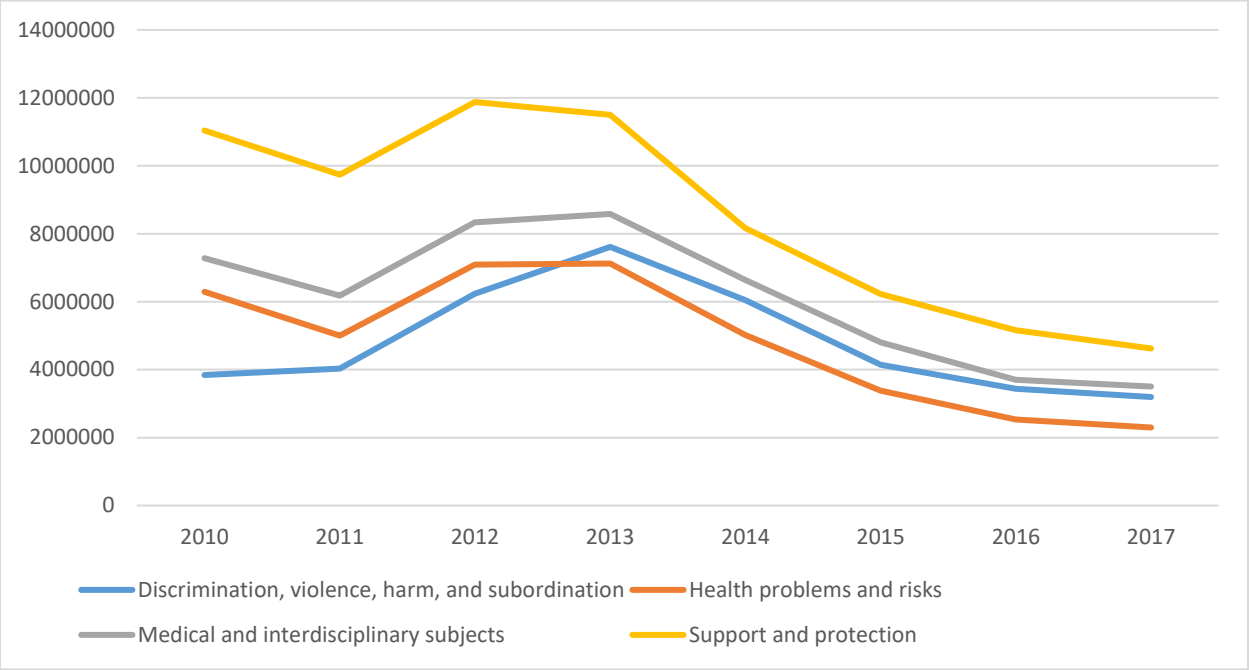


Figure 13. Numbers of Yearly Page Views for Each Theme of Women's Health

4.1.4. Descriptive Results Summary

Among all the three selected topics, *Child Maltreatment* was ranked the first place in terms of the numbers of entries and page views, while *Women's Health* and *Family Planning* occupied the second and third positions. Considering the number of the page edits, the ranking of the three topics changed frequently during the investigated time periods. For each topic or theme, its NYPE trend differed a lot from its NYPV trend from 2010 to 2017. These two types of the trends had different characteristics. It indicates that the user groups who created the page edits and the page views had different characteristics.

In this study, if a theme of a specific topic occupied the first position in most of the time according to the NYPE or the NYPV compared with the other themes, then it was named the salient theme of the topic. The trends of the selected three topics were consistent when looking at the yearly page views data. They rose and dropped in the same time periods. For each topic, the trend of its salient theme was consistent with the entire topic's trend. It implies that the Wikipedia viewers' interests in different family-health-related topics and themes changed in certain patterns and the Wikipedia viewers focused more on the salient themes than the other themes.

Different from the page views trends, the trends of the three topics and their themes varied a lot in terms of the NYPE. In other words, no consistent trend was found among the topics or themes. Among the three topics, only *Child Maltreatment* had a salient theme (*Abuse, violence, harm, and subordination*) which had the highest NYPE during the whole investigated time periods. These findings imply that the Wikipedia editors' interests on different family-health-related topics and themes varied from one to another and changed in different ways.

4.2. Results of Research Question One

Apart from the NYPEs and NYPVs of the selected topics, the changes of the text data of the selected topics can also reflect their evolutions. The first research question of this study aims to discover the associated entries and the emerged themes and subjects of each selected topic, and explore the connections among them. The associated entries and the themes of the selected topics were already demonstrated in Section 4.1. To discover the emerged subjects, the content of the associated entries for each topic was collected and examined.

4.2.1. Topics, Themes, and Associated Entries

As it was presented in Section 4.1, several themes were generated from the associated entries of each selected topic and the associated entries of each topic were collected. Appendix A lists the associated entries of each topic and theme. The numbers in the first and second column(s) of the table were the numbers of entries associated to the topics and themes. The abbreviations of the themes were contained in the second column. The entries of a specific theme were ordered alphabetically and numbered. The entries and their sequence numbers were shown in the third column of the table in Appendix A.

The historical versions of the associated entries were collected via RStudio and Wikipedia APIs. For each entry, its text of the last versions of 2011, 2013, 2015, and 2017 were collected. A special case emerged among all the associated entries, which was the "Wenatchee child abuse prosecutions" entry. All the versions created from May 2013 to November 2017 of this entry were removed and the content was reverted to an older version created in April 2013. Since the text data of this entry were incomplete, this entry was not included in the following subject analysis step.

4.2.2. Subjects Analysis Results

The SOM approach was applied to cluster the associated entries of each theme based on the matrices obtained. Each matrix was created for one theme, and it was utilized to generate one SOM display. Four, three, and four matrices were generated for the *Child Maltreatment* topic, the *Family Planning* topic, and the *Women's Health* topic, respectively. Therefore, in total eleven matrices were created in this study. Because the number of entries of

each theme differed from one another, so the size of the matrix obtained varied a lot. For example, the size of the *Abuse, violence, harm, and subordination* theme's matrix was 118×7570, while the size of the *Children, youth, families and friends* theme's matrix was 28×2343. Since the sizes of these matrices varied a lot, it was reasonable to process different matrices in different display sizes. Furthermore, the size of every SOM display was adjusted repeatedly so as to achieve the most interpretable results. Consequently, the sizes of the SOM displays obtained from the eleven matrices are different from each other.

Figures 14 to 24 are the SOM displays for the identified themes. The color bars on the right side of the figures represent different values of the U-matrix. Lower value means higher similarity. In the displays, every number stands for an entry, and the corresponding entry of each number is presented in Appendix A. The entries were grouped according to the criteria mentioned in Chapter 3. Every rectangle or polygon represents a cluster and the numbers in the same rectangle/polygon represent the entries belonging to the same cluster. The numbers not included in any rectangle/polygon stand for the isolated entries which were not grouped to any clusters. The clusters with more than three entries were the large clusters, while those with three or less entries were the small clusters. The large clusters were represented by purple rectangles or polygons. The small clusters were represented by red rectangles.

Appendix B lists the top 15 high-frequency terms and phrases in each theme. The terms and phrases in Appendix B illustrate that different themes had various high-frequency terms and phrases, which implies that the corresponding subjects could be different. For example, the terms/phrases in the *Abuse, violence, harm, and subordination* theme of *Child Maltreatment* were relevant to abuse and violence, while the terms/phrases in the *Children, youth, families*

and friends theme were about family issues, child protection, and child development. To further explore the subjects of each theme, the associated entries in each theme were clustered by the SOM approach and the high-frequency terms/phrases and the subjects of each cluster were extracted from the content of the entries.

Tables 10 to 20 display the high-frequency terms and phrases, and the subjects discovered within each large cluster. The high-frequency terms/phrases were extracted from the entries by the n-gram approach. This study only extracted the 2-word, 3-word, and 4-word phrases from the entries in each theme and cluster. The high-frequency terms and phrases are displayed in the second column of each table and the frequency of each term/phrase is included in the brackets following the term/phrase. The researcher proposed the subjects of each large cluster by examining the high-frequency terms and phrases of it. The small clusters and isolated entries were not included in these tables.

4.2.2.1. The Child Maltreatment Topic

(1) The Abuse, violence, harm, and subordination (AVHS) theme

Figure 14 presents the SOM display of the AVHS theme of *Child Maltreatment*. Seven large clusters and three small clusters were generated according to the clustering criteria. Cluster C2 was the largest cluster and Cluster C7 was the second largest one. Clusters C2, C3, C5, C6, and C7 were close to each other in the figure and were all located in the blue area, which implies that the entries in these clusters were relevant. Cluster C1 was a little far from the previous clusters, but since it was also in the same blue area, the entries in C1 were also relevant to those in the previous five clusters. Cluster C4 was far from the other large clusters

and there were yellow and red areas between it and the other large clusters. It implies that the entries in C4 had relatively weak connections with the entries in the other large clusters. Table 10 shows the high-frequency terms and phrases of the seven large clusters, as well as the subjects generated for them.

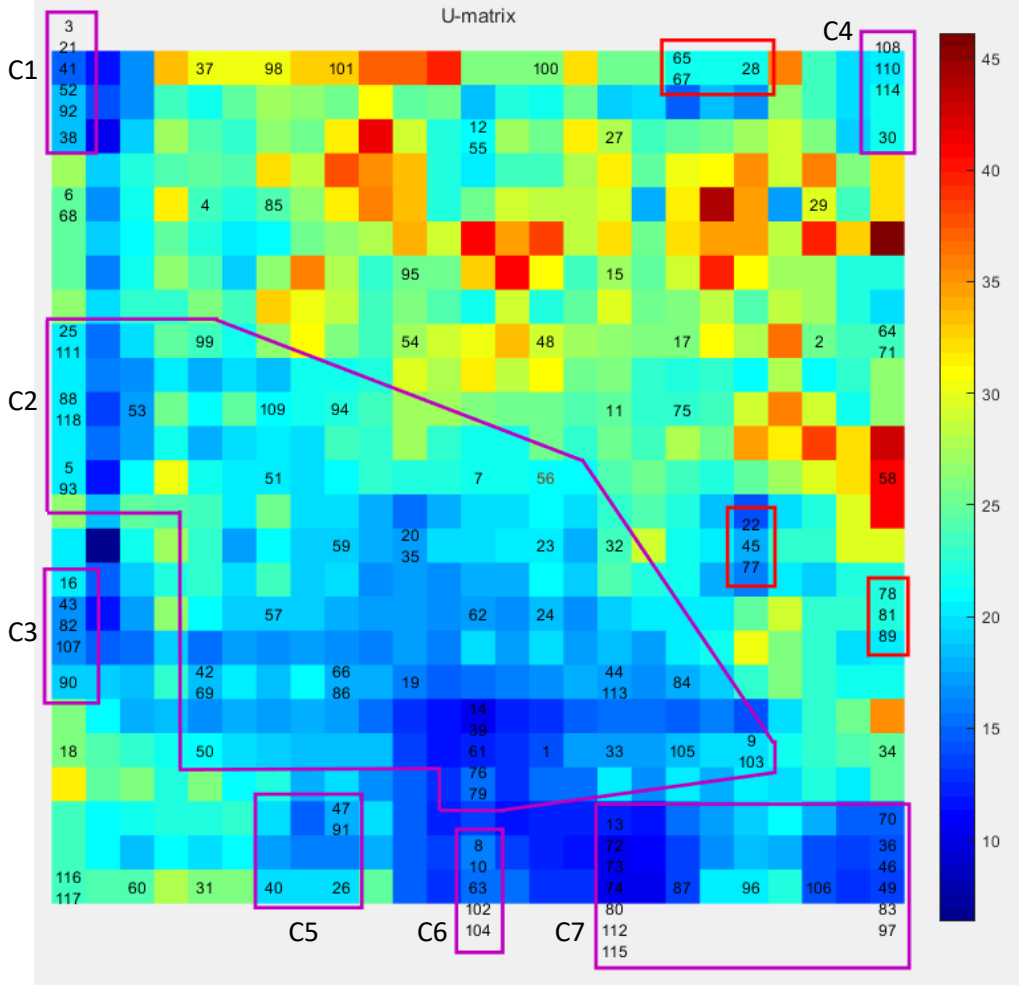


Figure 14. SOM Display of AVHS

Clusters	High-Frequency Terms and Phrases	Subjects
C1	domestic violence (282), child abuse (145), corporal punishment (123), violence against women (90), united states (88), human rights (70), intimate partner (67), partner violence (62), health organization (60), emotional abuse (56), intimate partner violence (55), World Health	Abuse and violence (domestic violence, physical abuse, emotional abuse),

	Organization (46), abuse and neglect (39), punishment of children (34), Council of Europe (26)	health issue (organization)
C2	sexual abuse (99), child abuse (94), social undermining (74), parental alienation (52), child neglect (45), child pornography (41), domestic violence (40), New York (36), mental health (34), social support (30), United States (24), child sexual abuse (19), the cruel mother (18), abuse and neglect (17), power and control (15)	Abuse and violence (sexual abuse, domestic violence, emotional abuse), social factor (social problem),
C3	sexual abuse (54), child sexual abuse (20), narcissistic abuse (16), domestic violence (16), physical abuse (11), breaking the cycle (10), child abuse (8), cycle of violence (6), child-on-child sexual abuse (6), family violence (5), stress disorder (5), mental health (4), abused children (4), narcissistic supply (4), abuse neglect (4), Alice Miller (4)	Abuse and violence (sexual abuse, domestic violence, physical abuse, emotional abuse), child and youth protection, health issue (problem)
C4	sex tourism (61), slave trade (59), child sex tourism (48), united states (75), sexual exploitation (35), human trafficking (32), sexual slavery (30), forced labour (29), sex slaves (27), human rights (26), World War (25), New York (24), child pornography (22), unfree labour (19), sex trafficking (18)	Abuse and violence (sexual abuse), slaves
C5	domestic violence (13), control domestic violence (8), Jill and Rob (7), Vanessa Jackson (6), power and control (5), attachment theory (4), Beth Thomas (4)	Abuse and violence (domestic violence), health issue (problem)
C6	sexual abuse (36), sexual activity (20), US Conference of Catholic (12), Child and Youth Protection (10), Fall River (9), child sex abuse (9), BBC News (7), Catholic diocese (6), abuse scandal (6), abuse cases (4), protection of children (4)	Abuse and violence (sexual abuse), child and youth protection (news report)
C7	sexual abuse (118), Jimmy Savile (87), child abuse (82), sexual exploitation (65), BBC News (57), child sex abuse (55), abuse scandal (37), child sexual exploitation (36), Daily Telegraph (35), North Wales (35), sexual activity (35), child sexual abuse (27), South Yorkshire Police (19), child abuse scandal (16), child sex abuse ring (14)	Abuse and violence (sexual abuse), child and youth protection (news report, organization)

Table 10. Subject Analysis of AVHS

The subjects listed in Table 10 shows that *abuse and violence* appeared in all the clusters, which means it was the most popular subjects among all the subjects of the AVHS theme. The abuse behaviors mentioned in this theme were mostly conducted to children. This

subject included several lower-level subjects, such as *sexual abuse*, *domestic violence*, *physical abuse*, and *emotional abuse*. Among these different types of abuse and violence, *sexual abuse* and *domestic violence* occurred in more clusters than the other types. Therefore, these two types of child abuse attracted more Wikipedia editors and viewers' attentions. For instance, the story of Beth Thomas, who was sexually abused as a child, was mentioned in an associated entry. Several entries were related to a series of the Catholic sex abuse cases or scandals, such as the sexual abuse scandal in Fall River diocese. *Domestic violence* included both physical (e.g. corporal punishment) and mental (e.g. neglect) abuse to children. Famous domestic violence cases, like the Collingswood Boys, were also demonstrated in the relevant entries of this theme. *Social factors*, such as social undermining, could be the potential cause of domestic violence.

Health issue and *child and youth protection* were other two popular subjects of the AVHS theme. The *health issue* subject had two lower-level subjects, which were *health problem* and *health organization*. Either physical or emotional abuse to children could cause health problems, such as stress disorder. To improve health conditions in the children, various organizations made their own contributions. The World Health Organization was one of the largest organizations helping developing child health world-widely. There were also various organizations for *child and youth protection*. The police (e.g. the South Yorkshire Police) and the news media (e.g. BBC News and the Daily Telegraph) played an important role in protecting children and youth. Usually the news media protected the children and youth by news reports. The remaining subject, *slaves*, was relatively special among all the subjects of this theme, because its associated entries were related to the history of specific countries. For instance, in ancient Rome it was common to employ female slaves for prostitution. During the World War II,

the government of Japanese organized a system of “comfort women” which was a euphemism of military sex slaves.

(2) The Children, youth, families and friends (CYFF) theme

Figure 15 presents the SOM display of the CYFF theme of *Child Maltreatment*, and two large clusters and three small clusters are illustrated in the figure. The two large clusters located quite far from each other and the orange, blue, and green colors filled the area between them, which implies that the entries in these two clusters had no strong association with each another. The high-frequency terms and phrases, and the subjects of the large clusters are demonstrated in Table 11.

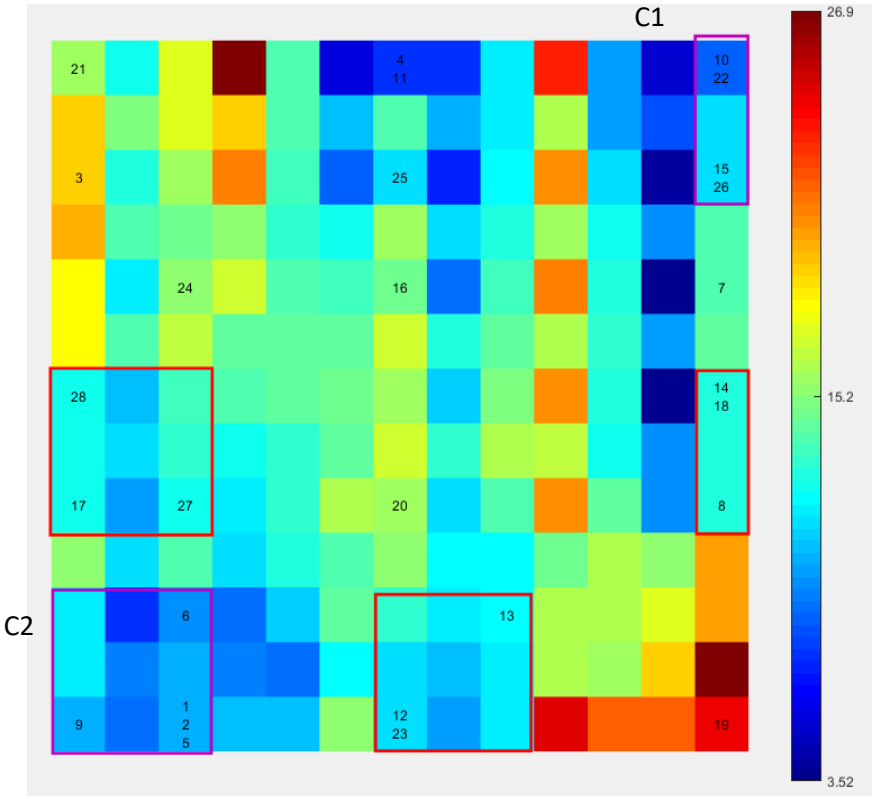


Figure 15. SOM Display of CYFF

Clusters	High-Frequency Terms and Phrases	Subjects
C1	United States (63), family members (24), immigrant families (16), men and women (16), nuclear family (15), domestic violence (13), gender equality (12), family issues (12), marriage and family (12), middle class (11), legal status (11), child care (11), American culture (10), family life (10), lower class (10), Journal of Family Issues (10)	Social factor (family issue)
C2	attachment styles (101), working models (53), attachment theory (24), relational schemas (21), attachment figure (20), personal relationships (18), adult attachment (16), Social Psychology (15), Personality and Social Psychology (15), secure attachment (13), attachment system (9), romantic relationships (8), child development (8), changes in attachment styles (8), Mikulincer Shaver (7), marital satisfaction (7)	Social factor (interpersonal relationship)

Table 11. Subject Analysis of CYFF

The subject of C1 was *social factor*, since all the high-frequency terms/phrases of Cluster C1 were about the society and culture. To be more specific, most of those terms/phrases of C1 related to certain *family issues*, such as immigrant families, domestic violence, lower class, and so on. These family issues might be caused by government policies, laws, economy status, culture, and so on. The family issues would potentially cause abuse and health problems.

All the high-frequency terms/phrases of C2 were associated with a psychology concept, attachment. The attachment theory is about interpersonal relationships. At first this theory only studied the context of child and parents, but in the past decades it was extended to adult relationships. Therefore, the subject of this cluster was also *social factor*, and to be more specific, the entries in this cluster focused on *interpersonal relationships*. Interpersonal relationship problems were potential causes of child maltreatment.

(3) The Health problems and risks (CM-HPR) theme

Figure 16 presents the SOM display of the CM-HPR theme and this figure shows three clusters. Clusters C1 and C2 were in the same blue area and although C3 located in another blue area, this area was connected to the C1 and C2's area. Therefore, the entries in these three clusters were relevant to each other. Table 12 displays the large clusters, the high-frequency terms/phrases, and the subjects of the large clusters.

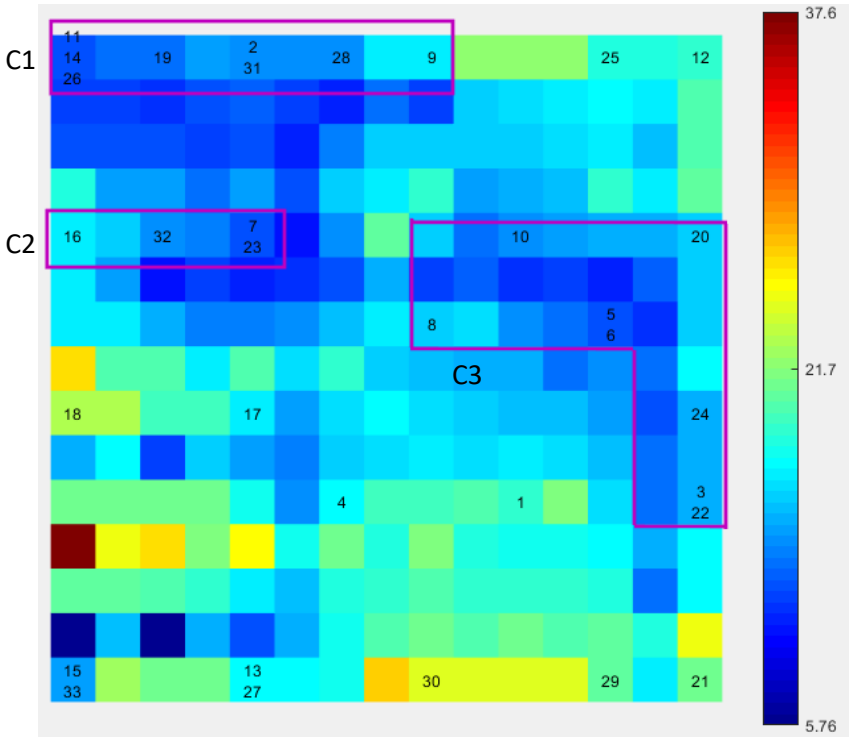


Figure 16. SOM Display of CM-HPR

Clusters	High-Frequency Terms and Phrases	Subjects
C1	emotional dysregulation (12), emotional regulation (8), mental health (10), personality disorder (7), traumatic bonding (6), narcissistic parents (6), self psychology (6), healthy narcissism (5), Geschwind syndrome (5), early childhood (5), spiritual crisis (5), bipolar disorder (4), borderline personality (4), spiritual emergency (4), narcissistic parenting (4), posttraumatic stress (4), narcissistic personality disorder (4), temporal lobe epilepsy (4)	Health issue (problem)

C2	vulnerable adult (23), neglected children (12), psychosomatic medicine (9), transactional analysis (8), child neglect (6), child abuse (5), psychosomatic disorders (4), family systems (4), Murray Bowen (4), stress disorder (4), posttraumatic stress (4), Karpman drama triangle (4), posttraumatic stress disorder (4), child abuse and neglect (4)	Health issue (problem, research), social factor (family issue), abuse and violence (emotional abuse)
C3	conduct disorder (100), borderline personality disorder (94), people with BPD (61), domestic violence (56), attachment disorder (52), emotion regulation (46), mental health (34), reactive attachment (32), reactive attachment disorder (30), traumatic stress (23), American Psychiatric Association (18), child and adolescent (18), antisocial personality (17), oppositional defiant disorder (17), posttraumatic stress disorder (15)	Health issue (problem, organization), abuse and violence (domestic violence)

Table 12. Subject Analysis of CM-HPR

The most salient subject of the CM-HPR theme was *health issue*, since it appeared in all the three large clusters. In all the three clusters, this subject had a lower-level subject, which was *health problem*. Many health problems were mentioned in the associated entries of C1, C2, and C3, including Geschwind syndrome, temporal lobe epilepsy, psychosomatic disorders, conduct disorder, and so on. Some entries in the three clusters were causes of health problems. For example, narcissistic parents and narcissistic parenting might affect children's behaviors and attitudes.

A lower-level subject of the *health issue* subject in C2 was *research*. It was found that an entry was psychosomatic medicine which explored the influence of social, psychological, and behavioral factors on humans and animals. The other two subjects of C2 were *social factor* and *abuse and violence*. Family issue was the lower-level subject of *social factor* and the high-frequency terms "family systems" and "Murray Bowen" were both relevant to family issue.

Murray Bowen proposed the theory of triangulation which was a part of Bowen's family systems theory. The health problems and social factors included in this cluster could be either the causes or results of child abuse.

The *health issue* subject of C3 had two lower-level subjects, *health problem* and *health organization*. The American Psychiatric Association mentioned in the associated entries of C3 was the largest professional psychiatric organization in the world. This organization not only trained psychiatrists but also published psychiatric journals and pamphlets so as to diagnose disorders and provide treatments. Another subject of this cluster, *abuse and violence*, could be the causes or results of disorders.

(4) The Support and protection (CM-SP) theme

Figure 17 presents the SOM display of the CM-SP theme and the clusters of this theme are illustrated in this figure. Four large clusters and four small clusters were generated for this theme. Cluster C1 and C3 were close to each other and located in the same blue area, which indicates their entries were also associated with each other. Cluster C2 was close to Cluster C1, but there was a small green area between them so that these two clusters were not quite similar. Cluster C4 was relatively far from the other three clusters and it located in the light blue area. Therefore, the entries in C4 had no strong connection with those in the other three clusters. Table 13 lists the clusters and their entries, high-frequency terms and phrases, and subjects.

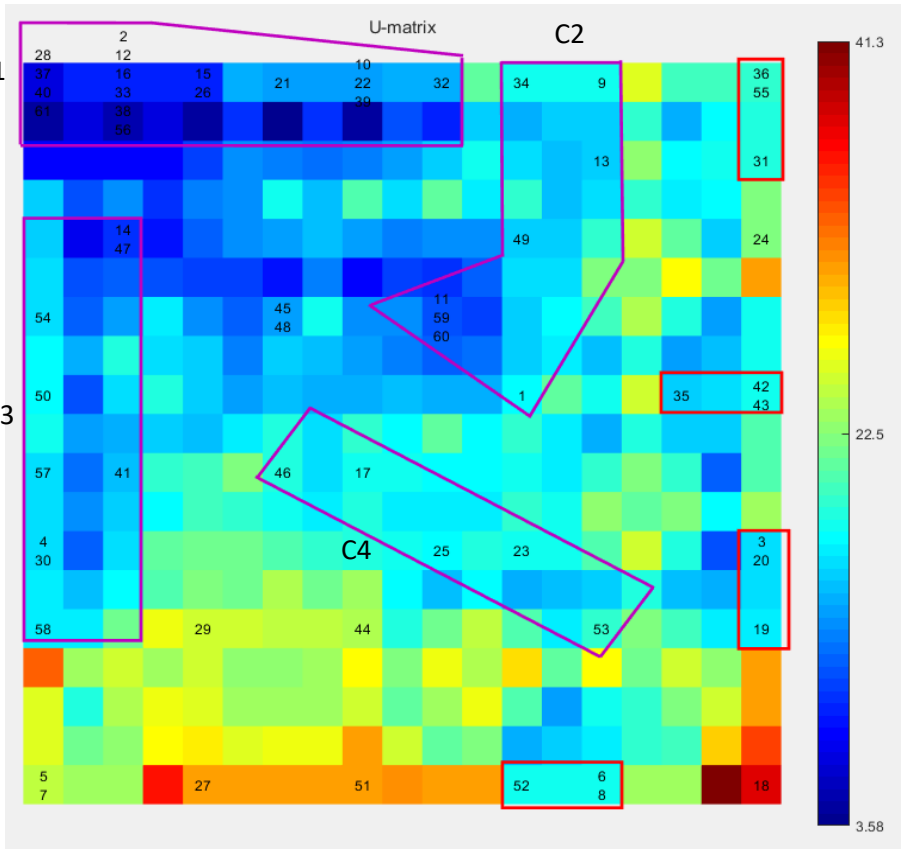


Figure 17. SOM Display of CM-SP

Clusters	High-Frequency Terms and Phrases	Subjects
C1	child abuse (57), child sexual abuse (29), human rights (28), child abuse and neglect (16), youth studies (15), Public Law (14), protection of children (13), Abuse Prevention and Treatment (12), Sexual Offences Act (11), Child Abuse Prevention (9), abuse laws (8), Mountain Goats (8), Save the Children (8), sexual abuse laws (8), Prevention and Treatment Act (8)	Abuse and violence (sexual abuse, emotional abuse), child and youth protection (law, work of art)
C2	corporal punishment (39), Prevention of Cruelty (23), Cruelty to Children (14), abuse excuse (11), abuse defense (8), vicarious liability (8), child protection (8), the Guardian (7), United States (7), National Society (7), content blocking (6), Supreme Court (6), ritual abuse (6), young people (5), Irish Society (5), social services (5), United Nations (5), criminal law (5), Internet Watch Foundation (5)	Abuse and violence (physical abuse, sexual abuse), child and youth protection (law, organization, technology, news report)
C3	child abuse (16), identified patient (8), mental health (7), Lloyd deMause (7), sexual abuse (7), mental disorders (7),	Abuse and violence (sexual abuse), child

	Journal of Psychohistory (7), trauma model (6), attachment theory (6), black women (6), child abuse investigation (6), traumatic experiences (5), personality disorder (4), attachment disorders (4), mental health professionals (4), Health and Human (4)	and youth protection (organization), health issue (research, organization)
C4	children's rights (75), human rights (28), false allegations (19), child sexual abuse (13), United Nations (12), child development (11), United States (10), child abuse (9), child wellbeing (9), parenting style (8), multisystemic therapy (8), mental health (7), health care (7), child rearing (6), child mortality (5), physical integrity (5), youth rights (5), American communities (5), parenting styles (5), mortality rate (5), child deprivation (5), Child Development Index (5), Council of Europe (5)	Child and youth protection (organization, research), abuse and violence (sexual abuse), health issue, social factor (family issue)

Table 13. Subject Analysis of CM-SP

The subjects, *child and youth protection* and *abuse and violence*, appeared in all the four large clusters, while different clusters had their own focuses. Cluster C1 focused more on the *laws* and *work of arts* relevant to child and youth protection. Some laws were enacted for children, such as the Child Abuse Prevention and Treatment Act, while some were not for specific groups, such as the Sexual Offences Act. In addition, an entry of C1 listed the songs about child abuse, including eight songs written by the Mountain Goats.

The *child and youth protection* subject of C2 was not only about *laws*, but also about *organizations*, *technology*, and *news report*. Different countries had various organizations for child and youth protection. For instance, there was an Irish charity (Irish Society for the Prevention of Cruelty to Children) that protecting the children and in the United States the Supreme Court of the United States also helped protect the children and youth. Some organizations like the Internet Watch Foundation provided technologies and information which contributed to child protection. For example, the Cleanfeed content blocking system used the

child abuse image content URL list offered by the Internet Watch Foundation to block child pornography. News media like the Guardian also made contributions to child and youth protection by news reporting.

Cluster C3 more focused on the *research* and *organization* of the *child and youth protection* subject. For instance, the child abuse investigation team in the United Kingdom was responsible for investigating child abuse and offenses related to minors. This cluster also referred to the *health issue* subject and this subject had two lower-level subjects, which were *health research* and *health organization*. Health professionals working in certain health organizations conducted research and published their findings (e.g. trauma model of mental disorders) in journals, such as the Journal of Psychohistory.

Compared with the other three clusters' subjects, C4's subjects were more diverse. Its entries were relevant not only to *child and youth protection*, but also to *abuse and violence*, *child development*, *health issue*, and *social factor*. The *child and youth protection* subject referred to both *organizations and research*. As it was mentioned before, the courts accused persons who acted child abuse. However, there were also false allegations which caused by revenge or other motivations. For the research perspective, the world's independent children's rights organization proposed the child development index to improve the wellbeing of children. Another subject of C4 was *health issue* and this subject related to the entries about health services, health care, and therapies.

4.2.2.2. The Family Planning Topic

(1) The Family planning and reproductive health (FPRH) theme

Figure 18 presents the SOM display of the FPRH theme and the clusters of this theme are illustrated in this figure. Nine large clusters and three small clusters were generated for this theme. Clusters C1, C2, C5, C6, C8, and C9 were close to each other and their locations were connected by blue areas, which means the entries in these clusters were quite relevant. Cluster C7 also located close to C6, C8, and C9, but there were yellow and green areas between C7 and the other three clusters. Meanwhile, C7 was far from C3 and C4. Therefore, the entries in C7 were not quite relevant to those in the other clusters. Clusters C3 and C4 located near each other but far from the remaining clusters. Therefore, the entries in these two clusters had connections with each other but did not share a lot of commons with the entries in the other seven clusters. Table 14 lists the large clusters and their high-frequency terms/phrases and subjects.

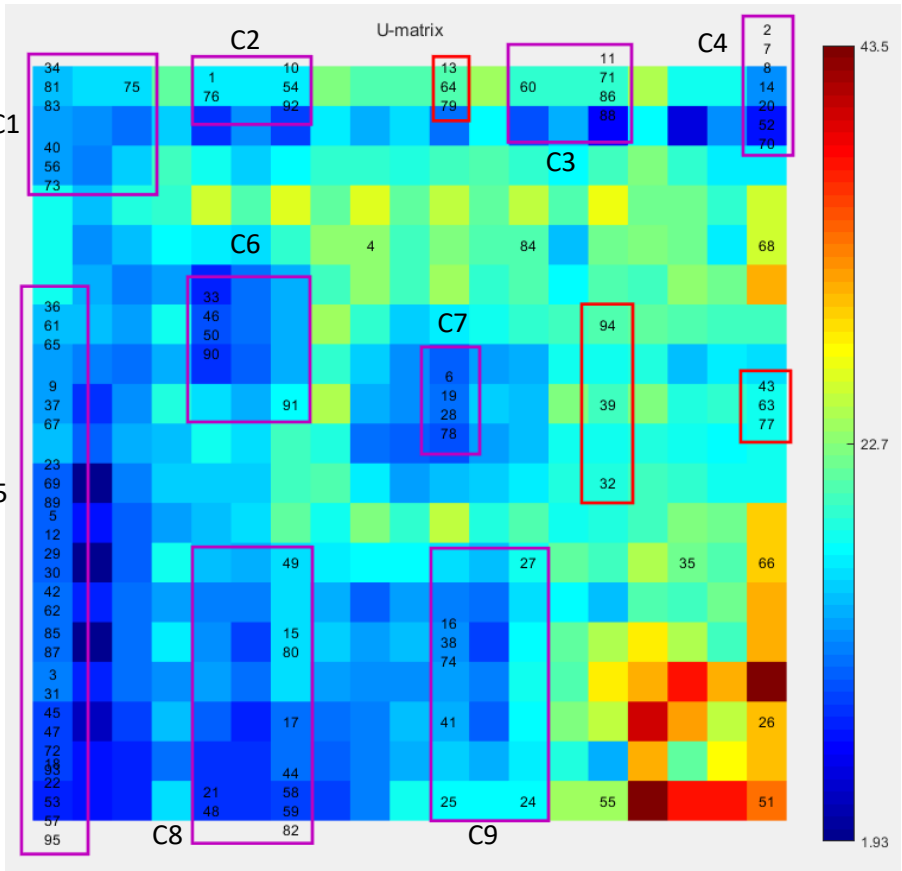


Figure 18. SOM Display of FPRH

Clusters	High-Frequency Terms and Phrases	Subjects
C1	female condom (47), reproductive health (40), sperm donation (33), donor sperm (33), sperm bank (30), sperm donor (21), birth control (21), human rights (19), potential donors (17), domestic violence (17), artificial insemination (15), teen dating violence (15), partner violence (13), United States (13), donor insemination (12), family planning (12)	Family planning and reproduction (method, device), abuse and violence (domestic violence, dating violence)
C2	reproductive rights (81), human rights (37), family planning (37), United States (33), oral health (29), health care (28), maternal mortality (24), birth control (22), maternal health (20), rhythm method (19), United Nations (19), health organization (19), sexual and reproductive health (17), maternal deaths (16), developing countries (14), calendar-based methods (14), standard days method (14)	Family planning and reproduction (method), health issue (problem, organization), woman protection (law)
C3	John Paul (61), birth control (53), family planning (51), Theology of the Body (39), Catholic Church (36), natural family (29), Supreme Court (27), natural family planning	Woman protection (law), family planning and

	(22), United States (18), reproductive health (10), sexual intercourse (10), Humanae Vitae (9), Quiverfull adherents (9), periodic abstinence (8), the Bible (8), reproductive rights (8)	reproduction (religion)
C4	birth control (364), United States (109), Margaret Sanger (92), New York (88), family planning (50), Planned Parenthood (36), reproductive health (35), health organization (33), induced abortion (23), latex condoms (18), first trimester (17), medical abortion (17), health care (17), unsafe abortion (17), Woman Rebel (16)	Family planning and reproduction (organization, method, device), health issue (problem), woman protection (law, news report)
C5	family planning (116), Title X (57), Planned Parenthood (44), health care (22), United States (22), reproductive health (19), sexual health (19), baby bonus (17), Hong Kong (16), family planning services (16), Family Planning Association (15), International Conference (14), sexuality education (14), Planned Parenthood Federation (14), sperm banks (12), public health (12), birth control (12)	Family planning and reproduction (policy, organization, method), health issue (service)
C6	Planned Parenthood (22), family planning (21), United Nations (18), father's quota (16), parental leave (10), paternity leave (8), United States (8), reproductive health (7), Conservative Party (5), Population Fund (5), Nations Fund for Population (5), Jens Stoltenberg (4), Trivers–Willard hypothesis (4), coercive abortion (4), Democratic Party (4), human rights (4), State Department (4)	Family planning and reproduction (organization, policy, method, research), population issue (organization)
C7	family planning (63), domestic violence (32), sex education (30), birth control (13), Billings Ovulation Method (12), reproductive health (9), violence during pregnancy (9), married women (8), cervical mucus (8), health survey (7), public health (7), research and development (7), development solutions (6), Family Planning Association (6), Research and Development Solutions (6)	Family planning and reproduction (education, method, policy), abuse and violence (domestic violence), health issue
C8	family planning (41), reproductive health (13), leave of absence (12), Family Planning Commission (9), Population and Family Planning (9), Health and Family Planning (8), Rama Rau (6), commuted leave (6), Journal of Family Planning (6), Commodity Security (5), reproductive healthcare (5), contraceptive security (5), International Development (5), National Population (5), Sexual and Reproductive Healthcare (5)	Family planning and production (organization, policy, research, technology), health issue (service), population issue
C9	reproductive health (44), family planning (39), Couple to Couple League (16), Deutsche Stiftung Weltbevoelkerung	Family planning and reproduction

	(15), Essential Medicines (8), Iran's population (7), United Nations (7), fertility rate (7), population growth (7), birth control (6), natural family planning (6), Reproductive Health Supplies Coalition (6), growth rate (5), birth rate (5), German foundation (5), sexual and reproductive health (5)	(organization), health issue (treatment), population issue (organization)

Table 14. Subject Analysis of FPRH

According to Table 15, the *family planning and reproduction* subject occurred in every cluster of the FPRH theme, so it was the salient subject of this theme. The associated entries of the *family planning and reproduction* subject covered information about the *family planning and reproduction methods* (e.g. sperm donation and artificial insemination), *devices* (e.g. condom), *woman protection* (e.g. reproductive rights), *organizations* (e.g. the Planned Parenthood Federation of America), *policies* (e.g. the father's quota), *research* (e.g. the Trivers-Willard hypothesis), *religion* (e.g. Catholicism), and *technologies* (e.g. the Strategic Pathway to Reproductive Health Commodity Security). Religion was a relatively special lower-level subject of the *family planning and reproduction* subject, since it only appeared in one cluster. Pope John Paul II gave a series of lectures, named the Theology of the Body, talking about the Catholic theology of sexuality. *Technology* was another special lower-level subject, which emerged from the entries in C8. A certain example was the Strategic Pathway to Reproductive Health Commodity Security tool that was developed to help countries improve contraceptive security.

Woman protection was another subject that had not been found in the previous topic and themes. This subject had two lower-level subjects, *woman protection law* (e.g. abortion rights) and *news report* (e.g. the Woman Rebel). Abortion rights were a kind of women's rights. The spread of women's rights showed the movements of woman protection. Furthermore,

news reporting played an important role in women's rights movements. For instance, the Woman Rebel was a newsletter that contained the discussion of contraception.

Apart from *family planning and reproduction* and *woman protection*, the entries of the FPRH theme related to *population issues*. There was a lower-level subjects generated from these entries, *population organization* (e.g. the United Nations Fund for Population Activities). The population organizations' goals included promoting population development, empowering women, protecting children and young people, improving the quality of people's life, and so forth.

Another subject, *health issue*, appearing in five clusters was also popular for the FPRH theme. Similar to the *health issue* subject of the themes for *Child Maltreatment*, this subject had *health problem* (e.g. unsafe abortion), *organization* (e.g. clinics), *service* (e.g. reproductive healthcare), and *treatment* (e.g. essential medicines) as the lower-level subjects. Moreover, two clusters had *abuse and violence* as a subject and the *abuse and violence* subject covered the information about *domestic violence* and *dating violence*. The generation of these two subjects shows the connection between the *Family Planning* topic and the *Child Maltreatment* topic: they had some common subjects.

(2) The Human and environment (HE) theme

Figure 19 presents the SOM display of the HE theme and the clusters of this theme are illustrated in this figure. Two clusters were generated for this theme. Since there were yellow and green areas between these two clusters, the entries of the two clusters did not have strong

connections. Table 15 displays these two clusters, the high-frequency terms/phrases and subjects of these clusters.

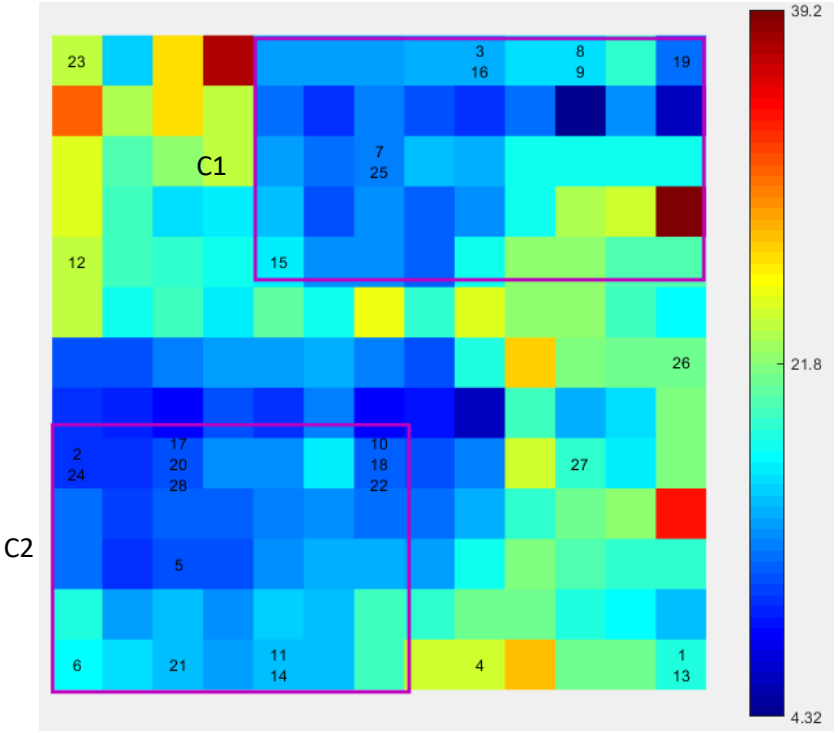


Figure 19. SOM Display of HE

Clusters	High-Frequency Terms and Phrases	Subjects
C1	futures studies (94), planetary boundaries (67), Gaia Hypothesis (62), climate change (46), tragedy of the commons (45), United States (28), earth system science (27), United Nations (21), carbon dioxide (16), global warming (14), existential risk (14), catastrophic risks (12), existential risks (11), artificial intelligence (11), population growth (11), human extinction (11)	Futures studies (technology), environment issue (research, problem), population issue
C2	United States (80), forced marriage (75), identity politics (42), financial crisis (35), fertility rate (31), human rights (25), reserve army (24), political demography (24), World Bank (23), men and women (23), maternity leave (22), people smuggling (21), New Feminists (21), migrant smuggling (19), Latin America (17)	Social factor, population issue, economy, military, politics

Table 15. Subject Analysis of HE

The subjects of the HE theme were quite diverse and were very different from the first theme of *Family Planning*. There was only one common subject found for the two clusters of the HE theme, which was *population issue*. Different population-related issues were mentioned by the associated entries of this subject, such as population growth, people smuggling, and so on.

In addition to *population issue*, Cluster C1 had two other subjects, *future studies* and *environment issues*. A widely discussed topic of *future studies* was artificial intelligence. For the *environment issue* subject, some entries were relevant to *research* (e.g. earth system science), while the others focused on existed or possible environment problems, like global warming and catastrophic risks. Cluster C2 had three special subjects, which were *economy* (e.g. World Bank and financial crisis), *military* (e.g. reserve army), and *politics* (e.g. political demography). The five subjects introduced in this paragraph were not found for any other themes, which indicates the uniqueness of the HE theme.

(3) The Population problems (PP) theme

Figure 20 presents the SOM display of the PP theme and the clusters of this theme are illustrated in this figure. Two large clusters and a small cluster were generated from this theme. The two large clusters located in the same large blue area, but there are some green cells between them, which means the entries of these two clusters were not quite relevant to each other. The two clusters' high-frequency terms and phrases, and subjects were shown in Table 16.

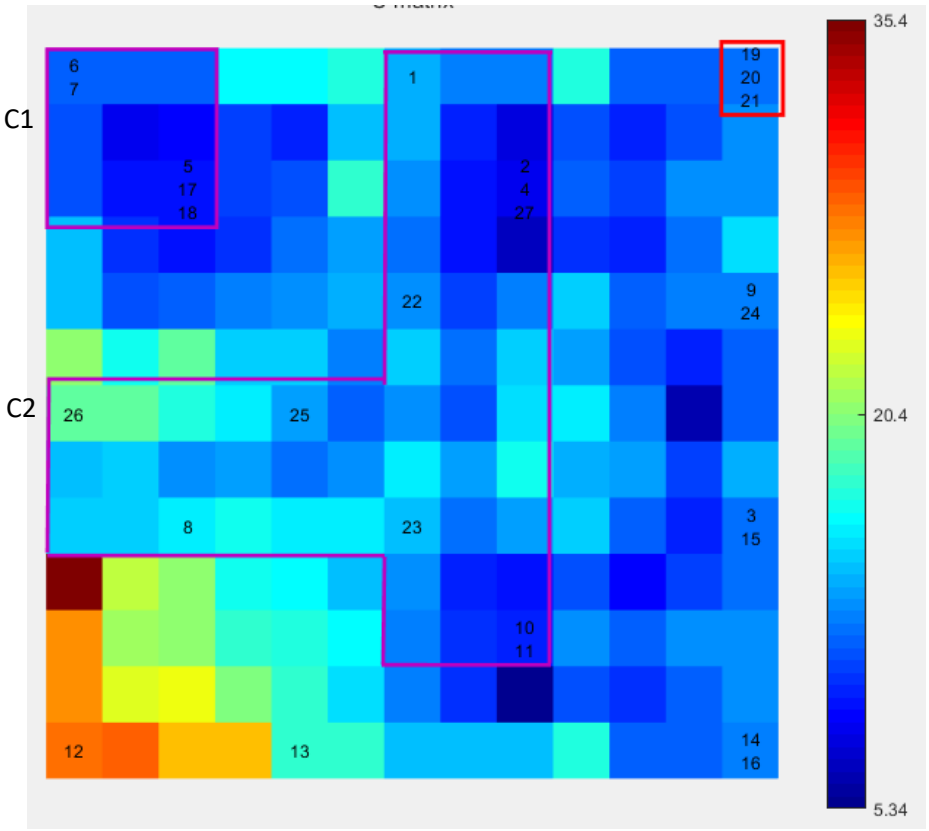


Figure 20. SOM Display of PP

Clusters	High-Frequency Terms and Phrases	Subjects
C1	population growth (75), world population (47), human population (33), United Nations (29), population planning (25), United States (21), birth control (16), family planning (16), developed countries (14), birth rates (14), carrying capacity (12), Club of Rome (12), developing countries (11), food production (11), population density (11), population control (11), international migration (11), human migration (11)	Population issue (problem, cause), family planning and reproduction (method)
C2	world population (100), sex ratio (91), population growth (63), official population (54), data templates (44), United Nations (29), country's population (28), United States (25), official sources (22), population density (22), manual calculation (22), population is not accurate (22), automatically calculate today's population (22), human population (19), population prospects (19), template without providing justification (19)	Population issue (statistics)

Table 16. Subject Analysis of PP

The subjects of the PP theme, *population issue* and *family planning and reproduction*, had been found within the previous themes, but some of its lower-level subjects were special and only appeared in this theme. For the *population issue* subject, the entries of this theme involved not only the population problems, but also the causes of the problems. For instance, the increase of food production and international migration could both cause population growth. Different from Cluster C1, Cluster C2 concentrated on population statistics. The entries in C2 related to data sources, calculation methods, and measures of population statistics, and the prediction of population as well.

4.2.2.3. The Women's Health Topic

(1) The Discrimination, violence, harm, and subordination (DVHS) theme

Figure 21 presents the SOM display of the DVHS theme and the clusters of this theme are illustrated in this figure. Four large clusters and two small clusters were discovered for this theme. Clusters C3 and C4 were all located in the same blue area. Therefore, the entries in these two clusters were relevant to one another. Table 17 lists the clusters and their high-frequency terms and phrases, and subjects.

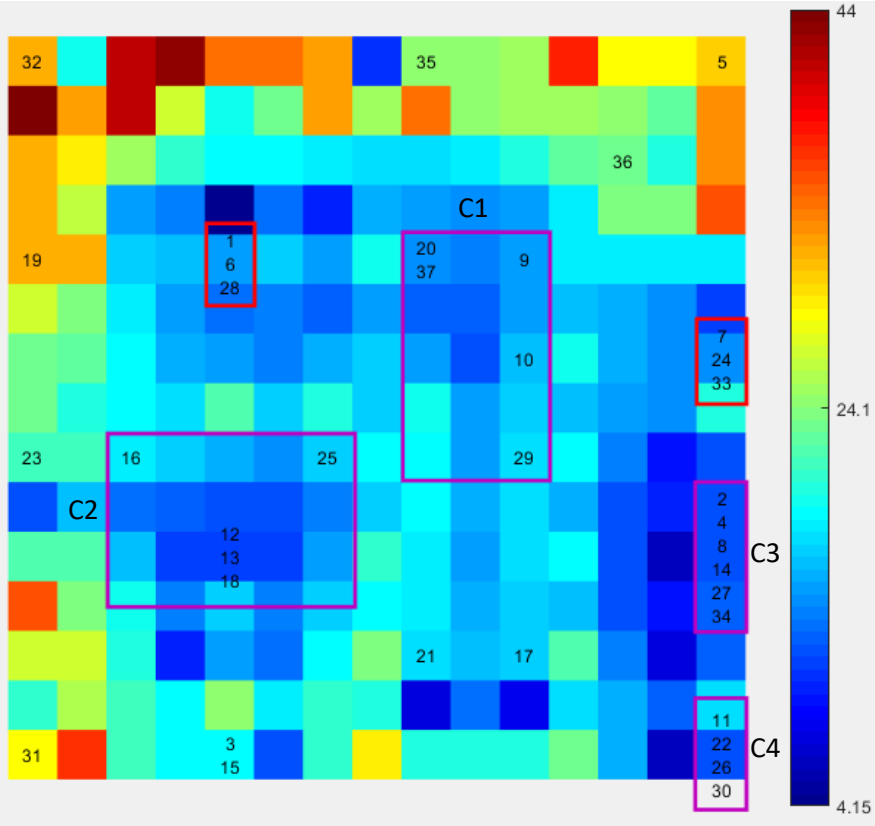


Figure 21. SOM Display of DVHS

Clusters	High-Frequency Terms and Phrases	Subjects
C1	sex ratio (81), missing women (65), gender inequality (47), gender gap (45), transgender people (43), men and women (40), United States (39), Gender Gap Report (25), sexual harassment (21), labor force (16), gender identity (16), female children (15), wage gap (15), World Economic Forum (15), birth sex (14)	Inequality and discrimination (healthcare, work), minority group (LGBT, woman), abuse and violence (sexual violence)
C2	rape culture (73), hegemonic masculinity (65), gay men (34), horror films (23), sexual violence (22), sexual assault (20), rape victims (17), sexual orientation (15), rape myths (15), South Africa (15), violence against women (11), slasher films (10), United States (9), victim blaming (9), mass media (9), LGBT people (9)	Minority group (LGBT, woman), abuse and violence (sexual violence, heterosexist violence)
C3	gender gap (22), law enforcement (20), female child (20), police officers (14), Wikimedia foundation (13), New York Times (13), Wikipedia editors (12), Silicon	Inequality and discrimination (work, research), minority

	Valley (11), gender bias (9), British Airways (8), New Zealand (8), black women (8), technology industry (7), female officers (6), sexual harassment (6)	group (woman), abuse and violence (sexual violence)
C4	glass cliff (28), triple oppression (20), black women (15), Communist Party (9), leadership positions (7), women executives (6), Michelle K (5), United States (5), black feminist (5), gender roles (4), occupational sexism (4), Claudia Jones (4), Socialist Party (4), reverse sexism (4), glass ceiling (4), Supreme Court (4), men and women (4)	Inequality and discrimination (work), minority group (black, woman), woman protection (organization, research)

Table 17. Subject Analysis of DVHS

Most of the subjects, except the *woman protection* subject, found within the DVHS theme were not revealed in the previous themes. The *minority group* subject occurred in all the four clusters of this theme, so it was the most dominant subject of the DVHS theme. The minority groups mentioned in the entries associated to this subject contained the LGBT people, women, and the black (African Americans).

The *inequality and discrimination* subject and the *abuse and violence* subject appeared in three clusters. The former subject had three lower-level subjects, which were *healthcare inequality and discrimination*, *inequality and discrimination in work*, and *inequality and discrimination in research*. *Healthcare inequality and discrimination* was reflected by the "missing women" phenomenon. As it was demonstrated in the "Missing women" entry, this phenomenon indicated that the number of women in a region was smaller than the expected number of women, which was caused by sex-selective abortion, female infanticide, and inadequate healthcare and nutrition for female children. *Inequality and discrimination in work* was revealed by the relatively low labor force participation rate of women. Ryan and Haslam (2005) proposed the concept of "glass cliff" which described the phenomenon that women

were likelier to be assigned precarious positions than men and thus had a higher risk of failure. Moreover, some research studies also focused on inequality. For example, a survey discovered that the Wikipedia editors were predominantly male. A general criticism of Wikipedia was its systemic gender bias that this platform provided less articles and information about women compared with men.

The *abuse and violence* subject had two lower-level subjects, sexual violence and heterosexist violence. The findings imply that these two types of violence were the most attractive violence-related subjects in the *Women's Health* topic. The associated entries of the *abuse and violence* subject mentioned not only different types of violence, but also the causes of the violence. For instance, rape culture was one of the main causes of high rape rates in certain countries, like India.

Women protection was another subject of the DVHS theme. Its associated entries were about the *organizations* (e.g. the Supreme Court of the United States) and *research* (e.g. the triple oppression theory) of women protection.

(2) The Health problems and risks (WH-HPR) theme

Figure 22 presents the SOM display of the WH-HPR theme and the clusters of this theme are illustrated in this figure. Two clusters were generated for this theme. These two clusters were close to each other and in the same blue area, which indicates that the entries in the two clusters shared some similarities. Table 18 presents the high-frequency terms/phrases and subjects of the two clusters.

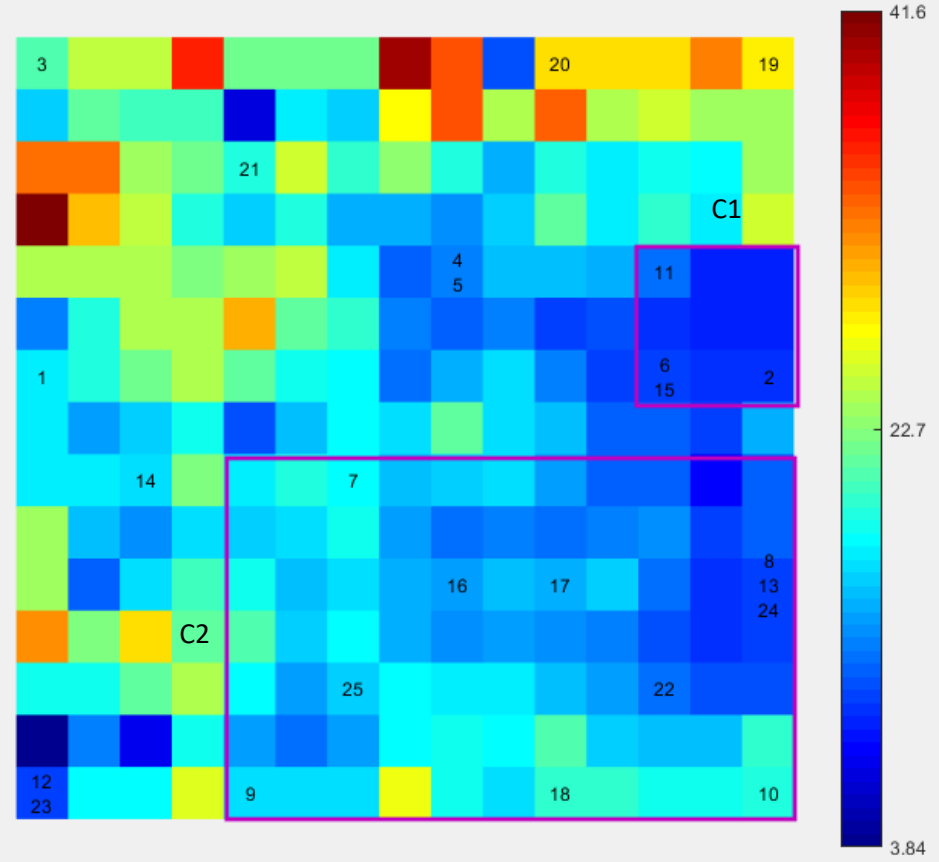


Figure 22. SOM Display of WH-HPR

Clusters	High-Frequency Terms and Phrases	Subjects
C1	blood pressure (20), high blood pressure (11), bacterial vaginosis (7), passive partner (7), active partner (6), chronic hypertension (5), diseases of affluence (5), sexually transmitted infections (4)	Health issue (problem)
C2	mental health (70), health care (53), United States (42), medical anthropology (39), public health (30), African Americans (23), heart disease (22), World Health Organization (19), gender equality (17), gender polarization (17), mental illness (16), sub-Saharan Africa (16), social class (15), men and women (15), women's health (14)	Health issue (service, research, organization, problem), inequality and discrimination (research)

Table 18. Subject Analysis of WH-HPR

The WH-HPR theme only had two subjects, *health issue* and *inequality and discrimination*. Cluster C1 mainly concentrated on *health problems*, such as high blood pressure

and sexually transmitted infections. In this cluster, a frequently-used synonym of high blood pressure was found, which was hypertension. The term "hypertension" was often used by health professionals and the terms "high blood pressure" were usually used by the lay people. Since Wikipedia is a user-generated platform, these two expressions were both utilized in the Wikipedia entries.

The *health issue* subject of Cluster C2 had four lower-level subjects, including *health service* (e.g. health care), *research* (e.g. medical anthropology), *organization* (e.g. the World Health Organization), and *problem* (e.g. heart disease). Another subject of this cluster was *inequality and discrimination* and this subject had a lower-level subject, *research*. For example, an entry of this subject was about the "gender polarization" concept proposed by American psychologist Sandra Bem (1995).

(3) The Medical and interdisciplinary subjects (MIS) theme

Figure 23 presents the SOM display of the MIS theme and the clusters of this theme are illustrated in this figure. Four large clusters and two small clusters emerged from all the entries of this theme. Two of the four clusters were either not close to each other or had green areas between them. These results show that the entries of the four clusters were not quite relevant. Table 19 lists the clusters, high-frequency terms and phrases, and subjects.

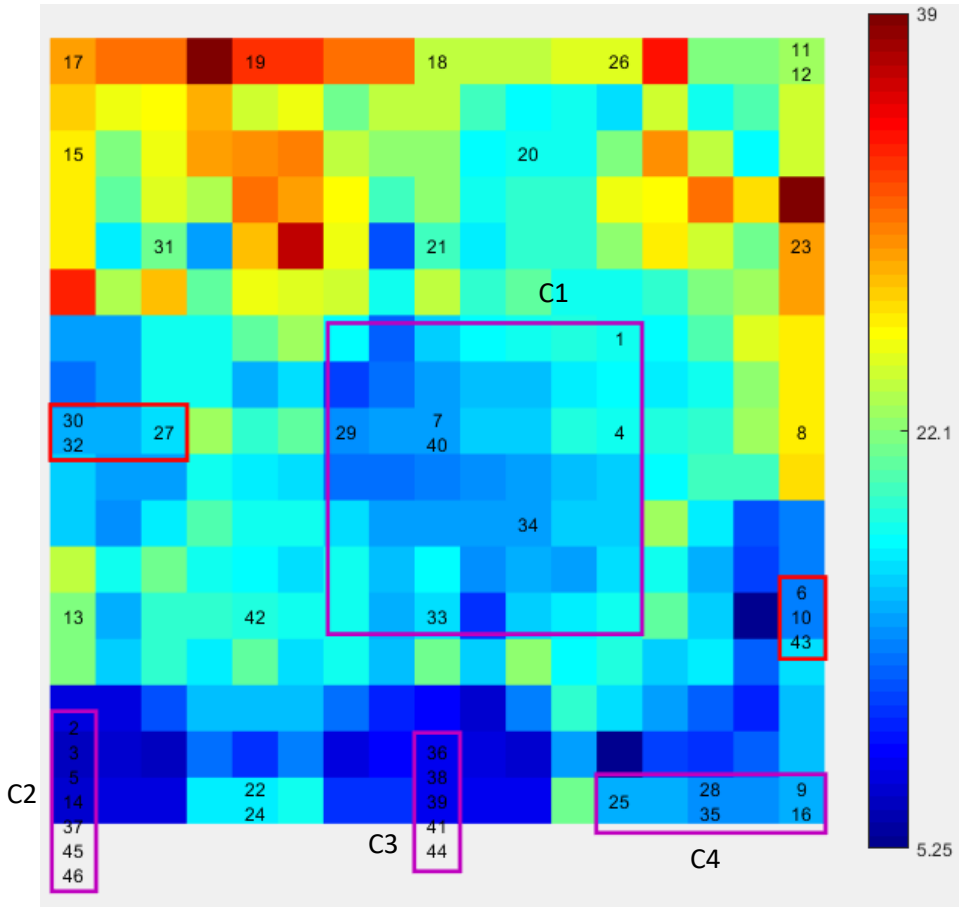


Figure 23. SOM Display of MIS

Clusters	High-Frequency Terms and Phrases	Subjects
C1	health care (68), family planning (68), United States (37), public health (32), health disparities (28), rural areas (24), health outcomes (24), structural violence (22), social determinants of health (21), World Health Organization (19), health equity (18), living conditions (16), spirit level (15), life expectancy (15), socioeconomic status (14), medical care (14), birth control (14)	Health issue (service, problem, research, organization), family planning and reproduction, abuse and violence (structural violence)
C2	United States (22), sexually transmitted disease (14), reproductive health (13), facial prominence (9), social epidemiology (8), intrauterine contraception (8), birth control (8), women's health (13), health and human (8), family medicine patients (8), hormonal contraceptives (7), CDC exploratory research (7), family planning (6), human services (5), reproductive age (5), bisexual women (5), sexual health (5), health and human services (5)	Health issue (problem, research, service), family planning and reproduction (method), minority group (LGBT)

C3	Whitehall Study (25), pelvic floor (23), reproductive health (21), Whitehall II (16), health care (16), reproductive rights (13), heart disease (13), Russian women (13), coronary heart disease (10), reproductive law and policy (8), Center for Reproductive Law (8), women's health (7), civil servants (6), social class (6), mortality rate (6), live births (6), risk factors (6), social determinants (6), blood pressure (6), pelvic floor muscles (6)	Health issue (research, problem, service), family planning and reproduction (law, organization), social factor
C4	social determinants of health (44), health care (43), population health (35), reproductive health (33), maternal mortality (30), oral health (29), World Health Organization (23), family planning (23), maternal health (20), public health (16), maternal deaths (14), health services (14), prenatal care (13), developing countries (12), United Nations (12), United States (12)	Health issue (problem, service, organization), population issue (problem), family planning and reproduction

Table 19. Subject Analysis of MIS

The *health issue* subject and the *family planning and reproduction* subject appeared in all the four clusters, but each cluster of the MIS theme had their own unique subject: C1 had the *abuse and violence* subject, C2 had the *minority group* subject, C3 had the *social factor* subject, and C4 had the *population issue* subject. All the subjects of this theme were discovered within the previous themes, while there was only one lower-level subject emerged from this theme, which was the *structural violence* subject. Different from the previous types of violence, structural violence was caused by social structure or social institution.

(4) The Support and protection (WH-SP) theme

Figure 24 presents the SOM display of the WH-SP theme and the clusters of this theme are illustrated in this figure. Eight large clusters and five small clusters were discovered for this theme. Clusters C1 to C7 were all located in the same blue area, it means that their entries had similarities to some extent. Cluster C8 stayed in another blue area, and the yellow and green areas separated it from the other clusters, which means its entries had no strong connections

with the entries of the other clusters. The eight large clusters and their high-frequency terms/phrases and subjects were displayed in Table 20.

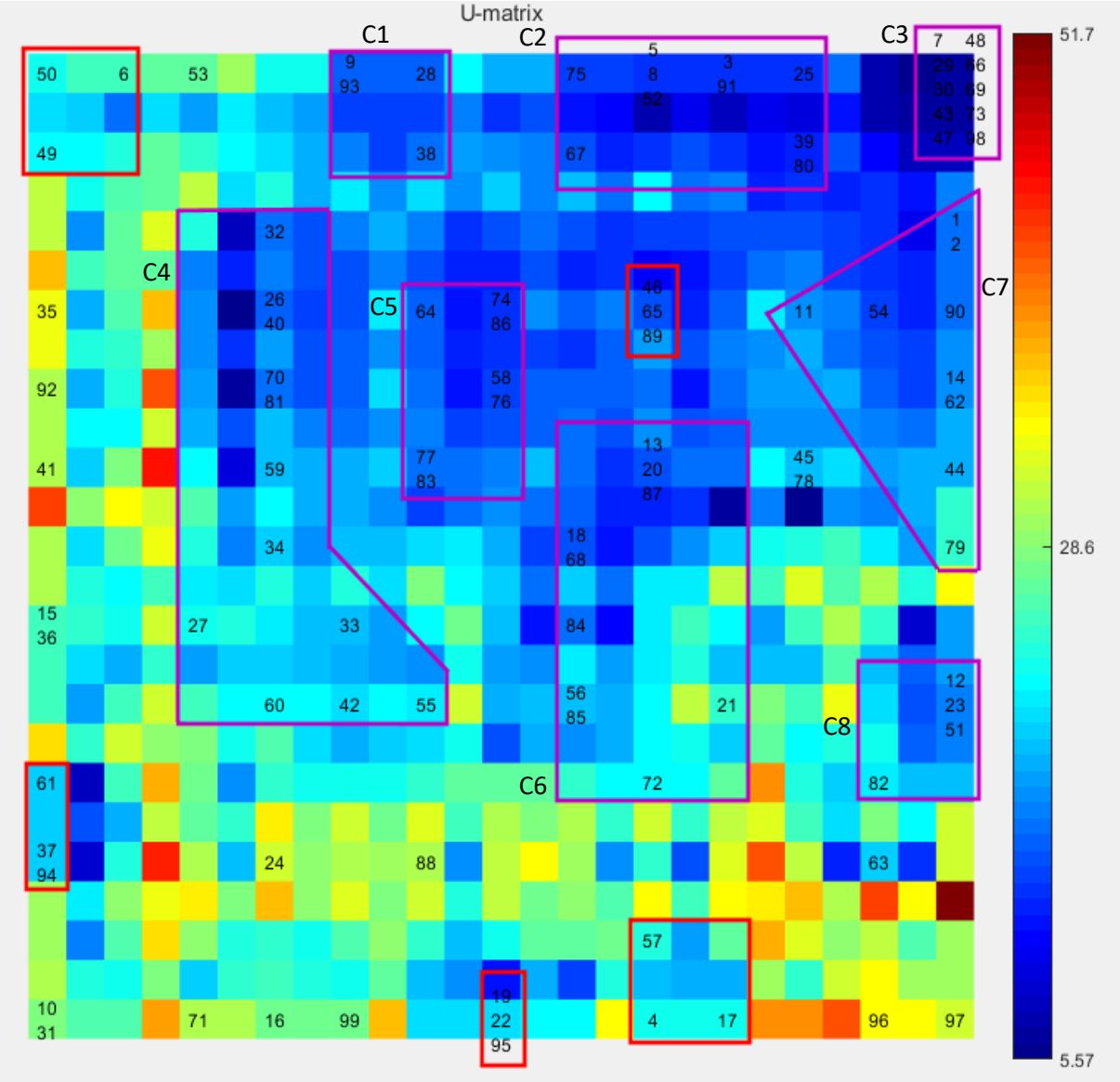


Figure 24. SOM Display of WH-SP

Clusters	High-Frequency Terms and Phrases	Subjects
C1	healthy people (17), Department of Health (12), black women (9), Health and Human Services (9), women's health (8), disease prevention (6), Health Promotion (5), Human Services Office (5), Office of Disease Prevention (5)	Health issue (organization, problem, service),

		minority group (black)
C2	Women's health (42), Center for Women's Health (14), Health Sciences (13), Health Centre (10), Women's Hospital (10), AnMed Health (7), health services (7), OHSU Center (6), women and newborns (6), Christie Street (5), Cancer Centre (4), World War (4), health care (4), St. Michael's Hospital (4), Honest Body (4), maternity hospital (4), The Huffington Post (4), Sunnybrook Health Sciences Centre (4), obstetric and neonatal nurses (4)	Health issue (organization, service)
C3	Women's health (38), health care (11), women's studies (10), Journal Citation Reports (8), Care for Women (6), healthcare journal (5), impact factor (5), women's studies journals (4)	Health issue (service, research)
C4	health education (138), health literacy (105), health disparities (83), public health (69), health system (44), reproductive health (43), University of Michigan (33), health care (32), Medical School (31), health promotion (28), United States (25), New York (24), Performance Indicators for Grades (24), school health (23), health supplies (23)	Health issue (education, problem, organization, service)
C5	United Nations (60), women's health (59), public health (41), health care (29), Sutter Health (22), UN Foundation (18), New York (15), health research (14), health network (12), University of Pittsburgh (11), San Francisco (10), Women's Health Research (10), Society for Women's Health (10), Medical Center (9), United States (9), Pittsburgh Graduate School (9), School of Public Health (9)	Health issue (service, organization, research, education), population issue (organization)
C6	women's health (98), United States (57), public health (40), health care (39), health education (35), Women's Health Center (34), Health and Human Services (26), social security (22), College Hospital (20), feminist health centers (20), United Nations (17), Department of Health Education (16), Federal Security (13), Our Bodies Ourselves (13), Health Education and Welfare (11), Boston Women's Health Book (11)	Health issue (service, education, organization, law)
C7	Women's health (24), Planned Parenthood (21), Heart Truth (20), Red Dress Collection (17), Conference on Planned Parenthood (10), American Medical Women's Association (10), United Nations (7), United States (7), New York (5), National Organization (5), First Ladies (5), sexual health (5), health policy (5), heart disease (5), National Organization for Men (5), Institute of Women's Health (5), Women's Health and Development (5)	Health issue (organization, law, problem, research), family planning and reproduction (organization)

C8	equity feminism (9), gender feminism (7), tobacco companies (5), United States (4), female physicians (4), Hoff Sommers (4), medical school (4)	Woman protection (activity, research)

Table 20. Subject Analysis of WH-SP

Table 20 shows that the *health issue* subject appeared in the first seven clusters (Clusters C1 to C7) of the WH-SP theme, more than any other subjects of this theme. Therefore, the *health issue* subject was the salient subject of the WH-SP theme. This subject had several lower-level subjects, including *health organizations* (e.g. OHSU Center for Women's Health), *services* (e.g. obstetric and neonatal nurses), *research* (e.g. women's studies journals), *problems* (e.g. heart disease), *education* (e.g. Performance Indicators), and *laws* (e.g. Social Security Act). Although all the lower-level subjects except *health laws* were found in the previous themes, these lower-level subjects were related to some different content compared with those lower-level subjects of the previous themes. For instance, the *health education* subject of this theme covered the content about the performance indicators which were used for student assessment. The *health law* subject, which only occurred in this theme, referred to the entries of health-related laws and policies, such as the Social Security Act and the policies developed by the European Institute of Women's Health.

The *minority group* subject, the *population issue* subject, the *family planning and reproduction* subject, and the *woman protection* subject only appeared in one cluster, respectively. Different from the other three subjects, the *woman protection* subject did not occur together with the *health issue* subject, which indicates that there was no strong connection between C8 and the other seven clusters. The entries in C8 were relevant to *woman protection activities* and *research*. For example, many theorists proposed a series of feminism

theories (e.g. liberal feminism and gender feminism) so as to fight against gender inequality. A certain instance was the history of women fighting for equal smoking rights.

4.2.3. Research Question One Results Summary

According to the results, *Child Maltreatment*, *Family Planning*, and *Women's Health* had 241, 150, and 207 associated entries, respectively. After examining the associated entries and their high-frequency terms/phrases of each theme, the subjects of each theme were generated. Tables 21 to 23 list the subjects of the identified themes and show the relations between the themes and subjects for each topic. Table 21 demonstrates that the AVHS theme and the CM-SP theme had various subjects, while the other two themes only had one or two subjects, respectively. The CYFF theme and the CM-HPR theme concentrated on specific subjects. Since the AVHS theme and the CM-SP theme had four common subjects, these two themes shared some similarities and their entries were relevant to each other.

Themes	**Subjects**				
	Abuse and violence	Health issue	Social factor	Children & youth protection	Slaves
Abuse, violence, harm, and subordination	√	√	√	√	√
Children, youth, families and friends			√		
Health problems and risks	√	√			
Support and protection	√	√	√	√	

Table 21. Subjects of Child Maltreatment

Table 22 shows that the FPRH theme and the HE theme had very diverse subjects, respectively. These two themes had only one common subject, which means that they were quite different from each other. Different from the other two themes, the PP theme had only two subjects, which implies that the associated entries of this theme focused on these two subjects. Among all the subjects, the *population issue* subject was the only one appearing in all

the themes. Therefore, *population issue* was the most popular subject of the *Family Planning* topic.

Themes	Subjects					
	Family planning and reproduction		Abuse and violence	Health issue	Woman protection	Population issue
Family planning and reproductive health	√		√	√	√	√
Human and environment						√
Population problems	√					√
-	Future studies	Environment issue	Social factor	Economy	Military	Politics
Family planning and reproductive health						
Human and environment	√	√	√	√	√	√
Population problems						

Table 22. Subjects of Family Planning

Table 23 shows that the DVHS theme, the MIS theme, and the WH-SP theme had more diverse subjects compared with the WH-HPR theme. In other words, the entries' subjects of the WH-HPR theme were more centralized than those of the other themes. Among the four themes, the MIS theme and the WH-SP theme had more common subjects. Meanwhile, every two of the four themes had one or more subjects in common with each other, which indicates that these themes were relevant to each other to some extent.

Themes	Subjects			
	Inequality and discrimination	minority group	violence	woman protection
Discrimination, violence, harm, and subordination	√	√	√	√
Health problems and risks	√			
Medical and interdisciplinary subjects		√	√	
Supports and protection		√		√
-	Health issue	family planning and reproduction	Social factor	population issue
Discrimination, violence, harm, and subordination				
Health problems and risks	√			
Medical and interdisciplinary subjects	√	√	√	√
Supports and protection	√	√		√

Table 23. Subjects of Women's Health

Comparing all the three family-health-related topics, the results illustrate that among the three topics, *Child Maltreatment* had less subjects than the other two topics, which means that its subjects were more centralized, while the other topics' subjects were more diverse. The *health issue* subject and the *social factor* subject occurred in all the three topics, which shows the connections among the three selected topics. Meanwhile, the other subjects reveal the special features of each topic.

4.3. Results of Research Question Two

Research question two intends to explore the evolution patterns of the internal characteristics and external popularities of each selected topic. The internal characteristics of a topic in a specific period was illustrated by the associated entries, subjects, and themes in the period. The external popularity of a topic in a specific period was presented by the total number of edits and views of its associated entries, and themes in the period. To explore the evolution patterns of a specific topic, its internal characteristics and external popularities in the four defined periods (2010 to 2011, 2012 to 2013, 2014 to 2015, and 2016 to 2017) were analyzed and compared.

4.3.1. Entry Growth in Each Period

As it was shown in Table 7 in Section 4.1, there were new entries created in each theme during the investigated eight years (2010 to 2017). Since no investigated entry was deleted from 2010 to 2017, the growth of the entries in each theme was revealed by the generation of the new entries. Therefore, Table 7 shows that the numbers of the associated entries in AVHS, CM-HPR, CM-SP, FPRH, DVHS, WH-HPR, MIS, and WH-SP increased in each defined period. The

number of the associated entries in HE kept the same in Period 1 (2010 to 2011). The numbers of the associated entries in CYFF, HE, and PP did not change during the fourth period (2016 to 2017). The increases of the entries in the defined themes caused the increases of the associated entries in the selected topics.

Appendix C displays the associated entries in each theme during each investigated period. Each number in the last four columns represents an entry in a specific theme and the corresponding entry is demonstrated in Appendix A. The data in Appendix C illustrates that the entries in all the themes increased from 2010 to 2017. These findings show the increases of the entries in all the selected topics. Therefore, both Appendix C and Table 7 reflect that the number of the associated entries in each topic kept increasing during the four periods. Appendix D demonstrates the new entries generated in each investigated time period. These new entries reveal the Wikipedia editors' new interests and focuses during each period.

4.3.1.1. The Child Maltreatment Topic

After examining the AVHS theme of the *Child Maltreatment topic*, it shows that the entries created during the first period mainly introduced different abuse types, such as narcissistic abuse, domestic violence, disability abuse, and so on. The other entries created in this period were related to child abuse cases or scandals reported in 2010 and 2011 or earlier (e.g. Collingswood Boys).

During the second, third, and fourth periods, there were two types of new entries: (1) the entries about child abuse cases or scandals reported in the three periods or earlier (e.g. Kasur child sexual abuse scandal); (2) the entries demonstrating the child abuse and child

protection status in specific regions (e.g. Child abuse in New Zealand). The rest entries presented some particular abusive behaviors (e.g. Isolation to facilitate abuse) or child abuse types (e.g. Athletes and domestic violence).

Regarding the CYFF theme, four new entries were created in the second period, much more than the other three periods. These four entries all concentrated on the attachment theory and family-related problems. The research of attachment theory was first introduced in 1960s and this theory has been developing for more than fifty years. The psychology of religion field adopted this theory in 1990 and subsequently its development expanded.

The CM-HPR theme had four new entries in the first period, and the rest three periods only had one new entry, respectively. The new entries in the first period were all relevant to the impacts of child abuse on children. The theories and knowledge included in these four new entries were all proposed much earlier than the time these entries created.

The new entries created in the CM-SP theme referred to research (e.g. journals), organizations (e.g. International Society for the Prevention of Child Abuse and Neglect), technologies (e.g. Child abuse image content list), works of arts (e.g. List of songs about child abuse), and laws and policies (e.g. Karly's Law) that protected children from being abused, as well as the treatment (e.g. Multisystemic therapy) that aimed to help the chronically violent youth. Among these entries, the "List of songs about child abuse" entry was special because it was a summary of child-abuse-related songs. Another special entry was National Child Abuse Prevention Month created in 2016. The National Child Abuse Prevention

Month was designated in 1983 and U.S. President Barack Obama issued a Presidential proclamation in 2016 which recommitted the observance.

4.3.1.2. The Family Planning Topic

Among the three themes of *Family Planning*, only the FPRH theme had relatively more new entries (37 entries) created from 2010 to 2017, the HE theme and the PP theme had only one and four new entries, respectively. For the PP theme, the entries created in the third period summarized the organizations and researchers (e.g. List of people that have expressed views relating to overpopulation as a problem, and List of population concern organizations) concerning population problems.

Regarding the FPRH theme, most of the entries generated in the four periods were about the family planning and reproduction organizations (e.g. Society for Family Health Nigeria), research (e.g. Journal of Family Planning and Reproductive Health Care), policies (e.g. Father's quota), and techniques and methods (e.g. Fertility monitor), and the family planning status in specific regions (e.g. Family planning in the United States) as well.

The background of the generation of the "Non-consensual condom removal" entry was special. This entry was created in 2017, but its synonym "stealthing" had been used in the gay community since 2014 or earlier. The concept and use of "non-consensual condom removal" and "stealthing" spread on Websites and forums since then. However, the public were not aware until an article about non-consensual condom removal was published and widely reported by news and media outlets in 2016 (Brodsky, 2017). The news reporting of this article triggered the generation of the Nonconsensual condom removal entry on Wikipedia.

4.3.1.3. The Women's Health Topic

Among the four themes of *Women's Health*, the WH-SP theme had much more new entries in the four investigated periods than the other three themes. The new entries in the DVHS theme were related to sexism, like sexism in workplace (e.g. Women in law enforcement) and sexism in specific regions (e.g. Discrimination against girls in India). The entries in the WH-HPR theme and the MIS theme focused on women's health issues, including the research of women's health issues (e.g. Women's Health Issues), the determinants of health issues (e.g. Social determinants of health in poverty), and women's health status in specific regions (e.g. Women's reproductive health in Russia).

The new entries in the WH-SP theme concentrated on the techniques and methods (e.g. Gynography), research (e.g. Black Women's Health Study), organizations (e.g. EuroHealthNet), training and education (e.g. Oregon Health and Science University Center for Women's Health), services (e.g. Midwife), and works of arts (e.g. The Honest Body Project) which aimed to support and protect women and improve women's health.

The generation of the "The Honest Body Project" entry was triggered by the creation of a project. This project was created by photographer Natalie McCain and the creator of the corresponding entry was NatalieRMcCain. The photographer's name and the creator's account indicate that it was Natalie McCain herself who generated the "The Honest Body Project" entry.

4.3.2. Changes of Subjects

The internal characteristics of the three selected topics in Period 4 (2016 to 2017) were demonstrated in Section 4.2. The entries in each theme were presented and the SOM display

and subjects of each theme were illustrated. To explore the internal characteristic evolution of each selected topic, the changes of the subjects in the topic from one period to another were explored. For each theme of a selected topic, the frequencies of its two-word/three-word/four-word terms and phrases were counted and the frequency difference of each term/phrase from one period to next period was calculated. In other words, for each term/phrase, the difference of its frequency in Period 2 and its frequency in Period 1, the difference of Periods 2 and 3, and the difference of Periods 3 and 4 were all calculated. The terms/phrases of each theme were ranked according to their frequency differences and the terms/phrases whose frequencies increased or decreased the most from one period to next were extracted from the rankings. Tables 24 to 34 display the top 20 terms/phrases of the rankings and only the terms/phrases whose frequencies increased or decreased more than 4 are included in the eleven tables. The subjects relevant to the terms/phrases were also included in these tables. The numbers in each table show the frequency differences. If a term's frequency decreased from Periods 1 to 2, its frequency difference would be negative and vice versa.

4.3.2.1. The Child Maltreatment Topic

(1) The Abuse, violence, harm, and subordination (AVHS) theme

Table 24 shows that although the frequencies of some terms/phrases (e.g. Catholic sexual abuse scandal and corporal punishment) about *abuse and violence* decreased in specific periods, the total frequency of the abuse-related terms/phrases increased rapidly during the investigated periods in the AVHS theme. Among all the lower-level subjects of *abuse and violence*, the total frequencies of the terms/phrases related to *sexual abuse* and *child abuse* increased in all the four periods. Some terms, whose frequencies increased, were related to

both two lower-level subjects, such as child sexual abuse, child pornography, and child prostitution.

Time Period	High-Frequency Terms and Phrases		Subjects
Period 1 VS. Period 2	Frequency Decreasing Terms	human trafficking (-135), South America (-12), Eastern Europe (-9), West Africa (-8), domestic violence (-7), Western Europe (-7), family violence (-6), male victims (-5)	Abuse and violence (domestic violence), human trafficking
	Frequency Increasing Terms	sexual abuse (126), sex ratio (119), child abuse (110), BBC News (98), Penn State (89), sex abuse (87), social undermining (70), United States (68), child sexual abuse (66), sex trafficking (64), New York (59), child sex (59), child pornography (37), sexual exploitation (32), Daily Telegraph (32), abuse cases (29), human rights (29), sex-selective abortion (26), law enforcement (25), trafficking victims (24), social support (24)	Abuse and violence (sexual abuse, child abuse), children and youth protection (news reporting), social factor
Period 2 VS. Period 3	Frequency Decreasing Terms	Catholic sexual abuse scandal (-9), social undermining (-7), trafficking victims (-6), Catholic sex abuse cases (-6), coercive persuasion (-6), forced labour (-6), New York Times (-5), Daily News (-5), high rates (-5), school corporal punishment (-5)	Abuse and violence (sexual abuse, emotional abuse, physical abuse), social factor, children and youth protection (news reporting)
	Frequency Increasing Terms	domestic violence (136), sexual abuse (131), child abuse (85), corporal punishment (79), human rights (71), violence against women (65), United Nations (62), sexual exploitation (50), New York (48), BBC News (42), childhood experiences (41), partner violence (35), child sexual abuse (34), family violence (33), adverse childhood experiences (32), World Health Organization (31), sex ratio (28), mental health (26), punishment of children (26), parental alienation (26)	Abuse and violence (domestic violence, sexual abuse, child abuse, physical abuse, woman abuse), children and youth protection (news reporting, organization), social factor (family issue)
Period 3 VS. Period 4	Frequency Decreasing Terms	corporal punishment (-18), Rochdale sex trafficking (-14), Taylor and Francis (-14), termination of pregnancy (-12), sex trafficking (-8), American Psychological Association (-7), abuse of children (-7), National Council on Family (-7), Council on Family Relations (-7), sex gang (-6), prisoners of war (-6), sex grooming (-6), human rights (-5), Global Initiative (-5), Daily Mail (-5), emotional abuse (-5), primary school (-5), intimate terrorism (-5), Child and Youth Protection (-5)	Abuse and violence (physical abuse, sexual abuse, emotional abuse), family planning and reproduction (method, organization), health issue (organization), children and youth protection (news reporting, organization), social factor (family issue)

	Frequency Increasing Terms	sexual abuse (156), child sexual abuse (62), Penn State (35), sexual exploitation (33), United States (32), sexual activity (26), Bryn Alyn (25), sexual assault (23), indecent assault (20), BBC News (19), child maltreatment (19), violence against children (18), South Yorkshire Police (17), adverse childhood experiences (15), abuse scandal (15), child prostitution (15), child neglect (15), localised grooming (15), human trafficking (14), sex-selective abortion (14)	Abuse and violence (sexual abuse, child abuse, emotional abuse), children and youth protection (news reporting, organization), human trafficking, family planning and reproduction (method)

Table 24. Changes of Subjects in the Four Periods in the AVHS Theme

Apart from the *abuse and violence* subject, the total frequency of *children and youth protection* rose from Periods 1 to 4. This subject had two lower-level subjects, which were *news reporting* and *organization*. The total frequency of the terms/phrases about *news reporting* grew during all the periods. BBC News appeared in all the three comparisons, which means that it played an important role in children and youth protection. These findings demonstrate that the Wikipedia editors paid more attentions to child abuse, sexual abuse, and child and youth protection from Periods 1 to 4.

(2) The Children, youth, families and friends (CYFF) theme

Table 25 shows that the total frequency of the terms/phrases about *social factor*, especially *family issue*, kept increasing from Periods 1 to 4 in the CYFF theme. Comparing of Periods 1 and 2, four of the frequency increasing terms were related to the attachment theory. As it was mentioned before, these were three entries about this theory created in the second period. The increases of the attachment-theory-related terms were caused by the generation of the corresponding entries.

Time Period	High-Frequency Terms and Phrases	Subjects

Period 1 VS. Period 2	Frequency Decreasing Terms	Oxford University (-11), antisocial behavior (-6), marriage and family (-6)	Health issue (problem), social factor (family issue)
	Frequency Increasing Terms	United States (52), family members (17), child support (16), Child Development (15), attachment figure (14), attachment security (13), extended family (11), domestic violence (8), nuclear family (6), men and women (6), attachment theory (5), family economics (5), individual differences (5)	Social factor (family issue), abuse and violence (domestic violence)
Period 2 VS. Period 3	Frequency Decreasing Terms	-	-
	Frequency Increasing Terms	men and women (17), United States (15), nuclear family (11), United Nations (11), domestic violence (10), family life (7), gender equality (7), gender roles (6), child abuse (5), nurture kinship (5), raising children (5), equal rights (5)	Social factor (family issue), abuse and violence (domestic violence, child abuse), inequality and discrimination
Period 3 VS. Period 4	Frequency Decreasing Terms	maternity leave (-8), men and women (-5), working mothers (-5)	Woman protection, inequality and discrimination (work)
	Frequency Increasing Terms	United States (10), marriage and family (8), father's rights (8), New York (5), father's rights movement (5)	Social factor (family issue), man protection

Table 25. Changes of Subjects in the Four Periods in the CYFF Theme

An interesting finding was that the frequencies of the terms about *inequality and discrimination* increased slightly from Periods 2 to 3 but decreased slightly from Periods 3 to 4. According to Table 25, the discussion about gender equality, gender roles, and equal rights rose from Periods 2 to 3. The discussion about maternity leave and working mothers reduced from Periods 3 to 4. To the contrary, the discussion about father's rights grew from Periods 3 to 4. These finding indicate that the Wikipedia editors' interests in inequality and discrimination increased first and then decrease, while their interests in father's rights rose in the last period.

(3) The Health problems and risks (CM-HPR) theme

The top frequency increasing/decreasing terms and phrases in the CM-HPR theme, and the associated subjects of these terms/phrases were displayed in Table 26. Considering both the frequency increasing and decreasing terms about *health issue*, the total frequency of terms related to this subject increased in the four periods, which implies that the Wikipedia editors' interests in health-related topics grew from 2010 to 2017. However, a particular finding illustrates that the term frequency of Disease Control and Prevention decreased rapidly from Periods 1 to 2 (the frequency difference was -46).

Time Period	High-Frequency Terms and Phrases		Subjects
Period 1 VS. Period 2	Frequency Decreasing Terms	Disease Control and Prevention (-46), traumatic stress (-15), sexual abuse (-12), heart disease (-11), population health (-10), social determinants of health (-9), complex trauma (-7), Sudden Infant Death Syndrome (-7), foster children (-6), risk factors (-6), physical abuse (-6), van Der Kolk (-6), conduct problems (-5), clinical psychology (-5), foster parent (-5)	Health issue (organization, problem, cause), abuse and violence (physical abuse), children and youth protection (law)
	Frequency Increasing Terms	people with BPD (77), infant mortality (68), emotion regulation (42), personality disorder (33), borderline personality disorder(31), mortality rate (21), foster care (19), blunted affect (15), substance abuse (14), birth weight (12), conduct disorder (11), bipolar disorder (10), domestic violence (9), oppositional defiant disorder(9), substance use order(9), emotional regulation (9), drug abuse (8), psychiatric association (8), diagnostic criteria (8), personality disorders (8), flat affect (8), New York(8)	Health issue (problem, service, organization, treatment), abuse and violence (domestic violence)
Period 2 VS. Period 3	Frequency Decreasing Terms	substance abuse (-9), oppositional defiant disorder (-6)	Health issue (problem)
	Frequency Increasing Terms	Washington DC (15), infant mortality (14), child molesters (12), conduct problems (11), sudden infant death (11), personality disorder (10), antisocial behavior (10), Sudden Infant Death Syndrome (10), conduct disorder (9), mortality rate (8), antisocial personality (8), sexual abuse (8), traumatic stress (8), women's health (7), assessment and treatment (7), chronic stress (7), determinants of health (7), borderline personality (6), Blanchard R (6), child pornography (6), sexual behavior (6)	Health issue (problem, cause, research), abuse and violence (child abuse, sexual abuse)

Period 3 VS. Period 4	Frequency Decreasing Terms	drug abuse (-6), Cantor JM (-6), van Der Kolk (-5), Blanchard R (-5)	Health issue (problem, research)
	Frequency Increasing Terms	women's health (50), United States (35), health care (21), United Nations (21), developing countries (18), World Health Organization (17), developed countries (16), substance use disorders (14), infant mortality (13), mental health (13), New York (12), borderline personality (12), public health (10), mortality rate (9), Child Health and Human Development (8), family members (8), substance abuse (8), health issues (7), disease control (7), personality disorder (6)	Health issue (service, organization, problem)

Table 26. Changes of Subjects in the Four Periods in the CM-HPR Theme

(4) The Support and protection (CM-SP) theme

Table 27 includes the terms/phrases whose frequencies increased/decreased the most from one period to next in the CM-SP theme. After examining all the frequency increasing/decreasing terms/phrases, it shows that the total frequencies of the *abuse and violence* subject, the *children and youth protection* subject, and the *social factor* subject all increased during the investigated periods. It reveals that the Wikipedia editors paid more and more attention to these three subjects from Periods 1 to 4.

Time Period	**High-Frequency Terms and Phrases**		**Subjects**
Period 1 VS. Period 2	Frequency Decreasing Terms	family members (-7), Jehovah's witnesses (-6), child wellbeing (-6), attachment parenting (-5), parenting style (-5)	Social factor (family issue), children and youth protection (organization)
	Frequency Increasing Terms	child abuse (50), sexual abuse (37), Amber Alert (27), child sexual abuse (20), cognitive development (17), child protection (16), Watch Tower Society (14), Child Abuse and Neglect (14), Department of Health (11), Haut de la Garenne (9), Catholic Church (8), United States (7), Supreme Court (7), Health and Human Services (7), marriage and family (7), family therapy (6), parental investment (6), child development (5), social work (5), United Kingdom (5), alert system (5)	Abuse and violence (child abuse, sexual abuse), children and youth protection (technology, organization, research), health issue (treatment), social factor (family issue)

Period 2 VS. Period 3	Frequency Decreasing Terms	family therapy (-18), attachment theory (-12), bodies of elders (-9), mandatory reporting (-9), substance abuse (-8), family members (-8), Journal of Family (-7), Juvenile Justice (-6), Cassidy J (-6), Developmental Psychology (-5), attachment theory research (-5)	Health issue (treatment, research), social factor (family issue), children and youth protection (law)
	Frequency Increasing Terms	sexual abuse (139), child sexual abuse (101), Commission into Institutional Responses (51), child abuse (40), Amber Alert (27), Royal Commission (27), corporal punishment (25), child protection (21), Children Act (20), attachment figure (15), Jehovah's witnesses (11), parenting styles (10), child's needs (10), Save the Children (10), child maltreatment (9), missing children (9), Human Development (8), Blehar M (8), human rights (8), parenting style (8), New South Wales (8)	Abuse and violence (sexual abuse, child abuse, physical abuse), children and youth protection (organization, law, technology), social factor (family issue)
Period 3 VS. Period 4	Frequency Decreasing Terms	Save the Children (-11), Force Report (-10), attachment style (-6)	Children and youth protection (organization), social factor (family issue)
	Frequency Increasing Terms	attachment parenting (143), sexual abuse (93), child sexual abuse (73), child abuse (58), Commission into Institutional Responses (53), New York (31), Royal Commission (27), parental investment (26), mandatory reporting (20), Child Abuse Prevention (17), child protection (16), Child Abuse and Neglect (16), United States (16), Catholic church (13), abuse inquiry (11), case study (11), foster care (10), child maltreatment (9), child care (9), sexual selection (9)	Social factor (family issue), abuse and violence (sexual abuse, child abuse), children and youth protection (organization, law)

Table 27. Changes of Subjects in the Four Periods in the CM-SP Theme

For the *children and youth protection* subject, many of the relevant frequency increasing terms/phrases of this subject were about the *organizations* and *laws* of children and youth protection, which implies that these organizations and laws played an increasingly important role in children and youth protection. Moreover, there were two phrases, "Amber Alert" and "alert system", referring to the technologies that helped protect children and youth. It reveals that the Wikipedia editors' interests in the children and youth protection organizations, laws, and technologies all increased from Periods 1 to 4.

4.3.2.2. The Family Planning Topic

(1) The Family planning and reproductive health (FPRH) theme

Table 28 contains the terms/phrases that increased/decreased the most in the FPRH theme during the investigated periods, and the associated subjects of these terms/phrases as well. Although some specific terms' frequencies reduced, the overall frequency increases of the terms about *health issue*, *family planning and reproduction*, and *population issue* revealed that the Wikipedia editors' interests in these subjects kept growing.

Time Period	High-Frequency Terms and Phrases		Subjects
Period 1 VS. Period 2	Frequency Decreasing Terms	Disease Control and Prevention (-46), the eugenics (-21), German Foundation for World (-14), World Population (-8), human heredity (-8), World War (-7), 20th century (-7), sperm donors (-6), Population Research (-6), birth control politics (-6), birth control movement (-6), teenage pregnancy (-5), Couple to Couple League (-5), University of California (-5), history of birth control (-5)	Health issue (organization), family planning and reproduction (research, method, organization), population issue (research), woman protection (activity)
	Frequency Increasing Terms	sex ratio (120), family planning (95), reproductive health (86), infant mortality (66), birth control (62), United States (59), Planned Parenthood (41), female condom (31), sex-selective abortion (28), reproductive coercion (27), domestic violence (25), health care (23), mortality rate (22), International Conference (19), United Nations (18), one-child policy (17), developing countries (17), health organization (16), prenatal care (16), ratio at birth (15)	Family planning and reproduction (law, policy, organization), health issue (problem, organization, service), abuse and violence (domestic violence), population issue (organization)
Period 2 VS. Period 3	Frequency Decreasing Terms	reproductive coercion (-17), paid leave (-17), domestic violence (-12), birth control sabotage (-6)	Family planning and reproduction (policy, method), abuse and violence (domestic violence)

	Frequency Increasing Terms	family planning (67), sex ratio (41), parental leave (40), United States (40), Planned Parenthood (31), New York (31), birth control (29), maternity leave (25), birth sex (22), United Nations Population Fund (19), human rights (18), infant mortality (17), sex selection (17), women's health (17), sex-selective abortion (16), reproductive rights (14), health organization (13), men and women (13), ovulation method (13), Billings Ovulation Method (11)	Family planning and reproduction (policy, organization), health issue (organization, problem), population issue (organization), woman protection
Period 3 VS. Period 4	Frequency Decreasing Terms	Catholic church (-15), termination of pregnancy (-10), ovulation method (-6), population control (-6), Roman Catholic church (-5)	Family planning and reproduction (method, religion), population issue
	Frequency Increasing Terms	family planning (82), Planned Parenthood (72), birth control (50), United States (42), reproductive health (37), maternity leave (32), one-child policy (31), health care (31), parental leave (30), sex education (29), human rights (23), United Nations Population Fund (20),Supreme Court (20), New York (16), mortality rate (16), infant mortality (14), sex-selective abortion (14), sex ratio (13), sexuality education (11), violence against women (11), Hong Kong (11)	Family planning and reproduction (organization, policy, method, education, law), health issue (organization, service), population issue (organization), abuse and violence, woman protection (law)

Table 28. Changes of Subjects in the Four Periods in the FPRH Theme

Among the lower-level subjects of *health issue* and *population issue*, the content about *health organizations* (e.g. the Planned Parenthood clinics) and *population organizations* (e.g. the United Nations Population Fund) were enhanced. The terms about *family planning and reproduction policies* and *laws* also grew from one period to next, especially the terms about parental leave, maternity leave, and abortion rights. There was only one lower-level subject whose related terms decreased from Periods 3 to 4 in the family planning and reproduction subject, which was *religion*. When examining the associated entries of this lower-level subject, it shows that some religious traditions regarded abortion as murder and allowed only natural methods of pregnancy avoidance. These findings demonstrate that the Wikipedia editors had

increasing interests in reproduction rights and benefits, while their interests in traditional reproduction methods decreased.

(2) The Human and environment (HE) theme

Table 29 shows that the subjects of the HE theme became more and more diverse during the investigated periods. During all the periods, the frequencies of the terms related to *social factor* kept increasing. Different from *social factor*, the content about other subjects in this theme did not keep increasing all the time but increased in specific periods. For instance, from Periods 2 to 3, the content about *human development* increased. From Periods 2 to 4, the content about *futures studies* kept growing. These results reveal that the Wikipedia editors' focuses of the HE theme became more diverse than before as time went by.

Time Period	High-Frequency Terms and Phrases		Subjects
Period 1 VS. Period 2	Frequency Decreasing Terms	parenting styles (-9), Gaia hypothesis (-8), social protection (-5)	Environment issue (research), social factor (family)
	Frequency Increasing Terms	a comedy series (28), planetary boundaries (22), forced marriage (21), United States (18), voluntary human extinction (16), New York (14), Jean Stapleton (13), financial crisis (12), global recession (11), Carroll O'Connor (10), United Nations (9), banking system (9), Rob Reiner (8), identity politics (7), maternity leave (7), Sally Struthers (7), New York Times (7), fertility rate (6), reserve army (6), middle class (5), conditional cash transfer (5)	Social factor (family issue), abuse and violence, population issue (problem), economy (policy), inequality and discrimination, family planning and reproduction (policy), military
Period 2 VS. Period 3	Frequency Decreasing Terms	All in the Family (-6)	Social factor (family)
	Frequency Increasing Terms	forced marriage (38), earth system science (26), human rights (17), United States (15), Gaia hypothesis (11), human development (10), existential risk (7), United Nations (6), reserve army (6), New York Times (6), United Kingdom (6), gender equality (6), Futures	Abuse and violence, environment issue (research), human development, military, futures studies, social factor (family issue)

		Research (5), catastrophic risks (5), British Columbia (5)	
Period 3 VS. Period 4	Frequency Decreasing Terms	work-life balance (-9), New York Times (-6)	Social factor (lifestyle)
	Frequency Increasing Terms	United States (18), futures studies (16), identity politics (13), fertility rate (12), tragedy of the commons (8), forced marriage (6), United Nations (5), European countries (5), Gaia hypothesis (5), Great Recession (5), New Brunswick (5), strategic foresight (5)	Futures studies, inequality and discrimination, family planning and reproduction, abuse and violence, social factor (family issue), environment issue (research), economy, population issue (problem)

Table 29. Changes of Subjects in the Four Periods in the HE Theme

(3) The Population problems (PP) theme

The high frequency increasing/decreasing terms/phrases and their associated subjects in different periods in the PP theme were demonstrated in Table 30. All the terms listed in this table are related to *population issue*, including *population problems*, *population policies*, and *censuses*. The result shows that the terms about *population problems* decreased from Periods 1 to 2 but increased from Periods 2 to 4. The terms about *population policies* increased from Periods 1 to 3, and then decreased from Periods 3 to 4. The *census* subject kept decreasing from Periods 2 to 4.

Time Period	**High-Frequency Terms and Phrases**		**Subjects**
Period 1 VS. Period 2	Frequency Decreasing Terms	list of countries (-6), Soviet Union (-5), population figures (-5)	Population issue (problem, policy)
	Frequency Increasing Terms	official population (24), sex ratio (27), birth sex ratio (18), population growth (14), world population (11), human sex (7), global population (7), population control (6), New York (6), human population (5), fertility rate (5)	Population issue (policy)

Period 2 VS. Period 3	Frequency Decreasing Terms	United States (-9), Census Bureau (-6), Population Clock (-5), 19th century (-5), 2008 census (-5), preliminary 2012 census (-5)	Population issue (census)
	Frequency Increasing Terms	population growth (24), human population (19), official population (14), population control (11), carrying capacity (8), world population (7), sovereign states (6)	Population issue (policy, problem)
Period 3 VS. Period 4	Frequency Decreasing Terms	population control (-28), 2010 census (-23), 2012 census (-21), 2011 census (-21), dependent territories (-7), census result (-6), youth bulge (-5), 2014 census (-5)	Population issue (policy, census)
	Frequency Increasing Terms	Country's population (22), population growth (19), world population (9), New York (9), list of countries (9), million people (9), United States (8), human overpopulation (7), Cambridge University (6), human population growth (6)	Population issue (problem)

Table 30. Changes of Subjects in the Four Periods in the PP Theme

4.3.2.3. The Women's Health Topic

(1) The Discrimination, violence, harm, and subordination (DVHS) theme

Table 31 presents the terms/phrases whose frequencies changed the most from one period to next in the DVHS theme. During all the periods, the terms about *abuse and violence*, *inequality and discrimination*, and *minority group* kept growing. One lower-level subject of *abuse and violence*, *sexual violence*, increased in all the periods, and another lower-level subject, *domestic violence*, increased from Periods 1 to 2, and Periods 3 to 4. The increase of the terms relevant to female genital mutilation mainly caused the growth of *domestic violence*.

Time Period	**High-Frequency Terms and Phrases**		**Subjects**
Period 1 VS. Period 2	Frequency Decreasing Terms	a history of women (-9), US Department (-5)	Inequality and discrimination (work), minority group (woman)

	Frequency Increasing Terms	female genital (92), female genital mutilation (73), female circumcision (40), hegemonic masculinity (39), human rights (38), New York (34), United States (29), gender apartheid (29), United Nations (27), gender identity (23), men and women (21), transgender people (20), age discrimination (18), rape culture (18), Glick P (16), World Health Organization (16), genital cutting (14), Oxford University (14), South Africa (14), gender roles (13), sexual assault (13), Islamic law (13)	Abuse and violence (domestic violence, sexual violence), inequality and discrimination (society, economy, work), minority group (LGBT), woman protection (organization), health issue (organization)
Period 2 VS. Period 3	Frequency Decreasing Terms	genital mutilation (-32), female circumcision (-30), female genital (-29), Islamic law (-13), South Carolina (-9), gender identity (-8), Type III (-7), Hosken Report (-6), Glick P (-5), Oxford University (-5), World Health Organization (-5), medical journal (-5), Agrarian System (-5)	Abuse and violence (domestic violence), inequality and discrimination (research), health issue (organization, research)
	Frequency Increasing Terms	New York (50), gender gap (39), sex ratio (31), first female (25), rape culture (23), gender inequality (22), age discrimination (19), missing women (19), New York Times (18), women and girls (18), United States (16), men and women (16), hegemonic masculinity (15), glass cliff (15), victim blaming (11), male privilege (11), sexual violence (10), gender equality (10), gender bias (10), transgender people (9), women and children (8)	Inequality and discrimination (healthcare, work), abuse and violence (sexual violence), minority group (LGBT, woman)
Period 3 VS. Period 4	Frequency Decreasing Terms	New Zealand (-5)	Inequality and discrimination (work)
	Frequency Increasing Terms	men and women (42), missing women (37), United States (30), gender gap (23), sexual assault (20), genital mutilation (20), transgender people (17), triple oppression (16), sexual violence (15), South Africa (13), women's rights (11), New York (9), gender bias (9), rape victims (9), global gender gap (9), gender roles (9), violence against women (9), labor force (9), sex differences (9), pay gap (9), gender gap report (9)	Inequality and discrimination (healthcare, work), abuse and violence (sexual violence, domestic violence), minority group (LGBT, woman), woman protection

Table 31. Changes of Subjects in the Four Periods in the DVHS Theme

The content about *inequality and discrimination* focused on different aspects in different time periods. For instance, the interests about inequality in *society*, *economy*, and *work* increased from Periods 1 to 2, while from Periods 2 to 4 the interests about inequality in

healthcare grew. For the *minority group* subject, the content about *LGBT* people increased in all the investigated periods. When examining the high-frequency terms/phrases about *LGBT*, it shows that "transgender people" occurred the most. In other words, from Periods 1 to 4, the Wikipedia editors paid increasing attention to the LGBT group, especially the transgender people.

(2) The Health problems and risks (WH-HPR) theme

The most highly increasing/decreasing terms/phrases and their associated subjects in the WH-HPR theme are listed in Table 32. This table shows that the terms about the *health issue* subject and the *family planning and reproduction* subject increased in all the periods. From Periods 1 to 2, the increasing terms about *family planning and reproduction* were related to *family planning and reproduction organizations* (e.g. United Nations Population Fund), while in the following periods, the terms were related to *family planning and reproduction methods* (e.g. induced abortion and medical abortion).

Time Period	High-Frequency Terms and Phrases		Subjects
Period 1 VS. Period 2	Frequency Decreasing Terms	Disease Control and Prevention (-43), medical dictionary (-5)	Health issue (organization)
	Frequency Increasing Terms	infant mortality (62), United States (40), mortality rate (38), health organization (36), health care (34), World Health Organization (29), public health (25), reproductive health (23), developing countries (18), infant mortality rate (17), Sub-Saharan Africa (14), family planning (12), urban development (10), men and women (10), ovarian cancer (9), New York (9), drinking water (9), determinants of health (9), heart disease (8), infant death (8), United Nations (8)	Health issue (problem, organization, service, cause), family planning and reproduction (organization)
Period 2 VS. Period 3	Frequency Decreasing Terms	Today's Evidence (-15), Tomorrow's Agenda (-15)	Health issue (organization),

			woman protection (organization)
	Frequency Increasing Terms	ovarian cancer (102), epithelial ovarian cancer (18), infant mortality (16), United States (13), gender polarization (12), risk of ovarian cancer (12), birth control (10), women's health (10), mental health (7), public health (6), live birth (6), health issues (6), side effects (6), health care (5), mortality rate (5), New York (5), infant mortality rate (5), infant death (5), induced abortion (5), million people (5), substance abuse (5), preterm birth (5), lymph node (5)	Health issue (problem, service), family planning and reproduction (method), inequality and discrimination (society)
Period 3 VS. Period 4	Frequency Decreasing Terms	embryo or fetus (-11), gynecologic oncology (-7), poor health (-6)	Health issue (research, problem), family planning and reproduction (method)
	Frequency Increasing Terms	mental health (67), health care (29), germ cell (26), blood pressure (21), United States (20), cell tumor (25), birth control (14), African Americans (12), socioeconomic status (12), infant mortality (11), mortality rate (8), health and human (8), health organization (7), burden of disease study (7), causes of death (7), ovarian cancer (11), infant mortality rate (6), systematic analysis (6), Stage I (6), health problems (5), health services (5), developed countries (5), medical abortion (5), systematic review (5), Cochrane Database (5)	Health issue (service, problem, organization, research, cause), family planning and reproduction (method)

Table 32. Changes of Subjects in the Four Periods in the WH-HPR Theme

The increasing terms of the health issue subject were related to different aspects in different periods. From Periods 1 to 2, the increasing terms covered various lower-level subjects, including *health problems*, *health organizations*, *health services*, and *causes of health problems*. Among these lower-level subjects, only *health problems* and *health services* attracted more attention than before in the next period. From Periods 3 to 4, a new lower-level subjects emerged, which was *health research*.

(3) The Medical and interdisciplinary subjects (MIS) theme

Table 33 shows that the frequency increasing terms covered more and more subjects as time went by in the MIS theme. From Periods 1 to 2, the terms were relevant to *health issue* and *family planning and reproduction*. It means that the Wikipedia editors' interests focused on these two subjects. In Period 3, a new interest about *population issue* emerged. In addition to the previous subjects, in Period 4, the Wikipedia editors had two more interests, *violence* and *inequality and discrimination*.

Time Period	High-Frequency Terms and Phrases		Subjects
Period 1 VS. Period 2	Frequency Decreasing Terms	population health (-12), Oxford University (-5), health status (-5), maternal deaths (-5), political economy (-5)	Population issue, health issue (problem), economy, politics
	Frequency Increasing Terms	public health (83), health care (72), sex segregation (70), world health (40), mental health (39), health organization (39), United States (37), men and women (28), social determinants (27), health outcomes (26), World Health Organization (26), history of medicine (25), rural areas (24), health services (22), determinants of health (21), global health (19), 19th century (19), family planning (18), maternal mortality (17), living conditions (17), sex differences (16)	Health issue (service, organization, cause, research, problem), family planning and reproduction
Period 2 VS. Period 3	Frequency Decreasing Terms	African American (-40), health disparities (-17), global health (-13), heart disease (-11), risk factors (-8), coronary heart disease (-8), Healthcare Research and Quality (-7), ethnic disparities (-6), Community Health (-5), physical activity (-5), racial and ethnic disparities (-5), racial differences (-5), universal health (-5)	Health issue (service, problem, organization, research), inequality and discrimination (healthcare)
	Frequency Increasing Terms	reproductive health (35), mental health (33), family planning (21), medical sociology (20), New York (19), health care (18), women's health (14), public health (13), sexually transmitted diseases (13), women's health (13), molecular pathology (13), United Nations (12), social science (12), reproductive age (12), health issues (11), live births (10), rural areas (9), determinants of health (9), mental illness (9), Social Science & Medicine (9), women of reproductive age (9)	Health issue (research, service, problem, cause), family planning and reproduction (organization), population issue (organization)

Period 3 VS. Period 4	Frequency Decreasing Terms	feminist theory (-10), women's studies (-8), Western medicine (-8), sexually transmitted diseases (-6), Chinese medicine (-6), Law Review (-5), myth of matriarchal prehistory (-5), Charlotte Perkins (-5), John Knox (-5), medical care (-5)	Woman protection (research, law), health issue (research, treatment, problem)
	Frequency Increasing Terms	mental health (72), women's health (54), United States (38), public health (34), mental illness (26), health care (25), social work (24), world health (24), female genital (24), family planning (23), United Nations (22), developing countries (22), reproductive health (20), sustainable development (20), violence against women (20), maternal mortality (16), developed countries (16), Millennium Development (16), health issues (15), social determinants (13), World Health Organization (13), health research (13), health disparities (13), intimate partner (13), cervical cancer (13)	Health issue (problem, organization, cause, research), abuse and violence (domestic violence), family planning and reproduction (organization), population issue (organization), inequality and discrimination (healthcare)

Table 33. Changes of Subjects in the Four Periods in the MIS Theme

(4) The Support and protection (WH-SP) theme

Table 34 displays the terms/phrases whose frequencies changed the most in each period in the WH-SP theme. This table illustrates that the terms about *health issue* and *woman protection* kept increasing from Periods 1 to 2, although in different periods these terms focused on different aspects of the two subjects. For instance, the terms about *treatment* only increased from Periods 1 to 2, while the terms about *health education* increased from Periods 1 to 3.

Time Period	**High-Frequency Terms and Phrases**		**Subjects**
Period 1 VS. Period 2	Frequency Decreasing Terms	health insurance (-42), Health Affairs (-13), care services (-12), women's college (-12), medical care (-10), health care services (-10), health care costs (-10), insurance coverage (-8), sex discrimination (-8), Medicaid Services (-8), Women's College Hospital (-8), Centers for Medicare (-7), Washington DC (-6), medical treatment (-6), National Organization for Women (-6), civil rights (-5), universal health (-5)	Health issue (insurance, research, service, organization, treatment), inequality and discrimination, woman protection (organization)

	Frequency Increasing Terms	women's health (163), health center (59), United States (50), reproductive health (50), women's suffrage (37), health initiative (28), public health (27), health centers (27), breast cancer (26), women's rights (25), men and women (22), Michigan Health System (22), Women's Health Initiative (20), hormone therapy (19), planned parenthood (18), health education (16), female condom (16), red dress (16), postmenopausal women (15), colorectal cancer (14), University of Michigan (14), red dress collection (14)	Health issue (organization, problem, treatment, education), inequality and discrimination (politics), woman protection, family planning and reproduction (method)
Period 2 VS. Period 3	Frequency Decreasing Terms	reproductive health (-8), ancient Rome (-5), University of Pittsburgh (-5), Journal of Obstetrics (-5), Stefanick ML (-5)	Family planning and reproduction (organization, research, treatment), health issue (organization, education)
	Frequency Increasing Terms	women's health (82), public health (40), health literacy (35), health education (33), human rights (33), New York (30), violence against women (30), gender equality (29), domestic violence (28), United Nations (27), health care (25), Department of Health (23), Oxford University (21), women's suffrage (20), breast cancer (20), right to vote (20), reproductive rights (20), social security (18), men and women (17), United States (15), health system (15), National Institute (15)	Health issue (education, service, organization), abuse and violence (domestic violence), inequality and discrimination, woman protection (politics, health)
Period 3 VS. Period 4	Frequency Decreasing Terms	hormone replacement (-20), health system (-16), hormone replacement therapy (-11), Michigan Health System (-10), red dress collection (-9), suffrage referendum (-5), equine estrogen (-5)	Health issue (treatment, organization), inequality and discrimination (politics)
	Frequency Increasing Terms	New York (84), United States (55), women's health (36), violence against women (35), United Nations (26), health care (25), reproductive health (25), gender equality (24), health organization (21), Medical Association (21), women's rights (20), women's suffrage (16), birth control (16), World Health Organization (16), sexual and reproductive health (15), Department of Health (14), health and human (14), United States Department (14), human rights (13), reproductive rights (13), family planning (13), women's education (13)	Health issue (service, organization), abuse and violence, women protection (politics, education), family planning and reproduction

Table 34. Changes of Subjects in the Four Periods in the WH-SP Theme

The woman protection subject had three lower-level subjects, which were *politics, health, and education*. The terms about *politics* increased from Periods 2 to 4, which indicates that the Wikipedia editors had increasing interests in this subject in the recent years. Furthermore, examination of the terms about *politics* demonstrates that the Wikipedia editors' interests increased the most in women's suffrage.

4.3.3. Changes of External Popularities

The external popularity of a topic/theme was defined as the numbers of the page edits and the numbers of the page views of its associated entries. To reveal the changes of the selected topics more in-depth, two hypotheses were proposed in Chapter 3 and this section presents the results of the hypothesis testing. The two hypotheses proposed in Chapter 3 are:

H01: There were no significant differences among the investigated time periods in terms of the number of page views of the entries relevant to each of the topics.

H02: There were no significant differences among the investigated time periods in terms of the number of page edits of the entries relevant to each of the topics.

According to *H01* and *H02*, the independent variable of the test was time period and the dependent variable was the number of the page views/edits of an entry relevant to a specific topic. Since three topics were selected in this study, six sub-hypotheses were generated:

H01a: There were no significant differences among the investigated time periods in terms of the number of page views of the entries relevant to Child Maltreatment.

H01b: There were no significant differences among the investigated time periods in terms of the number of page views of the entries relevant to Family Planning.

H01c: There were no significant differences among the investigated time periods in terms of the number of page views of the entries relevant to Women's Health.

H02a: There were no significant differences among the investigated time periods in terms of the number of page edits of the entries relevant to Child Maltreatment.

H02b: There were no significant differences among the investigated time periods in terms of the number of page edits of the entries relevant to Family Planning.

H02c: There were no significant differences among the investigated time periods in terms of the number of page edits of the entries relevant to Women's Health.

As it was mentioned in the Methodology chapter, the Friedman's Test was applied to test for the differences among the periods. Table 35 presents the results from *H01a* to *H01c* and *H02a* to H02c.

Hypotheses	Chi-square Value	P-value
H01a	$\chi2(3) = 110.660$	0.000
H01b	$\chi2(3) = 78.865$	0.000
H01c	$\chi2(3) = 73.384$	0.000
H02a	$\chi2(3) = 23.052$	0.000
H02b	$\chi2(3) = 26.672$	0.000
H02c	$\chi2(3) = 0.383$	0.944

Table 35. Hypothesis Testing Results of H01 and H02

The results show that *H01a*, *H01b*, *H01c*, *H02a*, and *H02b* were rejected, while *H02c* was not rejected. It means that: (1) there *were significant differences* among the four periods in

terms of the number of the page views of the entries relevant to each topic; (2) there *were significant differences* among the four periods in terms of the number of the page edits of the entries relevant to *Child Maltreatment*/*Family Planning*; (3) there *were no significant differences* among the four periods in terms of the number of the page edits of the entries relevant to *Women's Health*.

The Sign Test was used to explore the difference among every two periods. The comparisons intended to reveal the differences from one period to next in order to show the temporal changes of external popularities. Hence, only the adjacent periods were compared. Since the result of *H02c* was not significant, no follow-up test was conducted for this hypothesis. The results of the follow-up tests for *H01a*, *H01b*, *H01c*, *H02a*, and *H02b* were presented in Table 36. In this table, P1, P2, P3, and P4 stand for Period 1, Period 2, Period 3, and Period 4, respectively.

Topics	Measures	Values	P1 VS. P2	P2 VS. P3	P3 VS. P4
Child Maltreatment	No. of Edits	Z-value	-2.487	-3.076	-1.016
		P-value	0.013	0.002	0.310
	No. of Views	Z-value	-7.951	-4.843	-7.988
		P-value	0.000	0.000	0.000
Family Planning	No. of Edits	Z-value	-1.814	-3.344	0.000
		P-value	0.070	0.001	1.000
	No. of Views	Z-value	-6.929	-1.025	-7.640
		P-value	0.000	0.305	0.000
Women's Health	No. of Views	Z-value	-7.363	-1.331	-6.134
		P-value	0.000	0.183	0.000

Table 36. Pairwise Comparison Results of H01 and H02

4.3.3.1. Changes of Page Edits

According to Table 35, the number of the page edits of the entries in the *Women's Health* topic had no significant change during the investigated periods. To the contrary, the other two topics had significant changes in terms of the number of edits.

The results in Table 36 reveal that for *Child Maltreatment*, there were significant differences among Periods 1 and 2 (p-value = 0.013 < 0.05), and Periods 2 and 3 (p-value = 0.002 < 0.05) in terms of the associated entries' number of the page edits. However, there was no significant difference among Periods 3 and 4 in terms of the number of the page edits (p-value = 0.310 > 0.05). When examining the detailed results, there were 81 (much less than half) positive signs obtained from the comparison among Periods 1 and 2. In other words, there were 81 entries' numbers of the page edits of Period 2 less than Period 1. Therefore, the overall number of the page edits of Period 2 was less than that of Period 1. After examining the positive and negative signs obtained from the comparison among Periods 2 and 3, the results demonstrate that the number of the page edits of Period 3 was less than Period 2 (84 positive signs VS. 130 negative signs). These findings show that the number of the page edits of the entries in the *Child Maltreatment* topic dropped a lot from Periods 1 to 3, but had no significant change after Period 3.

For the *Family Planning* topic, there were no significant differences among Periods 1 and 2, or Periods 3 and 4 in terms of the associated entries' number of the page edits. However, there was a significant difference among Periods 2 and 3. The examination of the comparison between Periods 2 and 3 presents that the number of the page edits of Period 3 was less than Period 2 (48 positive signs VS. 88 negative signs). Therefore, the number of the

page edits of the entries in the *Family Planning* topic did not change a lot from Periods 1 to 2 or from Periods 3 to 4, but decreased from Periods 2 to 3.

4.3.3.2. Changes of Page Views

The results in Table 36 reveal that for *Child Maltreatment*, there were significant differences among Periods 1 and 2, Periods 2 and 3, and Periods 3 and 4 in terms of the number of the page views. Examining the positive and negative signs obtained from the comparisons illustrates that the number of the page views of Period 2 was larger than Period 1 (158 positive signs VS. 44 negative signs), the number of the page views of Period 2 was less than Period 3 (74 positive signs VS. 147 negative signs), and the number of the page views of Period 4 was less than that of Period 3 (58 positive signs VS. 183 negative signs). These findings show that the number of the page views of the entries associated to *Child Maltreatment* increased significantly from Periods 1 to 2, while decreased significantly from Periods 2 to 4.

Regarding the *Family Planning* topic, the results reveal that there were significant differences among Periods 1 and 2, and Periods 3 and 4. However, there were no significant differences among Periods 2 and 3. The results of the positive and negative signs obtained from the comparisons reveal that the number of the page views of Period 2 was larger than Period 1 (105 positive signs VS. 25 negative signs), and the number of the page views of Period 4 was less then Period 3 (26 positive signs VS. 119 negative signs). These results show that the number of the views of the associated entries in *Family Planning* increased significantly from Periods 1 to 2, kept stable from Periods 2 to 3, and then decreased significantly from Periods 3 to 4.

Similar to *Family Planning*, the results for *Women's Health* show that there were significant differences among Periods 1 and 2, and Periods 3 and 4, but no significant difference was found among Periods 2 and 3 in terms of the number of the page views. When investigating the detailed results obtained from the pairwise comparisons, it shows that the number of the page views of Period 2 was larger than Period 1 (129 positive signs VS. 34 negative signs) and the number of the page views of Period 4 was smaller than Period 3 (53 positive signs VS. 139 negative signs). Therefore, the number of the page views of the associated entries in *Women's Health* grew from Periods 1 to 2, remained stable from Periods 2 to 3, and dropped from Periods 3 to 4.

4.3.4. Research Question Two Results Summary

The internal characteristics of the three selected topics changed during the investigated periods. The changes were demonstrated from two perspectives: entry growths and changes of subjects. The results show that the number of the associated entries in each selected topic kept growing from 2010 to 2017. Among the three topics, the number of the associated entries of *Child Maltreatment* increased the most during the four periods, while the number of the associated entries of *Family Planning* increased less than the other two topics. The number of the associated entries in each defined theme also grew, but in some specific periods, no new entry was generated. For example, no new entry was created in the fourth period in the CYFF theme of *Child Maltreatment*.

For the *Child Maltreatment* topic, a part of the entries generated in the investigated periods mainly referred to specific abuse types, abuse behaviors, cases or scandals, and regions.

The entries created in early years mainly focused on child maltreatment in the United States, while more and more entries about other countries or regions were created in the next few years. These entries usually summarized the child maltreatment and child protection status in certain countries/regions. Some other entries presented a collection of related items about child maltreatment, such as the "List of child abuse cases featuring long-term detention" entry and the "List of songs about child abuse" entry.

Similar to the entries created in the *Child Maltreatment* topic, some of the entries created in the *Family Planning* topic referred to specific organizations, policies, laws, techniques, and regions. Some other entries, like the "List of people that have expressed views relating to overpopulation as a problem" entry and the "List of population concern organizations" entry, listed a series of related items. The rest entries contained the content about general topics, like the "Family planning policy" entry.

Regarding the *Women's Health* topic, similar findings were obtained. A part of the entries created during the investigated periods in this topic mainly concentrated on specific sexism types, regions, organizations, and research. Some other entries focused on more general topics, like gender polarization, women's health issues, and so on. The rest entries displayed a collection of associated items, such as the "List of women's studies journals" entry.

The changes of the subjects in each theme of a topic can also reflect the evolutions of the topic's internal characteristics. Table 37 summarizes the growing, diminishing, and fluctuating subjects of the three selected topics. The growing/diminishing subjects were the subjects whose associated terms and phrases kept increasing/decreasing during the

investigated periods. In other words, the growing/diminishing subjects attracted increasing/decreasing attention during the investigated periods. The fluctuating subjects were the subjects whose associated terms and phrases increased in some periods but decreased in the other periods.

Topic	Subjects	
Child Maltreatment	Growing subjects	Abuse and violence, children and youth protection, family planning and reproduction, health issue, man protection, social factor
	Diminishing subjects	Woman protection
	Fluctuating subjects	Human trafficking, inequality and discrimination
Family Planning	Growing subjects	Abuse and violence, economy, family planning and reproduction, futures studies, health issue, human development, inequality and discrimination, military, population issue, social factor
	Diminishing subjects	-
	Fluctuating subjects	Environment issue, woman protection
Women's Health	Growing subjects	Abuse and violence, family planning and reproduction, health issue, inequality and discrimination, minority group, woman protection
	Diminishing subjects	Economy, politics
	Fluctuating subjects	Population issue

Table 37. Growing, Diminishing, and Fluctuating Subjects

Table 37 shows that there were six subjects kept growing in the *Child Maltreatment* topic. Five of the six subjects grew from Periods 1 to 4 except the *man protection* subject. The *man protection* subject rose only from Periods 3 to 4. It implies that the Wikipedia editors' interests not only focused on women's rights and woman protection, but extended to other groups in recent years.

For the *Women's Health* topic, the growing subjects' associated terms/phrases kept increasing from Periods 1 to 4. A typical example was the *minority group* subject. This subject became more and more important from Periods 1 to 4. In other words, the Wikipedia editors paid increasing attention to the minority groups from 2010 to 2017.

Similar to the internal characteristics, the external popularities of the three selected topics changed during the investigated periods. The changes were explored from two aspects: number of the page edits and number of the page views. The results of hypothesis testing show that the numbers of the page edits of *Child Maltreatment* and *Family Planning* declined substantially from 2010 to 2017, while the number of the page edits of *Women's Health* had no significant changes during the four periods. The number of the page views of *Child Maltreatment* increased rapidly from 2010 to 2013, while decreased rapidly after that. The numbers of the page views of the other two topics increased quickly from 2010 to 2013, kept stable from 2013 to 2015, and declined quickly from 2015 to 2017, respectively.

4.4. Results of Research Question Three

Research question three aims to discover the commonalities and differences among the selected topics' evolution patterns. It explores the commonalities and differences among the evolution patterns from both the internal and external perspectives. The changes of the internal characteristics and external popularities of each topic were demonstrated in Section 4.1 and Section 4.3.

4.4.1. Differences and Commonalities among the External Popularity Evolution Patterns

4.4.1.1. External Popularity Differences among Topics

As it was mentioned before, the external popularity of a topic was represented by the total number of the page edits and the total number of the page views of its associated entries. To explore whether the number of the page edits/views was influenced by a topic, two hypotheses were proposed:

H03: There were no significant differences among the selected topics in terms of the number of the page edits of the associated entries.

H04: There were no significant differences among the selected topics in terms of the number of the page views of the associated entries.

The independent variable of *H03* and *H04* was topic and the independent variables were the number of the page edits and the number of the page views respectively. As it was mentioned in the Methodology chapter, the Kruskal-Wallis H Test was employed for *H03* and *H04*. Table 38 illustrates the hypothesis testing results of *H03* and *H04*.

Hypotheses	Chi-square Value	P-value
H03	$\chi2(2) = 39.988$	0.000
H04	$\chi2(2) = 38.329$	0.000

Table 38. Hypothesis Testing Results of H03 and H04

The results of the Kruskal-Wallis H Tests show that there were significant differences among the topics in terms of the number of the page edits (P-value = 0.000 < 0.05). It indicates that the Wikipedia editors did not pay same attention to the three topics. Similarly, there were significant differences among the topics in terms of the number of the page views (P-value = 0.000 < 0.05). Therefore, the Wikipedia viewers did not have same interests in the three topics.

To figure out the difference among every two topics, the post-hoc pairwise comparisons were conducted for *H03* and *H04*. For *H03*, the results of pairwise comparisons demonstrate that there were significant differences between *Family Planning* and the other two topics (P-value = 0.000 < 0.05), while there was no significant difference among *Child Maltreatment* and *Women's Health* (P-value = 0.088 > 0.05). Since *Child Maltreatment* (Mean Rank = 178.19) and *Women's Health* (Mean Rank = 103.30) had higher mean ranks than *Family Planning* (Mean Rank = 152.02), the numbers of the page edits of *Child Maltreatment* and *Women's Health* were significantly larger than *Family Planning*. It indicates that the Wikipedia editors had more interests in *Child Maltreatment* and *Women's Health* than *Family Planning*.

For *H04*, the pairwise comparisons demonstrate that there was a significant difference among *Child Maltreatment* and *Family Planning* (P-value = 0.000 < 0.05), *Child Maltreatment* and *Women's Health* (P-value = 0.022 < 0.05), and *Family Planning* and *Women's Health* (P-value = 0.001 < 0.05). *Child Maltreatment* had the highest mean rank (Mean Rank = 179.98) among the three topics, *Women's Health* had the second highest mean rank (Mean Rank = 147.75), and *Family Planning* had the lowest mean rank (Mean Rank = 105.77). These findings imply that the Wikipedia viewers had the most interests in *Child Maltreatment*, the second most interests in *Women's Health*, and the least interests in *Family Planning*.

4.4.1.2. Commonalities among the External Popularity Evolution Patterns

According to the results stated before, several commonalities of the external popularity patterns were found among the selected topics. The descriptive statistical results reveal that the trends of the page edits and the trends of the page views were not consistent for the

selected topics. This phenomenon implies that the Wikipedia editors and viewers were two groups with different interests.

Another commonality was that the three topics' trends of the page views were consistent in the figures in Section 4.1. The three topics' numbers of the page views all declined from 2010 to 2011, climbed from 2011 to 2013, and dropped from 2013 to 2017. Therefore, the Wikipedia viewers' interests on the three topics changed in the same pattern. This result indicates that the Wikipedia viewers' interests on the family-health-related topics changed in the same pattern.

4.4.1.3. Differences among the External Popularity Evolution Patterns

Differences were observed among the external popularity evolution patterns of the selected topics. The trends of the number of the page views from Periods 1 to 4 and the results of *H04* illustrate that *Child Maltreatment* attracted much more attention than the other two topics, *Women's Health* attracted the second most attention among the three topics, while *Family Planning* received the least attention. The inferential statistical tests confirmed these results.

The trends of the three topics in terms of the number of the page edits varied in the topics. Although the hypothesis testing results reveal that in general *Child Maltreatment*, *Women's Health*, and *Family Planning* reached the first, second, and third places according to their numbers of the page edits, the changes of the three topics did not follow the same pattern. Therefore, the Wikipedia editors' interests on the three topics changed in different patterns.

4.4.2. Differences and Commonalities among the Internal Characteristic Evolution Patterns

4.4.2.1. Commonalities among the Internal Characteristic Evolution Patterns

As it was mentioned before, the evolutions of the internal characteristics of a topic were reflected by the entry growths and the changes of the subjects. The examination of the entry growths of each topic in Section 4.3.1 shows a commonality among the three topics: the entries in all the three topics kept increasing from 2010 to 2017. It means that the Wikipedia editors' interests in these topics grew from 2010 to 2017.

Another commonality was discovered when investigating the emerged entries: the content of some emerged entries became more specialized. For instance, the "Attachment theory" entry, introducing the general knowledge of the attachment theory, was created in 2004. In 2012 two associated entries of this theory were generated, which were the "Attachment theory and psychology of religion" entry and the "Fathers as attachment figures" entry. These two entries focused on specific aspects of the attachment theory.

Apart from the specialized entries, some other emerged entries provided summaries of other entries, such as the "Family planning policy" entry and the "List of women's studies journals" entry. The "Family planning policy" entry offered a summary of the one-child policy and the two-child policy and the "List of women's studies journals" entry listed a collection of journals concerning women's studies.

The third was that the content of the emerged entries became more internationalized. The entries created in the early years mainly focused on the family-health-related content in the United States, but as time went by, more and more entries about other countries and

regions were generated. In the three selected topics, there were emerged entries about specific countries and regions rather than the United States, including New Zealand, Sierra Leone, Sri Lanka, India, Hong Kong, and so forth.

Commonalities were also found by examining the changes of the subjects in each topic. The fourth commonality was that there were common subjects emerged in the three topics, including *abuse and violence*, *family planning and reproduction*, *health issue*, *inequality and discrimination*, and *woman protection*. Among the five common subjects, the *abuse and violence* subject, the *family planning and reproduction* subject, and the *health issue* subject kept growing from Periods 1 to 4. In other words, consistent growths of the subjects were found across the three topics. It also implies that the Wikipedia editors had increasing interests in these subjects.

In addition to the growing subjects, there were diminishing subjects and fluctuating subjects in each topic. These findings reveal that different subjects changed in different patterns. Since this phenomenon was observed in all the three topics, it infers that the variation of subject changes was common for the family-health-related topics.

The last commonality was that new subjects emerged during the investigated periods and the subjects in each topic became more and more diverse over time. For instance, a lower-level subject, *health education*, occurred in Period 3, and the *man protection* subject emerged in Period 4. These findings show that the Wikipedia editors' interests extended to new areas of family health.

4.4.2.2. Differences among the Internal Characteristic Evolution Patterns

After comparing the entry growths of the three topics, it shows that although the numbers of the associated entries in the topics all kept growing, the number of entries of *Child Maltreatment* climbed much faster than the other two topics. In other words, the speed of the entry growth varied from one topic to another. Moreover, the trends of the entry growths were also different among the three topics. The entries of *Child Maltreatment* increased the fastest during 2010 to 2011, while the entries of *Family Planning* increased the fastest during 2011 to 2012. Different from the other two topics, the *Women's Health* topic's associated entries emerged the most in 2013 and 2015.

Regarding the different types of the subjects in each topic, more differences were discovered. The *woman protection* subject kept diminishing in the *Child Maltreatment* topic, but kept growing in the *Women's Health* topic. Meanwhile, this subject was a fluctuating subject in *Family Planning*. These findings reveal that one subject could have totally different developing trajectories in different topics. The different developing trends of the subjects finally formed the different internal characteristic and external popularity evolution patterns of the family-health-related topics.

The third difference was that each topic had their own unique subjects. For instance, *Child Maltreatment* had the *human trafficking* subject and the *man protection* subject; *Family Planning* involved *futures studies*, *human development*, *military*, and *environment issue*; *Women's Health* contained *minority group*, *and politics*. Each of these subjects belonged to only one topic. The Wikipedia editors and viewers paid different attention to these subjects according to the external popularities of the defined theme and their associated subjects. These subjects and their related entries were the potential causes of the differences among the

external popularities of the topics. For instance, the diminishment of the women protection subject could be a reason for the declines in the numbers of yearly page edits and views of *Child Maltreatment*.

4.4.3. Research Question Three Results Summary

The inferential statistical results of *H03* and *H04* illustrate that the *Child Maltreatment* topic and the *Women's Health* topic were more popular than the *Family Planning* topic among the Wikipedia editors. The *Child Maltreatment* topic, *Women's Health* topic, and *Family Planning* topic were the most, second most, and least popular among the Wikipedia viewers, respectively.

Regarding the external popularity evolution patterns, two commonalities were found among the three topics: (1) their page edits trends were not consistent with their page views trends; (2) their page views trends were similar to each other. Considering the internal characteristic evolution patterns, six commonalities were discovered for the topics: (1) the associated entries of the topics all kept increasing from 2010 to 2017; (2) a part of the new entries became more specialized, while the other entries offered summaries of certain topics for the Wikipedia users; (3) the content of the emerged entries became more internationalized; (4) the three topics contained common growing subjects; (5) each subject had its own growing pattern, which might differ from other subjects; (6) the subjects in each topic became more and more diverse.

Two differences among the external popularity evolution patterns of the topics were observed: (1) the popularities of the three topics were different among the Wikipedia viewers;

(2) the Wikipedia editors' interests in the topics changed follow different patterns. Moreover, three differences were found in terms of the internal characteristic evolution patterns of the topics: (1) the topics' entry growth trends were different from each other; (2) the same subjects had different developing trajectories in the different topics; (3) each topic had its own unique subjects and these subjects could cause the differences in the topics' external popularities.

4.5. Chapter Four Summary

This chapter demonstrates the results of the three research questions proposed in the previous chapters. The results show that *Child Maltreatment*, *Family Planning*, and *Women's Health* had 241, 150, and 207 associated entries, respectively. *Child Maltreatment* had four themes and five subjects, *Family Planning* had three themes and eleven subjects, and *Women's health* had four themes and eight subjects. *Child Maltreatment* was the most popular among the three topics in terms of the numbers of page edits and page views, while *Women's Health* and *Family Planning* ranked the second and third places.

The entries and subjects of each topic kept increasing from 2010 to 2017. The subjects were assigned to three groups in terms of their developing trajectories during the four defined periods, which were the growing subject group, the diminishing subject group, and the fluctuating subject group. However, the external popularities of the topics declined during the investigated periods.

Based on the evolution patterns obtained for each topic, the commonalities and differences among the three topics' evolution patterns were discovered. Six commonalities and three differences were found among the internal characteristic evolution patterns of the three

topics. Additionally, two commonalities and two differences were observed among the external popularity evolution patterns of the three topics.

5. DISCUSSION & IMPLICATIONS

5.1. Discussion

Family-health-related topics are widely discussed in academic research and people's daily life. Family health information is a prominent component of information posted online, especially on social media platforms. This study investigated the evolutions of the internal characteristics and external popularities of three family-health-related topics. The implications of this study covered three areas: the theoretical implication, the practical implication, and the methodological implication.

5.1.1. External Popularity Evolution Patterns

5.1.1.1. Trends of Online Health Information Generation, Seeking, and Use

The use of social media kept rising during the past decades. According to *Social Media Usage: 2005-2015* (Perrin, 2015), American adults who used social network sites rose from 7% to 65% during 2005 to 2015. From 2012 to 2018, Smith and Anderson's (A. Smith, Monica, & erson, 2018) report illustrated that American adults who used Facebook, Instagram, Pinterest, Snapchat, LinkedIn, Twitter, and WhatsApp also kept growing. As one of the most well-known social media platforms, Wikipedia attracted increasing users and received increasing accesses from 2010 to 2017. The registered Wikipedia users was 1.06 million in December 2009 and reached 2.51 million in December 2017 (Wikimedia, 2018). Moreover, the Wikipedia page views increased 1595.68 billion from 2010 to 2017 (Wikimedia, 2018).

To the contrary, the use of online health information did not achieve sustainable growth during the past decades. Tu (2011) reported that considering all the health information sources, the proportion of online health consumers kept increasing from 2001 (15.9%) to 2010 (32.6%). A survey (Fox & Jones, 2009) demonstrated that 39% of the online health consumers used social network sites for information seeking and sharing. In 2010, 42.5% of the online health information seekers retrieved the online information about other people's health experience (Fox, 2011). The proportion dropped to 36.11% in 2012 (Fox, 2014). The proportion of the online health consumers who sought the people having the same health concerns with themselves also decreased from 24.32% to 22.22% during 2011 to 2012 (Fox, 2011 & 2014).

The Social Life of Health Information (Fox & Jones, 2009) survey also showed that 5% of the online health consumers posted health-related comments on blogs and 8% of them participated in online discussion. In 2010, 8.45% of those people had posted health-related information on social network sites (Fox, 2011). The proportion reached 8.49% in 2012 (Fox, 2014). These studies demonstrate that the contributors of health-related content on social media increased during the past years. Meanwhile, the trends of online health information use and online health information generation were not consistent.

A similar result was obtained from this study, which is that the trends of the page edits (online health information generation) and the page views (online health information use) were not consistent. However, different from the results obtained from the previous research studies, the findings of this study reveal that the family-health-related topics' page edits and page views both declined during the investigated periods.

Figure 25 illustrates the trends of the total number of page edits and page views of the selected topics. This figure shows that the overall trends of the two measures declined from 2010 to 2017. The total number of the page edits decreased from 2010 to 2014, rose slightly from 2014 to 2015, but continued to decrease after 2015. The total number of the page edits declined from 2010 to 2011, climbed from 2011 to 2013, and dropped dramatically after 2013.

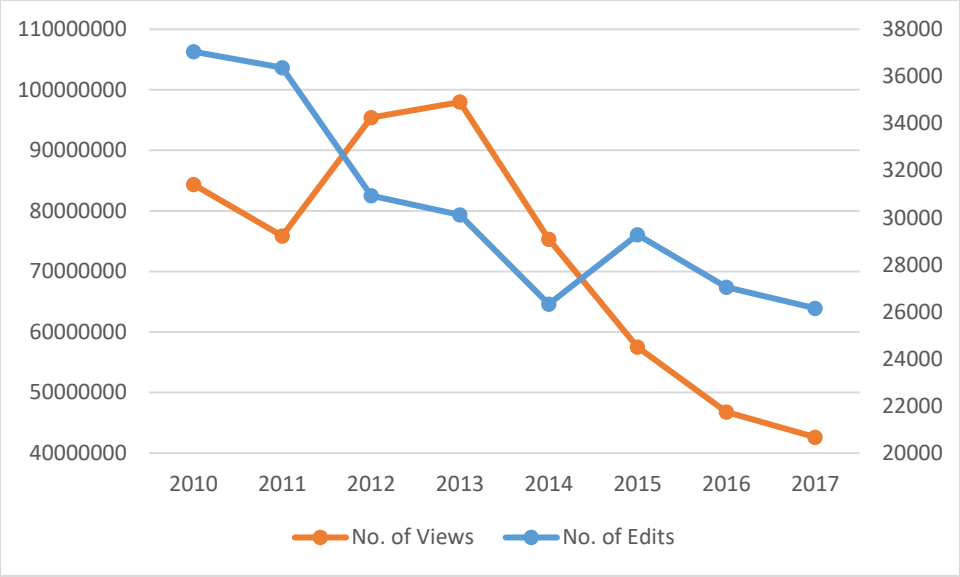


Figure 25. Trends of the Numbers of Page Edits and Page Views

The different findings demonstrate that although the overall trend of online health information generation rose, the trends of particular health topics (e.g. family-health-related topics) on particular social media platforms (e.g. Wikipedia) might be different from the general trend. Similarly, the online health information use trends of specific topics on specific platforms could differ from the overall online health information use trend.

5.1.1.2. Characteristics of Page Views

The number of the Wikipedia page views of an entry reflects the Wikipedia viewers' interests in the entry and its relevant concepts. According to the previous literature, it can reflect the Web search trend of the corresponding search term of the entry, which means that it can reflect the public's interests in the terms and the relevant concepts (Yoshida, Arase, Tsunoda, & Yamamoto, 2015).

Yoshida et al. (2015) conducted correlation tests to explore the association between Google search frequency and Wikipedia page views. They collected the Google Trends data and Wikipedia page views data of almost ten thousand personal names and utilized the Pearson product-moment correlation tests to investigate the associations. High correlations were found between the search frequency and the page views of the personal name keywords because the correlation coefficient was 0.72.

Same tests were conducted for three of the selected family-health-related entries, including the *Child abuse* entry, the *Family planning* entry, and the *Women's health* entry. The entries were also used as search terms to obtain Google search frequency data on Google Trends. Figure 26 displays the trends of the Google search frequency and the Wikipedia page views from 2010 to 2017 for *Child abuse*. The *X*-axis represents the month from January 2010 to December 2017. The left *Y*-axis stands for the Google search frequency and the right *Y*-axis stands for the ratio of the number of the Wikipedia page views of a month to the highest number of monthly page views of *Child abuse*. The Google search frequency of a search term was the proportion of its number of searches at a time point to the highest number of searches over the selected time period on Google (Nuti et al., 2014). The figure shows that the two curves fluctuate at a same frequency.

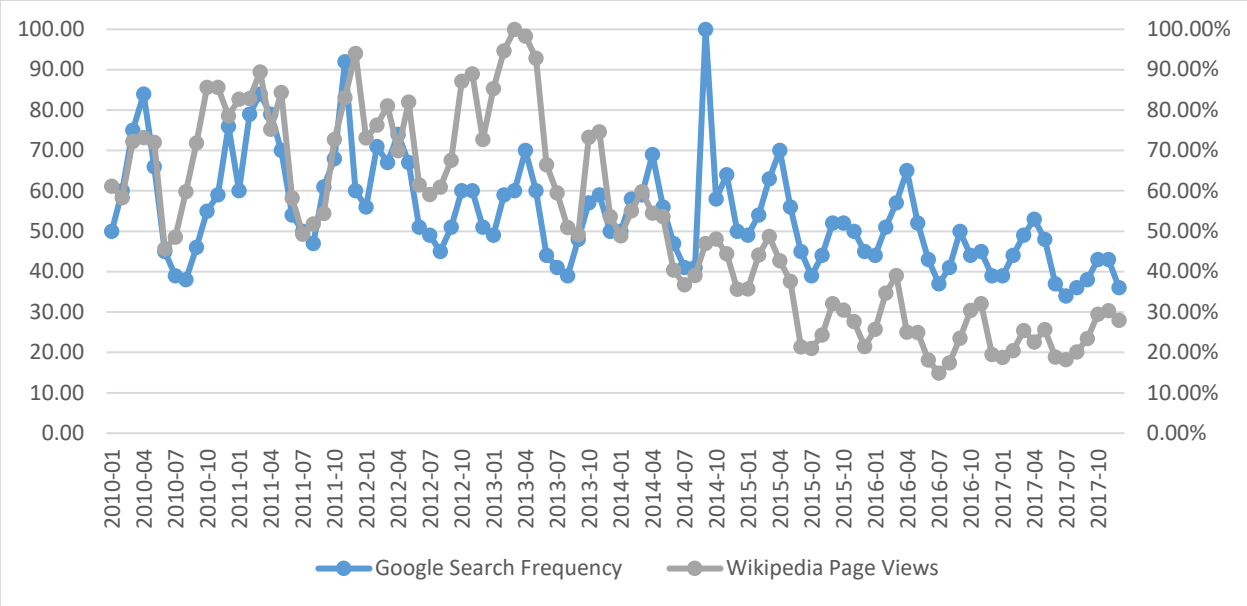


Figure 26. Trends of Google Search Frequency and Wikipedia Page Views for Child Abuse

The correlation coefficient between the search frequency and the number of the page views of *Child abuse* was 0.608 and the P-value was 0.000, smaller than the 0.05 significant level. For the *Family planning* entry, the correlation coefficient was 0.704 and the P-value was 0.000, smaller than the 0.05 significant level. However, the correlation coefficient of *Women's health* was 0.102 and the P-value was 0.323, larger than the 0.05 significant level. Therefore, there were significant associations between the search frequency and the number of the page views for *Child abuse* and *Family planning*, but no significant association was found for *Women's health*.

The results of *Child abuse* and *Family planning* confirm Yoshida et al.'s (2015) results, but the result of *Women's health* did not. The Google search frequency shows the general public's interests, but some Wikipedia entries' page view trends and Google search trends were inconsistent (Yoshida et al., 2015). It infers that the Wikipedia page views cannot always reflect

the public's interests. In other words, the Wikipedia viewers involved not only the public or lay people, but also the other user groups.

Thij, Volkovich, Laniado, and Kaltenbrunner (2012) analyzed the Wikipedia page view trend and proposed a model for popularity prediction of a promoted Wikipedia entry (an entry shown on the Main Page of Wikipedia as "today's features articles" and sent to subscribers). They found that a promoted entry received decreasing page views from the second hour to the eighth hour after it was promoted, then rose until the 18th hour, and the fell rapidly until the 24th hour. After the first day, the promoted entry's page views decayed exponentially with a constant rate. Figure 27 is an example cited from the Thij et al.'s work, which displays the real page view trend and the predicted page view trend of the *Augustus* entry.

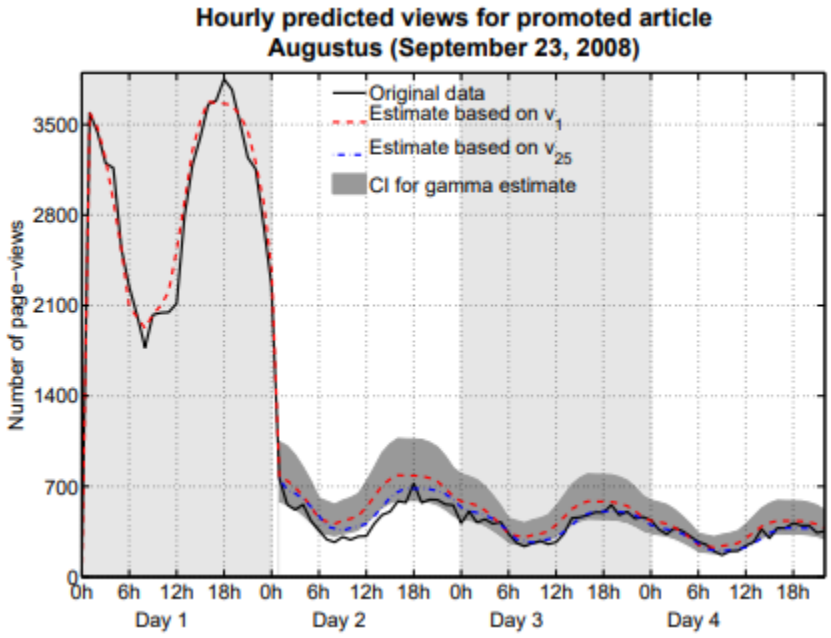


Figure 27. Example of the Prediction of the Page Views for a Promoted Entry (Thij et al., 2012)

Thij et al. (2012) explored the Wikipedia page view trend on a micro level, while this study illustrated the trend on a relatively macro level. Figure 10 in Section 4.1.3 demonstrates the trends of the selected topics in terms of their yearly number of the page views. Comparison of the two types of trends shows they had similarities to some extent. They both declined at first, then climbed to their highest points, and then dropped quickly. However, the fluctuation ranges of the two types of trends were different. Therefore, both similarities and differences were discovered among the two types of Wikipedia page view trends.

5.1.1.3. Characteristics of Page Edits

The Wikipedia page edit is another measure of the external popularity of an entry. Suh, Convertino, Chi, and Pirolli (2009) collected the monthly Wikipedia edits data generated by five user classes. A user was assigned to a class according to his/her contribution to Wikipedia entries. Figure 28, cited from the Suh et al.'s work, shows the monthly Wikipedia page edits (in thousands) generated by each user class. This figure shows that the page edits increased rapidly from 2005 to 2007, but decreased slightly after 2007.

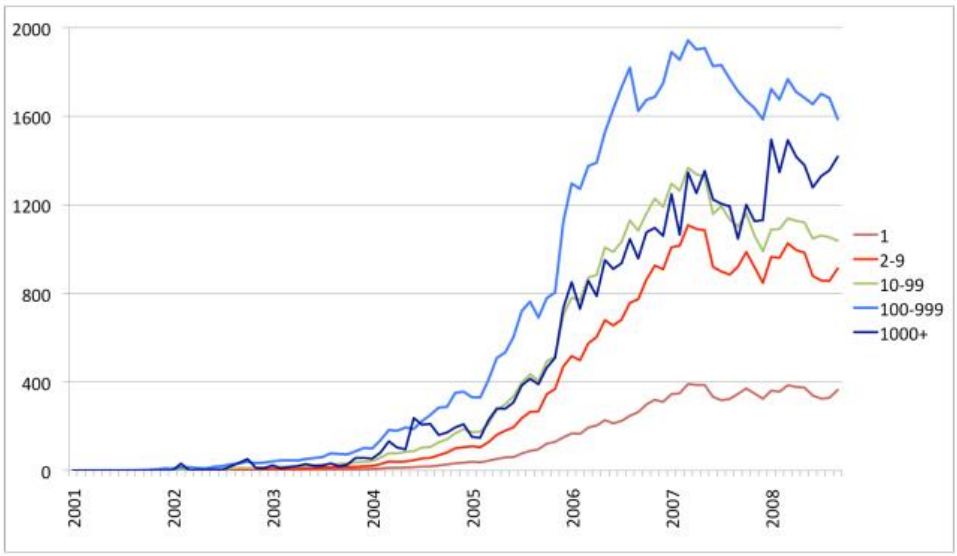


Figure 28. Monthly Edits by User Class (Suh et al., (Suh et al., 2009)

The trend of Wikipedia page edits obtained in this study was illustrated in Figure 29. The *X*-axis represents the year. The left *Y-axis* stands for the number of page edits of the whole English Wikipedia, while the right one stands for the total number of page edits of the selected topics. This figure confirms the results obtained by Suh et al., since the trend rose rapidly from 2005 to 2007 and then fell slow until 2014. It also shows the consistency between the page edits data of the English Wikipedia and the selected family-health-related topics. These results indicate that the family-health-related topics followed the general page edit trend, and moreover, potentially followed the page edit trends proposed by Suh et al. (2009).

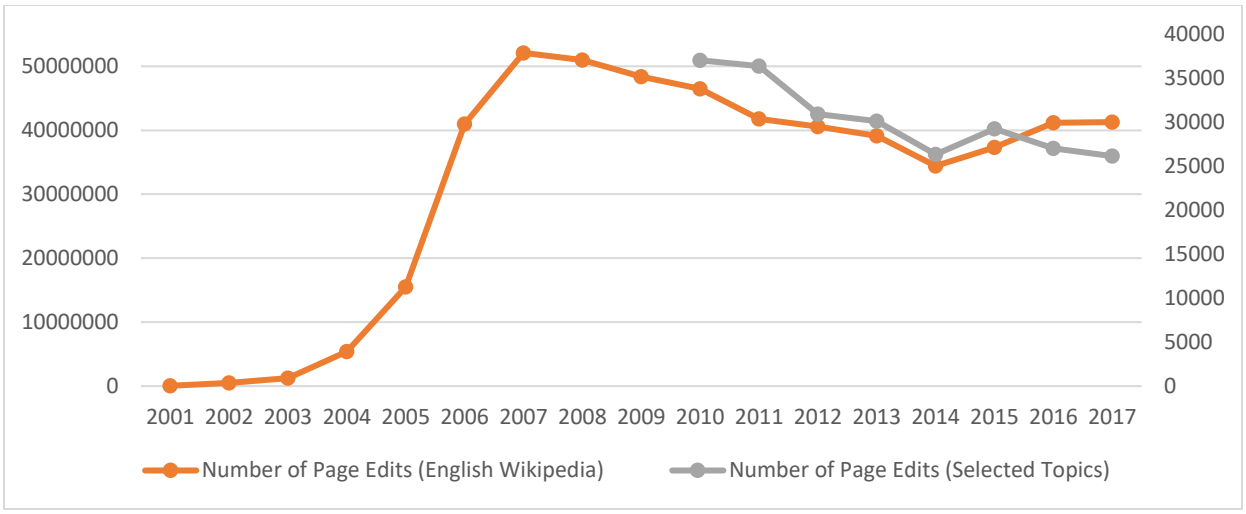


Figure 29. Numbers of Yearly Wikipedia Page Edits

5.1.1.4. Page Edits VS. Page Views

As it was mentioned before, the previous studies investigated the characteristics of the Wikipedia page edits and views and proposed models for them. However, few studies paid attention to the association between these two factors. This study collected the monthly page edits and views data of the selected topics and applied the simple linear regression tests to explore whether there was significant association between the two variables, the number of the page edits and the number of the page views.

A series of regression tests were conducted and the R-square values and P-values obtained from the tests are demonstrated in Figure 30. The monthly numbers of the page views (dependent variable) from January 2012 to December 2017 were collected and the monthly number of the page edits (independent variable) from 2012 to 2014 were collected. Considering the effect of time on the association, the monthly page views data were matched to the monthly page edits data generated 36-month earlier to the same month. The *X*-axis

represents the time difference between the independent variable and dependent variable. For instance, 8 means that the page views data used were 8 month later than the page edits data used in the test. The left *Y-axis* stands for the R-square value and the right *Y*-axis stands for the P-value of the regression tests.

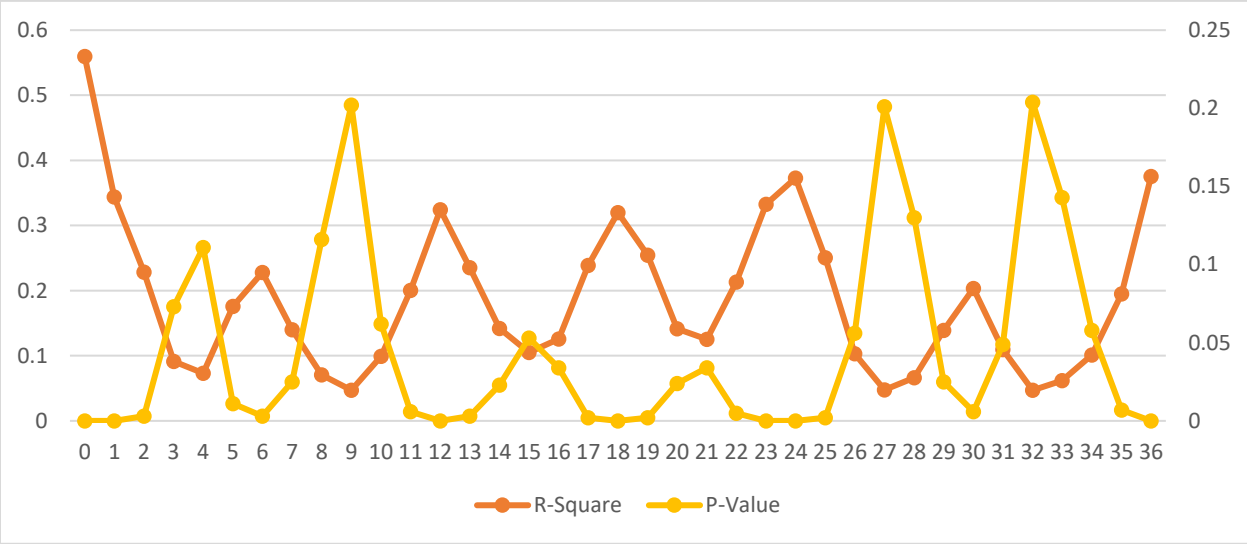


Figure 30. R-Square Values and P-Values of Regression Tests

The results show that although significant associations were found between the monthly page edits and the monthly page views when the page views data were later than the page edits data, the values of R-square were all smaller than 0.4, which means the influence of page edits on page views were not strong. The highest R-square value (0.56) was found between the two variables when the data were both collected from 2012 to 2014, but the R-square value was smaller than 0.7 which was considered as the borderline of strong effect size (Moore, Notz, & Flinger, 2013). Therefore, there was no strong association between the Wikipedia page edits and views. The Wikipedia users' editing behaviors and the viewing behaviors do not influence each other.

5.1.2. Internal Characteristic Evolution Patterns

5.1.2.1. Reasons for Entry Growth

The previous studies emphasized the importance of the editors' intrinsic and Extrinsic motivation and in the content generation on Wikipedia (Yang & Lai, 2010; Zhang & Zhu, 2006). Zhang and Zhu (2006) proposed that editing of an article decreased the creator's passion in further contribution. However, Yang and Lai's (2010) declaimed that intrinsic motivation had no significant effect on the contribution. They tested the effects of intrinsic motivation, extrinsic motivation, external self-concept motivation, and internal self-concept motivation in Wikipedia knowledge sharing behavior and found that only self-concept motivation had significantly positive effect on the knowledge sharing behavior. These studies concentrated on the reasons why content was generated on Wikipedia from the contributors' perspective, while this study explored the reasons of entry generation by examining the content of the emerged entries.

After investigating all the entries created in the four investigated periods and the characteristics of these entries, four reasons for generating new entries were discovered. The first reason was that there were certain "triggers" inducing the generation of new entries. The triggers included establishing organizations, events happening, proposing concepts and theories, publishing journals, enacting laws, promulgate policies, and so on. The creation time of the "triggered" entries were usually close to the time when the corresponding triggers appeared. Some of the "triggers", such as popular culture events, deaths, and breaks of diseases also affected the external popularities of the corresponding entries (Generous,

Fairchild, Deshpande, Valle, & Priedhorsky, 2016; McIver & Brownstein, 2014; West & Milowent, 2013).

The second reason was the spread of an existing concept. A good example was the creation of the "Non-consensual condom removal" entry. This concept emerged no later than 2014 but was not used or accepted by the mainstream. It was the news reporting of this concept that led to the creation of the corresponding entry on Wikipedia later.

The third reason was the need of a summary of related entries. There were hundreds of entries relevant to specific topics on Wikipedia, like the entries about organizations concerning the same topics, the entries about a specific region, the entries about a particular form of arts, and so forth. As the relevant entries increased, the need for a summary of these entries rose. For example, the "List of population concern organizations" entry listed the world-wide famous population organizations. The "Misogyny in horror films" entry summarized the types and characteristics of films which degraded women. These summaries provided rich information of a specific topic to the Wikipedia users.

The fourth reason was the engagement of the people who concerned about certain topics, such as the domain experts, the founders or members of organizations, and so on. For instance, the "The Honest Body Project" entry was generated by the creator of the project when the project began. Some entries on Wikipedia were not generated when the "triggers" appeared, but they were produced when the people concerned about these "triggers". For instance, there were three entries of the attachment theory created in the second period, but the time this theory and relevant studies proposed were much earlier than 2012. These three

entries were created by the people who were interested in this theory and had enough domain knowledge.

5.1.2.4. Featured Subjects

(1) The man protection subject

The *man protection* subject started to grow from Periods 3 to 4, especially the phrases "father's rights" and "father's rights movement" which increased fast during these periods. Examining the associated entries shows that the observed *man protection* subject mainly centered on father's parental and reproductive rights.

The advocate of fatherhood emerged in the 1990s since the fatherlessness became "the most critical social issue" (Baskerville, 2004, p.485). The researchers claimed that many serious social problems (e.g. violence, crime, unwed pregnancy, and so on) and health problems (e.g. mental disorders) were strongly correlated to fatherlessness. A primary cause of fatherlessness was the gender bias in family law. The researchers examined the previous cases and proposed suggestions for protecting father's rights to the government and the father's rights organizations (e.g. Fathers 4 Justice) were built all over the world (Baskerville, 2004). The research studies about father's rights also grew in the past decades.

Figure 31 presents the number of articles about father's rights published from 2000 to 2017 and the yearly number of page edits and views of the "Father's rights movement" entry on Wikipedia. The *X*-axis is the year, the left *Y-axis* is the number of articles and page edits, and the right *Y-axis* is the number of page views. The articles were retrieved from WOS with the search query "TS=(father's rights)" (TS means topic in the database).

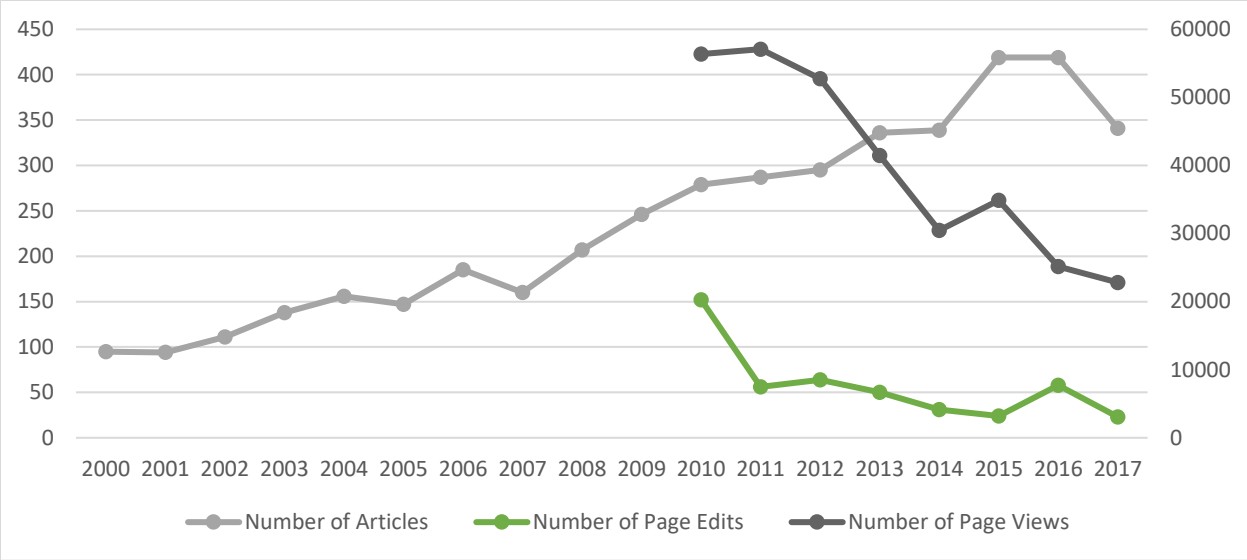


Figure 31. Trends of Father's Rights Related Articles and Wikipedia Page Edits and Views

The figure illustrates that the number the related articles increased rapidly from 2014 to 2015, kept stable till 2016, but decreased rapidly from 2016 to 2017. Moreover, no obvious growths were found for the other two factors. Therefore, these trends were not consistent with the changes of *man protection*.

A relevant entry of *man protection* was the "Father's quota" entry. Father's quota is the paternity leave for fathers. This concept began to expand in the 1990s as it was introduced by the Norwegian welfare state, the studies about the paternity leave for fathers dated back to the early 1990s, and the studies are still active in the recent years (Jensen & McKee, 2003; Klinth, 2008). The researchers studied how fathers used the paternity leave, whether the paternity leave affected fathers' work and income, the influences of fatherhood on children, and so forth (Cools, Fiva, & Kirkebøen, 2015; Jensen & McKee, 2003; Klinth, 2008). The associated Wikipedia entries focused on the history and policies of father's quota and paternity leave.

Figure 32 presents the number of articles about the paternity leave for fathers published from 2000 to 2017 and the yearly number of page edits and views of the "Father's quota" entry on Wikipedia. The *X*-axis is the year, the left *Y-axis* is the number of articles and page edits, and the right *Y-axis* is the number of page views. The articles were retrieved from WOS with the search query "TS=(paternity leave AND (men OR father))".

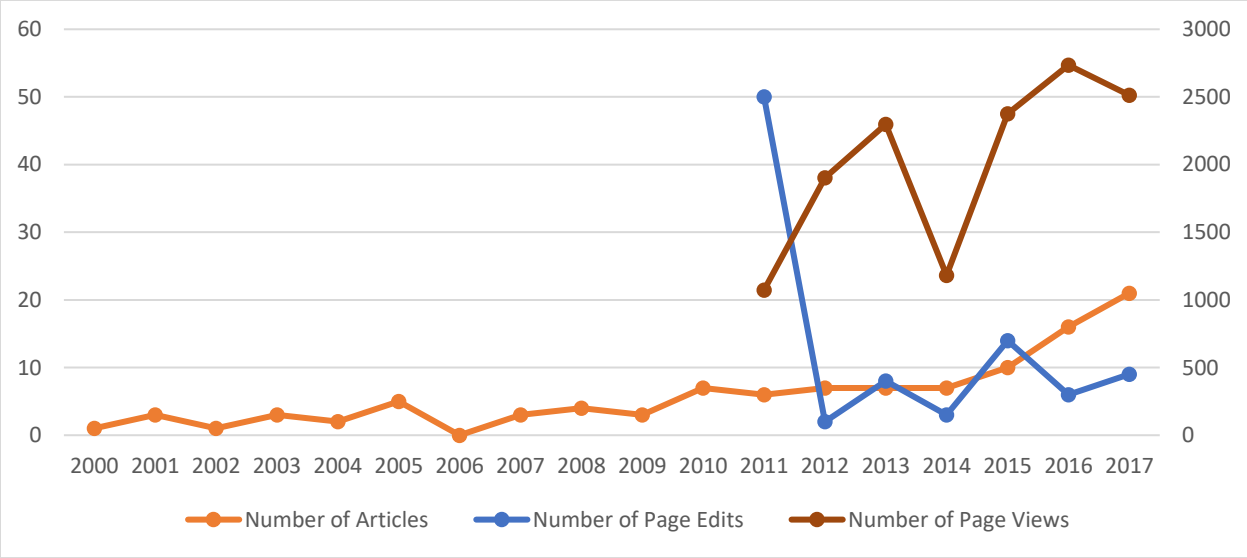


Figure 32. Trends of Father's Quota Related Articles and Wikipedia Page Edits and Views

The figure shows that the number of relevant articles rose rapidly from 2014 to 2017 (Periods 3 to 4), which confirmed the changes of the *man protection* subject. The number of page views had two substantial growths, 2011 to 2013 and 2014 to 2016, which did not match the subject changes. Meanwhile, the number of page edits climbed quickly from 2014 to 2015, but no notable increase was observed during 2015 to 2017.

(2) The minority group subject

The Wikipedia editors paid increasing attention to the minority groups from 2010 to 2017, particularly the LGBT people. The collected entries like "LGBT stereotypes" and "Healthcare and the LGBT community" centered on the LGBT people. The associated entries of the *abuse and violence* subject, the *family planning and reproduction* subject, and the *health issue* subject also gave their eyes on the LGBT group.

The research papers studied both family health and LGBT topics grew steadily growth from 2010 to 2016, but decreased in 2017 in WOS (retrieved with search query "TS=(LGBT AND (family health))"). The research subjects covered family planning and reproduction, health problems and diseases, healthcare, interpersonal relationships, social interactions, abuse and violence, and so forth (Croghan, Moone, & Olson, 2014; Klein et al., 2018; S. T. Russell, Ryan, Toomey, Diaz, & Sanchez, 2011; Shields et al., 2012; Snapp, Watson, Russell, Diaz, & Ryan, 2015; Willging, Salvador, & Kano, 2006). These subjects were also found among the selected Wikipedia entries. Therefore, a consistency was discovered between the research subjects and the user-generated subjects on Wikipedia.

The findings of the featured subjects imply that there was association between research studies and Wikipedia entries' content to some extent. However, not all the changes of subjects were reflected by the changes of relevant studies, which indicates that the Wikipedia editors were not only researchers or scholars.

(3) The diminishing subjects

There were three diminishing subjects discovered in this study, which were the *woman protection* subject in *Child Maltreatment*, and the *economy* subject and the *politics* subject in

Women's Health. The *women protection* subject in the *Children, youth, families and friends* theme of *Child Maltreatment* diminished from Period 3 to Period 4. It was caused by the decreasing discussion of maternity leave and working mothers.

The *economy* subject and the *politics* subject diminished from Period 1 to Period 2 in the *Medical and interdisciplinary subjects* theme of *Women's Health*. The diminishing was caused by the decrease of the phrase "political economy". In the early version of the *Global health* entry, political economy was included as a factor that influenced global health, while in the late versions, the content of political economy was removed. It means that the Wikipedia editors do not regard political economy as an important factor for global health in the recent years.

5.1.3. Internal Characteristics VS. External Popularities

Figure 33 displays the trends of the yearly numbers of the new generated entries, the page edits, and the page views for the selected topics from 2010 to 2017. The *X*-axis represents the year, the left *Y-axis* represents the number of the new entries every year, the right *Y*-axis represents the yearly number of page edits (thousands) and page views (billions). This figure demonstrates that the trends of the new entries and the page views were similar, while the trend of the page edits varied from the others. The probable reason was that the generation of new entries attracted the users to access and view the Wikipedia pages.

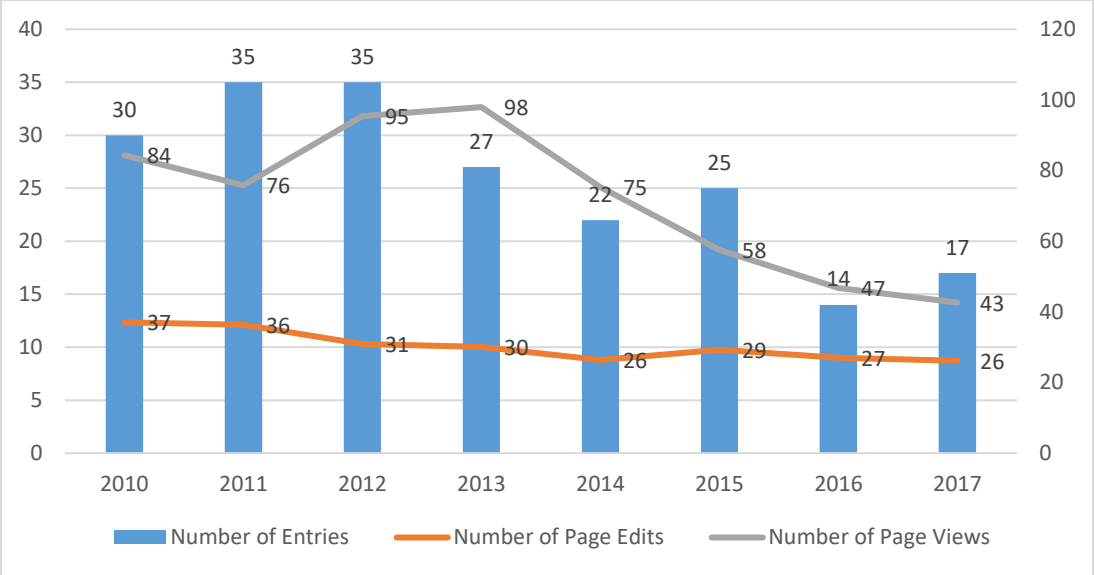


Figure 33. Trends of Entries, Page Edits, and Page Views for the Selected Topics

Regarding the content of the selected topics, the average length of the entries increased from 1945.83 terms to 2838.11 terms and the subjects of each topic became increasingly diverse from Periods 1 to 4. However, the decreasing page edits and views show totally different trends. In other words, the content and subject changes had no association with the external popularity changes.

5.1.4. Wikipedia VS. Other Platforms

As it was mentioned before, there were common subjects discovered from the Wikipedia entries and the academic research. Staub and Hodel (2015) also declared that widely-accepted research findings were cited on Wikipedia and accurate terminologies were provided for the viewers. To the contrary, the external popularity evolution patterns of the entries was not consistent with the increase of research papers. This finding confirms König's (2013) conclusion that both experts and lay people made contributions to Wikipedia entries.

König proposed that the experts completed most of the knowledge production, while the lay people contributed more in the content hierarchy organization.

The relation between Google search frequency and Wikipedia page views was investigated by Yoshida et al. (2015), but the results of this study rejected their findings. Not all the family-health-related topics had high correlated Google search frequencies and Wikipedia page views.

Health-related topics were also widely discussed on social media platforms other than Wikipedia, such as social network sites (e.g. Facebook), media sharing sites (e.g. YouTube), blogs (e.g. WordPress), microblogs (e.g. Twitter) and so on (Grajales III, Sheps, Ho, Novak-Lauscher, & Eysenbach, 2014). The social media platforms were utilized to surveil diseases and educate the lay public, and moreover, these platforms also attracted the health professionals to share their knowledge and built the bridge between them and the lay public (Hagg, Dahinten, & Currie, 2018). Wikipedia shared the common functionalities as the other social media platforms.

Wikipedia was one of the prominent resources of health information and knowledge (Heilman et al., 2011). Considering the results obtained, the entries growth was faster than the increase of research papers. The number of the page views for each Wikipedia entry was huge, which illustrates the popularity of the entry to the Wikipedia viewers, both the lay people and experts. Heilman et al. (2011) reported that 50% to 70% physicians used the Wikipedia as a health information source and Allahwala, Nadkarni, and Sebaratnam (2013) discovered that 94% medical students regarded Wikipedia as the most used information source. Meanwhile,

large amounts of the lay public read the health-related Wikipedia entries extensively (Heilman & West, 2015).

According to the previous research papers, Wikipedia provided a variety of health-related information and knowledge, including health care, medicine, illnesses and diseases, health research and education, and so forth (Azer, 2014; Generous et al., 2016; Heilman et al., 2011; McIver & Brownstein, 2014). Based on the previous findings, this study investigated the family-health-related topics in-depth and found the prevalent subjects (e.g. *abuse and violence*, *family planning and reproduction*, *health issue*, and so on) and lower-level subjects (e.g. *sexual abuse*, *domestic violence*, *family planning and reproduction method*, *LGBT group*, *health organization*, and so on) for the Wikipedia editors and viewers.

5.1.5. Discussion Summary

This section discussed the unique findings and compared them with the previous research papers. The previous findings showed that the trends of online health information generation and seeking kept growing during the past decades, but a decrease was found for the trend of the online health information use. Different from the previous findings, the external popularity evolution patterns observed in this study both fell during the investigated periods. The previous models of Wikipedia page views did not work well for the selected topics, while a consistency was found for the page edit trends between the previous literature and this study. This phenomenon confirms the results obtained in Chapter 4.

For the internal characteristic evolution patterns, the previous literature explored the factors influencing the users' contribution on social media from the users' perspective, while

this study investigated the reasons for entry growth by examining the content of the entries themselves. Four reasons were proposed, which were the "triggers", the spread of an existing concept, the need of summarization, and the engagement of the people who concerned about typical topics.

The external popularity evolution patterns were not associated with the internal characteristic evolution patterns. However, there was association between research studies and Wikipedia entries' content to some extent, since health professionals had made contributions to the content on Wikipedia, similar increases were discovered between research articles and Wikipedia subjects for specific entries, and common subjects were studied in both academic research papers and Wikipedia entries.

5.2. Implications

This study explored the associated entries, themes, and subjects of the Child Maltreatment topic, the Family Planning topic, and the Women's Health topic on Wikipedia, and examined the evolution patterns of the three topics, which included not only the evolutions of their internal characteristics but also external popularities. The implications of this study consist of three layers: the theoretical implications, the practical implications, and the methodological implications.

5.2.1. Theoretical Implications

5.2.1.1. Family Health Topics

Family-health-related topics are widely discussed in research studies and people's daily life. Family health information is a prominent component of information posted online, especially on social media platforms. This study demonstrated that the overall external popularities of the investigated family-health-related topics declined from 2010 to 2017, but their associated entries and terms increased and the associated subjects became more diverse. There were three types of subjects for the topics, which were growing subjects (e.g. *abuse and violence*, *family planning and reproduction*, and *health issue*), fluctuating subjects (e.g. *human trafficking* in *Child Maltreatment*, and *environment issue* in *Family Planning*), and diminishing subjects (e.g. *economy* in *Women's Health*). The findings can help health professionals, patients, and general users gain insights into the history and development of the topics from a general public's perspective and understand the related concepts, subjects, and themes better.

5.2.1.2. Wikipedia

The family-health-related topics' internal characteristic and external popularity evolution patterns were demonstrated based on the data obtained from Wikipedia. The evolution patterns of the selected topics can reflect the evolution of the Wikipedia content to some extent. For instance, the general trends of the page edits and views for the entries went down in the long run, while their content and related subjects became rich and diverse as time went by. The Wikipedia contributors and users can deepen their knowledge about the characteristics of Wikipedia and the user-generated content on it.

This study also compared Wikipedia with other platforms, including Google Trends, academic databases, and other social media platforms from both internal and external aspects.

The findings revealed the commonalities and differences among the different platforms, which enables the researchers and the general users to know more about the characteristics of these platforms.

5.2.1.2. Ontology Development

In the philosophical sense, ontology is defined as "a particular system of categories accounting for a certain vision of the world" (Guarino, 1998, p.82). In the area of artificial intelligence and computer science, ontology refers to "an engineering artifact, constituted by a specific vocabulary used to describe a certain reality, plus a set of explicit assumptions regarding the intended meaning of the vocabulary words." (Guarino, 1998, p.82). In the second definition, vocabulary words are concepts and relations. Therefore, to develop an ontology, it is necessary to include all the related concepts of a specific topic and identify the relations among them.

The results of Research Question One demonstrated the associated entries, themes, subjects, and high-frequency terms/phrases of the selected topics, and the relations among them as well. The entries, subjects, and terms obtained for the topics could be recognized as the vocabulary words of the family health ontology. The relations among the topics, themes, subjects, entries, and terms/phrases could be the references for developing the relations among the concepts of the ontology.

For example, the *Child Maltreatment* topic had 241 associated entries assigned to four themes and the majority of the entries were relevant concepts of *Child Maltreatment*, like *Child grooming*, *Abuse defense*, and so on. Moreover, a part of the high-frequency terms and phrases

of this topic were the relevant concepts, such as *immigrant families*, *neglected children*, and so forth. The entries and terms/phrases obtained in this study, as well as the relations among them, could be used for ontology development of family health and child maltreatment.

Additionally, since this study presented the subject changes of the topics, it is possible to develop temporal dynamic ontologies based on the findings. The emerging or diminishing concepts in each period could be extracted from the emerging entries and the frequency-increasing or frequency-decreasing terms and phrases.

5.2.1.3. Consumer Health Vocabularies

The consumer health vocabulary problem which describes the mismatch between the terms used by health professional and the ones used by consumers of health information has existed for a long time (C. A. Smith & Stavri, 2005; Zeng et al., 2007). Since general users usually have different educational background and do not have professional health training, it is impossible to require them to use professional terms. However, in general, consumer terms are not well defined, which causes problems in seeking and understanding health information (Zielstorff, 2003). On the other hand, the existing health vocabularies do not cover all the consumer terms. To fill the vocabulary gap between health professionals and consumers, health scientists have started to develop consumer health vocabularies (Patrick, Monga, Sievert, Hall, & Longo, 2001; Tse, 2011). For instance, Patrick et al. (Patrick, Monga, Sievert, Hall, & Longo, 2001b) used folk medical terms to extend three controlled vocabulary resources of technical medical terms and with the extended resources they linked consumer diabetes-related terms to their related terms used by family physicians.

Consumer health vocabulary development includes two steps: "(1) the identification of candidate strings (i.e., words or phrases) in a domain and (2) the determination of which of these should be included in a vocabulary as 'valid' terms" (Zeng et al., 2007, p.1). To identify the candidate strings, there are two criteria: the terms should be "commonly used by consumers" and the terms should allow for "computer processing of consumer language" (Zeng et al., 2007, p.2).

The family-health-related entries, subjects, terms, and phrases investigated in this study were collected from Wikipedia, created by the experts and lay people. Both professional terminologies and consumer health vocabularies were found from the content of the entries. For instance, the terms "hypertension" and "high blood pressure" were used interchangeably in the "Complications of pregnancy" entry. The health-related terms and phrases obtained from the selected entries were associated with family health, so they could be regarded as the candidates of the consumer health vocabularies. For example, the terms "child grooming" and "immigrant family" obtained from the entries could be added to the open-access and collaborative consumer health vocabulary. This procedure will contribute to the first step of consumer health vocabulary development.

5.2.2. Practical Implications

The theoretical implications mentioned before have relations to the practical implications in this section. The implications in ontology development and consumer health vocabulary are not only theoretical but also practical. Researchers can build ontology systems based on the ontologies obtained from research studies, and consumer health vocabulary can

be applied to optimizing information retrieval systems. Apart from these practical implications, the other implications were classified into two groups: the user-oriented implications and the system-oriented implications.

5.2.2.1. User-Oriented Practical Implications

The illustration of the selected topics' internal characteristics enables health professionals and general users to get a more comprehensive understanding of family health. The associated entries, subjects, and themes obtained in this study offered a whole picture of the three selected topics, demonstrated the relations among the relevant concepts, and can help different user groups distinguish relevant or similar concepts. The findings can improve the communications among different groups, such as patients, healthcare takers, and medical professionals. Moreover, in health information seeking, especially online information searching, they can use accurate search terms and related terms to present their information needs. Based on the changes of the subjects, and the frequency-increasing and frequency-decreasing terms/phrases discovered in the investigated periods, the users can modify their search strategies so as to retrieve more relevant information in specific periods.

5.2.2.2. System-Oriented Practical Implications

(1) Information organization

The system-oriented practical implications include an information organization aspect and an information retrieval aspect. The internal characteristics and external popularities obtained for the topics provide a way for information organization from the general public's view.

Wikipedia allows editors to classify entries into categories and sub-categories, and organize them in tree-like structures (Wikipedia, 2018). Navigation pages have been created to displays the categories, sub-categories, and entries. However, the relations among the categories, sub-categories, and entries are very complex. The categories and sub-categories are not exclusive. A sub-category might belong to multiple categories on different levels.

The *Health* category is one of the top-level categories on Wikipedia and has 45 sub-categories, including *Health by continent*, *Health care*, *Sexual health*, *Women's health*, and so on (Category:Health, 2018), while family health is not identified as a sub-category. This study collected and examined three family-health-related topics and their associated entries, which will supplement the entries in the existing sub-categories (e.g. the *Women's health* sub-category) and contribute to the generation of the *Family health* sub-category in the future. The methods used in this study could be applied to generate and enrich the other categories or sub-categories.

For Websites which aim to offer health information to the general public, their designers can consult the internal characteristics of the selected topics when organizing the family-health-related topics. The external popularities could be references for selecting the popular topics. Additionally, in order to assist users in exploring more related information, the Websites can add links of the information according to the related entries, subjects, and high-frequency terms and phrases detected in this study. For example, *abuse and violence* and *family planning and reproduction* were two popular subjects of *Family Health*, so these two subjects could be displayed in a prominent area on the family health page. The links of *domestic*

violence, *sexual abuse*, and other popular relevant lower-level subjects could be created and connected to the *abuse and violence* subject.

(2) Information retrieval

From the information retrieval aspect, the related entries, subjects, and terms/phrases obtained could be considered as related search terms by the recommendation system. An instance was that *women protection laws* and *women protection news report* could be related search terms of *women protection*. In query search, these items could be provided by information retrieval systems to help users modify their search queries.

Another potential implication is that observing the evolution patterns of the selected topics will support temporal information retrieval. A challenge in developing temporal information retrieval is the changes of terms and topics. New terms emerge and their meanings change as time goes by; the related concepts and terms of specific topics also change over time. As a result, in different ages the terms about the same topic vary a lot. This phenomenon causes the difficulty in retrieving all related documents in different time periods of a topic with same terms. This study demonstrated the internal characteristic changes of the selected topics in the determined periods. These findings provide the temporal relevant terms of the selected topics for information retrieval system, which will help information retrieval system judge the relevance between search queries and documents in a specific time period and return more relevant search results.

A typical example was the emergence of the “Non-consensual condom removal” entry and the term “stealthing”. Stealthing emerged in gay community in 2014 or earlier, while non-

consensual condom removal, a formal expression of stealthing, appeared later than stealthing and was widely used after 2016. To retrieve the results on “non-consensual condom removal” after 2016, the information retrieval system can return the items related to “non-consensual condom removal”. However, to retrieve the results before 2016 or 2014, the system can return the items about “stealthing” instead.

5.2.3. Methodological Implications

The third layer of implications lies in the research design. Since this study is a mixed research method study, different types of data were collected, and various methods and approaches were adopted for data collection and analysis. The uniqueness of this study included two aspects: data and methods applied.

5.2.3.1. Wikipedia Historical Data

Different data sets have been used in the Wikipedia studies, such as the text of Wikipedia entries, the users’ profiles, the editing records of entries, and the talk pages of entries. However, after reviewing plenty of research papers about Wikipedia, few of them collected and used the historical data of Wikipedia entries, like text of historical versions. Therefore, the data used in this study were unique.

The Wikipedia data dump was the data source of the page edits and page views data and the view historical pages on Wikipedia were the data source of the page creation time data and the historical revision data. The tools, r and RStudio, and the packages, WikipediR and pageviews, were utilized to collect and process these data. These were efficient and effective

means to acquire and process the Wikipedia data. The data from Google Trends and Web of Science were also collected and compared with the Wikipedia data in the Discussion section.

The findings reveal that the partition of the four time periods in this study was reasonable. The two-year time interval was long enough to show the subject changes for the selected topics. The experience of collecting and analyzing Wikipedia historical data would make a contribution to the future studies.

5.2.3.2. SOM Approach

The SOM approach is a popular unsupervised learning approach that projects high-dimensional data onto low-dimensional output (Kohonen, 1990). It is a widely used neural network method which can measure similarities between items of input data so as to form similarity graphs. The SOM approach has been applied to diverse research fields, such as finance, industry, biology, and so on (Deboeck & Kohonen, 2013; Ernst, Kellis, Hardison, Myers, & Wold, 2013; Fuertes et al., 2010; Sarlin, Yao, & Eklund, 2012). In the field of library and information science, this approach is usually employed in document cluster analysis and information retrieval algorithm, and to explore and extract information from documents (X. Lin, Soergel, & Marchionini, 1991; Suchanek et al., 2009; J. Zhang, 2007). However, few research studies employed this approach to analyze Wikipedia data.

This study applied the SOM approach to cluster Wikipedia entries, which was a new attempt in the information science field. Compared with other clustering approaches, the SOM approach has several advantages: (1) its outputs, the SOM displays, illustrate clear boundaries of clusters; (2) the colors on the SOM displays demonstrate the similarities between the

investigated items in addition to the distances between them; and (3) the SOM toolbox allows users to customize the data processing procedure of SOM based on the users' needs. According to the qualitative analysis results, the entries in one resultant cluster from the SOM analysis had high similarities, which means the SOM approach performed well for the Wikipedia data. Additionally, the SOM approach was combined with the coding method and the subject analysis method to reveal the internal characteristics of the selected topics.

5.2.3.3. Temporal Analysis Methods

Not only one quantitative analysis method (SOM) was utilized in this study, descriptive statistical analysis, inferential statistical analysis, and natural language processing methods were used as well. These methods were all applied for the temporal analysis.

Temporal analysis covers various data analysis methods that demonstrate the temporal changes of the research objects. This study used the bar charts and line charts to reveal the trends of the page edits and views, the Google search frequencies, and the research articles of the selected topics and entries. Four hypotheses and six sub-hypotheses were posted and tested to illustrate the differences between the four identified periods (Friedman's Test and Wilcoxon Signed-Rank Test) and the differences between the topics (Kruskal-Wallis H Test). The n-gram approach, a natural language processing method, was used to analyze the historical data of each period, and extract the terms/phrases changed the most from one period to next. The combination of temporal analysis methods in this study was unique compared with the previous research studies.

5.2.4. Implication Summary

The findings of this study could enable the experts and lay people understand family health and Wikipedia better, and benefit the development of ontologies and consumer health vocabularies. The practical implications have two aspects: the user-oriented and system-oriented aspects. The system-oriented implications will improve the information organization of online health information, and optimize the temporal information retrieval of health information. Regarding the methodological implications, the data collected and the data analysis methods used in this study will both contribute to future studies, particularly the methodologies.

5.3. Chapter Five Summary

This chapter includes two sections, Discussion and Implications. The first section further discussed the unique findings obtained in Chapter Four, and compared them with the previous literature. The second section demonstrated the theoretical implications, the practical implications, and the methodological implications of this study.

6. CONCLUSION

This chapter reviews the research problem of this study, and summarizes the primary findings obtained in the Results section and the Discussion & Implications section. The limitations of this study and the future directions are also demonstrated in this chapter.

6.1. Research Problem and Primary Findings

With the development of Web 2.0 and social media, the volume of information grows much more rapidly than the previous decades. New terms and concepts emerge and existing terms and concepts change much quicker on social media platforms than on conventional media. Family-health-related topics, terms, and concepts also change a lot during the past decades. This study investigated and discovered the evolutions of three family-health-related topics derived from the social media Website Wikipedia from 2010 to 2017. The entire time span was divided to four periods, which were 2010 to 2011 (Period 1), 2012 to 2013 (Period 2), 2014 to 2015 (Period 3), and 2016 to 2017 (Period 4).

Three family-health-related topics were selected from the WHO Website, which were *Child Maltreatment*, *Family Planning*, and *Women's Health*. The associated entries of the topics were retrieved from Wikipedia, and the numeric and text historical data of the entries were collected from Wikipedia data dump and the Wikipedia Web pages. Coding, subject analysis, descriptive statistical analysis, inferential statistical analysis, SOM approach, and n-gram approach were employed to explore the internal characteristic and external popularity evolutions of the selected topics. The primary findings of this study are demonstrated in the following paragraphs.

6.1.1. Findings of Child Maltreatment

Child Maltreatment had 241 associated entries on Wikipedia and the entries referred to four themes: (1) *Abuse, violence, harm, and subordination*; (2) *Children, youth, families and friends*; (3) *Health problems and risks*; and (4) *Support and protection*. Six subjects attracted increasing attention from the Wikipedia editors, which were *abuse and violence*, *children and youth protection*, *family planning and reproduction*, *health issue*, *man protection*, and *social factor*. However, the editors' interests in *Woman protection* diminished during the investigated periods. This topic was the most popular among the three topics from both the Wikipedia editors' and viewers' perspectives. Its popularity among the Wikipedia editors decreased significantly from Periods 1 to 3, while stayed stable after Period 3.

6.1.2. Findings of Family Planning

One hundred and fifty associated entries of *Family Planning* were found on Wikipedia and were assigned to three themes, *Family planning and reproductive health*, *Human and environment*, and *Population problems*. This topic had more diverse subjects than the other two topics. Among the subjects, *abuse and violence*, *economy*, *family planning and reproduction*, *futures studies*, *health issue*, *human development*, *inequality and discrimination*, *military*, *population issue*, and *social factor* received increasing attention from the editors. This topic was less popular than the other two topics among the Wikipedia editors and viewers. The popularity among the Wikipedia editors dropped rapidly from Periods 2 to 3, but kept stable in other periods.

6.1.3. Findings of Women's Health

Two hundred and seven associated entries of *Women's Health* were retrieved on Wikipedia and four themes emerged from these entries, which were (2) *Discrimination, violence, harm, and subordination*; (2) *Health problems and risks*; (3) *Medical and interdisciplinary subjects*; and (4) *Support and protection*. For this topic, the editors paid more and more attention to *abuse and violence*, *family planning and reproduction*, *health issue*, *inequality and discrimination*, *minority group*, and *woman protection*, while their interests in *economy* and *politics* decreased. Different from the other two topics, no significant change was found in terms of this topic's popularity among the editors.

6.1.4. Internal Characteristic and External Popularity Evolution Patterns

No association was found between the internal characteristic evolution patterns and the external popularity evolution patterns. From the internal characteristic's aspect, the family-health-related topics' content (e.g. entries, subjects, terms, and phrases) on Wikipedia kept increasing from 2010 to 2017. However, the topics' entry growth trends were different from each other. Three features were discovered for the emerged associated entries, which were specialization, summarization, and internationalization. The reasons for the entry growth included: (1) the "triggers"; (2) the spread of an existing concept; (3) the need of summarization; and (4) the engagement of the people who concerned about typical topics.

For some of the entries, the increases of their relevant articles were consistent with the growth of the relevant subjects. Meanwhile, similar family health subjects were found from the Wikipedia entries and research papers. Therefore, academic research had an association with the Wikipedia content to some extent.

The subjects in each topic became increasingly diverse as time went by. The common subjects (e.g. *abuse and violence*, *family planning and reproduction*, and *health issue*) of the topics had different developing trajectories in each topic.

From the external popularity's aspect, the overall popularity of the family-health-related topics declined from 2010 to 2017, contrary to the growth of their content and the growth of extensive online health information seeking.

The selected topics had similar trends of popularities among the Wikipedia viewers. Their popularity all grew rapidly from Periods 1 to 2, remained stable from Periods 2 to 3, and fell dramatically from Periods 3 to 4. However, the trends of the popularities among the viewers were not consistent with those among the editors. Therefore, the members in these two groups did not totally overlap each other. Moreover, the trends of the topics' popularities among the editors were not consistent with each other. It means that the editors of different family health topics varied from one to another.

6.2. Limitations

The first limitation is that due to the limitation of time, only the data of three family-health-related topics from 2010 to 2017 were collected and analyzed. Since the number of selected topics was relatively small and the time span was relatively short, it is possible that the results obtained from the three topics cannot be generalized to the whole health science field.

The *Gender bias on Wikipedia* entry on Wikipedia claimed that the majority of Wikipedia editors were male, which might lead to a potential bias in the Wikipedia content. Since male editors were much more than female editors, the user-generated content on Wikipedia more

reflected the male's opinions than the female's opinions, which might skew the findings obtained in this study.

Another limitation is that only one clustering approach and one data mining approach were employed in this study. In data analysis, different clustering and data mining approaches may lead to different results. However, it is hard to find and choose the most suitable approaches to get the best results. The researcher could only apply a limited number of approaches and acquire the most reasonable and reliable results by using these approaches.

6.3. Future Directions

To overcome the limitations, the data of more family-health-related topics from Wikipedia or other platforms (e.g. Twitter, YouTube, academic databases, and digital libraries) should be gathered in future studies. Examination of the evolution patterns of other family-health-related topics will show whether the patterns obtained in this study would be confirmed in other topics. The time duration of the investigation will be extended as well. In this case, the data obtained will be more complete and the research scope will be larger.

The data of the health-related topics collected could be compared with those obtained from the other platforms, such as WebMD, PubMed, and academic data sources. Since the content from these health Websites and academic data sources are mostly created by health professionals, comparing them with the data collected from this study can reveal the differences between the public and the health professionals' interests.

Another future research direction is to employ other data collection and analysis methods to achieve, present, and interpret results from different angles. As it was mentioned

before, there are various methods and tools for data collection and analysis. Although this study used r and RStudio for data collection, other tools (e.g. Python) are also useful for online data crawling. Additionally, apart from SOM and n-gram, there are many other clustering approaches and text mining approaches, such as Hierarchical Clustering methods, Partitioning Clustering methods, topic modeling methods, and so on. Using different data analysis methods will lead to different results and the corresponding interpretation will be different. Therefore, the researcher could compare the different results.

REFERENCES

Agarwal, N., & Yiliyasi, Y. (2010, November). Information quality challenges in social media. In *International Conference on Information Quality (ICIQ)* (pp. 234-248).

Aggarwal, N., Kumar, A., Khatter, H., & Aggarwal, V. (2012). Analysis the effect of data mining techniques on database. *Advances in Engineering Software*, *47*(1), 164–169. https://doi.org/10.1016/j.advengsoft.2011.12.013

Ahlqvist, T., Bäck, A., Halonen, M., & Heinonen, S. (2008). Social media road maps exploring the futures triggered by social media. *VTT Tiedotteita-Valtion Teknillinen Tutkimuskeskus*, *2454*, 13.

Ahmed, S. R. (2004). Applications of data mining in retail business. In *International Conference on Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004.* (Vol. 2, pp. 455-459 Vol.2). https://doi.org/10.1109/ITCC.2004.1286695

Alias-i (2008) Lingpipe 4.1.0. Retrieved from http://alias-i.com/lingpipe.

Aliev, R. A., Aliev, R. R., Guirimov, B., & Uyar, K. (2008). Dynamic data mining technique for rules extraction in a process of battery charging. *Applied Soft Computing*, *8*(3), 1252–1258. https://doi.org/10.1016/j.asoc.2007.02.015

Allahwala, U. K., Nadkarni, A., & Sebaratnam, D. F. (2013). Wikipedia use amongst medical students - new insights into the digital revolution. *Medical Teacher*, *35*(4), 337. https://doi.org/10.3109/0142159X.2012.737064

Alonso, O., Gertz, M., & Baeza-Yates, R. (2007). On the Value of Temporal Information in Information Retrieval. *SIGIR Forum*, *41*(2), 35–41. https://doi.org/10.1145/1328964.1328968

Arikan, I., Bedathur, S., & Berberich, K. (2009). Time will tell: Leveraging temporal expressions in ir. In *In WSDM*.

Azer, S. A. (2014). Evaluation of gastroenterology and hepatology articles on Wikipedia: Are they suitable as learning resources for medical students? *European Journal of Gastroenterology & Hepatology*, *26*(2), 155. https://doi.org/10.1097/MEG.0000000000000003

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. *ICWSM*, *8*, 361-362.

Bansal, N., & Koudas, N. (2007). BlogScope: A System for Online Analysis of High Volume Text Streams. In *Proceedings of the 33rd International Conference on Very Large Data Bases*

(pp. 1410–1413). Vienna, Austria: VLDB Endowment. Retrieved from http://dl.acm.org/citation.cfm?id=1325851.1326028

Baskerville, S. (2004). Is There Really a Fatherhood Crisis? *The Independent Review*, *8*(4), 485–508.

Bazeley, P., & Jackson, K. (Eds.). (2013). *Qualitative data analysis with NVivo*. Sage Publications Limited.

Bem, S. L. (1995). Dismantling gender polarization and compulsory heterosexuality: should we turn the volume down or up?. *Journal of Sex Research, 32*(4), 329-334.

Benkler, Y. (2006). *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press.

Benkler, Y., & Nissenbaum, H. (2006). Commons-based Peer Production and Virtue*. *Journal of Political Philosophy*, *14*(4), 394–419. https://doi.org/10.1111/j.1467-9760.2006.00235.x

Berson, A., & Smith, S. J. (2002). *Building Data Mining Applications for CRM*. New York, NY, USA: McGraw-Hill, Inc.

Black, K., & Lobo, M. (2008). A Conceptual Review of Family Resilience Factors. *Journal of Family Nursing*, *14*(1), 33–55. https://doi.org/10.1177/1074840707312237

Blitzer, J., Dredze, M., & Pereira, F. (2007, June). Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In*ACL* (Vol. 7, pp. 440-447).

Bolton, R. N., Parasuraman, A., Hoefnagels, A., Migchels, N., Kabadayi, S., Gruber, T., … Solnet, D. (2013). Understanding Generation Y and their use of social media: a review and research agenda. *Journal of Service Management*, *24*(3), 245–267. http://dx.doi.org/10.1108/09564231311326987

Bomar, P. J. (2004). *Promoting Health in Families: Applying Family Research and Theory to Nursing Practice*. Elsevier Health Sciences.

Borgatti, S., Everett, M., & Freeman, L. (2002). *Ucinet for Windows: Software for Social Network Analysis*. Analytic Technologies.

Böttcher, M., Höppner, F., & Spiliopoulou, M. (2008). On Exploiting the Power of Time in Data Mining. *SIGKDD Explor. Newsl.*, *10*(2), 3–11. https://doi.org/10.1145/1540276.1540278

Boyd, D. M., & Ellison, N. B. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, *13*(1), 210–230. https://doi.org/10.1111/j.1083-6101.2007.00393.x

Brasil, P., Pereira, J. P. J., Moreira, M. E., Ribeiro Nogueira, R. M., Damasceno, L., Wakimoto, M., … Nielsen-Saines, K. (2016). Zika Virus Infection in Pregnant Women in Rio de

Janeiro. *New England Journal of Medicine*, *375*(24), 2321–2334. https://doi.org/10.1056/NEJMoa1602412

Bredl, K., Hünniger, J., & Jensen, J. L. (2012). Methods for Analyzing Social Media: Introduction to the Special Issue. *Journal of Technology in Human Services*, *30*(3–4), 141–144. https://doi.org/10.1080/15228835.2012.750218

Brin, S., & Page, L. (2012). Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer Networks*, *56*(18), 3825–3833. https://doi.org/10.1016/j.comnet.2012.10.007

Brodsky, A. (2017). "Rape-Adjacent": Imagining Legal Responses to Nonconsensual Condom Removal. *Social Science Electronic Publishing*, *32*. Retrieved from http://papers.ssrn.com/sol3/papers....act_id=2954726

Bruns, A. (2006). Towards Produsage: Futures for User-Led Content Production. In F. Sudweeks, H. Hrachovec, & C. Ess (Eds.), *Creative Industries Faculty* (pp. 275–284). Tartu, Estonia: Murdoch University. Retrieved from http://eprints.qut.edu.au/4863/

Bruns, A. (2008). *Blogs, Wikipedia, Second Life, and Beyond: From Production to Produsage*. Peter Lang.

Burgess, J., & Green, J. (2013). *YouTube: Online Video and Participatory Culture*. John Wiley & Sons.

Burton, K., Java, A., & Soboroff, I. (2009, May). The icwsm 2009 spinn3r dataset. In *Proceedings of the Third Annual Conference on Weblogs and Social Media (ICWSM 2009), San Jose, CA*.

Burton, K., Kasch, N., & Soboroff, I. (2011). The icwsm 2011 spinn3r dataset. In *Proceedings of the Annual Conference on Weblogs and Social Media (ICWSM 2011)*.

C. Ross, M. Terras, C. Warwick, & A. Welsh. (2011). Enabled backchannel: conference Twitter use by digital humanists. *Journal of Documentation*, *67*(2), 214–237. https://doi.org/10.1108/00220411111109449

Campos, R., Dias, G., Jorge, A. M., & Jatowt, A. (2014). Survey of Temporal Information Retrieval and Related Applications. *ACM Comput. Surv.*, *47*(2), 15:1–15:41. https://doi.org/10.1145/2619088

Carboni, O. A., & Russu, P. (2015). Assessing Regional Wellbeing in Italy: An Application of Malmquist–DEA and Self-organizing Map Neural Clustering. *Social Indicators Research*, *122*(3), 677–700. https://doi.org/10.1007/s11205-014-0722-7

Chakraborty, G., Miyanishi, Y., Mizuno, K., Yamamoto, M., Togashi, A., & Noguchi, S. (2006). Collection and analysis of life-style data - a novel approach to improve health-

consciousness. In *HEALTHCOM 2006 8th International Conference on e-Health Networking, Applications and Services* (pp. 174–179). https://doi.org/10.1109/HEALTH.2006.246442

Chmiel, A., Sobkowicz, P., Sienkiewicz, J., Paltoglou, G., Buckley, K., Thelwall, M., & Hołyst, J. A. (2011). Negative emotions boost user activity at BBC forum. *Physica A: Statistical Mechanics and Its Applications*, *390*(16), 2936–2944. https://doi.org/10.1016/j.physa.2011.03.040

Chorev, N. (2012). *The World Health Organization between North and South*. Cornell University Press. Retrieved from https://muse-jhu-edu.ezproxy.lib.uwm.edu/book/24105/

Coates, T. (2003). My working definition of social software. Retrieved from https://www.researchgate.net/publication/246913269_My_working_definition_of_Social_Software

Cognizant. (2014). *Social network analysis: Bringing visibility to your connections.*

Cohen, J. J., Blevins, M., Mapenzi, A., Reppart, L., Reppart, J., Mainthia, R., & Wester, C. W. (2014). Overcoming the perceived barriers to health care access among single mothers in coastal Kenya. *International Journal of Public Health*, *59*(1), 189–196. https://doi.org/10.1007/s00038-013-0511-0

Collins, G., & Quan-Haase, A. (2014). Are Social Media Ubiquitous in Academic Libraries? A Longitudinal Study of Adoption and Usage Patterns. *Journal of Web Librarianship*, *8*(1), 48–68. https://doi.org/10.1080/19322909.2014.873663

Cools, S., Fiva, J. H., & Kirkebøen, L. J. (2015). Causal Effects of Paternity Leave on Children and Parents. *The Scandinavian Journal of Economics*, *117*(3), 801–828. https://doi.org/10.1111/sjoe.12113

Cooper, W. S. (1971). A definition of relevance for information retrieval. *Information Storage and Retrieval*, *7*(1), 19–37. https://doi.org/10.1016/0020-0271(71)90024-6

Craft-Rosenberg, M., & Pehler, S.-R. (2011). *Encyclopedia of Family Health*. SAGE.

Crimson Hexagon. (2014). *Crimson Hexagon: Social media monitoring and analysis for consumer brands.*

Croghan, C. F., Moone, R. P., & Olson, A. M. (2014). Friends, Family, and Caregiving Among Midlife and Older Lesbian, Gay, Bisexual, and Transgender Adults. *Journal of Homosexuality*, *61*(1), 79–102. https://doi.org/10.1080/00918369.2013.835238

Damaske, S., Bratter, J. L., & Frech, A. (2017). Single mother families and employment, race, and poverty in changing economic times. *Social Science Research*, *62*, 120–133. https://doi.org/10.1016/j.ssresearch.2016.08.008

Dawson, J. (2010). Doctors join patients in going online for health information. *New Media Age*, 7.

Deboeck, G., & Kohonen, T. (2013). *Visual Explorations in Finance: with Self-Organizing Maps*. Springer Science & Business Media.

Deering, M. J., & Harris, J. (1996). Consumer health information demand and delivery: implications for libraries. *Bulletin of the Medical Library Association*, *84*(2), 209–216.

DeNavas-Walt, C. (2010). *Income, Poverty, and Health Insurance Coverage in the United States (2005)*. DIANE Publishing.

Deursen, A. J. A. M. van, & Dijk, J. A. G. M. van. (2015). Internet skill levels increase, but gaps widen: a longitudinal cross-sectional analysis (2010–2013) among the Dutch population. *Information, Communication & Society*, *18*(7), 782–797. https://doi.org/10.1080/1369118X.2014.994544

DiClemente, R. J., Hansen, W. B., & Ponton, L. E. (2013). *Handbook of Adolescent Health Risk Behavior*. Springer Science & Business Media.

Ding, C., & Patra, J. C. (2007). User modeling for personalized Web search with self-organizing map. *Journal of the American Society for Information Science and Technology*, *58*(4), 494–507. https://doi.org/10.1002/asi.20497

Efron, M. (2013). Query Representation for Cross-temporal Information Retrieval. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 383–392). New York, NY, USA: ACM. https://doi.org/10.1145/2484028.2484054

Elkan, C. (2001). Magical Thinking in Data Mining: Lessons from CoIL Challenge 2000. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 426–431). New York, NY, USA: ACM. https://doi.org/10.1145/502512.502576

Ernst, J., Kellis, M., Hardison, R. C., Myers, R. M., & Wold, B. J. (2013). Integrating and mining the chromatin landscape of cell-type specificity using self-organizing maps.

Eynon, R., Schroeder, R., & Fry, J. (2009). New techniques in online research: challenges for research ethics. *Twenty-First Century Society*, *4*(2), 187–199. https://doi.org/10.1080/17450140903000308

Eysenbach, G. (2008). Medicine 2.0: Social Networking, Collaboration, Participation, Apomediation, and Openness. *Journal of Medical Internet Research*, *10*(3), e22. https://doi.org/10.2196/jmir.1030

Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2012). *Applied Longitudinal Analysis*. John Wiley & Sons.

Flewelling, R. L., & Bauman, K. E. (1990). Family Structure as a Predictor of Initial Substance Use and Sexual Intercourse in Early Adolescence. *Journal of Marriage and Family*, *52*(1), 171–181. https://doi.org/10.2307/352848

Fort, J. C., Letremy, P., & Cottrell, M. (2002). Advantages and drawbacks of the Batch Kohonen algorithm. In *ESANN* (Vol. 2, pp. 223-230).

Fox, S. (2011, May 12). The Social Life of Health Information, 2011. Retrieved December 31, 2016, from http://www.pewinternet.org/2011/05/12/the-social-life-of-health-information-2011/

Fox, S. (2014, January 15). The social life of health information. Retrieved June 16, 2018, from http://www.pewresearch.org/fact-tank/2014/01/15/the-social-life-of-health-information/

Fox, S., & Jones, S. (2009, June 11). The Social Life of Health Information. Retrieved June 16, 2018, from http://www.pewinternet.org/2009/06/11/the-social-life-of-health-information/

Fu, T. (2011). A review on time series data mining. *Engineering Applications of Artificial Intelligence*, *24*(1), 164–181. https://doi.org/10.1016/j.engappai.2010.09.007

Fu, T. C., Chung, F. L., Ng, V., & Luk, R. (2001, August). Pattern discovery from stock time series using self-organizing maps. In *Workshop Notes of KDD2001 Workshop on Temporal Data Mining* (pp. 26-29).

Fuemmeler, B. F., Behrman, P., Taylor, M., Sokol, R., Rothman, E., Jacobson, L. T., … Tercyak, K. P. (2017). Child and family health in the era of prevention: new opportunities and challenges. *Journal of Behavioral Medicine*, *40*(1), 159–174. https://doi.org/10.1007/s10865-016-9791-1

Fuertes, J. J., Domínguez, M., Reguera, P., Prada, M. A., Díaz, I., & Cuadrado, A. A. (2010). Visual dynamic model based on self-organizing maps for supervision and fault detection in industrial processes. *Engineering Applications of Artificial Intelligence*, *23*(1), 8–17. https://doi.org/10.1016/j.engappai.2009.06.001

Generous, N., Fairchild, G., Deshpande, A., Valle, S. Y. D., & Priedhorsky, R. (2016). Global Disease Monitoring and Forecasting with Wikipedia. *Online Journal of Public Health Informatics*, *8*(1). https://doi.org/10.5210/ojphi.v8i1.6530

Ghapanchi, A. H. (2015). Predicting software future sustainability: A longitudinal perspective. *Information Systems*, *49*, 40–51. https://doi.org/10.1016/j.is.2014.10.005

Gilbert, E., & Karahalios, K. (2009). Predicting Tie Strength with Social Media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 211–220). New York, NY, USA: ACM. https://doi.org/10.1145/1518701.1518736

Goodrum, A. A., Bejune, M. M., & Siochi, A. C. (2003). A State Transition Analysis of Image Search Patterns on the Web. In E. M. Bakker, M. S. Lew, T. S. Huang, N. Sebe, & X. S. Zhou (Eds.), *Image and Video Retrieval* (pp. 281–290). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-45113-7_28

Grajales III, F. J., Sheps, S., Ho, K., Novak-Lauscher, H., & Eysenbach, G. (2014). Social Media: A Review and Tutorial of Applications in Medicine and Health Care. *Journal of Medical Internet Research*, *16*(2). https://doi.org/10.2196/jmir.2912

Granka, L. A., Joachims, T., & Gay, G. (2004). Eye-tracking Analysis of User Behavior in WWW Search. In *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 478–479). New York, NY, USA: ACM. https://doi.org/10.1145/1008992.1009079

Gravetter, F., & Forzano, L.-A. (2015). *Research Methods for the Behavioral Sciences*. Cengage Learning.

Gravetter, F. J., & Forzano, L.-A. B. (2011). *Research Methods for the Behavioral Sciences*. Cengage Learning.

Gregori, D., Petrinco, M., Bo, S., Rosato, R., Pagano, E., Berchialla, P., & Merletti, F. (2011). Using Data Mining Techniques in Monitoring Diabetes Care. The Simpler the Better? *Journal of Medical Systems*, *35*(2), 277–281. https://doi.org/10.1007/s10916-009-9363-9

Guarino, N. (1998). *Formal Ontology in Information Systems: Proceedings of the First International Conference (FOIS'98), June 6-8, Trento, Italy*. IOS Press.

Guillemin, M., & Gillam, L. (2004). Ethics, Reflexivity, and "Ethically Important Moments" in Research. *Qualitative Inquiry*, *10*(2), 261–280. https://doi.org/10.1177/1077800403262360

Hagg, E., Dahinten, V. S., & Currie, L. M. (2018). The emerging use of social media for health-related purposes in low and middle-income countries: A scoping review. *International Journal of Medical Informatics*, *115*, 92–105. https://doi.org/10.1016/j.ijmedinf.2018.04.010

Halfon, N., Larson, K., Lu, M., Tullis, E., & Russ, S. (2014). Lifecourse Health Development: Past, Present and Future. *Maternal and Child Health Journal*, *18*(2), 344–365. https://doi.org/10.1007/s10995-013-1346-2

Hamm, M. P., Chisholm, A., Shulhan, J., Milne, A., Scott, S. D., Klassen, T. P., & Hartling, L. (2013). Social Media Use by Health Care Professionals and Trainees: A Scoping Review. *Academic Medicine*, *88*(9), 1376–1383. https://doi.org/10.1097/ACM.0b013e31829eb91c

Han, J., & Kamber, M. (2006). *Data Mining : Concepts and Techniques (2nd Edition)*. Burlington, MA, USA: Elsevier Science & Technology. Retrieved from http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10399307

Han, S., He, D., Yue, Z., Jiang, J., & Jeng, W. (2012). IRIS-IPS: An Interactive People Search System for HCIR Challenge. Presented at the 2012 Human-Computer Information Retrieval Symposium, Boston, MA: University of Pittsburgh. Retrieved from http://d-scholarship.pitt.edu/19002/

Hannan, M. T., & Tuma, N. B. (1979). Methods for Temporal Analysis. *Annual Review of Sociology*, *5*, 303–328.

Hansen, D., Shneiderman, B., & Smith, M. A. (2010). *Analyzing Social Media Networks with NodeXL: Insights from a Connected World*. Morgan Kaufmann.

Hansstein, F. V. (2016). The Impact of Breastfeeding on Early Childhood Obesity: Evidence From the National Survey of Children's Health. *American Journal of Health Promotion*, *30*(4), 250–258. https://doi.org/10.1177/0890117116639564

Harzing, A.-W. (2014). A longitudinal study of Google Scholar coverage between 2012 and 2013. *Scientometrics*, *98*(1), 565–575. https://doi.org/10.1007/s11192-013-0975-y

Hassan, A., Jones, R., & Klinkner, K. L. (2010). Beyond DCG: User Behavior As a Predictor of a Successful Search. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining* (pp. 221–230). New York, NY, USA: ACM. https://doi.org/10.1145/1718487.1718515

Hatzivassiloglou, V., & McKeown, K. R. (1997). Predicting the Semantic Orientation of Adjectives. In *Proceedings of the Eighth Conference on European Chapter of the Association for Computational Linguistics* (pp. 174–181). Stroudsburg, PA, USA: Association for Computational Linguistics. https://doi.org/10.3115/979617.979640

He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management*, *33*(3), 464–472. https://doi.org/10.1016/j.ijinfomgt.2013.01.001

Heidelberger CA. Health Care Professionals' Use of Online Social Networks. 2011. URL: http://cahdsu.wordpress.com/2011/04/07/infs-892-health-care-professionals-use-of-online-social-networks/ [accessed 2016-12-20]

Heilman, J. M., Kemmann, E., Bonert, M., Chatterjee, A., Ragar, B., Beards, G. M., … Laurent, M. R. (2011). Wikipedia: A Key Tool for Global Public Health Promotion. *Journal of Medical Internet Research*, *13*(1), e14. https://doi.org/10.2196/jmir.1589

Heilman, J. M., & West, A. G. (2015). Wikipedia and Medicine: Quantifying Readership, Editors, and the Significance of Natural Language. *Journal of Medical Internet Research*, *17*(3). https://doi.org/10.2196/jmir.4069

Hermida, A. (2010). *Twittering the News: The Emergence of Ambient Journalism* (SSRN Scholarly Paper No. ID 1732598). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=1732598

Hey, T. (2012). The Fourth Paradigm – Data-Intensive Scientific Discovery. In S. Kurbanoğlu, U. Al, P. L. Erdoğan, Y. Tonta, & N. Uçak (Eds.), *E-Science and Information Management* (pp. 1–1). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-33299-9_1

Hill, C. A., Dean, E., & Murphy, J. (2013). *Social Media, Sociality, and Survey Research*. John Wiley & Sons.

Himberg, J., Alhoniemi, E., & Parhankangas, J. (2000). SOM toolbox for Matlab 5. *Helsinki University of Technology, Helsinki*.

Hoffart, J., Suchanek, F. M., Berberich, K., & Weikum, G. (2013). YAGO2: A Spatially and Temporally Enhanced Knowledge Base from Wikipedia (Extended Abstract). In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence* (pp. 3161–3165). Beijing, China: AAAI Press. Retrieved from http://dl.acm.org/citation.cfm?id=2540128.2540600

Hölscher, C., & Strube, G. (2000). Web search behavior of Internet experts and newbies. *Computer Networks*, *33*(1–6), 337–346. https://doi.org/10.1016/S1389-1286(00)00031-1

Hossain, L., Karimi, F., Wigand, R. T., & Crawford, J. W. (2015). Evolutionary longitudinal network dynamics of global zoonotic research. *Scientometrics*, *103*(2), 337–353. https://doi.org/10.1007/s11192-015-1557-y

Hosseini, M., & Abolhassani, H. (2007). Mining Search Engine Query Log for Evaluating Content and Structure of a Web Site. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence* (pp. 235–241). Washington, DC, USA: IEEE Computer Society. https://doi.org/10.1109/WI.2007.77

Hughes, B., Joshi, I., Lemonde, H., & Wareham, J. (2009). Junior physician's use of Web 2.0 for information seeking and medical education: A qualitative study. *International Journal of Medical Informatics*, *78*(10), 645–655. https://doi.org/10.1016/j.ijmedinf.2009.04.008

Iba, T., Nemoto, K., Peters, B., & Gloor, P. A. (2010). Analyzing the Creative Editing Behavior of Wikipedia Editors. *Procedia - Social and Behavioral Sciences*, *2*(4), 6441–6456. https://doi.org/10.1016/j.sbspro.2010.04.054

International Institute for Population Sciences. (2017). In *National Family Health Survey, India*. Retrieved from http://rchiips.org/nfhs/index.shtml.

Iris Xie, & Jennifer Stevenson. (2014). Social media application in digital libraries. *Online Information Review*, *38*(4), 502–523. https://doi.org/10.1108/OIR-11-2013-0261

Jain, M., Rajyalakshmi, S., Tripathy, R. M., & Bagchi, A. (2013). Temporal Analysis of User Behavior and Topic Evolution on Twitter. In V. Bhatnagar & S. Srinivasa (Eds.), *Big Data Analytics* (pp. 22–36). Springer International Publishing. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-03689-2_2

Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, *60*(11), 2169–2188. https://doi.org/10.1002/asi.21149

Jatowt, A., & Au Yeung, C. (2011). Extracting Collective Expectations About the Future from Large Text Collections. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management* (pp. 1259–1264). New York, NY, USA: ACM. https://doi.org/10.1145/2063576.2063759

Jatowt, A., Kanazawa, K., Oyama, S., & Tanaka, K. (2009). Supporting Analysis of Future-related Information in News Archives and the Web. In *Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 115–124). New York, NY, USA: ACM. https://doi.org/10.1145/1555400.1555420

Jatowt, A., Kawai, Y., & Tanaka, K. (2007). Detecting Age of Page Content. In *Proceedings of the 9th Annual ACM International Workshop on Web Information and Data Management* (pp. 137–144). New York, NY, USA: ACM. https://doi.org/10.1145/1316902.1316925

Jensen, A.-M., & McKee, L. (2003). *Children and the Changing Family: Between Transformation and Negotiation*. Psychology Press.

Joseph, S. (2012). Social Media, Political Change, and Human Rights. *Boston College International and Comparative Law Review*, *35*, 145.

Julien, H., Tan, M., & Merillat, S. (2013). Instruction for Information Literacy in Canadian Academic Libraries: A Longitudinal Analysis of Aims, Methods, and Success. *L'enseignement Visant Les Compétences Informationnelles Dans Les Bibliothèques*

Universitaires Canadiennes: Une Analyse Longitudinale Des Objectifs, Des Méthodes et Du Succès Obtenu., *37*(2), 81–102.

Kaakinen, J. R., Coehlo, D. P., Steele, R., Tabacco, A., & Hanson, S. M. H. (2014). *Family Health Care Nursing: Theory, Practice, and Research*. F.A. Davis.

Kahle, B. (1997) Preserving the Internet. *Scientific American*, 276 (3): 82–83.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, *53*(1), 59–68. https://doi.org/10.1016/j.bushor.2009.09.003

Kawai, H., Jatowt, A., Tanaka, K., Kunieda, K., & Yamada, K. (2010). ChronoSeeker: Search Engine for Future and Past Events. In *Proceedings of the 4th International Conference on Uniquitous Information Management and Communication* (pp. 25:1–25:10). New York, NY, USA: ACM. https://doi.org/10.1145/2108616.2108647

Keogh, E., Lonardi, S., & Chiu, B. "Yuan-chi." (2002). Finding Surprising Patterns in a Time Series Database in Linear Time and Space. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 550–556). New York, NY, USA: ACM. https://doi.org/10.1145/775047.775128

Keyes. (2017). Package 'WikipediR'. Retrieved from http://www.stats.bris.ac.uk/R/web/packages/WikipediR/WikipediR.pdf

Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Business Horizons*, *54*(3), 241–251. https://doi.org/10.1016/j.bushor.2011.01.005

Kim, S.-S., Kim-Godwin, Y. S., & Koenig, H. G. (2016). Family Spirituality and Family Health Among Korean-American Elderly Couples. *Journal of Religion and Health*, *55*(2), 729–746. https://doi.org/10.1007/s10943-015-0107-5

Klein, D. A., Berry-Bibee, E. N., Baker, K. K., Malcolm, N. M., Rollison, J. M., & Frederiksen, B. N. (2018). Providing quality family planning services to LGBTQIA individuals: a systematic review. *Contraception*, *97*(5), 378–391. https://doi.org/10.1016/j.contraception.2017.12.016

Klinth, R. (2008). The Best of Both Worlds?: Fatherhood and Gender Equality in Swedish Paternity Leave Campaigns, 1976-2006. *Fathering*, *6*(1), 20–38. http://dx.doi.org.ezproxy.lib.uwm.edu/10.3149/fth.0601.20

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43, 59-69.

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, *78*(9), 1464–1480. https://doi.org/10.1109/5.58325

Kohonen, T. 1995. Self-organizing Maps. Berlin: Springer-Verlag.

Kohonen, T., Kaski, S., Lagus, K., Salojarvi, J., Honkela, J., Paatero, V., & Saarela, A. (2000). Self organization of a massive document collection. *IEEE Transactions on Neural Networks*, *11*(3), 574–585. https://doi.org/10.1109/72.846729

Kolaczyk, E. D., & Csárdi, G. (2014). *Statistical Analysis of Network Data with R* (Vol. 65). New York, NY: Springer New York. Retrieved from http://link.springer.com/10.1007/978-1-4939-0983-4

König, R. (2013). Wikipedia. *Information, Communication & Society*, *16*(2), 160–177. https://doi.org/10.1080/1369118X.2012.734319

Kotch, J. (2005). *Maternal and Child Health: Programs, Problems, and Policy in Public Health*. Jones & Bartlett Learning.

Kräenbring, J., Penza, T. M., Gutmann, J., Muehlich, S., Zolk, O., Wojnowski, L., … Sarikas, A. (2014). Accuracy and Completeness of Drug Information in Wikipedia: A Comparison with Standard Textbooks of Pharmacology. *PLOS ONE*, *9*(9), e106930. https://doi.org/10.1371/journal.pone.0106930

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a Social Network or a News Media? In *Proceedings of the 19th International Conference on World Wide Web* (pp. 591–600). New York, NY, USA: ACM. https://doi.org/10.1145/1772690.1772751

Lagus, K., Honkela, T., Kaski, S., & Kohonen, T. (1996). Self-organizing maps of document collections: a new approach to interactive exploration. *International Conference on Knowledge Discovery and Data Mining*(Vol.79, pp.238-243). AAAI Press.

Law, M. H., & Kwok, J. T. (2000). Rival penalized competitive learning for model-based sequence clustering. In *15th International Conference on Pattern Recognition, 2000. Proceedings* (Vol. 2, pp. 195–198 vol.2). https://doi.org/10.1109/ICPR.2000.906046

Laxman, S., & Sastry, P. S. (2006). A survey of temporal data mining. *Sadhana*, *31*(2), 173–198. https://doi.org/10.1007/BF02719780

Lewis, D., Eysenbach, G., Kukafka, R., Stavri, P. Z., & Jimison, H. (2006). *Consumer Health Informatics: Informing Consumers and Improving Health Care*. New York, NY: Springer Science & Business Media.

Li, R., Lei, K. H., Khadiwala, R., & Chang, K. C. C. (2012, April). Tedas: A twitter-based event detection and analysis system. In *Data engineering (icde), 2012 ieee 28th international conference on* (pp. 1273-1276). IEEE.

Li, C.-T., Shan, M.-K., & Lin, S.-D. (2011). Context-based People Search in Labeled Social Networks. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management* (pp. 1607–1612). New York, NY, USA: ACM. https://doi.org/10.1145/2063576.2063809

Liao, S.-H., Chu, P.-H., & Hsiao, P.-Y. (2012). Data mining techniques and applications – A decade review from 2000 to 2011. *Expert Systems with Applications*, *39*(12), 11303–11311. https://doi.org/10.1016/j.eswa.2012.02.063

Lin, J., Wang, B., Wang, N., & Lu, Y. (2013). Understanding the evolution of consumer trust in mobile commerce: a longitudinal study. *Information Technology and Management*, *15*(1), 37–49. https://doi.org/10.1007/s10799-013-0172-y

Lin, X., Soergel, D., & Marchionini, G. (1991). A Self-organizing Semantic Map for Information Retrieval. In *Proceedings of the 14th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 262–269). New York, NY, USA: ACM. https://doi.org/10.1145/122860.122887

Lingras, P., Hogo, M., Snorek, M., & West, C. (2005). Temporal analysis of clusters of supermarket customers: conventional versus interval set approach. *Information Sciences*, *172*(1–2), 215–240. https://doi.org/10.1016/j.ins.2004.12.007

Lipizzi, C., Iandoli, L., & Ramirez Marquez, J. E. (2015). Extracting and evaluating conversational patterns in social media: A socio-semantic analysis of customers' reactions to the launch of new products using Twitter streams. *International Journal of Information Management*, *35*(4), 490–503. https://doi.org/10.1016/j.ijinfomgt.2015.04.001

Liu, H., & Kešelj, V. (2007). Combined mining of Web server logs and web contents for classifying user navigation patterns and predicting users' future requests. *Data & Knowledge Engineering*, *61*(2), 304–330. https://doi.org/10.1016/j.datak.2006.06.001

Liu, X., Jiang, T., & Ma, F. (2013). Collective dynamics in knowledge networks: Emerging trends analysis. *Journal of Informetrics*, *7*(2), 425–438. https://doi.org/10.1016/j.joi.2013.01.003

López-Rubio, E., & Díaz Ramos, A. (2014). Grid topologies for the self-organizing map. *Neural Networks*, *56*, 35–48. https://doi.org/10.1016/j.neunet.2014.05.001

Lu An, Jin Zhang, & Chuanming Yu. (2011). The Visual Subject Analysis of Library and Information Science Journals with Self-Organizing Map. *Knowledge Organization*, *38*(4), 299–320.

Subramaniam, M., Prasad, R. O., Abdin, P., Vaingankar, J. A., & Chong, S. A. (2017). Single mothers have a higher risk of mood disorders. Retrieved from https://open-access.imh.com.sg/handle/123456789/4754

Malone, T. W., Laubacher, R., & Dellarocas, C. (2009). *Harnessing Crowds: Mapping the Genome of Collective Intelligence* (SSRN Scholarly Paper No. ID 1381502). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=1381502

Maness, J. M. (n.d.). Library 2.0 Theory: Web 2.0 and its Implications for Libraries [text]. Retrieved October 4, 2015, from http://www.webology.org/2006/v3n2/a25.html

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge: Cambridge university press.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity. Retrieved from http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation

Marcus, S. M., Flynn, H. A., Blow, F. C., & Barry, K. L. (2003). Depressive Symptoms among Pregnant Women Screened in Obstetrics Settings. *Journal of Women's Health*, *12*(4), 373–380. https://doi.org/10.1089/154099903765448880

Matta, R., Doiron, C., & Leveridge, M. J. (2014). The Dramatic Increase in Social Media in Urology. *The Journal of Urology*, *192*(2), 494–498. https://doi.org/10.1016/j.juro.2014.02.043

Matthews, M., Tolchinsky, P., Blanco, R., Atserias, J., Mika, P., & Zaragoza, H. (2010, August). Searching through time in the New York Times. In *Proc. of the 4th Workshop on Human-Computer Interaction and Information Retrieval* (pp. 41-44).

Maxwell, J. A. (2005a). *Qualitative Research Design: An Interactive Approach*. SAGE.

Maxwell, J. A. (2005b). *Qualitative Research Design: An Interactive Approach*. SAGE.

Maxwell, J. A. (2012). *Qualitative Research Design: An Interactive Approach: An Interactive Approach*. SAGE.

McCallum AK (2002) Mallet: a machine learning for language toolkit. Retrieved from http://www.cs.umass.edu/ mccallum/mallet.

McCown, F., & Nelson, M. L. (2008). Recovering a Website's Server Components from the Web Infrastructure. In *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 124–133). New York, NY, USA: ACM. https://doi.org/10.1145/1378889.1378911

McIver, D. J., & Brownstein, J. S. (2014). Wikipedia Usage Estimates Prevalence of Influenza-Like Illness in the United States in Near Real-Time. *PLoS Computational Biology*, *10*(4), 1–8. https://doi.org/10.1371/journal.pcbi.1003581

McNab, C. (2009). What social media offers to health professionals and citizens. *Bulletin of the World Health Organization*, *87*(8), 566–566. https://doi.org/10.1590/S0042-96862009000800002

Metzger, M. J. (2007). Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology*, *58*(13), 2078–2091. https://doi.org/10.1002/asi.20672

Meyer, D., Hornik, K., & Feinerer, I. (2008). Text Mining Infrastructure in R. *Journal of Statistical Software*, *25*(5), 1–54.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Team, T. G. B., … Aiden, E. L. (2011). Quantitative Analysis of Culture Using Millions of Digitized Books. *Science*, *331*(6014), 176–182. https://doi.org/10.1126/science.1199644

Miguel A.P.M. Lejeune. (2001). Measuring the impact of data mining on churn management. *Internet Research*, *11*(5), 375–387. https://doi.org/10.1108/10662240110410183

Milne, D., & Witten, I. H. (2008). Learning to Link with Wikipedia. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management* (pp. 509–518). New York, NY, USA: ACM. https://doi.org/10.1145/1458082.1458150

Milne, D., & Witten, I. H. (2013). An open-source toolkit for mining Wikipedia. *Artificial Intelligence*, *194*, 222–239. https://doi.org/10.1016/j.artint.2012.06.007

Miner, G. (2012). *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications*. Academic Press.

Mitrović, M., Paltoglou, G., & Tadić, B. (2011). Quantitative analysis of bloggers' collective behavior powered by emotions. *Journal of Statistical Mechanics: Theory and Experiment*, *2011*(02), P02005. https://doi.org/10.1088/1742-5468/2011/02/P02005

Montiel-Overall, P., & Grimes, K. (2013). Teachers and librarians collaborating on inquiry-based science instruction: A longitudinal study. *Library & Information Science Research*, *35*(1), 41–53. https://doi.org/10.1016/j.lisr.2012.08.002

Moore, D. S., Notz, W., & Fligner, M. A. (2013). *The basic practice of statistics*. WH Freeman.

Moorhead, S. A., Hazlett, D. E., Harrison, L., Carroll, J. K., Irwin, A., & Hoving, C. (2013). A New Dimension of Health Care: Systematic Review of the Uses, Benefits, and Limitations of Social Media for Health Communication. *Journal of Medical Internet Research*, *15*(4). https://doi.org/10.2196/jmir.1933

Moussiades, L., & Vakali, A. (2009). Mining the Community Structure of a Web Site. In *Informatics, 2009. BCI '09. Fourth Balkan Conference in* (pp. 239–244). https://doi.org/10.1109/BCI.2009.12

Mudambi, S. M., & Schuff, D. (2010). What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com. *MIS Quarterly*, *34*(1), 185–200.

Naaman, M. (2010). Social multimedia: highlighting opportunities for search and mining of multimedia data in social media applications. *Multimedia Tools and Applications*, *56*(1), 9–34. https://doi.org/10.1007/s11042-010-0538-7

Naaman, M., Becker, H., & Gravano, L. (2011). Hip and trendy: Characterizing emerging trends on Twitter. *Journal of the American Society for Information Science & Technology*, *62*(5), 902–918. https://doi.org/10.1002/asi.21489

National Library of Medicine. (2014, December 6). Health Information [List of Links]. Retrieved December 6, 2014, from http://www.nlm.nih.gov/hinfo.html

Neviarouskaya, A., Prendinger, H., & Ishizuka, M. (2007). Textual Affect Sensing for Sociable and Expressive Online Communication. In A. C. R. Paiva, R. Prada, & R. W. Picard (Eds.), *Affective Computing and Intelligent Interaction* (pp. 218–229). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-74889-2_20

Nooy, W. de, Mrvar, A., & Batagelj, V. (2011). *Exploratory Social Network Analysis with Pajek*. Cambridge University Press.

Nunes, S., Ribeiro, C., & David, G. (2007). Using Neighbors to Date Web Documents. In *Proceedings of the 9th Annual ACM International Workshop on Web Information and Data Management* (pp. 129–136). New York, NY, USA: ACM. https://doi.org/10.1145/1316902.1316924

Nuti, S. V., Wayda, B., Ranasinghe, I., Wang, S., Dreyer, R. P., Chen, S. I., & Murugiah, K. (2014). The Use of Google Trends in Health Care Research: A Systematic Review. *PLOS ONE*, *9*(10), e109583. https://doi.org/10.1371/journal.pone.0109583

Oliver, P. (2010). *The Student's Guide to Research Ethics*. McGraw-Hill Education (UK).

Olson, D. H. (2000). Circumplex Model of Marital and Family Sytems. *Journal of Family Therapy*, *22*(2), 144–167. https://doi.org/10.1111/1467-6427.00144

Oteng-Ntim, E., Tezcan, B., Seed, P., Poston, L., & Doyle, P. (2015). Lifestyle interventions for obese and overweight pregnant women to improve pregnancy outcome: a systematic review and meta-analysis. *The Lancet*, *386*, S61. https://doi.org/10.1016/S0140-6736(15)00899-5

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999, November 11). The PageRank Citation Ranking: Bringing Order to the Web. [Techreport]. Retrieved September 10, 2015, from http://ilpubs.stanford.edu:8090/422/

Paltoglou, G. (2014). Sentiment Analysis in Social Media. In N. Agarwal, M. Lim, & R. T. Wigand (Eds.), *Online Collective Action* (pp. 3–17). Springer Vienna. Retrieved from http://link.springer.com/chapter/10.1007/978-3-7091-1340-0_1

Paltoglou, G., Thelwall, M., & Buckely, K. (2010, May). Online textual communications annotated with grades of emotion strength. In *Proceedings of the 3rd International Workshop of Emotion: Corpora for research on Emotion and Affect* (pp. 25-31).

Pandit, N. (1996). The Creation of Theory: A Recent Application of the Grounded Theory Method. *The Qualitative Report*, *2*(4), 1–15.

Pang, B., & Lee, L. (2005). Seeing Stars: Exploiting Class Relationships for Sentiment Categorization with Respect to Rating Scales. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics* (pp. 115–124). Stroudsburg, PA, USA: Association for Computational Linguistics. https://doi.org/10.3115/1219840.1219855

Pasca, M. (2008). Towards Temporal Web Search. In *Proceedings of the 2008 ACM Symposium on Applied Computing* (pp. 1117–1121). New York, NY, USA: ACM. https://doi.org/10.1145/1363686.1363946

Passmore, D. L. (2011). Social network analysis: Theory and applications.*Tersedia: http://code. pediapress. com/[12 Juni 2014]*.

Patrick, T. B., Monga, H. K., Sievert, M. C., Hall, J. H., & Longo, D. R. (2001a). Evaluation of Controlled Vocabulary Resources for Development of a Consumer Entry Vocabulary for Diabetes. *Journal of Medical Internet Research*, *3*(3), e24. https://doi.org/10.2196/jmir.3.3.e24

Patrick, T. B., Monga, H. K., Sievert, M. C., Hall, J. H., & Longo, D. R. (2001b). Evaluation of Controlled Vocabulary Resources for Development of a Consumer Entry Vocabulary for Diabetes. *Journal of Medical Internet Research*, *3*(3). https://doi.org/10.2196/jmir.3.3.e24

Pelechrinis, K., & Krishnamurthy, P. (2012). Location-based Social Network Users Through a Lense: Examining Temporal User Patterns. *AAAI SNSC*. Retrieved from http://www.aaai.org/ocs/index.php/FSS/FSS12/paper/view/5555

Peng, T.-Q. (2015). Assortative mixing, preferential attachment, and triadic closure: A longitudinal study of tie-generative mechanisms in journal citation networks. *Journal of Informetrics*, *9*(2), 250–262. https://doi.org/10.1016/j.joi.2015.02.002

Perrin, A. (2015). Social Media Usage: 2005-2015. Retrieved from http://ictlogy.net/bibliography/reports/projects.php?idp=2894

Petrović, S., Osborne, M., & Lavrenko, V. (2010). Streaming First Story Detection with Application to Twitter. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 181–189). Stroudsburg, PA, USA: Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=1857999.1858020

Pivovarov, R., Albers, D. J., Hripcsak, G., Sepulveda, J. L., & Elhadad, N. (2014). Temporal trends of hemoglobin A1c testing. *Journal of the American Medical Informatics Association*, *21*(6), 1038–1044. https://doi.org/10.1136/amiajnl-2013-002592

Puolamäki, K., Salojärvi, J., Savia, E., Simola, J., & Kaski, S. (2005). Combining Eye Movements and Collaborative Filtering for Proactive Information Retrieval. In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 146–153). New York, NY, USA: ACM. https://doi.org/10.1145/1076034.1076062

Radford, M. L., & Connaway, L. S. (2013). Not dead yet! A longitudinal study of query type and ready reference accuracy in live chat and IM reference. *Library & Information Science Research (07408188)*, *35*(1), 2–13. https://doi.org/10.1016/j.lisr.2012.08.001

Radinsky, K., Diaz, F., Dumais, S., Shokouhi, M., Dong, A., & Chang, Y. (2013). Temporal Web Dynamics and Its Application to Information Retrieval. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining* (pp. 781–782). New York, NY, USA: ACM. https://doi.org/10.1145/2433396.2433500

Räihä, K.-J., Aula, A., Majaranta, P., Rantala, H., & Koivunen, K. (2005). Static Visualization of Temporal Eye-Tracking Data. In M. F. Costabile & F. Paternò (Eds.), *Human-Computer Interaction - INTERACT 2005* (pp. 946–949). Springer Berlin Heidelberg. https://doi.org/10.1007/11555261_76

Rattenbury, T., Good, N., & Naaman, M. (2007). Towards Automatic Extraction of Event and Place Semantics from Flickr Tags. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 103–110). New York, NY, USA: ACM. https://doi.org/10.1145/1277741.1277762

Rosenbaum, P. R. (2002). Observational Studies. In *Observational Studies* (pp. 1–17). Springer New York. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4757-3692-2_1

Ross, C., Orr, E. S., Sisic, M., Arseneault, J. M., Simmering, M. G., & Orr, R. R. (2009). Personality and motivations associated with Facebook use. *Computers in Human Behavior*, *25*(2), 578–586. https://doi.org/10.1016/j.chb.2008.12.024

Ruocco, M., & Ramampiaro, H. (2015). Geo-temporal distribution of tag terms for event-related image retrieval. *Information Processing & Management*, *51*(1), 92–110. https://doi.org/10.1016/j.ipm.2014.09.001

Russell, M. A. (2013). *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More*. O'Reilly Media, Inc.

Russell, S. T., Ryan, C., Toomey, R. B., Diaz, R. M., & Sanchez, J. (2011). Lesbian, Gay, Bisexual, and Transgender Adolescent School Victimization: Implications for Young Adult Health and Adjustment. *Journal of School Health*, *81*(5), 223–230. https://doi.org/10.1111/j.1746-1561.2011.00583.x

Ryan, M. K., & Haslam, S. A. (2005). The glass cliff: Evidence that women are over-represented in precarious leadership positions. *British Journal of management*, *16*(2), 81-90.

Saif, H., He, Y., & Alani, H. (2012). Semantic Sentiment Analysis of Twitter. In P. Cudré-Mauroux, J. Heflin, E. Sirin, T. Tudorache, J. Euzenat, M. Hauswirth, … E. Blomqvist (Eds.), *The Semantic Web – ISWC 2012* (pp. 508–524). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-35176-1_32

SalahEldeen, H. M., & Nelson, M. L. (2013). Carbon Dating the Web: Estimating the Age of Web Resources. In *Proceedings of the 22Nd International Conference on World Wide Web* (pp. 1075–1082). Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee. Retrieved from http://dl.acm.org/citation.cfm?id=2487788.2488121

Sarlin, P., Yao, Z., & Eklund, T. (2012). A Framework for State Transitions on the Self-Organizing Map: Some Temporal Financial Applications. *Intelligent Systems in Accounting, Finance and Management*, *19*(3), 189–203. https://doi.org/10.1002/isaf.1328

Sayyadi, H., Hurst, M., & Maykov, A. (2009). Event Detection and Tracking in Social Streams. *International Conference on Weblogs and Social Media*.

Seale, C., Gobo, G., Gubrium, J. F., & Silverman, D. (2004). *Qualitative Research Practice*. SAGE.

Sebastiani, P., Ramoni, M., Cohen, P., Warwick, J., & Davis, J. (1999). Discovering Dynamics Using Bayesian Clustering. In D. J. Hand, J. N. Kok, & M. R. Berthold (Eds.), *Advances in Intelligent Data Analysis* (pp. 199–209). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-48412-4_17

Shi, Z., Rui, H., & Whinston, A. B. (2013). *Content Sharing in a Social Broadcasting Environment: Evidence from Twitter* (SSRN Scholarly Paper No. ID 2341243). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=2341243

Shields, L., Zappia, T., Blackwood, D., Watkins, R., Wardrop, J., & Chapman, R. (2012). Lesbian, Gay, Bisexual, and Transgender Parents Seeking Health Care for Their Children: A

Systematic Review of the Literature. *Worldviews on Evidence-Based Nursing*, *9*(4), 200–209. https://doi.org/10.1111/j.1741-6787.2012.00251.x

Singh, B., & Singh, H. K. (2010). Web Data Mining research: A survey. In *2010 IEEE International Conference on Computational Intelligence and Computing Research* (pp. 1–10). https://doi.org/10.1109/ICCIC.2010.5705856

Smith, A., Monica, & erson. (2018, March 1). Social Media Use in 2018. Retrieved June 16, 2018, from http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/

Smith, C. A., & Stavri, P. Z. (2005). Consumer Health Vocabulary. In D. L. E. MPH RN, G. E. M. MPH, R. K. D. CHES MA, P. Z. S. MLS, & H. B. Jimison (Eds.), *Consumer Health Informatics* (pp. 122–128). Springer New York. https://doi.org/10.1007/0-387-27652-1_10

Smith, M. A., Shneiderman, B., Milic-Frayling, N., Mendes Rodrigues, E., Barash, V., Dunne, C., … Gleave, E. (2009). Analyzing (Social Media) Networks with NodeXL. In *Proceedings of the Fourth International Conference on Communities and Technologies* (pp. 255–264). New York, NY, USA: ACM. https://doi.org/10.1145/1556460.1556497

Snapp, S. D., Watson, R. J., Russell, S. T., Diaz, R. M., & Ryan, C. (2015). Social Support Networks for LGBT Young Adults: Low Cost Strategies for Positive Adjustment. *Family Relations*, *64*(3), 420–430. https://doi.org/10.1111/fare.12124

Staub, T., & Hodel, T. (2015, January). WIKIPEDIA vs. ACADEMIA: An investigation into the role of the Internet in education, with a special focus on collaborative editing tools such as Wikipedia. In *The International Scientific Conference eLearning and Software for Education* (Vol. 1, p. 13). " Carol I" National Defence University.

St. Jean, B. (2014). Devising and implementing a card-sorting technique for a longitudinal investigation of the information behavior of people with type 2 diabetes. *Library & Information Science Research*, *36*(1), 16–26. https://doi.org/10.1016/j.lisr.2013.10.002

Stenmark, D. (2008). Identifying clusters of user behavior in intranet search engine log files. *Journal of the American Society for Information Science and Technology*, *59*(14), 2232–2243. https://doi.org/10.1002/asi.20931

Stevenson, J. A., & Zhang, J. (2015). A temporal analysis of institutional repository research. *Scientometrics*, 1–35. https://doi.org/10.1007/s11192-015-1728-x

Strapparava, C., & Mihalcea, R. (2008). Learning to Identify Emotions in Text. In *Proceedings of the 2008 ACM Symposium on Applied Computing* (pp. 1556–1560). New York, NY, USA: ACM. https://doi.org/10.1145/1363686.1364052

Street, 1615 L., NW, Washington, S. 800, & Inquiries, D. 20036 202 419 4300 | M. 202 419 4349 | F. 202 419 4372 | M. (2013, December 16). Health Fact Sheet. Retrieved December 29, 2016, from http://www.pewinternet.org/fact-sheets/health-fact-sheet/

Strötgen, J., Alonso, O., & Gertz, M. (2012). Identification of Top Relevant Temporal Expressions in Documents. In *Proceedings of the 2Nd Temporal Web Analytics Workshop* (pp. 33–40). New York, NY, USA: ACM. https://doi.org/10.1145/2169095.2169102

Suchanek, F. M., Sozio, M., & Weikum, G. (2009). SOFIE: A Self-organizing Framework for Information Extraction. In *Proceedings of the 18th International Conference on World Wide Web* (pp. 631–640). New York, NY, USA: ACM. https://doi.org/10.1145/1526709.1526794

Suess, S. (2001). Consumer Health Information. *Journal of Hospital Librarianship*, *1*(4), 41–52.

Suh, B., Convertino, G., Chi, E. H., & Pirolli, P. (2009). The Singularity is Not Near: Slowing Growth of Wikipedia. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration* (pp. 8:1–8:10). New York, NY, USA: ACM. https://doi.org/10.1145/1641309.1641322

Surowiecki, J. (2005). *The Wisdom of Crowds*. Knopf Doubleday Publishing Group.

Taba, S. T., Hossain, L., Atkinson, S. R., & Lewis, S. (2015). Towards understanding longitudinal collaboration networks: a case of mammography performance research. *Scientometrics*, *103*(2), 531–544. https://doi.org/10.1007/s11192-015-1560-3

Takeuchi, A., & Amari, S. (1979). Formation of topographic maps and columnar microstructures in nerve fields. *Biological Cybernetics*, *35*(2), 63–72. https://doi.org/10.1007/BF00337432

Takizawa, Y., Davis, P., Kawai, M., Iwai, H., Yamaguchi, A., & Obana, S. (2006). Self-Organizing Location Estimation Method using Ad-hoc Networks. In *7th International Conference on Mobile Data Management (MDM'06)* (pp. 53–53). https://doi.org/10.1109/MDM.2006.140

Tapscott, D., & Williams, A. D. (2008). *Wikinomics: How Mass Collaboration Changes Everything*. Penguin.

Thackeray, R., Neiger, B. L., Hanson, C. L., & McKenzie, J. F. (2008). Enhancing promotional strategies within social marketing programs: use of Web 2.0 social media. *Health Promotion Practice*, *9*(4), 338–343. https://doi.org/10.1177/1524839908325335

Thelwall, M., Buckley, K., & Paltoglou, G. (2011). Sentiment in Twitter events. *Journal of the American Society for Information Science & Technology*, *62*(2), 406–418. https://doi.org/10.1002/asi.21462

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, *61*(12), 2544–2558. https://doi.org/10.1002/asi.21416

Thij, M. ten, Volkovich, Y., Laniado, D., & Kaltenbrunner, A. (2012). Modeling page-view dynamics on Wikipedia. *ArXiv:1212.5943 [Physics]*. Retrieved from http://arxiv.org/abs/1212.5943

Toyoda, M., & Kitsuregawa, M. (2006). What's Really New on the Web?: Identifying New Pages from a Series of Unstable Web Snapshots. In *Proceedings of the 15th International Conference on World Wide Web* (pp. 233–241). New York, NY, USA: ACM. https://doi.org/10.1145/1135777.1135815

Trafalis, T. B., & White, A. (2003). Data Mining Techniques for Pattern Recognition: Tornado Signatures in Doppler Weather Radar Data. *International Journal of Smart Engineering System Design*, *5*(4), 347–359. https://doi.org/10.1080/10255810390224107

Tran, D.-H., Gaber, M. M., & Sattler, K.-U. (2014). Change Detection in Streaming Data in the Era of Big Data: Models and Issues. *SIGKDD Explor. Newsl.*, *16*(1), 30–38. https://doi.org/10.1145/2674026.2674031

Tse, T. (2011). Identifying and characterizing a "consumer medical vocabulary". *Advances in Classification Research Online*, *11*(1), 141-143.

Tseng, Y.-H., Lin, Y.-I., Lee, Y.-Y., Hung, W.-C., & Lee, C.-H. (2009). A comparison of methods for detecting hot topics. *Scientometrics*, *81*(1), 73–90. https://doi.org/10.1007/s11192-009-1885-x

Tu, H. T. (2011). *Surprising decline in consumers seeking health information*. Retrieved from: http://www.hschange.com/CONTENT/1260/

Turban, E., Sharda, R., Delen, D., & Efraim, T. (2007). *Decision support and business intelligence systems*. Pearson Education India.

Turner, R. A., Irwin, C. E., Tschann, J. M., & Millstein, S. G. (1993). Autonomy, relatedness, and the initiation of health risk behaviors in early adolescence. *Health Psychology*, *12*(3), 200–208. https://doi.org/10.1037/0278-6133.12.3.200

Ultsch, A., & Siemon, H. (1990). {K}ohonen's Self Organizing Feature Maps for Exploratory Data Analysis (pp. 305–308). Presented at the Proc. INNC'90, Int. Neural Network Conf., Kluwer.

Uricchio, T., Ballan, L., Bertini, M., & Bimbo, A. D. (2013). Evaluating Temporal Information for Social Image Annotation and Retrieval. In A. Petrosino (Ed.), *Image Analysis and Processing – ICIAP 2013* (pp. 722–732). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-41181-6_73

Van de Sompel, H., Nelson, M. L., Sanderson, R., Balakireva, L. L., Ainsworth, S., & Shankar, H. (2009). Memento: Time Travel for the Web. *ArXiv:0911.1112 [Cs]*. Retrieved from http://arxiv.org/abs/0911.1112

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Family Medicine*, *37*(5), 360–363.

Wang, H., Wang, W., Yang, J., & Yu, P. S. (2002). Clustering by Pattern Similarity in Large Data Sets. In *Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data* (pp. 394–405). New York, NY, USA: ACM. https://doi.org/10.1145/564691.564737

Wang, J., Clements, M., Yang, J., de Vries, A. P., & Reinders, M. J. T. (2010). Personalization of tagging systems. *Information Processing & Management*, *46*(1), 58–70. https://doi.org/10.1016/j.ipm.2009.06.002

Wang, P., Berry, M. W., & Yang, Y. (2003). Mining longitudinal web queries: Trends and patterns. *Journal of the American Society for Information Science and Technology*, *54*(8), 743–758. https://doi.org/10.1002/asi.10262

Warren Liao, T. (2005). Clustering of time series data—a survey. *Pattern Recognition*, *38*(11), 1857–1874. https://doi.org/10.1016/j.patcog.2005.01.025

West, AG. & Milowent, E. (2013). Examining the popularity of Wikipedia articles: Catalysts, trends, and applications. *Wikipedia Signpost*. Retrieved from https://en.wikipedia.org/wiki/Wikipedia:Wikipedia_Signpost/2013-02-04/Special_report.

White, H. D., Lin, X., & Buzydlowski, J. (2004). An Associative Information Visualizer. In *IEEE Symposium on Information Visualization* (pp. r8–r8). https://doi.org/10.1109/INFVIS.2004.4

Widrow, B. (1988). DARPA Neural Network Study. *Armed Forces Communication and Electronics Association, Fairfax, VA1988*.

Wiebe, J., Wilson, T., & Cardie, C. (2006). Annotating Expressions of Opinions and Emotions in Language. *Language Resources and Evaluation*, *39*(2–3), 165–210. https://doi.org/10.1007/s10579-005-7880-9

Wikipedia. (2018). Category:Health. In *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/Category:Health

Wikipedia. (2018). Help:Category. In *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/Help:Category

Wikipedia. (2017). In *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/Wikipedia.

Wikimedia. (2018). In *Wikipedia Statistics: New Wikipedians*. Retrieved from https://stats.wikimedia.org/EN/TablesWikipediansNew.htm

Wikimedia. (2018). In *Page Views for Wikipedia, Both sites, Normalized*. Retrieved from https://stats.wikimedia.org/EN/TablesPageViewsMonthlyCombined.htm

Willging, C. E., Salvador, M., & Kano, M. (2006). Pragmatic help seeking: How sexual and gender minority groups access mental health care in a rural state. *Psychiatric Services*, *57*(6), 871–874. https://doi.org/10.1176/appi.ps.57.6.871

Willshaw, D. J., & Malsburg, C. V. D. (1976). How Patterned Neural Connections Can Be Set Up by Self-Organization. *Proceedings of the Royal Society of London B: Biological Sciences*, *194*(1117), 431–445. https://doi.org/10.1098/rspb.1976.0087

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing Contextual Polarity in Phrase-level Sentiment Analysis. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing* (pp. 347–354). Stroudsburg, PA, USA: Association for Computational Linguistics. https://doi.org/10.3115/1220575.1220619

Winter, J., Zadpoor, A., & Dodou, D. (2014). The expansion of Google Scholar versus Web of Science: a longitudinal study. *Scientometrics*, *98*(2), 1547–1565. https://doi.org/10.1007/s11192-013-1089-2

World Health Organization. *Who we are, what we do*. Retrieved June 20, 2017, from http://www.who.int/about/en/

Xi, X., Keogh, E., Shelton, C., Wei, L., & Ratanamahatana, C. A. (2006). Fast Time Series Classification Using Numerosity Reduction. In *Proceedings of the 23rd International Conference on Machine Learning* (pp. 1033–1040). New York, NY, USA: ACM. https://doi.org/10.1145/1143844.1143974

Yang, H.-L., & Lai, C.-Y. (2010). Motivations of Wikipedia content contributors. *Computers in Human Behavior*, *26*(6), 1377–1383. https://doi.org/10.1016/j.chb.2010.04.011

Yoo, J. S., & Shekhar, S. (2009). Similarity-Profiled Temporal Association Mining. *IEEE Transactions on Knowledge and Data Engineering*, *21*(8), 1147–1161. https://doi.org/10.1109/TKDE.2008.185

Yoshida, M., Arase, Y., Tsunoda, T., & Yamamoto, M. (2015). Wikipedia Page View Reflects Web Search Trend (pp. 1–2). ACM Press. https://doi.org/10.1145/2786451.2786495

Zeng, Q., Tse, T., Divita, G., Keselman, A., Crowell, J., Browne, A., … Ngo, L. (2007). Term Identification Methods for Consumer Health Vocabulary Development. *Journal of Medical Internet Research*, *9*(1), e4. https://doi.org/10.2196/jmir.9.1.e4

Zhang, H., Ho, T. B., & Lin, M. S. (2004). A Non-parametric Wavelet Feature Extractor for Time Series Classification. In H. Dai, R. Srikant, & C. Zhang (Eds.), *Advances in Knowledge*

Discovery and Data Mining (pp. 595–603). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-24775-3_71

Zhang, J. (2007). *Visualization for Information Retrieval*. Springer Science & Business Media.

Zhang, J., & An, L. (2010). Visual component plane analysis for the medical subjects based on a transaction log / L'analyse visuelle selon les plans de composants (component plane analysis) dans le cas de sujets médicaux à partir d'un journal de transactions. *Canadian Journal of Information and Library Science*, *34*(1), 83–111. https://doi.org/10.1353/ils.0.0006

Zhang, J., An, L., Tang, T., & Hong, Y. (2009). Visual health subject directory analysis based on users' traversal activities. *Journal of the American Society for Information Science and Technology*, *60*(10), 1977–1994. https://doi.org/10.1002/asi.21153

Zhang, X., & Zhu, F. (2006, January). Intrinsic motivation of open content contributors: The case of Wikipedia. In *Workshop on Information Systems and Economics* (Vol. 10, p. 4).

Zhao, Q., Mitra, P., & Chen, B. (2007). Temporal and information flow based event detection from social text streams. *National Conference on Artificial Intelligence* (Vol.2, pp.1501-1506). AAAI Press.

Zielstorff, R. D. (2003). Controlled vocabularies for consumer health. *Journal of Biomedical Informatics*, *36*(4–5), 326–333. https://doi.org/10.1016/j.jbi.2003.09.015

APPENDIX A:

Topics, Themes, and Associated Entries

Selected Topics	Themes	Entries
Child Maltreatment: 241 entries	Abuse, violence, harm, and subordination (AVHS): 118 entries	(1) 2009 Plymouth child abuse case, (2) A Modest Proposal, (3) Abortion, (4) Abuse, (5) Abusive power and control, (6) Adverse Childhood Experiences Study, (7) Athletes and domestic violence, (8) Aylesbury child sex abuse ring, (9) Baby farming, (10) Banbury child sex abuse ring, (11) Blackmail, (12) Brainwashing, (13) Bristol child sex abuse ring, (14) Candace Newmaker, (15) Catholic Church sexual abuse cases, (16) Child-on-child sexual abuse, (17) Child abduction, (18) Child abuse (skin signs), (19) Child abuse in China, (20) Child abuse in New Zealand, (21) Child abuse, (22) Child erotica, (23) Child grooming, (24) Child murder, (25) Child neglect, (26) Child of Rage, (27) Child pornography, (28) Child prostitution, (29) Child sacrifice, (30) Child sex tourism, (31) Child sexual abuse accommodation syndrome, (32) Child sexual abuse by UN peacekeepers, (33) Child sexual abuse in Australia, (34) Child sexual abuse in New York City religious institutions, (35) Child sexual abuse in Nigeria, (36) Child sexual abuse in the United Kingdom, (37) Child sexual abuse, (38) Cinderella effect, (39) Cleveland child abuse scandal, (40) Collingswood Boys, (41) Corporal punishment in the home, (42) Covert incest, (43) Cycle of violence, (44) Dave Pelzer, (45) Day-care sex-abuse hysteria, (46) Death of Baby P, (47) Death of Daniel Valerio, (48) Debate on the causes of clerical child abuse, (49) Derby child sex abuse ring, (50) Destabilisation, (51) Disability abuse, (52) Domestic violence, (53) Dysfunctional family, (54) Early infanticidal childrearing, (55) Exploitation of labour, (56) Extortion, (57) Female perversion, (58) Feral child, (59) Filicide, (60) Flying monkeys (psychology), (61) Franklin child prostitution ring allegations, (62) Haleigh Poutre, (63) Halifax child sex abuse ring, (64) Hostage, (65) Human trafficking, (66) Infant exposure, (67) Infanticide, (68) Institutional abuse, (69) Isolation to facilitate abuse, (70) Jimmy Savile sexual abuse scandal, (71) Jonathan Swift, (72) Kasur child sexual abuse scandal, (73) Keighley child sex abuse ring, (74) Kern County child abuse cases, (75) Kidnapping, (76) List of child abuse cases featuring long-term detention, (77) List of satanic ritual abuse allegations, (78) Margaret Garner, (79) Miyuki Ishikawa, (80) Mormon abuse cases, (81) Murder of Sylvia Likens, (82) Narcissistic abuse, (83) North Wales child abuse scandal, (84) Orkney child abuse scandal, (85) Outline of domestic violence, (86) Overlaying, (87) Oxford child sex abuse ring, (88) Parental alienation, (89) Penn State child sex abuse scandal, (90) Physical abuse, (91) Power and Control- Domestic Violence in America, (92) Psychological abuse, (93) Psychological manipulation, (94) Relationship between child

		pornography and child sexual abuse, (95) Religious abuse, (96) Rochdale child sex abuse ring, (97) Rotherham child sexual exploitation scandal, (98) Satanic ritual abuse, (99) School bullying, (100) School corporal punishment, (101) Sex-selective abortion, (102) Sexual abuse scandal in Fall River diocese, (103) Sexual abuse scandal in the Catholic archdiocese of Boston, (104) Sexual abuse scandal in the Catholic diocese of Orange, (105) Sexual abuse scandal in the Congregation of Christian Brothers, (106) Sexual abuse scandal in the English Benedictine Congregation, (107) Sexual abuse, (108) Sexual slavery, (109) Sibling abuse, (110) Slavery, (111) Social undermining, (112) Telford child sexual exploitation scandal, (113) The Cruel Mother, (114) Unfree labour, (115) USA Gymnastics sex abuse scandal, (116) Verbal abuse, (117) Victim playing, (118) Victimisation
	Children, youth, families and friends (CYFF): 28 entries	(1) Attachment in adults, (2) Attachment theory and psychology of religion, (3) Child soldiers in Sierra Leone, (4) Child, (5) Cinderella complex, (6) Enmeshment, (7) Extended family, (8) Family economics, (9) Family nexus, (10) Family, (11) Fathers' rights movement, (12) Fathers as attachment figures, (13) Human bonding, (14) Hypergamy, (15) Inequality within immigrant families in the United States, (16) Juvenile delinquency, (17) Maternal bond, (18) Nuclear family, (19) Nurture kinship, (20) Parental abuse by children, (21) Paternity fraud, (22) Vocal school, (23) Rotten kid theorem, (24) Runaway (dependent), (25) School discipline, (26) Sociology of the family, (27) Teenage rebellion, (28) Work-family balance in the United States
	Health problems and risks (CM-HPR): 33 entries	(1) Alcoholism in family systems, (2) Atlas personality, (3) Borderline personality disorder, (4) Child sexuality, (5) Complex post-traumatic stress disorder, (6) Conduct disorder, (7) Developmental impact of child neglect in early childhood, (8) Effects of domestic violence on children, (9) Emotional dysregulation, (10) Emotional self-regulation, (11) Enabling, (12) Externalizing disorders, (13) Foster care, (14) Healthy narcissism, (15) Infant mortality, (16) Karpman drama triangle, (17) List of countries by infant and under-five mortality rates, (18) Men's health, (19) Narcissistic parent, (20) Oppositional defiant disorder, (21) Pedophilia, (22) Pseudobulbar affect, (23) Psychosomatic medicine, (24) Reactive attachment disorder, (25) Reduced affect display, (26) Self psychology, (27) Social determinants of health, (28) Spiritual crisis, (29) Substance abuse, (30) Sudden infant death syndrome, (31) Traumatic bonding, (32) Vulnerable adult, (33) Women's health
	Support and protection (CM-SP): 62 entries	(1) Abuse defense, (2) Abuse prevention program, (3) AMBER Alert, (4) Attachment-based psychotherapy, (5) Attachment in children, (6) Attachment parenting, (7) Attachment theory, (8) Attachment therapy, (9) Barnardo's, (10) Bikers Against Child Abuse, (11) Campaigns against corporal punishment, (12) Child Abuse & Neglect, (13) Child abuse image content list, (14) Child abuse investigation team, (15) Child Abuse Prevention and Treatment Act,

		(16) Child Abuse Review, (17) Child Development Index, (18) Child development, (19) Child protection, (20) Child Protective Services, (21) Child sexual abuse laws in India, (22) Child sexual abuse laws in the United States, (23) Children's rights, (24) Commission to Inquire into Child Abuse, (25) False allegation of child sexual abuse, (26) Family law, (27) Family therapy, (28) George Hosking, (29) Harry Stack Sullivan, (30) Identified patient, (31) Independent Inquiry into Child Sexual Abuse, (32) International Federation for Human Rights, (33) International Society for the Prevention of Child Abuse and Neglect, (34) Irish Society for the Prevention of Cruelty to Children, (35) Jehovah's Witnesses' handling of child sex abuse, (36) Jersey child abuse investigation 2008, (37) Journal of Child Sexual Abuse, (38) Karly's Law, (39) Laws regarding child sexual abuse, (40) List of songs about child abuse, (41) Lloyd deMause, (42) Mandated reporter, (43) Mandatory reporting in the United States, (44) Masculism, (45) Mothers' rights, (46) Multisystemic therapy, (47) National Center on Child Abuse and Neglect, (48) National Child Abuse Prevention Month, (49) National Society for the Prevention of Cruelty to Children, (50) Othermother, (51) Parental investment, (52) Parenting styles, (53) Parenting, (54) Paternal bond, (55) Royal Commission into Institutional Responses to Child Sexual Abuse, (56) Save the Children International, (57) Theraplay, (58) Trauma model of mental disorders, (59) Vicarious liability, (60) WAVE Trust, (61) Youth studies
Family Planning (150)	Family planning and reproductive health (FPRH): 95 entries	(1) Abortion-rights movements, (2) Abortion, (3) American Family Planning, (4) Avabai Bomanji Wadia, (5) Baby bonus, (6) Billings ovulation method, (7) Birth control movement in the United States, (8) Birth control, (9) British Pregnancy Advisory Service, (10) Calendar-based contraceptive methods, (11) Catholic theology of sexuality, (12) Cecile Richards, (13) Childbirth in Sri Lanka, (14) Condom, (15) Contraceptive security, (16) Couple to Couple League, (17) Dhanvanthi Rama Rau, (18) Directorate General of Family Planning, (19) Domestic violence and pregnancy, (20) Eugenics, (21) Faculty of Sexual and Reproductive Healthcare, (22) Family Planning Association of India, (23) Family Planning Association, (24) Family planning in Bangladesh, (25) Family planning in Hong Kong, (26) Family planning in India, (27) Family planning in Iran, (28) Family planning in Pakistan, (29) Family planning in the United States, (30) Family planning policy, (31) Family Planning Queensland, (32) Family planning, (33) Father's quota, (34) Female condom, (35) Fertility and intelligence, (36) Fertility testing, (37) Forced abortion, (38) Forced pregnancy, (39) Gender inequality in Bolivia, (40) Genetic diagnosis of intersex, (41) German Foundation for World Population, (42) Individual Family Service Plan, (43) Infant mortality, (44) International Perspectives on Sexual and Reproductive Health, (45) International Planned Parenthood Federation, (46) Irish Family Planning Association, (47) Japan Family Planning Association, (48) Journal of Family Planning and Reproductive Health Care, (49) Leave

		of absence, (50) Lila Rose, (51) List of sovereign states and dependent territories by birth rate, (52) Margaret Sanger, (53) Marie Stopes International, (54) Maternal health, (55) Maternity leave and the Organisation for Economic Co-operation and Development, (56) Natalism, (57) National Alliance for Optional Parenthood, (58) National Health and Family Planning Commission, (59) National Population and Family Planning Commission, (60) Natural family planning, (61) Non-consensual condom removal, (62) Office of Population Affairs, (63) One-child policy, (64) Only child, (65) Onselling of sperm, (66) Parental leave, (67) Pharmaceutical fraud, (68) Planned Parenthood, (69) POPLINE, (70) Pregnancy from rape, (71) Quiverfull, (72) Red Triangle (family planning), (73) Reproductive coercion, (74) Reproductive Health Supplies Coalition, (75) Reproductive health, (76) Reproductive rights, (77) Sex-selective abortion, (78) Sex education in India, (79) Sex selection, (80) Society for Family Health Nigeria, (81) Sperm donation, (82) Studies in Family Planning, (83) Teen dating violence, (84) Teenage pregnancy, (85) The Family Planning Association of Hong Kong, (86) Theology of the Body, (87) Time bind, (88) Timeline of reproductive rights legislation, (89) Title X, (90) Trivers–Willard hypothesis, (91) United Nations Population Fund, (92) United States pro-choice movement, (93) US Family Health Plan, (94) Voluntary childlessness, (95) Yayasan Cipta Cara Padu
	Human and environment (HE): 28 entries	(1) All in the Family, (2) Conditional cash transfer, (3) Earth system science, (4) Family Day (Canada), (5) Family, (6) Forced marriage, (7) Futures studies, (8) Gaia hypothesis, (9) Global catastrophic risk, (10) Great Recession, (11) Identity politics, (12) International development, (13) Koyaanisqatsi, (14) New feminism, (15) Overconsumption, (16) Overexploitation, (17) Parenting, (18) People smuggling, (19) Planetary boundaries, (20) Political demography, (21) Pregnancy discrimination, (22) Reserve army of labour, (23) Societal collapse, (24) Sub-replacement fertility, (25) Tragedy of the commons, (26) Voluntary Human Extinction Movement, (27) What a Way to Go- Life at the End of Empire, (28) Work-life balance
	Population problems (PP): 27 entries	(1) Antinatalism, (2) Behavioral sink, (3) Category-Overpopulation fiction, (4) Demographic trap, (5) Human migration, (6) Human overpopulation, (7) Human population planning, (8) Human sex ratio, (9) International Conference on Population and Development, (10) List of countries and dependencies by population density, (11) List of countries and dependencies by population, (12) List of countries by population growth rate, (13) List of countries by sex ratio, (14) List of organisations campaigning for population stabilisation, (15) List of people who have expressed views relating to overpopulation as a problem, (16) List of population concern organizations, (17) Malthusian trap, (18) Malthusianism, (19) Optimum population, (20) Overpopulation in domestic pets, (21) Overshoot (population), (22) Population ageing, (23) Population

		density, (24) Population ethics, (25) Population pyramid, (26) World population, (27) Zero population growth
Women's Health (207)	Discrimination, violence, harm, and subordination (DVHS): 37 entries	(1) Ageism, (2) Airline seating sex discrimination controversy, (3) Ambivalent sexism, (4) Discrimination against girls in India, (5) Female genital mutilation, (6) Femicide, (7) Gender apartheid, (8) Gender bias on Wikipedia, (9) Gender inequality in India, (10) Gender inequality, (11) Glass cliff, (12) Hegemonic masculinity, (13) Heterosexism, (14) Husband stitch, (15) Hypermasculinity, (16) LGBT stereotypes, (17) Male privilege, (18) Misogyny in horror films, (19) Misogyny, (20) Missing women, (21) Occupational segregation, (22) Occupational sexism, (23) Patriarchy, (24) Pink-collar worker, (25) Rape culture, (26) Reverse sexism, (27) Sexism in the technology industry, (28) Sexism, (29) Transphobia, (30) Triple oppression, (31) Victim blaming, (32) Wife selling, (33) Women in firefighting, (34) Women in law enforcement, (35) Women in medicine, (36) Women in Pakistan, (37) Women in the workforce
	Health problems and risks (WH-HPR): 25 entries	(1) Abortion, (2) Anilingus, (3) Birth control, (4) Complications of pregnancy, (5) Disease, (6) Diseases of affluence, (7) Diseases of poverty, (8) Drift hypothesis, (9) Gender disparities in health, (10) Gender polarization, (11) Hypertensive disease of pregnancy, (12) Incarceration of women in the United States, (13) Inequality in disease, (14) Infant mortality, (15) List of bacterial vaginosis microbiota, (16) Medical anthropology, (17) Mental health inequality, (18) Misandry, (19) Molar pregnancy, (20) Ovarian cancer, (21) Schistosomiasis, (22) Unnatural Causes: Is Inequality Making Us Sick?, (23) Water supply and sanitation in India, (24) Women's Health Issues (journal), (25) Women and smoking
	Medical and interdisciplinary subjects (MIS): 46 entries	(1) Epidemiology, (2) Etiology, (3) Face-ism, (4) Family planning, (5) Gender-blind, (6) Global health, (7) Health equity, (8) Health in China, (9) Health in India, (10) Health, (11) History of medicine, (12) History of nursing, (13) Immigrant paradox, (14) International Conference on Population and Development, (15) Intersectionality, (16) Maternal health, (17) Matriarchy, (18) Medical sociology, (19) Menstruation, (20) Mental health, (21) Molecular pathological epidemiology, (22) Pathogenesis, (23) Pathology, (24) Population Health Forum, (25) Population health, (26) Public health, (27) Race and health, (28) Reproductive health, (29) Richard G. Wilkinson, (30) Sex differences in humans, (31) Sex segregation, (32) Sexual division of labour, (33) Social determinants of health in Mexico, (34) Social determinants of health in poverty, (35) Social determinants of health, (36) Social determinants of obesity, (37) Social epidemiology, (38) Vaginal tightening, (39) Whitehall Study, (40) Women's health in China, (41) Women's health in Ethiopia, (42) Women's health in India, (43) Women's health, (44) Women's reproductive health in Russia, (45) Women's reproductive health in the United States, (46) Women who have sex with women
	Support and protection	(1) Alexandria Regional Center for Women's Health and Development, (2) American Medical Women's Association, (3)

	(WH-SP): 99 entries	AnMed Health Women's & Children's Hospital, (4) Antifeminism, (5) Association of Women's Health, Obstetric and Neonatal Nurses, (6) Australian Longitudinal Study on Women's Health, (7) Australian Women's Health Network, (8) B.C. Women's Hospital & Health Centre, (9) Black Women's Health Study, (10) Condom, (11) Dennis Raphael, (12) Equity feminism, (13) EuroHealthNet, (14) European Institute of Women's Health, (15) Female condom, (16) Female education, (17) Feminism, (18) Feminist health centers, (19) Feminist movement, (20) Feminist Women's Health Center (Atlanta, Georgia), (21) Florence Hartley, (22) Gender equality, (23) Gender feminism, (24) Gender neutrality, (25) Global Library of Women's Medicine, (26) Global Task Force on Expanded Access to Cancer Care and Control in Developing Countries, (27) Gynaecology, (28) Gynography, (29) Health (magazine), (30) Health Care for Women International, (31) Health care in the United States, (32) Health Disparities Center, (33) Health education, (34) Health literacy, (35) Health professional, (36) Healthcare and the LGBT community, (37) Healthcare in Canada, (38) Healthy People program, (39) HealthyWomen, (40) Hopkins Center for Health Disparities Solutions, (41) Hormone replacement therapy (menopause), (42) Howard Atwood Kelly, (43) International Journal of Women's Health, (44) International Planned Parenthood Federation, (45) International Women's Health Coalition, (46) Ipas (organization), (47) Journal of Midwifery & Women's Health, (48) Journal of Women's Health, (49) Kegel exercise, (50) Laura W. Bush Institute for Women's Health, (51) List of first female physicians by country, (52) List of health and fitness magazines, (53) List of medical journals, (54) List of women's studies journals, (55) Madsen v. Women's Health Center, Inc., (56) Martha Ballard, (57) Men and feminism, (58) Michael Marmot, (59) Michigan Medicine, (60) Midwife, (61) Midwifery, (62) National Organization for Men Against Sexism, (63) National Organization for Women, (64) National Women's Health Network, (65) New Space for Women's Health, (66) Office on Women's Health, (67) Oregon Health and Science University Center for Women's Health, (68) Our Bodies, Ourselves, (69) Psychology of Women Quarterly, (70) Reproductive Health Supplies Coalition, (71) Reproductive rights, (72) Separatist feminism, (73) Sex Roles (journal), (74) Society for Women's Health Research, (75) Sunnybrook Health Sciences Centre, (76) Sutter Health, (77) Sybil Shainwald, (78) Tamika D. Mallory, (79) The Heart Truth, (80) The Honest Body Project, (81) The NeuroGenderings Network, (82) Torches of Freedom, (83) United Nations Foundation, (84) United Nations Population Fund, (85) United States Department of Health and Human Services, (86) University of Pittsburgh Graduate School of Public Health, (87) Women's College Hospital, (88) Women's empowerment, (89) Women's Health (magazine), (90) Women's Health Action and Mobilization, (91) Women's Health Care Nurse Practitioner-Board Certified, (92) Women's Health

		Initiative, (93) Women's health nurse practitioner, (94) Women's medicine in antiquity, (95) Women's rights in Iran, (96) Women's rights, (97) Women's suffrage, (98) Women & Health, (99) Women in India

APPENDIX B:

High-Frequency Terms and Phrases in Each Theme

Topic	Themes	High-Frequency Terms and Phrases
Child Maltreatment	Abuse, violence, harm, and subordination	sexual abuse (1236), child abuse (522), domestic violence (441), child sexual abuse (278), New York (260), corporal punishment (259), child pornography (251), Penn State (196), sexual exploitation (173), human trafficking (151), human rights (147), sex ratio (130), abuse scandal (126), violence against women (114), United Nations (112)
	Children, youth, families and friends	United States (111), fathers' rights (66), working models (54), child support (53), New York (48), nuclear family (41), fathers' rights movement (37), family members (34), Sierra Leone (34), extended family (29), attachment theory (28), child development (27), domestic violence (26), Journal of Family (26), men and women (24), attachment figure (21)
	Health problems and risks	personality disorder (167), infant mortality (154), borderline personality disorder (142), conduct disorder (105), United States (96), New York (86), mental health (71), women's health (68), infant death (64), mortality rate (59), domestic violence (59), social determinants of health (58), attachment disorder (53), emotion regulation (51), sexual abuse (48)
	Supports and protection	sexual abuse (339), child abuse (307), child sexual abuse (209), attachment parenting (168), attachment theory (118), Amber Alert (114), New York (111), institutional responses (110), responses to child sexual (109), commission into institutional responses (108), child protection (94), United States (93), attachment therapy (76), commonsense guide (75), nurturing your baby (75), parenting book a commonsense (75)
Family Planning	Family planning and reproductive health	family planning (602), birth control (556), United States (337), planned parenthood (335), reproductive health (272), sex ratio (222), New York (198), infant mortality (159), parental leave (121), health care (115), reproductive rights (111), United Nations (108), one-child policy (93), human rights (90), maternity leave (90)
	Human and environment	United States (130), New York (96), futures studies (83), forced marriage (75), planetary boundaries (67), earth system (58), Gaia Hypothesis (50), conditional cash transfer (50), United Nations (46), tragedy of the commons (45), identity politics (42), cash transfer (41), climate change (38), human rights (38), comedy series (37)
	Population problems	population growth (150), sex ratio (128), world population (114), United Nations (67), human population (57), official population (54), United States (53), New York (34), population density (34), global population (30), list of countries (30), million people (29), country's population (28), 2011 census (28), world's population (28)
Women's Health	Discrimination, violence, harm, and subordination	New York (150), United States (137), genital mutilation (112), men and women (91), female genital mutilation (86), gender gap (81), human rights (77), rape culture (73), missing women (67), hegemonic masculinity (66), sex ratio (64), age discrimination (62), gender inequality (60), first female (58), transgender people (54)
	Health problems and risks	ovarian cancer (197), infant mortality (146), United States (132), birth control (100), water supply and sanitation (91), mental health (88), health care (85), health organization (77), mortality rate (59), world health (58),

		public health (54), reproductive health (41), World Health Organization (41), New York (40), medical anthropology (40)
	Medical and interdisciplinary subjects	public health (350), health care (263), mental health (240), United States (185), world health (123), social determinants of health (115), health organization (112), family planning (109), reproductive health (93), women's health (89), sex segregation (83), New York (78), World Health Organization (78), United Nations (64), population health (62)
	Supports and protection	health care (418), women's health (307), United States (299), New York (222), public health (195), health education (178), United Nations (140), women's suffrage (133), reproductive health (132), women's health (132), women's rights (120), human rights (117), health literacy (106), right to vote (105), reproductive rights (103)

APPENDIX C:

Entries in Each Theme during Each Time Period

Selected Topics	Themes	2010-2011	2012-2013	2014-2015	2016-2017
Child Maltreatment	Abuse, violence, harm, and subordination (AVHS)	1, 2, 3, 4, 9, 11, 12, 14, 15, 16, 17, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 34, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 48, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 61, 62, 64, 65, 66, 67, 68, 71, 74, 75, 77, 78, 79, 81, 82, 83, 84, 85, 86, 88, 89, 90, 91, 92, 93, 94, 95, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 113, 114, 116, 117, 118	1, 2, 3, 4, 5, 6, 9, 11, 12, 14, 15, 16, 17, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 34, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 61, 62, 64, 65, 66, 67, 68, 70, 71, 74, 75, 77, 78, 79, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 116, 117, 118	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 61, 62, 64, 65, 66, 67, 68, 70, 71, 72, 74, 75, 77, 78, 79, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 116, 117, 118	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118
	Children, youth, families and friends (CYFF)	1, 4, 5, 7, 8, 9, 10, 11, 13, 14, 16, 17, 18, 19, 20, 21, 23, 24, 25, 26, 27, 22, 28	1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 23, 24, 25, 26, 27, 22, 28	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 23, 24, 25, 26, 27, 22, 28	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28
	Health problems and risks (CM-HPR)	1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 33	1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33	1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33

	Support and protection (CM-SP)	1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 14, 15, 17, 18, 19, 20, 22, 23, 24, 25, 26, 27, 28, 29, 30, 32, 33, 34, 35, 36, 37, 39, 40, 41, 42, 44, 45, 46, 47, 49, 50, 51, 52, 53, 54, 56, 57, 58, 59, 60, 61	1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 14, 15, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 44, 45, 46, 47, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61
Family Planning	Family planning and reproductive health (FPRH)	2, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 19, 20, 21, 23, 26, 27, 28, 31, 32, 33, 34, 35, 36, 41, 42, 43, 45, 47, 49, 50, 51, 52, 53, 54, 56, 57, 59, 60, 62, 63, 64, 66, 67, 68, 69, 71, 72, 75, 76, 77, 79, 81, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94	1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 19, 20, 21, 22, 23, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 41, 42, 43, 45, 46, 47, 49, 50, 51, 52, 53, 54, 56, 57, 58, 59, 60, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 79, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 19, 20, 21, 22, 23, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 41, 42, 43, 44, 45, 46, 47, 49, 50, 51, 52, 53, 54, 56, 57, 58, 59, 60, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95
	Human and environment (HE)	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 21, 22, 23, 24, 25, 26, 27, 28	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28
	Population problems (PP)	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27
Women's Health	Discrimination, violence, harm, and subordination (DVHS)	1, 2, 3, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, 17, 19, 20, 21, 22, 23, 24, 25, 26, 28, 29, 30, 31, 32, 33, 35, 36, 37	1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, 17, 19, 20, 21, 22, 23, 24, 25, 26, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31,

					32, 33, 34, 35, 36, 37
	Health problems and risks (WH-HPR)	1, 2, 3, 4, 5, 6, 7, 8, 10, 13, 14, 16, 18, 19, 20, 21, 22, 23, 25	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 16, 18, 19, 20, 21, 22, 23, 25	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 19, 20, 21, 22, 23, 24, 25	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25
	Medical and interdisciplinary subjects (MIS)	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 17, 18, 19, 20, 22, 23, 24, 25, 26, 27, 28, 29, 31, 32, 35, 36, 37, 38, 39, 40, 43, 46	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 42, 43, 46	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46
	Support and protection (WH-SP)	1, 2, 3, 4, 5, 8, 10, 11, 12, 14, 15, 16, 17, 22, 23, 24, 25, 26, 27, 28, 29, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 44, 46, 47, 49, 50, 52, 53, 56, 57, 58, 59, 61, 63, 64, 65, 68, 69, 71, 72, 73, 74, 75, 76, 79, 82, 83, 84, 85, 86, 87, 89, 90, 91, 92, 94, 95, 96, 97, 98, 99	1, 2, 3, 4, 5, 8, 9, 10, 11, 12, 14, 15, 16, 17, 18, 19, 20, 22, 23, 24, 25, 26, 27, 28, 29, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 57, 58, 59, 61, 62, 63, 64, 65, 66, 68, 69, 70, 71, 72, 73, 74, 75, 76, 79, 82, 83, 84, 85, 86, 87, 89, 90, 91, 92, 94, 95, 96, 97, 98, 99	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 70, 71, 72, 73, 74, 75, 76, 79, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99

APPENDIX D:

Entries Created in Each Theme during Each Time Period

Selected Topics	Themes	Time Period	New Entries
Child Maltreatment	Abuse, violence, harm, and subordination (AVHS)	2010-2011	Child sexual abuse in New York City religious institutions Collingswood Boys Franklin child prostitution ring allegations Narcissistic abuse Outline of domestic violence Penn State child sex abuse scandal Power and Control: Domestic Violence in America Child sexual abuse accommodation syndrome Destabilisation Disability abuse Institutional abuse Sexual abuse scandal in the English Benedictine Congregation Social undermining Victim playing
		2012-2013	Abusive power and control Adverse Childhood Experiences Study Derby child sex abuse ring Oxford child sex abuse ring Rotherham child sexual exploitation scandal Telford child sex abuse ring Child abuse (skin signs) Death of Daniel Valerio Jimmy Savile sexual abuse scandal Rochdale child sex abuse ring
		2014-2015	Athletes and domestic violence Aylesbury child sex abuse ring Banbury child sex abuse ring Kasur child sexual abuse scandal Bristol child sex abuse ring Child abuse in New Zealand Child sexual abuse in Australia Child sexual abuse in Nigeria Child sexual abuse in the United Kingdom
		2016-2017	Child abuse in China USA Gymnastics sex abuse scandal Flying monkeys (psychology) Halifax child sex abuse ring Isolation to facilitate abuse Keighley child sex abuse ring

			List of child abuse cases featuring long-term detention Mormon abuse cases
	Children, youth, families and friends (CYFF)	2010-2011	Nurture kinship Work-Family Balance in the United States
		2012-2013	Enmeshment Attachment theory and psychology of religion Fathers as attachment figures Inequality within immigrant families (United States)
		2014-2015	Child soldiers in Sierra Leone
		2016-2017	-
	Health problems and risks (CM-HPR)	2010-2011	Developmental impact of child neglect in early childhood Effects of domestic violence on children Healthy narcissism Traumatic bonding
		2012-2013	Vulnerable adult
		2014-2015	Externalizing disorders
		2016-2017	Atlas personality
	Support and protection (CM-SP)	2010-2011	List of songs about child abuse Child sexual abuse laws in the United States International Society for the Prevention of Child Abuse and Neglect Multisystemic therapy (MST)
		2012-2013	Karly's Law Child sexual abuse laws in India Royal Commission into Institutional Responses to Child Sexual Abuse
		2014-2015	Bikers Against Child Abuse Child Abuse Review Mandatory reporting in the United States Child abuse image content list Independent Inquiry into Child Sexual Abuse
		2016-2017	Child Abuse & Neglect National Child Abuse Prevention Month
Family Planning	Family planning and reproductive health (FPRH)	2010-2011	Birth control movement in the United States Birth in Sri Lanka Domestic violence and pregnancy Faculty of Sexual and Reproductive Healthcare Family planning in Pakistan Father's quota Fertility monitor Japan Family Planning Association Pharmaceutical fraud

			Family Planning Queensland Lila Rose
		2012-2013	National Health and Family Planning Commission Studies in Family Planning Abortion-rights movements American Family Planning Family planning in Hong Kong Family planning in the United States Family planning policy Forced abortion Forced pregnancy Women in Bolivia Irish Family Planning Association Onselling of sperm Pregnancy from rape Reproductive coercion Reproductive Health Supplies Coalition
		2014-2015	Avabai Bomanji Wadia International Perspectives on Sexual and Reproductive Health Society for Family Health Nigeria Yayasan Cipta Cara Padu
		2016-2017	Directorate General of Family Planning Family planning in Bangladesh Genetic diagnosis of intersex Journal of Family Planning and Reproductive Health Care Non-consensual condom removal Maternity leave and the Organisation for Economic Co-operation and Development Sex education in India
	Human and environment (HE)	2010-2011	-
		2012-2013	Political demography
		2014-2015	-
		2016-2017	-
	Population problems (PP)	2010-2011	Overpopulation fiction (category)
		2012-2013	-
		2014-2015	List of organisations campaigning for population stabilisation List of people that have expressed views relating to overpopulation as a problem List of population concern organizations

		2016-2017	-
Women's Health	Discrimination, violence, harm, and subordination (DVHS)	2010-2011	Wife selling Ambivalent sexism
		2012-2013	Women in law enforcement Discrimination against girls in India
		2014-2015	Misogyny in horror films Gender bias on Wikipedia Sexism in the technology industry
		2016-2017	Husband stitch
	Health problems and risks (WH-HPR)	2010-2011	Gender polarization
		2012-2013	Gender disparities in health Incarceration of women in the United States
		2014-2015	Hypertensive disease of pregnancy List of bacterial vaginosis microbiota Women's Health Issues
		2016-2017	Mental health inequality
	Medical and interdisciplinary subjects (MIS)	2010-2011	Sexual division of labour
		2012-2013	Molecular pathological epidemiology Sex differences in humans Social determinants of health in Mexico Women's health in India Social determinants of health in poverty
		2014-2015	Women's health in Ethiopia Women's reproductive health in Russia Women's reproductive health in the United States
		2016-2017	Immigrant paradox
	Support and protection (WH-SP)	2010-2011	Gynography Journal of Midwifery & Women's Health Psychology of Women Quarterly Women's Health Care Nurse Practitioner-Board Certified Association of Women's Health, Obstetric and Neonatal Nurses Dennis Raphael Global Task Force on Expanded Access to Cancer Care and Control in Developing Countries Ipas (organization) Laura W. Bush Institute for Women's Health Torches of Freedom Women's Health (magazine)
		2012-2013	Black Women's Health Study Feminist movement

			National Organization for Men Against Sexism Office on Women's Health Feminist health centers Feminist Women's Health Center (Atlanta, Georgia) International Journal of Women's Health Madsen v. Women's Health Center, Inc. Reproductive Health Supplies Coalition
		2014-2015	EuroHealthNet Health Care for Women International Journal of Women's Health List of first female physicians by country List of women's studies journals Midwife Oregon Health and Science University Center for Women's Health Women's empowerment Australian Longitudinal Study on Women's Health Australian Women's Health Network Women's health nurse practitioner
		2016-2017	Florence Hartley Tamika D. Mallory Neurosexism International Women's Health Coalition Sybil Shainwald The Honest Body Project

CURRICULUM VITA

Yanyan Wang

EDUCATION

University of Wisconsin-Milwaukee Ph.D. Candidate (ABD), 2018

- Major: Information Retrieval
- Minor: Social Media, Data Science, Health Informatics

Renmin University of China Master of Library & Information Science, 2013

Renmin University of China B.S. in Management, 2011

- Major: Information Management System

TEACHING EXPERIENCE

Adjunct Instructor, School of Information Studies, UW-Milwaukee

- Course: Database Information Retrieval Systems

Teaching Assistant, SOIS, UW-Milwaukee

- Course: [1] Database Information Retrieval Systems (Guest Lecturer)
  [2] Information Access & Retrieval (Guest Lecturer)
  [3] Preserving Information Media

RESEARCH EXPERIENCE

Research Assistant, School of Information Studies, UW-Milwaukee

Research Projects

- Data Science Projects
  [1] Evolutions of Health-Related Topics on Wikipedia (Dissertation)
  [2] Comparison of Big Data Topics in Research-Oriented Journals and Wikipedia
  [3] Explore General Public's Perceptions of Data Science

- Users' Information Seeking Behavior Projects
  [1] Users' Eye Movements in Online Information Seeking
  [2] YouTube Users' Attitude on Diabetes-Related Videos
  [3] User Interactions through Weak Ties on Social Media

- Statistical Method Project
  Applications of Inferential Statistical Methods in Library and Information Science

- Undergraduate Projects

[1] Forest Pest Control System Based On 3G Technology Project
[2] Survey of the Development of Eco-agriculture

PUBLICATIONS

[1] Zhang, J., Wang, Y., & Zhao, Y. (2018). Applications of inferential statistical methods in library and information science. Data and Information Management. (Accepted)

[2] Wang, Y. & Zhang, J. (2017). Exploring topics related to data mining on Wikipedia. The Electronic Library, 35(4), 667-688.

[3] Zhang, J., Wang, Y., & Zhao, Y. (2017). Investigation on the statistical methods in research studies of library and information science. The Electronic Library, 35(6), 1070-1086.

[4] Wang, Y., & Zhao, Y. (2017). Seeking Information through Weak Ties on Facebook. In iConference 2017 Proceedings (pp. 899-903).

[5] Zhao, Y., Zhang, J., & Wang, Y. (2017). Social Media and Autism Support: Investigation on Autism Support Groups on Facebook. In iConference 2017 Proceedings (pp. 869-875).

[6] Zhang, J., Zhao, Y., & Wang, Y. (2016). A Study on Statistical Methods Used in Six Journals of Library and Information Science. Online Information Review, 40(3), 416-434.

[7] Wang, Y. (2016). Explore General Public's Perceptions of Data Mining: A Pilot Study. Proceedings of the 8th International Conference on Qualitative and Quantitative Methods in Libraries - Book of Abstracts (pp.149).

[8] Wang, Y., Xie, I., & Lee, S. (2015). Explore Eye Movement Patterns in Search Results Evaluation and Individual Document Evaluation. Proceedings of the Association for Information Science and Technology, 52(1), 1-4. DOI: 10.1002/pra2.2015.1450520100144.

[9] Xie, I., Wang, Y., & Lee, S. (2015). Search Result Evaluation across Different Systems and Tasks: An Eye Tracking Analysis. Proceedings of the 7th International Conference on Qualitative and Quantitative Methods in Libraries - Book of Abstracts.

[10] Wang, Y., Joo, S., & Lu, K. (2014). Exploring Topics in the Field of Data Science by Analyzing Wikipedia Documents: A Preliminary Result. Proceedings of the Association for Information Science and Technology, 51(1), 1-4. DOI: 10.1002/meet.2014.14505101116.

[11] Wang, Y. (2009). Application and improvement of Cadastral Information System for Ecological Forest in Quzhou. JOURNAL OF ZHEJIANG FORESTRY SCIENCE AND TECHNOLOGY, 29(6), 80-84.

CONFRENCE PRESENTATIONS

[1] ALISE 18 Annual Conference
Denver, CO, 2018

What Influence YouTube Users' Attitude on Diabetes-Related Videos: A Preliminary Result

[2] ALISE 17 Annual Conference
Atlanta, GA, 2017
(1) Explore Topics about Big Data from Academic Journals and Wikipedia
(2) Evolution of Health-Related Topics on Social Media Website Wikipedia
(3) Social Media and Autism Support: Investigation on the Spread of Emotion in Online Autism Support Groups

[3] iConference 2017 Annual Conference
Wuhan, China, 2017
(1) Seeking Information through Weak Ties on Social Media
(2) Social Media and Autism Support: Investigation on Autism Support Groups on Facebook

[4] 8th International Conference on Qualitative and Quantitative Methods in Libraries
London, UK, 2016
Explore General Public's Perceptions of Data Mining: A Pilot Study

[5] 78th Annual Meeting of the Association for Information Science and Technology
St. Louis, MO, 2015
Explore Eye Movement Patterns in Search Results Evaluation and Individual Document Evaluation

[6] 7th International Conference on Qualitative and Quantitative Methods in Libraries
Paris, France, 2015
Search Result Evaluation across Different Systems and Tasks: An Eye Tracking Analysis

[7] 77th Annual Meeting of the Association for Information Science and Technology
Seattle, WA, 2014
Exploring Topics in the Field of Data Science by Analyzing Wikipedia Documents: A Preliminary Result

[8] WAAL 2014 Annual Conference
Wisconsin Dells, WI, 2014
Assessment Model of Academic Library Productivity Using Multiple Regression

INTERNSHIP

Data Scientist, Jintel Health, Inc.

- Clinical data process and analysis, natural language processing, & database management

System Maintenance Intern, System Department, Library of Renmin University of China

- System maintenance & web development

Functional Supporting HR, Human Resources Department, Novartis Pharma Ltd.

- Database system management, data management, & data analysis

HONORS

R1 Distinguished Dissertation Fellowship (2018-2019)
Chancellor's Graduate Student Award (2017 - 2018, 2016 - 2017, 2014 -2015, & 2013 - 2014)
Graduate Student Travel Award (2016 & 2015)
Guanghua Scholarship (2012)

SERVICES

- Secretary of Doctoral Student Organization
  School of Information Studies, University of Wisconsin-Milwaukee