Spring 2019

Evaluating Stakeholder Bias in Stakeholder Analysis In Social Media

Ahmad A. Bajarwan
Old Dominion University

EVALUATING STAKEHOLDER BIAS IN STAKEHOLDER ANALYSIS

IN SOCIAL MEDIA

by

Ahmad A. Bajarwan
B.S. December 2004, King Abdulaziz University, Saudi Arabia
M.S. January 2014, Stevens Institute of Technology, United States

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

ENGINEERING MANAGEMENT AND SYSTEM ENGINEERING

OLD DOMINION UNIVERSITY
May 2019

Approved by:

Mamadou Seck (Director)

Ariel Pinto (Member)

Fatou Diouf (Member)

ABSTRACT

EVALUATING STAKEHOLDER BIAS IN STAKEHOLDER ANALYSIS
IN SOCIAL MEDIA

Ahmad A. Bajarwan
Old Dominion University, 2019

Stakeholder analysis is the first step in the planning of most infrastructure projects. Selecting and then applying the best method for a project's stakeholder analysis is extremely important for correctly assessing stakeholder opinions. Social media platforms allow stakeholders to participate directly in analysis. However, as with most other analysis methods, social media introduces inherent biases.

Social media is a powerful tool for communication and networking, and it also provides a valuable source of information for analyzing user opinions about infrastructure projects. By using data collected from Twitter, analysts can create networks to represent connections among users, quantify their similarities, and then use those values to predict public opinion. We can also use this information to measure bias – that is, the impact the social media has on the opinions of its users.

Research and analysis show a correlation between user similarity and user opinion that indicates bias. Additionally, I observed that disagreement was stronger than agreement – if users disagreed, they would disagree strongly; if they agreed, they had varying levels of agreement strength. In other words, disagreement was fairly polarizing, but agreement tended not to invoke strong emotions one way or another.

The nearly universal use of social media is a powerful tool to both predict and shape public opinion. Stakeholder managers can predict stakeholder opinion by using their social

network connections to determine conformity. And although social media has its own biases, its value as a data source for preliminary planning analysis should not be discounted.

To my Wife.
Without you, this would only be a dream.

ACKNOWLEDGEMENTS

As I complete my dissertation and reflect on the journey so far, it has been a period of intense learning on both a personal level and also in the area of science. Working on this dissertation has impacted me positively and I would like to show my gratitude to those people who have supported and helped me through this period.

Thank you, Allah, for your infinite blessings. Thank you for giving me the courage and strength to complete this dissertation. I give thanks to Him and I pray for his continual guidance through life. Without His help and support I would not have been able to come to this point of completion. I thank Him for the people He has put around me who have been a massive source of support. Without them I would not have been able to complete this piece of work.

I would also like to thank Dr. Seck Mamadou, my advisor and mentor, for his endless support, guidance, and invaluable training which has always spurred me on to chase my goals and objectives in this PhD program. I thank him for taking me under his wing as a student and making my experience in the program both stimulating and rewarding. On a personal level, he has always been a source of support and encouragement despite the rigors of the program and his busy schedule. It is very well-known to everyone who knows Dr. Seck Mamadou that he is a great person both personally and professionally.

I also would like to thank my dissertation committee: Dr. Ariel Pinto and Dr.Fatou Diouf, for their constructive criticism and guidance. Their discussions, advice, and constructive feedback

have been invaluable and key to the completion of this project.

Thank you to my father Dr. Abdullah Bajarwan who passed away. Thank you for the memories. Thank you for encouraging me to be the best and to do the best I can in everything I wish to achieve. Thank you to my mother, Noor Hantoush, for your support, affection, prayers, and patience to see me as a professor one day.

To the best support system, one could ask for, my lovely wife Dr. Duaa Bafail, I thank you for encouraging me while pursuing this degree and being my pillar of strength to rely on. I deeply appreciate the love and support all along the way. Thanks for bringing perspective and making things look easier than I thought in times of desperation.

I would also like to offer a special thanks to my kids, Elyana, Abdullah, and Tala, for their patience and support as well. I appreciate your understanding on nights where you wanted me to stay with you, but I had to study. The memory of my daughter, Elyana, and my son Abdullah walking in on me studying and saying they hope I get an A, pass, and be done typifies how difficult it must have been, but they have all been wonderful and made my hard studies tolerable with their beautiful smiles.

I thank my brother and sisters for their support and my mother-in-law and father-in-law for keeping me in prayer and also for the selfless support. Last but not least, I would like to thank the Ministry of Higher Education, Saudi Arabia, for the scholarship and their financial support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I – STAKEHOLDER ANALYSIS METHODS

1.1 Introduction and Background

Networking technologies such as social media have become a daily part of the majority of people's lives. They are used by people all across the globe, in many cultures and social levels. In the recent years, social media has become a popular means of communication. Increasingly, public and private organizations are using platforms like Facebook and Twitter to connect with their customers and constituencies. This raises the potential of using social media for stakeholder analysis, which aims to discover and meet the needs of the stakeholders (Waters, Burnett, Lamm, & Lucas, 2009). However, the potential influence of social media biases on the outcomes of the stakeholder analysis remains unclear. The purpose of this research is to evaluate the biases that the use of social media introduces when it is used as the primary tool for carrying out a stakeholder analysis.

Stakeholder analysis is pivotal in assisting infrastructure planning teams with not only meeting the project objectives, but also understanding the role of stakeholders throughout the execution of the project. This is key, as stakeholders directly affect the completion of projects, their outcomes, and their profits (Aaltonen, Jaakko, & Tuomas, 2008; Ruairi Brugha & Varvasovszky, 2000; Elias, Cavana, & Jackson, 2002; J. Yang, Shen, Bourne, Ho, & Xue, 2011). Large infrastructure projects usually involve numerous stakeholders, such as the communities and the governments who might have vested and have differing interests in the infrastructure. While governments might be primarily motivated by providing services that are both good and relevant for its populace, the private sectors are usually profit-driven (Fraser, Dougill, Mabee, Reed, & McAlpine, 2006). Additionally, the local communities and their members might be interested in getting better, newer services (Bryson, 2004).

The ever-growing adoption and use of stakeholder analysis indicates an increased understanding of the relevance of stakeholder populations in the decision-making process during a project (Aaltonen, 2011; Prell, Hubacek, & Reed, 2007). Formulating a project is a complex process that requires understanding of the context in which it is to take place. The future of a project depends on how the relationship between different stakeholders is handled, with stakeholder management providing a driving force for affecting the outcome (Fraser et al., 2006; Prell et al., 2007). The current stakeholder analysis methods have weaknesses which may affect the results of the analysis. Specifically, various biases could potentially be introduced in the way that participants are interviewed or the way the data is handled (Boyce & Neale, 2006). The use of a social media platform to carry out the early stakeholder analysis could help eliminate some of these biases. However, this novel method could potentially introduce biases of its own and this is what this research project aims to evaluate.

1.2 General Ideas and Reasons for Using Stakeholder Analysis

1.2.1 What is a Stakeholder?

Stakeholders are persons, institutions, or groups of people whose opinions and interests affect projects, their outcomes, or their profits (Aaltonen et al., 2008; Aaltonen & Sivonen, 2009; ODA, 1995; Sharp, Finkelstein, & Galal, 1999). Through stakeholder analysis, it is possible to identify and categorize the relationships of these stakeholders and use them for the benefit of projects (Jepsen & Eskerod, 2009).

1.2.2 Why do a Stakeholder Analysis?

Stakeholder analysis involves finding the opinions of stakeholders in relation to the problems that are being addressed in a project in order to make adjustments in a way that satisfies the stakeholders and maintains their interests in the projects (Jepsen & Eskerod, 2009;

ODA, 1995). It also allows managers to identify any conflicts of interests by showing any possible relationships that may exist between the stakeholders, which will allow the stakeholders to identify any possible coalitions (Bryson, 2004; Mitchell, Agle, & Wood, 1997). The preliminary analysis would allow the project managers to identify which stakeholders could be involved at which stages of the project (Fraser et al., 2006).

This type of analysis allows the managers to pinpoint the needs of the stakeholders early on and work to satisfy them. The objective is to maintain happy stakeholders that will support the project every step of the way. Their continuing support will also help to attain support for future projects. Understanding their needs via the use of social media will also offer a better means of maintaining proper communications with the stakeholders. Finally, it will also help anticipate the reactions that stakeholders will show to the project (Baccarini, 1999; Crosby, 1991).

1.2.3 When Should It Be Done?

It is best to perform an analysis at the beginning of a project in order to get an overview of the possible stakeholders and their potential interests (ODA, 1995). This overview allows the managers to form an idea of where their stakeholders' initial opinions toward a project lie and what relationships to investigate. Stakeholder analysis is also especially useful if conducted throughout the life of a project, in order to maintain awareness of stakeholder opinions at various stages of completion (R. Brugha, 2000; Bryson, 2004; Jepsen & Eskerod, 2009; ODA, 1995).

1.3 Review of Stakeholder Analysis Methods

A wide range of stakeholder analysis methods has been presented in the existing literature. These can be classified into methods that are utilized for i) identification of stakeholders and their interest; ii) categorization of stakeholders; and iii) investigation of the relationships among stakeholders (Dougill et al., 2006). Table 1 includes the methods, authors, strengths, and weaknesses. It is possible to utilize some methods for several objectives; Social Network Analysis, for example, is principally used in the investigation of relationships between stakeholders, but also to initially identify and categorize those stakeholders (Reed et al., 2009a).

1.3.1 Methods for Identifying Stakeholders and Their Interests

In order to gain access to the minds of stakeholders, it is first important to identify who these stakeholders are (Currie, Seaton, & Wesley, 2009). To identify stakeholders, it is necessary to first define the problems or the issues for the study. It is difficult to determine which stakeholders should be involved if problems or issues for a project are not identified (Prell et al., 2007). Focus groups, semi-conservative interviews, and snowballing samples are a few of the ways in which analysis is conducted in order to help identify the stakeholders and their interests.

1.3.1.1 Focus Groups

A focus group is a research method that utilizes the collection of data from a group of participants. Focus groups are ideal for measuring participants' views, attitudes, and experiences regarding the topic under discussion. As such, the purpose of a focus group is to use the insights provided by participants to feed into survey findings. Led by a facilitator, participants sharing similar experiences or backgrounds are guided through questions that give them an opportunity to share their views and attitudes (Freitas, Oliveira, Jenkins, & Popjoy, 1998). This method offers a fast and cost-effective means of gathering information and gaining access to the minds and ideas of the stakeholders. However, much like any other method, it offers a source of bias. The information gathered using this method could be problematic due to the fact that its quality is largely dependent on the facilitators and their leadership. Facilitators play an important role in directing the narrative and direction of the groups, thus their performance has a significant impact on the quality of the focus groups and the resulting data attained from them (Freitas et al., 1998; Harrell & Bradley, 2009; Reed et al., 2009b; J. Yang et al., 2011).

1.3.1.2 Semi-Structured Interviews

Semi-structured interviews are formulated questions presented in questionnaire format and administered to respondents. Semi-structured interviews may employ open-ended questions to generate answers in narrative form rather than simple yes/no responses (Harrell & Bradley, 2009; Louise Barriball & While, 1994). Semi-structured interviews that consist of open-ended questions elicit more information from the initial answers given by a respondent. While the questions are usually predetermined, there is also the potential for impromptu questions to be incorporated during the interview (Louise Barriball & While, 1994). However, this process is both costly and time-consuming (Mathern, B., Bellet, T., & Mille, 2010). It involves gathering a large amount of information from many groups and then cross-sectioning that information in order to broaden the gaps in the data. For this reason, it is not a popular method for gathering information. Another problem is in finding individuals who have the proper training to conduct the interviews, assuring confidentiality, and training people to properly analyze the results. A deficiency in any of these areas would result in faulty or unusable data (Harrell & Bradley, 2009; Louise Barriball & While, 1994; Reed et al., 2009a).

1.3.1.3 Snowball Sampling

Snowballing is a type probability sampling method that focuses on referrals from the initial contact person to other potential participants with shared experiences. Snowballing technique is used when the judgment of primary data determines that the necessary research subjects are uncommon and/or difficult to locate within the identified sample areas. The initial connection leads the researcher to further subjects with similar relationships as the contact nominates or recruits other individuals in the same category (Noy, 2008). The biggest advantage of this study type is the ability to conduct a study that would have otherwise been impossible to conduct due to lack of subject availability. The biggest disadvantage is the source of bias, which is primarily determined by the initial contact, and to the limitations in analyzing the results. Specifically, the sample will be biased because all participants are recruited using the original sample of stakeholders (Biernacki & Waldorf, 1981; Reed et al., 2009a; J. Yang et al., 2011).

1.3.2 Methods for Categorizing Stakeholders

The next step after identifying the stakeholders is to categorize them. This is done using analytical categorization or reconstructive categorization (Dryzek & Berejikian, 1993).

1.3.2.1 Analytical Categorizations

Bryson, Cunningham, & Lokkesmoe (2002), Dale & Lane (1994), and ODA (1995) all use the analytical categorization referred to as Strategic Perspectives Analysis, which involves conducting interviews with stakeholders within their various places of employment. In this regard, stakeholders' goals were identified and comparisons were made between different groups, the perception of opportunities, as well as constraints stakeholders had for their goal accomplishments. Within the Strategic Perspectives Analysis method, researchers utilized an iterative method of interviews to identify stakeholders who have common goals. Different methods of the analytical categorizations developed and discussed by various researchers are further described as follows.

1.3.2.1.1 Interest Influence Matrices

De Lopez (2001) and Eden & Ackermann (2013) observe that interest influence is one of the most common analytical methods used by researchers to classify stakeholders as either context-setters, key players, or subjects. Within the influence and interest stakeholder categorization, the developed methods by the researchers play a crucial role in specification of the stakeholders involved in the literature analytical review process (Rowley & Moldoveanu, 2003). For instance, from an analytical perspective, key players are viewed as stakeholders who must be groomed since their interest is high and they can easily influence a certain phenomenon among other stakeholders. On the other hand, researchers also discuss how context-setters can also be highly influential— however, there interest is minimal. With the limited interest of the context-setters, they are likely to cause more harm than good, hence they must be closely supervised. Additionally, subjects are regarded to have high interest, but a significant lack of influence with other stakeholders. Arguably, despite the fact that the definition appears to back the influencing potential of context-setters, most of them lack the influential capacities that can be found in key players (Bryson, 2004; Newcombe, 2003; Reed et al., 2009a).

1.3.2.1.2 Radical Transitiveness

Radical transitiveness involves snowball sampling in order to identify any fringe stakeholders. This method helps to pinpoint issues that would have otherwise been missed and also helps minimize any risks that these issues would have posed to the future of the project. However, this method is time consuming and costly, so it is not one of the more popular methods that companies use to help them categorize their stakeholders (Hart, 2016; Reed et al., 2009a).

1.3.2.2 Reconstructive Categorizations

Reconstructive categorization involves using stakeholder-led stakeholder categorization. While this method is often used, this method involves the stakeholders categorization of each other, which has far more flaws than advantages (Hare & Pahl-Wostl, 2002). Some stakeholders put different stakeholders in the same category and since the entire process is done using the subjective perceptions that these stakeholders have of each other, it is also a very biased method of categorizing stakeholders. As it is possible for various respondents to place different stakeholders in incorrect or inappropriate categories, it is easy for the categories to become meaningless and therefore unusable for analysis (Dale & Lane, 1994; Reed et al., 2009b).

1.3.3 Relationships Methods

Finally, no stakeholder analysis is considered complete without investigating the relationships that exist between the stakeholders. This involves methods such as actor-linkage matrices, social network analysis, and knowledge mapping.

1.3.3.1 Actor-Linkage Matrices

According to Biggs & Matsaert (1999) and ODA (1995), actor-linkage matrices provide a comprehensive way of explaining stakeholder interrelations. The outline of rows and columns where stakeholders are listed is an essential process of defining any formed interrelations. The grid helps categorize identified relationships as a conflict, complementary, or cooperation. The flexibility of this approach also proves that research can still be conducted without the use of computers, as it requires only a pen and paper.

1.3.3.2 Social Network Analysis

Just as in the actor-linkage matrices method, social network analysis utilizes matrices to have the data organized based on rational ties that promote relations among stakeholders. Rather than utilizing the key matrix cell words, social network analysis uses numbers in its representation (Stanley Wasserman & Faust, 1994). The numbers used depend on the following two factors;

- The absence or presence of ties
- The relative tie strength

In this case, each matrix is designed to represent a relationship that is unique, such as friendship, communication, conflict, advice, and trust. Data collection can be done through the use of questionnaires, interviews, and observation (S. Wasserman & Faust, 1994). Hence, social network analysis not only utilizes different kinds of relationships, but also illustrates strengths pertaining to rational ties. In addition, it involves storage of information in a quantitative form that is already summarized, thus facilitating easy analysis. As such, the structure of the network analysis of the stakeholders can assist in the identification of central stakeholders. Such central stakeholders are considered to be significant, since they have good relationships which can hold other participants together (Prell et al., 2007; Stanley Wasserman & Faust, 1994; J. Yang et al., 2011).

1.3.3.3 Knowledge Mapping

This method of knowledge analysis was developed from the charts that were designed for organizational purposes and used as control and planning tools. However, to successfully manage a natural system of resources, which is subjected to a number of changes, feedback, or responses from different societal sectors, more flexible approaches should be designed to promote both communication and learning. Presently, according to recent research carried out by Nissen & Levitt, 2004, it is evident that modern businesses emphasize the relevance of knowledge management, which is where knowledge mapping can be used.

1.3.3.4 The Power/Predictability Matrix

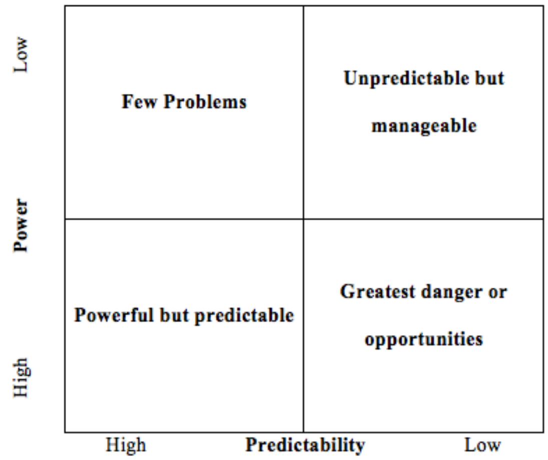

Figure 1. The Power/Predictability Matrix (Newcombe, 2003)

The allocation of stakeholder's groupings into the four zones allows the project managers to evaluate the amount of problems that are encountered by the stakeholders. Resistance from the stakeholders with the greatest danger can be overcome or influenced through decisions that will be acceptable to the zone of stakeholders that are powerful but predictable. Despite the fact that stakeholders in predictable, but manageable and few, problems zones have less authority, this does not indicate that they are insignificant as these stakeholders' support can have a robust effect on powerful stakeholders' attitudes (Hardy, Wickham, & Gretzel, 2013; Newcombe, 2003).

1.3.3.5 Power/Interest Matrix

This type of matrix categorizes stakeholders via their interest level and the authority they hold in the project. All the four zones indicate the kind of relationship that a project manager should require so as to establish and uphold connections with every kind of stakeholder group (Newcombe, 2003; J. Yang et al., 2011).

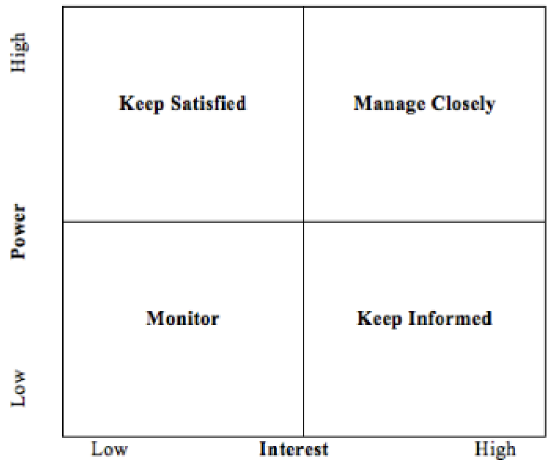

Figure 2. The Power/Interest Matrix (Newcombe, 2003)

Table 1. Stakeholder Analysis Methods - Strengths and Weakness

Methods	Authors	Strength	Weakness
Step 1: In order to gain access to the minds of stakeholders, it is important to identify who these stakeholders are.			
Focus groups	(Harrell & Bradley, 2009; Kitzinger, 1995; Reed et al., 2009a; J. Yang et al., 2011)	○ Fast and cost-effective ○ Gain access to the minds	○ Source of bias ○ Problematic information ○ Big impact on the quality
Semi-structured interviews	(Harrell & Bradley, 2009; Louise Barriball & While, 1994; Reed et al., 2009a)	○ Useful for in-depth insights to stakeholder relationships and to triangulate data collected in focus groups	○ Time-consuming and costly ○ Difficult to reach consensus over stakeholder categories
Snowball sampling	(Biernacki & Waldorf, 1981; Reed et al., 2009a; J. Yang et al., 2011)	○ Being able to conduct a study that would have otherwise been impossible to conduct	○ Source of bias
○ **Step 2:** The next step is to categorize them.			
Interest influence matrices	(Bryson, 2004; Newcombe, 2003; Reed et al., 2009a)	○ Identify the real stakeholders ○ Stakeholder power Effective communication ○ Crucial role in specification of the stakeholders and their priorities	○ Fails to show the actual attitudes of the stakeholder ○ Marginalization of certain groups ○ Very subjective
Radical transitiveness	(Hart, 2016; Reed et al., 2009b)	○ Help find issues that would have been missed ○ Helps minimize any risks posed on the future of the project	○ Time-consuming and hence costly
Stakeholder-led categorization	(Hare & Pahl-Wostl, 2002)	○ Stakeholders categorization each other ○ Perceptions of stakeholders	○ Source of bias ○ Categories become meaningless
○ **Step 3:** No stakeholder analysis is complete without investigating the relationships that exist between the stakeholders.			

Actor linkage matrices	(Biggs & Matsaert, 1999; ODA, 1995)(Biggs & Matsaert, 1999; ODA, 1995)	○ Easy in terms of the resources.	○ The analysis could become very confusing and tedious.
Social network analysis	(Prell et al., 2007; Stanley Wasserman & Faust, 1994; J. Yang et al., 2011)	○ Identifies influential stakeholders and peripheral stakeholders	○ Limited by the ability to analyze the questionnaires
Knowledge mapping	(Nissen & Levitt, 2004)	○ Shows the power level for each stakeholder ○ Presents the knowledge for stakeholders	○ Gaps in knowledge ○ Fail to fulfill the needs of all stakeholders
The Power/Predictability Matrix	(Hardy et al., 2013; Newcombe, 2003)	○ Locates power in the project ○ Better project decisions	○ Subjective ○ Must be performed on a regular basis
The Power/Interest Matrix	(Newcombe, 2003; J. Yang et al., 2011)	○ Discovers real power and interests ○ Improves the process of execution ○ Provides the correct communication	○ Stakeholder selection can be subjective ○ To benefit, it must be performed on regular basis ○ Does not guarantee positive attitude

In the analysis of stakeholders, several methods are used in steps 1, 2, and 3. In step 1, focus groups, semi-structured interviews, and snowball sampling are used. In step 2, analytical categorization including interest influence matrices and radical transitiveness, or reconstructive categorizations, which includes stakeholder-led method, can be used. In step 3, combinations of methods are used, including actor-linkage matrices, social network analysis, knowledge mapping, power/predictability matrix, and power/interest matrix.

Focus groups and snowball sampling introduce bias into the study, while semi-structured interviews are more costly. The interest influence matrices fail to bring out the qualitative aspects of research, such as stakeholders' attitudes toward the projects, besides contributing to the marginalization of certain groups. On the other hand, actor linkage matrices can be tedious and confusing if large datasets have to be analyzed, while social network analysis requires the use of knowledgeable and qualified data. Knowledge mapping does not consider the diversity of stakeholders, while the predictability matrix is challenging when used for the first time. Stakeholder commitment matrix provides limited information and may be counter-productive to research objectives.

1.4 Conclusion

Selection of a particular identification method or procedure will be determined by the context of the project, the stage of the project, and the existing resources. Involvement of the stakeholders is one way of obtaining effective input and contributions in the first stage of analysis. Conway & Lance (2010) also explain that failure to recognize some stakeholders may create bias in the series of phases of the method chosen.

Normally, there is no commonly agreed upon method to both the selection and the usage of the analysis. The researchers rely on the objectives and context of the project, along with clear expectations. An exceptional consideration must be taken to avoid involving stakeholders or their direct representatives in the analytical process, as their inquiry and decisions may be personal rather than data-driven. Podsakoff, MacKenzie, & Podsakoff (2012) explain that in order to reduce potential biases from individuals such as those who represent the will of the people, either stakeholders themselves or specialists need to classify all stakeholders via a particular method. Various biases could potentially be introduced in the way that participants are interviewed or the way the data is handled (Boyce & Neale, 2006).

CHAPTER II – BIAS IN STAKEHOLDER ANALYSIS METHODS

2.1 Introduction

Experienced research specialists are aware that no research program can be 100% free from bias (Levine & Safer, 2002). So, when does this inherent issue become a menace? And how do researchers recognize and prevent the emergence of bias to create the utmost quality and highest value of research? The objective of decreasing bias is not to make everyone become the same, but to ensure that questions are completely reviewed and presented in a manner that enables participants to discover their genuine feelings without bias (Tracy, 2010). The problem of bias is presented in all dimensions of qualitative research and can be derived from the questions, the participants, and the researcher. This is important to understand so as to minimize bias from all three sources and thus conduct superior research.

Dan Ariely (a behavioral economist) explains that humans normally think of themselves as being in the driver's seat, having overall control over decisions they create and the direction their life takes (Marshall & Rossman, 2015). But it is unfortunate that this thinking is connected with human desires regarding how they want to see themselves rather the reality. When an individual holds a one-sided perspective, he or she can be accused of being biased or prejudiced. On a personal level, humans tend to interpret things in a prejudiced manner, primarily founded in their cultural beliefs and values. However, there is another form of bias identified as cognitive bias which all human beings collectively share (Sekaran & Bougie, 2016). Cognitive bias is a human habit in which humans tend to create common decisions on certain things founded on cognitive considerations instead of evidence.

2.2 Bias Types

Bias in stakeholder analysis is defined as a tendency that prevents unprejudiced consideration of that which is under study. In stakeholder analysis, bias arises when systematic errors are present in testing or sampling by researchers choosing or preferring one outcome that suits the hypothesis of the study over others. Bias can occur at any step of stakeholder analysis, including in sample selection, data collection, and data analysis, as well as in publication. In essence, bias in stakeholder analysis is not a dichotomous variable. That is, interpretation of biases is not limited to the simple inquisition as to whether bias is present or absent. Instead, stakeholder analysts must consider the extent to which bias was prevented through the use of proper study implementation and designs. Nonetheless, some degrees of bias are present in every step of stakeholder analysis, as well as in the published research; therefore, stakeholder analysts must consider how bias might affect study conclusions.

Table 2. Bias Types

Source of Bias	Author	Description
Sampling Bias	(Boria, Olson, Goodman, & Anderson, 2014; Panzeri, Senatore, Montemurro, & Petersen, 2007; Signor & Lipps, 1982)	This form of bias happens when a prejudice or stereotype exists in selecting a particular population to sample and thus the respondents do not represent the wider population. Therefore, the findings cannot be generalized to the wider population.
Selection Bias	(Emran, Greene, & Shilpi, 2015)	This prejudice or stereotype is created by choosing data, groups, or persons for evaluation in a manner in which effective randomization is not accomplished, thus contributing to creating a sample that does not represent the wider population targeted to be evaluated. In certain cases, it is identified as the selection effect.
Response Bias	(Holbrook,	It is a habit of an individual to answer questions dishonestly or

	Green, & Krosnick, 2003)	falsely on a survey. For instance, a person may feel pressure to provide answers which are socially acceptable. Or else, it may occur when only certain kinds of people respond after being invited to take part in a specified activity. The group that participates does not therefore represent the wider population.
Performance Bias	(Juni, Altman, & Egger, 2001; Moses, Villate, Binns, Davidson, & Ryan, 2008; Pannucci & Wilkins, 2010)	This form of bias happens when something affects the delivery of interventions or treatments. Normally this happens when the researchers or respondents behave differently since they are an active component of a study. For instance, if a researcher perceives treatment X is more efficient than treatment Y, he/she may pay more attention to the respondents getting treatment X.
Confirmation Bias	(Hernandez & Preston, 2013; Jonas, Schulz-Hardt, Frey, & Thelen, 2001; Knobloch-Westerwick, Johnson, & Westerwick, 2015; Mynatt, Doherty, & Tweney, 1977; Nickerson, 1998; Oswald & Grosjean, 2004)	In research, this is the one of the most extensively identified and most common types of bias. This bias happens when a researcher creates an assumption or a hypothesis and utilizes participants' information to verify or confirm that hypothesis. This happens when a researcher determines and evaluates responses which verify their hypotheses as valid and significant while rejecting evidence which does not validate that hypothesis.
Culture Bias	(Fischer & Derham, 2016)	Beliefs regarding influences and motivations which are founded on cultural points of view (on the perspective of cultural relativity or ethnocentricity) lead to the creation of cultural bias. Ethnocentrism means evaluating another culture only by the standards and values of an individual's own culture. On the other hand, cultural relativism is an idea that a person's actions and beliefs need to be comprehended by others based upon that person's own culture.
Question-Order Bias	(Bard & Weinstein, 2017; Jackson & Greene, 2017)	A question can influence the following series of other questions, thus causing question-order bias. Participants are influenced by ideas and words illustrated in the questions which influence their attitudes, emotions, and thinking on other questions.
Leading Questions and Wording Bias	(Choi & Pak, 2005; Dodd & Bradshaw,	Expounding on a participant's answers puts words in their mouth that they would not otherwise have spoken. While leading questions and wording are not forms of bias

	1980)	themselves, they cause biasness or are the outcome of bias. Researchers normally create that bias since they try to confirm or validate a hypothesis, develop support, or overrate their understanding of the participants.

2.3 Stakeholder Analysis and Bias Types

Many research studies fail to make attempts in differentiating the analysis and identification processes as far as stakeholder analysis and biasness are concerned. In essence, stakeholder analysis is the work of identifying stakeholders who may be affected by any proposed initiative and assessing their participation and interest via research studies. It involves stakeholders, the goals they have, their specific interests, as well as their relationships, functions, and characteristics in understanding their actions and opinions. The researcher could also need to assess which investigator can best fulfill the role of stakeholder representative in the process of identifying stakeholders. Specifically, the kinds of stakeholders who are at least somewhat important to the project's success are analyzed. The following tables will illustrate stakeholder analysis methods and the corresponding bias types.

Table 3. Bias in Each Stakeholder Analysis Method

Bias types / Methods	Sampling Bias	Selection Bias	Performance Bias	Sponsor bias	Confirmation bias	Culture bias	Question-order bias	Leading questions
Focus groups	It essentially occurs when the sample statistics deviate from the reflection of the true estimate of the targeted population. Typically, sampling bias in focus groups focuses on either the ratios or the average, which are the two basic types of statistics. The sampling biases that arise from average derive from imperfect sampling frames, measurement errors, or nonresponse bias. Nonetheless, mathematical statisticians	In most cases, selection bias arises when the participants in the research are volunteers. Essentially, those who chose to participate may share common traits that make them different from the non-participants from the get-go. For instance, a study that needs to assess the program for boosting habits of eating of night shift workers may put up flyers and invite night shift workers to participate. Importantly, those participants	Typically, performance bias arises when researchers have two groups, treatment group and control group, yet he/she does not give equal attention to both of them. The variation in care level results in systematic differences within the two groups that make it difficult to derive a conclusion on the effects of interventions and drugs as compared to level of care. Similar bias may arise when the consequences	Because a person may feel pressure to create answers which are analyst acceptable. Therefore, the biased answer created by the participant is somewhat untrue and sometimes partially true sentiments. The bias essentially influences the skew answers and masks the truth. The untrue answer may be intentional or unintentional especially when the member of the focus group is familiar with the attitudes, identity, and characters of	Confirmation bias is applicable in focus groups when it comes to decision making where participants have tendencies for favors and use information that tend to support pre-existing views on a certain topic in the focus group discussion. This bias happens when a researcher creates an assumption or a hypothesis and utilizes participants' information to verify or confirm that hypothesis. This happens when a researcher	Applicable when the questions in the focus group discussion rise culturally sensitive subjects that the participant would rather not talk about in the research study. Participant may give false feedback to hide the secrets.	Yes because it may depend on the sequence of questions that the investigator uses. Typically, general questions come first before specific questions, then followed by unaided questions before asking the aided ones. Besides, positive questions should start before negative questions, then behavior questions before those that touch on attitude.	Leading question bias is applicable when the investigator tries to confirm validity of the hypothesis in developing support that he/she will document in the final report. In most cases, the investigator will overwrite the responses provided by the members of the group and cooks assumptions that are consistent with the hypothesis in the study.

	consider biases those sources such as sample size (n).	who sign in may vary greatly from those who do not participate. Basically, they may be more health conscious which make them likely to enroll in the program. Good researchers may overcome selection bias through making the study more representatived by incorporating different groups with different characteristics.	are more likely to be derived from the treatment group where the investigators know the identity of groups of each participant. In essence, performance bias explains, in focus groups, when the members of the focus group change their behavior in cases where they realize the groups they have been allocated in the study.	the investigator in the focus group. The respondent may appear to be consistent with their feedback as the previous response usually affects the later sentiments even when the feedback masks the truth. In addition, dominant members in the group take time to vocalize their biased responses that may eventually influence the feedback of others members in the group.	determines and evaluates responses which verify their hypotheses as valid and significant, whereas rejecting evidence which does not validate a hypothesis.			
Semi structured interviews	Typically, the study tends to have high validity and reliability when the number of participants	The bias will most likely occur when the interviewer is not adequately skilled in proper techniques of	It is applicable because the analyst will bias the interviews depending on his thinking. When the	Yes because a person may feel pressure to create answers which are analyst acceptable.	This type of bias is applicable at the time when a researcher determines and evaluates responses	It's applicable when the interviewer touches on cultural sensitive questions that the participant	Question biased bias applies when the investigator distorts the usual sequence of questions	Yes it is applicable when the interviewer uses questions that tend to guide the respondents on

	who took part in the interview is large.	interviewing the study participant.	interviewer has personal assumption in the study, biasness will occur when he/she tries to validate the study despite the fact that the response from the participants deviates from his/her assumptions.		which verify their hypotheses as valid and significant, whereas rejecting evidence which does not validate a hypothesis.	opts to mask the truth by giving biased answers.	with sometimes starting with negative questions instead of positive question that tend to disorient the participant.	the answers that he/she intends to arrive at depending on the hypothesis of the study.
Snowball sampling	It is not applicable because it applies to the specific scope that has specific traits under investigation. It essentially occurs when the sample statistics deviate from the reflection of the true estimate of the targeted population.	It is not applicable as the method is a non-probabilistic technique that applies when the traits under investigation are rare and difficult to find through random selection.	It is not applicable as this method is a respondent driven technique having a mathematical model that tends to gauge the sample in compensating for facts that sample selection was in a non-random way. Therefore, the delivery of the treatment will be to this specific group collected that have the	It is not applicable as the study participants rarely have the knowledge about the attitudes and the mood of the investigator thus the answers that they give does not mask the truth. It is worth to note that the method is non-probabilistic hence the responses provided correspond to	Confirmation bias rarely happens in snowball sampling as the investigator is sure of the traits that he/she is going to examine hence the hypothesis addresses the specific characteristics of the study participant. Notably, the sample under investigation has specific traits that the researcher	It is applicable depending on the cultural difference between the study participant and the investigator. Typically, ideologies regarding influences and motivation which are founded on cultural point of views (on the perspective of cultural relativity or ethnocentricity) lead to the creation of	Yes, as it depends on who is doing the investigation. When the investigator is less qualified in conducting snowball sampling, he/she will end up using questions that can influence a series of other questions, thus causing question-order bias. Participants are clued-up by ideas and words	Yes, depending on how the investigator phrases the questions under the study. Typically, expounding on a participant's answers puts words in their mouth. Whereas leading questions and wording are not forms of bias themselves, they cause biasness or are the outcome of

			targeted traits.	the thoughts of the participant.	wants to study, hence there is no doubt of encountering some odd characteristics that were not anticipated.	cultural bias.	illustrated in the questions which influence their attitudes, emotions, and thinking on other questions.	bias. Researchers normally create that bias since they try to confirm or validate a hypothesis, develop support, or overate their understanding of the participants.
Interest Influence matrices	Comes into play depending on the number of stakeholders selected. When the number of all the stakeholders selected is low, generalizing the research findings tend to be difficult as that would not represent the view of the majority.	Applicable depending on who is doing the selection. Primarily, prejudice or stereotype is created by choosing data, groups, or persons for evaluation in a manner which effective randomization is not accomplished, thus contributing to creating a sample that does not represent the wider population targeted to be	Yes, this kind of bias is applicable when there are factors that affect the delivery of interventions or treatments. In essence, it arises when the investigator as well as the respondents behaves differently since they are components of a study. For instance, in a case study where an investigator views treatment C to be more	Tend to apply when the stakeholders have knowledge about the mood and the attitude of the investigator and thus give biased responses that essentially deviate from the truth of the matter under investigation.	Confirmation bias is applicable in interest influence matrices when it comes to decision making where participant have tendencies to favor and use information that support pre-existing views on a certain topic under investigation.	Cultural bias may be applicable when the questions in the matrix discussion rise culturally sensitive subjects that the participant would rather not talk about in the research study. Participant may give false feedbacks to hide the secrets.	Question biased bias applies when the researchers distort the usual sequence of questions with sometimes starting with negative questions instead of positive question that tend to disorient the stakeholders.	Leading question bias is applicable when investigation tries to confirm validity of the hypothesis in developing support that he/she will document in the final report. In most cases, the investigator will overwrite the responses provided by the members of the group and cooks assumptions that are consistent with the hypothesis

		evaluated. In certain cases, it is identified as the selection effect.	efficient than treatment K, he/she may pay more attention to the participants getting treatment C.					in the study.
Radical transitiveness	Applies when the number of participants in the sample size is extremely small to allow generalization of findings.	It may apply depending on the person who is selecting the stakeholders deciding to ignore randomization, thus the selection biasness.	Comes into play when the investigator deviates from the hypothesis of the study and forces the data collected to fit the problem statement.	It is applicable when the participant in the method develops their own assumptions about the investigator, thus they give unrealistic answers.	This type of bias is applicable at the time when a researcher determines and evaluates responses which verify their hypotheses as valid and significant, whereas rejecting evidence which does not validate a hypothesis.	Yes it is applicable when there is some sort of culturally sensitive questions that may force the participant to give wrong responses in hiding the truth.	Applicable depending on the investigator as he may choose to ask negative questions that are more likely to cause biasness on the responses that the participants give.	Yes, depending on the on how the researchers phrase the study questions.
Stakeholder-led categorization	Applicable depending on the scope size. Notably, generalization becomes difficult when the number of stakeholders selected in the categorization.	Common when there is lack of randomization. Depending on the investigator qualification, biasness will arise when he/she decides to select a certain group that fit the	Yes, it is applicable depending on the interaction of the investigator and the stakeholders. Common when the researcher tries to modify the data collected to fit	Rarely happens especially when the respondents are not familiar with the researcher.	Rare, especially when the researcher is consistent with the stakeholder's perspectives.	Applicable when the stakeholders view some issues in the study as culturally sensitive.	Yes, depending on the qualification of the investigator.	Applicable when the researcher uses phrases that lead the stakeholders to the answers that fit the study.

		study.	the hypothesis.						
Actor linkage matrices	Happens depending on the sample size. In essence, this form of bias happens when a prejudice or stereotype exists in selecting a particular population to sample and thus the respondents do not represent the wider population. Therefore, the findings cannot be generalized to the wider population.	Depends on randomization. The linkage between the stakeholders in the study become realistic when the investigator takes randomization into consideration with different traits of stakeholders critically analyzed.	Depends on the hypothesis of the study. This form of bias happens when something affects the delivery of interventions or treatments. Normally this happens when the researchers or respondents behave differently since they are components of a study.	Knowledge about the investigator affects the outcome. Yes, it is applicable when the stakeholders take a certain assumption about the investigator hence giving biased responses that misguide the research.	This type of bias is applicable at the time when a researcher determines and evaluates responses which verify their hypotheses as valid and significant, whereas rejecting evidence which does not validate a hypothesis.	Culturally sensitive links affect the outcomes. When culturally sensitive aspects are introduced in establishing the linkage matrices, the stakeholders under investigation become somewhat biased.	Applicable depending on the sequence of the questions.	Phrases that lead to answers lead to biasness. Depending on the investigator, he/she might decide to use questions that lead the stakeholders to answer, thus biasness.
Social Network Analysis	Depends on the sample size and tends to apply when the sample size is small.	Yes, depending on who is the investigator.	Yes, depending on the hypothesis. This form of bias happens when something affects the delivery of interventions or treatments. Normally this happens when the researchers	Applicable depending on the attitude and moods of the analyst that affect the answers provided by the participant.	Applicable when the findings are not consistent with the research questions.	Applicable when some issues are culturally sensitive.	Yes, depending on the positivity and negativity of the questions.	Leading phrases affect the outcomes.

			or respondents behave differently since they are components of a study.					
Knowledge Mapping	Applies when the sample size is extremely small to allow generalization of findings.	Common when there is lack of randomization. Therefore, it occurs when only some kind of people respond after being invited to take part in a specified activity. The group that participates does not therefore represent the wider population.	Yes, because answers may be cooked to be consistent with the hypothesis.	Yes, because a person may feel pressure to create answers which are analyst acceptable	Yes, because some information may erode the cultures of some participants.	Applicable when the purpose of the study is inconsistent with the data collected.	Affected by the nurture of the mapping structure.	Yes, as the mapping may decide to use specific phrases that lead to targeted answers.
The Power/Predictability Matrix	Depends on the sample size and tends to apply when the sample size is small.	Yes, because it may depend on who is selecting the stakeholders. The bias will most likely occur when the interviewer is not adequately skilled in proper techniques on how to define the population	Yes, depending on the hypothesis. Applies when some other factors come into place to affect the delivery of interventions or treatments. Normally this happens when the researchers or respondents	Applicable depending on the attitude and moods of the analyst that affect the answers provided by the participant.	Applicable when the findings are not consistent with the research questions.	Applicable when some issues are culturally sensitive. Beliefs regarding influences and motivation which are founded on cultural point of views (on the perspective of cultural	Yes, depending on the positivity and negativity of the questions.	Leading phrases affect the outcomes.

		that would yield a result that can generalize the entire population.	behave differently since they are components of a study.			relativity or ethnocentricity) lead to the creation of cultural bias.		
The Power/Interest Matrix	Comes into play depending on the number of stakeholders selected. Essentially, stereotype exists in selecting a particular population to sample and thus the respondents do not represent the wider population. Therefore, the findings cannot be generalized to the wider population.	Applicable depending on who is doing the selection. Primarily, prejudice or stereotype is created by choosing data that are biased.	Yes, this kind of bias is applicable when there are factors that affect the delivery of interventions or treatments.	Tend to apply when the stakeholders have knowledge about the mood and the attitude of the investigator and thus give biased responses.	Confirmation bias is applicable in interest influence matrices when it comes to decision making.	Cultural bias may be applicable when the questions in the matrix discussion are culturally sensitive.	Question biased bias applies when the researchers distort the usual sequence of questions with sometimes starting with negative questions.	Leading question bias is applicable when investigation tries to confirm validity of the hypothesis.

2.4 Table 3 Summary

Biasness is almost always present in all stakeholder analysis methods. The table above shows the relationship between different stakeholder analysis methods and the type of bias that is applicable to each. Sampling bias is also applicable in all the stakeholder analysis methods. In essence, this form of bias happens when a prejudice or stereotype exists in selecting a particular population to sample and thus the respondents do not represent the wider population. Therefore, the findings cannot be generalized to the wider population. Selection bias is a prejudice or stereotype that is created by choosing data, groups, or persons for evaluation in a manner in which effective randomization is not accomplished, thus contributing to creating a sample that does not represent the wider population targeted to be evaluated. In certain cases, it is identified as the selection effect. Performance bias occurs in many stakeholder analysis methods, except in the snowball method. This form of bias happens when something affects the delivery of research interventions or treatments. Normally this happens when the researchers or respondents behave differently since they are a component of a study. The bias essentially influences or skews answers and masks the truth. The untrue answer may be intentional or unintentional especially, when the member of the focus group is familiar with the attitudes, identity, and characters of the investigator in the focus group. The respondent may appear to be consistent with their feedback, as the earlier response usually affects the later sentiments. Additionally, as dominant members in the group take time to vocalize their biased responses, that may eventually influence the feedback of other members in the group. Sponsor bias occurs in many stakeholder analysis methods because a person may feel pressure to create answers which are analyst-acceptable. Therefore, the feedback provided by the participant may be untrue and sometimes partially true sentiments. This bias also results in skewed answers.

2.5 Conclusion

Bias in stakeholder analysis is defined as a tendency that prevents unprejudiced consideration of the information that is under study. In stakeholder analysis, bias arises in almost all steps, for example, during testing or sampling. Sampling bias is the most common type of bias that is applicable to small group and small population data collection. It essentially occurs when the sample statistics deviate from the reflection of the true estimate of the targeted population.

Increased levels of public engagement are in most cases referred to as central constituent of an efficient analyzing process for infrastructural projects. The engagement of major stakeholders is extensively perceived as the most significant aspect of an efficient outcome and thus it is imperative that researchers work to overcome bias (Grimble, 1998). Stakeholders' engagement from the onset of a project also improves trust and comprehension, in addition to directly supporting the outcome of the project (Reed et al., 2009a).

CHAPTER THREE – STAKEHOLDER PARTICIPATION

3.1 Introduction

Interviewing individual stakeholders and practitioners in a face-to-face setting is often biased by either the way that participants are interviewed or the way the resulting data is handled. One key theme evident in the existing literature is the requirement to replace the 'tool-kit' interviewing technique, which highlights choosing the right 'tools' for the task, with a method that perceives engagement as a process. A more precise way to describe this perception is "direct participation" through both in-person engagement and via online platforms. This technique requires being strengthened by a suitable philosophy, and considering how best to incorporate the key stakeholders at the most suitable times, directly and in a method that enables them to shape their decisions in a fair and efficient manner (Project Management Institute, 2013; Warner, 2006). It is also important to analyze and represent key stakeholders systematically (Aaltonen, 2011; C. H. Yang, Motohashi, & Chen, 2009). Stakeholder analysis may be exercised with the active engagement of the stakeholders themselves, particularly in the case where there is significant documentary proof or where analysts possess personal knowledge about the groups and individuals with a stake in the system being investigated (such as through intervention, issue or organization) (Reed et al., 2009b).

Nonetheless, engagement may be essential, given there is no clarity over the most important issues as far as the investigation is concerned, or given partial knowledge on the population acting as the representatives of all stakeholders (Aaltonen, 2011; Bourne & Walker, 2005). The engagement level in stakeholder analysis may also vary extensively. This may comprise of passive discussion, where stakeholders simply offer information without being directly solicited to do so. The opposite type of engagement, active participation, can result in source bias due to a 2-way information exchange between analysts and stakeholders as equal partners, in a procedure which is intended to enable stakeholders to affect those engaged in conducting the analysis (Beringer, Jonas, & Kock, 2013).

3.1.1 Participatory Approach

While a small number of the claims devised for stakeholder engagement have undergone testing, there is proof that further testing may improve the value of environmental decisions, likely as a result of more thorough information inputs (Fageha & Aibinu, 2013). Nonetheless, the value of decisions made via stakeholder engagement is highly dependent on the method used in analyzing the data. Scarcities in this procedure are in most cases held responsible for the failures leading to disenchantment in stakeholder engagement. In most cases, this has resulted from a focus on the tools of engagement, instead of the procedure considered in utilizing these tools (Fageha & Aibinu, 2013).

By concentrating on engagement as a procedure, this review has established several best practice aspects from the existing literature. A variety of methods have been created to comprehend the basis for stakeholder engagement and may be used in selecting and tailoring techniques to the decision-making framework, taking into consideration objectives, forms of participants, and suitable engagement level. It is also considered that stakeholder engagement requires being supported by a philosophy that underscores empowerment and direct engagement.

3.1.2 Stakeholder Participation

This section offers an assessment of the significance of stakeholder engagement and alliance in the first step of the analysis and planning process to avoid some analysis bias.

3.1.3 Why Participatory?

Through the identification of individuals, groups, and facilities with an interest in the project, the stakeholder analysis holds the likelihood to offer solutions to conflicts by aiding in the identification of those who may be productive in the solution-making process (Clulow, 2005). The approach was initially created and utilized as a research tool, and as such, it has been unable to engage and enable stakeholders to participate in the analysis (Gao & Zhang, 2006). The following are the benefits of stakeholders taking part in the analysis:

- Improved trust in decisions (OECD, 2001; Richards, C., Blackstock, K.L. e Carter, 2004)
- Enhancing project design utilizing local knowledge
- Improved comprehension of projects and issues
- Incorporation of a variety of interests and outlooks
- Optimizing execution of plans and projects
- Public reception of the decisions
- Cultivating and creating social learning

3.1.4 Choice of Participatory Methods

So as to uphold a feasible procedure of stakeholder engagement in the analysis, intended participatory methods must be established (Johansson, 2008). In most cases, they are simply settled upon once the objectives and level of stakeholder engagement have been defined. The present literature identifies a plethora of methods; hence there exists no single standardized technique to select the pertinent participatory method. The preference is dependent on a number of factors, such as:

- Level of engagement
- Classification of stakeholders

3.2 Direct Participation Through Social Media Platforms

The relationship between networking technology and the public sector is a new way of communication (Waters et al., 2009). This connection between social media and the people also offers a new method of communication between the public, stakeholders, and planners (Taylor, Kent, & White, 2001). It becomes an unobstructed pathway between all. The foundation for social networking sites is relationships, which means that direct participation from the stakeholders through social media networks obtain more meaningful results and allows planners to overcome weaknesses effectively. For the purposes of equitability, efficiency, and competence of natural resource management, all the different stakeholders ought to be considered in the management and decision-making processes, particularly in the initial phase of the analysis (Bourne & Walker, 2006). This is key, as stakeholder analysis tries to meet the needs of the stakeholders (Waters et al., 2009).

Networking technology can provide a simple method to additionally support the decision-making process through easy communication. By using social media technology, participation in the decision-making process becomes easier for others and could help reduce some types of bias (Mustajoki, Hämäläinen, & Marttunen, 2004). Since transparency and openness of the planning process is important, technology would allow more people to observe the possibilities. For instance, people with limited mobility or those who are away on a business trip are still able to participle in a meeting (Mustajoki et al., 2004). It can essentially be used to make interactions easier and possible when they previously were not possible (Kirkman & Gibson, 2004; Warkentin, Sayeed, & Hightower, 1997).

Social networking also allows interactions with those both currently involved and others who may not know about the organization. They are able to connect with all stakeholders, allowing more potential participants to be informed about the goals of the organization or project. (Kaplan & Haenlein, 2010).

3.3 Social Media Applications

3.3.1 Facebook

There are approximately 16.5 million users of Facebook every month, and Facebook is among those platforms which are most popular and enable users to share updates, photos, and general news with the people who “like” or follow them.

One important step on Facebook that organizations should start with is building a fan base. This can be done by posting a link to publicize the Facebook page and by adding icons of social media onto the organization’s website. It is also important to post things that will make the audience engage with what other people have posted. Users will comment, "like," click, and even share. By appearing in the timelines of others more frequently when engaging more people, the organization can gather even more followers for their fan base.

Keeping in mind that Facebook is used by many to connect with friends, using the personal network to connect is essential. Naturally fitting into the atmosphere of personal network sharing is necessary for reaching people who are interested in the organization’s posts.

3.3.2 Twitter

Twitter is an easy, concise, and fast-paced way of connecting with the audience. It is a social media/networking site allowing readers to read and send short (i.e., 280 characters) messages which are in real-time, called "Tweets." It is applied in different domains because it is a fast way of disseminating information; the fields that utilize Twitter include smart cities, disaster recovery, military scenario, business, and intelligent transportation. Twitter generates approximately half a million Tweets daily. Some of these Tweets are available through public APIs of Twitter to developers and researchers.

The leading platform among social media forms of "microblogging," Twitter provides a way for brief posts to be broadcasted. Twitter users can use their accounts to post original Tweets and they can also "reTweet" by posting what has already been posted by another user while crediting the originator. Profile pages describe and indicate followers of Twitter users and also give an indication of whom they follow. Choosing to follow a person makes him or her also receive the follower's Tweets. It has often been the case for people to follow and reciprocate for those users who follow them.

3.3.3 Twitter Versus Facebook

First, in comparing them with other channels of social media, the social interaction enabled by Twitter consumes less time and social dynamic, yet its entrepreneurs' network is less relationship-oriented than that of Facebook. It is believed by the entrepreneurs in our study that there is less exposure to the public and more commitment which is personal on Facebook as compared to Twitter.

Engaging in Facebook makes users surrounded by social interaction because it is like an online cocktail party, and in some instances, it makes you be peripherally surrounded by the people that are known to you and also that are part of the social groups in your life. It is not that way on Twitter, which is more akin to someone who stands with a megaphone on the corner of the street saying, "Hey, check out this paper if you have interest in the project!" and I – as a user – may or may not check it out depending on whether I am interested or not; if not, I will just walk right by.

Moreover, Twitter can be used by entrepreneurs in broadcasting blogs and sending messages automatically to their pages on Facebook, making it augment other channels of social media. Finally, Twitter, as a channel of social media, helps in rendering the entrepreneur as central to interaction and hence suits the study of effectuation and also the process which is individualized (Sarasvathy, 2004).

3.4 Using Twitter as a Platform

For all of the aforementioned reasons, I will use Twitter as the platform for my stakeholder analysis. Twitter is dynamic and consumes less time. It also provides a way of broadcasting brief posts and allows for "reTweeting." Twitter is also easy to use and provides a fast way of connecting with audiences and thus can generate information to a large number of viewers.

What distinguishes the platform from other networks is its capability of providing new means of communication. It is also capable of delivering data over multiple channels of delivery to interested users, which is a distinguishing factor of such applications and smaller networks. For example, Twitter users can receive Tweets as text messages on their cell phones.

3.4.1 Advantages of Using Twitter as a Network platform

- Twitter would reach different stakeholders more effectively
- Twitter allows people who are away on business trips to participate
- Allow more people in general to participate
- Allow stakeholders to participate directly
- It can be used to make interactions easier and possible
- Participation in the decision-making process becomes easier

- Communication is open and dynamic
- Online networks can help increase public relations as well as networking with key stakeholders.

3.4.2 Twitter Data Collection and Data Analysis

3.4.2.1 Data Collection

The module is responsible for collecting data downloads from several social platforms. The data is then stored in the database accordingly. The storage is done depending on the type of application and parameters specified by the API call. The data modeling process defines and analyses data demands for the purposes of supporting the application procedure (He, Zha, & Li, 2013). Data is then modeled in various forms to match the application's nature.

3.5.2.2 Twitter API

An interface for searching information on Twitter is called Search API. It allows a search of Tweet contents and specific users. This API has imposed limits which are higher, and thus advantageous for research; it is also independent of the rest of API's limit.

I will rely on API functions that are provided by Twitter when collecting data. I will also gather information which is detailed on the list of users and the users' profiles. Most of my focus will be on a different number of stakeholders and their issues. There is a direction for Twitter relationships, but no methods of gathering the sets of information are available. I can consider using the "public timeline" API method, for example.

The database cache of Twitter API's purpose is simple. It only needs decoupling of your application of Twitter from the Twitter API. The approach was founded to separate the process of how new Tweets are gathered from their layer of presentation that displays data which is displayed quickly on the website.

3.5 Database

Collection of Tweets from Twitter streaming API and distribution of data to tables supporting the Twitter framework is done by the core Twitter databases. Population of the database is done in different steps. We then conduct data analytics in different ways, such as:

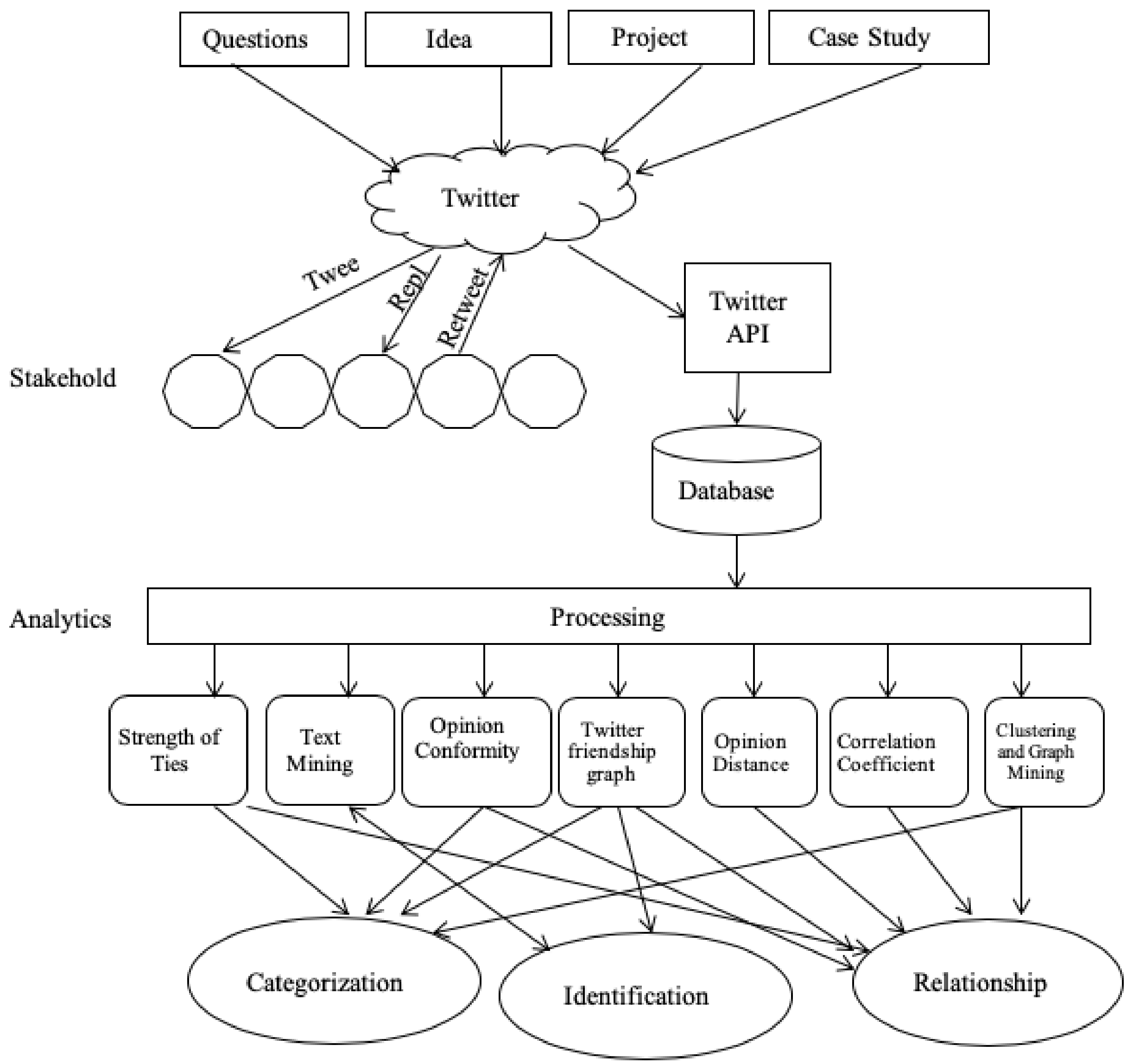

Figure 3. Databases

3.5.1 Social Media Analytics

According to Zeng et al. (2010), it is the duty of social media analytics to provide frameworks and tools for collecting, monitoring, analyzing, summarizing, and visualizing data in social media in a way which is automated, as social media produces massive amounts of data, most of which is usually unstructured.

There has been increasing relevance for social media analytics not only in the government sector and political institutions (Kavanaugh et al.2011) but also for other stakeholders (e.g., Gruhl et al. 2010). A rich platform for providing information for the relationship between management and stakeholders, social media networks are being tapped into for private and commercial stakeholder analysis.

The stakeholder data gathering is implemented in two scripts. The first script collects Tweets, user profiles and simple statistics. The second stakeholder script collects social network relations, i.e. lists of friends and followers of a given set of users.

With the increased use of the internet, there has been an avid interest in social media and how it affects influence. One of the articles on this area highlights online social networks (OSNs) as platforms on which ample data and complex ties among various users come together (Riquelme & González-Cantergiani, 2016). This makes social network analysis quite useful in various fields such as information dissemination, viral marketing, and customer relationship management. Despite the millions of users on these widely available social media platforms, only a handful are genuine influencers. Another article defines influencers as a group of people with a large following who consume and spread the content of the influencer further (Anger & Kittl, 2011). As a result, the reach and effectiveness of these influencers increases the further their content is spread. Previous studies on blogs report that the most influential bloggers were not necessarily the most active bloggers (Agarwal, Liu, Tang, & Yu, 2008). On Twitter, reTweets and mentions correlated well between themselves, and there was an observed correlation with the number of followers of the users. Following this observation, it was postulated that the number of followers represent the level of influence accurately. The use of topic-sensitive PageRank has also been suggested by a different study (Weng, Lim, Jiang, & He, 2010). This proposed measure is based on an observed high reciprocity among followers in the study set. This was attributed to homophily. Other studies have contradicted these findings, as reciprocity has been found to be low on Twitter in general.

3.5.2 Text Analytics

Text analytics is known as text mining and refers to the techniques of extracting information from textual context. Converting large volumes of text generated into meaningful summaries supporting decision making which is evidence-based is enabled by text analytics. For instance, text analytics can be used in extracting information from financial news to predict activity in the stock market (Chung, 2014). Text mining can also be referred to as gathering data from large collections. Additionally, this process automatically identifies the unique patterns of textual information. On a broader perspective, both text and data mining are related. Numerous applications, especially question answering systems, rely on text mining.

3.5.3 Sentiment Analysis (Opinion Mining)

Sentiment analysis (opinion mining) techniques are used in analyzing text which is opinionated, containing opinions of the people towards entities such as events, individuals, organizations, and products. Capturing of data about sentiments of customers leading to sentiment analysis proliferation is increasingly being done by businesses (Liu, 2012). The major application areas of sentiment analysis include finance, marketing, social, and political sciences. Further division of sentiment analysis creates three sub-groups, namely; aspect-based, sentence-level, and document-level. Techniques for document-level analysis help in determining whether the whole document expresses a positive or negative sentiment.

3.5.4 Twitter Users Network

The Twitter user network portrays linked user accounts based on their relatedness. For example, a basic system highlights users who have been mentioned or made replies to the other users' Tweets. In this case, the combination of friends, followers, and the basic network facilitates the capturing of information relating to the users and those they interact with (Aggarwal, 2011). Also, the unchecking option that imports information from the user one is interested in increasing the aspect of information capture. This information includes data relating to followers and replies to mentions.

3.5.5 Tweets Mentioning Synbio

The Gephi program is significant in analyzing and drawing networks. It includes built-in tools that conduct data clustering and analysis. Specifically, the Force Atlas design is utilized in restructuring and resizing the network's nodes. This task is undertaken depending on the number of available nodes. The outcome, results in a network graph that showcases the people receiving many mentions and the ones whose Tweets have synbio. The synthetic, synbio, and biology emerged on one group (Aggarwal, 2011). For efficiency purposes, I removed the nodes from the Gephi, thereby managing to fine-tune the entire graph.

3.5.6 Graph Mining and Clustering

Cluster analysis is usually done through either automatic or semi-automatic procedures. High quantities of data are used to create data records. Also, the detection of anomalies takes place in order to identify and eliminate all the events that are contrary to the expected outcome.

3.6 Conclusion

Social networking sites can create new opportunities for connections with stakeholders and the expression of ideas. It can be used for the flow of information, involving both needs and expectations of stakeholders, as well as finding places that need improvement (Kaplan & Haenlein, 2010). There are countless opportunities, as technology is always changing. Social media is, as such, a powerful tool for organizations to interact with stakeholders and the public in the future of advancing technology. Networking technology specifically is key to keeping social media connections.

So as to come up with a highly efficient and suitable participatory approach, it is also important to have the stakeholder take part directly in the analysis. Utilizing the types of best practice lessons resulting from the present literature, these analyses require working with stakeholders directly to have a systematic evaluation of participatory technique against standards acquired from both theory and stakeholders. However, with the increased use of the internet, there has been avid interest in social media and how it affects influence.

CHAPTER IV - METHODOLOGY

This chapter focuses on the proposed hypothesis, the aims, and the step-by-step processes that were followed in order to determine a robust research conclusion. The ultimate aim of research design methods is to accumulate the right information, and then analyze that information effectively (Lewis, 2015; van den Akker, 1999). All of this is done in order to help answer the research questions and shed more light on the topic being investigated.

The more similar a population is socially the more likely they are to conform. Considering that people often compensate for the fear of social rejection by imitating people around them, individuals are more likely to conform to those that are similar to themselves (Giles & Oxford, 1970); (Lacey Ganser, 2006). Interestingly, it has been observed that when participants are given prior information that a majority of the in-group agree with a stereotype, they are more likely to follow suite (Castelli, Arcuri, & Zogmaister, 2003).

4.1 Objectives

This research is aimed at evaluating the biases that arise from the use of social media, specifically Twitter, as the primary tool for carrying out a stakeholder analysis. Various biases could potentially be introduced through the way that participants are interviewed or the way the data is handled (Boyce & Neale, 2006). The use of a social media platform to carry out the early stakeholder analysis could help eliminate other stakeholder analysis methods. However, this novel method could potentially introduce biases of its own, and this is what the current research aims to evaluate.

In social media, there are many different stakeholders who are not related to each other. As a result, users cannot directly influence each other's opinions, since there are no connections between those stakeholders. Therefore, different companies use social media to get their audience's opinion about certain projects or issues (Culnan, McHugh, & Zubillaga, 2010). Using social media to carry out stakeholder analysis has become more common in the process of collecting data about public opinion. However, social media's platforms can build relationships between users in the social media platforms (Baird & Parasnis, 2011). Social media offers some features, like "follow," which allow users to follow different accounts on social media based on their interests (Kietzmann, Hermkens, McCarthy, & Silvestre, 2011). This kind of relationship can build some similarity between users. For example, if user A and user B both follow "parenting tips" accounts, we may expect that these two users are both parents. If user C follows college admission pages, we expect a different type of user, such as a student. We can use this similarity network when trying to predict those users' response tweets. Focusing specifically on Twitter as one of the social media platforms, there are similarities between users on Twitter that divided them into groups based on their interests (Kietzmann et al., 2011). As such, we can demarcate groups depending on the subjects that they are interested in, based on the accounts that they follow or unfollow. This similarity may affect user opinions and make them more likely to give the same opinions as their network connections. As a result, the more network similarity there is, the more likely there is to be a similar answer from each member within that network. Thus, similarity results in conformity between users on the social media platforms based on their network connections. Conformity can be quantified in terms of scale, based on how similar answers are. As a result, the opinions on social media can be directly influenced by their similarity within interest groups. That causes a source of bias for the social media stakeholder analysis method because people give opinions based on their network connection or their similarity rather than individual perspective. Therefore, the first research question is: does social media network similarity introduce any bias into stakeholder analysis with regards to conformity? To answer this question, I will use the correlation between this similarity and the conformity between pairs of Twitter users.

When we compare Twitter responses about a topic between different users, we may find a positive or negative relationship between responses. A positive relationship would indicate that users have the same opinion, whereas the negative relationship indicates that users have different opinions. This relationship influence can be either positive or negative, depending on the topic. However, it seems that people are more likely to be disagreeable than agreeable on social media. On one hand, conformity is useful for seeing if people are influenced by each other, but to determine if they are more disagreeable than agreeable, we have to consider another basis for comparison between users. Looking at their average answers can provide a good metric to see which side carries more influence. Calculating the average between opinions will preserve the sign of these values. I will then identify the correlation between conformity and the average answers to respond to the research question: do social media platforms influence users to agree or disagree with each other?

Since we may have a correlation between network similarity and conformity, the opinion of the users can be obtained by measuring similarity. It may be that there is no similarity, or weak similarity between users, in which case the user opinions can be either very similar or totally different. Opinion does not appear to be affected by moderate similarity or weak similarity, but once that similarity becomes strong, the opinion may be affected more. As such, there is a point when the similarity starts to become strong and where opinion starts to become affected. When the stakeholder manager obtains answers from the audience or the users, there is a high chance of obtaining answers that are similar after users reach that point of strong similarity. We can use this information to predict responses from users who are similar to users who are unresponsive for the stakeholder manager. This point can answer the research question: does network similarity allow a stakeholder manager to predict the opinions of other stakeholders?

In the research, we examine networks of stakeholder communication on Twitter through Tweets and Retweets about new infrastructure projects related to the stakeholder analysis. We provide evidence that these social network structures affect individual opinion through network similarity, and provide statistical evidence in support of this hypothesis. From all of these objectives, I will answer the following research questions:

1- Does social media network similarity introduce any bias into stakeholder analysis with regards to conformity?

- 1.1-How can we measure similarity between two social media profiles that are not connected?
- 1.2- Do social media platforms influence users to agree or disagree with each other?

2- Does network similarity allow a stakeholder manager to predict the opinions of other stakeholders?

4.2 Hypothesis

The alternative and null hypotheses of this research are:

Hypotheses 1

H1: The network conformity of opinions is correlated with the similarity during the stakeholder analysis process in social media.

H0: The network conformity of opinions is not correlated with the similarity during the stakeholder analysis process in social media.

Hypotheses 2

H1: The network conformity is correlated with the average of opinions during the stakeholder analysis process in social media.

H0: The network conformity is not correlated with the average of opinions during the stakeholder analysis process in social media.

A- Assessing the similarity among stakeholders between every two stakeholders.

B- Assessing the conformity of opinions among stakeholders between every two stakeholders.

This can be done via three steps:

- Evaluate the opinion for each stakeholder in each question and then the total.

- Evaluate conformity in opinions by the distance between opinions.

- Evaluate the average opinion among stakeholders between every two stakeholders.

D. Assessing the interplay between similarity and conformity of opinions among stakeholders.

Use the coefficient of correlation to determine the correlation between strength ties and opinion conformity between individual stakeholders.

4.3 Extracting Topics-Specific Twitter Responses

Four Tweets from a company about an infrastructural project were selected from the thousands of different Tweets based on infrastructure projects in order to assess the similarities of user opinion on particular projects (Table 4). These companies included transport companies running new projects. The selected Tweets were those that contributed the most to the stakeholder analysis. Tweets of various organizations that were aimed at attracting users to offer their opinions on new projects were sought out on Twitter. To define a set of users, four different sources from the organization's Twitter handle were used as an input for screening and coding.

Table 4. Twitter Cases

Case	**Organization**	**Tweet**	**Number of Replies**
1	@MTA	"Look for our 10 new electric buses we're testing in the @MTA's fleet. It's another step towards a cleaner, greener future for New York"	47
2	@NYCTSubway	"Today we released a comprehensive plan to modernize all aspects of NYC's transit system and transform everything we do. Read the #fastforwardny plan here:"	60
3	@wmata	"Exciting news! With more than 500 7000-series railcars delivered, Metro is imagining the NEXT generation of railcars, designed using customer feedback. New 8000-series railcars will replace the 2000/3000 series, which will be 40 years old and due for retirement in 2024. #wmata"	55
4	@NYCTSubway	"A very special thank you to all who attended our R211 prototype design open house. If you visited us but didn't have the chance to give feedback, please leave your comments here on our feed using #R211 by 6pm, December 11. For info on the new cars, visit: http://web.mta.info/nyct/R211OpenHouseFeedback.html …"	27

It was deemed necessary to collect at least 25 responses; therefore, 25 responses were collected and the responses were classified as "strongly disagree," "disagree," "neutral," and "agree." The data was used to derive a set of stakeholders via the Python coding process. The main aim and intent of this phase of the research was to find a pool of users. The initial source of data for the stakeholder analysis was Tweets from users who replied to the original case. From this phase, I printed out all users and their Tweets. Each user was sorted with his Tweets in an Excel file for each case. All Twitter responses are from unique users.

4.4 Quantifying Twitter Responses

There are many different analysis methods to determining whether the Tweet is classified as positive, negative, or neutral; some examples include sentiment analysis and emotional mining (Bing Liu, 2010). To classify Twitter users' opinions in the research, I invited public participants to assist me in this phase. The data collection in this phase involved collection from 22 participants who were part of the Likert scale classification survey. A significant amount of raw, unanalyzed data was amassed from this stage. Excel was then used to do the analysis of the collected data and a coding procedure was used to change answers to numbers in order to allow for quantitative analysis. After the initial data collection from Twitter was conducted, the 22 participants were invited to help in the classification of each Tweet.

Table 5. Participant Process

Data Collection	**Description**	**Interaction with participants**
Purpose	Survey to help in classifying Twitter users' opinions	Using Likert Item Key to find the agreement level for each case from all users. The Likert scale, or the Likert-type scale, was used to classify and categorize the Twitter user replies. The scale consists of the following categories: Not related, strongly agree, agree, neutral, disagree, and strongly disagree. The Likert scale is a commonly used scale in the area of research and it is so widely used in questionnaires that the name is often used interchangeably with rating scale despite there being other types of rating scales.
Method	Focus group	1- Invitation was posted in the university announcements to invite individuals from different domains to participate in the research endeavor. 2- Catered lunch and a small gift were offered. 3- Reminder email sent via e-mail to all participants.

		4- A five minute presentation was provided to let all participants know what was expected. 5- Participants took all replies for the four cases for approximately 45 minutes. 6- Check box was used to mark responses for each participant on the Tweets instrument. There is no identifying information that can link the participant to their response. 7- The results of the data analysis were anonymous, without traceability to any participant.

I translated classification for all participants to the following Likert scale:

Table 6. Likert Scale (Matell & Jacoby, 1972)

Opinion	**Value**
Strongly Disagree	1
Disagree	2
Neutral	3
Agree	4

I then transcribed all participant surveys into an Excel file for comparison purposes. Scatter plots were used to look for a relationship between the 22 participant classifications for each Tweet. The scatter plots, which have data points on the horizontal and vertical axes, illustrate how participants classify the Tweet and where most of the classifications fall along the scale. Drawing a scatter diagram shows where the most answers fall in order to show the relationship between the 22 participants' classifications for each Tweet. To have better results, I used the aggregation function by calculating the participants' classification average for all participants in the same area from the scatter chart.

I filtered Tweets that were not related to the cases and I used a code to denote those Tweets as not related. Considering that Twitter allows access to a wide range of replies and answers, users with replies which were deemed unrelated and unclear were excluded. The selected 22 participants were used to remove users whose reply was classified as 'not related' in order to avoid bias on the part of the researcher. With 22 participants and six different categories, an average of four participants per category is derived [22/6 = 3.67]. Therefore, if more than four participants deemed a Tweet not related, then this Tweet was excluded from the study.

4.5 Data Processing

The first step in this phase was to find all possible pair combinations between users in each case. There are many ways to do that, including programming or online platform usage. I used Python code (Appendix E) to help me to automate this step, in order to find all relevant information and print it out to an Excel file. The reasoning behind this step is to use it to test the following variables for comparing between users: conformity, similarity measures, and average of opinion.

4.5.1 Conformity

Conformity is a type of social influence that can involving a change in belief or behavior in order to fit in with a group (Cialdini & Goldstein, 2004). If there is a pair of users sharing the same opinion, then both users fit within a group – that could be an example of conformity, if one of the users initially held a different opinion, but shifted views in order to join the group. On the other hand, if we have a pair of users whose answers are totally different, then we can say that they are not agreeable to each other and they are not within the same group; this, by contrast, represents nonconformity. To come up with a definition for this step, I used the distance dimension to find the distance between stakeholders' answers in each case. Since we are using scales from 1 to 5 (the Likert scale provided), I will measure the distance between users' answers by the equation below, where 0 (the smallest number) means that they are sharing the same opinion, whereas 4 (the biggest number) means that they are totally different, and no conformity shows here from using the distance dimension formula for all four cases between all possible relation stakeholders. As in the following (Lambers, 2009):

$$d = \sqrt{(x_2 - x_1)^2} = |x_2 - x_1|$$

4.5.2 Similarity Measures

One of the research questions asked is: 1.1 how can similarity between two social media profiles that are not connected be measured? By using Tanimoto measure (Lourenço, Lobo, & Bação, 2004), I developed this measure of calculating the similarity between pairs based on two unordered sets, A and B. The similarities between A and B can be measured by the ratio of common elements to all the different elements. I find that this applies very well to the Twitter platform, since many accounts can be shared between different users. This sharing is not just coincidence – this method also compares the different elements. In this step, I am measuring the network similarity between all pairs in each case.

The other reason to use this method is to determine the correlation between similarity and conformity, in order to identify whether social media platforms can influence opinion. Twitter is a social media platform with a number of famous accounts that enable users to have network connections with each other. A similarity can be said to exist when two people, persons A and B, follow a mutual account. When users A and B both follow certain accounts and abstain from following other accounts, it means they have certain interests in common and these similarities link the users together. As part of the research, the proportion of similarities between stakeholders were measured; stronger similarity between people is associated with a higher influence of opinion among those people.

Most popular Twitter accounts act as a base to find the similarity between users. These accounts divide users based on their thoughts, affiliations, and interests (Kaplan & Haenlein, 2010). Thus, they create new kinds of bonds between Twitter users. For instance, Democrats could follow mutual accounts that support their views and opinions – such as Hillary Clinton, Bernie Sanders, and Barack Obama – while Republicans might do the same – following people like Mitt Romney, Donald Trump, and Marco Rubio. Additionally, newspapers might classify users based on their subscribing to newspapers such as the *New York Times* and the *Washington Post*.

Furthermore, some of these accounts' interests, such as sports, music, or politics, could create similarity among users – for example, conservative users are generally more in favor of coal mining, natural gas drilling, and construction of nuclear reactors, whereas progressive users are usually more likely to support wind, solar, and geothermal energy options. The opinion of the user depends on the group they belong to. Twitter accounts act as a source of information that might be a common factor among different users. Thus, they have been used to determine the amount of similarity between the users.

The Tanimoto measure (Lourenço et al., 2004) was the measure that was used to find the similarity between the selected users on Twitter. It is also referred to as Jaccard (Suphakit Niwattanakul*, Jatsada Singthongchai, 2013). The Jaccard coefficient measures similarity between sample sets, and that can be defined as the size of the intersection divided by the size of the union of the sample sets:

$$S_T(A,B)=\frac{n(A\cap B)}{n(A\cup B)}=\frac{n(A\cap B)}{n(A)+n(B)-n(A\cap B)},$$

The value of the coefficients ranges from +1 which shows the highest similarity to 0 which shows no similarity.

4.5.3 Average of Opinion

Another research question: is the influence of the user's social network dependent on whether users agree or disagree? The average opinion variable, which is essentially the average opinion, tells us the average of opinion for two different users in reference to the same question (Vicente, Martins, & Caticha, 2009). In contrast, the previously-discussed opinion conformity variable tells us how similar two answers are that are given by two different users to the same question – effectively, it represents a "disagreement level" between two users: the higher this variable is, the more disagreement there is between the two users. There is a negative correlation between opinion conformity and average of opinion absolute opinion, presumably because if two people are in disagreement, their disagreement is equally strong (since negative emotions are usually stronger than positive ones), but if two people are in agreement, then they have a wider range of agreement. The formula for calculating average opinion is:

$$d = \frac{x_1 + x_2}{2}$$

4.6 Examples

In these ten example cases shown below, four users give their opinions.

The Likert Item Key is used to find the user's opinion for each case:

Table 7. User Opinion

	Case1	Case2	Case3	Case4	Case5	Case6	Case7	Case8	Case9	Case10
A	5	3	4	1	5	3	5	2	4	1
B	2	5	2	4	1	5	3	4	4	5
C	2	1	1	4	3	5	1	4	2	4

D	1	3	5	1	5	3	4	1	2	3

Calculating distant dimension between nodes to find conformity:

Table 8. Distance Dimension (Conformity)

	X1 - X2										Distance dimension
	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9	Case 10	Total
A-B	2	-2	2	-3	4	-2	2	-2	0	-4	8.06
A-C	3	2	3	-3	2	-2	4	-2	2	-3	8.49
A-D	4	0	-1	0	0	0	1	1	2	-2	5.20
B-C	1	4	1	0	-2	0	2	0	2	1	5.57
B-D	2	2	-3	3	-4	2	-1	3	2	2	8
C-D	1	-2	-4	3	-2	2	-3	3	0	1	7.55

Calculating Pearson's Correlation "Individual Influence":

I used a small sample size of 4 only for the application. My final sample size will be more than that, because the correlation hypothesis test is only valid for n greater or equal to 30. For two data sets, such as in my study, where the case strength tie is 'x' and opinion conformity is y, the correlation can be calculated as shown (Pierce, 2017):

I. The mean of the case strength ties (x) is calculated along with the opinion conformity values (y)

II. The mean of each of the datasets is subtracted from each value in the dataset

III. The subtracted means of 'x' is termed as 'a' while the subtracted means of 'y' is called b

IV. The values ‘ab,’ ‘a2,’ and ‘b2’ are calculated for each value and then summed up

V. The total of ab is divided by the square root of the sum of a2 multiplied by the sum of b2

Table 9. Calculating Pearson’s Correlation

	Similarity	Opinion conformity	a	b	a*b	a^2	b^2
AB	0	8.06	-1	0.91	-0.91	1	0.83
AC	1.5	8.49	0.5	1.34	0.67	0.25	1.8
AD	2.5	5.20	1.5	-1.95	-2.93	2.25	3.8
BC	0.5	5.57	-0.5	-1.58	0.79	0.25	2.5
BD	0	8	-1	0.85	-0.85	1	0.72
CD	1.5	7.55	0.5	0.4	0.2	0.25	0.16
	1	7.15			3.03	5	9.81

From the table: r = 0.43

There is no evidence of correlation between the strength of ties and conformity of opinion (Appendix B).

4.7 Computational Analysis “Python Coding”

Python code is primarily used to automate the process of calculating similarity. Compared to manual analysis, this will help minimize errors. Python, a popular choice among researchers for analysis and coding, was the chosen language for this study in order to manage and automate the vast amount of data that was generated. The use of Python through the Twitter API required prior approval from Twitter, which was attained along with a key and a token. (Raschka, 2015).

In order to find similarities between users, the Tanimioto formula was used (Suphakit Niwattanakul*, Jatsada Singthongchai, 2013). Then, this formula was written in deep detail within the Python code, along with all the requirements for the code to run. The similarity requires steps to be taken before we can run this formula in Python, which are (Appendix B and C):

1. Print all following users for each user (in other words, all accounts that a user follows)
2. Provide the list of famous accounts for the test
3. Used code to remove any following account not in the famous accounts list
4. Print all accounts from the famous account for each user
5. Count the number of Twitter accounts that are shared between every two users
6. Count the total number of accounts across both users (shared and un-shared)
7. Divide the number of shared members (5) by the total number of members (6)
8. Multiply the number found in (7) by 100.

I excluded any users who follow very few famous accounts; depending on similarity between users, I specifically excluded users based on a minimum-threshold criterion where users who follow three or fewer famous accounts were excluded. This was possible to assess because the Python code can find how many accounts each user follows and print it. Therefore, the threshold for the minimum number of famous accounts to be followed was three. There are other reasons to remove users from the data pool - such as the deletion, closure, or protection of the account.

4.8 Analysis Phases

4.8.1 Correlation Analysis

Two research questions will be answered in the analysis research question: 1. does social media network similarity introduce any bias in stakeholder analysis in the side of conformity? And research question 2. does the social media platform influence users to agree or disagree with each other? We specifically need to know what the bias factors are that can occur in social media participatory stakeholder analysis. I used correlation analysis to analyze and compare variables to find whether the similarity and conformity were correlated or not (Cohen, Cohen, West, & Aiken, 2014). For the correlation analysis, I have three things to analyze: the dependent variable (conformity) and two independent variables (similarity and average of opinion).

In order to test my hypotheses of a negative correlation between similarity and conformity, correlation coefficients were calculated in all five cases for all users.

The formulae to perform this correlation calculation is shown below (Mukaka, 2012):

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\,(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Σ is the summation symbol – it indicates that we are summing all x/y values

$(x_i - \bar{x})$ is each x-value minus the mean of x (called "A" above)

$(y_i - \bar{y})$ is each y-value minus the mean of y (called "B" above)

4.8.2 Probability Analysis (Bayesian Probability)

To answer research question: 2. does the network similarity give a chance for the stakeholder manager to predict opinions of other stakeholders? I used Bayesian probability to find what is the probability to know the conformity if we know similarity. Bayes' theorem is a rule in probability and statistical theory that calculates an event's probability based on related conditions or events (Cheeseman, 1983). Scientists and researchers may use these calculations to determine the likelihood of various outcomes (Zhaxybayeva & Gogarten, 2002). As such, I can also use these calculations to determine the point of similarity at which the chance to predict stakeholder opinion becomes possible. The Bayesian probability is calculated using the standard Bayesian conditional probability formula (Jaeger, 1997):

$$P(B|A) = \frac{P(A|B) * P(B)}{P(A|B) * P(B) + P(A|B^{-}) * P(B^{-})}$$

The Bayesian probability formula shows the cumulative probabilities of answers for a given value of similarity – in other words, what is the probability that a given conformity will be above the threshold (1 or 1.5), given that the network connection is equal to or less than a given point (starting from 0 to the highest similarity score)?

4.9 Conclusion

I designed this analysis to assess and compare stakeholder opinion through Twitter responses. Stakeholders in a social network could choose to align their opinion to other, similar stakeholders – or, on the other hand, they could take the opposite opinion of those other stakeholders. Similarity will indicate conformity within the sub-network or group as well. The use of a social media platform to carry out early stakeholder analysis could help eliminate some biases, as it is a direct participatory method, but at the same time it has its own biases. Using social media as a method for performing stakeholder analysis could potentially introduce some new types of biases, and this is what the project aims to evaluate. There are many other factors to consider, such as individual differences, characteristics of the situation, self-esteem, location, or transportation, among others (Campbell,1990 ; Cherry, 2017; Cherry, 2017; Codol, 1975). However, since social media is one of the factors and the correlation could be significant, then stakeholder managers should consider using it to predict the opinions of other users.

CHAPTER V – DATA ANALYSIS AND RESULTS

This document shows the results of my research and has been divided into five sections. In each section, I provide all of the data and the analysis, including the results and the conclusion. Each case goes through the three-analysis phases: the two correlation coefficients and the probability analysis.

5.1 Case One

Table 10. Case One Summery

Text	Company	Date	Replies	All Users	Users Excluded	Users for the Analysis
"Look for our 10 new electric buses we're testing in the @MTA's fleet. It's another step towards a cleaner, greener future for New York"	MTA	April 2017	47	32	18	14

The first case in Table 10 was about public transportation in New York City. Business is booming in the renewable energy sector. As solar and wind power become ever more affordable, electric transit options become more and more competitive with traditional, gas-fueled alternatives. This Tweet was originally made by the New York governor, but it was retweeted by Metropolitan Transportation Authority (MTA) in April 2017. As an infrastructural project, MTA wanted to create new electric buses or run existing ones using clean and green power instead of the old-fashioned buses that run on gas. MTA mentions in their Tweet that they want a "clean, green future for New York." About 50 Twitter users replied to this tweet to give their opinion; some of these users support and agree with this project, and others disagree and prefer to not have this new green system. Different actions and replies were generated by this Tweet, and some of these reply Tweets are not directly related – for example, people are sarcastic, humorous, rude, etc., without contributing to a productive discussion.

Table 11. User Opinions

Users	Opinion
B	4.45
C	1.45
D	1.75
F	4.25
H	4.24
J	4.18
K	3.00
L	1.73
P	3.29
T	3.50
U	4.46
Y	2.61
Z	3.09
A2	3.59
B2	4.27
C2	3.11
E2	3.05
F2	3.32
Total pairs	**117**

5.1.1 Correlation Analysis (Conformity and Similarity)

The table below shows the level of similarity and conformity for all combinations of any two selected Twitter users in our final dataset, regarding the responses of the previously mentioned Twitter users to the Case 1 Tweet. There are 117 total pairs of users (Table 11). The below table contains a few basic descriptive analyses regarding the data (Table 12).

Table 12. Similarity and Conformity

Users	Similarity	Conformity
J,U	0.06	0.28
J,Z	0.05	1.09
J,C2	0.02	1.07
J,A2	0.13	0.59
J,T	0.05	0.68
J,L	0.05	2.45
J,B2	0.05	0.08
J,P	0.04	0.90
J,K	0.05	1.18
J,Y	0.02	1.57
J,F2	0.05	0.87
U,Z	0.13	1.37
U,C2	0.04	1.35
U,A2	0.05	0.87
U,B2	0.02	0.19
U,E2	0.02	1.41
U,Y	0.04	1.85
U,F2	0.02	1.15

D,U	0.02	2.71
D,T	0.09	1.75
D,L	0.03	0.02
D,F	0.02	2.50
D,H	0.07	2.49
Z,C2	0.07	0.02
Z,A2	0.06	0.50
Z,B2	0.07	1.18
Z,E2	0.03	0.04
Z,F2	0.11	0.22
A2,B2	0.06	0.68
A2,F2	0.13	0.27
T,U	0.06	0.96
T,Z	0.03	0.41
T,A2	0.12	0.09
T,B2	0.11	0.77
T,E2	0.10	0.45
T,Y	0.20	0.89
T,F2	0.22	0.18
L,U	0.14	2.73
L,Z	0.05	1.36
L,C2	0.01	1.38
L,A2	0.02	1.86
L,T	0.01	1.77
L,B2	0.01	2.54
L,P	0.03	1.56
L,E2	0.01	1.33
L,Y	0.01	0.88
B2,C2	0.14	1.16
B2,F2	0.29	0.95
P,U	0.11	1.18
P,Z	0.03	0.19
P,A2	0.04	0.30
P,T	0.11	0.21
P,B2	0.06	0.98
P,E2	0.05	0.23
P,Y	0.05	0.67
P,F2	0.11	0.03
K,U	0.06	1.46
K,Z	0.07	0.09
K,C2	0.13	0.11

K,A2	0.06	0.59
K,T	0.22	0.50
K,L	0.03	1.27
K,B2	0.13	1.27
K,P	0.11	0.29
K,E2	0.11	0.05
K,Y	0.10	0.39
K,F2	0.11	0.32
C,J	0.13	2.73
C,U	0.08	3.01
C,Z	0.05	1.64
C,C2	0.03	1.66
C,A2	0.14	2.13
C,T	0.08	2.05
C,L	0.06	0.27
C,B2	0.05	2.81
C,P	0.08	1.83
C,K	0.02	1.55
C,F	0.11	2.80
C,H	0.07	2.78
C,Y	0.05	1.16
C,F2	0.05	1.86
F,J	0.09	0.07
F,U	0.20	0.21
F,Z	0.09	1.16
F,C2	0.03	1.14
F,A2	0.09	0.66
F,T	0.08	0.75
F,L	0.12	2.52
F,B2	0.08	0.02
F,P	0.11	0.96
F,K	0.03	1.25
F,H	0.13	0.01
F,F2	0.08	0.93
B,U	0.04	0.01
B,Z	0.10	1.36
B,T	0.08	0.95
B,L	0.02	2.73
B,K	0.08	1.45
B,C	0.02	3.00
B,Y	0.08	1.84

B,F2	0.08	1.14
E2,F2	0.25	0.26
H,J	0.09	0.05
H,U	0.04	0.23
H,Z	0.10	1.14
H,C2	0.08	1.12
H,A2	0.10	0.65
H,T	0.15	0.74
H,L	0.04	2.51
H,B2	0.30	0.03
H,P	0.04	0.95
H,K	0.08	1.24
H,F2	0.27	0.92
Y,Z	0.03	0.48
Y,A2	0.06	0.98
Y,E2	0.22	0.44
Y,F2	0.10	0.70

Descriptive Analysis		
Mean	0.08	1.07
Median	0.07	0.96
Max	0.30	3.01
Min	0.01	0.01
St Dev.	0.06	0.82
Variance	0.00	0.68
t-test and p-value		
	H_0: ρ= 0 , H_1: ρ≠ 0	
α is 0.05, df= n-2	117-2=115	
$t_{0.025,df}$	$t_{0.025,115}$=1.9799	
r	-0.24	
t	10.40 > 1.96	
p-value	0.000<0.01	
Reject H_0 Fail to reject	Rejected H_0	
There is evidence of a linear relationship at 5% level of significance between **conformity and similarity**		

From Table 12, the p-value for this t-test is <0.01. Given that the p-value falls below our selected significance level of 0.05, we will reject the null hypothesis that there is no evidence of a linear relationship at 5% level of significance between conformity and similarity; we will conclude that the data provides evidence that there is a statistically significant relationship between conformity and similarity (Figure 4).

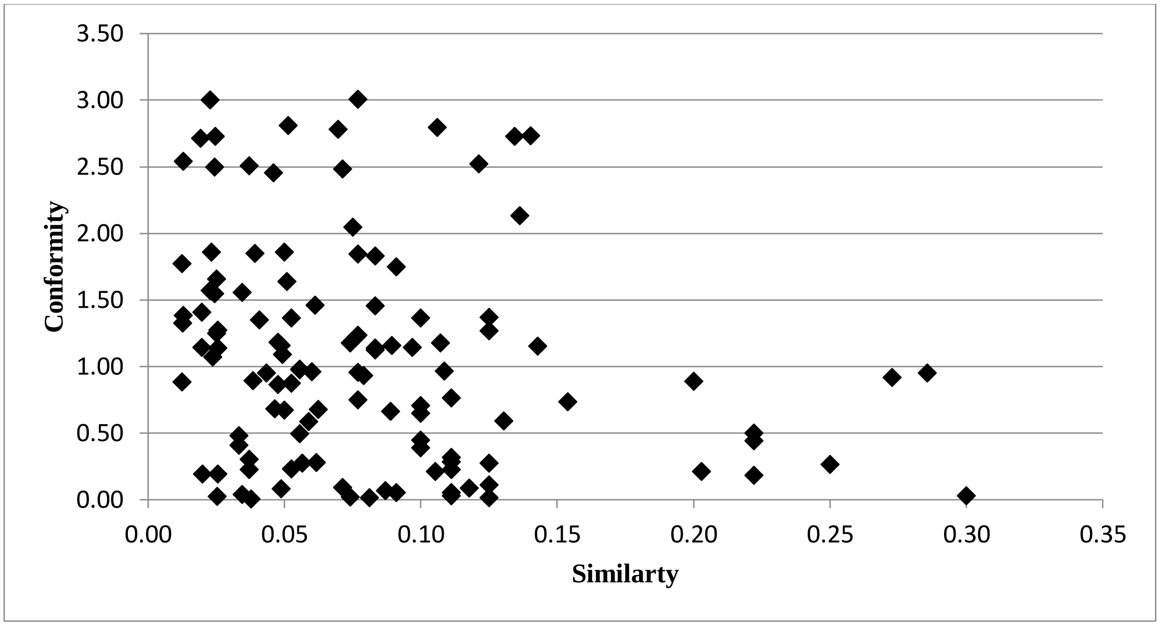

Figure 3. Conformity and Similarity Correlation

The data points in Figure 4 are negatively correlated – based on the distribution of data points and the correlation coefficient (-0.25), there appears to be a statistically significant negative relationship between the two factors. Most of the data points have network connections less than 0.15, while the values on the y-axis (answers on how close data points are) are fairly evenly distributed, although they tend to cluster a bit near the bottom of the chart. There are extremely few points with network connections greater than 0.15, and of the points that have network connections greater than 0.15, there are NO answers that have >1.00, which indicates that these two methods agree with each other, and are thus correlated/similar.

5.1.2 Correlation Analysis (Conformity and Average of Opinion)

This table below shows the level of average opinion and conformity for all combinations of any two selected Twitter users in our final dataset, regarding the responses of the previously mentioned Twitter users to the Case 1 Tweet. There are 117 total pairs of users. The below table contains a few basic descriptive analyses regarding the data (Table 13).

Table 13. Conformity and Average of Opinion

Users	Conformity	Average of opinion
J,U	0.28	4.32
J,Z	1.09	3.64
J,C2	1.07	3.65
J,A2	0.59	3.89
J,T	0.68	3.84
J,L	2.45	2.95
J,B2	0.08	4.22
J,P	0.90	3.73
J,K	1.18	3.59
J,Y	1.57	3.40
J,F2	0.87	3.75
U,Z	1.37	3.78
U,C2	1.35	3.79
U,A2	0.87	4.02
U,B2	0.19	4.36
U,E2	1.41	3.76
U,Y	1.85	3.54
U,F2	1.15	3.89
D,U	2.71	3.11
D,T	1.75	2.63
D,L	0.02	1.74
D,F	2.50	3.00
D,H	2.49	2.99
Z,C2	0.02	3.10
Z,A2	0.50	3.34
Z,B2	1.18	3.68
Z,E2	0.04	3.07
Z,F2	0.22	3.20
A2,B2	0.68	3.93
A2,F2	0.27	3.45
T,U	0.96	3.98
T,Z	0.41	3.30
T,A2	0.09	3.54
T,B2	0.77	3.88
T,E2	0.45	3.28
T,Y	0.89	3.06

T,F2	0.18	3.41
L,U	2.73	3.09
L,Z	1.36	2.41
L,C2	1.38	2.42
L,A2	1.86	2.66
L,T	1.77	2.61
L,B2	2.54	3.00
L,P	1.56	2.51
L,E2	1.33	2.39
L,Y	0.88	2.17
B2,C2	1.16	3.69
B2,F2	0.95	3.79
P,U	1.18	3.87
P,Z	0.19	3.19
P,A2	0.30	3.44
P,T	0.21	3.39
P,B2	0.98	3.78
P,E2	0.23	3.17
P,Y	0.67	2.95
P,F2	0.03	3.30
K,U	1.46	3.73
K,Z	0.09	3.05
K,C2	0.11	3.06
K,A2	0.59	3.29
K,T	0.50	3.25
K,L	1.27	2.36
K,B2	1.27	3.63
K,P	0.29	3.14
K,E2	0.05	3.03
K,Y	0.39	2.81
K,F2	0.32	3.16
C,J	2.73	2.82
C,U	3.01	2.96
C,Z	1.64	2.27
C,C2	1.66	2.28
C,A2	2.13	2.52
C,T	2.05	2.48
C,L	0.27	1.59
C,B2	2.81	2.86
C,P	1.83	2.37
C,K	1.55	2.23

C,F	2.80	2.85
C,H	2.78	2.84
C,Y	1.16	2.03
C,F2	1.86	2.39
F,J	0.07	4.22
F,U	0.21	4.36
F,Z	1.16	3.67
F,C2	1.14	3.68
F,A2	0.66	3.92
F,T	0.75	3.88
F,L	2.52	2.99
F,B2	0.02	4.26
F,P	0.96	3.77
F,K	1.25	3.63
F,H	0.01	4.24
F,F2	0.93	3.78
B,U	0.01	4.46
B,Z	1.36	3.77
B,T	0.95	3.98
B,L	2.73	3.09
B,K	1.45	3.73
B,C	3.00	2.95
B,Y	1.84	3.53
B,F2	1.14	3.89
E2,F2	0.26	3.18
H,J	0.05	4.21
H,U	0.23	4.35
H,Z	1.14	3.66
H,C2	1.12	3.67
H,A2	0.65	3.91
H,T	0.74	3.87
H,L	2.51	2.98
H,B2	0.03	4.25
H,P	0.95	3.76
H,K	1.24	3.62
H,F2	0.92	3.78
Y,Z	0.48	2.85
Y,A2	0.98	3.10
Y,E2	0.44	2.83
Y,F2	0.70	2.96
Descriptive Analysis		

Mean	1.07	3.33
Median	0.96	3.39
Max	3.01	4.46
Min	0.01	1.59
St Dev	0.82	0.61
Variance	0.68	0.38
t-test and p-value		
	H_0: ρ= 0 , H_1: ρ ≠ 0	
α is 0.05, df= n-2	117-2=115	
$t_{0.025,df}$	$t_{0.025,115}$=1.9799	
r	-0.26	
t	10.40 > 1.96	
p-value	0.000<0.01	
Reject H_0 Fail to reject	Rejected H_0	
There is evidence of linear relationship at 5% level of significance between **Conformity and average of opinion**		

From Table 13, the p-value for this t-test is <0.01. Given that the p-value falls below our selected significance level of 0.05, we will reject the null hypothesis that there is no evidence of a linear relationship at 5% level of significance between conformity and average opinion; we will conclude that the data provides evidence that there is a statistically significant relationship between conformity and average opinion (Figure 5).

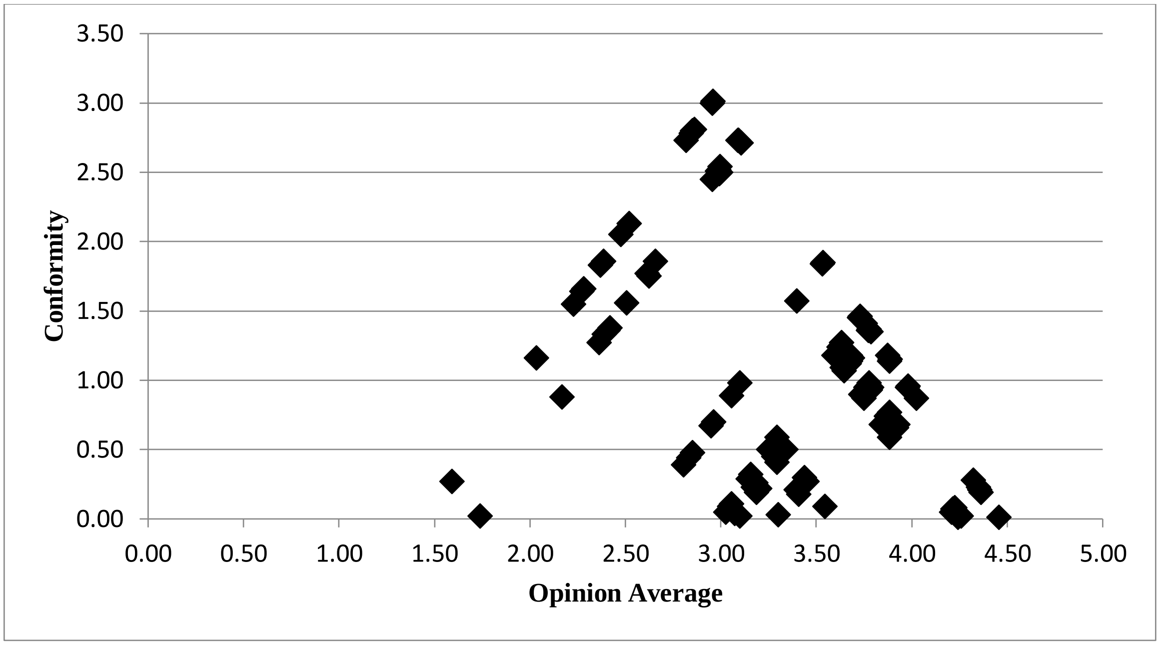

Figure 4. Conformity and Opinion Average Correlation

In Figure 5, there is a -0.36 correlation, which is a significant but fairly weak correlation, between opinion conformity and average absolute opinion for this dataset; this is presumably because if two people are in disagreement, their disagreement is equally strong (since negative emotions are usually stronger than positive ones), but if two people are in agreement, then they have a wider range of agreement. It is also worth noting that there are more people who agree than people who disagree, so if a user has a low opinion on a particular question, he is more likely to experience disagreement from a fellow user than if he has a high opinion on that question.

The chart's data is pyramid-shaped because if the average of opinion answer is very small/large, then most of the underlying data points are also small/large, so there cannot be much average (Answer Distance). Conversely, if the average of opinion answer is medium-sized, then the answers could be grouped together in the middle (i.e., lots of people who do not have strong opinions either way), or the answers could be spread out (i.e., lots of people with opinions all over the map). In other words, our correlation coefficient of -0.36 does not tell the whole story – there is a non-linear (pyramid) pattern in the data.

5.1.3 Probability Analysis (Bayesian Probability)

The below chart (Bayesian Probability) shows the cumulative probabilities of answers for a given value of network connection – in other words, what is the probability that a given answer value will be above the threshold (1 or 1.5), given that the network connection is equal to or less than a given point?

In this case:

- B: network connection less than or equal to 0.1
- A: a given answer is greater than or equal to 1.0

For instance, if a given answer is greater than or equal to 1.0 (blue line), the odds that the network connection is less than or equal to 0.1 is about 70%. Approximately 70% of the answers greater than 1.0 have network connections that are less than or equal to 0.1.

The blue line shows the probability that a given network connection will be less than or equal to 0.1, assuming that the corresponding answer is greater than 1.0.

The red line shows the probability that a given network connection will be less than or equal to 0.1, assuming that the corresponding answer is greater than 1.5.

Why do we designate answer cutoffs of 1.0 and 1.5? The reasoning is somewhat subjective – we picked these cutoffs because the underlying probabilities for these two cutoffs differ significantly, and between these two cutoffs, the entire range of data is moderately well-represented. These cutoffs help us understand the data without making the data set excessively complicated (Figure 6).

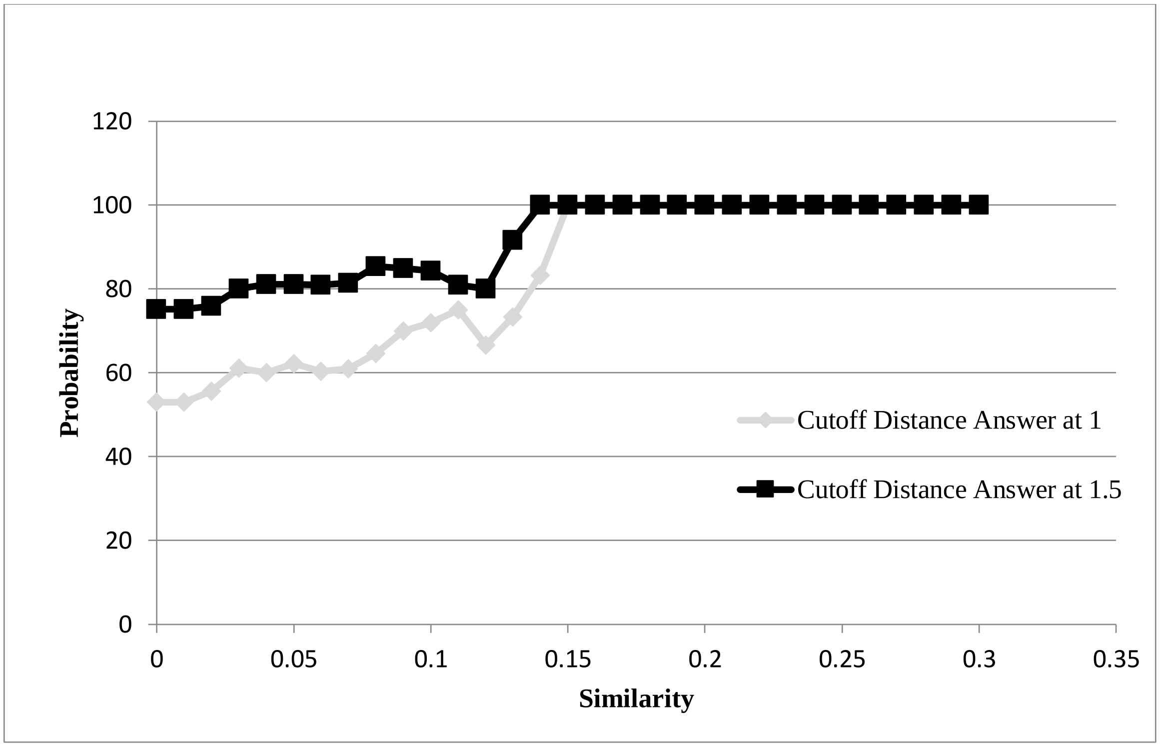

Figure 5. Bayesian Analysis

We can see from Figure 6 that in the strength of ties between 0.1 and 0.15, we start seeing an impact. Also, if there is a low strength of ties that could be not correlated – for instance, in the first pairs of points of the ties where it shows that the value is horizontal – this is called the Thrash Effect. The chart points indicate the chance to agree and to get the same answer between the two methods. The chance is 50% (where the cutoff for Answer is 1.0), or 88% (where the cutoff for Answer is 1.5). From that point onward, the strength of ties rises and the chance to share the same opinion rises concurrently. At the end, where answers and ties are both large, everybody has this strength of ties – in other words, the users share the same opinion and they give the same answer. Before this point, our graph is horizontal – the chance does not really increase – but when the data reaches a certain threshold, it starts to increase significantly. In this case, the curve starts rising between 0.1 and 0.15 for both cutoffs.

5.1.4 Conclusion

The following analyses have been conducted: two correlation analyses and one Bayesian analysis. All of the analyses show a negative correlation between the two data points, which illustrates an inverse relationship between how much people agree with the case (or how high their case values are) and how much people agree with each other about the case. We have determined that there is a negative correlation between the conformity and the network connection, and also a negative correlation between the conformity and the average of opinion answer. For the probability, if we have a high network connection or high ties, the probability of getting the same answer is very high after a certain point. In conclusion, these data points are correlated, as there is a relationship between all three data points, and the stakeholder's social media analysis is not without its own biases.

5.2 Case Two

Table 14. Case Two Summery

Text	Company	Date	Replies	All Users	Users Excluded	Users for the Analysis
"Today we released a comprehensive plan to modernize all aspects of NYC's transit system and transform everything we do. Read the #fastforwardny plan here:"	NYCT Subway	May 2018	60	46	18	28

The second case (Table 14) was also related to public transportation in New York City. This Tweet was written by the New York City subway company in May 2018. As an infrastructural project, NYC subway wanted to modernize the entire transit system. The NYC subway company used the hashtag/link #fastforwardnyc – under this hash tag, they provided a whole plan for the new updates that would affect the city. About 60 Twitter users replied to this Tweet to give their opinion about this project plan. Some of the users support and agree with this project, and others disagree – preferring to not have the updated transit system. Different actions and replies were posted in response to this Tweet, and some of these reply Tweets are unrelated – for example, people being sarcastic, humorous, rude, etc., without contributing to a productive discussion.

Table 15. Users Opinion

Users	Opinion
A	4.25
B	3.53
C	2.95
D	2.95
E	4.19
F	1.64
G	1.73
J	3.11
L	1.73
O	3.59
R	2.78
V	2.21
W	3.05
X	3.25
Z	4.27
A2	4.09
B2	1.50
C2	3.70
D2	4.41
F2	2.82
H2	2.90
I2	1.73
K2	2.95
N	2.74
O2	2.74
P2	2.38
Q2	1.64
S2	2.61
Total pairs	**284**

5.2.1 Correlation Analysis (conformity and similarity)

The table below shows the level of similarity and conformity for all combinations of any two selected Twitter users in our final dataset, regarding the responses of the previously mentioned Twitter users to the Case Two Tweet. There are 284 total pairs of users (Table 15). The below table contains a few basic descriptive analyses regarding the data (Table 16).

Table 16. Similarity and Conformity

Users	Similarity	Conformity
B,C	0.05	0.57
B,D	0.06	0.57
B,Z	0.03	0.75
B,P2	0.33	1.15
B,J	0.07	0.42
B,N2	0.17	0.79
B,D2	0.02	0.89
B,B2	0.03	2.03
B,F2	0.04	0.71
B,C2	0.14	0.17
B,S2	0.04	0.92
B,I2	0.05	1.80
B,H2	0.07	0.62
B,W	0.12	0.48
B,V	0.11	1.32
B,K2	0.01	0.58
B,Q2	0.06	1.89
B,E	0.07	0.66
B,G	0.03	1.80
B,O2	0.04	0.79
C,D	0.07	0.00
C,Z	0.10	1.32
C,P2	0.05	0.58
C,J	0.17	0.16
C,N2	0.08	0.22
C,D2	0.13	1.46
C,X	0.06	0.30
C,F2	0.05	0.13
C,C2	0.11	0.75
C,S2	0.14	0.34
C,A2	0.04	1.14
C,I2	0.18	1.23
C,H2	0.04	0.05
C,W	0.10	0.10
C,V	0.10	0.74
C,K2	0.12	0.00
C,Q2	0.18	1.32
C,E	0.07	1.24
C,G	0.04	1.23
C,O	0.15	0.64
C,O2	0.03	0.22
D,L	0.10	1.23
D,Z	0.11	1.32
D,P2	0.06	0.58
D,J	0.26	0.16
D,N2	0.15	0.22
D,D2	0.14	1.46
D,B2	0.14	1.45
D,F2	0.09	0.13
D,C2	0.13	0.75
D,S2	0.09	0.34
D,A2	0.04	1.14
D,I2	0.12	1.23
D,H2	0.06	0.05
D,W	0.17	0.10
D,V	0.12	0.74
D,K2	0.06	0.00
D,Q2	0.08	1.32
D,E	0.14	1.24
D,G	0.05	1.23
D,O	0.08	0.64
L,Z	0.06	2.55
L,D2	0.02	2.68
L,X	0.03	1.52
L,B2	0.15	0.23
L,F2	0.14	1.09
L,S2	0.03	0.88
L,A2	0.05	2.36
L,H2	0.05	1.18
L,K2	0.04	1.22
L,Q2	0.03	0.09

L,O2	0.15	1.01
Z,P2	0.03	1.90
Z,N2	0.13	1.54
Z,D2	0.34	0.14
Z,B2	0.11	2.77
Z,F2	0.05	1.45
Z,C2	0.05	0.57
Z,S2	0.11	1.66
Z,A2	0.08	0.18
Z,I2	0.14	2.55
Z,H2	0.12	1.37
Z,K2	0.25	1.32
Z,Q2	0.19	2.64
Z,O2	0.03	1.54
P2,S2	0.04	0.24
P2,Q2	0.06	0.74
J,L	0.05	1.38
J,Z	0.13	1.16
J,P2	0.15	0.74
J,N2	0.31	0.37
J,D2	0.22	1.30
J,X	0.03	0.14
J,B2	0.11	1.61
J,F2	0.06	0.29
J,C2	0.15	0.59
J,S2	0.25	0.50
J,A2	0.10	0.98
J,I2	0.17	1.38
J,H2	0.10	0.21
J,W	0.18	0.06
J,V	0.21	0.90
J,K2	0.08	0.16
J,Q2	0.21	1.47
J,O	0.25	0.48
J,O2	0.03	0.37
N2,P2	0.27	0.36
N2,S2	0.17	0.13
N2,Q2	0.11	1.10
N2,O2	0.07	0.00
D2,P2	0.04	2.04
D2,N2	0.17	1.67

D2,F2	0.05	1.59
D2,S2	0.17	1.80
D2,I2	0.18	2.68
D2,H2	0.13	1.51
D2,K2	0.19	1.46
D2,Q2	0.30	2.78
D2,O2	0.03	1.67
X,Z	0.05	1.02
X,N2	0.07	0.51
X,D2	0.05	1.16
X,B2	0.06	1.75
X,F2	0.07	0.43
X,C2	0.04	0.45
X,S2	0.07	0.64
X,A2	0.03	0.84
X,I2	0.06	1.52
X,H2	0.08	0.35
X,K2	0.06	0.30
X,Q2	0.06	1.61
X,O2	0.08	0.51
B2,P2	0.03	0.88
B2,N2	0.11	1.24
B2,D2	0.09	2.91
B2,F2	0.13	1.32
B2,C2	0.06	2.20
B2,S2	0.08	1.11
B2,I2	0.05	0.23
B2,H2	0.10	1.40
B2,K2	0.07	1.45
B2,Q2	0.09	0.14
B2,O2	0.11	1.24
F2,P2	0.04	0.44
F2,N2	0.06	0.08
F2,S2	0.04	0.21
F2,I2	0.08	1.09
F2,H2	0.05	0.09
F2,K2	0.09	0.13
F2,Q2	0.02	1.18
F2,O2	0.22	0.08
C2,P2	0.14	1.33
C2,N2	0.40	0.96

C2,D2	0.09	0.71
C2,F2	0.08	0.88
C2,S2	0.17	1.09
C2,I2	0.11	1.97
C2,H2	0.05	0.80
C2,K2	0.03	0.75
C2,Q2	0.03	2.06
A2,D2	0.06	0.32
A2,B2	0.03	2.59
A2,F2	0.03	1.27
A2,S2	0.03	1.48
A2,I2	0.04	2.36
A2,H2	0.01	1.19
A2,K2	0.05	1.14
A2,Q2	0.05	2.45
I2,P2	0.11	0.65
I2,N2	0.18	1.01
I2,S2	0.08	0.88
I2,K2	0.14	1.22
I2,Q2	0.18	0.09
I2,O2	0.03	1.01
H2,P2	0.05	0.53
H2,N2	0.14	0.17
H2,S2	0.08	0.29
H2,I2	0.04	1.18
H2,K2	0.13	0.05
H2,Q2	0.12	1.27
H2,O2	0.05	0.17
W,Z	0.06	1.22
W,P2	0.12	0.68
W,N2	0.14	0.31
W,D2	0.13	1.36
W,X	0.03	0.20
W,B2	0.07	1.55
W,F2	0.09	0.23
W,C2	0.12	0.65
W,S2	0.11	0.44
W,A2	0.08	1.04
W,I2	0.11	1.32
W,H2	0.03	0.15
W,K2	0.07	0.10

W,Q2	0.15	1.41
W,O2	0.06	0.31
F,Z	0.02	2.64
F,A2	0.06	2.45
V,Z	0.05	2.06
V,P2	0.11	0.16
V,N2	0.23	0.53
V,D2	0.06	2.20
V,B2	0.03	0.71
V,F2	0.07	0.61
V,C2	0.25	1.49
V,S2	0.20	0.40
V,I2	0.10	0.48
V,H2	0.03	0.69
V,W	0.11	0.84
V,K2	0.02	0.74
K2,P2	0.02	0.58
K2,N2	0.05	0.21
K2,S2	0.08	0.34
K2,Q2	0.13	1.31
K2,O2	0.04	0.21
A,C	0.04	1.30
A,D	0.10	1.30
A,Z	0.08	0.02
A,J	0.05	1.14
A,D2	0.04	0.16
A,X	0.03	1.00
A,B2	0.06	2.75
A,A2	0.05	0.16
A,I2	0.04	2.52
A,H2	0.02	1.35
A,W	0.04	1.20
A,K2	0.05	1.30
A,Q2	0.03	2.61
A,G	0.03	2.52
A,O	0.05	0.66
Q2,S2	0.13	0.97
E,Z	0.08	0.09
E,P2	0.07	1.81
E,J	0.21	1.08
E,N2	0.16	1.45

E,D2	0.10	0.22
E,X	0.03	0.94
E,B2	0.08	2.69
E,F2	0.09	1.37
E,C2	0.14	0.49
E,S2	0.09	1.58
E,A2	0.04	0.10
E,I2	0.12	2.46
E,H2	0.06	1.28
E,V	0.13	1.98
E,K2	0.05	1.24
E,Q2	0.11	2.55
E,G	0.02	2.46
E,O	0.08	0.60
G,Z	0.20	2.55
G,P2	0.06	0.65
G,J	0.05	1.38
G,N2	0.08	1.01
G,D2	0.23	2.68
G,X	0.04	1.52
G,B2	0.07	0.23
G,C2	0.03	1.97
G,S2	0.04	0.88
G,A2	0.05	2.36
G,I2	0.12	0.00
G,H2	0.06	1.18
G,W	0.07	1.32
G,K2	0.08	1.22
G,Q2	0.17	0.09
G,O	0.10	1.86
G,O2	0.04	1.01
O,Z	0.07	0.68
O,P2	0.06	1.21
O,N2	0.09	0.85

O,D2	0.12	0.82
O,X	0.09	0.34
O,B2	0.05	2.09
O,F2	0.06	0.77
O,S2	0.12	0.98
O,A2	0.13	0.50
O,I2	0.07	1.86
O,H2	0.04	0.68
O,W	0.12	0.54
O,K2	0.07	0.64
O,Q2	0.16	1.95
O,O2	0.06	0.85
O2,P2	0.09	0.36
O2,S2	0.05	0.13
O2,Q2	0.09	1.10
Descriptive Analysis		
Mean	0.09	1.02
Median	0.08	0.97
Max	0.40	2.91
Min	0.01	0.00
St Dev	0.06	0.73
Variance	0.00	0.53
t-test and p-value		
	H_0: ρ= 0 , H_1: ρ ≠ 0	
α is 0.05, df= n-2	284-2=282	
$t_{0.025,df}$	$t_{0.025,282}$=1.96	
r	-0.04	
t	16.78 > 1.96	
p-value	0.000<0.01	
Reject H_0 Fail to reject	Rejected H_0	
There is evidence of a linear relationship at 5% level of significance between **conformity and similarity**		

From Table 16, the p-value for this t-test is <0.01. Given that the p-value falls below our selected significance level of 0.05, we will reject the null hypothesis that there is no evidence of a linear relationship at 5% level of significance between conformity and similarity; we will conclude that the data provides evidence that there is a statistically significant relationship between conformity and similarity (Figure 7).

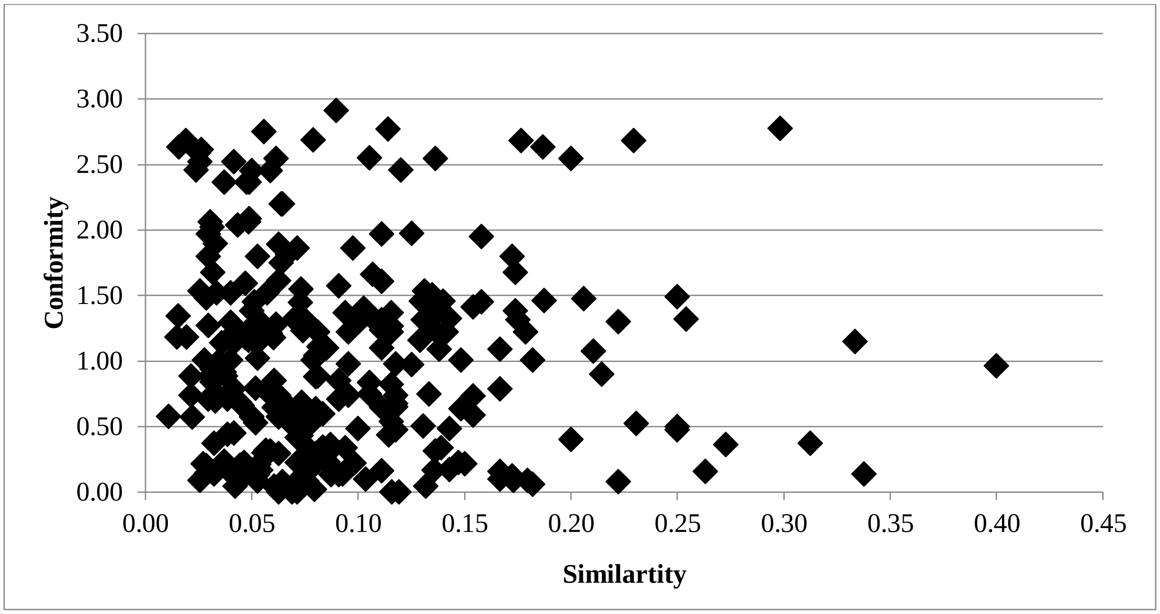

Figure 6. Conformity and Similarity Correlation

The data points in Figure 7 are also weakly correlated; based on the distribution of data points and the correlation coefficient (-0.04), there appears to be a statistically significant relationship between the two factors. Most of the data points have low values: both network connection and answers have most of their values located below the midrange. Based on the chart, the data points do not appear to be correlated, as both the high network connection/low answer and low network connection/high answer parts of the graph are equally populated. There also appears to be no meaningful trends in the data that would indicate a significant correlation between the two variables.

5.2.2 Correlation Analysis (Conformity and Average of Opinion)

The table below shows the level of average opinion and conformity for all combinations of any two selected Twitter users in our final dataset, specifically regarding the responses of the previously mentioned Twitter users to the Case Two Tweet. There are 284 total pairs of users. The below table contains a few basic descriptive analyses regarding the data (Table 17).

Table 17. Conformity and Average of Opinion

Users	Average of opinion	Conformity
B,C	0.57	3.24
B,D	0.57	3.24
B,Z	0.75	3.90
B,P2	1.15	2.95
B,J	0.42	3.32
B,N2	0.79	3.13
B,D2	0.89	3.97
B,B2	2.03	2.51
B,F2	0.71	3.17
B,C2	0.17	3.61
B,S2	0.92	3.07
B,I2	1.80	2.63
B,H2	0.62	3.22
B,W	0.48	3.29
B,V	1.32	2.87
B,K2	0.58	3.24
B,Q2	1.89	2.58
B,E	0.66	3.86
B,G	1.80	2.63
B,O2	0.79	3.13
C,D	0.00	2.95
C,Z	1.32	3.61
C,P2	0.58	2.66
C,J	0.16	3.03
C,N2	0.22	2.84
C,D2	1.46	3.68
C,X	0.30	3.10
C,F2	0.13	2.89
C,C2	0.75	3.33
C,S2	0.34	2.78
C,A2	1.14	3.52
C,I2	1.23	2.34
C,H2	0.05	2.93
C,W	0.10	3.00
C,V	0.74	2.58
C,K2	0.00	2.95
C,Q2	1.32	2.29
C,E	1.24	3.57
C,G	1.23	2.34
C,O	0.64	3.27
C,O2	0.22	2.84
D,L	1.23	2.34
D,Z	1.32	3.61
D,P2	0.58	2.66
D,J	0.16	3.03
D,N2	0.22	2.84
D,D2	1.46	3.68
D,B2	1.45	2.23
D,F2	0.13	2.89
D,C2	0.75	3.33
D,S2	0.34	2.78
D,A2	1.14	3.52
D,I2	1.23	2.34
D,H2	0.05	2.93
D,W	0.10	3.00
D,V	0.74	2.58
D,K2	0.00	2.95
D,Q2	1.32	2.29
D,E	1.24	3.57
D,G	1.23	2.34
D,O	0.64	3.27
L,Z	2.55	3.00
L,D2	2.68	3.07
L,X	1.52	2.49
L,B2	0.23	1.61
L,F2	1.09	2.27
L,S2	0.88	2.17
L,A2	2.36	2.91
L,H2	1.18	2.32
L,K2	1.22	2.34
L,Q2	0.09	1.68

L,O2	1.01	2.23
Z,P2	1.90	3.32
Z,N2	1.54	3.50
Z,D2	0.14	4.34
Z,B2	2.77	2.89
Z,F2	1.45	3.55
Z,C2	0.57	3.99
Z,S2	1.66	3.44
Z,A2	0.18	4.18
Z,I2	2.55	3.00
Z,H2	1.37	3.59
Z,K2	1.32	3.61
Z,Q2	2.64	2.95
Z,O2	1.54	3.50
P2,S2	0.24	2.49
P2,Q2	0.74	2.01
J,L	1.38	2.42
J,Z	1.16	3.69
J,P2	0.74	2.74
J,N2	0.37	2.92
J,D2	1.30	3.76
J,X	0.14	3.18
J,B2	1.61	2.31
J,F2	0.29	2.96
J,C2	0.59	3.41
J,S2	0.50	2.86
J,A2	0.98	3.60
J,I2	1.38	2.42
J,H2	0.21	3.01
J,W	0.06	3.08
J,V	0.90	2.66
J,K2	0.16	3.03
J,Q2	1.47	2.37
J,O	0.48	3.35
J,O2	0.37	2.92
N2,P2	0.36	2.56
N2,S2	0.13	2.67
N2,Q2	1.10	2.19
N2,O2	0.00	2.74
D2,P2	2.04	3.39
D2,N2	1.67	3.57

D2,F2	1.59	3.61
D2,S2	1.80	3.51
D2,I2	2.68	3.07
D2,H2	1.51	3.66
D2,K2	1.46	3.68
D2,Q2	2.78	3.02
D2,O2	1.67	3.57
X,Z	1.02	3.76
X,N2	0.51	2.99
X,D2	1.16	3.83
X,B2	1.75	2.38
X,F2	0.43	3.03
X,C2	0.45	3.48
X,S2	0.64	2.93
X,A2	0.84	3.67
X,I2	1.52	2.49
X,H2	0.35	3.08
X,K2	0.30	3.10
X,Q2	1.61	2.44
X,O2	0.51	2.99
B2,P2	0.88	1.94
B2,N2	1.24	2.12
B2,D2	2.91	2.96
B2,F2	1.32	2.16
B2,C2	2.20	2.60
B2,S2	1.11	2.06
B2,I2	0.23	1.61
B2,H2	1.40	2.20
B2,K2	1.45	2.23
B2,Q2	0.14	1.57
B2,O2	1.24	2.12
F2,P2	0.44	2.60
F2,N2	0.08	2.78
F2,S2	0.21	2.71
F2,I2	1.09	2.27
F2,H2	0.09	2.86
F2,K2	0.13	2.88
F2,Q2	1.18	2.23
F2,O2	0.08	2.78
C2,P2	1.33	3.04
C2,N2	0.96	3.22

C2,D2	0.71	4.06
C2,F2	0.88	3.26
C2,S2	1.09	3.16
C2,I2	1.97	2.71
C2,H2	0.80	3.30
C2,K2	0.75	3.33
C2,Q2	2.06	2.67
A2,D2	0.32	4.25
A2,B2	2.59	2.80
A2,F2	1.27	3.45
A2,S2	1.48	3.35
A2,I2	2.36	2.91
A2,H2	1.19	3.50
A2,K2	1.14	3.52
A2,Q2	2.45	2.86
I2,P2	0.65	2.05
I2,N2	1.01	2.23
I2,S2	0.88	2.17
I2,K2	1.22	2.34
I2,Q2	0.09	1.68
I2,O2	1.01	2.23
H2,P2	0.53	2.64
H2,N2	0.17	2.82
H2,S2	0.29	2.76
H2,I2	1.18	2.32
H2,K2	0.05	2.93
H2,Q2	1.27	2.27
H2,O2	0.17	2.82
W,Z	1.22	3.66
W,P2	0.68	2.71
W,N2	0.31	2.89
W,D2	1.36	3.73
W,X	0.20	3.15
W,B2	1.55	2.28
W,F2	0.23	2.93
W,C2	0.65	3.38
W,S2	0.44	2.83
W,A2	1.04	3.57
W,I2	1.32	2.39
W,H2	0.15	2.98
W,K2	0.10	3.00

W,Q2	1.41	2.34
W,O2	0.31	2.89
F,Z	2.64	2.95
F,A2	2.45	2.86
V,Z	2.06	3.24
V,P2	0.16	2.29
V,N2	0.53	2.47
V,D2	2.20	3.31
V,B2	0.71	1.86
V,F2	0.61	2.51
V,C2	1.49	2.96
V,S2	0.40	2.41
V,I2	0.48	1.97
V,H2	0.69	2.56
V,W	0.84	2.63
V,K2	0.74	2.58
K2,P2	0.58	2.66
K2,N2	0.21	2.84
K2,S2	0.34	2.78
K2,Q2	1.31	2.29
K2,O2	0.21	2.84
A,C	1.30	3.60
A,D	1.30	3.60
A,Z	0.02	4.26
A,J	1.14	3.68
A,D2	0.16	4.33
A,X	1.00	3.75
A,B2	2.75	2.88
A,A2	0.16	4.17
A,I2	2.52	2.99
A,H2	1.35	3.58
A,W	1.20	3.65
A,K2	1.30	3.60
A,Q2	2.61	2.94
A,G	2.52	2.99
A,O	0.66	3.92
Q2,S2	0.97	2.12
E,Z	0.09	4.23
E,P2	1.81	3.28
E,J	1.08	3.65
E,N2	1.45	3.46

E,D2	0.22	4.30
E,X	0.94	3.72
E,B2	2.69	2.84
E,F2	1.37	3.50
E,C2	0.49	3.94
E,S2	1.58	3.40
E,A2	0.10	4.14
E,I2	2.46	2.96
E,H2	1.28	3.55
E,V	1.98	3.20
E,K2	1.24	3.57
E,Q2	2.55	2.91
E,G	2.46	2.96
E,O	0.60	3.89
G,Z	2.55	3.00
G,P2	0.65	2.05
G,J	1.38	2.42
G,N2	1.01	2.23
G,D2	2.68	3.07
G,X	1.52	2.49
G,B2	0.23	1.61
G,C2	1.97	2.71
G,S2	0.88	2.17
G,A2	2.36	2.91
G,I2	0.00	1.73
G,H2	1.18	2.32
G,W	1.32	2.39
G,K2	1.22	2.34
G,Q2	0.09	1.68
G,O	1.86	2.66
G,O2	1.01	2.23
O,Z	0.68	3.93
O,P2	1.21	2.98
O,N2	0.85	3.16
O,D2	0.82	4.00

O,X	0.34	3.42
O,B2	2.09	2.54
O,F2	0.77	3.20
O,S2	0.98	3.10
O,A2	0.50	3.84
O,I2	1.86	2.66
O,H2	0.68	3.25
O,W	0.54	3.32
O,K2	0.64	3.27
O,Q2	1.95	2.61
O,O2	0.85	3.16
O2,P2	0.36	2.56
O2,S2	0.13	2.67
O2,Q2	1.10	2.19
Descriptive Analysis		
Mean	1.02	2.96
Median	0.97	2.95
Max	2.91	4.34
Min	0.00	1.57
St Dev	0.73	0.59
Variance	0.53	0.34
t-test and p-value		
	H_0: ρ= 0 , H_1: ρ≠ 0	
α is 0.05, df= n-2	284-2=282	
$t_{0.025,df}$	$t_{0.025,282}$=1.96	
r	-0.04	
t	16.78 > 1.96	
p-value	0.000<0.01	
Reject H_0 Fail to reject	Rejected H_0	
There is evidence of a linear relationship at 5% level of significance between **conformity and average answer**		

From Table 17, the p-value for this t-test is <0.01. Given that the p-value falls below our selected significance level of 0.05, we will reject the null hypothesis that there is no evidence of a linear relationship at 5% level of significance between conformity and average opinion; we will conclude that the data provides evidence that there is a statistically significant relationship between conformity and average opinion (Figure 8).

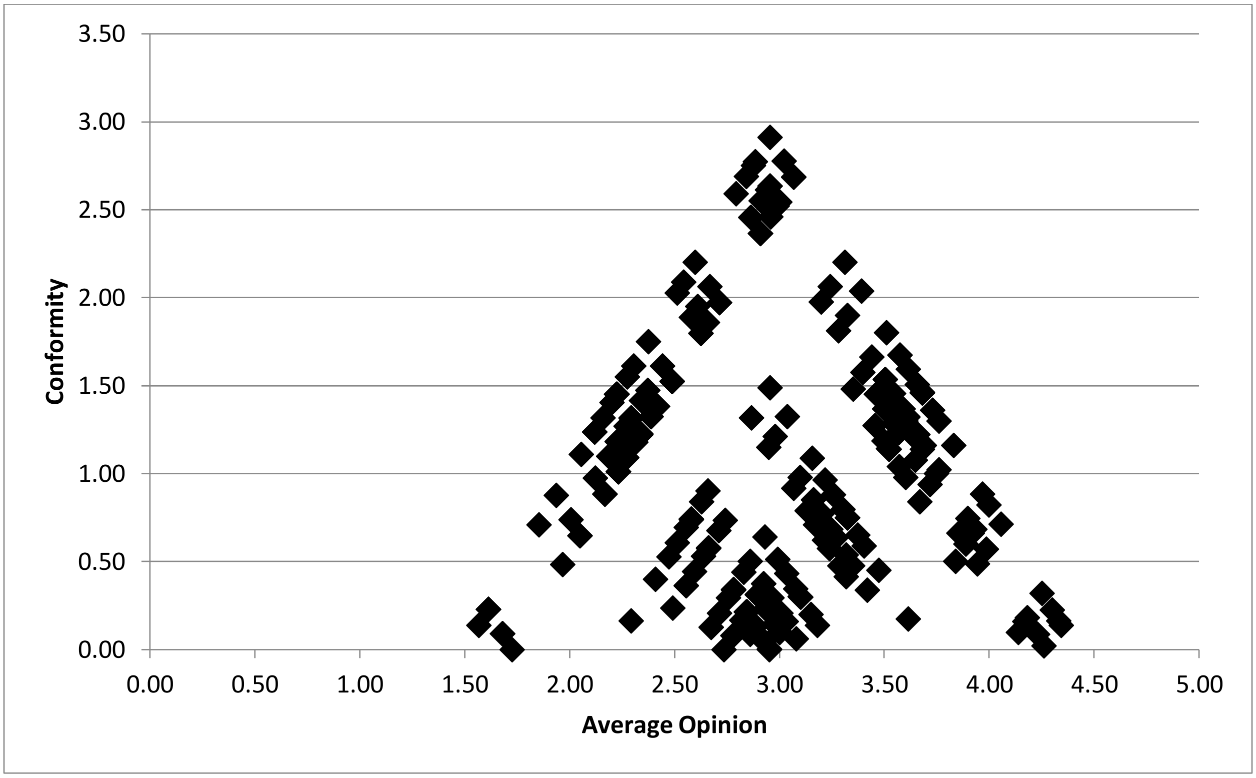

Figure 7. Conformity and Average Opinion Correlation

In Figure 8, there is a -0.51 correlation, which is statistically significant, between opinion conformity and average absolute opinion for this dataset, presumably because if two people are in disagreement, their disagreement is equally strong (since negative emotions are usually stronger than positive ones), but if two people are in agreement, then they have a wider range of agreement. It is also worth noting that there are more people who agree than people who disagree, so if a user has a low opinion on a particular question, he is more likely to experience disagreement with a fellow user than if he had had a high opinion on a particular question.

As in the previous example, the chart's data is still somewhat pyramid-shaped, but the pattern is not as pronounced because the data is much more clustered together. The data is (somewhat) pyramid-shaped because if the average of opinion answer is very small/large, then most of the underlying data points are also small/large, so there cannot be much average (Answer Distance). Conversely, if the average of opinion answer is medium-sized, then the answers could be grouped together in the middle (i.e., people who do not have strong opinions either way), or the answers could be spread out (i.e., people with opinions all over the map). In other words, our correlation coefficient of -0.51 does not tell the whole story, as there is a non-linear (pyramid) pattern in the data.

5.2.3 Probability Analysis (Bayesian Probability)

The following chart (Bayesian Probability Case Two) shows the cumulative probabilities of answers for a given value of network connection, or what is the probability that a given answer value will be above the threshold (1 or 1.5), given that the network connection is equal to or less than a given point.

- B: network connection less than or equal to 0.1
- A: a given answer is greater than or equal to 1.0

For instance, if a given answer is greater than or equal to 1.0 (blue line), the odds that the network connection is less than or equal to 0.1 is about 70%. In other words, about 70% of the answers greater than 1.0 have network connections that are less than or equal to 0.1.

The blue line shows the probability that a given network connection will be less than or equal to 0.1, in the instance that a specific answer is greater than 1.0.

The red line shows the probability that a given network connection will be less than or equal to 0.1, in the instance that a specific answer is greater than 1.5.

Why do we designate answer cutoffs of 1.0 and 1.5? The reasoning is somewhat subjective – we picked these cutoffs because the underlying probabilities for these two cutoffs differ significantly, and between these two cutoffs, the entire range of data is moderately well-represented. These cutoffs help us to understand the data, without causing the data set to be unnecessarily complicated (Figure 10).

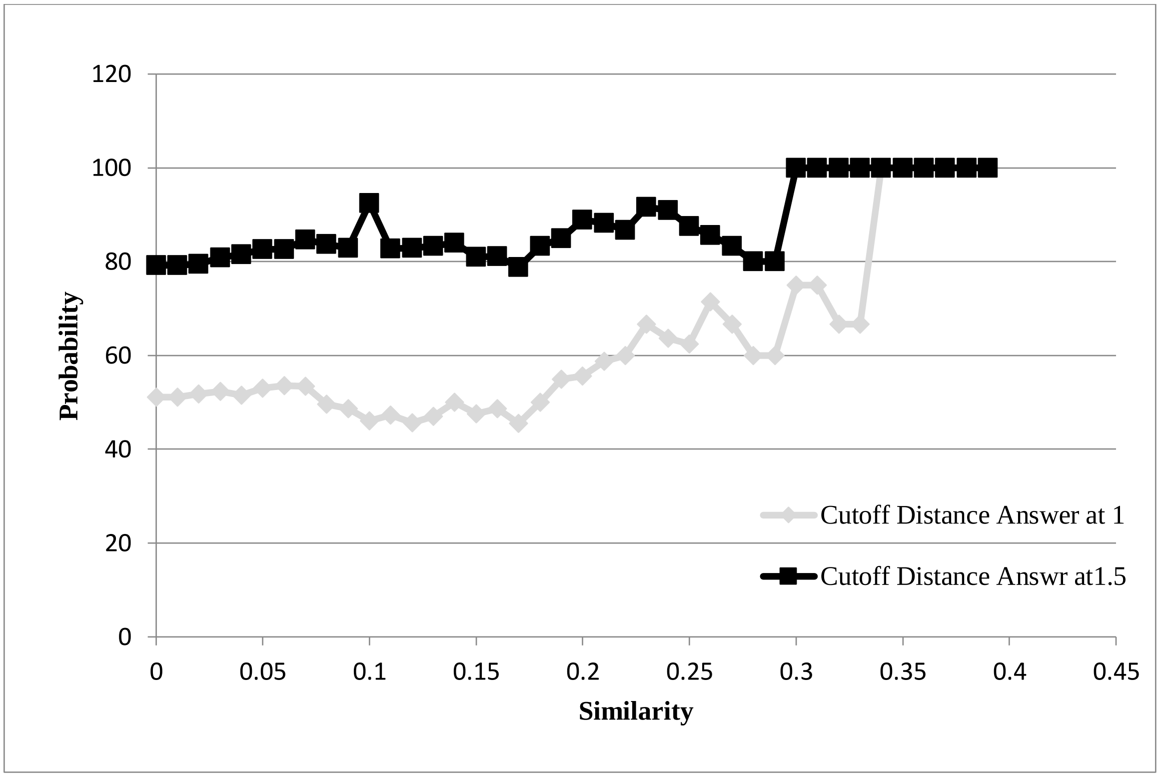

Figure 8. Bayesian Probability Case Two

We can see from (Figure 9) that starting in the strength of ties range between 0.15 and 0.20, we begin to see an impact. Also, if there is a low strength of ties that could not be correlated – for instance, in the first few pairs of points of the ties, the lines are horizontal – this is called the Thrash Effect. The chart points indicate the chance to agree and to get the same answer between the two methods. The chance is 50% (where the cutoff for answer is 1.0) or 80% (where the cutoff for answer is 1.5); from then on, the strength of ties rises and the chance to share the same opinion rises, as well. At the end, where answers and ties are both large, everybody has this strength of ties – in other words, they share the same opinion and the same answer. Before this point, our graph is horizontal –the chance does not really increase – but when the data reaches a certain threshold, it starts to increase significantly. In this case, the curve begins rising between 0.15 and 0.20 for both cutoffs.

5.2.4 Conclusion

The following analyses have been conducted: two correlation analyses and one Bayesian analysis. All of the analyses show a negative (but weak) correlation between the two data points, as there is an inverse relationship between how much people agree with the case (or how high their case values are) and how much people agree with each other in the case. We have determined that there is a negative (but weak) correlation between the conformity and the network connection, and also a negative correlation between the conformity and the average of opinion answer. For the probability, if we have a high network connection or high ties, the probability of getting the same answer is very high after a certain point. It is also worth noting that this data is fairly consistent compared to the other datasets – the cases tend to agree, instead of varying their levels of agreement. In conclusion, these data points are correlated, as there is a relationship between all three data points, and the stakeholder's social media analysis is not without its own biases.

5.3 Case Three

Table 18. Case Three Summery

Text	Company	Date	Replies	All Users	Users Excluded	Users for the Analysis
"Exciting news! With more than 500 7000-series railcars delivered, Metro is imagining the NEXT generation of railcars, designed using customer feedback. New 8000-series railcars will replace the 2000/3000 series, which will be 40 years old and due for retirement in 2024. #wmata"	Metro	Apr 2018	55	46	27	19

The third case focused on public transportation in Washington D.C.. The above Tweet was made by Washington Metropolitan Area Transit Authority (WMATA) in April 2018. As an infrastructural project, Washington Metro Company wanted to redesign the transit system using customer feedback. The Washington Metro Company used the hashtag #fWMATA – under this hashtag they provided a whole plan for the updating of the transit system. In this case, about 55 users replied to this Tweet to give their opinion about this project plan. Some of the users support and agree with this project, and others disagree, preferring to not have this updated transit system. Different actions and replies were posted in response to this Tweet, and some of these reply Tweets are unrelated – for example, people are sarcastic, humorous, rude, etc., without contributing to a productive discussion.

Table 19. Users Opinion

Users	Opinion
D	3.86
J	3.21
N	3.00
O	3.22
Q	3.82
R	4.00
S	3.89
T	2.94
X	1.91
Y	3.29
Z	4.13
D2	4.27
G2	3.00
J2	3.80
L2	3.38
M2	3.26
O2	3.22
S2	2.95
T2	2.63
Total pairs	**119**

5.3.1 Correlation Analysis (Conformity and Similarity)

This table below shows the level of similarity and conformity for all combinations of any two selected Twitter users in our final dataset, which is based on the responses of the previously mentioned Twitter users to the Case 3 Tweet. There are 119 total pairs of users (Table 19). The below table contains a few basic descriptive analyses regarding the data (Table 20).

Table 18. Similarity and Conformity

Users	Similarity	Conformity
D2,G2	0.10	1.27
D2,M2	0.03	1.00
D2,S2	0.21	1.32
D2,T2	0.05	1.64
D2,J2	0.06	0.47
O2,S2	0.08	0.27
O2,T2	0.08	0.59
X,D2	0.02	2.36
X,O2	0.09	1.31
X,G2	0.08	1.09
X,M2	0.13	1.35
X,L2	0.05	1.48
X,S2	0.10	1.04
X,Z	0.02	2.22

X,T2	0.12	0.72
X,J2	0.05	1.89
G2,O2	0.03	0.22
G2,M2	0.14	0.26
G2,S2	0.14	0.05
G2,T2	0.06	0.37
G2,J2	0.08	0.80
R,D2	0.07	0.27
R,X	0.03	2.09
R,G2	0.04	1.00
R,Y	0.09	0.71
R,M2	0.10	0.74
R,S	0.08	0.11
R,S2	0.15	1.05
R,T2	0.07	1.37
R,J2	0.13	0.20
R,T	0.09	1.06
Y,G2	0.03	0.29
M2,O2	0.06	0.04
M2,S2	0.15	0.31
M2,T2	0.11	0.63
Q,D2	0.11	0.44
Q,O2	0.04	0.60
Q,X	0.07	1.91
Q,G2	0.03	0.82
Q,R	0.18	0.18
Q,Y	0.06	0.54

Q,M2	0.12	0.56
Q,S	0.10	0.07
Q,S2	0.25	0.87
Q,Z	0.07	0.31
Q,T2	0.11	1.19
Q,J2	0.08	0.02
D,D2	0.10	0.41
D,O2	0.09	0.63
D,X	0.10	1.95
D,G2	0.11	0.86
D,R	0.09	0.14
D,M2	0.22	0.59
D,Q	0.16	0.03
D,N	0.12	0.86
D,S	0.16	0.03
D,J	0.38	0.65
D,S2	0.24	0.91
D,Z	0.04	0.28
D,T2	0.15	1.23
D,J2	0.09	0.06
D,T	0.12	0.91
N,D2	0.09	1.27
N,O2	0.06	0.22
N,X	0.10	1.09
N,G2	0.16	0.00
N,R	0.03	1.00
N,Y	0.04	0.29

N,M2	0.13	0.26
N,Q	0.08	0.82
N,S	0.17	0.89
N,S2	0.12	0.05
N,O	0.04	0.22
N,Z	0.04	1.13
N,T2	0.10	0.37
N,J2	0.05	0.80
N,T	0.08	0.06
S,D2	0.13	0.38
S,O2	0.05	0.67
S,X	0.07	1.98
S,G2	0.15	0.89
S,M2	0.13	0.63
S,S2	0.25	0.94
S,Z	0.03	0.24
S,T2	0.13	1.26
S,J2	0.04	0.09
S,T	0.14	0.94
J,D2	0.09	1.06
J,O2	0.12	0.01
J,X	0.09	1.30
J,G2	0.06	0.21
J,R	0.13	0.79
J,M2	0.23	0.05
J,Q	0.22	0.61
J,N	0.14	0.21

J,S	0.23	0.68
J,S2	0.33	0.26
J,T2	0.20	0.58
J,J2	0.13	0.59
J,T	0.24	0.27
S2,T2	0.16	0.32
O,X	0.04	1.31
O,G2	0.09	0.22
O,S	0.03	0.67
O,S2	0.04	0.27
O,Z	0.33	0.91
Z,G2	0.07	1.13
Z,S2	0.06	1.18
J2,M2	0.14	0.54
J2,S2	0.07	0.85
J2,T2	0.14	1.17
T,O2	0.04	0.28
T,X	0.02	1.04
T,G2	0.03	0.06
T,M2	0.13	0.32
T,S2	0.12	0.01
T,T2	0.06	0.31
Descriptive Analysis		
Mean	0.11	0.71
Median	0.09	0.63
Max	0.38	2.36
Min	0.02	0.00

St Dev	0.07	0.53
Variance	0.00	0.29
t-test and p-value		
	H_0: ρ= 0 , H_1: ρ≠ 0	
α is 0.05, df= n-2	119-2=117	
$t_{0.025,df}$	$t_{0.025,117}$=1.9799	
r	-0.15	

t	16.78 > 1.96
p-value	0.000<0.01
Reject H_0 Fail to reject	Rejected H_0
There is evidence of a linear relationship at 5% level of significance between **conformity and similarity**	

The p-value for this t-test is <0.01. Given that the p-value falls below our selected significance level of 0.05, we will reject the null hypothesis that there is no evidence of a linear relationship at 5% level of significance between conformity and similarity; we will conclude that the data provides evidence that there is a statistically significant relationship between conformity and similarity (Figure 10).

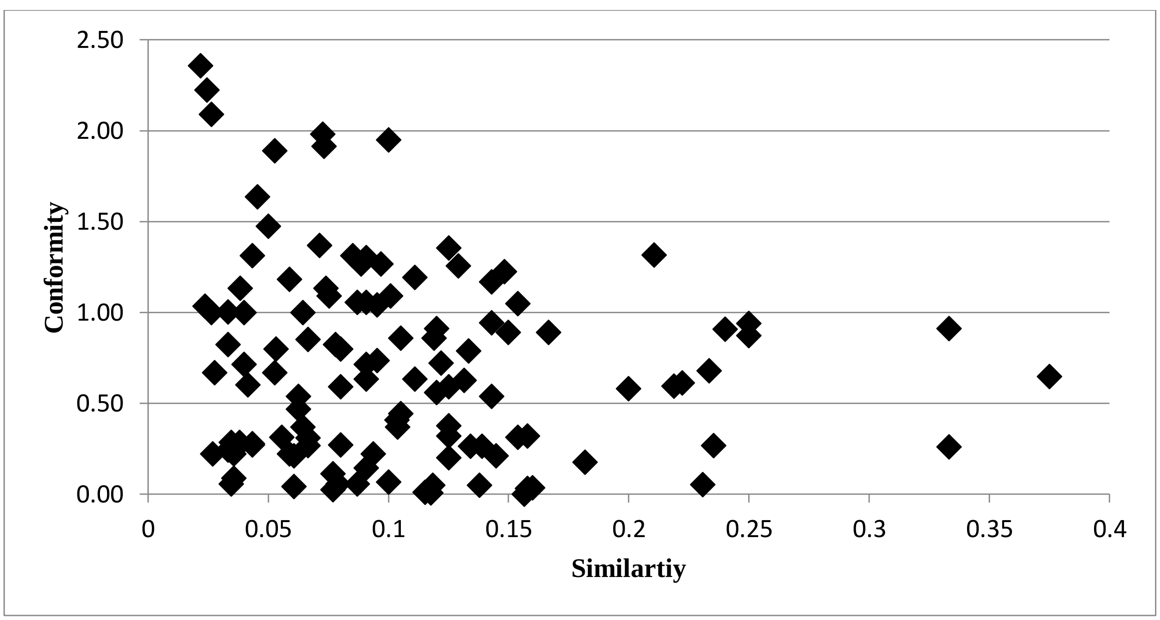

Figure 9. Conformity and Similarity Correlation

In Figure 10, the data points are negatively correlated; based on the distribution of data points and the correlation coefficient (-0.15), there appears to be a negative relationship between the two factors. Most of the data points have network connections less than 0.20, while the values on the y-axis (answers on how close data points are) are fairly evenly distributed, although they tend to cluster a bit near the bottom of the chart. There are very few points with network connections greater than 0.20 – and of the points that have network connections greater than 0.20, there is just ONE answer that has >1.20, which indicates that these two methods agree with each other, and are thus correlated/similar.

5.3.2 Correlation Analysis (Conformity and Average of Opinion)

This table below shows the level of average opinion and conformity for all combinations of any two selected Twitter users in our final dataset, which consists of the responses of the previously mentioned Twitter users to the Case 3 Tweet. There are 119 total pairs of users. The below table contains a few basic descriptive analyses regarding the data (Table 21).

Table 19. Conformity and Average of Opinion

Users	Conformity	Average of opinion
D2,G2	1.27	3.63
D2,M2	1.00	3.76
D2,S2	1.32	3.61
D2,T2	1.64	3.45
D2,J2	0.47	4.03
O2,S2	0.27	3.09
O2,T2	0.59	2.93
X,D2	2.36	3.09
X,O2	1.31	2.57
X,G2	1.09	2.45
X,M2	1.35	2.59
X,L2	1.48	2.65
X,S2	1.04	2.43
X,Z	2.22	3.02
X,T2	0.72	2.27
X,J2	1.89	2.85
G2,O2	0.22	3.11
G2,M2	1.00	3.50
G2,S2	0.26	3.13
G2,T2	0.05	2.98

G2,J2	0.37	2.82
R,D2	0.80	3.40
R,X	0.27	4.13
R,G2	2.09	2.95
R,Y	1.00	3.50
R,M2	0.71	3.64
R,S	0.74	3.63
R,S2	0.11	3.94
R,T2	1.05	3.48
R,J2	1.37	3.32
R,T	0.20	3.90
Y,G2	1.06	3.47
M2,O2	0.71	3.64
M2,S2	0.29	3.14
M2,T2	0.04	3.24
Q,D2	0.31	3.11
Q,O2	0.63	2.95
Q,X	0.44	4.05
Q,G2	0.60	3.52
Q,R	1.91	2.87
Q,Y	0.82	3.41
Q,M2	0.18	3.91
Q,S	0.54	3.55
Q,S2	0.56	3.54
Q,Z	0.07	3.86
Q,T2	0.87	3.39
Q,J2	0.31	3.98

D,D2	1.19	3.23
D,O2	0.02	3.81
D,X	0.41	4.06
D,G2	0.63	3.54
D,R	1.95	2.88
D,M2	0.86	3.43
D,Q	0.14	3.93
D,N	0.59	3.56
D,S	0.03	3.84
D,J	0.86	3.43
D,S2	0.03	3.87
D,Z	0.65	3.53
D,T2	0.91	3.40
D,J2	0.28	4.00
D,T	1.23	3.24
N,D2	0.06	3.83
N,O2	0.91	3.40
N,X	1.27	3.63
N,G2	0.22	3.11
N,R	1.09	2.45
N,Y	0.00	3.00
N,M2	1.00	3.50
N,Q	0.29	3.14
N,S	0.26	3.13
N,S2	0.82	3.41
N,O	0.89	3.44
N,Z	0.05	2.98

N,T2	0.22	3.11
N,J2	1.13	3.57
N,T	0.37	2.82
S,D2	0.80	3.40
S,O2	0.06	2.97
S,X	0.38	4.08
S,G2	0.67	3.56
S,M2	1.98	2.90
S,S2	0.89	3.44
S,Z	0.63	3.58
S,T2	0.94	3.42
S,J2	0.24	4.01
S,T	1.26	3.26
J,D2	0.09	3.84
J,O2	0.94	3.42
J,X	1.06	3.74
J,G2	0.01	3.22
J,R	1.30	2.56
J,M2	0.21	3.11
J,Q	0.79	3.61
J,N	0.05	3.24
J,S	0.61	3.52
J,S2	0.21	3.11
J,T2	0.68	3.55
J,J2	0.26	3.08
J,T	0.58	2.92
S2,T2	0.59	3.51

O,X	0.27	3.08
O,G2	0.32	2.79
O,S	1.31	2.57
O,S2	0.22	3.11
O,Z	0.67	3.56
Z,G2	0.27	3.09
Z,S2	0.91	3.68
J2,M2	1.13	3.57
J2,S2	1.18	3.54
J2,T2	0.54	3.53
T,O2	0.85	3.38
T,X	1.17	3.22
T,G2	0.28	3.08
T,M2	1.04	2.43
T,S2	0.06	2.97
T,T2	0.32	3.10
D2,G2	0.01	2.95
D2,M2	0.31	2.79
Descriptive Analysis		
Mean	**0.71**	**3.32**
Median	**0.63**	**3.40**
Max	**2.36**	**4.13**
Min	**0.00**	**2.27**
St Dev	**0.53**	**0.42**
Variance	**0.29**	**0.17**
t-test and p-value		
	H_0: $\rho = 0$, H_1: $\rho \neq 0$	

α is 0.05, df= n-2	117-2=115
$t_{0.025,df}$	$t_{0.025,115}$=1.9799
r	-0.27
t	10.32 > 1.9799
p-value	0.000<0.01

Reject H_0 Fail to reject	Rejected H_0
There is evidence of a linear relationship at 5% level of significance between **conformity and average opinion**	

From Table 21, the p-value for this t-test is <0.01. Given that the p-value falls below our selected significance level of 0.05, we will reject the null hypothesis that there is no evidence of a linear relationship at 5% level of significance between conformity and average opinion; we will conclude that the data provides evidence that there is a statistically significant relationship between conformity and average opinion (Figure 11).

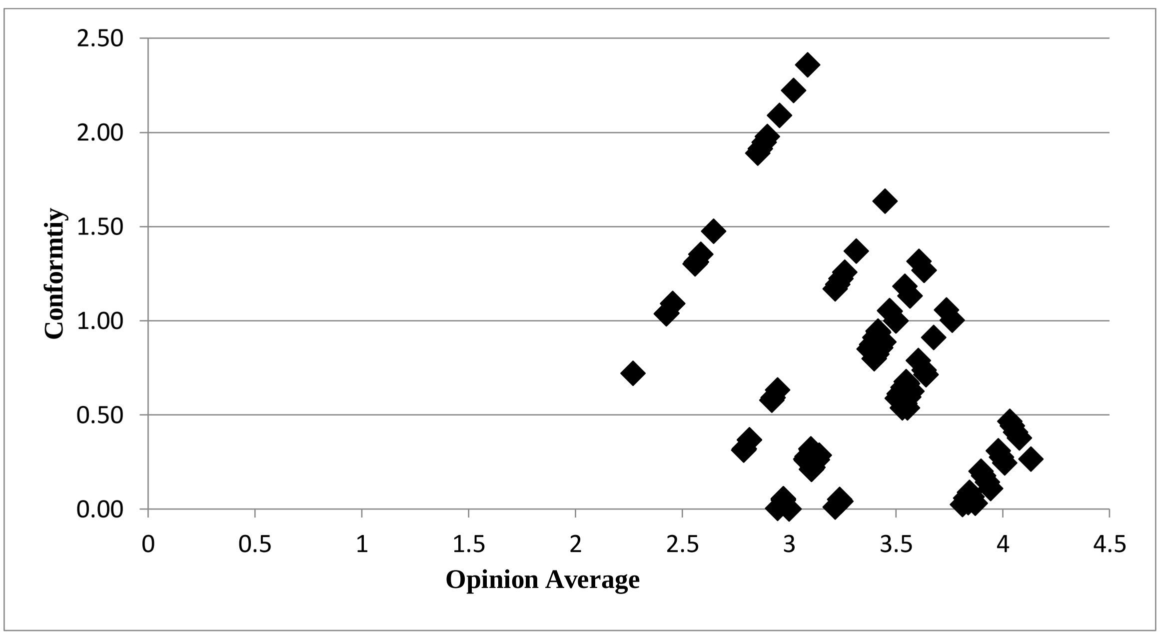

Figure 10. Opinion Average and Conformity Correlation

In Table 21, there is a -0.27 correlation, which is statistically significant, between opinion conformity and average absolute opinion for this dataset, presumably because if two people are in disagreement, their disagreement is equally strong (since negative emotions are usually stronger than positive ones), but if two people are in agreement, then they have a wider range of agreement. It is also worth noting that there are more people who agree than people who disagree, so if a user has a low opinion on a particular question, he is more likely to experience disagreement from a fellow user than if he had had a high opinion on a particular question.

The chart's data is pyramid-shaped because if the average of opinion answer is very small/large, then most of the underlying data points are also small/large, so there cannot be much average (answer distance). Conversely, if the average of opinion answer is medium-sized, then the answers could be grouped together in the middle (i.e., people who do not have strong opinions either way), or the answers could be spread out (i.e., people with opinions all over the map). In other words, our correlation coefficient of -0.27 does not tell the whole story, as there is a non-linear (pyramid) pattern in the data.

5.3.3 Probability Analysis (Bayesian Probability)

The following chart (Bayesian Probability) shows the cumulative probabilities of answers for a given value of network connection, or the probability that an answer value will be above the threshold (1 or 1.5), given that the network connection is equal to or less than a given point.

In this case:

- B: network connection less than or equal to 0.1
- A: a given answer is greater than or equal to 1.0

For instance, if a given answer is greater than or equal to 1.0 (blue line), the odds that the network connection is less than or equal to 0.1 is about 85%. In other words, about 85% of the answers greater than 1.0 have network connections that are less than or equal to 0.1.

The blue line shows the probability that a given network connection will be less than or equal to 0.1, in the instance that a specific answer is greater than 1.0.

The red line shows the probability that a given network connection will be less than or equal to 0.1, in the instance that a specific answer is greater than 1.5.

Why do we designate answer cutoffs of 1.0 and 1.5? The reasoning is somewhat subjective – we picked these cutoffs because the underlying probabilities for these two cutoffs differ significantly, and between these two cutoffs, the entire range of data is moderately well-represented. These cutoffs help us understand the data by producing a workable set of data that is not overly complicated (Figure 12).

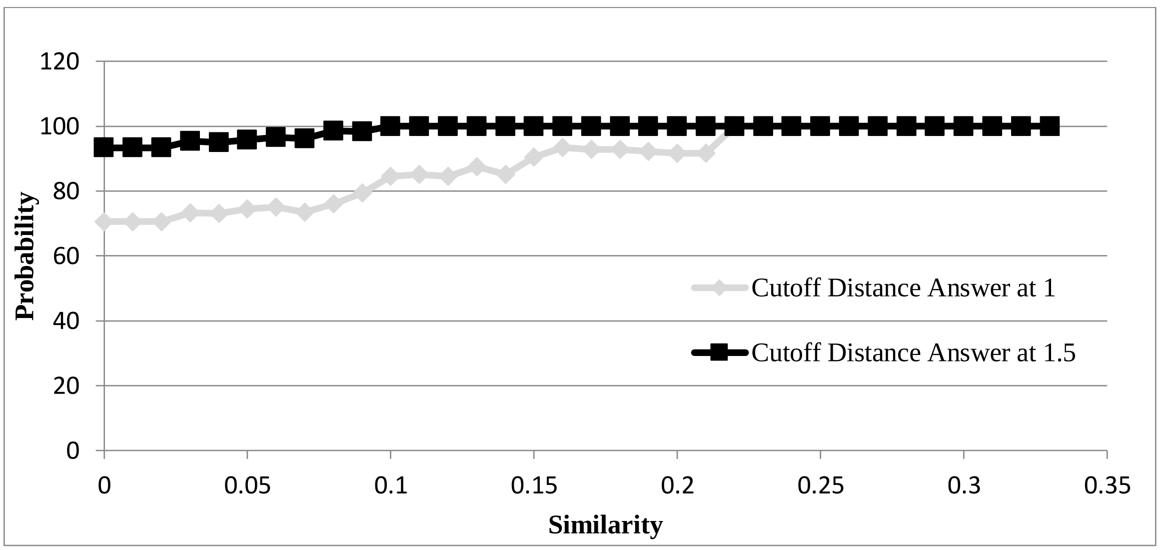

Figure 11. Bayesian Probability

Figure 12 shows that, starting in the strength of ties between 0.05 and 0.1, we begin observing an impact. Also, if there is a low strength of ties that could be not correlated – for instance, in the first pairs of points, the line is horizontal – this is called the Thrash Effect. The chart points indicate the chance to agree and to receive the same answer between the two methods. The chance is 70% (where the cutoff for answer is 1.0) or 98% (where the cutoff for Answer is 1.5). From then on, the strength of ties rises, and the chance to share the same opinion also rises. At the end, where answers and ties are both large, everybody has this strength of ties; in other words, they share the same opinion and the same answer. Before this point, our graph is horizontal, illustrating that the chance does not really increase, but when the numbers reach a certain threshold, the chance starts to increase significantly. In this case, the curve starts rising between 0.05 and 0.1 for both cutoffs.

5.3.4 Conclusion

The following analyses have been conducted: two correlation analyses and one Bayesian analysis. All of the analyses show a negative (but weak) correlation between the two data points, as there is an inverse relationship between how much people agree with the case (or how high their case values are) and how much people agree with each other in the case. We have determined that there is a negative (but weak) correlation between the conformity and the network connection, and also a negative correlation between the conformity and the average of opinion answer. For the probability, if we have a high network connection or high ties, the probability of getting the same answer is very high after a certain point. In conclusion, these data points are correlated, as there is a relationship between all three data points, and the stakeholder's social media analysis is not without its own biases.

5.4 Case Four

Table 22. Case Four Summery

Text	Company	Date	Replies	All Users	Users Excluded	Users for the Analysis
"A very special thank you to all who attended our R211 prototype design open house. If you visited us but didn't have the chance to give feedback, please leave your comments here on our feed using #R211 by 6pm, December 11. For info on the new cars, visit: http://web.mta.info/nyct/R211OpenHouseFeedback.html …	NYCT Subway	Dec 2017	27	21	8	13

The fourth case was again related to public transportation in New York City. This Tweet was made by the New York City subway company in December 2017. As an infrastructural project, NYCT Subway wanted to provide new cars. Under the #R211 hashtag they provided a whole plan for the updates that they planned to implement and asked their audience to give feedback about the project. About 30 users replied to the Tweet to give their opinion about this project plan. Some of the users support and agree with the plan for new cars, and others disagree, preferring to not have these updated cars. The Tweet resulted in many different actions and replies, and some of these reply Tweets are unrelated – for example, people are sarcastic, humorous, rude, etc., without contributing to a productive discussion (Table 23).

Table 23. User Opinions

User	Opinion
A	4.10
D	1.82
E	1.82
H	4.25
I	3.74
J	3.00
K	1.82
M	1.70
P	3.22
Q	4.27
R	4.18
S	4.17
U	2.89
Total pairs	**55**

5.4.1 Correlation Analysis (Conformity and Similarity)

The table below shows the level of similarity and conformity for all combinations of any two selected Twitter users in our final dataset, which consists of the responses of the previously mentioned Twitter users to the Case 4 Tweet. There are 55 total pairs of users (Table 23). The below table contains a few basic descriptive analyses regarding the data (Table 24).

Table 20. Conformity and Similarity

Users	Similarity	Conformity
K,M	0.03	0.12
K,R	0.16	2.36
K,Q	0.05	2.45
K,S	0.01	2.35
K,P	0.02	1.40
K,U	0.15	1.07
H,K	0.09	2.43
H,M	0.20	2.55
H,I	0.33	0.51
H,R	0.04	0.07
H,Q	0.13	0.02
H,S	0.08	0.08
H,P	0.08	1.03
H,U	0.18	1.36
M,R	0.02	2.48
M,Q	0.17	2.57
M,S	0.13	2.47

M,P	0.25	1.52
M,U	0.09	1.19
D,K	0.03	1.92
D,M	0.23	2.04
D,R	0.04	0.44
D,Q	0.14	0.54
D,S	0.10	0.43
D,P	0.09	0.51
D,U	0.12	0.85
I,S	0.02	0.02
I,U	0.07	1.29
E,K	0.06	0.00
E,H	0.04	2.43
E,M	0.02	0.12
E,I	0.06	1.92
E,R	0.07	2.36
E,Q	0.06	2.45
E,S	0.04	2.35
E,P	0.04	1.40
E,U	0.07	1.07
Q,R	0.09	0.09
Q,S	0.13	0.11
Q,U	0.13	1.38
S,U	0.03	1.28
P,R	0.02	0.96
P,Q	0.25	1.05
P,S	0.25	0.94

P,U	0.07	0.33
A,K	0.04	2.28
A,H	0.14	0.15
A,M	0.20	2.40
A,I	0.17	0.36
A,R	0.07	0.08
A,E	0.06	2.28
A,Q	0.50	0.17
A,S	0.17	0.07
A,P	0.33	0.88
A,U	0.10	1.21
Descriptive Analysis		
Mean	0.11	1.20
Median	0.09	1.07
Max	0.50	2.57
Min	0.01	0.00
St Dev	0.09	0.91
Variance	0.01	0.83
t-test and p-value		
	H_0: $\rho = 0$, H_1: $\rho \neq 0$	
α is 0.05, df= n-2	55-2=53	
$t_{0.025,df}$	$t_{0.025,53}$=2.0057	
r	-0.12	
t	7.23 >2.0040	
p-value	0.000<0.01	

Reject H_0 Fail to reject	Rejected H_0	There is evidence of a linear relationship at 5% level of significance between **conformity and similarity**

The p-value for this t-test is <0.01. Given that the p-value falls below our selected significance level of 0.05, we will reject the null hypothesis that there is no evidence of a linear relationship at 5% level of significance between conformity and similarity; we will conclude that the data provides evidence that there is a statistically significant relationship between conformity and similarity (Figure 13).

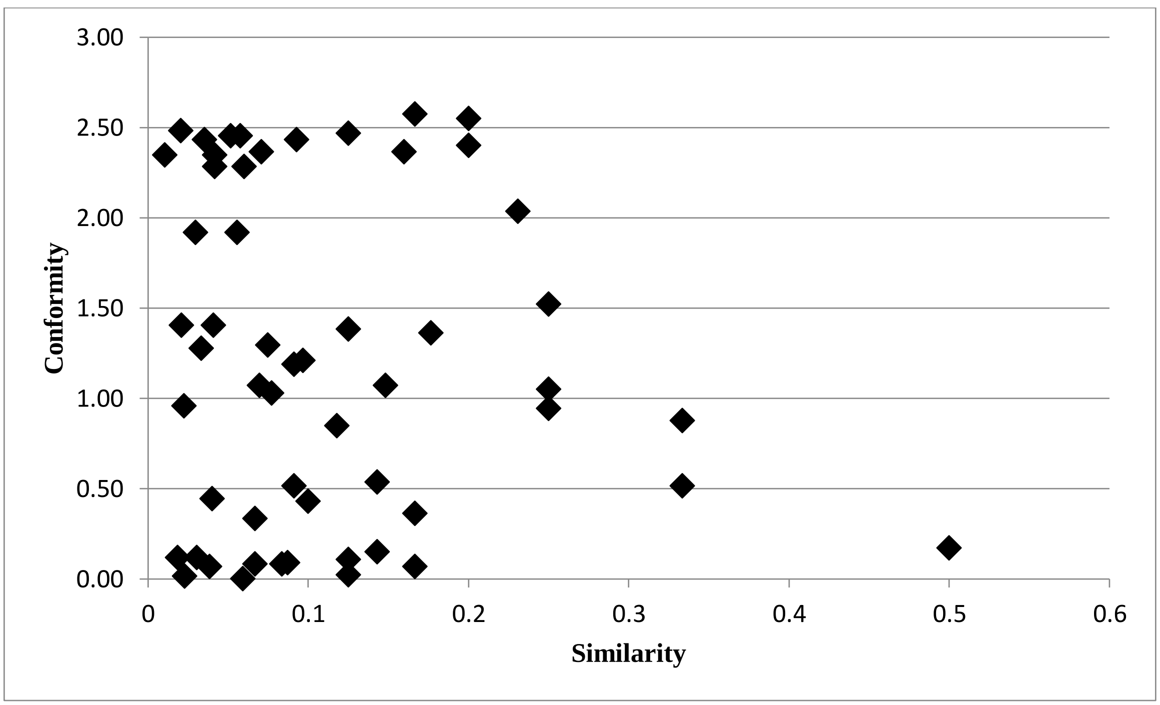

Figure 13. Conformity and Similarity Correlation

In Figure 13, the data points are negatively correlated, and based on the distribution of data points and the correlation coefficient ($r = -0.12$, $t = 7.23$, $p < 0.01$), there appears to be a statistically significant negative relationship between the two factors. Most of the data points have network connections less than 0.25, while the values on the y-axis (answers on how close data points are) are fairly evenly distributed, although they appear to be slightly clustered at y = 0.0, y = 1.2 and y = 2.4. There are very few points with network connections greater than 0.25, and of the points that have network connections greater than 0.25, there are NO answers that have >1.00, which indicates that these two methods agree with each other, and are thus correlated/similar.

5.4.2 Correlation Analysis (Conformity and Average of Opinion)

The following table illustrates the level of average opinion and average opinion for all combinations of any two selected Twitter users in our final dataset, which consists of the responses of the previously mentioned Twitter users to the Case 4 Tweet. There are 55 total pairs of users. The below table contains a few basic descriptive analyses regarding the data (Table 25).

Table 21. Conformity and Average of Opinion Answers

Users	Conformity	Average of opinion
K,M	0.12	1.76
K,R	2.36	3.00
K,Q	2.45	3.05
K,S	2.35	2.99
K,P	1.40	2.52
K,U	1.07	2.35
H,K	2.43	3.03
H,M	2.55	2.98
H,I	0.51	3.99
H,R	0.07	4.22
H,Q	0.02	4.26
H,S	0.08	4.21
H,P	1.03	3.74
H,U	1.36	3.57
M,R	2.48	2.94
M,Q	2.57	2.99
M,S	2.47	2.93
M,P	1.52	2.46
M,U	1.19	2.29
D,K	1.92	2.78
D,M	2.04	2.72
D,R	0.44	3.96
D,Q	0.54	4.00
D,S	0.43	3.95
D,P	0.51	3.48
D,U	0.85	3.31

I,S	0.02	4.17
I,U	1.29	3.54
E,K	0.00	1.82
E,H	2.43	3.03
E,M	0.12	1.76
E,I	1.92	2.78
E,R	2.36	3.00
E,Q	2.45	3.05
E,S	2.35	2.99
E,P	1.40	2.52
E,U	1.07	2.35
Q,R	0.09	4.23
Q,S	0.11	4.22
Q,U	1.38	3.58
S,U	1.28	3.53
P,R	0.96	3.70
P,Q	1.05	3.75
P,S	0.94	3.69
P,U	0.33	3.06
A,K	2.28	2.96
A,H	0.15	4.18
A,M	2.40	2.90
A,I	0.36	3.92
A,R	0.08	4.14

A,E	2.28	2.96
A,Q	0.17	4.19
A,S	0.07	4.13
A,P	0.88	3.66
A,U	1.21	3.49
Descriptive Analysis		
Mean	1.20	3.29
Median	1.07	3.06
Max	2.57	4.26
Min	0.00	1.76
St Dev	0.91	0.69
Variance	0.83	0.47
t-test and p-value		
	H_0: ρ= 0 , H_1: ρ ≠ 0	
α is 0.05, df= n-2	55-2=53	
$t_{0.025,df}$	$t_{0.025,53}$=2.0040	
r	-0.44	
t	5.24 >2.0057	
p-value	0.000<0.01	
Reject H_0 Fail to reject	Rejected H_0	
There is evidence of a linear relationship at 5% level of significance between **conformity and average opinion**		

The p-value for this t-test is <0.01. Given that the p-value falls below our selected significance level of 0.05, we will reject the null hypothesis that there is no evidence of a linear relationship at 5% level of significance between conformity and average opinion; we will conclude that the data provides evidence that there is a statistically significant relationship between conformity and average opinion (Figure 14).

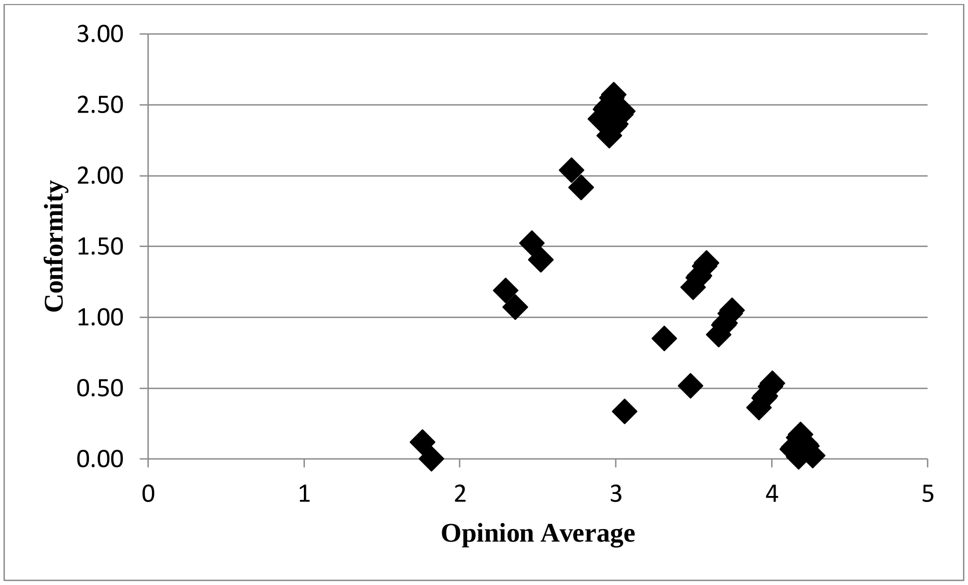

Figure 14. Conformity and Opinion Average Correlation

Figure 14 shows that there is a -0.44 correlation, which is a significant correlation between opinion conformity and average absolute opinion for this dataset, presumably because if two people are in disagreement, their disagreement is equally strong (since negative emotions are usually stronger than positive ones), but if two people are in agreement, then they have a wider range of agreement. It is also worth noting that there are about as many people who agree as people who disagree, so if a user has a low opinion on a particular question, he is about as likely to disagree with a fellow user as if he has a high opinion on a particular question.

The chart's data is pyramid-shaped because if the average of opinion answer is very small/large, then most of the underlying data points are also small/large, so there cannot be much average (answer distance). Conversely, if the average of opinion is medium-sized, then that means that the answers could be grouped together in the middle (i.e., people who do not have strong opinions either way), or the answers could be spread out (i.e., people with opinions all over the map). In other words, our correlation coefficient of -0.44 does not tell the whole story, as there is a non-linear (pyramid) pattern in the data.

5.4.3 Probability Analysis (Bayesian Probability)

The following chart (Bayesian Probability) shows the cumulative probabilities of answers for a given value of network connection, or what the probability is that a specific answer value will be above the threshold (1 or 1.5), given that the network connection is equal to or less than a given point.

In this case:

- B: network connection less than or equal to 0.1
- A: a given answer is greater than or equal to 1.0

For instance, if a given answer is greater than or equal to 1.0 (blue line), the odds that the network connection is less than or equal to 0.2 is about 45%. In other words, about 45% of the answers greater than 1.0 have network connections that are less than or equal to 0.2.

The blue line shows the probability that a given network connection will be less than or equal to 0.2, in the instance that a specific answer is greater than 1.0.

The red line shows the probability that a given network connection will be less than or equal to 0.2, in the instance that a specific answer is greater than 1.5.

Why do we designate answer cutoffs of 1.0 and 1.5? The reasoning is somewhat subjective – we picked these cutoffs because the underlying probabilities for these two cutoffs differ significantly, and between these two cutoffs, the entire range of data is moderately well-represented. These cutoffs help us understand the data without overcomplicating the data set (Figure 15).

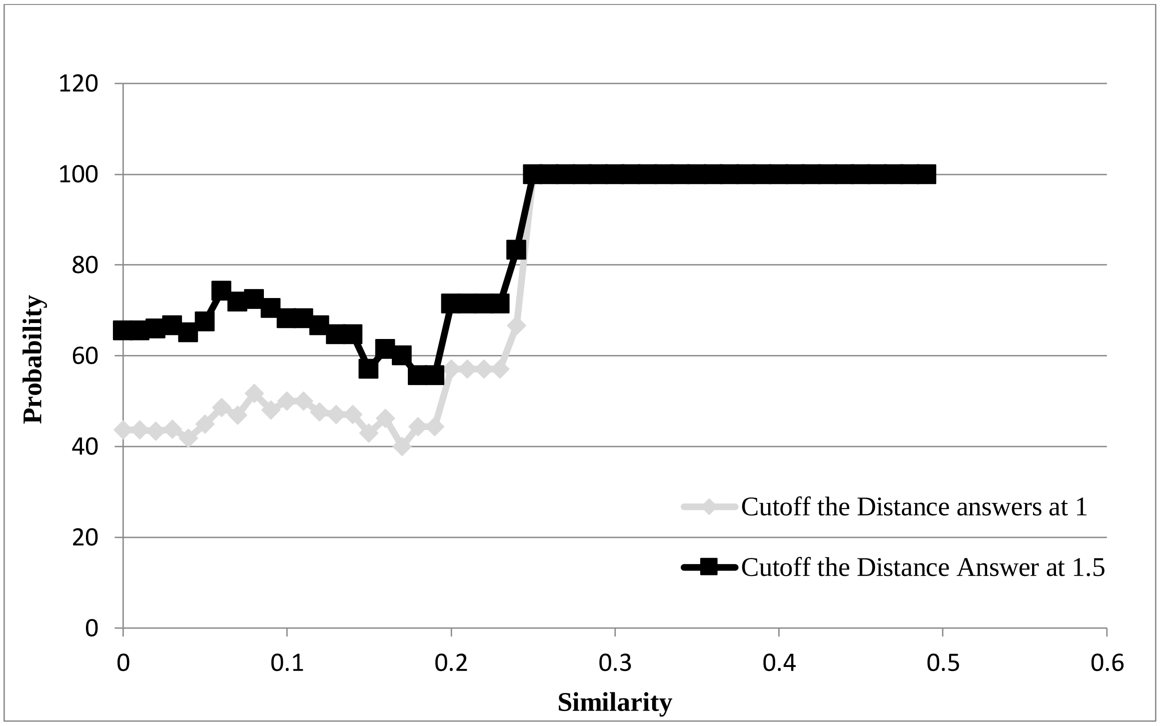

Figure 15. Bayesian Probability

In Figure 15, the range distance between stakeholders' answers is between 0 to 4, where 0 mean is the same and 4 mean is positive opinion. I used Bayesian probability for all cases for the users who have similar answers and a distance less than 1. On the other hand, the probability to have that distance less than 1 in all ties levels from 0 to the highest ties number in the case. I started from 0, 0.1, .02, 0.03 to the highest ties value in the case.

5.4.4 Conclusion

The following analyses have been conducted: two correlation analyses and one Bayesian analysis. All of the analyses show a negative correlation between the two data points, as there is an inverse relationship between how much people agree with the case (or how high their case values are) and how much people agree with each other in the case. We have determined that there is a negative correlation between the conformity and the network connection, and also a negative correlation between the conformity and the average of opinion answer. For the probability, if we have a high network connection or high ties, the probability of getting the same answer is very high after a certain point. It is also worth noting that, even though there are not as many data points in this dataset as there are in other datasets, the patterns in this dataset still resemble the patterns in the other datasets. In conclusion, these data points are correlated, as there is a relationship between all three data points, and the stakeholder's social media analysis is not without its own biases.

5.5 Descriptive Analysis

Table 22. All Cases Descriptive Analysis

Analysis	Description	Case 1	Case 2	Case 3	Case 4
Sample size n	For all variables	117	284	119	55
Correlation Analysis	Conformity and similarity	-0.24	-0.04	-0.15	-0.12
	Conformity and average of opinion	-0.36	-0.02	-0.27	-0.44
t-test	Conformity and similarity	10.4	16.78	10.70	7.23
	Conformity and average of opinion	9.98	16.79	10.41	5.24
p-value	Conformity and similarity	< 0.01	< 0.01	< 0.01	< 0.01
	Conformity and average of opinion	< 0.01	< 0.01	< 0.01	< 0.01
Probability Analysis	The similarity point where probability of conformity starts to become higher after the cutoff of 1 for conformity	0.12	0.17	0.03	0.17
	The similarity point where probability of conformity starts to become higher after the cutoff of 1.5 for conformity	0.12	0.16	0.07	0.19
Bayesian Probability	Starting probability of getting conformity at cutoff 1 for conformity	65	50	99	55.6
	Starting probability of getting conformity at cutoff 1.5 for conformity	80	80	87	44

In all four of these cases, the test statistics (t-tests) are highly significant ($t > 5$, $n > 30$), which indicates that there is strong evidence that the correlations that were calculated above are statistically significant. The correlations are all between -0.50 and 0.00, which indicates that the sample correlations are likely to be moderately impactful. By using the Tanimoto measure and Bayesian probability formula calculations on the four cases in the above table, we found that if there is a similarity of 0.19, the probability of conformity starts to rise sharply. Thus, if there is a similarity greater than 0.20 between two users, it is very likely that those users will share the same opinion.

5.6 Conclusion

Based on this study, there is a bias that comes from our social network similarity. Strong ties can be used to predict the similarity of various responses, which was shown in two different ways: first by linear correlation, and then – to understand whether there is a threshold affect – we performed a Bayesian Analysis to figure out the specific relationship. In our linear correlation analysis, we found a negative linear correlation in every study between strong opinions and similar answers. All of our linear correlations were between -0.50 and -0.20, which indicates that the negative correlations were significant, but not excessively strong.

We found in our Bayesian analysis that beyond a certain threshold of similarity there is a high chance to have a really strong opinion, and we consistently find that the probability of event A (to get the same answer) is more than the probability of A given B (to get the same answer given that there is a tie), therefore B can be used as a predictor for A.

However, the results of the study should not be considered definitive, as there may be other factors that are influencing the results of the analysis that this research did not consider, such as types of transportation that the person has used, the location of the person's residence or workplace, and random sampling error.

In conclusion, using social media as a primary tool for stakeholder analysis can provide biased analysis results. Therefore, we should be cautious when using social media to make conclusions that drive business decisions. On the other hand, a project manager can use the strength of ties to predict user responses at a certain point of strength of ties –the stronger the ties, the better the prediction. Additionally, based on the correlation charts (for example: "Correlation 'Case 1'"), most network connections are strong (network connections cluster around the left/low end of the chart), and the answer distance values are fairly evenly distributed.

CHAPTER VI – CONCLUSION

This chapter provides an overall summary of this research project. I will provide an individual summary for each chapter and make recommendations for future research based on the findings and results from each chapter.

In Chapter 1, stakeholder analysis is defined and discussed: what it is, and why stakeholder analysis is important for industrial projects. Some project processes start by performing stakeholder analysis, as it is seen as the first phase of any public or commercial project. To do that, different methods were provided in three phases for the analysis. Each of these methods has specific strengths and weaknesses.

Chapter 2 provides the biases that were introduced in the current stakeholder analysis methods. Each method has its own bias types and each of them has multiple different types of bias, but biases differ from one method to another. Bias occurs during almost all steps of stakeholder analysis.

In Chapter 3, direct participation is used as a new way of communication, as it minimizes the number of biases in each step because it can be done without direct contact occurring between the users and the researcher. The social media network is one of the world's newest forms of communication, and one of the newest methods for direct participation. This method may reduce some of the biases, but it has its own biases; in this case, the goal of this research is to find if use of the social media platform can invite some bias into the stakeholder analysis. Reasons are also discussed as to why I picked a social media platform – Twitter, specifically – as the focus for my study about social media bias. This chapter also highlights the potential of the Twitter API to be used to collect data.

Chapter 4 outlines the three phases in which we conducted the research in order to achieve a rigorous research design. The Methodology chapter elaborates on the specific research objectives, and why social media can introduce bias into the stakeholder analysis. I provided three variables for this analysis: conformity as the dependent variable, and similarity and average answers of opinion as the two independent variables. We used Bayesian probability calculations and scatterplot charts to compute conditional probabilities, with the resulting probability of conformity being high, given that the answers provided by any two users are similar.

In Chapter 5, I present the detailed results of the research. I used two different analysis methods: the correlation coefficient (to compare two variables) and probability analysis (to compare the similarity and conformity measures). In order to determine if there is a significant chance of similarity, we must first calculate and analyze the conformity. I rejected the null hypothesis that there is not a statistically significant relationship between network similarity with the conformity of opinions during stakeholder analysis process in the social media, and thus accepted the alternative hypothesis that there is a statistically significant relationship.

6.1 Research Validation

In order to ascertain validity, the following validity were conducted: face validity, external validity, and conclusion validity.

6.1.1 Face validity

According to the analysis of variables, more specifically the correlation analysis, this was achieved as the variables seemed to measure the intended target. This statement is supported by the correlation between the variables. The correlation between the variables suggests the high reliability and accuracy of the results derived from the measure of how social media influences opinion.

6.1.2 Content Validity

How well a test measures the intended behavior is assessed by content validity. This is quite an important research methodology term. The collection of opinions randomly from relevant stakeholders on social media and the study of the degree of influence by other users of social media on them all via social media platforms hold much promise to have random data (Rothman, Gnanaskathy, Wicks, & Papadopoulos, 2015). In order to ensure a fair representation of the general population, various random cases were utilized for collecting data. It is worth noting that the advantage of remaining anonymous conferred by online platforms was maintained.

6.1.3 External Validity

The applicability of the results of a study to a population is assessed by the external validity of the study (Trochim, 2000). The generalizability of a study to various groups, times, and settings is based on the external validity. The findings of studies should have an impact and be applicable to other individuals at different settings and at different times (Trochim, (2000) p. 22). In this study, trends or seasonality was eliminated by the randomization of the data set. The four tweets that were used were related to the project management of transport projects within the USA. Public opinion was collected at the first part of the analysis which was prior to the start of the project hence making it valid for that phase of the analysis. Social media relationships and links were used to assess social media similarity. Different results could have been obtained if the similarity was assessed using a different method or other data sets were used such as opinion prediction.

6.1.4 Conclusion Validity

The null hypothesis was rejected by the researcher based on the obtained results.

6.2 Limitations of the Research

Public opinion is an interplay of the dynamic processes which are present within the minds of individuals due to social interactions and communication. The various areas of the self-organization process have been emphasized by social scientists at different levels, with a focus on the contribution of social influence factors to public opinion. The outcome of this area of research has been the elucidation of the many factors which affect opinion (Krueger, Szwabiński, & Weron, 2017), one of which is the strength of ties.

A key question to consider: what makes people declare their opinions publicly? The nature of the question seems to be instrumental in this study. For example, if I publicly asked for people's opinions on a basketball game, they are more forthcoming with their opinions, as this question is relatively less critical and socially significant when compared to a question based on political opinions, such as questions related to the current president. People might be less willing to volunteer their opinions publicly for such controversial questions. The nature of the question helps to determine conformity. However, the presence of conformity can also be due to other factors, such as:

- Self-esteem: people with a low self-esteem tend to yield to group pressure, as compared to their counterparts with a higher self-esteem (Campbell, 1990).

- Cultural differences: generally, people from cultures with a collective mindset are more likely to conform when compared to other individuals from more individualistic cultural backgrounds (Cherry, 2017).
- Task difficulty: difficult tasks can result in people being more likely to conform as a result of not knowing how to perform the task. On the other hand, difficult tasks can also result in people accepting various responses, hence leading to lower conformity. Difficult tasks therefore can lead to both an increase or a decrease in conformity, depending on the situation (Klein, 1972; Rosander & Eriksson, 2012).
- Individual differences: the differences in personal character, such as having the drive to achieve and strong leadership qualities are associated with decreased conformity (Cherry, 2017).
- Characteristics of the situation: The situation also plays a role in how likely people are to conform. People are more likely to conform in situations of uncertainty, compared to situations where they are sure how to respond based on previous experience (Cherry, 2017; Codol, 1975).

6.2 Future Study

6.2.1 Using Similarity to Measure Opinion Dynamics

The study was conducted as a cross-sectional study – data was gathered at a specific time for all users rather than across several points in time. Future studies could produce more thorough datasets if conducted longitudinally, comparing conformity in all users during different periods of time to get a broader picture of the opinion dynamic.

Various theories explain the formation of opinion dynamics. Mathematical and computational models exist which are predominantly used in modeling opinion dynamics. These models, which allow for theoretical and numerical analysis, operate with assumptions that simplify the spreading process to allow a focus of the represented opinions on a wider level.

Communicating mediums such as websites and blogs have been studied with regards to their effects on opinion conformity (Krueger et al., 2017). More focus has been placed on social media in recent years, as it is a popular avenue for communication and it brings people from various backgrounds together. Regarding social media platforms, there are two main schools of thought regarding the dynamics of opinions. One states that people will usually avoid perspectives which are different from theirs, preferring instead to expose themselves to like-minded opinions (Krueger et al., 2017). The inherent features of various social media platforms such as filtering and recommendations further enhance this possibility. Although new information is available to allow people to have well-rounded points of view across various topics, studies have shown that people are more likely to select information sources that are in line with their underlying beliefs and principles.

6.2.2 Using Strength of Ties to Measure Opinion Dynamics

Different ties between people create unique types of similar-thinking groups (Kennedy &

Weimann, 2011). However, these ties could also cause embarrassment for some individuals when expressing their views differently. As a result, an individual tends to think about the surrounding people based on the tie, not his or her personal thoughts on the issue. He might also find it difficult to give opinions that differ from the surrounding peoples' opinions, or from those with whom he has direct or indirect ties. By using these two variables – strength of tie and opinion dynamics – a future study could investigate whether these connections are significant or not.

REFERENCES

Aaltonen, K. (2011). Project stakeholder analysis as an environmental interpretation process. *International Journal of Project Management*, *29*(2), 165–183. https://doi.org/10.1016/j.ijproman.2010.02.001

Aaltonen, K., Jaakko, K., & Tuomas, O. (2008). Stakeholder salience in global projects. *International Journal of Project Management*, *26*(5), 509–516. https://doi.org/10.1016/j.ijproman.2008.05.004

Aaltonen, K., & Sivonen, R. (2009). Response strategies to stakeholder pressures in global projects. *International Journal of Project Management*, *27*(2), 131–141. https://doi.org/10.1016/j.ijproman.2008.09.007

Agarwal, N., Liu, H., Tang, L., & Yu, P. S. (2008). Identifying the influential bloggers in a community. *Proceedings of the 2008 International Conference on Web Search and Data Mining*, 207–218.

Anger, I., & Kittl, C. (2011). Measuring influence on Twitter. In *Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies - i-KNOW '11* (p. 1). https://doi.org/10.1145/2024288.2024326

Baccarini, D. (1999). The Logical Framework Method for Defining Project Success. *Project Management Journal*, *Volume 30*(Issue 4), 25–32. https://doi.org/10.1086/250095

Baird, C. H., & Parasnis, G. (2011). From social media to social customer relationship management. *Strategy and Leadership*. https://doi.org/10.1108/10878571111161507

Bard, G., & Weinstein, Y. (2017). The effect of question order on evaluations of test performance: Can the bias dissolve? *Quarterly Journal of Experimental Psychology*, *70*(10), 2130–2140. https://doi.org/10.1080/17470218.2016.1225108

Beringer, C., Jonas, D., & Kock, A. (2013). Behavior of internal stakeholders in project portfolio management and its impact on success. *International Journal of Project Management*, *31*(6), 830–846. https://doi.org/10.1016/j.ijproman.2012.11.006

Biernacki, P., & Waldorf, D. (1981). e124d58ffe3f519ffe963c3ea8dfd122f44f0d59.pdf. Retrieved from http://ftp.columbia.edu/itc/hs/pubhealth/p8462/misc/biernacki_lect4.pdf

Biggs, S., & Matsaert, H. (1999). An actor-oriented approach for strengthening research and development capabilities in natural resource systems. *Public Administration and Development*, *19*(3), 231–262. https://doi.org/10.1002/(SICI)1099-162X(199908)19:3<231::AID-PAD71>3.0.CO;2-E

Bing Liu. (2010). Sentiment Analysis and Subjectivity. In *Handbook of Natural Language Processing*. https://doi.org/10.1145/1772690.1772756

Boria, R. A., Olson, L. E., Goodman, S. M., & Anderson, R. P. (2014). Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling*, *275*, 73–77. https://doi.org/10.1016/j.ecolmodel.2013.12.012

Bourne, L., & Walker, D. H. T. (2005). Visualising and mapping stakeholder influence. *Management Decision*, *43*(5), 649–660. https://doi.org/10.1108/00251740510597680

Bourne, L., & Walker, D. H. T. (2006). VISUALIZING STAKEHOLDER INFLUENCE—TWO AUSTRALIAN EXAMPLES. *Project Management Journal*, *37*(1), 5–22.

Boyce, C., & Neale, P. (2006). CONDUCTING IN-DEPTH INTERVIEWS : A Guide for Designing and Conducting In-Depth Interviews for Evaluation Input. *Pathfinder International Tool Series*, *Monitoring*(2), 1–12. https://doi.org/10.1080/14616730210154225

Brugha, R. (2000). Stakeholder analysis: a review. *Health Policy and Planning*, *15*, 239–246. https://doi.org/10.1093/heapol/15.3.239
Brugha, R., & Varvasovszky, Z. (2000). Stakeholder analysis: a review. *Health Policy and Planning*, *15*, 239–246. https://doi.org/10.1093/heapol/15.3.239
Bryson, J. M. (2004). What to do when stakeholders matter: a guide to stakeholder identification and analysis technique. *Public Management Review*, (October 2003), 9–11. https://doi.org/10.1080/14719030410001675722
Campbell, J. D. (1990). Self-esteem and clarity of the self-concept. *Journal of Personality and Social Psychology*, *59*(3), 538–549. https://doi.org/10.1037/0022-3514.59.3.538
Castelli, L., Arcuri, L., & Zogmaister, C. (2003). Perceiving ingroup members who use stereotypes: Implicit conformity and similarity. *European Journal of Social Psychology*. https://doi.org/10.1002/ejsp.138
Cheeseman, P. (1983). A method of computing generalized Bayesian probability values for expert systems. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*.
Cherry, K. (2017). What Is Conformity? How Does It Influence Behavior?
Choi, B. C. K., & Pak, A. W. P. (2005). A catalog of biases in questionnaires. *Preventing Chronic Disease*, *2*(1), A13. https://doi.org/A13 [pii]
Cialdini, R. B., & Goldstein, N. J. (2004). Social Influence: Compliance and Conformity. *Annual Review of Psychology*. https://doi.org/10.1146/annurev.psych.55.090902.142015
Clulow, V. (2005). Futures dilemmas for marketers: can stakeholder analysis add value? *European Journal of Marketing*, *39*(9/10), 978–997. https://doi.org/10.1108/03090560510610671
Codol, J. -P. (1975). On the so-called 'superior conformity of the self' behavior: Twenty experimental investigations. *European Journal of Social Psychology*, *5*(4), 457–501. https://doi.org/10.1002/ejsp.2420050404
Cohen, J., Cohen, P., West, S., & Aiken, L. (2014). *Applied Multiple Regression/ Correlation Analysis for the Behavioral Sciences*. *Igarss 2014*. https://doi.org/10.2307/2064799
Crosby, B. (1991). Stakeholder Analysis : A Vital Tool for Strategic Managers. *U.S Agency for International Development*, (2), 1–6. https://doi.org/10.1155/2011/953047
Culnan, M. J., McHugh, P. J., & Zubillaga, J. I. (2010). How large U.S. companies can use Twitter and other social media to gain business value. *MIS Quarterly Executive*. https://doi.org/10.1108/07378830510636300
Currie, R. R., Seaton, S., & Wesley, F. (2009). Determining stakeholders for feasibility analysis. *Annals of Tourism Research*, *36*(1), 41–63. https://doi.org/10.1016/j.annals.2008.10.002
Dale, A. P., & Lane, M. B. (1994). Strategic perspectives analysis: A procedure for participatory and political social impact assessment. *Society & Natural Resources*, *7*(3), 253–267. https://doi.org/10.1080/08941929409380863
Dodd, D. H., & Bradshaw, J. M. (1980). Leading questions and memory: Pragmatic constraints. *Journal of Verbal Learning and Verbal Behavior*, *19*(6), 695–704. https://doi.org/10.1016/S0022-5371(80)90379-5
Dougill, A. j., Fraser, E. D. G., Holden, J., Hubacek, K., Prell, C., Reed, M. S., … Stringer, L. C. (2006). Learning from doing participatory rural research: Lessons from the Peak Distrinct National Park. *Journal of Agricultural Economics*, *57*(2), 259–275. Retrieved from http://homepages.see.leeds.ac.uk/~lecajd/papers/Dougilletal2006relu.pdf
Dryzek, J. S., & Berejikian, J. (1993). Reconstructive Democratic Theory. *The American*

Political Science Review, *87*(1), 48–60. https://doi.org/10.2307/2938955
Elias, A. a, Cavana, R. Y., & Jackson, L. S. (2002). Stakeholder analysis for R & D project management. *R & D Management*, *32*(4), 301–310. https://doi.org/10.1111/1467-9310.00262
Emran, M. S., Greene, W. H., & Shilpi, F. (2015). *When measure matters: coresident sample selection bias in estimating intergenerational mobility in developing countries*. *MPRA*. https://doi.org/10.2139/ssrn.2639343
Fageha, M. K., & Aibinu, A. A. (2013). Managing Project Scope Definition to Improve Stakeholders' Participation and Enhance Project Outcome. *Procedia - Social and Behavioral Sciences*, *74*, 154–164. https://doi.org/10.1016/j.sbspro.2013.03.038
Fischer, R., & Derham, C. (2016). Is in-group bias culture-dependent? A meta-analysis across 18 societies. *SpringerPlus*, *5*(1), 1–9. https://doi.org/10.1186/s40064-015-1663-6
Fraser, E. D. G., Dougill, A. J., Mabee, W. E., Reed, M., & McAlpine, P. (2006). Bottom up and top down: Analysis of participatory processes for sustainability indicator identification as a pathway to community empowerment and sustainable environmental management. *Journal of Environmental Management*, *78*(2), 114–127. https://doi.org/10.1016/j.jenvman.2005.04.009
Freitas, H., Oliveira, M., Jenkins, M., & Popjoy, O. (1998). The Focus Group, A Qualitative Research Method. *Isrc*, (010298), 1–22. https://doi.org/10.1016/j.watres.2013.02.032
Gao, S. S., & Zhang, J. J. (2006). Stakeholder engagement, social auditing and corporate sustainability. *Business Process Management Journal*, *12*(6), 722–740. https://doi.org/10.1108/14637150610710891
Giles, H., & Oxford, G. S. (1970). Towards a multidimensional theory of laughter causation and its social implications. *Bulletin of the British Psychological Society*.
Grimble, R. (1998). SOCIO-ECONOMIC METHODOLOGIES BEST PRACTICE GUIDELINES STAKEHOLDER METHODOLOGIES IN NATURAL RESOURCE MANAGEMENT Robin Grimble. *Practice*, 12. Retrieved from http://www.nri.org/publications/bpg/bpg02.pdf
Hardy, A., Wickham, M., & Gretzel, U. (2013). Neglected stakeholder groups: Conceptualising a dynamic model for neglected stakeholder analysis and engagement. *Corporate Ownership and Control*, *11*(1 D), 348–359.
Hare, M., & Pahl-Wostl, C. (2002). Stakeholder Categorisation in Participatory Integrated Assessment Processes. *Integrated Assessment*, *3*(1), 50–62. https://doi.org/10.1076/iaij.3.1.50.7408
Harrell, M., & Bradley, M. (2009). *Data Collection Methods Semi-Structured Interviews and Focus Groups*. *Distribution*. https://doi.org/978-0-8330-4889-9
Hart, S. L. (2016). Engaging fringe stakeholders, (August 2015). Retrieved from https://www.researchgate.net/profile/Sanjay_Sharma75/publication/280805409_AME_04/links/55c7c56708aeca747d667d1d.pdf
Hernandez, I., & Preston, J. L. (2013). Disfluency disrupts the confirmation bias. *Journal of Experimental Social Psychology*, *49*(1), 178–182. https://doi.org/10.1016/j.jesp.2012.08.010
Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). Telephone versus face-to-face interviewing of national probability samples with long questionnaires. *Public Opinion Quarterly*, *67*(1), 79–125. https://doi.org/10.1086/346010
Jackson, A., & Greene, R. L. (2017). Question order bias in retrospective evaluations of item and associative recognition. *Memory*, *25*(4), 481–486.

https://doi.org/10.1080/09658211.2016.1188964

Jaeger, M. (1997). Relational bayesian networks. *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*. https://doi.org/10.1145/1273496.1273543

Jepsen, A. L., & Eskerod, P. (2009). Stakeholder analysis in projects: Challenges in using current guidelines in the real world. *International Journal of Project Management*, *27*(4), 335–343. https://doi.org/10.1016/j.ijproman.2008.04.002

Johansson, P. (2008). Implementing stakeholder management: a case study at a micro-enterprise. *Measuring Business Excellence*, *12*(3), 33–41. https://doi.org/10.1108/13683040810900386

Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology*, *80*(4), 557–571. https://doi.org/10.1037/0022-3514.80.4.557

Juni, P., Altman, D. G., & Egger, M. (2001). Systematic reviews in health care: Assessing the quality of controlled clinical trials. *Bmj*, *323*(7303), 42–46. https://doi.org/10.1136/bmj.323.7303.42

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

Kennedy, J., & Weimann, G. (2011). The strength of Weak Terrorist Ties. *Terrorism and Political Violence*. https://doi.org/10.1080/09546553.2010.521087

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*. https://doi.org/10.1016/j.bushor.2011.01.005

Kirkman, B. L., & Gibson, C. B. (2004). THE IMPACT OF TEAM EMPOWERMENT ON VIRTUAL TEAM PERFORMANCE : THE MODERATING ROLE OF FACE-TO-FACE INTERACTION University of North Carolina at Chapel Hill, *47*(2), 175–192. Retrieved from http://cor.web.uci.edu/files/2013/02/Gibson_et_al_AMJ.pdf

Kitzinger, J. (1995). Qualitative research: Introducing focus groups. *BMJ : British Medical Journal*, *311*(7000), 299–302. https://doi.org/10.1136/bmj.311.7000.299

Klein, R. L. (1972). Age, sex, and task difficulty as predictors of social conformity. *Journal of Gerontology*, *27*(2), 229–236. https://doi.org/10.1093/geronj/27.2.229

Knobloch-Westerwick, S., Johnson, B. K., & Westerwick, A. (2015). Confirmation bias in online searches: Impacts of selective exposure before an election on political attitude strength and shifts. *Journal of Computer-Mediated Communication*, *20*(2), 171–187. https://doi.org/10.1111/jcc4.12105

Krueger, T., Szwabiński, J., & Weron, T. (2017). Conformity, anticonformity and polarization of opinions: Insights from a mathematical model of opinion dynamics. *Entropy*, *19*(7). https://doi.org/10.3390/e19070371

Lacey Ganser, C. Z. (2006). The Effects of Perceived Similarity on Conformity in Non-Humorous Situations. *UW-L Journal of Undergraduate Research IX*.

Lambers, J. (2009). *Three-Dimensional Coordinate Systems*.

Levine, L. J., & Safer, M. a. (2002). Sources of Bias in Memory for Emotions. *Current Directions in Psychological Science*, *11*, 169–173. https://doi.org/10.1111/1467-8721.00193

Lewis, S. (2015). Qualitative Inquiry and Research Design: Choosing Among Five Approaches. *Health Promotion Practice*. https://doi.org/10.1177/1524839915580941

Louise Barriball, K., & While, A. (1994). Collecting data using a semi???structured interview: a discussion paper. *Journal of Advanced Nursing*, *19*(2), 328–335. https://doi.org/10.1111/j.1365-2648.1994.tb01088.x

Lourenço, F., Lobo, V., & Bação, F. (2004). Binary-based similarity measures for categorical data and their application in Self- Organizing Maps 1 Introduction. *Measurement*.

Marshall, C., & Rossman, G. B. (2015). *Designing Qualitative Research*. *PhD Proposal* (Vol. 1). https://doi.org/10.1017/CBO9781107415324.004

Matell, M. S., & Jacoby, J. (1972). Is there an optimal number of alternatives for Likert-scale items? Effects of testing time and scale properties. *Journal of Applied Psychology*. https://doi.org/10.1037/h0033601

Mathern, B., Bellet, T., & Mille, A. (2010). No An Iterative Approach to Develop a Cognitive Model of the Driver for Human Centred Design of ITS., 85–95.

Mitchell, R. K., Agle, B. R., & Wood, D. J. (1997). Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts. *Academy of Management Review*, *22*(4), 853–886. https://doi.org/10.5465/AMR.1997.9711022105

Moses, S. N., Villate, C., Binns, M. A., Davidson, P. S. R., & Ryan, J. D. (2008). Cognitive integrity predicts transitive inference performance bias and success. *Neuropsychologia*, *46*(5), 1314–1325. https://doi.org/10.1016/j.neuropsychologia.2007.12.009

Mukaka, M. M. (2012). Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal*, *24*(3), 69–71. https://doi.org/10.1016/j.cmpb.2016.01.020

Mustajoki, J., Hämäläinen, R. P., & Marttunen, M. (2004). Participatory multicriteria decision analysis with Web-HIPRE: A case of lake regulation policy. *Environmental Modelling and Software*, *19*(6), 537–547. https://doi.org/10.1016/j.envsoft.2003.07.002

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology*, *29*(1), 85–95. https://doi.org/10.1080/00335557743000053

Newcombe, R. (2003). From client to project stakeholders: a stakeholder mapping approach." Construction Management and Economics, (September 1999), 5–10. Retrieved from http://www.irbnet.de/daten/iconda/CIB3398.pdf

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

Nissen, M. E., & Levitt, R. E. (2004). Agent-based modeling of knowledge dynamics. *Knowledge Management Research & Practice*, *2*(3), 169–183. https://doi.org/10.1057/palgrave.kmrp.8500039

Noy, C. (2008). Sampling Knowledge: The Hermeneutics of Snowball Sampling in Qualitative Research. *International Journal of Social Research Methodology*, *11*(4), 327–344. https://doi.org/10.1080/13645570701401305

ODA. (1995). Guidance Not On How To Do Stakeholder Analysis of Aid Projects and Programmers. *Overseas Development Administration*.

OECD. (2001). Citizens as Partners: Information, Consultation and Public Participation in Policy-Making. *OECD Handbook*, *20*(424473), 163–178. https://doi.org/10.1177/0963662509336713

Oswald, M. E., & Grosjean, S. (2004). Confirmation bias. *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgement and Memory*, (August), 79–96. https://doi.org/10.13140/2.1.2068.0641

Pannucci, C. J., & Wilkins, E. G. (2010). Identifying and Avoiding Bias in Research. *Plastic and*

Reconstructive Surgery, *126*(2), 619–625. https://doi.org/10.1097/PRS.0b013e3181de24bc

Panzeri, S., Senatore, R., Montemurro, M. A., & Petersen, R. S. (2007). Correcting for the Sampling Bias Problem in Spike Train Information Measures. *Journal of Neurophysiology*, *98*(3), 1064–1072. https://doi.org/10.1152/jn.00559.2007

Pierce, R. (2017). Correlation. Retrieved from http://www.mathsisfun.com/data/correlation.html

Prell, C., Hubacek, K., & Reed, M. (2007). Stakeholder Analysis and Social Network Analysis in Natural Resource Management. *Society and Natural Resources*, *22*(6), 501–518. https://doi.org/10.1080/08941920802199202

Project Management Institute. (2013). *A guide to the project management body of knowledge (PMBOK ® guide)*. *Project Management Institute*. https://doi.org/10.1002/pmj.20125

Raschka, S. (2015). *Python Machine Learning*. *Igarss 2014*. https://doi.org/10.1007/s13398-014-0173-7.2

Reed, M. S., Graves, A., Dandy, N., Posthumus, H., Hubacek, K., Morris, J., … Stringer, L. C. (2009a). Who’s in and why? A typology of stakeholder analysis methods for natural resource management. *Journal of Environmental Management*, *90*(5), 1933–1949. https://doi.org/10.1016/j.jenvman.2009.01.001

Reed, M. S., Graves, A., Dandy, N., Posthumus, H., Hubacek, K., Morris, J., … Stringer, L. C. (2009b). Who’s in and why? A typology of stakeholder analysis methods for natural resource management. *Journal of Environmental Management*, *90*(5), 1933–1949. https://doi.org/10.1016/j.jenvman.2009.01.001

Richards, C., Blackstock, K.L. e Carter, C. E. (2004). Practical Approaches to Participation. *SERG Policy Brief No. 1;*, (1), 23. https://doi.org/ISBN 0-7084-0661-0

Riquelme, F., & González-Cantergiani, P. (2016). Measuring user influence on Twitter: A survey. *Information Processing and Management*, *52*(5), 949–975. https://doi.org/10.1016/j.ipm.2016.04.003

Rosander, M., & Eriksson, O. (2012). Conformity on the Internet - The role of task difficulty and gender differences. *Computers in Human Behavior*, *28*(5), 1587–1595. https://doi.org/10.1016/j.chb.2012.03.023

Rothman, M., Gnanaskathy, A., Wicks, P., & Papadopoulos, E. J. (2015). Can we use social media to support content validity of patient-reported outcome instruments in medical product development? *Value in Health*. https://doi.org/10.1016/j.jval.2014.10.001

Rowley, T., & Moldoveanu, M. (2003). When Will Stakeholder Groups Act ? *The Academy of Management Review*, *28*(2), 204–219. Retrieved from https://www.researchgate.net/profile/Mihnea_Moldoveanu/publication/235418948_When_Will_Stakeholder_Groups_Act/links/02bfe5117b362b4229000000.pdf

Sekaran, U., & Bougie, R. (2016). Research Methods for Business: A Skill Building Approach. In *Research methods for business* (p. 436).

Sharp, H., Finkelstein, a., & Galal, G. (1999). Stakeholder identification in the requirements engineering process. *Proceedings. Tenth International Workshop on Database and Expert Systems Applications. DEXA 99*, 1–5. https://doi.org/10.1109/DEXA.1999.795198

Signor, P. P. W., & Lipps, J. J. H. (1982). Sampling bias, gradual extinction patterns and catastrophes in the fossil record. *Geological Society of America Special Papers*, *Mass Extin*(Special Paper 190), 291–296. https://doi.org/10.1130/SPE190-p291

Suphakit Niwattanakul*, Jatsada Singthongchai, E. N. and S. W. (2013). Using of Jaccard Coefficient for Keywords Similarity. *Proceedings of the International MultiConference of Engineers and Computer Scientists, Hong Kong*. https://doi.org/ISBN 978-988-19251-8-3

Taylor, M., Kent, M. L., & White, W. J. (2001). How activist organizations are using the Internet to build relationships. *Public Relations Review, 27*(3), 263–284. https://doi.org/10.1016/S0363-8111(01)00086-8

Tracy, S. J. (2010). Qualitative Quality: Eight “Big-Tent” Criteria for Excellent Qualitative Research. *Qualitative Inquiry*, *16*(10), 837–851. https://doi.org/10.1177/1077800410383121

van den Akker, J. (1999). Principles and Methods of Development Research. In *Design Approaches and Tools in Education and Training* (pp. 1–14). https://doi.org/10.1007/978-94-011-4255-7

Vicente, R., Martins, A. C. R., & Caticha, N. (2009). Opinion dynamics of learning agents: Does seeking consensus lead to disagreement? *Journal of Statistical Mechanics: Theory and Experiment*. https://doi.org/10.1088/1742-5468/2009/03/P03015

Warkentin, M. E., Sayeed, L., & Hightower, R. (1997). Virtual Teams versus Face-to-Face Teams: An Exploratory Study of a Web-based Conference System. *Decision Sciences*, *28*(4), 975–996. https://doi.org/10.1111/j.1540-5915.1997.tb01338.x

Warner, J. F. (2006). More sustainable participation? Multi-Stakeholder Platforms for integrated catchment management. *International Journal of Water Resources Development*, *22*(1), 15–35. https://doi.org/10.1080/07900620500404992

Wasserman, S., & Faust, K. (1994). *Social network analysis : methods and applications*. *American Ethnologist* (Vol. 24). https://doi.org/10.1525/ae.1997.24.1.219

Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications. *Cambridge University Press*, *1*, 116. https://doi.org/10.1525/ae.1997.24.1.219

Waters, R. D., Burnett, E., Lamm, A., & Lucas, J. (2009). Engaging stakeholders through social networking: How nonprofit organizations are using Facebook. *Public Relations Review*, *35*(2), 102–106. https://doi.org/10.1016/j.pubrev.2009.01.006

Weng, J., Lim, E. P., Jiang, J., & He, Q. (2010). Twitterrank: Finding topic-sensitive influential twitterers. *Proceedings of the 3rd ACM International Conference on Web Search and Data Mining (WSDM 2010)*, 261–270. https://doi.org/10.1145/1718487.1718520

Yang, C. H., Motohashi, K., & Chen, J. R. (2009). Are new technology-based firms located on science parks really more innovative?. Evidence from Taiwan. *Research Policy*, *38*(1), 77–85. https://doi.org/10.1016/j.respol.2008.09.001

Yang, J., Shen, G., Bourne, L., Ho, C., & Xue, X. (2011). This is the Pre-Published Version . A Typology of Operational Approaches for Stakeholder Analysis and Engagement for Practitioners, *29*, 145–162. Retrieved from http://ira.lib.polyu.edu.hk/bitstream/10397/6807/1/Yang_Typology_Operational_Stakeholder.pdf

Zhaxybayeva, O., & Gogarten, J. P. (2002). Boostrap, Bayesian probability and maximum likelihood mapping: Exploring new tools for comparative genome analyses. *BMC Genomics*. https://doi.org/10.1186/1471-2164-3-4

APPENDICES

Appendix A – Statistical Hypothesis Testing

The 6 steps process to statistically test my hypothesis
Step 1: I will state the null hypothesis and the alternative hypothesis H_0: ρ= 0 , H_1: ρ≠ 0
Step 2: Set α The typical value of α is 0.05 the level of significance df= 4-2=2 $t_{0.025,df}$= $t_{0.025,2}$=4.3027
Step 3: Collect Data
Individual Influence Analysis
Step 4: Calculate a test statistic $t=(r- \rho)/(1-r^2/(n-2))^{1/2}$ r= sample correlation coefficient = 0.1 ρ = Population correlation coefficient n= The sample size $t=0.1/((1-(0.1)^2/(4-2))^{1/2}= 0.142$
Step 5: Construct Acceptance / Rejection regions t value= 0.142 t > 4.3027 H_0 Rejected t < 4.3027 H_0 Fail to reject
Step 6: Based on steps 5 and 6, draw a conclusion about H_0 ○ Reject or fail to reject the null hypothesis? ○ The null hypothesis ○ The alternative hypotheses

Appendix B - Python "Code 1" (Mamadou Seck, 2018)

```
# -*- coding: utf-8 -*-

"""

Created on Thu Sep 20 12:13:17 2018



@author: Mamadou

"""



# -*- coding: utf-8 -*-
"""
Created on Fri Mar 16 22:25:56 2018

@author: Mamadou
"""
import csv

users_to_test = []
with open('Case 1.csv', 'rU') as csvfile:

    n = 0

    spamreader = csv.reader(csvfile)
    for row in spamreader:
        #print row[0]
        users_to_test.append(row[0])

print users_to_test
famous_accounts = []

with open('Most popular Twitter accounts.csv', 'rU') as csvfile:

    n = 0

    spamreader = csv.reader(csvfile)
    for row in spamreader:
        #print row[0]
```

```
    famous_accounts.append(row[0])

print famous_accounts
import tweepy
import os

# Consumer keys and access tokens, used for OAuth
consumer_key       = '----------------'
consumer_secret     = '-----------------'
access_token        = '-----------------
access_token_secret = '---------------------'



# OAuth process, using the keys and tokens
auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)

# Creation of the actual interface, using authentication
api = tweepy.API(auth, wait_on_rate_limit = True)

def get_friends(user_id):
  users = []
  page_count = 0
  for user in tweepy.Cursor(api.friends, id=user_id, count=200).pages():
    page_count += 1
    print 'Getting page {} for friends'.format(page_count)
    users.extend(user)
  return users

#users = get_friends('mdseck')


dict_users_to_test = {}


for element in users_to_test:
  percent = 100 * users_to_test.index(element)/len(users_to_test)
  print "------------- %s " %percent
  friends = get_friends(element)
  dict_users_to_test.update({element : [f.screen_name for f in friends]})

print dict_users_to_test
```

```
for u in dict_users_to_test.keys():
   list_to_remove = []
   for i in dict_users_to_test[u]:
      if "@"+i not in famous_accounts:
         list_to_remove.append(i)
   for i in list_to_remove:
      dict_users_to_test[u].remove(i)

for i in dict_users_to_test:
   print "-----------Friends of %s" %i
   print dict_users_to_test[i]


dict_similarity = {}
for i in users_to_test:
   for j in users_to_test:
      if i != j and j+'_'+i not in dict_similarity.keys() :
         dict_similarity.update ({i+'_'+j : 0})
         for k in famous_accounts:
            if k[1:] in dict_users_to_test[i] and k[1:] in dict_users_to_test[j]:
               dict_similarity [i+'_'+j] += 1
            elif k[1:] not in dict_users_to_test[i] and k[1:] not in dict_users_to_test[j]:
               dict_similarity [i+'_'+j] += .01
            else:
               pass
               #dict_similarity [i+'_'+j] += -.


with open('similarity_Case_1.csv', mode='w') as similarity_file:
   similarity_writer = csv.writer(similarity_file, delimiter=',', quotechar='"',
lineterminator = '\n', quoting=csv.QUOTE_MINIMAL)

   for i in dict_similarity.keys():
      similarity_writer.writerow([i,dict_similarity[i]])


with open ('caseoneCase_1_followers.csv', mode='w') as case_file:
   case_writer = csv.writer(case_file, delimiter = ',', quotechar = '"', lineterminator = '\n')
   for i in dict_users_to_test.keys():
      a = []
      a.append(i)
      a.extend(dict_users_to_test[i])
      case_writer.writerow (a)
```

Appendix C – Python "Code 2" (Mamadou Seck, 2018)

```
# -*- coding: utf-8 -*-
"""
Created on Thu Sep 20 12:13:17 2018

@author: Mamadou
"""


# -*- coding: utf-8 -*-
"""
Created on Fri Mar 16 22:25:56 2018

@author: Mamadou
"""
import csv

ordered_list = []
with open('Case 6.csv', 'rU') as csvfile:

    n = 0

    spamreader = csv.reader(csvfile)
    for row in spamreader:
        #print row[0]
        ordered_list.append (row[0])

user_followers = {}
with open('caseoneCase_6new_followers.csv', 'rU') as csvfile:

    n = 0

    spamreader = csv.reader(csvfile)
    for row in spamreader:
        print len (row)
        if len(row) > -1:
            #print "TOO MANY FRIENDS"
            user_followers.update ({row[0]:row[1:]})
        else:
            ordered_list.remove(row[0])
            print row[0]


famous_accounts = []
with open('Most popular Twitter accounts.csv', 'rU') as csvfile:

    n = 0

    spamreader = csv.reader(csvfile)
    for row in spamreader:
        #print row[0]
        famous_accounts.append(row[0])
```

```
for u in user_followers.keys():
    list_to_remove = []
    for i in user_followers[u]:
        if "@"+i not in famous_accounts:
            list_to_remove.append(i)
    for i in list_to_remove:
        user_followers[u].remove(i)
keep = 0
delete = 0
for u in user_followers.keys():

    print user_followers[u]
    if len (user_followers[u] ) < 10:
        #print "DELETE"
        delete += 1
    else:
        #print "KEEP"
        keep += 1
print "KEEP %s" %keep
print "DELETE %s" %delete


dict_similarity = {}
for i in ordered_list:
    for j in ordered_list:
        if i != j and j+'_'+i not in dict_similarity.keys() :
            dict_similarity.update ({i+'_'+j : 0})
            a = 0
            for k in famous_accounts:
                if k[1:] in user_followers[i] and k[1:] in user_followers[j]:
                    a += 1
#             print a
#
#             print len(user_followers[i])
#             print len(user_followers[j])
#             print "\n"
            if (len(user_followers[i]) + len(user_followers[j]) - a) != 0:
                sim = a / (1.0*(len(user_followers[i]) + len(user_followers[j]) - a))
                dict_similarity.update ({i+'_'+j : sim})
            else:
                dict_similarity.update ({i+'_'+j : 0})

#for i in  dict_similarity:
#    print "%s : %s" %(i , dict_similarity[i])


with open('Similarty_6new_followers.csv', mode='w') as similarity_file:
    similarity_writer = csv.writer(similarity_file, delimiter=',', quotechar='"', lineterminator = '\n',
quoting=csv.QUOTE_MINIMAL)

    for i in dict_similarity.keys():
        similarity_writer.writerow([i,dict_similarity[i]])
```

VITA

Ahmad Bajarwan was born on December 22, 1979 in Settle, USA. He graduated with a bachelor of science degree in civil engineering from King Abdullaziz University, Jeddah, Saudi Arabia in 2005. He worked as a Field Engineer for Fluor Arabia Limited for six years regarding the Abdul-Aziz University Project in Saudi, Arabia. In 2013, he graduated with Master of Science degree in Construction Management from Stevens Institute of Technology. He came to Old Dominion University in 2014 to do his PhD degree in Engineering Management. He is planning to work as a teaching assistant for engineering management.