8-2017

Effects of Gatekeeping on the Diffusion of Information

Rebecca Anderson
University of Tennessee, Knoxville, rander32@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Rebecca Anderson entitled "Effects of Gatekeeping on the Diffusion of Information." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Communication and Information.

Suzie Allard, Major Professor

We have read this dissertation and recommend its acceptance:

Michael Kotowski, Bruce MacLennan, Carol Tenopir

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Effects of Gatekeeping on the Diffusion of Information

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Rebecca Anderson
August 2017

DEDICATION

To my parents, without whom, I never could.

To my brother, without whom, I never would.

To my husband, because of whom, I did.

ABSTRACT

This study proposes a theoretical model of information diffusion using the conceptual framework of Gatekeeping Theory (Shoemaker & Vos, 2009). Diffusion is a process by which elements are distributed through a social system (Rogers, 2003; Kadushin, 2012). This model builds on previous diffusion research and incorporates constructs of authority and vivid information, novel to the domain. To test the fit of the model, Twitter data derived using data mining techniques are utilized. Specifically, messages posted to Twitter relating to the 2013 Consumer Electronics (CES) conference are mined. Essentially, this study focuses on the diffusion of technology information through a popular social medium, Twitter. From these messages, the network was be visualized and diffusion paths were determined using network analysis. A test of the model was conducted to determine fit using structural equation modeling.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION AND GENERAL INFORMATION

Diffusion, as it relates to society, is the process by which elements are distributed through social systems (Bass, 1969; Rogers, 2003; Kadushin, 2012). Elements can be tangible or intangible, such as artwork, technology, behavior, information, and even disease. The study of the diffusion process can reveal the evolution of culture within a social system. Cosmides and Tooby (1989) define the study of culture as "…how the different kinds of information from each individual's environment, especially from his or her social environment, can be expected to affect that individual's behavior" (pg. 51). Essentially, culture is the aggregation of generations of social information that affects an individual's behavior. So, observing the diffusion process over time reveals the convergence and divergence of the culture of a social system.

At the heart of this process is the social network construct. The study of social networks stems from graph theory, which utilizes mathematical structures to model relationships between objects from some source or pool (Tutte, 2001). In physical sciences networks emerge through the interactions of particles, such as Bose-Einstein condensation (Bianconi &Barabasi, 2001). In computer sciences networks emerge through the connectivity of machines, such as local area networks and the Internet (Wellan, Salaff, Dimitrova, Garton, Gulia, Haythornthwaite, 1996; Chong & Kumar, 2003). In biological sciences networks emerge through the structure and linkage of cells and/or organs, such as metabolic networks and gene regulatory networks (Wheelock, Wheelock, Kawashima, Diez, Kanehisa, van Erk, Kleeman, Haeggstrom, & Goto, 2009). In social sciences networks, emerge through the relationships between humans and/or organizations, such as social networks (White, Boorman, & Breiger, 1976).

In a social network, the variables of interest are actors, often people or organizations, connected by social relationships, such as friendship, romance, acquaintance, colleague, etc (Haythornthwaite, 1996). The advantage of taking a network approach in the diffusion process is that allows one to observe the structure of the social environment of interacting units. As Kadushin (2012) describes, some element may be diffused in one of three ways: 1) contact with one or more influential or persuasive actors; 2) imitation of an actor with whom there is direct contact; 3) imitation of actor with whom there is not direct contact. In order for diffusion to occur, some decision or action is required on the part of the receiving actor.

Social Network Analysis

Social networks are studied using social network analysis (SNA) (Wasserman & Faust, 2008). In SNA the actors are referred to as nodes and the relationships are referred to as links (Haythornthwaite, 1996). Further, there are four main relational concepts, according to Wasserman and Faust (2008):

1. Actors and actors' behaviors are viewed as interdependent.
2. Linkages between actors are viewed as channels by which resources are exchanged.
3. At the actor level, network models view the structural environment as providing opportunities or constraints on the actor.
4. Network structure is conceptualized as lasting patterns of relationships among actors.

Inherent in the name, the unit of analysis in SNA is the network itself. As such, SNA posits specific relational patterns and network structures (Otte & Rousseau, 2002). This differs from other social science methods where the unit of analysis is the individual actor. Essentially, SNA focuses on the characteristics of the relational patterns that create the structure of the network.

Relationships are measured in two ways: directionality and strength (Haythornthwaite, 1996; Otte & Rousseau, 2002; Wasserman & Faust, 2008). Directionality is a reference to a type of relationship. Haythornthwaite (1996, pg. 326) “when information is passed, it flows in a certain direction.” The directionality of a relationship can be symmetrical or asymmetrical. Symmetrical relationships occur when there is a flow of network resources between both connecting actors, while asymmetrical relationships occur when resources flow only one way between connecting actors. Strength is a reference to the intensity of the relationship (Haythornthwaite, 1996). For instance, relationships in which actors frequently exchange resources have a higher intensity than relationships in which actors rarely exchange resources.

Network structure in defined by cohesion (Otte & Rousseau, 2002) or how closely knit the network is. Cohesion is measured in three ways: density, centrality, and cliques (Otte & Rousseau, 2002; Wasserman & Faust, 2008). Density is a measure of the interconnectivity of the network (Otte & Rousseau, 2002). Essentially, a network in which all or most actors are connected is considered high density, while a network in which few actors are connected is considered low density. Centrality is a measure of the number of links for any given network actor (Wasserman & Faust, 2008). Therefore, an actor with a large amount of network connections is considered more central to the network than an actor with a small amount of network connections. Individual centrality can, then, be aggregated to define a global measure of cohesion for the social network (Otte & Rousseau, 2002). Cliques are a measure of subgroups within a social network (Wasserman & Faust, 2008). Cliques are highly connected network actors within a larger network of actors. In other words, they can be considered local social networks with a global network.

Social Network Analysis and Diffusion

Based on these concepts, SNA allows one to observe diffusion for numerous tangible and non-tangible social elements. Tangible social elements include pottery, clothing, tools, etc. and are often observed in anthropological research (Bentley & Shennan, 2003). Intangible elements include verbal communication, non-verbal physical behaviors, digital data, etc. and are observed in multifarious fields of research (Haythornthwaite, 1996; Wellman, Salaff, Dimitrova, Garton, Gulia, Haythornthwaite, 1996; Reagans & McEvily, 2003). Specifically, it allows one to observe the convergence and divergence of ideas, behaviors, and cultures. Convergence and divergence relate to the patterns of diversification within a social group (Langerhans & DeWitt, 2004). From a social network perspective, convergence is observed at the point at which two or more local social networks within a larger social network adopt a behavior, are privy to a rumor, or purchase a new technology. Conversely, divergence is observed at the point at which one or more local social networks within a larger social network fail to adopt a behavior, are not privy to a rumor, or fail to purchase a new technology.

For example, using SNA one could follow the evolution of a technological idea, such as social networking websites, from initial concept, to the launch of Friendster, the fall of Myspace, and the public offering of Facebook. A further example, using SNA could allow one to observe the rise and fall of the parachute pants trend by following the patterns of relationships as actors in a social system either adopted the fashion choice or neglected to adopt the fashion choice. As an example of the diffusion of intangible elements, SNA also allows one to observe the beginning and end of a social revolution, as well as the influential actors involved in the process. For instance, SNA could be used from an historical perspective to understand movements such as civil right or women's right by tracking patterns of information exchange through the linkages of actors. In the same way, the formation of terrorist cells can be observed, as well as the flow of resources, such as information, money, weapons, etc. Observing the pattern of linkages over time provides insight into types of connections and the utility of connections between various network actors. For instance, one could observe that there are a few actors with more linkages than the rest and resources tend to flow through these actors, making them influential in the network. Also, observing the pattern of linkages over time allows one to detect the point at which connections between actors are made and when they dissolve. This provides insight into the evolution of social groups.

Not only is the study of diffusion necessary to observe the evolution of social systems and cultures, but knowledge of the process within specific networks allows one to strategically place actors in the network to disseminate, regulate, or prevent the spread of the information. Burt (1992) refers to this type of actor as a bridge, while Shoemaker and Vos (2009) refer to this type of actor as a gatekeeper. Essentially, this actor is needed to allow novel information to flow into a network, or keep novel information out of a network (Burt, 1992). In other words, the gatekeeper is the control agent of the network.

While the study of diffusion is of theoretical and practical importance, Schnettler (2009) points out that the diffusion process of social networks is not well understood. While this process has been theorized (Bass, 1969; Rogers, 2003), there are challenges to empirical studies. One reason for this is because the phenomenon is difficult to observe in the field, as knowledge of the boundaries of a social network is necessary and often unknown. Further, behavioral studies conducted in the laboratory require large sample sizes, as the unit of analysis is the network.

Further, behavioral studies that have been conducted, demonstrate the difficulty of observing the process. For instance, anthropological diffusion research often utilizes historical data analytic techniques (Bentley & Shennan, 2003). However, using historical analysis measures adoption of behavior through the convergence of artifacts, such as eating utensils or combat machinery, over time which doesn't allow for the observation of the transmission process. In other words, the process is not directly observed, so the mechanisms of diffusion are unknown. Ethnographic techniques (Wellin, 1955; Pelto & Muller-Wille, 1972) and interviewing techniques (Erikson, Nosanchuk, Mostacci & Dalrymple, 1978; Richardson, Erikson & Nosanchuk, 1979) are also used in diffusion research. Due to the nature of these methodological designs, generalizations cannot be made from the results beyond the observational groups. Therefore, further empirical research is needed in this area. Computational methods also allow for the observation of the diffusion process, as well as comparison of differing models under theorized constraints (Dodds & Watts, 2007). However, this method is really theoretical modeling as behavioral observations are made of computer-based agents, as opposed to human behavioral observation.

The advent of social media, or technologies created social interaction, allows for the observation of diffusion using historical research methods, yet allows one to observe the direct transmission process. This is because the social interactions are published on the Web and saved in databases. So, researchers can retrieve information from social media websites and analyze the interactions among users.

Studies of this type are beginning to be conducted. Yang and Counts (2010) used this method to study the speed and rate of diffusion through social media. Lam, Lo, Yeung, and McNaught (2010) used this method to examine the diffusion of education strategies in social media. Bakshy, Hofman, and Mason (2011) used this method to measure influence in social media. Romero, Meeder, and Kleinbery (2011) used this method to measure diffusion of differing topics in social media. Lerman, Intagorn, Kang, and Ghosh (2012) used this method to use proximity as a predictor of activity in social media.

Social media does require users to opt-in to use the media type, which can lead to biases similar to those discussed using other methods and analytic techniques. However, social media usage is rapidly increasing amongst all adults (Smith, 2011). So, as usage becomes more predominant and studies become more frequent, a program of research will emerge that will allow for generalization beyond the social medium. This study seeks to develop a theoretical model of diffusion guided by the concepts of gatekeeping theory using retrieval and analysis methods of a social media website.

Need for Theoretical Modeling

There are many ways of gaining and organizing knowledge. This study takes a social scientific approach. From this perspective, the role of theory is used to describe, predict, and explain natural phenomena (Pavitt, 2001). A good theory is valuable to the basic researcher, as well as the practitioner. Because theory of objective reality is acontextual by definition, there is utility in applying theory across contexts and domains. So, when faced with a research question one must look to theory to provide an answer. If an answer doesn't exist, theory is used to derive predictions about nature. If theory doesn't exist or is incomplete, theory is developed with reason and logic and is generated deductively (Chalmers, 1999).

Theoretical modeling is used in theory construction in the social sciences (Jaccard & Jacoby, 2010). Model construction can take many forms, including causal, mathematical, and computational. This study takes a causal modeling approach. This approach seeks to explain the causes of variation in constructs of interest (Jaccard & Jacoby, 2010). In other words, causal modeling defines the relationships between constructs in which one construct influences the variance on the other construct. This type of model is generally represented in a path diagram, with rectangles representing constructs and arrows representing causal relationships.

It is important to note that theoretical modeling is only the initial step in the theory construction process. The model does not represent a complete scientific theory. While a complete scientific theory describes, predicts, and explains nature (Pavitt, 2001), a theoretical model only provides predictions of nature (Jaccard & Jacoby, 2010). It is, therefore, an important step in the process of theory construction, but does not provide a complete definition of the theory. According to Lakatos (1978), science progresses through programs of research. A program of research consists of generating theory, deriving falsifiable hypotheses from theory, and objectively collecting observations based on the hypotheses. In this way, science progresses by continually refining theory through testing novel hypotheses predicted by theory. Theoretical modeling is the beginning step of the program of research.

CHAPTER TWO

LITERATURE REVIEW

Search in Social Networks

In the main, investigation of the structure of social networks has been in the area of social search (Milgram, 1967; Granovetter, 1973; Killworth & Bernard, 1978; Adamic & Adar, 2005). Social search is the process of using the social network in order to locate an actor of resource. For instance, Milgram's (1967) seminal study of social networks found that any given person is connected to another given person by a median of six connecting people. His experiment capitalized on relational connections between actors to produce the effect and provided a foundation for the structure of social network, often referred to as small-world networks.

Milgram (1967) demonstrated that any given person could successfully search a global social network with only local knowledge of the network. Basically, when presented with a search query, a network actor, or node, has some probability of successful retrieval, with knowledge of only their social circle. The question is, how is this done? Travers and Milgram (1969) reported a 5% success rate of randomly started an information packet with one actor and having it successfully delivered to a randomly selected target actor. Guiot (1976) reported a success rate of 85% and Dodds, Muhamad and Watts (2003ab) reported a 1.5% success rate in replicated studies. While there is a large amount of variance between the results of the studies, there are common search characteristics; specifically participants were more likely to use geographic location and professional occupation as search qualifiers.

Homophily

According to McPherson, Smith-Lovin and Cook (2001) homophily is the principle that people are more likely to connect with others who are similar to themselves as opposed to others who are dissimilar. It is this principle that helps explain the likely ease with which search occurs in global social networks. However, homophily is a general principle that refers to similarities such as location, ethnicity, education, taste in music, etc. So, which characteristics or combination of characteristics lead to a greater likelihood of a successful search?

Killworth and Bernard (1978) were the first to tackle this problem. They had 58 participants identify a person from their social circle most likely to know a fictitious target. The participants were given the target's name, sex, location, occupation and ethnicity. Along with the identity of a person from their social circle, they were also asked to provide their relationship with the person, the person's sex and the reason why that person was chosen. Killworth and Bernard (1978) found that 47% of participants listed occupation for the reason they chose a particular person from their social circle. Location was listed as the reason by 45% of participants and ethnicity and other was listed by 7% of the participants. They further found that friends and acquaintances were chosen 82% of the time and males were chosen 64% of the time.

Travers and Milgram (1969) and Dodds et al (2003ab) found results consistent with Killworth and Bernard (1978). Dodds et al (2003ab) reported that 50% of the senders in the study chose a particular person in their social circle due to location or occupation. In fact, they found occupational ties to be so strong that 34% of successful chains, from initial sender to target involved occupational ties. Further, 67% of people were chosen due to a friend relationship, as opposed to relatives, co-worker or significant other.

Travers and Milgram (1969) reported similar findings. Qualitatively, they reported that location was a primarily used as an initial search qualifier. However, the message would traverse many participants in the location before finding a suitable path to the target. Conversely, once the message reached a participant in the target's occupational field, the route to the target was quicker. This may be that a given occupational field is smaller in terms of social distance that a geographic location.

Homophily versus Network Structure

It is important to note that structure also impacts network search. Watts and Strogatz (1998) demonstrated the presence of small-world networks by simulating completely ordered networks and rewired them uniformly at random. However, the random rewiring didn't exhibit searchability (Kleinberg, 2000) because social networks don't originate from a completely ordered state. As suggested by Granovetter (1973), social networks will structurally exhibit local clusters, or strong tie relationships, and global bridges, or weak tie relationships. This structure creates a scale-free network, resulting in network hubs (Barabási & Albert, 1999).

Travers and Milgram (1969) found that 48% of completed searches went through 1 of 3 people immediately to the target. Guiot (1976) 27% of completed searches went through 1 person immediately to the target. This suggests that hubs play an important role in successful searches.

To get a better understanding of the importance of hubs and homophily, Adamic and Adar (2005) compared network structure, occupation and location in two distinct social networks by simulating information paths. They first utilized a network dataset constructed of email chains within an organization. Within the email network, they simulated a network structure search strategy by setting a rule to select a sender based on their connectivity to others; essentially targeting a network hub first. While, all targets in the simulation were found, there was a median of 16 steps between an initial sender and target. Next, they simulated an occupation search strategy by setting a rule to select a sender based on position in the organizational hierarchy. In this case, there was a median of 4 steps between initial sender and target. Finally, they simulated a location search strategy by setting a rule to select a sender based on location in the building, including floor and cubicle location. In this case, there was a median of 6 steps between initial sender and target.

Adamic and Adar (2005) repeated the experiment using a network dataset constructed from links in a university social networking site. Network structure was simulated using the same rules, occupation was simulated using department and year in school and location was simulated using location in a dormitory, including floor and room. The experiments yielded results similar to those of the email network.

These results indicate that hubs are necessary in order to search a network (Travers & Milgram, 1969; Guiot, 1976; Kleinberg, 2000), but utilizing connectivity as a search strategy is not likely to be the most successful search strategy.

Social Networks and Diffusion

Social networks are groups of people connected to one another by some kind of relationship. When these relationships are visualized, a network of actors connected by links (Haythornthwaite, 1996) can be seen. A prominent area of research centering on social networks is structure. Specifically, Reagans and McEvily (2003) found that the more connected the network is, the more likely it is for information to reach a majority of a network. This is the concept of cohesion (Haythornthwaite, 1996; Reagans & McEvily, 2003).

Haythornthwaite (1996) defines cohesion as the degree of connectedness within a social network. In other words, a highly cohesive network will consist of the majority of network members connected to the majority of other network members. Granovetter (1973) argues that network actors in a highly cohesive network demonstrate strong tie connections. Examples of strong tie connections are the relationships are those maintained between family and close friends. Conversely, the less connected network members are to others in the network, the less cohesive the network is. Granovetter (1973) also argues that network actors in a low cohesive network demonstrate weak tie connections. Examples of weak tie connections are the relationships are those maintained between acquaintances.

If a network is too densely connected, the network can become closed (Gargiulo & Benassi, 2000). In this kind of network, it becomes difficult for information to come into the network or leave the network. To overcome this obstacle, bridging relationships are necessary (Burt, 1992). Bridging relationships are those that connect disparate groups. Granovetter (1973) argues that weak tie connections are necessary for novel information flow in and out of cohesive networks.

Centola (2010) conducted a behavioral experiment looking at network structure in relationship to behavioral spread and found that highly dense networks were not suitable for diffusion. Instead, it’s necessary to have highly cohesive clusters with some network members bridging the clusters. This was also found in computer simulations conducted by Watts (2004). This type of network structure is defined as a small-world network (Milgram, 1967).

Milgram (1967) focused on both in a seminal study of small-world networks. In this study, subjects had to search for other network actors in a global network solely by using their local social networks. In this instance, a local network refers to a network of actors known to a subject, while a global network refers to a network of actors unknown to the subject. In this study it was observed that are people are connected to one another by a median of six other people. This began to reveal the underlying structure of networks, later theorized by Granovetter (1973) and Burt (1992).

A process that is less well understood regarding social networks is that of diffusion (Schnettler, 2009). This is a relevant process to study because it allows one to observe the extent to which information, behaviors, opinions, etc. converge or diverge within a network of actors. Rogers (2003) demonstrates that when the diffusion process is completed in a social network, it is graphically represented by an ogive. This suggests that at some point there is a cascade, after which the majority of network actors have been privy to the diffused information. Boster, Kotowski, Andrews, and Serota (2011) argue that superdiffusers, or network actors who are highly connected, highly persuasive, and mavens, are able to exert influence in a network. This suggests that superdiffusers are likely to be at or near the tipping point of a diffusion cascade.

Diffusion is often measured in completeness, or the extent to which all network actors are privy to the information. Therefore, diffusion is considered successful if the majority of network actors have been involved in the process. So, during the diffusion process, the majority of network will take on the role of gatekeeper. However, Shoemaker and Vos (2009) argue that previous gatekeepers can influence gatekeepers and that not all gatekeepers are significant to the diffusion process.

Social Media

"Social media is a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user generated content" (Kaplan & Haenlein, 2010). In other words, social media is an online social network, where relationships are publicly acknowledged and real-time conversations become information that is digitized and stored within an Internet-based platform. As boyd and Ellison (2007) argue, the unique feature of social media is that they allow users to make social connections, and thus larger social networks, visible. If the social connections and conversations are visible, it stands to reason that the diffusion process is visible as well. Therefore, social media offers a lens to visualize social network dynamics and the information diffusion process.

From a practical perspective, both public and private sectors turn to social media for a more tangible connection to constituents/customers. Luo, Zhang, and Duan (2013) find consumer ratings on social media to be predictive of a company's equity rating, indicating the need for positive social media connections. Pfeffer, Zorbach, and Carley (2014) define the dynamics of online word-of-mouth for consumer brands, specifically when word-of-mouth is negative. Lovejoy and Saxton (2012) demonstrate how nonprofit organizations provide information and build community through social media. Mergel (2013) describes social media adoption tactics or government agencies stemming from a need to communicate more directly with constituents. It is clear that social media is being used to disseminate information. As social media has theoretical and practical importance, it is a valuable research tool in investigating the dynamics of social network diffusion.

Theories of Diffusion

Diffusion of Innovations

Rogers (2003) introduced the Diffusion of Innovations theory based on work done on the dissemination of farming technology innovations. The theory has since been expanded to include innovation diffusion in consumer markets (Ghoshal & Bartlett, 1988; Frambach, 1993), political systems (Walker, 1969), health systems (Greenhalgh, Robert, Bate, Macfarlane, & Kyriakidou, 2007; Oldenburg & Glanz, 2008) and other social systems (Valente, 1996; Valente & Davis, 1999; Wejnert, 2002; Watts & Dodds, 2007). In this theory Rogers (2003) proposes five categories of adopters, five stages to the adoption process, and four main elements of diffusion.

The five categories of adopters include: innovators, early adopters, early majority, late majority and laggards (see Figure 1). Innovators are the first to adopt and innovation. Early adopters are the opinion leaders; they tend to be the most influential group in terms of the success or failure of an innovation. This group is also identified as superdiffusers within their social networks (Boster, Kotowski, Andrews & Serota, 2010). The early majority tend to be slower in the adoption process and rarely hold positions of opinion leadership. The late majority tend to be highly skeptical and only tend to have contact with the early majority and others in their category. They don't adopt until the majority of a population has adopted an innovation. Finally, the laggards tend to be averse to change and so are the last to adopt an innovation.

Further, Rogers (2003) argues that there are five stages to the adoption process: knowledge, persuasion, decision, implementation, and confirmation. The knowledge stage provides first exposure to the innovation. In the persuasion stage, the individual actively seeks information about the innovation.

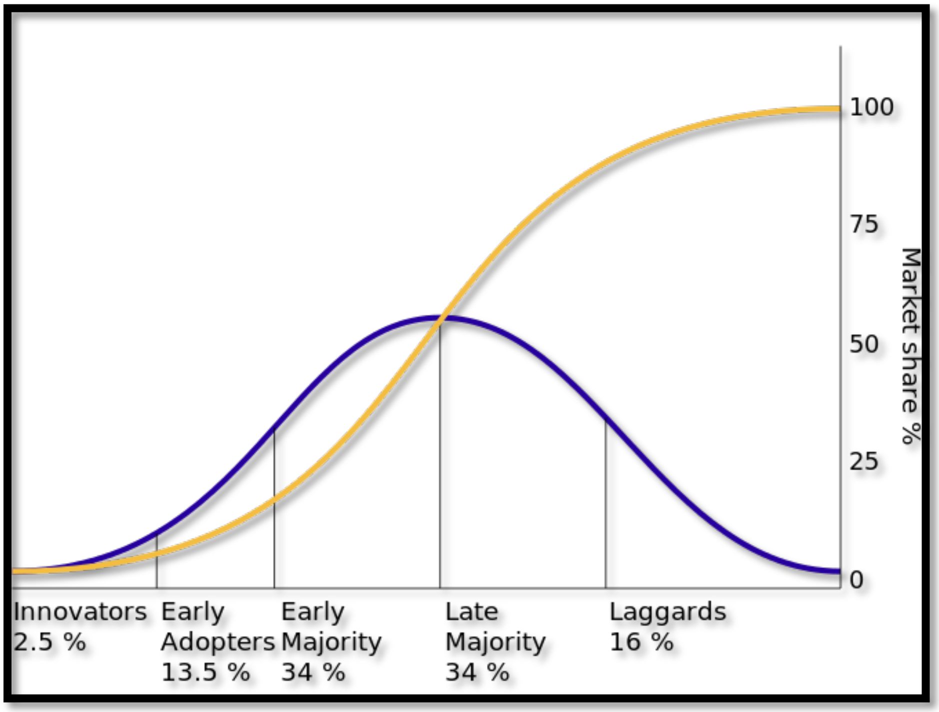

Figure 1. Diffusion of Innovations Model Innovation Curve. Rogers Everett, via Wikimedia Commons. Used under CC-BY-SA-3.0

In the decision stage, the individual conducts a risk benefit analysis of this innovation. If the individual decides to reject the innovation, the process does not continue to the next two stages. In the implementation stage, the individual utilizes the innovation. In the confirmation stage, the individual makes a final judgment of the innovation having already utilized it. Following this stage, the individual either continues use or disregards the innovation.

Finally, Rogers (2003) identifies four main elements of diffusion: the innovation, communication channels, time, and a social system. The innovation is an object or idea that is perceived as new by an adopting actor. Communication channels are the means by which information is transmitted between actors in a social system. Time is the period from the introduction of the innovation to the adoption decision. A social system is the social network of actors into which the innovation has been introduced.

SIR Model of Diffusion

The SIR model was introduced by Kermack and McKendrick (1927) to explain the process of transmitting communicable diseases through a population. S represents susceptibility and is defined as the individuals not yet infected. I represents infected and is defined as the individuals who are infected and are capable of transmission. R represents recovered and is defined as the individuals who have been infected, but have recuperated and are no longer susceptible for some period of time. Since that time, SIR models diffusion processes (Gruhl, Guha, Liden-Nowell, & Tomkins, 2004; Mahajan, Muller, & Bass, 1995; Watts & Dodds, 2007) have been proposed.

The SIR model integrates Rogers (2003) theory of Diffusion of Innovations such that the five categories of adopters are split into two categories: innovators and imitators. The SIR model retains the innovator category and aggregates the remaining four categories into a newly created imitator category (see Figure 2). Bass (1969) describes imitators as being affected by adoption pressure of the social network, while innovators are not.

The process of the SIR model is described by the equation

$$P(T) = p + (q/m)Y(T),$$

where p is the fraction of all adopters who are innovators, m is the initial adoption of the behavior, and, therefore, q/m is the pressure on non-adopters as the number of adopters increases. Essentially, the SIR model proposes that the number of individuals engaging in a behavior increases based on the absolute number of contacts in the individual's network engaged in the behavior. Assuming a normal distribution in the population, the diffusion path will display the traits of a concave curve, with a gradual increase.

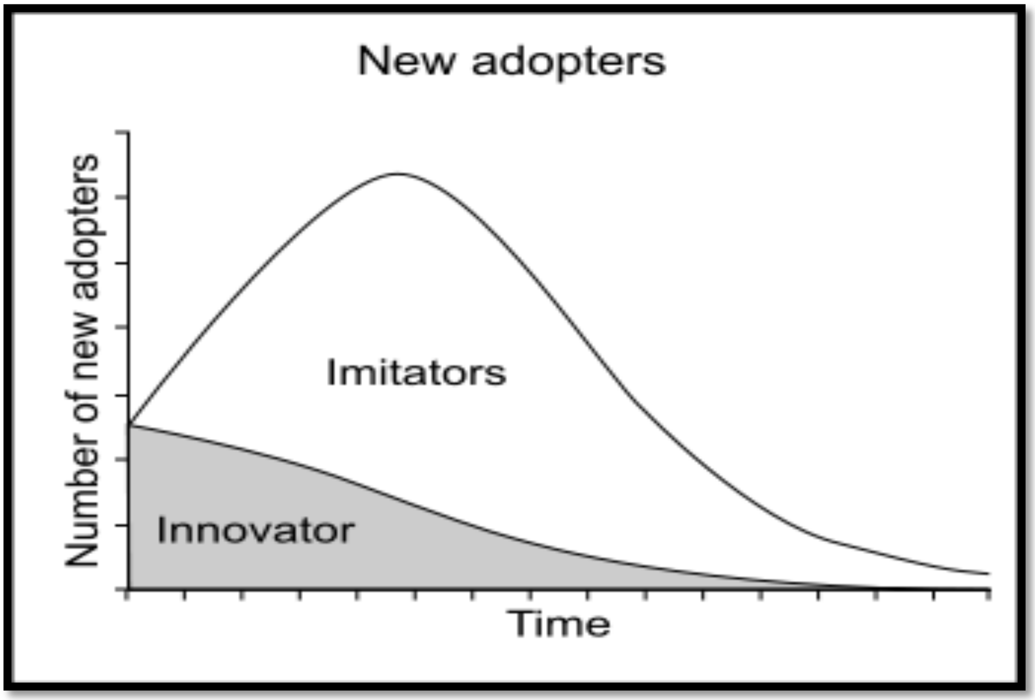

Figure 2. SIR Model of Diffusion. Apdevries, via Wikimedia Commons. Used under CC-BY-SA-3.0

Limitations of the Current Diffusion Theories

Diffusion of Innovations doesn't account for authority or information control. While the theory defines five categories of adopters, it doesn't account for the specific role of any of the adopters. The definition of an innovator is that this actor is the first to adopt and innovation. This implies that the innovator, the first adopter type in the Diffusions of Innovations theory, is not necessarily the creator of the innovation. The innovator could be the creator, but, as inherent in the definition, it isn't required that the innovator also be the creator. Therefore, this adopter type is not responsible for introducing the innovation to the social system.

Further, this theory is dependent on the diffusion of an innovation. It may not work in acontexual applications. For instance, diffusion of a message via a rumor mill may not adhere to the constructs and relationships defined in Diffusion of Innovations theory. Crisis communication is another diffusion context that may not follow the tenets of the theory. Crisis information doesn't always include an innovation or practice to adopt. Also, this type of information needs to be delivered quickly and from an authoritative source. Based on these definitions, Diffusion of Innovations, likely, cannot be extended to account for this context. In other words, the theory is too narrowly defined to specific contexts.

One of the major problems with the SIR model is that it doesn't require a decision on the part of the adopting actor. The SIR model was developed to model the diffusion of communicable diseases. Succumbing to disease does not require a decision on the part of the receiver. Arguably, most actors likely try to avoid acquiring a communicable disease. While an actor can engage in behaviors that increase the likelihood of catching a disease, such as not washing hands during flu season or engaging in unprotected casual sex acts, the body acquires a disease without the actor actually choosing to do so. Also, as with Diffusion of Innovations, this model does not account for authority or information control.

Gatekeeping Theory

Gatekeeping theory provides a framework from which to study the process of information diffusion. Gatekeeping is the process of information control (Lewin, 1951). Gates are decision or action points and gatekeepers are those individuals responsible for making the information control decisions (Shoemaker & Vos, 2009). Also, forces exist at each gate that influence the decision process (Lewin, 1951). In terms of diffusion, Chaffee (1975) argues that actions that aid or hinder the flow of information will alter the diffusion distribution, such that it will deviate from the normal ogive, or S-curve. Further, Lewin (1951) argues that any network actor can be a gatekeeper. This suggests that being a gatekeeper is a role one takes on. Shoemaker and Vos (2009) argue that gatekeepers are not equally influential in the process of information control.

In the main, gatekeeping theory is used to describe, predict, and explain processes of mass communication, specifically journalism (Shoemaker & Vos, 2009). Barzilai-Nahon (2008) extended the theory to the field of information science, specifically to the role of technology in gatekeeping. However, this theory can be further extended to the general process of information control.

White (1950) undertook the first behavioral investigation of the gatekeeping process through a case study of 'Mr. Gates'. 'Mr. Gates' was an editor at a newspaper, who documented his decision process, specifically regarding the inclusion or exclusion of news articles from wire press. The decisions often centered on Mr. Gates' personal opinions of the submitting press or his personal opinions of the content. Some decisions were made using no strict criteria for evaluation. From the case study White (1950) suggested that gatekeepers actively control information into the media channel.

Gieber (1956), however, found contradictory results from a study of telegraph editors. In this study, the gatekeepers were found to have a passive role in the process. Instead of actively deciding what information to pass along and what to refuse, the gatekeepers were following a selection process set out by the organization.

The different results are likely due to different levels of analysis within gatekeeping theory. Shoemaker and Vos (2009) layout four levels of analysis: individual, routines, organizational, and social systems. At each level of analysis forces exist (Shoemaker & Vos, 2009). As stated earlier, forces can be either positive or negative and their polarity is not constant

(Lewin, 1951). Further, multiple forces can exist for any gate (Lewin, 1951) and these forces can be in competition with one another (Shoemaker & Vos, 2009).

At the individual level, personality and communication characteristics are important (Hickey, 1968; Henningham, 1997). For instance, Hickey (1968) found three types of gatekeepers: the communication handler, the channel mediator, and the content manipulator. The communication handler is one who transmits information. The channel mediator is one who maintains information channels. The content manipulator is one who has both of the preceding characteristics and who, also, shapes the information. Henningham (1997) further found a difference between introverted and extroverted gatekeepers, such that introverts tend to be more reflective and analytic regarding the shaping of information, while extroverts tend to be more concerned with disseminating the information.

At the routines level, communication practices are important (Shoemaker & Vos, 2009). Communication practices are impartial rules, or norms, that are followed by network actors (Shoemaker & Vos, 2009). For example, in some Southern American cultures it is considered impolite for men to use particular words or phrases while in the company of women. Another example is when one adheres to the principle of political correctness. Entman (2007) argues that the more one identifies with and is immersed in a group, the more likely they are to abide by the routines of that group. In terms of the social network, this is Granovetter's (1973) concept of strong ties.

At the organizational level, characteristics of the institution are important (Shoemaker & Vos, 2009). This is the level at which Gieber (1956) was studying when it was observed that the telegraph editors were inhibited by the selection mechanisms of the news organization. Finally, at the social system level, cultural characteristics are important (Shoemaker & Vos, 2009). The social system in gatekeeping theory refers to the society in which the information is created and transmitted (Shoemaker & Vos, 2009). This means that characteristics such as culture and ideology are relevant.

Information channels are also relevant to gatekeeping theory (Shoemaker & Vos, 2009). Sigal (1973) defines three types of information channels: routine, informal, and enterprise. Routine channels are those that are recognized and scheduled, such as public demonstrations or broadcasts. Informal channels are those that exist behind the scene or off the record. Enterprise channels are those that occur intrapersonally, by critical thinking, or arising through serendipitous interactions with other network actors. Information diffusing through routine or information channels enters a network through a boundary person. In terms of the social network, this is Burt's (1992) concept of a bridge.

Finally, the information itself is relevant to gatekeeping theory (Shoemaker & Vos, 2009). Nisbett and Ross (1980) argue that vivid information is more likely to be transmitted. Vivid information is that which relates directly, spatially, or temporally to oneself, a known person, or event. Further, vivid information violates a norm in some way, either positively or negatively. However, Shoemaker and Vos (2009) note that information doesn't shape itself; someone is responsible for transforming data collected from sources into information. To that end, Shoemaker and Vos (2009) define a news item as the final product to be transmitted. This definition works in a mass communication context. More broadly, however, Lewin (1951) argues that information can change form as it passes through any gate.

Barzilai-Nahon (2008) extended gatekeeping theory into an information science context by defining the relationships between the gatekeeper and the gated, specifically introducing four constructs: power, information production, relationship, and alternatives. Power refers to the difference in authority between the gatekeeper and the gated. Information production refers to the volume and quality of information produced by the gated and provided to the gatekeeper. Relationship refers to the relationship between the gatekeeper and the gated. Alternatives refers to varying channels the gated has to choose from. Extending the theory in this way was intended to allow for the role of technology as a gatekeeper. For instance, a search engine can be considered a gatekeeper of information. Further, Web 2.0 technologies, such as data sharing infrastructures and social networking infrastructures, have become types of information gatekeepers.

Proposed Theoretical Model

A causal model of information diffusion through social networks is proposed (see Figure 3). This model incorporates the constructs of gatekeeping theory within the framework of social networks. Specifically, the proposed model emphasizes authority in the role of the gatekeeper and purposeful diffusion of information by the actor (see list of definitions Table A-1 in Appendix).

Inherent in gatekeeping theory, is the concept of authority, because the gatekeeper is an agent of information control. Traditionally, the gatekeeper title has been attributed to control agents such as journalists (Robinson, 2006) or librarians (Ovadia, 2007). Under this paradigm, the public received information from an authority, who was perceived to be either more knowledgeable than oneself or had access to information that one could not otherwise retrieve. However, with the advent of the content aggregators on the World Wide Web, there has been a paradigm shift. As Lewin (1951) noted, anyone can be a gatekeeper. Now that the Web has gone social the ability to identify an information authority is difficult.

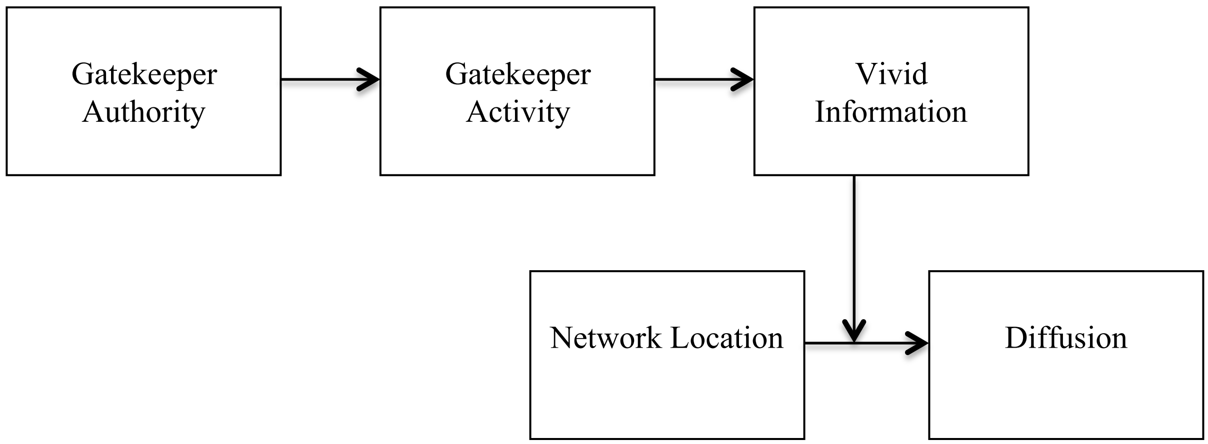

Figure 3. Gatekeeping Model of Diffusion

It is becoming more common to take a PageRank approach to authority, such that each link or follower accumulated on the Web is counted as a vote of authority (Marlow, 2004). Essentially, authority appears to be defined through social aggregation on the Web (Ovadia, 2007). To that end, authority is often measured in one of two ways: as popularity metric or as an influence metric (Marlow, 2004). However, these measures are distinct constructs on their own; to be used as measures of authority conflates the constructs. For instance, one can be an agent of information control without being influential in the diffusion of information. As Burt (2004) demonstrates, network actors most closely linked with network bridges tend to have more novel ideas and tend to be considered more creative. This suggests that, although a bridge may introduce novel information, the integration of that information with current knowledge and dissemination of that new idea is done by a different network actor. Therefore, the gatekeeper is not necessarily the influential.

Hypothesis 1: The more authority a gatekeeper is perceived to have, the more active the gatekeeper will be in the network.

At the point that information is received the actor must decide whether or not to pass the information along through the network. At each level of analysis identified in Gatekeeping Theory forces exist (Shoemaker & Vos, 2009). Forces can be either positive or negative and their polarity is not constant (Lewin, 1951). Further, multiple forces can exist at any gate (Lewin, 1951) and these forces can be in competition with one another (Shoemaker & Vos, 2009). At the point information is received, the network actor must decide whether or not to pass the information along through the network. Two forces that are relevant to the diffusion process are the activity level of the gatekeepers and the network location of the gatekeepers.

Activity level is the extent to which a gatekeeper actively seeks out and distributes information. Internal forces, such as personality characteristics, can affect activity levels. External forces, such as organizational demands, can also affect activity levels. Location refers to where the gatekeeper is positioned in the structure of the social network. In other words, location refers to the network actors to whom the gatekeeper is connected. So, a gatekeepers' location will impact the network actors who receive the information. Network actors who are highly connected act as bridges between highly cohesive, otherwise unconnected, local networks (Burt, 1992). Granovetter (1973) argues that these relationships are necessary for novel information to flow through global social networks.

Hypothesis 2: The more active the gatekeeper is, the more likely the gatekeeper will be to shape the form of the information.

Hypothesis 3: The more closely connected the gatekeepers are to network bridges the more likely the information will diffuse quickly and completely through the network.

Shoemaker and Vos (2009) argue that not all gatekeepers are influential in the diffusion process. Boster et al (2011) find that superdiffusers are more likely to be highly connected, and astute at persuasively shaping information. Shoemaker and Vos (2009) note that information doesn't shape itself; someone is responsible for collecting data and transforming it into information. Lewin (1951) argues that information can change form as it passes from gate to gate. Nisbett and Ross (1980) note that vivid information is more likely to be transmitted through the network. Vivid information is that which relates directly, spatially, or temporally to a known person or event.

Hypothesis 4: Information form will interact with network receivers such that more vivid information will more quickly and completely diffuse through the network.

CHAPTER THREE

MATERIALS AND METHODS

A data mining approach was taken for data collection. According to Larose (2005) data mining is the process of finding meaningful patterns and trends in large amounts of data using pattern recognition, statistical, and/or mathematical trends. Essentially, data mining is a data analytic technique used to extract meaning from large datasets. A large dataset is necessary to observe the diffusion process. The goal of disseminating information is to inform a population of interest. Therefore, any study of diffusion requires the observation of a population, or large sample of a population, which results in a large dataset. The first part of the data mining process requires programming scripts to be written to extract data. The second part of the process requires parsing and analyzing the data.

To measure the diffusion process, information surrounding a technology event, the Consumer Electronics Show (CES) 2013, was mined from Twitter data using Python (http://www.python.org/) scripts. Python is an open source programming language that will be used to interact with Twitter's APIs (application programming interface) to retrieve data. Python was chosen because it is open source, so there are many libraries with freely available scripts, simple to use, and works well with other programming languages.

Consumer Electronics Show 2013

The Consumer Electronics Show is a yearly, international, technology related trade show held in Las Vegas, Nevada. The 2013 show took place over a three day period, from January 8 thru January 13. The purpose of the trade show is to preview to technology products and innovations.

The Consumer Electronics Association (CEA) produces the CES trade show. The CEA is a standards and trade organization for the consumer electronics industry and has a goal to grow the consumer electronics industry. The Association represents over 2,000 business members. In order to gain membership, businesses must be involved in the consumer electronics industry, such as manufacturing, distribution, development, retail, and more. It is also possible to gain associate membership for businesses providing products and services to the industry, such as advertising firms, financial institutions, and more.

Typically, many new electronics products are debuted at the Consumer Electronics Show. For instance, at past events products such as the videocassette recorder, HDTV, Microsoft Xbox, Blu-ray Discs, and Android devices. At the 2013 event some of the popular product debuts were the Razer Edge tablet PC, YotaPhone, and Fitbit Flex. Since many developers and manufacturers choose to debut products during the Consumer Electronics Show, it is a highly followed event by industry insiders and technology enthusiasts. The 2013 CES trade show had over 150,000 attendees and was covered by technology centered news sources as well as the popular press.

Twitter

Twitter is a microblogging, social network website. It was launched in July 2006 and quickly grew in popularity. As of April 2012, Twitter has over 500 million active users. It is available worldwide with the heaviest usage in North America, Europe, and Australia. The service is available to users on the Web and through applications for mobile devices. In order to use Twitter, one must create a profile that consists of a username and password. Once the profile is created users can choose to upload a photo, add a short biography, provide a location, and provide a link to a homepage or weblog.

Twitter is considered a microblog because the service allows users to create messages with no more than 140 characters, called tweets. A user can only post 1,000 tweets per day. Tweets can occur in two ways: the tweet and the retweet. A tweet consists of a message created by a user and posted to Twitter. A retweet consists of a reposting of another user's tweet. Further, tweets can either be public or protected. Public tweets are visible to anyone, while protected tweets are only visible to followers approved by the user. Avid Twitter users have developed a process called live tweeting. When a user is live tweeting, that user is creating new tweets, focused around a set topic, for a continuous period of time. The topic of focus tends to be some event, such as the Presidential debate, an academic conference, or even a birthday party. Often users attach entities to their tweets. Entities provide contextual information about the content of a tweet. There are four types of entities defined by Twitter: hashtags, media, URLs, and user mentions.

Hashtags are words or phrases embedded in the content of a tweet, prefixed with the hash (#) symbol. Hashtags can occur anywhere in the tweet. They are used to categorize tweets and make them more easily searchable. Popular hashtags are used by Twitter as a means to show topics trending in real time. Media are photos or images uploaded with a tweet. Media can be uploaded anywhere in the content. URLs (uniform resource locator) are links that point away from the current to some type of Web content and are included in the tweet text. They can also be included anywhere in the tweet. User mentions are other Twitter users mentioned in the content of the tweet. Mentions must contain a Twitter username prefixed with the at (@) symbol. They can also be included anywhere in the content of the tweet.

While Twitter is considered a social network, it follows an asymmetric model. One user can follow the tweets of a second user, but the second user doesn't have to follow the tweets of the first user. In other words, the Twitter network is made up of a series of one-way relationships with no requirement of mutuality. This is different from the symmetric model followed by many social networks, which requires two-way relationships. The asymmetric model allows for four types of relationships:

1. User A follows User B, but User B does not follow User A.
2. User B follows User A, but User A does not follow User B.
3. User A follows User B, and User B follows User A.
4. User A does not follow User B, and User B does not follow User A.

The entire Twitter network type of relationship model does allow for extremes, such as a user follows this or a user is followed by none of the Twitter network.

There are a few advantages to using Twitter for data collection. First, Twitter provides a bounded network. This allows one to know the scope of the network and observe the dynamics of the entire network. Second, Twitter provides a publicly available API to retrieve user data and tweet data. Third, the Twitter entities allow for the categorization and retrieval of pertinent keywords and/or phrases within individual tweets and across the Twitter network. Fourth, live tweeting allows for real time observation of the diffusion of an event, such as the CES trade show. Finally, the asymmetric relationship structure provides a valuable gage of user influence as relationships are essentially based on attention. If one has limited attention to give, one chooses to give attention to the most important information. So, the more followers a user has, the more likely the information provided by that user is perceived to be important.

User Demographics

Twitter does not collect demographic data from users directly. Specifically, Twitter does not ask users to provide sex, birthdate, or location to utilize the service. It is reasonable to assume that Twitter is able to extract some of this data (i.e. through geo-location), though the company does not make this information publicly available. Pew Research Center, an independent research institute, has investigated Twitter usage by adults in the United States as part of the Center's Internet & American Life Project. While, this only represents a subset of the Twitter population, it does provide some insight into the demographics of Twitter users.

Smith and Brenner (2012) conducted a telephone survey of American adults, aged 18 and older, using a combination of landline and cellular random digit dial samples. The sample was then weighted in two ways. First the sample was weighted to correct for probabilities of telephone usage based on the number of adults in the household, as well as an overlap with landline and cellular phone sample frames. Second the sample was weighted to balance sample demographics to population demographics. A sample of Internet users (n=1,729) was asked questions about usage of Twitter within the past year.

Survey responses (Smith & Brenner, 2012) indicate that 15% of the sample use Twitter and 8% use Twitter daily. The predominant age group to use Twitter is 18-29 year olds, representing 26%; followed by 30-49 year olds, representing 14%; then 50-64 years, representing 9%; and, lastly, 65+ year olds, representing 4%. The predominant ethnicity to use the service identifies as Black Non-Hispanic, representing 28%; followed by Hispanic, representing 14%, and, lastly, White, Non-Hispanic, representing 12%. 14% of men and 15% of women uses Twitter. Finally, 19% of Twitter users reside in urban areas, while 14% reside in suburban areas, ad 8% reside in rural areas.

Data Description

To measure the diffusion process, tweets will be mined between the dates of January 2, 2013 and January 18, 2013. Since the Consumer Electronics Show began on January 8, 2013 and ended on January 13, 2013, the three week time frame provides enough time to see a ramp up and slow down of communication regarding the landing. As Romero, Meeder and Kleinberg (2011) show, information about innovations has a rapid spike and quick decay. It is expected that this information will follow the same diffusion pattern. Both tweet data and user data were mined.

Only tweets related to the CES trade show were utilized. User data will be retrieved based on username. Tweet data was retrieved based on tweet entities, specifically hashtags and mentions. According to Topsy.com, a social search and social analytics website, the four most popular entities used during the conference were #CES, #CES2013, #2013CES, #CES13, CES, @intlCES. These are the terms that were used to begin the analysis. All user data was retrieved along with the tweets.

Measurement

Gatekeeper Authority

Twitter verifies the profiles of companies and public figures whose identities could otherwise be falsely portrayed in an anonymous social medium. Therefore, diffusion paths of verified authorities and non-authorities will be separated. The verified authority for CES is @intlCES which is the Twitter profile created by CES to distribute information from and about the organization to the Twitter population. Additional verified Twitter sources were also identified in the dataset. These verified sources include media sources, such as @Wired, technology companies, such as @Cisco, and technology evangelists, such as @JasonSilva. Any Twitter user that is not using a verified profile is considered a non-authority.

Gatekeeper Activity Level

Gatekeeper activity level was measured by the amount of tweets a gatekeeper produces. The amount of tweets was measured in quantity and frequency. In other words, activity level measurement shows how much and how often the gatekeeper produces content.

Vivid Information

Vivid information was measured using emotional, spatial, and temporal indicators. Examples of emotional language include some reference to mood or emotional state, such as ‘super excited’, ‘pissed off’, or ‘underwhelmed’, etc. Examples of spatial language include some reference to location, such as ‘close by’ or ‘too far’, etc. Examples of temporal language include some reference to time, such as ‘launching soon’, ‘arriving in an hour’, or ‘minute-by-minute countdown’, etc.

Gatekeeper Location

Gatekeeper location was measured by retrieving user data from Twitter. The first attempt at retrieving user data was to retrieve all followers of CES (@intlCES). This may not capture all Twitter users who tweeted about the 2013 Consumer Electronics Show, however, as Twitter users could have heard about the event from sources outside of CES and/or Twitter. Therefore, to capture any additional Twitter users tweeting about the mission, tweet data was retrieved based on tweet content. Tweets were retrieved based on the hashtag entities relating to the trade show (#CES, #CES2013, #2013CES, #CES13) as well as on the as the keyword 'CES'. From the tweet data, the authoring user account was be parsed. Followers of the parsed authoring user accounts were then retrieved.

Betweenness centrality, closeness centrality and eigenvector centrality were calculated along with follower count to measure location. Betweenness centrality measures a node's ability to bridge subnetworks within a closed network (Wasserman & Faust, 2008). Closeness centrality measures the average shortest distance between nodes (Wasserman & Faust, 2008). Eigenvector centrality measures the degrees of each node that is connected to the focal node (Wasserman & Faust, 2008). In other words, eigenvector centrality measures the number of number of nodes connected to each node that is connected to the focal node.

Diffusion

Diffusion was measured according the method proposed by Cosley, Huttenlocher, Kleinberg, Lan, and Suri (2010). This calculation defines user *X*, entity *H*, and neighbor *k,* such that users *X*, who haven't mentioned *H*, have some *k* neighbors who have mentioned *H.* Therefore, *p(k)* is the fraction of users who mention *H* before a $(\boldsymbol{k}+\mathbf{1})^{\boldsymbol{st}}$ neighbor mentions *H.* In other words, if user *X* hasn't yet adopted an entity, then *p(k)* is the proportion of *X* who will adopt *H* after their $\boldsymbol{k}^{\boldsymbol{th}}$ exposure to it.

This equation measures diffusion actively. In other words, it requires that the Twitter user actually tweet about the CES Conference. It doesn't account for users who may have viewed tweets about the conference, but did not tweet or retweet about it themselves. This is in accordance the Kadushin's (2012) definition of diffusion, which requires an action or behavior in order for the process to occur. A metric of users who could have seen tweets about the conference, but not tweeted about it will be taken as *k/X*. While this metric does provide information regarding potential exposure to the event, it doesn't provide any information as to whether a user actually viewed the tweet or not. For instance, a given user may not have logged into Twitter within the three week time frame. Also, a user may follow so many other users, that tweets about the CES conference could have been overlooked. Therefore, the most useful metric to measure diffusion must incorporate the act of tweeting or retweeting the event.

CHAPTER FOUR

RESULTS AND DISCUSSION

Data was gathered from Twitter by querying the representational transfer state (REST) application programming interface (API) between the dates January 2, 2013 and January 18, 2013. The REST API allowed for the mining of core Twitter data, including tweets and user information. In total 923,315 tweets were captured from a total of 284,605 Twitter users. Of the captured users 9,314 had no followers. This means those users had no way to forward the flow of information and, therefore, considered to be outside scope of the study population. These users were removed from the dataset leaving 275,291 users and 908,609 tweets. Over the three-week time period, the tweets about the CES conference follow a normal distribution, with the largest number of tweets occurring on the first day of the conference, January 8, 2013 (see Figure 4). The total tweets generated about the CES conference, during the given time period, ranged from 1 tweet per user to 4,184 tweets per user, with χ^2 = 3.24 tweets per user (see Figure 5).

Gatekeeper Authority

The dataset was divided by verified users and non-verified users in order to observe the diffusion patterns of the two groups. A total of 1,515 verified users, defined as Twitter authorities, created tweets about the CES conference within the three week time frame. The verified users produced a total of 13,082 tweets. A total of 273,776 non-verified users, defined as Twitter non-authorities, created tweets about the CES conference within the three week time frame.

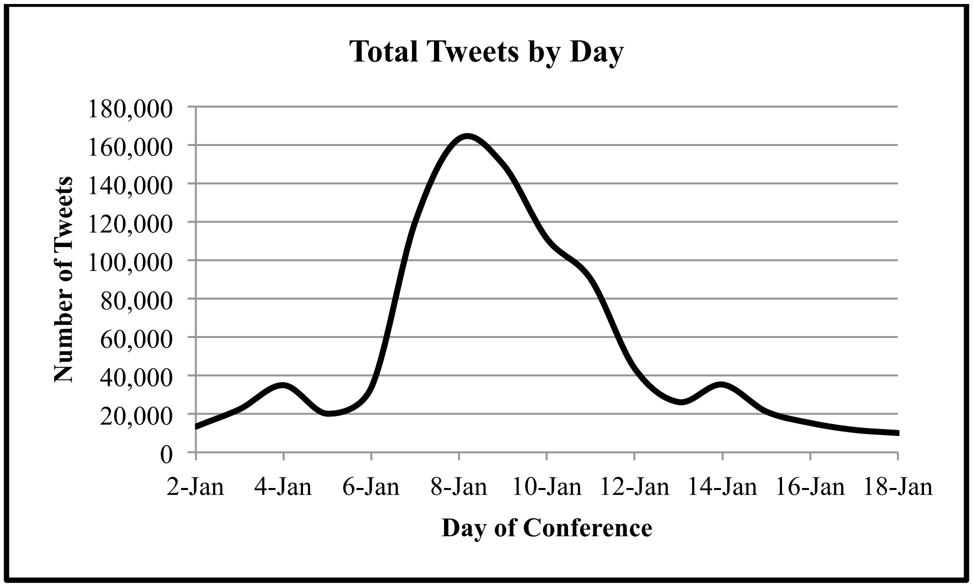

Figure 4. Conversation Curve for Total CES Tweets

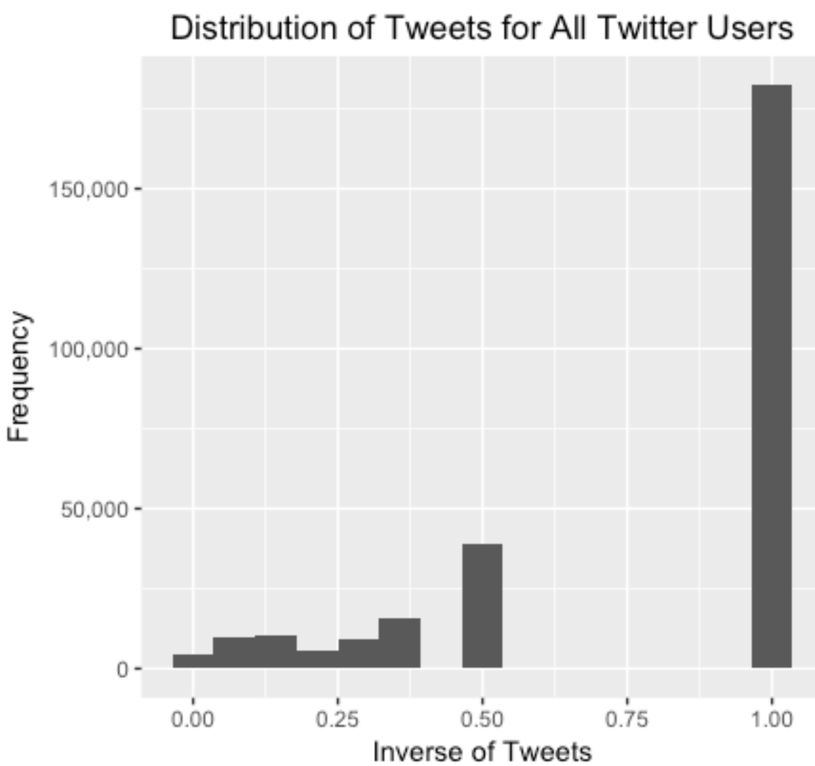

Figure 5. Inverse Transformation of Tweets Generated by All Users

The non-verified users produced a total of 895,527 tweets. Proportionally, verified users accounted for 0.01% of the total users that tweeted about the CES conference.

Further, tweets created by verified users accounted for 0.01% of the total CES related tweet content that was created during the three week period. Over the three week time period, the tweets from both verified and non-verified users, about the CES conference follow a normal distribution (see Figure 4), with the largest number of tweets occurring on the first day of the conference, January 8, 2013, with a slightly less kurtosis curve generated by verified users (see Figure 6).

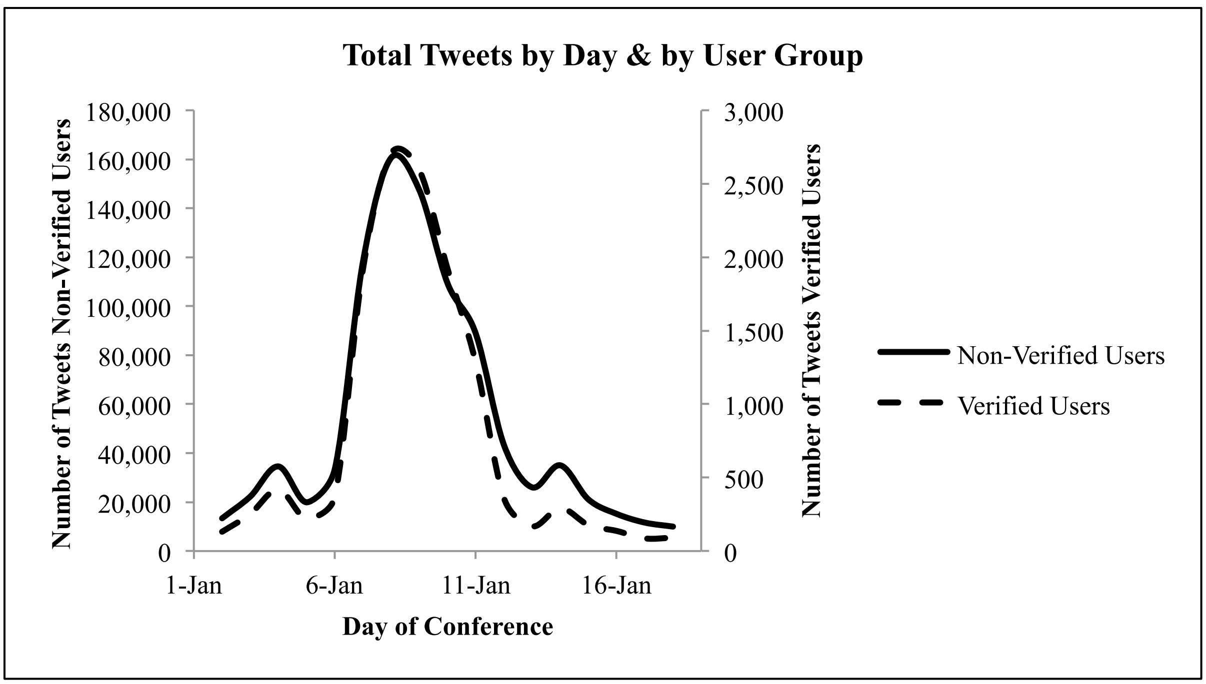

Figure 6. Conversation Curves for Verified and Non-Verified User CES Tweets

Gatekeeper Activity Level

Verified users generated a total of 13,082 tweets. Individual verified users generated tweets ranging from 1 tweet per user to 78 tweets per user, with χ^2 = 6.94 tweets per user (see Figure 7). The distribution of tweets per verified user follows a power law distribution, where a small subset of users sends the majority of the tweets. Non-verified users generated a total of 895,527 tweets. Individual non-verified users generated tweets ranging from 1 tweet per user to 4,184 tweets per user, with χ^2 = 3.2 tweets per user (see Figure 8).

According to Kutner, Nachtsheim, Neter, and Li (2005), when normality distributions are violated, transformations are useful. The activity datasets naturally follow a Poisson distribution, where the likelihood of followers is small. Specifically, the likelihood is 1 tweet per user. Due to the inherent skew in the data an inverse transformation was applied to reduce the skew and normalize the data.

Vivid Information

An open source machine-learning algorithm, Easy Text Classification with Machine Learning (etcML), was used to classify tweets (Socher, Paulus, McCann, Tai, Hu & Ng, 2013) by type of vivid information. The dataset was trained to identify and categorize tweets by type of vivid information, emotional, spatial, or temporal (see Table A-2 in Appendix). To compare the categories of vivid information, each type was assigned a number: emotional = 1, spatial = 2, temporal = 3.

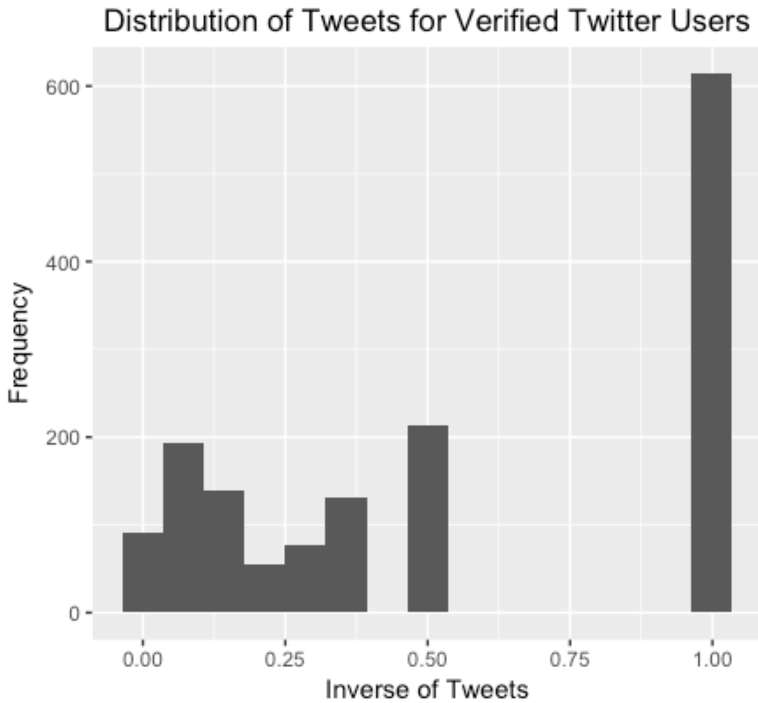

Figure 7. Inverse Transformation Tweets Generated by Verified Users

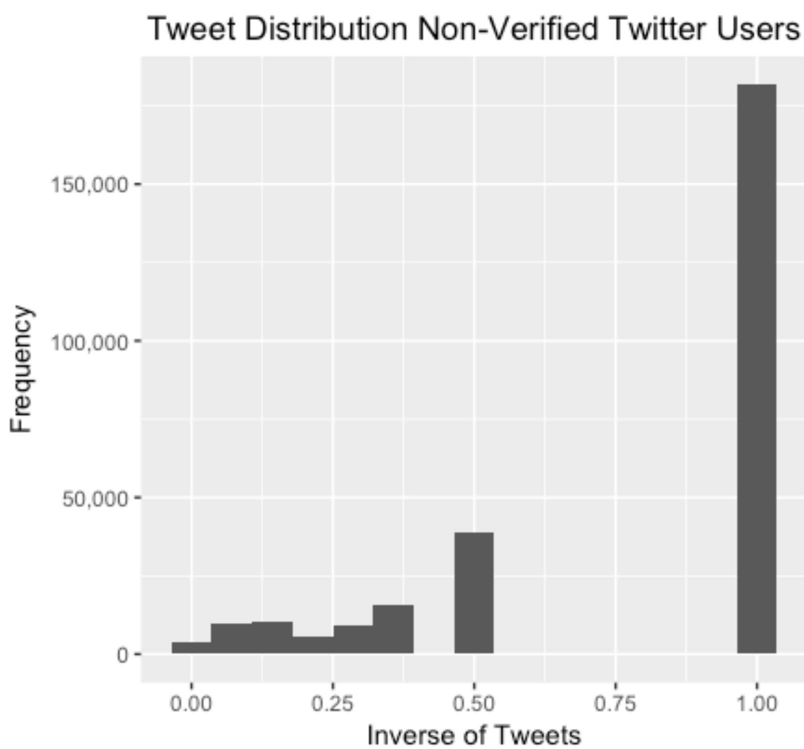

Figure 8. Inverse of Tweets Generated by Non-Verified Users

For both verified users and non-verified users, emotional information was the primary type of vivid information used in tweets, while spatial information was the least used type of vivid information (see Figure 9 and Figure 10). Further for verified users, emotional information accounted for 53% of all vivid information content, while spatial information accounted for 13% and temporal information accounted for 34%. For non-verified users, emotional information accounted for 71% of all vivid information content, while spatial information accounted for 7% and temporal information accounted for 22%.

Gatekeeper Location

The follower distributions were highly skewed for both verified and non-verified users, so median and mode were calculated to further measure centrality.

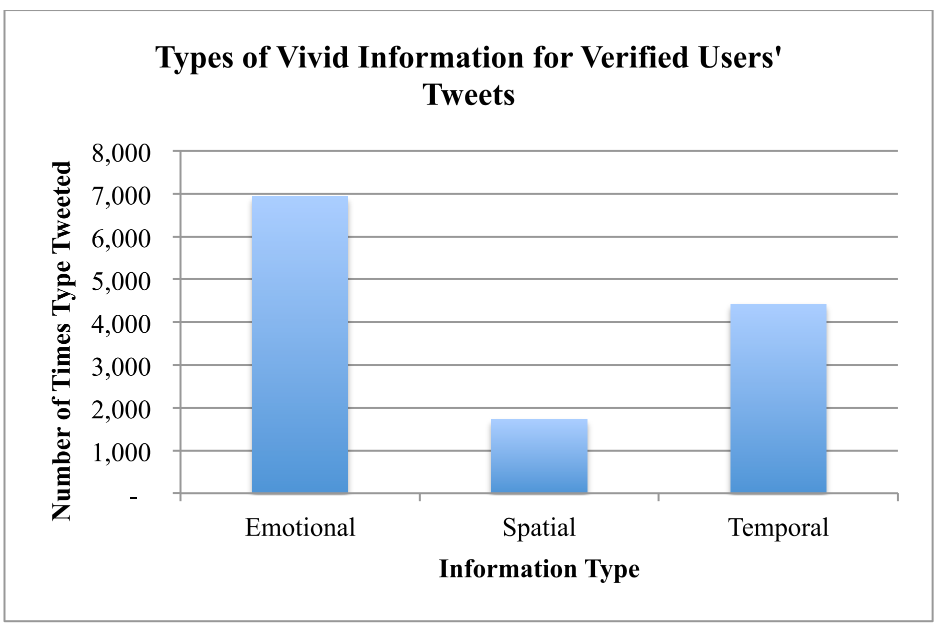

Figure 9. Distribution of Vivid Information Topics for Verified Users

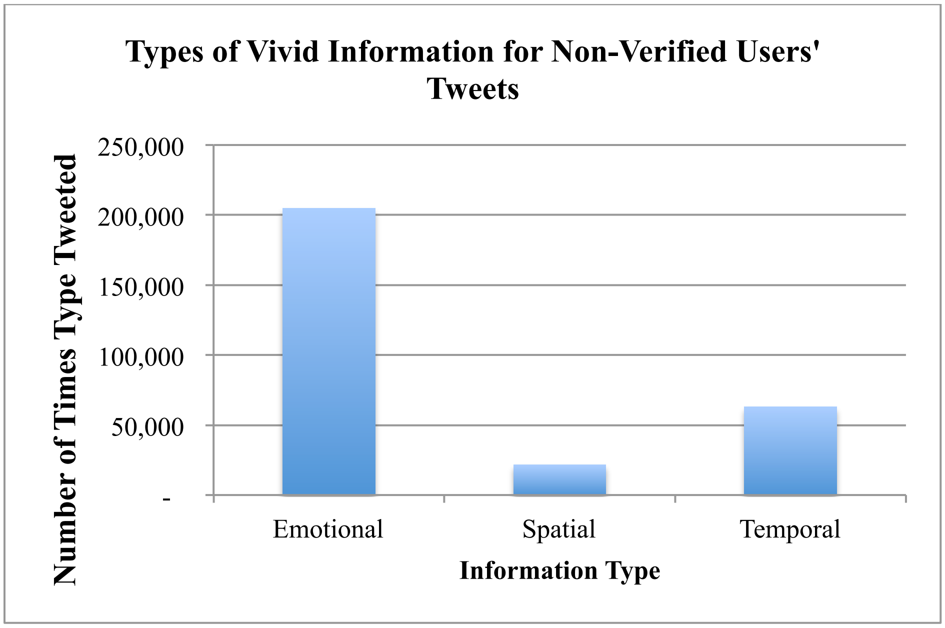

Figure 10. Distribution of Vivid Information Topics for Non-Verified Users

As a whole, the CES Twitter chatter network had a mean betweenness centrality measure of 948, a mean closeness centrality measure of 0.052, and a mean eigenvector centrality measure of 0.0019. For networks of verified users, the mean betweenness centrality measure was 167, the mean closeness centrality measure was 0.0069, and the mean eigenvector centrality measure was 0.0029.Further, individual, verified users had follower counts ranging from 103 follower to 342,541,563 followers, with mean = 2,208,711, median = 104,163, mode = 103 (see Figure 11). For networks of non-verified users, the mean betweenness centrality measure was 0.038, the mean closeness centrality measure was 0.028, and the mean eigenvector measure was 0.0031. Further, individual, non-verified users had follower counts ranging from 1 follower to 183,191,924 followers, with mean = 10,226, median = 258, mode = 1 (see Figure 12).

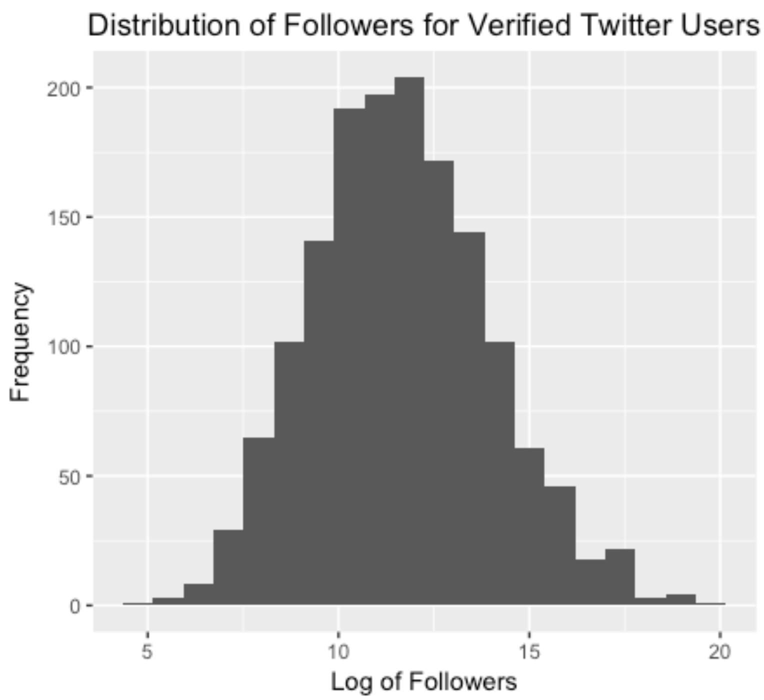

Figure 11. Log Transformation of Followers for Verified Users

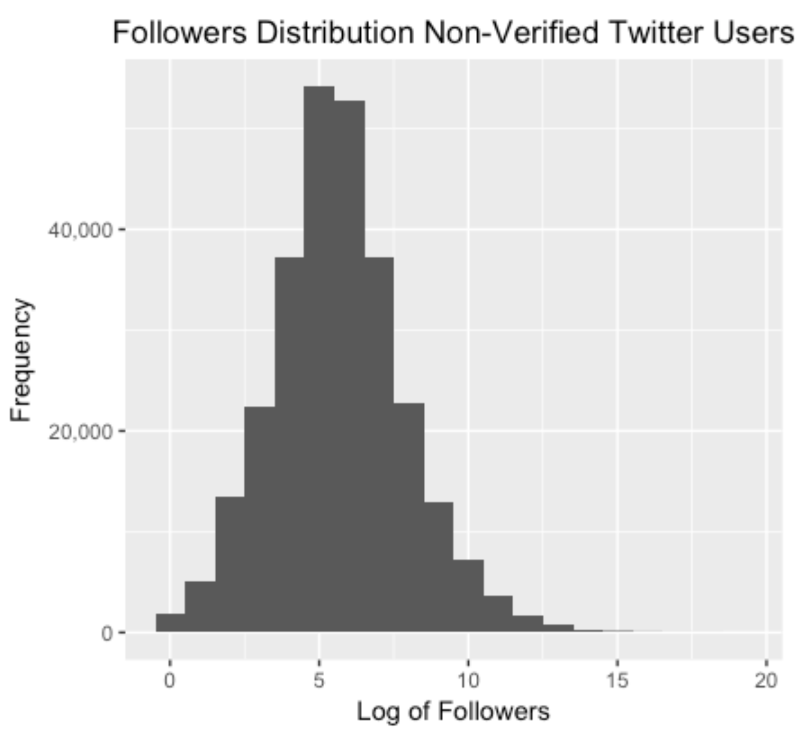

Figure 12. Log Transformation of Followers for Non-Verified Users

According to Kutner et al (2005), when normality distributions are violated, transformations are useful. The location datasets naturally follow a Poisson distribution, where the likelihood of followers is small. Due to the inherent skew in the data a log transformation was applied to reduce the skew and normalize the data.

Diffusion

Tweets from verified users had a greater likelihood of diffusion. Comparing figures 13 and 14, the diffusion curve plateauing more quickly for non-verified users than for verified users. This shows a slower rate of diffusion for non-verified users compared to verified users.For both groups the ramp up in diffusion began on the first day of the CES conference. Also, the diffusion curve continued to increase after the conference ended for verified users. This was not true for non-verified users.

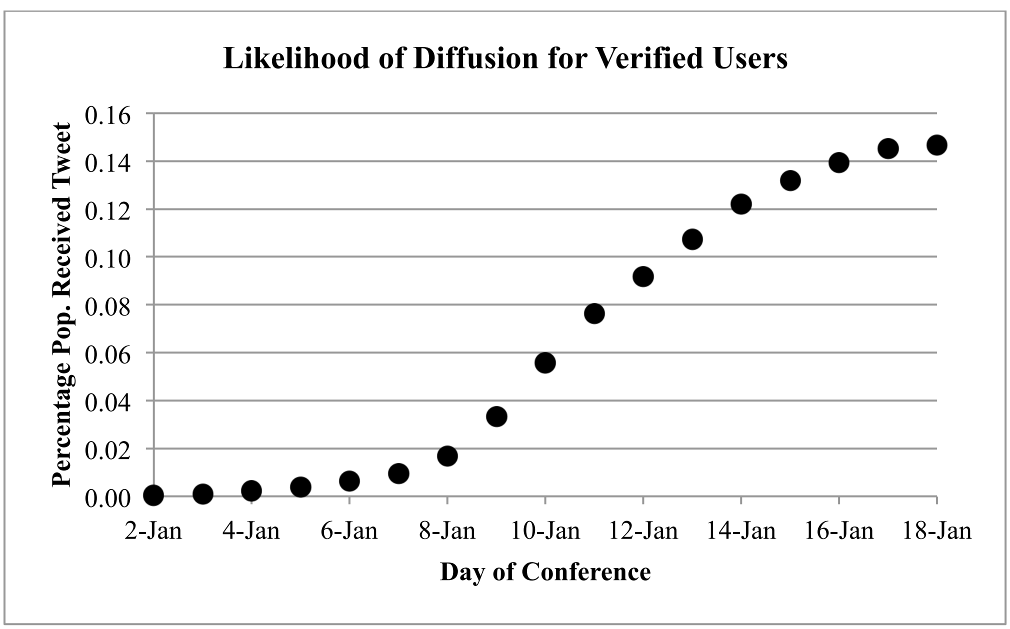

Figure 13. Cumulative Probability Plot of Diffusion by Verified Users

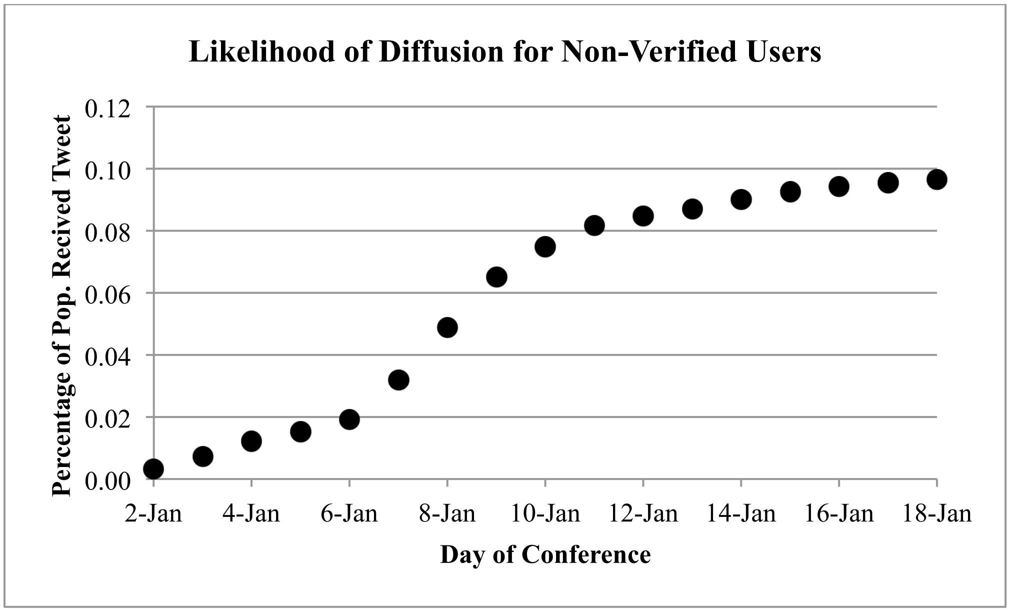

Figure 14. Cumulative Probability Plot of Diffusion by Non-Verified Users

Structural Equation Model

The proposed model hypothesized that there were significant causal relationships between five constructs of gatekeeper authority, gatekeeper activity level, vivid information, gatekeeper location, and diffusion. The causal relationship represented the four hypotheses in the path model. Structural equation modeling (SEM) was used to evaluate the causal relationships hypothesized in the proposed model. SEM is a confirmatory technique used to determine the validity of the theorized model. SEM defines a structure of the covariance matrix. Once the model's parameters are estimated, the model-defined covariance matrix is compared to the empirical matrix to determine the probability of the theoretical model.

To estimate the hypothesized interaction effect of vivid information and gatekeeper location, all variables were standardized and an interaction variable, location multiplied by information, was created (see Figure 15). Chi square was chosen as it assesses the degree of divergence between the predicted correlation matrix and the sample correlation matrix (Hooper, Coughlan & Mullen, 2008). The hypothesized model is not consistent with the data, as shown through a chi square test of global fit, χ^2 = (7, N = 275,291) = 622,957.526, p < .000.

Interaction Effect → Diffusion and Receiver Location → Diffusion parameter estimates are not statistically significant, while Gatekeeper Authority → Gatekeeper Activity, Gatekeeper Activity → Vivid Information, and Vivid Information → Diffusion parameter estimates are statistically significant (see Table 1). Figure 15 shows the standardized model as estimated by AMOS.

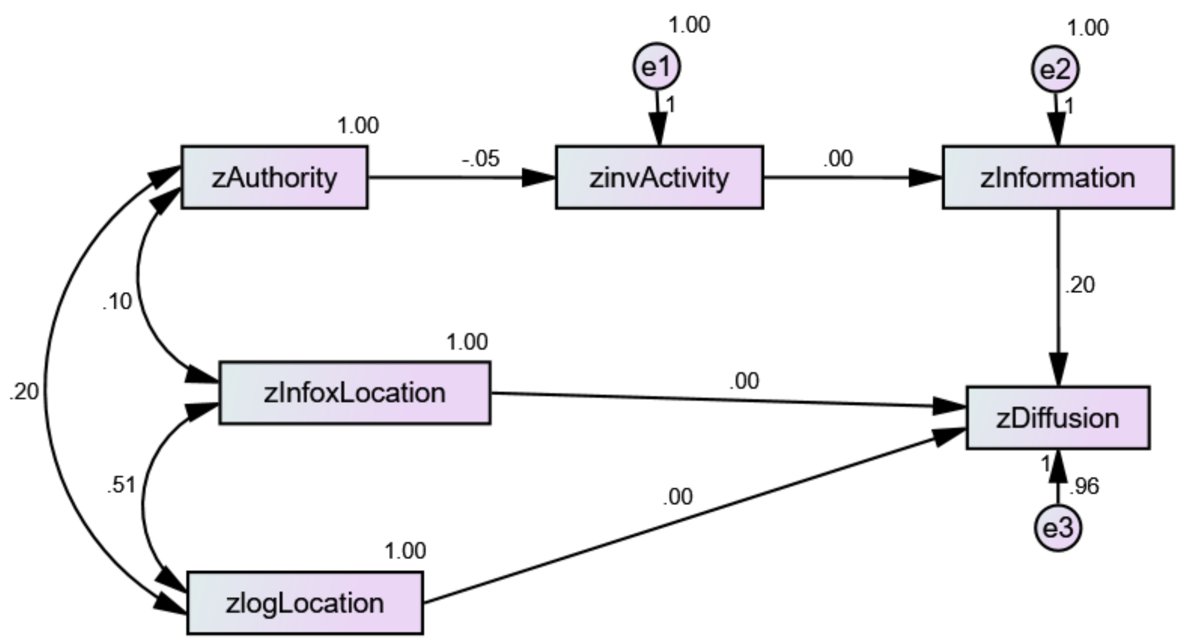

Figure 15. Structural Equation Model Estimates of Gatekeeper Effects on Information Diffusion (Standardized Solution; N = 275,291)

Table 1. Structural Equation Model Path Estimates

			Estimate	S.E.	C.R.	P	Label
zinvActivity	<---	zAuthority	-.050	.002	-26.265	***	
zInformation	<---	zinvActivity	.004	.002	2.016	.044	
zDiffusion	<---	zInformation	.198	.002	105.983	***	
zDiffusion	<---	zInfoxLocation	.002	.002	.866	.386	
zDiffusion	<---	zlogLocation	.000	.002	.000	1.000	

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

The study of diffusion in social systems is the study of the dissemination of culture. Understanding the processes of diffusion provides insight into how tangible and intangible cultural objects are spread within and across social groups. This is an important phenomenon to understand as this is how human behavior is transmitted. More broadly, diffusion is the process by which cultures evolve or die out. Social networks are a means by which to observe the diffusion process as social networks are the structures that connect people together. These connections are relationships, such as familial, friendship, work, etc. The connections can be strong or weak and both types are necessary for diffusion to occur.

This study took a network approach to hypothesize a theoretical model of diffusion within the Twitter social network (see Figure 3). A Gatekeeping Theory Framework was used. Four hypotheses were proposed:

1. The more authority a gatekeeper is perceived to have, the more active the gatekeeper will be in the network.

2. The more active the gatekeeper is, the more likely the gatekeeper will be to shape the form of the information.

3. The more closely connected the gatekeepers are to network bridges the more likely the information will diffuse quickly and completely through the network.

4. Information form will interact with network receivers such that more vivid information will more quickly and completely diffuse through the network.

The hypothesized model is not consistent with the data. Three of the hypothesized paths, Gatekeeper Authority → Gatekeeper Activity, Gatekeeper Activity → Vivid Information, and Vivid Information → Diffusion were found to be statistically significant. However, the parameter estimates for those paths are small, -.05, .004, and .198 respectively. This could indicate that the paths are of little practical significance. This could also indicate that mediating variables exist between the variables that were not considered. Descriptive analysis of the Twitter dataset shows that differences exist between verified and non-verified users. Further inferential investigation could reveal causal explanations for the differences.

It could also be the case that diffusion behavior is different within the Twitter social network than within other digital social networks or within offline social networks. Relationships on Twitter are asynchronous, so information diffusion within the network could behave similarly to information diffusion of mass media. With mass media information is diffused to unidentified recipients, as opposed to identified recipients within social networks with reciprocal social relationships. Therefore, it could be the case that the vividness of information is crafted differently for unknown recipients than it would be for known recipients. Further investigation replicating this study across different digital and offline social networks is needed to explore this hypothesis.

Three types of vivid information were explored in this study: emotional, spatial, and temporal. Future work should explore these individually. Parsing the effects of the three vivid information types individually was not the focus of this study. These three types could be negatively interacting in combination in this study. It is also possible that one or more types of vivid information not identified in this study have a greater impact within the Twitter network.

Future work should explore the impact of vivid information on diffusion within Twitter and other digital and offline social networks.

Limitations of the Approach

This method used historical information; therefore there is no control or manipulation of any of the variables of interest. There may be unknown mediating variables that are interacting with the variables of interest to this study. This study also focuses on diffusion of information about a well-established, technology driven, event that occurs within a discreet time frame. It could be the case that diffusion patterns differ outside of the technology industry or within more niche and/or less well-known events. Future studies should replicate and extend this study to conversations beyond technology including niche topic areas. Further, as the event observed in this study took place within a discreet time frame, it is difficult to generalize the results of this study to conversations occurring outside the bounds of a given time frame.

Also, Twitter doesn’t provide demographic information of its user base. Since one can self-select to use the service, it could be the case that Twitter attracts a population that is different in some meaningful way from the general population. It is, therefore, difficult to generalize the results of this study beyond the Twitter user base. Future studies should extend this study to additional social networks, both online and offline.

Future Areas of Research

It is important to understand diffusion of information in Twitter, and social media more broadly, because it is an increasingly significant communication tool. Gottfried and Shearer

(2016) found that 62% of U.S. adults get news on social media, and increase of 14 percentage points over a four-year period. Further, Gottfried and Shearer (2016) also found that 59% of Twitter users get news on that platform. Fan and Gordon (2014) found that users spend 20% of their time the Internet on social media and Twitter users send 340 million tweets. This means that more and more U.S. adults are turning to Twitter and other social media platforms as a primary source of information. It is, therefore, necessary to understand how information spreads on Twitter and other platforms.

This study considered vivid information of the tweet, which consists of only 140 characters. Tweets can also include images, video, and outbound links, none of which were considered for this study, but could have a causal link with information diffusion. For instance, Bandari, Asur, and Huberman (2012) found characteristics of the news articles, specifically the source, category, language within the article, and named entities in the article, linked to within a tweet to be predictive of spread of the news articles. It is worth exploring tweet content holistically in future research.

The current study used the Consumer Electronics Show as a diffusion topic to follow. Guille, Hacid, Favre and Zighed (2013) found that bursty topics, or topics made up of popularity patterns within a time interval, are necessary to study information diffusion in social media. Figure 4 shows a burst of popularity for CES, but no successive bursts. It could be the case that the topic was not popular enough to generate diffusion through Twitter. The proposed model should be applied to a series of topics to better know if there is fit with the data.

Sentiment of a tweet can impact the diffusion of a tweet (Ferrara & Yang, 2015; Wang, Lin, Jin, Cheng, & Yang, 2015). The current study did consider emotion as a component of vivid information. However, emotion was measured as a binary – yes the tweet conveyed emotion or no the tweet did not convey emotion. Sentiment measures the differing emotions of a tweet, such as happy, angry, sad, neutral, etc. Ferrara and Yang (2015) found that negative tweet diffused more rapidly, but positive tweets diffused more broadly. Wang et al (2015) also found that happy tweets diffuse more broadly and that angry tweets are unlikely to be re-tweeted. Sentiment of tweets should be considered in further research of the proposed model.

Regression Model to Explore

A regression model was briefly explored to identify potential areas of exploration. Multiple linear regression was to develop a model for predicting diffusion from vivid information, location, and the interaction of vivid information and location. The three predictor model was able to account for 36% of the variance in diffusion, $F(3, 1{,}511) = 98.20$, $p < .001$, Adjusted $R^2 = .36$, 90% CI [.83, 1.24].T

This model is created from verified users only (see Table 2). Vivid information and location account for 36% of the variance in the model. Additional variables should be explored to further account for the variance in the model and more accurately predict diffusion. It could be the case that different variables are important for diffusion originating with verified users than diffusion originating with non-verified users.

Diffusion of Innovations

Diffusion of Innovations is a theory in which Rogers (2003) proposes five categories of adopters, five stages to the adoption process, and four main elements of diffusion. It is a robust theory that has been extended into multiple disciplines. Taking a Diffusion of Innovations approach to the proposed theoretical model should be explored.

Table 2. Verified User Multiple Linear Regression Model Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
	(Constant)	7.749E-15	.021		.000	1.000	-.040	.040
	zlogLocation	.837	.039	.837	21.690	.000	.761	.913
	zInfoxLocation	-1.487	.109	-1.487	-13.616	.000	-1.701	-1.273
	zInformation	1.032	.104	1.032	9.909	.000	.828	1.237

a. Dependent Variable: zDiffusion

Rogers (2003) defines the functions of the actors and the network in the diffusion process. Specifically, the five categories of adopters, innovators, early adopters, early majority, late majority and laggards, each have a different role in the diffusion process. The innovators and the early adopters drive the success or failure of diffusion. For these actors, the proposed model could be modified to further explain their knowledge and persuasion processes. The knowledge stage provides first exposure to the innovation. In the persuasion stage, the individual actively seeks information about the innovation. It could be hypothesized that innovators and early adopters act as gatekeepers, exposing information to their social networks. Differences in the quantity and quality compared to other groups should be explored. Further, differences information seeking behaviors between groups, defined in the persuasion stage, should be explored.

Specifically, the role of vivid information created by the innovators and early adopters should be explored. It could be the case that innovators and early adopters produce more content using more vivid information than other actors in the network. Further, Rogers (2003) identifies the communication channel as a main element of the diffusion process. Communication channels are the means by which information is transmitted between actors in a social system. It could be the case that a social media communication channel behaves differently than offline or other web based communication channels. It is worth testing the inclusion of content production, vivid information, and social media channels within the Diffusion of Innovations framework.

Implications

The diffusion process is difficult to measure because, even within a defined network such as Twitter, behavior is complex. Social trends rapidly change. Network connections are quickly created and easily dissolved. Mediated communication, specifically web-based communication, provides a feeling of anonymity to the user that may lead to differences in behavioral patterns from in-person interactions. It is, therefore, difficult to account for every variable involved in a diffusion process. Social media has a unique variable that needs to be considered in all diffusion research, the content algorithm.

Each social media platform uses an algorithm to determine what content to show to whom and when to show it. The algorithm is basically a personalized recommendation of the user-generated content of a given user's network. For example, the algorithm may prioritize content created by a significant other over content created by extended family members. Theoretically, a given user could never see content from a user they are following if it is not prioritized by the algorithm. More likely, content is de-prioritized such that it is improbable that it will be seen. The priorities given are individual to each user and predictive models used to generate the priorities are constantly being modified.

Though the intricacies of the algorithms are unknown, there are some generalities that are understood. For instance, sponsored content, or targeted content, generally has some weighted priority because the advertisers or brands that create the content pay the social media sites to target it to particular user groups. Sponsored content is labeled as such, though the labels can be subtle. Also, content that a user frequently interacts with has some weighted priority. For instance, a user might frequently like or retweet content from a particular user they are following. When that is the case, the content from that user they are following will be given some priority. Factors such as how often users log into the platform, how connected users are within a network, whom the user is interacting with (brands, news, people, etc.), and the length of time since the last interaction with a user are also given some priority. The algorithms are unique to each platform.

A real world example of the role algorithms play in the diffusion process is the proliferation of fake news. Fake news consists of deliberate misinformation, conspiracy theories, and hoaxes. Throughout the 2016 presidential election cycle news articles making false or hyperbolized claims proliferated social media platforms. For example, a conspiracy theory colloquially referred to as ‘pizzagate’ went viral on Twitter, 4chan, and Facebook. The conspiracy theory posited that the campaign manager for the Democratic presidential candidate used coded messages in his email communications referring to human trafficking of children through local Washington, D.C. pizza restaurants. The messages were so believable to some users that one person took it upon himself to self-investigate the incident resulting in gunshots fired into a pizzeria.

Pizzagate exemplifies the role of the algorithm. Pizzagate articles were introduced to networks by popular, highly followed, conspiracy theorists. Due to the scandalous nature of the content it was rapidly shared and diffused through social networks. Reputable news articles reporting the debunked conspiracy theory also diffused through social networks. However, those reputable articles didn't diffuse into all social networks on all social media platforms. The reason they didn't penetrate all networks is because the algorithm didn't evenly prioritize to the story to all users. Users don't have to follow reputable news organizations on social media. Therefore, all users don't see content that is introduced by those organizations, even if it has diffused through much of the social platform.

Because the algorithm plays such an important, and ever increasing, role in what content is shown to which users and when, it is important to include as a variable in diffusion research. Information can only be diffused if it is seen. More broadly, it is important to consider the algorithm in social media research as there are potential algorithmic effects on mediated social behavior, such as behavioral implications from content that was or wasn't seen.

As digital media continues to supplant traditional media as the go-to source for news, information, and entertainment the role of the algorithm will only increase. There must be some way to organize ever-increasing volumes of content. The algorithm has the ability to showcase or suppress any information on the Web. Information-prioritizing algorithms proliferate the Web, from e-commerce sites, such as Amazon, to search engines, such as Google, to social media, such as Twitter. With so much content housed on these sites, it is impossible for human beings to police the algorithms and ensure content prioritization is consistent across users. The algorithm relies on end users to know what they are looking and to be literate about sources. This is a paradigm shift in the media landscape. Researchers must seek to understand how algorithms are impacting human behaviors.

LIST OF REFERENCES

Adamic, L. & Adar, E. (2005). How to search a social network. *Social Networks*, 27, 187-203.

Apdevries (Creator). (2009). *Bass diffusion model* [digital image]. Retrieved from Wikimedia Commons website: https://commons.wikimedia.org/wiki/File:Bass_diffusion_model.svg

Bakshy, E., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011). Everyone's an influencer: Quantifying influence on Twitter. *Proceedings of the 4th ACM International Conference on Web Search and Data Mining.*

Barabasi, A. L. & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286, 509-512.

Barzilai-Nahon, K. (2008). Toward a theory of network gatekeeping: A framework for exploring information control. *Journal of the American Society of Information Science and Technology*, 59, 1493-1512.

Bass, F. M. (1969). A new product growth for model consumer durables. *Management Science*, 15, 215-227.

Bianconi, G. & Barabasi, A. L. (2001). Bose-Einstein condensation in complex networks. *Physical Review Letters*, 86, 5632-5635.

Bentley, R. A. & Shennan, S. J. (2003). Cultural transmission and stochastic network growth. *American Antiquity*, 68, 459-485.

Boster, F. J., Kotowski, M. R., Andrews, K. R., & Serota, K. (2011). Identifying influence: Development and validation of the connectivity, persuasiveness, and maven scales. *Journal of Communication*, 61, 178-196.

Burt, R. S. (1992). *Structural Holes*. Harvard University Press: Cambridge, MA.

Burt, R. S. (2004). Structural holes and good ideas. *American Journal of Sociology*, 110, 349-399.

Centola, D. (2010). The spread of behavior in an online social network experiment. *Science*, 329, 1194-1197.

Chaffee, S. H. (1975). *Political Communication: Issues and Strategies for Research.* Sage Publications: Beverly Hills, CA.

Chalmers, A. F. (1999). *What is this thing called Science?* University of Queensland Press: St. Lucia, Queensland.

Chong, C. Y., & Kumar, S. P. (2003). Sensor networks: Evolution, opportunities, and challenges. *Proceedings of the IEEE*, 91, 1247-1256.

Cosley, D., Huttenlocher, D., Kleinberg, J., Lan, X., & Suri, S. (2010). Sequential influence models in social networks. *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*.

Cosmides, L. & Tooby, J. (1989). Evolutionary psychology and the generation of culture, part II. *Ethology and Sociobiology*, 10, 51-97.

Entman, R. M. (2007). Framing bias: Media in the distribution of power. *Journal of Communication*, 57, 163-173.

Erikson, B. H., Nosanchuk, T. A., Mostacci, L. & Dalrymple, C. F. (1978). The flow of crisis information as a probe of work relations. *Canadian Journal of Sociology*, 3, 71 – 87.

Everett, R. (Creator). (2012). *Diffusion of ideas* [digital image]. Retrieved from Wikimedia Commons website: https://commons.wikimedia.org/wiki/File:Diffusion_of_ideas.svg

Framback, R. T. (1993). An integrated model of organizational adoption and diffusion of innovations. *European Journal of Marketing*, 27, 22-41.

Gargiulo, M. & Benassi, M. (2000). Trapped in your own net? Network cohesion, structural holes, and the adaptation of social capital. *Organization Science*, 11, 183-196.

Ghoshal, S. & Bartlett, C. A. (1988). Creation, adoption, and diffusion of innovation by subsidiaries of multinational corporations. *Journal of International Business Studies*, 19, 365-388.

Gieber, W. (1956). Across the desk: A study of 16 telegraph editors. *Journalism & Mass Communication Quarterly*, 33, 423-432.

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 78, 1360-1380.

Greenhalgh, T., Robert, G., Bate, P. MacFarlane, F., & Kyriakidou, O. (2007). *Diffusion of Innovations in Health Services Organisations: A Systematic Literature Review.* Blackwell Publishing, Ltd.: Oxford, UK.

Gruhl, D., Guha, R., Liben-Nowell, D., & Tomkins, A. (2004). Information diffusion through blogspace. *Proceedings of the 13th International Conference on the World Wide Web, ACM*.

Haythornthwaite, C. (1996). Social network analysis: An approach and technique for the study of information exchange. *Library & Information Science Research*, 18, 323-342.

Henningham, J. (1997). The journalist's personality: An exploratory study. *Journalism & Mass Communication Quarterly*, 74, 615-624.

Hickey, J. R. (1968). The effects of information control on perception of centrality. *Journalism Quarterly*, 45, 49-54.

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, 6, 53-60.

Jaccard, J. & Jacoby, J. (2010). *Theory Construction and Model-Building Skills: A Practical Guide for Social Scientists.* The Guilford Press: New York, NY.

Kadushin, C. (2012). *Understanding Social Networks: Theories, Concepts, and Findings.* Oxford University Press: New York, NY.

Kermack, W. O. & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 115, 700-721.

Kilworth, P. D. & Bernard, H. R. (1979). A pseudomodel of the small world problem. *Social Forces*, 58, 477-505.

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied Linear Statistical Models* (5^{th} ed.). McGraw-Hill/Irwin: New York, NY.

Lakatos, I. (1978). *The methodology of scientific research programmes: Philosophical Papers Volume 1*. Cambridge: Cambridge University Press.

Lam, P., Lo, J., Yeung, A., & McNaught, C. (2010). Examining diffusion and sustainability of e-learning strategies through weblog data. *International Journal of E-Adoption*, 2, 39-52.

Langerhans, R. B. & DeWitt, T. J. (2004). Shared and unique features of evolutionary diversification. *The American Naturalist*, 164, 335-349.

Larose, D. T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. John Wiley & Sons, Inc.: Hoboken, NJ.

Lerman, K., Integorn, S., Kang, J., & Ghosh, R. (2012). Using proximity to predict activity in social networks. *Proceedings of the 21^{st} International Conference Companion on World Wide Web.*

Lewin, K. (1951). *Field Theory in Social Science: Selected Theoretical Papers*. Harper: Oxford, England.

Mahajan, V., Miller, E., & Bass, F. M. (1995), Diffusion of new products: Generalizations and managerial uses. *Marketing Science*, 14, 79-88.

Marlow, C. (2004). Audience, structure and authority in the weblog community. *International Communication Association Conference.*

Milgram, S. (1967). The small world problem. *Psychology Today*, 1, 61-67.

Nisbett, R. E. & Ross, L. (1980). *Human Inference: Strategies and Shortcomings of Social Judgment*. Prentice-Hall: Englewood Cliffs, NJ.

Oldenburg, B. & Glanz, K. (2008). Diffusion of innovations. In Glanz, K., Rimer, B. K., & Viswanath, K. (4th Ed.), *Health Behavior and Health Education: Theory, Research, and Practice.* Jossey-Bass: San Francisco, CA.

Otte, E. & Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information sciences. *Journal of Information Science*, 28, 441-453.

Ovadia, S. (2007). Digg.com and socially driven authority. *Library Philosophy and Practice*, 9, 1-8.

Pavitt, C. (2001). *The Philosophy of Science and Communication Theory*. Huntington, NY: Nova Science Publishers, Inc.

Pelto, P. J. & Muller-Wille, L. (1972), The snowmobile revolution: Technology and social change in the Artic. In Bernard, R. H. and Pelto, P. J. eds., *Technology and Social Change*. Macmillan: New York, NY.

Reagans, R. & McEvily, B. (2003). Network structure and knowledge transfer: The effects of cohesion and range. *Administrative Science Quarterly*, 48, 240-267.

Richardson, R. J., Erikson, B. H., Nosanchuk, T. A. (1979). Community size, network structure, and the flow of information. *Canadian Journal of Sociology*, 4, 379 – 392.

Robinson, S. (2006). Gateway or gatekeeper: The institutionalization of online news in creating an altered technological authority. *International Symposium in Online Journalism*.

Rogers, E. M. (2003). *Diffusion of Innovations* (5th Ed.). The Free Press: New York, NY.

Romero, D. M., Meeder, B., & Kleinberg, J. (2011). Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on Twitter. *Proceedings of the 20th International Conference on World Wide Web.*

Russell, M. A. (2011). *Mining the Social Web*. O'Reilly Media: Sebastopol, CA.

Schnettler, S. (2009). A structured overview of 50 years of small-world research. *Social Networks*, 31, 165-178.

Shoemaker, P. J. & Vos, T. P. (2009). *Gatekeeping Theory*. Routledge: New York, NY.

Sigal, L. V. (1973). *Reporters and Officials: The Organization and Politics of Newsmaking*. D. C. Heath: Lexington, MA.

Smith, A. (2011). Why Americans use social media. Pew Internet & American Life Project, November 15, 2011, http://www.pewinternet.org/Reports/2011/Why-Americans-Use-Social-Media/Main-report.aspx, accessed on October 29, 2012.

Smith, A. & Brenner, J. (2012). Twitter use 2012. Pew Internet & American Life Project, May 31, 2012, http://www.pewinternet.org/Reports/2012/Twitter-Use-2012/Findings.aspx, accessed on October 29, 2012.

Socher, R., Paulus, R., McCann, B., Tai, K. S., Hu, J., Ng, A. Y. (2013). Demonstration: etcML,com – Easy text classification with machine learning. *Advances in Neural Information Processing.*

Tutte, W. T. (2001). *Graph Theory*. Cambridge University Press: New York, NY.

Valente, T. W. (1996). Social network thresholds in diffusion of innovations. *Social Networks,* 18, 69-89.

Valente, T. W. & Davis, R. L. (1999). Accelerating the diffusion of innovations using opinion leaders. *The Annals of the American Academy of Political and Social Science*, 566, 55-67.

Walker, J. L. (1969). The diffusion of innovations among the American states. *The American Political Science Review*, 63, 880-899.

Wasserman, S. & Faust, K. (2008). *Social Network Analysis: Methods and Applications.* Cambridge University Press: New York, NY.

Watts, D. J. (2004). The "new" science of networks. *Annual Review of Sociology*, 30, 243-270.

Watts, D. J. & Dodds, P. S. (2007). Influentials, networks, and public opinion formation. *Journal of Consumer Research*, 34, 441-458.

Wejnert, B. (2002). Integrating models of diffusion of innovations: A conceptual framework. *Annual Review of Sociology*, 28, 297-326.

Wellin, E. (1955). Water boiling in a Peruvian town. In Paul, B. D., ed., *Health, Culture and Community*. Russell Sage Foundation: New York, NY.

Wellman, B., Salaff, J., Dimitrova, D., Garton, L., Gulia, M., Haythornthwaite, C. (1996). Computer networks as social networks: Collaborative work, telework, and virtual community.

Wheelock, C., E., Wheelock, A. M., Kawashima, S., Diez, D., Kanehisa, M., van Erk, M., Kleeman, R., Haeggstrom, J. Z., Goto, S. (2009). Systems biology approaches and pathway tools for investigating cardiovascular disease. *Molecular BioSystems*, 5, 588-602.

White, D. M. (1950). The “gate keeper”: A case study in the selection of new. *Journalism Quarterly*, 27, 383-390.

White, H. C., Boorman, S. A., & Breiger, R. L. (1976). Social structure from multiple networks. I. Blockmodels of roles and positions. *American Journal of Sociology*, 81, 730-780.

Yang, J. & Counts, S. (2010). Predicting the speed, scale, and range of information diffusion in Twitter. *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*.

APPENDIX

Table A-1. List of Variable Definitions in Theoretical Model

Variable	***Definition***
Authority	Each link or follower accumulated on the Web.
Activity	The extent to which a gatekeeper seeks out and distributes information.
Vivid Information	That which relates directly, spatially, or temporally to a known person or event.
Location	The network actors to whom the gatekeeper is connected.
Diffusion	The extent to which all network actors are privy to the information

Table A-2. Vivid Information Training Dataset

Vivid Information Type	***Training Tweet***
Emotional	Samsung's CES TV surprise is blowing our minds: http://t.co/VDpQvYSv
Emotional	trying not to get overwhelmed by the awesomeness.
Emotional	#MissingYou MT @StinaSanDiego: Slightly sad but relieved that I won't be attending #CES.
Emotional	@mashable Good God Someone Please Make CES Interesting.
Emotional	These are so awesome.
Emotional	You can do better than "awesomeness" but we think its a good start RT @gbengasosan: @intlCES ,1 week of "awesomeness".
Emotional	Now I'm anxious.
Emotional	Via @RWW 6 Reasons This Could Be The Most Boring CES Ever
Emotional	RT @arcieri: We love this industry because of its constant change. It's addictive:
Emotional	Mind just blown by the IllumiRoom concept Microsoft has shown off at CES.
Emotional	That sucks.
Emotional	Your HDTV sucks now http://t.co/9ugcsYxR
Emotional	While all of you are enjoying #CES, I will be applying to game devs. You could say I'm slightly nervous. #WHATIFTHEYHATEME
Emotional	My mRobo killed it . Is it weird that I'm proud of a robot ? #CES #Tosy http://t.co/sQjce9xg
Emotional	Excited to give @IntlCES opening speech, pumped to get show officially started!
Emotional	Two happy dudes #2013CES http://t.co/wwUvdmej
Emotional	@rikkiends so sad that consumers enable brands that practice such #slimeballmarketing
Emotional	Was surprise/happy to see at #ces the ubiquity of @android
Emotional	We're happy to grow our partnerships and as a company. Glad to showcase at #2013ces.
Emotional	you might even get angry at @NinaFrazier's favorite photos from CES 2013
Spatial	I'll be easy to find - in the #AMDSurround tent (next to the Registration tent) directly in front of the LVCC.
Spatial	Coming soon to a home near you
Spatial	There are 2 @RadioShack locations near the LVCC
Spatial	Finally inside the Samsung #2013CES press conference. Poised for a big announcement to match the queues outside.
Spatial	Find Newegg TV & the Samsung SSD Angels Airstream Trailer outside Central Hall
Spatial	The first 2 people to come to the Nokia bus outside LVCC will win

Table A-2. (continued)

Vivid Information Type	*Training Tweet*
Spatial	@drew am around
Spatial	is you are around don't be shy and come hang out with us!
Spatial	CES 2013: what's around, what's up and what's down
Spatial	We'll be there though! --PG #CES13
Spatial	@intlCES I mean I'll be there....
Spatial	Vegas here I come.
Spatial	Near Field Comm? @broadcom's there!
Spatial	I'll probably be inside the AMD tent out in front of the convention center
Spatial	I'm sandwiched between #NMX #CES and #NAIAS. :-0
Spatial	This outdoor walkway between North and South Halls is a tremendous respite
Spatial	Logitech shuttles today for FREE rides between the @TIvegas, @AriaLV hotels & LVCC.
Spatial	Want to see inside the DTS #CES Cinema booth?
Spatial	stage is in the South Hall, above the Starbucks. see you there!
Spatial	it's @JenFriel behind her fancy-pants door to the master suite.
Spatial	We're in front of the main CES entrance at Central Hall.
Spatial	Here's @gpatricksmith in front of the Sanus booth #ces
Temporal	Only 5 more days and a bit till the largest consumer electronics show.
Temporal	Save the date: January 9th: http://t.co/KDECGoX6 #CES2013
Temporal	Today feels like Monday, and then we leave for the CES time warp on Friday.
Temporal	Counting the seconds! RT @AeroMobile: Anyone else counting down to #2013CES? @intlCES
Temporal	Need a day count? #5days #2013CES
Temporal	Register by tomorrow, Jan. 4, for the Entertainment Matters Party at #2013CES
Temporal	Social Hour is the new Tweet Up, Jan 9, 4-6 pm in the V Bar, Venetian
Temporal	CES kicks off this weekend
Temporal	Join Us Today at 11 a.m. EST for a Google Hangout About CES 2013 http://t.co/P7BOSOJQ
Temporal	Yesterday, Samsung unveiled the NX300
Temporal	Joining @mashable for Google+ Hangout today.
Temporal	I'll have @AdamSessler on the show later today
Temporal	@DuffingtonQC Saturday and then every day after that during CES!
Temporal	Tomorrow I'm participating in the @IntlCES panel
Temporal	Did you tune into yesterday's #SamsungCES Press Conference?
Temporal	@DuffingtonQC Saturday and then every day after that during CES!

Table A-2. (continued)

Vivid Information Type	***Training Tweet***
Temporal	live from #2013CES later today. Until then, watch the 1st Verge Cast
Temporal	Bit of a night owl?
Temporal	saw a lot of innovative things today at #ces2013 #2013CES
Temporal	@engadget went hands-on with the Pebble today at #CES

VITA

Rebecca Anderson was born in Manchester, Connecticut to the parents of Linda and Gregory Anderson. She has one older brother, Nathan. She attended St. Petersburg High School in St. Petersburg, FL, and later attended Flagler College where she earned her Bachelor's degree in Psychology. After working in the media industry for three years, Rebecca entered graduate school at the University of South Florida where she earned a Master's of Science degree in Management Information Systems with a focus on the social web. Following completion of her Master's degree, she entered the Ph.D. program in Information Sciences at the University of Tennessee. At the University of Tennessee she held positions as both a teaching associate and research assistant. Rebecca is a Data Scientist in the media industry.