The Influence of Network Structures and Information Seeking Uncertainty on Information Seeking Behavior

By

Ziad Albert Matni

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the Degree of

Doctor of Philosophy

Graduate Program in Communication and Information

written under the direction of

Chirag Shah, Ph.D.

And approved by

New Brunswick, New Jersey

October, 2018

ABSTRACT OF THE DISSERTATION

The Influence of Network Structures and Information Seeking Uncertainty on Information Seeking Behavior

By Ziad Albert Matni

Dissertation Director:

Chirag Shah

People utilize their social networks to get to resources, tangible or otherwise, that aid them in their everyday lives. Information scientists have shown that network characteristics of information structures can indeed influence human information searching and browsing behavior. However, we do not have enough detail on what particular network characteristics may influence information seeking behavior. There is an incomplete picture of the how network structures influence people's information seeking interactions over time.

In this dissertation, I will look at some quantifiable behavioral dynamics of individuals who are seeking information using different social network structures over time. This research can shed light on our understanding of the interplay between human behavior and the environmental structures that people find themselves both being influenced by and influencing.

This study utilizes a custom-built Web-based tool that simulates an information-seeking scenario via various network structures and has participants utilize it to achieve a stated goal of collecting answers to questions from others in their network. The tool allows a finite amount of interactions, thus limiting the participants' engagements to a defined set of allowable actions. As all participants go through the simulation, the system logs their actions over time and measurements are taken in timed intervals of certain information seeking behaviors of the participants and changes that they create in their network topologies. The participants run through two types of networks: one with a scale-free topology one node has a disproportionate high number of connections compared to the other nodes and another with two sub-networks connected to one another via two structural holes. Both networks differ significantly in structure, but are very similar in network density and in average node degree centrality.

This dissertation aims to contribute to the theories of information seeking in social network environments, as well as to social network theory as it pertains to human information behavior. From a practical standpoint, this dissertation aims at giving scholars another way to study human behavior through the lens of social networks by providing them with a sophisticated computer-mediated platform to collect log-based data of human behavior in simulated networked environments.

ACKNOWLEDGEMENTS

This dissertation has been a labor of love that began 7 years ago when I decided that engaging in academic research and university teaching was my calling. The journey has been challenging and exhilarating for me and I am a wiser and happier man for it.

I would like to thank my academic advisor, Dr. Chirag Shah, for his mentorship, encouragement, and guidance throughout my years at Rutgers. Dr. Shah has been instrumental in my development as a scholar and researcher in information science and is always a source of great ideas and conversation. My work on data gathering tools like SIMPSON and SOCRATES would not have been possible without Dr. Shah's guidance. I would like to also thank the members of my dissertation committee, Dr. Nicholas Belkin, Dr. Marya Doerfel, and Dr. Winter Mason, who were very generous with their time and advice. I feel lucky to have had, not only such prestigious academics of the highest caliber on this dissertation committee, but also ones so thoughtful and kind.

I would like to expressly thank Mr. Kevin Albertson who helped immensely with the development of the SIMPSON tool and who saved my data gathering efforts more than once over long-distance phone calls between Santa Barbara and New York City. My deep gratitude also goes out to my colleagues at the School of Communication and Information (SC&I), including my wonderful colleagues in the InfoSeeking Lab: Dr. Dongho Choi, Dr. Erik Choi, Dr. Roberto González-Ibáñez, and Dr. Matthew Mitsui. Of the many faculty and staff at SC&I whose friendship, wisdom, and advice were immeasurable over the course of my time at Rutgers as a doctoral student and part-time lecturer, I'd like to particularly thank Dr. Keith Hampton, Dr. Mor Naaman, Dr. Mark Aakhus, Dr. Craig Scott, Dr. Laurie Lewis, Dr. Claire McInerney, Dr. Stew Mohr, Dr.

Sharon Stoerger, Dr. Matthew Weber, Dr. Jennifer Theiss, Dr. Marija Dalbello, Dr. Marie Radford, Dr. Anselm Spoerri, Dr. Karen Novick, Mr. Jon Oliver, Dr. Jonathan Potter, and, of course, my good friends Dr. Marie Haverfield, Dr. Amanda Carpenter, Dr. Lora Appel, and Mr. Punit Dadlani.

There were certain students at the University of California, Santa Barbara, who assisted me with data collection tasks that I'd like to acknowledge here as well: Mr. Vincent Chan, Ms. Samantha Perez, Ms. Angeline Benitez, and Ms. Vanessa Chacon.

I want to express my heartfelt gratitude to my former professor and mentor, the late Dr. Ronald P. Olowin, who inspired me many years ago during my undergraduate studies at Saint Mary's College to become engrossed in science and research. He continues to be the ultimate blueprint for what I consider a great teacher and well-rounded university professor to be.

I never had less than super-enthusiastic support from my mother and father, Sanaa and Albert, and my sister, Nada, and for that, I love them and thank them dearly. Last, but certainly not least, I would like to express my deep gratitude, admiration, and love to my wife Dr. Jennifer Gibbs who unconditionally supported me in every way imaginable and believed in my academic adventure from the very start. She was, and continues to be, my biggest cheerleader throughout the successes and the challenges that I invariably encounter. I'd like to also thank my darling children Maxwell and Zeina for their unreserved love and general cuteness. May this work inspire them one day to achieve academic accomplishments of their own.

Table of Contents

List of Tables

List of Figures

CHAPTER 1: INTRODUCTION

1.1. Problem statement

We know that people often seek information through access to their network-structured ties (social or otherwise), but most studies focus on explaining the dynamics therein through the qualities of relationships (for example, strong versus weak ties) or the characteristics of the potential for future support, most notably using social capital to explain those particular dynamics. Some scholars in Information Science have shown that network characteristics of information structures (like the World Wide Web) can influence human information searching and browsing behavior, while others have theorized that social network topologies have an influence as well, but without much detail on what network characteristics may be at play. So scholars do not have a complete picture of the interplay of people's information seeking interactions over time within the social networks that they utilize.

We have many user-centric models that inform us about information seekers' needs when looking for information (Belkin, Oddy, & Brooks, 1982; Wilson, 1981), how individuals interact with their information sources (Ellis, 1989), what processes information seekers go through to acquire information (Bates, 1989; Kuhlthau, 1991; Marchionini, 1997), how people's temperaments guide their information seeking activities into their daily lives (Savolainen, 1995; Savolainen & Kari, 2004), and what are some of the information effects of network structures (Haythornthwaite, 1996). Most of these established theories are highly descriptive and have stood the test of time, however they do not specifically describe information seeking behavior in the context of socially networked environments.

Other studies in Information Science have examined information in networked environments. The examination of networked documents and their influence on information seeking, for example, is done with link analysis, which has been used to describe the importance of a Web document based on its visibility on the Web, through metrics like the Web Impact Factor (WIF) (Ingwersen, 1998), but these studies are mostly divorced from human behavior. We can also find studies of how to best utilize user-created information, such as implicit feedback or tagged documents, in networked environments (Agichtein, Brill, & Dumais, 2006; Zhou, Lawless, & Wade, 2012), but they are focused more on information retrieval than on information seeking.

More recent studies have described how people search for certain information through a combination of online search engines and their online social ties (Morris, Teevan, & Panovich, 2010b; Rainie & Wellman, 2012), and described the popularity of online social networks as information sources regardless of the quality of the information seeking and retrieval tool (Hirsh & Dinkelacker, 2004), but there is a gap in that area of the literature describing not just how people seek information in their social networks, but also how their places in their social networks influence their information seeking.

I pose two general questions that the literature does not answer: How do certain structures of social networks influence information seeking behavior? And how do information seekers' states of uncertainty influence what strategies they employ in order to get answers to their questions in a social network?

1.2. Background

The mesh of our relationships to one another is a big part of what defines us as social beings. The sociologist Georg Simmel claims that society itself is nothing more than a web of relations (Marin & Wellman, 2010; Simmel, 2011). There is something that rings true about the claim that "the science of networks is the science of the real world" (Watts, 2003, p. 13) despite its clichéd nature.

Social network analysis (SNA) can be a useful tool to study networks as it provides a theoretical alternative to the notion of independent social actors. It gives researchers a framework for testing theories about structured social relationships that is a fundamentally different perspective than that adopted by individualist social scientists (Marin & Wellman, 2010; Wasserman & Faust, 1994). Social network theory espouses the idea that a social actor's position in a network partially determines the constraints and opportunities that he or she will encounter. This is why identifying and analyzing that position is useful for predicting actor outcomes, such as performance or behavior (Borgatti, Everett, & Johnson, 2013). Actors in networks are always discussed vis-à-vis the links or relationships that exist between them. These relationships are the fundamental component of network theory and distinguish network analysis from other research approaches. The theoretical concepts, the data under study, and the analysis performed is all about the relationships among the units in the studied network (Wasserman & Faust, 1994).

When individuals create social ties, they often see them as investments in the accumulation of social resources (Katz, Lazer, Arrow, & Contractor, 2004). Social resources are those embedded in social environments (like social networks) that can lead to successful instrumental action (Lin, 1982). In online settings, useful examinations of social networks can thus provide us with useful theories and methodologies that enable us to better answer questions of how people's social ties, as expressed in their online social networks and on information and communication technologies (ICTs) such as social media, influence their *behaviors* online, such as when they seek social resources. Information is important in providing a basis for action and it can be acquired by the use of social relations whether physical or virtual (i.e. online) (Coleman, 1988; Evans, Kairam, & Pirolli, 2010; Granovetter, 1973). Hence it is reasonable to call information a social resource, a "valuable currency" even, especially in the context of a networked environment (Gruzd, Wellman, & Takhteyev, 2011). Moreover, the *topology* of the social network structure can have an impact on these behaviors as well, given that individuals' places in a network is suggestive of their influence on others in their network (Aral & Walker, 2012) and can be an indication of how effectively they can connect people to each other (Milgram, 1967).

Scholars in LIS have shown that network characteristics of certain information structures can influence certain human behavior, like searching for information. These studies have explored the different structures of information repositories' environments such as scholarly journals (in the form of citation networks, for instance) and Web-based hypertext documents (Björneborn, 2006; Björneborn & Ingwersen, 2004; Park & Thelwall, 2003; Thelwall, Vaughan, & Björneborn, 2005). There is also interesting theoretical work about the information effects of network structures (Haythornthwaite, 1996, 2002) and "social search" for information in online social networks (DiMicco et al., 2008; Evans & Chi, 2008, 2010; Morris et al., 2010b).

In online contexts, we know that social networks play a role in individuals' information seeking activities. People turn to others for help when seeking information – not just to professionals like librarians, but also to people they know like colleagues or good friends. Given that a substantial proportion of interactions between individuals happens online via social networking sites (SNS) and social media, it should be no surprise to know that people utilize their social networks online to seek information (Evans & Chi, 2008; Morris, Teevan, & Panovich, 2010a).

Social network analysis (SNA) is one of the preferred set of methods that scholars use in order to better examine network dynamics. SNA has been developed, by no small measure, by many scholars in the field of Sociology. Sociological research has, for example, examined how the quality of ties in a social network (for example, weak versus strong) influences the dynamics of resource exchanges that take place (Granovetter, 1973, 1983; Krackhardt, 1992). Others have examined the phenomenon of "social capital", a form of value associated with the outcomes of social participation that produces tangible goods, both public and private (Lin, 1999; Lin & Dumin, 1986), while still others have examined how certain network structure characteristics, like "structural holes", play an important role in the flow of social capital and other resources (like information) across networks (Burt, 1992, 1997).

1.3. Significance of the research

This research sheds light on our understanding of the dynamic between human behavior and their environmental structures. People are influenced by the social structures they find themselves in. This dissertation aims to contribute to both information seeking models and theories around networked environments. This research combines concepts and theories of both information seeking and social networks, two fields that should have more overlapping theories in common than they currently do, given the growing importance of the role of online social networks in helping people find information, whether through friends and acquaintances (e.g. Facebook), or through online communities of similarly-minded people (e.g. Reddit), or through corporate "knowledge networks" that aid employees find information from subject-matter experts in their larger organization.

The Web-based tool, called the SIMulated social-computational Platform with a SOcial Network environment (or SIMPSON for short), can be further adapted to help academics and others research human information and communication behaviors in social networks. SIMPSON has been modeled with certain real-world online social networking sites that lend themselves well to people seeking information from others they are linked (or can link) to, such as certain community-oriented sites like Reddit, but on a smaller scale (smaller networks) and on a more limited basis (less functionality and choices of connection dynamics than most online social networking sites). SIMPSON has also been modeled with Web-based knowledge sharing tools, where users are made aware of "who knows what" and are therefore guided to certain individuals in their social network in order to get answers or gain knowledge.

CHAPTER 2: LITERATURE REVIEW

In the literature review section, I will introduce some important models and concepts from the LIS literature in information seeking and searching. Additionally, I will review literature that has dealt with information exchanges in network structures and, more generally, in social media. I will then introduce concepts regarding seeking information in an online social context and introduce concepts of knowledge sharing networks and attempt to tie them to concepts in social networking.

I will also give a background on social network theory that will discuss what social networks are. In addition, I will discuss important phenomena of and concepts on social networks like the “small world” concept, the scale-free network concept, tie strength, social capital, and structural holes. This is followed by a brief review of a key network characterization measure on node centrality (degree centrality) and one important whole network centralization concept, namely network density.

2.1. Information seeking

Most scholars active in information science research today would likely agree to a framework showing information seeking coming from realizing a need for information (Belkin et al., 1982; Dervin, 1992; Kuhlthau, 1993; Taylor, 1968) and that *both* the need for information and the outcomes of seeking information stem from a cognitive perspective involving communication, sensing, or thinking (Ingwersen, 1996). Many people do thus agree that information seeking is a subjective process that individuals approach with prior knowledge and differing levels of cognitive development (Weiler, 2005).

The definition of "information need", however, can prove to be just as elusive as the definition of "information" thanks in big part to the truism that needs, unlike wants, are often contestable (Case, 2016). In Belkin, Oddy and Brooks' (1982) important paper, we are introduced to the concept of the anomalous state of knowledge (or ASK) which comes into being when a person realizes that he or she has a gap (i.e. an anomaly) in their state of knowledge. This creates an uncertainty in the person who may then attempt to fill in this gap by seeking information or knowledge. Belkin et al. remind us that people seeking information do not always *know* what they are looking for. This admonition, in part, is a response to earlier work in information science that focused on the information system rather than the user – a trend that reversed, in part thanks to scholars like Taylor, Belkin, Kuhlthau, and Dervin to name but a few. For example, Kuhlthau focuses on the uncertainty of someone at the beginning of a search for information and Dervin's (1992) sense-making model starts with the premise that people have a need to make sense of their experiences. These concepts of anomaly or gap in knowledge, sometimes called the "problematic situation in information science", have injected new perspectives in the field (and that have held up for the last 35 years or so) because they point out that an information seeker's problem is usually not topical, but rather cognitive and needs to be understood within the larger situation of tasks and goals, which are best drawn out through interaction (Cool, 2001).

To further add to the vast dimensionality of what information seeking is, we should realize that information seeking is not just *one* thing (Courtright, 2007). It is an iterative process (Byström & Järvelin, 1995; Marchionini, 1997; Taylor, 1968), it does not need to include directed search (Bates, 1989), it can be about scanning (Choo, Detlor, & Turnbull, 2000), it may not include the discovery of a need by the user (Courtright, 2007), and it can be a pleasurable or leisurely activity (Fulton, 2009; Hartel, 2010; Matni & Shah, 2014).

2.2. Information seeking models

Models and frameworks precede theories and focus on specific problems and can describe processes or systems (Case, 2016). Early on, information and communication studies adopted positivist models of communication, like Shannon and Weaver's (1949), to try and explain how people communicated with one another, but while the mathematical model per se works very well with computer and communication systems (it's still one of the gold standards in communication systems engineering research and development), it has proven itself to be a limited model in that it is unable to take into account the contexts and nuances of human (i.e. non-systematic, constructivist) communication and information interaction. It was never intended as an information science model and can therefore not tell us anything of substance about information needs (Wilson, 1981). Since the 1960s and 1970s, the field of information science has gradually moved away from system-centric model to user-centric ones.

T.D. Wilson (1981) developed a model of information seeking behavior that is informed by several of the user's needs, including physical, cognitive and emotional ones.

He has subsequently revised the model over the next two decades. He identified 11 (in later models, 12) components emanating from the information user, who upon discovering a need, is led to a choice of several activities. The user would then place demands on information systems or other sources of information and, should the user be successful in his or her task, use the information. Along the way of information seeking and information use, Wilson's model takes into account that information is exchanged with other people. At some point after information use, the individual decides whether he or she is satisfied or not and if he or she needs to modify his or her need.

David Ellis' model (1989) focuses on the behavioral aspects of information systems, especially information retrieval (IR) systems, and how people interact with their information sources, with the reasoned hope that if one could understand researchers' information seeking patterns, then most IR users should be able to also understand their own information seeking patterns while interacting with the system. In today's modern information systems, this is a feature that is mostly taken for granted. Many of Ellis' researchers were seasoned in the techniques and art of professional information seeking. After doing several interviews and coding them using a grounded theory approach (and explaining rather agreeably how he went about with his qualitative study), Ellis presented six characteristics or stages of behavioral information searching: *starting*, *chaining*, *browsing*, *differentiating*, *monitoring*, *extracting*, *verifying*, and *ending*. Optimally, of course, his model works best in the same context in which it was generated, that is, with individuals with experience searching scholarly material search for information. In a testament to the interdisciplinary reach of Ellis' model, recent research based on it includes studies on how design engineers search for technical information (Robinson, 2010) and how technology-savvy users search for tourism information (Ho, Lin, & Chen, 2012). Ellis advocates the design of information systems that reinforce these six stages and aid the user in navigating through them. Present-day online bibliographic IR and search systems (like Google Scholar, for example) seem to implement rather well many of his goals and recommendations. Other heavily cited models of information behavior that were inspired by the Belkin et al. (1982) ASK concept and Ellis' model include Kuhlthau's (1991) model of students' library search process and Bates' (1989) berrypicking model.

Carol Kuhlthau's (1991) model of the "information search process" (ISP) is based on theories of learning and can be universally applicable to any domain (Case, 2016). It describes both cognitive and affective behaviors that people go through as they evaluate information in their search processes and has 6 chronological stages: *initiation*, *selection*, *exploration*, *formulation*, *collection*, and *presentation* (later publications show a seventh stage, *assessment*). In every stage, Kuhlthau examines the feelings, thoughts, and actions that individuals take. At the initiation stage, the person has feelings of uncertainty and his or her thoughts are vague. Once action is taken in the selection phase, however, feelings of uncertainty give way to optimism. This oscillation of *feelings* is evident as people go through confusion, clarity, confidence, satisfaction, and finally a sense of accomplishment, respectively. Likewise, people go through different *cognitive* processes from vague thoughts, to focused attention with increased interest, until thoughts of increased self-awareness become prevalent in the final stage. People's *actions* vary from seeking then exploring relevant information to documenting pertinent information. The steps described are very intuitive and make the most sense when describing a learning process, however, unlike most other popular models of information seeking, the need for information does not get a mention in the ISP model. Nor does the ISP model provide for any feedback loops for re-assessing and starting over again. As such, it could be used as a subset model, where a more macroscopic one might provide for assessing user need and feedback for iteration of the process. The strength of Ellis's and Kuhlthau's models is that they are based on empirical research and have been tested in subsequent studies (Wilson, 1999). Kuhlthau's ISP has informed a great amount of research, including studies on children's information seeking (Bilal, 2000) and generating a theory of tasked-based information retrieval (Vakkari, 2001), among others.

Gary Marchionini (1997) presents a similar step-by-step construct for his information seeking model in electronic environments as Kuhlthau's ISP, but he provides feedback paths from almost every stage in the process to other stages and he takes information needs into account. Marchionini's model begins with users *recognizing and accepting* the information problem after wrestling with whatever gap or anomaly they are confronted with. Then, users *define* the problem, *select* a source of information, *formulate* a query, *execute* it, *examine* the results, *extract* the information, and *reflect* upon it. Should the users be satisfied with the results of their search, they can stop at this point, otherwise the model provides for 5 feedback paths going to the information extraction stage, or the examination stage, or the query formulation stage, or the problem definition stage, or the recognition of the user need stage. Similarly, all 7 stages in Marchionini's model show some feedback paths to other stages in the model (Marchionini, 1997, 2010), allowing for a great deal of flexibility and re-assessment of action for the information seeker along every stage of the process.

Reijo Savolainen (1995) presented a model on information seeking that eschews a particular kind of seeker, but instead focuses on a “way of life” approach. “Way of life” borrows heavily from the concept of *habitus* first proposed by Pierre Bourdieu (1984). Habitus is a system of temperaments by which people integrate their experiences and evaluate the importance of different choices. It ties strongly into concepts of social structure as well, like race and gender. Savolainen’s user-centric model is for “everyday life information seeking” (ELIS) and was developed by analyzing interviews he did with ordinary people (specifically, teachers and industrial workers) doing “nonwork information seeking” on both electronic and printed media. He found that the more the quantity of electronic media a person used, the more affective his or her orientation was in behavior, whereas the lighter the quantity of media consumed, the more cognitive the behavior. While Savolainen (1995) describes the use of radio and television – two technologies whose users experience much less active interaction compared to the modern-day social media user (Matni & Shah, 2014), he eventually does apply the ELIS model to come up with conceptions of the Internet as a source for information seekers (Savolainen & Kari, 2004).

Although Bates (1989) introduced the concept of berrypicking as a *search* technique, per se, I mention it here since it explicates a dynamic process that includes information seeking. Bates claims that real information searching does not always work with one query, one response. Instead, she points out, real life queries evolve during the course of the search and the query is typically satisfied by a series of choices and bits of information at each stage. The latter phenomenon is what Bates called “berrypicking”. Additionally, Bates points out that searching by subject was just one way to perform a query to find a document. There was also footnote searching, citation searching, journal runs, area scanning, and author searching – and not just all in one domain or database. Bates’ ideas have also given rise to the notion of “orienteering” that Teevan and her colleagues have articulated (Teevan, Alvarado, Ackerman, & Karger, 2004). Orienteering information seekers go about their search bit-by-bit, not by issuing an initial query that gets them the answer, but rather by getting to approximately the right area of information through several queries (Hearst, 2011). When done iteratively, this technique helps users eventually get to a satisfactory conclusion. Research that has utilized Bates’ berrypicking notions and their off-shoots, include studies done on the cost structure analysis (i.e. the trade-offs in the value of information gained against the costs of performing the information search) of foraging for information (Pirolli & Card, 1995) and research done on personal information management (Jones, 2007).

2.3. Principles and theories for use in information seeking

Beyond the models discussed here, there are a number of paradigms that have been associated with information seeking research. Case (2016) points out that theories originating in education, psychology, and sociology have and continue to inform most research done in information science. Case classifies these theories into two generally distinct camps: objectivist (like Zipf's principle of least effort) and interpretive (like phenomenology).

Zipf (1949) found that many relationships in the human world showed patterns of preference of one resource used over another and he attributes this to the economy of effort by humans. This phenomenon is observable in many other areas, like citation and bibliographic analysis (Brookes, 1973) or computer router networks or the World Wide Web (WWW) (Huberman, Pirolli, Pitkow, & Lukose, 1998), or indeed, people's social networks (Barabási & Albert, 1999; Watts, 2003).

Probably one of the more complete philosophies tied to information science is phenomenology. Phenomenology's main tenet is that subjectivity and socialization are the common elements in how we contextualize reality, experience life, and build up and use knowledge (Burger & Luckmann, 1966; Schutz & Luckmann, 1973). It ties into general systems theory that looks at human behavior as something found in a larger interconnected system and that to understand it, one must understand the dynamics of the larger global society that those individuals live in (Boulding, 1956). Who we are and what we know are intricately enmeshed. Build-ups and changes in one's stock of knowledge, according to Schutz and Luckmann, are made through an integrative process that may not always be based on logical progressions, but rather on socialized "taken-for-granted" knowledge and on personal practicalities. Even "meaning" is not an objective thing; rather it comes about as a result of past experiences that are anchored in a valid reference schema, like one's culture or education. When we research people's use of information and communication technologies (ICTs), we must keep in mind what their motives are and realize that the way they use ICTs does not always conform to a prescribed logical sense, but rather to subjective practical realities.

2.4. Social and information exchange in network structures

Moving past the models and principles of information seeking behavior, I want to explore how existing research shows how people interact socially and how they exchange information especially in the context of network structures.

Networks, as will be explained further in this chapter, are made up of two main components: the actors and the relationships they have among one another. Social exchange theory is concentrated on the relationships in networks. One of this theory's fundamental ideas is that relationships evolve over time into trusting, loyal, and mutual commitments and are the basis for all exchange of social resources. These exchanges are facilitated by *rules of exchange*, that is, norms of behavior that are based mostly on reciprocity. Exchanges require a bidirectional transaction, that is, something given and something returned (Cropanzano & Mitchell, 2005). Beyond reciprocity, other factors of rules of exchange are rationality, altruism, group gain, and competition (Meeker, 1971). Relationships are the key to the *exchange of social resources* between people and this would include information, as well as a multitude of economic (e.g. money) and socio-emotional (e.g. love) resources (Foa & Foa, 1980). Social exchange theory also explains that while exchange relationships are sometimes altruistic in nature, they often demand repayment in a set time period and can be motivated by personal self-interest (Cropanzano & Mitchell, 2005).

Social exchange theory explains well *how* and *why* we interact with one another in order to get social resources, but Haythornthwaite (1996) expounds further on information as a social resource and shows how regular patterns of information exchange expose social networks, with actors as nodes in the network and information exchange relationships as connectors between nodes. Network structures can limit access to information, but they can also enhance that access. This is mostly due to the fact that networks emphasize who is connected to whom. This is why a “well structured” network can provide informational benefits to a user in terms of timely access to information. Moreover, informational opportunities in a network are influenced by who can make contact with whom and what information can be provided (Haythornthwaite, 1996).

2.5. Social media and online social networks

While social and information exchange concepts do not need to be rooted in online environments, my study seeks to simulate an online social network where people interact and exchange information, so it behooves me to spend some time on social media and online social networks.

Social media (assumed to be online) allow users to seek information with relatively little effort. The interfaces are easy to use for navigation of the information and the information itself comes in inherently rich structures. Because of this, social media sources have an inherent advantage over more traditional collections of documents when performing information retrieval tasks, even if the quality of the content varies (Agichtein, Castillo, Donato, Gionis, & Mishne, 2008). Social media’s ubiquitous presence in people’s online activities is also a large contributor to their motivations of everyday information seeking behaviors (Matni & Shah, 2014).

Social media has deep roots in the Internet, starting with the creation of Usenet in 1979 up through the various technologies of the “Web 2.0” in the late 1990s and early 2000s, including social networking sites (SNS) and social media sites that sprung up in the last decade (Ellison & Boyd, 2013; Kaplan & Haenlein, 2010; Treem & Leonardi, 2012). Treem and Leonardi (2012) think that social media distinguishes itself from other computer mediated communication technologies because they *afford* new types of behaviors that were previously difficult to realize in the workplace. Treem and Leonardi encourage researchers in search of a theory behind social media to consider basing it on the affordances (i.e. the perceptions of an object’s utility) that social media offers its users.

Although social media and its use in information seeking, searching, and retrieval is a relatively new phenomenon, I believe that we must allow for it to expand in its role and applications (it may very well metamorphose into something else in the coming decades). The combination of the ubiquity and ease-of-access of social media, its inherent richness of information, and its natural facilitation of social networking make it a powerful emerging way to seek and share information.

When seeking information people turn to others for help – not just to professionals like librarians (Taylor, 1968), but also to people they know like colleagues or good friends. This seems to be true ever more with the advent of online social networks through social media and other outlets. The concept is called “social search” and has been shown to give information seekers cognitive benefits (Evans & Chi, 2008). Asking questions to one’s social network also validated other search results that the information seeker might have obtained and gave them personalized answers to their particular questions (Morris et al., 2010b). However, some research suggests that social searchers’ views of what makes a “good answer” might not be the same for how experts might see it (Shah & Kitzie, 2012) unless the criteria for certain components of a “good answer” (like quality) is carefully defined (Shah & Pomerantz, 2010).

Social media offers the social search advantage to information seekers by giving them access to a social network. Elements of the social network can be important measures of utility and usefulness in these instances. For example, when seeking information on Facebook, users find the information they seek particularly useful if their bridging social capital and their engagement with their network meet minimum thresholds (Lampe, Vitak, Gray, & Ellison, 2012). Another example that Lampe and his colleagues illustrate is that older people with higher perceived levels of bonding social capital are less likely to use Facebook (Lampe, Vitak, & Ellison, 2013). To be sure, not everyone uses social media to seek information and non-users typically express their concerns about privacy, context collapse, limited time, and channel effects when conveying reasons why they do not use Facebook (Lampe et al., 2013).

Recent studies have shown that certain social network characteristics, like *tie strengths* (Gilbert & Karahalios, 2009; Panovich, Miller, & Karger, 2012) and both *bonding* and *bridging social capital* on various social media, like Facebook (Burke, Kraut,

& Marlow, 2011; Lampe et al., 2012) or Twitter (Ghosh et al., 2012; Naaman, Boase, & Lai, 2010), are enablers for people to obtain information from their social networks. Bonding social capital is present in close relationships, like with kinships and companionships, and enables reciprocity and emotional support (Wellman & Wortley, 1990). Bridging social capital can enable access to *novel* information because one's closest and strongest ties are likely to have redundant information (Granovetter, 1973).

2.5.1. Knowledge sharing networks

There are specialized online social networks where the people know what everyone else knows, at least in a general sense, and connect them to these resources of information and knowledge. These are called knowledge sharing networks. Knowledge sharing online environments have been around for a couple of decades, sometimes known as collaborative filters, or "communityware". These tools are used to make work communities' social networks visible to users and serve as repositories of the work organization's knowledge network, i.e. not only knowing "who knows what?" but also "who knows who knows what?" (Contractor, Zink, & Chan, 1998). Scholars of knowledge sharing networks have shown how organizations benefit from using knowledge residing in its different sub-units (Hansen, 2002) and how correlating the knowledge of two or more users can help identify sources of organizational misperceptions (Pathak, Mane, Srivastava, & Contractor, 2006).

The advent of social media has made the presence of knowledge networks better known and many such environments or platforms reside within an organization's intranet, that is, they aren't always accessible to the public since they could point out the intellectual property of some companies, or other such proprietary work. McLure-Wasko and Faraj (2005) describe how these types of networks are found in several professional online communities. There are several motivations of knowledge network users, such as their perception that their participation can enhance their professional reputations, or because they want to altruistically share their experiences with their communities of practice. Majchrzak and her colleagues (Majchrzak, Faraj, Kane, & Azad, 2013) theorized that there are four affordances of social media that represent different ways to engage in publicly visible knowledge conversations, one of which is network-informed associating, which is the engagement of online knowledge conversations informed by relational ties. This is something that social media uniquely brings to knowledge-sharing networks. However, other scholars have critiqued the "ideology of openness" that social media is seen to bring to knowledge sharing networks and have found that social media are also used strategically to limit information and knowledge in an organization (Gibbs, Rozaidi, & Eisenberg, 2013).

2.6. What are social networks?

One way we can describe social interactions and the connections people have with one another is through a social structure made up of actors and the ties they have to one another. Viewing social interactions through social networks gives scholars a set of methods to analyze both this structure and the patterns of interactions observed in these structures (Wasserman & Faust, 1994).

2.6.1. Actors and relations

In analyzing social systems, it is useful to think of two things: the social actors in the system and the relationships they have with one another. Social network analysis is therefore grounded on the notion of the importance of relationships among interacting social actors (Wasserman & Faust, 1994). This is in line with a notable shift in social science scholarship beginning in the second half of the 20th century, away from individualist rationalizations and more towards relational, contextual and systemic understandings (Borgatti & Foster, 2003). We are generally not interested in what the actor might do; only that the actor is part of a social structure. In fact, the term "actor" does not necessarily mean that the entity has the ability to "act". Moreover, a social network is often studied as a snapshot (or series of snapshots) of the structure of relations in one particular point in time (Wasserman & Faust, 1994; Garton, Haythornthwaite, & Wellman, 1997). As it happens, a common criticism of social network research is that not enough attention is paid to network dynamics (Watts, 2003). Actors (sometimes referred to as "nodes" or "vertices" in network topologies) are social entities that can represent an individual, a social group, an organization, or a population of groups. Their distinguishing characteristics (attributes) can be any number of qualifications such as age, being male or female, or being a widget manufacturer. When a network describes actors that have *n* attributes, it is called an "*n* mode network". The most common type of network studied, for example, is a one-mode network where actors have a single attribute of research interest (Wasserman & Faust, 1994; Watts, 2003).

Actors in a *social* network are often people or groups of people. The ties between nodes can be seen as channels though which things flow, for instance, material goods, such as money or diseases, or non-material ones, such as information or ideas (Borgatti et al., 2013). Relationships structure the flow of resources in a social environment (Haythornthwaite, 1996). Actors can be members of various social networks, each one based on different kinds of relationships.

Actors are linked or tied to one another via links or "relations" (sometimes referred to as "edges" in network topologies) whose characteristics can also be scalar or ordinal values (for example, person A is person B's boss, or person X has been married to person Y for at least 10 years, or company M is a client of company N). Relations are a specific kind of interaction between actors as determined by the researcher. There are usually three types of attributes given to relations or ties: content, direction, and strength (Haythornthwaite, 1996; Wasserman & Faust, 1994).

2.7. Social network phenomena

2.7.1. The "Small-World" problem

Milgram was among the first to formulate the small-world problem, asking what the probability was of any two people in the world knowing each other. Milgram conducted two experiments that consisted of sending letters to random people in Nebraska and Kansas and asking them to forward the letters to target recipients in Massachusetts either directly (if they knew them) or through an intermediary that *they believed* was more likely to know the target. Milgram was able to trace the routes that the correspondence took and saw that the median length of the routes was between 5 and 6 links (the shortest was 3; the longest was 10). Thus was born Milgram’s famous “six degrees of separation” concept, which said that any person in the world was only six connections away (on average) from anyone else. Milgram also showed that there was a convergence of these links to the target through common channels. These were nodes on the network occupied by people who seemingly knew a disproportionate amount of others (Milgram, 1967).

2.7.2. Subsequent models of complex networks

While there are several limitations to his study, Milgram’s small world experiment was enormously influential (the original article has been cited 5,772 times, per Google Scholar, as of this writing) and opened the doors for a large amount of research in the area of social networks that goes on today. Milgram’s study has been criticized, for example, because the number of data points was low (Barabási & Bonabeau, 2003), but also because it relies on people’s often wrong guesses as to who to extend the link to (Killworth, McCarty, Bernard, & House, 2006).

Erdös and Rényi developed one of the first models of a complex network, today known as the Erdös-Rényi, or ER, model (Erdös & Rényi, 1960) which represented a random graph. Most nodes in ER types of networks have the same number of connections (low heterogeneity) and show a degree distribution of a Gaussian bell-shaped curve. ER random graphs have short average paths between nodes and exhibit low clustering of linked nodes. Many years later, Watts and Strogatz (1998) proposed a model where the connections between the nodes in a regular graph were rewired with a certain probabilistic (Poisson) distribution. The resulting graphs were between the regular and random in their structure and are referred to as small-world (SW) networks. SW networks were meant to be closely structural to social networks, in that they have higher clustering than random networks (like the ER model), but almost the same average path lengths, given the same number of nodes and edges. Around the same time, Barabási and his colleagues (Barabási & Albert, 1999; Barabási & Bonabeau, 2003) proposed another model characterized by a highly heterogeneous degree distribution and which follows a "power-law". A small number of nodes have a majority of total connections and many other nodes with very few connections. They called these scale-free (SF) networks since zooming in on any part of the distribution does not change its shape.

Watts (2004) himself admits to limitations to his initial model when used to study social networks, chiefly because as Kleinberg and his colleagues have demonstrated, social networks are searchable (Kleinberg, 2000). This denies the application of a normal-like distribution that Watts-Strogatz presents on most social networks. Using Watts and Strogatz's own data, Barabási was able to show that many social networks in fact followed a power-law distribution (Barabási & Albert, 1999; Watts, 2003), which meant that a few of the nodes showed a very large amount of linkages to all other nodes, while the remaining nodes exhibited very few. This type of relationship has sometimes been called the "Matthew Effect" (after a passage in the New Testament) or the "rich get richer" effect, which explains how early nodes in the networks have a disproportionate advantage over later entries into the network. These early nodes with a lopsided number of links are often referred to as *network hubs*. The scale-free network model applies not only to social networks, but to economic networks, transportation networks, citation networks, Internet router networks, and the World Wide Web as well. Scale-free networks are resilient to random accidental failures, but are very vulnerable to directed attacks on the hubs (Barabási & Bonabeau, 2003).

2.8. Relationship characterizations in social networks

There are countless relationship characterizations in social networks. Of interest to this dissertation are the concepts of tie strength and structural holes, which I will discuss in the following sub-sections.

2.8.1. Tie strength

The gestalt of a social network is that of actors linked together by some relationship and that that relationship is paramount to understanding what these actors can do and how they can do it. The ties that bind any two actors can be seen to symbolize the exchange or sharing of resources such as goods, services, information or social support (Haythornthwaite, 2002). The strength of a tie has been described as the availability of the amount of time, the emotional intensity, the mutual confiding, and the reciprocal services between the two connected actors (Granovetter, 1973).

Usually we ascribe “strong ties” to relations between people of a core network: like members of a family, or close friends. Amongst each other, strongly tied people exhibit higher levels of intimacy, more emotional exchanges, more self-disclosure, more frequent interactions, and higher reciprocities (Granovetter, 1983; Haythornthwaite, 1996, 2002). There is a range to all of these characteristics and where they delineate between different tie strengths is open to interpretation by different scholars (Gilbert & Karahalios, 2009; Granovetter, 1983; Haythornthwaite, 2002; Krackhardt, 1992). For instance, the frequency of interactions between strongly tied social actors does not seem to be as important in kinship ties as it is with other “inner-circle” people that need certain maintenance of the tie strength, such as the case of friends and work colleagues. Ties and their strength attributes can therefore vary over time, exhibiting growth as people get to know each other better or decline as the reason for some strong associations reaches some conclusions (Haythornthwaite, 2002). Strongly tied social actors are usually self-motivated, often because of positively affective feelings for one another (“philos relationship”), to share resources with each other and usually make themselves available to one another (Krackhardt, 1992).

It is easy therefore to intuitively understand “weak ties” as those between mere acquaintances, for instance. Strongly tied pairs provide high velocity paths to information already circulating in their tightly knit network, which means that these actors have access to the same resources. If they wanted new information or fresh resources, they would necessarily have to go outside their strong tie network (Burt, 1992; Haythornthwaite, 1996). The “strength of weak ties”, then, is that they provide connections to others outside the strong tie network along with their new resources (Burt, 1992; Granovetter, 1973). These weak ties can help someone generate creative ideas, (Granovetter, 1973), find a job (Granovetter, 1974), or get information on the competition (Burt, 1992). Weak ties also require less time, less energy, and fewer overall costs to maintain, which releases time to do other things (Levin, Walter, & Murnighan, 2011).

When people search for informational resources, they will need to access networks beyond their strong-tie ones (Haythornthwaite, 1996). As my research interests lie in the realm of accessing information via ICTs, especially social media, I need to take into account the strength of people's ties in their social networks. Social media does not incorporate tie strength or its lessons, but instead all users are treated the same, whether friend or stranger, with little or nothing in between (Gilbert & Karahalios, 2009).

2.8.2. Structural holes

While Coleman (1990) questioned whether social capital can be useful as a quantitative concept in social science, Lin (1999) posited that the trouble lay in extending the notion of social capital beyond its theoretical roots in social networks and the difficulty in predicting social capital for every individual case. Ron Burt (1992, 2001) attempted to better understand this by proposing that social capital was an asset that owes its being to *location effects* in differentiated markets. As information flowed through a network structure, diffusion of information occurred over an interval of time. This meant that individuals informed early on had an advantage, even if the information eventually reached everyone.

Burt (1992) further pointed out that the structure of the relationships between individuals in social networks is sometimes constructed such that very few individuals tie two groups weakly together. These kinds of weaker connections are *holes* in the social structure of the network. These "structural holes" create a competitive advantage for someone whose relationships span the holes. In other words, these individuals have an opportunity to broker the flow of resources (like information) between people. Burt's (1992) theory of structural holes builds on Granovetter's (1973) theory of the role of weak ties as important resource bridges, but goes beyond it by also focusing on the role of strong ties in bridging social networks, especially how they facilitate access and trust between nodes and by providing a conduit between non-redundant network benefits.

Burt (1992) recognized the importance of network density in this regard, especially as it can define the potential of the value buried in structural holes. So, the stronger the ties within each of the disconnected networks “A” and “B”, the more a bridge between “A” and “B” will be useful, whether this bridge is a weak tie or not, and the more advantage and power the broker/entrepreneur who bridges that structural hole can obtain. This is how Burt sees social capital as a function of brokerage opportunities.

2.9. Characteristics of social network structures

There are myriads of social network structure characteristics that are employed in social network analysis. Of interest to this dissertation are the concepts of node degree centrality and network density, which I will discuss in the following sub-sections.

2.9.1. Node degree centrality

The flow of how resources like information move in a network typically comes down to the concept of *centrality* in network analysis. There are several types of centrality measures, such as *degree centrality*, *closeness*, *betweenness*, or *eigenvector centrality* (Borgatti, 2005). Borgatti cautions researchers that the types of flow processes must be first identified before the type of centrality measure is decided upon because of the different assumptions made by these different measures. Centrality is a property of a node’s position in a network. The centrality of a node is, loosely speaking, about the contribution the node makes to the structure of the network (Borgatti et al., 2013). It is a common way to find out the “most important” actors in a network (Wasserman & Faust, 1994). The simplest measure of centrality is degree centrality, which is merely the number of ties of a given type that a node has. It can be further delineated as *in-degree* and *out-degree* centrality measures, which classifies incoming versus outgoing links to and from a node, respectively. Degree centrality is very popularly used when using measures of network characteristics (be they social, organizational, or other kinds of networks) to inform the researchers about the actors. Examples include studies that located influential people within a group vis-à-vis their social structure (Gould, 1989), that used a person's social network to determine if he or she was a potential bully (Faris & Felmlee, 2011), that examined inter-organizational dynamics of "cooperative competitors" (Doerfel & Taylor, 2004), or that examined the hyperlinking practices of newspaper organizations in order to predict the likelihood of failure of the business (Weber & Monge, 2013). Degree centrality has been criticized as not adhering to a strict definition because it does not take into account any measures of the whole network beyond the adjacent nodes (Borgatti et al., 2013), but it is nevertheless easy to calculate and popularly used.

2.9.2. Network density

In addition to characterizing the nodes and edges of a network, one can characterize the *whole network* as well. Centralization is a measure that characterizes the entire network and can be expressed in terms of *network density*, which typically measures some centralization value in proportion to a total network term (Borgatti et al., 2013; Wasserman & Faust, 1994).

Network density is a measure of cohesion of the network and offers a general picture of the network (Borgatti et al., 2013; Doerfel & Taylor, 2004). It is a single number that is calculated simply by dividing the number of all ties in the network by the total number of possible ties (a.k.a. Metcalfe's number, $n(n-1)/2$, where n is the number of nodes in the network). Network density is almost always used as a comparative tool between multiple networks, but if the relative sizes of the compared networks are too far apart, some researchers prefer to use the *average degree of the network* instead, which is merely the mean of all the nodes' degree centrality.

2.10. Summary

This chapter has introduced various published research related to information seeking and social networks. I have introduced important models and concepts from the LIS literature in information seeking and searching. In order to tie structural concepts with information seeking behavior, I have reviewed the literature on information effects of network structures. I have also reviewed concepts of seeking information in an online social context (as in social media), and then tied them to concepts in social networking, explaining what it means to look for information via social networks and social media. This transitioned into a deep background on social networks that discussed what they are and the important ideas and theories on network structures that apply to social networks, like the "small world" concept, the scale-free network concept, tie strength, and structural holes. Finally, I examined degree centrality, a key network characterization measure, as well as ideas on how to measure whole networks through network density.

CHAPTER 3: CONCEPTUAL FRAMEWORK

In this chapter, I will present the theoretical frameworks that guide the proposed research. My generalized research questions are also presented. The framework that I propose is that which brings together theories on the role of uncertainty in information seeking behavior and those on network structure (especially structural holes) and its effects on effective access to information. In addition, in order to ascertain possible costs of successful and unsuccessful interactions between people (assuming rational players) in a networked setting, I've turned to aspects of game theory, especially cooperative/competitive games in small network settings.

3.1. The role of uncertainty in information seeking behavior

Uncertainty in information seeking is useful to understand information-seeking behavior (Kuhlthau, 1993; Wilson, Ford, Ellis, Foster, & Spink, 2002). As mentioned in Chapter 2, Kuhlthau's (1991, 1993) model of the "information search process" (ISP) describes both cognitive and affective behaviors that people go through as they evaluate information in their search processes and has 6 chronological stages: *initiation*, *selection*, *exploration*, *formulation*, *collection*, and *presentation* (later publications show a seventh stage, *assessment*). In every stage, Kuhlthau examines the feelings, thoughts, and actions that individuals take and shows that uncertainty at the start of the process is characterized by vague thoughts, anxious feelings and exploratory actions and is therefore uncertainty is quite pronounced. In the subsequent processes, understanding (characterized by clear thoughts and confident feelings) overtakes much of the uncertainty, and although some uncertainty persists, it is noticeably less than at the beginning of the information seeking process (if the information seeking task is assessed as a successful one). So, uncertainty in information seeking gives way to certainty a little bit at a time as the information-seeking task is successfully concluded. Uncertainty is usually reduced, but not eliminated (Wilson et al., 2002), but the ISP model strongly suggests that it is advantageous to information seekers for uncertainty to be reduced as they progress through their tasks and head towards their goals.

In the social networks literature, the role of uncertain behavior is acknowledged as well (Granovetter, 1983; Krackhardt, 1992). The theory of weak and strong ties claims that people develop ties to reduce uncertainty, especially strong ties. These ties reduce resistance to the actions taken between people. In this dissertation, I will not focus on strong versus weak ties, but rather I will use as proxies the presence of ties versus the absence of ties.

3.2. Network structure and its effects on access to information

Haythornthwaite (1996) discussed information effects of network structure thusly: network structures constrain and limit access to information, based on the network topology, that is based on who is connected to whom. A "well structured" network provides informational benefits in terms of access to information *in an efficient and effective manner*. Additionally, informational opportunities in a network are influenced by who can make contact with whom (I'll add to this: and how much effort or cost each contact might take/have) and what information can be provided.

Another social network theory that contributes a theoretical frame to explain information-seeking behavior is the aforementioned concept of structural holes. Burt (1992) explained that people with well-structured networks obtain higher rates of return for their efforts in getting resources. How individuals are connected in a social structure is indicative of the volume of resources they hold and the volume to which they each are connected. A sparse network, or indeed a social network where few or no relationships are expressed, provides more information benefits than a dense network. This is because it can connect people to information in separate areas of their usual social activity. A dense network is a “virtually worthless monitoring device” (Burt, 1992, p. 74) for finding new resources (such as information) because each person in it knows what the other people know, so they can all more or less simultaneously discover the same opportunities at the same time. When sparse networks have a node, or a set of nodes, that ensure the connection of two groups of non-redundant sets of connections, the possibilities of accessing diverse information by everyone in the network become a lot higher. Such a node is called a “structural hole” and the person who occupies that node has an outsized advantage of controlling and accessing diverse information in an efficient and effective manner.

Structural holes appear frequently in actual social networks in many situations where almost-distinct and separate networks have a few overlapping nodes that can act as go-betweens for resource exchanges between the distinct groups. An example of this might be two separate groups or organizations, such as two separate university classes or two distinct professional associations, which have a small number of people that belong to both groups.

Scale-free networks also appear extensively in several situations in economic and social networks as well as in information systems, most famously, the World Wide Web (Barabasi, 2009; Barabási & Bonabeau, 2003). The network structure of the WWW developed as a scale-free network partly because of the dynamics of the behavior of people around providing and seeking information online. For example, when people create new pages on the Web, they tend to include hyperlinks to hubs rather than to pages that hardly anyone knows. Similar links between behavior and network structure had been described in earlier works on citation analysis (Nicolaisen, 1981). While there might not be adequate scholastic research on if and how the presence of a scale-free network influences a person's information seeking behavior, one might plausibly make the connection between efficient and effective information seeking behavior and a network structure that has been shown to be present in existing and well-used information system architectures, such as the connections of routers on the Internet, or the hyperlinks between online documents on the WWW. Thus, I posit that networks that have structural holes as well as those that have scale-free topologies are useful for studying information seeking behavior.

As previously explained, researchers have, for some time now, qualified social ties as either "weak" or strong (Granovetter, 1973, 1983; Krackhardt, 1992). In explaining strong ties, Granovetter (1983) found that people in insecure positions are more likely to resort to the development of strong ties for protection and uncertainty reduction. People resist change and are uncomfortable with uncertainty. Strong ties therefore constitute a center of trust between people that can reduce resistance and provide comfort in the face of uncertainty (Krackhardt, 1992). In fact, in their study on wired communities, Hampton and Wellman (2003) find no evidence that online social networks (among other ICTs) make neighbors close socially, thus confirming older findings by Granovetter that closeness and strong ties are the most significant defining characteristics of helpful intimate relationships, whereas weak ties are very important for accessing information and resources. One conclusion from this is that connections amongst people who “know” each other are less uncertain, and maybe less “costly”, than those amongst people who are not close to begin with.

3.3. The costs and payoffs of connecting with others

This leaves trying to understand *how* to determine these possible costs of connections with others. In experimental studies where participants played cooperative/competitive games in small network settings, Hanaki and his colleagues have accounted for the burden or cost of the interactions by defining a quantifiable total cost:

$$\boldsymbol{\gamma(k) = c_{i,j}k^{\alpha}},$$

where γ is the total cost of interacting with k partners, c is a probability of making the connection between participant *i* and participant *j*, and α is a number larger than or equal to 1 that represents a power-law cost factor (Hanaki, Peterhansl, Dodds, & Watts, 2007). Note that, if α and c are constants, the total cost, $\gamma(k)$, is a power-law relationship that increases the total cost as the number of partners in a network increases.

Hanaki et al. (2007) also claim that participants *i* and *j* will commence a relationship if both their expected payoffs from interacting exceed their respective costs incurred by adding one more neighbor. Put another way, we can say:

$$\mathbf{Min.\ E = \gamma(k+1) - \gamma(k)},$$

or the minimum expected payoff for *i* and *j* to connect is proportional to the number of others that they are connected to, and is equal to the cost of adding one more connection. If we assume that a connection will certainly be made (c = 1) and that the power-law cost factor α is minimal, but bigger than 1 (α = 2), the above equation reduces to

$$\mathbf{Min.\ E = k^2 - (k\text{-}1)^2 = 2k - 1}.$$

Therefore, if participant *i*, who has a degree centrality (or total number of relationships) equal to L_i wants to connect to participant *j*, who has a degree centrality equal to L_j, then participant *i* will have to pay a cost *less than* $2L_i - 1$ (for $L_i \geq 1$) and, likewise, participant *j* will have to pay a cost *less than* $2L_j - 1$ (for $L_j \geq 1$). So the cost of making that dyadic connection is:

$$\mathbf{CDC_{\mathit{i,j}} \leq Min\ E_{\mathit{i,j}} = 2L_{i,j} - 1}, \text{ where } L_{i,j} = \text{Min. } (L_i, L_j).$$

Similarly, if participant *i* wants to connect to participant *j*, then participant *i* will receive a payoff of *at least* $2L_i - 1$ (for $L_i \geq 1$) and, likewise, participant *j* will receive a payoff of *at least* $2L_j - 1$ (for $L_j \geq 1$). The payoff of successfully making that dyadic connection is thus:

$$\mathbf{PDC_{\mathit{i,j}} \geq Min\ E_{\mathit{i,j}}}.$$

If we make the payoff minimally higher than the cost (i.e. $\mathbf{PDC_{\mathit{i,j}} = 2L_{\mathit{i,j}}}$), then the net benefit, which is the difference between the payoff and the cost,

$$\mathbf{BDC_{i,j} = PDC_{i,j} - CDC_{i,j} = 1}.$$

This means that, assuming a successful connection is made, both parties gain a token benefit, equal to 1. This means that lower degree centrality and higher degree centrality participants are on equal footing, in other words, none have an advantage over the other if a successful connection is made. This also means that in an exercise where connections with others in a social network are encouraged in order to meet a specified goal (like finding information), there is no bias for one network topology (e.g. a scale-free network) to emerge over another type. Additionally, this framework encourages cooperation between participants in order for connections between participants to be made.

In the case of an *unsuccessful* connection attempt initiated by participant *i* and targeting participant *j*, then *i* would still incur the minimal cost of connection without any payoff, meaning:

$$\mathbf{BDC_{i} \leq 2L_{i} - 1}.$$

Participant *j*, however, would not incur either a cost or a payoff, meaning

$$\mathbf{BDC_{j} = 0}.$$

In an exercise where connections with others in a social network are encouraged in order to meet a specified goal (like finding information), this framework means that one way that participants can be competitive with one another is to deny a connection request, and thus delaying the meeting of the specified goal.

3.4. General framework overview

This dissertation examines ways in which network structures influence information-seeking behavior, especially when the uncertainty of the information seeker is a factor. The high-level context is that of information seekers interacting with one another in an online social network. My proposed framework defines some aspects of information seeking behavior around networks, such as how people connect and get access to information, and how much information they gather in a specified amount of time (i.e. effectiveness). Effectiveness is a common measure of how successful an information seeker is in meeting a desired information seeking goal. In the information science literature, effectiveness is an important metric used in evaluating both information retrieval systems (Baeza-Yates & Ribeiro-Neto, 2011) as well as users' cognitive traits and decision making when using information systems (Saracevic & Kantor, 1988a, 1988b).

The interplay between information seeking participants in their networked environment can also be examined as a series of dyadic interactions that can be seen to have costs and payoffs. My framework explains how one might quantify those costs and payoffs and borrows from certain areas of game theory to do so. These three main areas of my framework are illustrated in Figure 1.

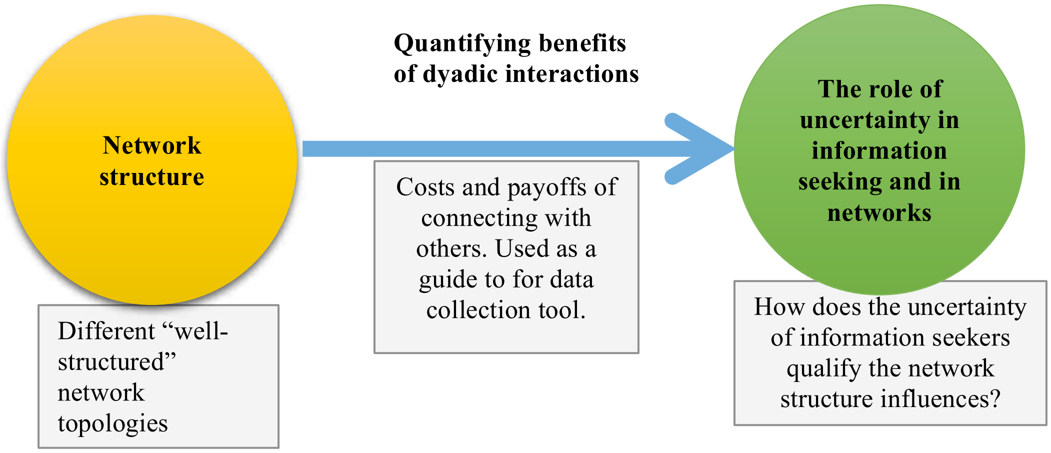

Figure 1: General framework of the dissertation borrows from 3 theoretical areas.

3.5. General research goals

I pose two general research questions that I pursue in this dissertation. Given that well-structured networks aid in effectively acquiring informational benefits and given that there exist multiple topologies of "well-structured" networks (let us for the purposes of this dissertation consider only two of those: networks with structural holes, and networks with very few highly-centralized hubs, that is, scale-free networks):

RQ1: How do these two different social network topologies influence the effectiveness of information seeking and gathering activities of individuals who experience different levels of information seeking uncertainty in their task?

RQ2: How do information seekers' states of uncertainty influence what strategies they employ in order to get answers to their questions in different well-structured social network topologies?

In the following section, I will discuss my methodology, describe my tool design, and expand on my general research questions.

CHAPTER 4: METHODOLOGY

4.1. Research design

To conduct this research, I created a data collection tool that had participants interacting with each other in a series of exercises that moderated their interactions via different social network structures and different settings of uncertainty levels. The data collection tool thus created a simulated environment in which the participants sought information from others in a simplified networked environment. The simulation was designed to be simple in comparison to a real life situation because the parameters of human behavior are many (some might say endless) and real social networks can be very complex, varied, and dynamic structures. Therefore, while the tool allowed for a limited set of actions to be taken by the participants (such as making connections with others and asking for answers), it necessarily eliminated other possibilities of action to be taken by them (such as answering questions on their own without interacting with others, or breaking already established relationships in their social networks).

Furthermore, the participants in this study earned "points" in order to motivate their participation. These points were determined by a tool-maintained metric which ascertained if some participants used the tool "better" than others. This metric was reliant on the calculation of "benefit points" based on the participants' interaction choices. The generation of these points is based on research done by Hanaki et al. (2007) and others who have proposed ways to calculate possible costs of successful and unsuccessful interactions between people. This system of quantifying benefits makes some assumptions on the rationality of the actors. When people seek information, they may not necessarily act rationally or methodically, however, this scheme is used only to calculate these "benefit points" and was not revealed to the participants in order not to influence their behavior in the study (that is, only a running total on the "points" were shared with the participant, but not the methodology behind it).

4.1.1. Exercise scenarios

The tool was designed for 4 separate types of scenarios that will stand in for permutations of two types of well-structured networks: a topology with 2 structural holes (topology type: SH) and a scale-free network topology (topology type: SF), and two types of participant uncertainty: a situation where the information seeking participants are given a lot of knowledge about who-knows-what in the network, resulting in them having low uncertainty in their information seeking (uncertainty type: LU) vs. a situation where the information seeking participants are given very little knowledge about who-knows-what in the network, resulting in them having high uncertainty in their information seeking (uncertainty type: HU). Thus the 4 scenarios or exercises will be referred to in short-hand as: SH-LU, SF-LU, SH-HU, and SF-HU.

The experiment was thus designed as a 2x2 factorial design, where the 2 independent variables are type of network (SF vs. SH) and level of uncertainty (HU vs. LU). See Table 1. The relevant hypotheses are clarified later on in this chapter.

Table 1: 2x2 Factorial Design of the Experiment

2x2 Factorial Design	Scale-Free Network	Structural Hole Network
High Uncertainty	*Scenario A*	*Scenario C*
Low Uncertainty	*Scenario B*	*Scenario D*

The network type SF was modeled as a scale-free model following Barabási's work (Barabási & Albert, 1999). The network type SH was one where there are two sub-network groups of non-redundant connections that are connected to one another via 2 structural holes, following models studied by Burt (1992). Table 2 summarizes the scenario differences.

Table 2: Description of the 4 scenarios

<table>
<tr><th>Scenario</th><th>Network Topology</th><th>Description</th></tr>
<tr><td>SH-LU</td><td>Network with structural holes with low uncertainty in the information seeking task</td><td rowspan="4">Participants completed their task by forming new connections (undirected links) and asking other participants they are connected to if they have a particular answer for a particular question. They were also allowed to connect to other participants that were once removed from them via an intermediary participant.
The networks were dynamic and changed accordingly as the tasks progressed.
Participants had a visual representation of their network and were aware of who knows who/what when in the LU scenarios.
Participants were not told how their scores are calculated other than in the introduction when that was explained as a high-level generality.</td></tr>
<tr><td>SH-HU</td><td>Network with structural holes with high uncertainty in the information seeking task</td></tr>
<tr><td>SF-LU</td><td>Scale-free network with low uncertainty in the information seeking task</td></tr>
<tr><td>SF-HU</td><td>Scale-free network with high uncertainty in the information seeking task</td></tr>
</table>

I thus designed the SH and SF network topologies to be different in structure, but in order to reduce the number of differences in network parameters between them, I ensured that they had identical network densities. This meant that I had to run them with the same number of nodes and links (since network density is a function of both of these). My ideal was to have 10 participants in each run, who would then have 12 links (regardless of SH or SF network type). In case I had fewer than the signed-up 10 participants per run, (which did, in fact, happen sometimes), I also designed contingencies where I had as few as 7 nodes and 6 links. Table 3 illustrates all four designs:

Table 3: Network designs for SH and SF types as used in the experiment

Number of nodes (N)	Number of Links (L)	Network Density (D)
7	6	0.286
8	8	0.286
9	10	0.278
10	12	0.267

The SH and SF networks (for N = 10) are illustrated in Figure 2 and Figure 3.

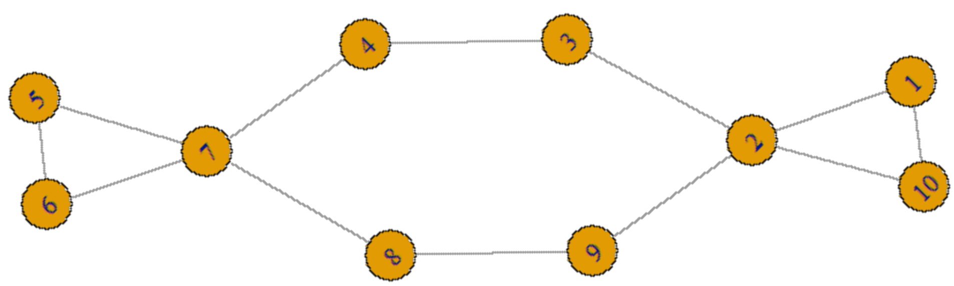

Figure 2: Network structure SH, showing two structural holes (nodes 3 and 8)

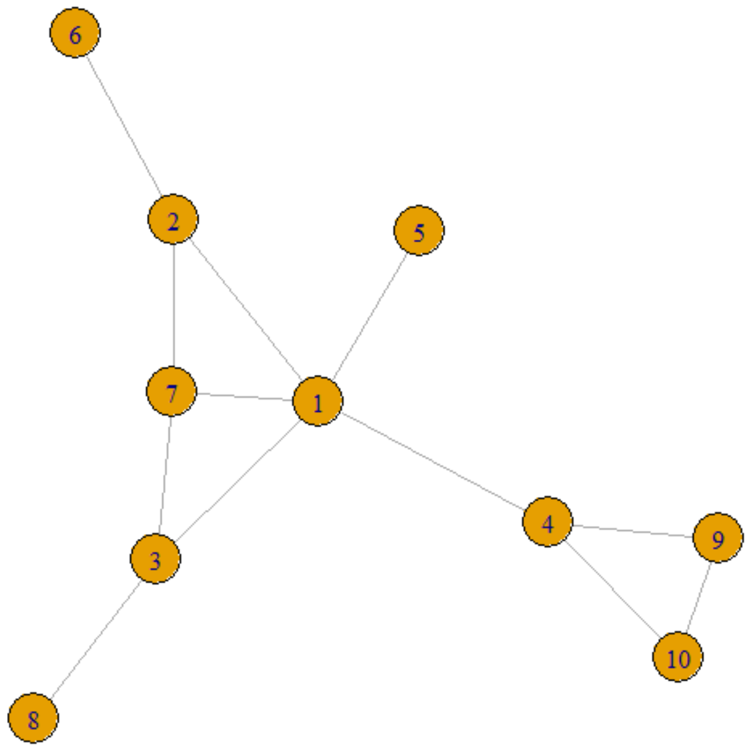

Figure 3: Network type SF, showing a scale-free topology (hub is node 1)

The participants were tasked with finding as many answers to specific questions as they could in a limited amount of time (10 minutes for each of the 2 sessions they did in one sitting). All participants got the same list of 30 questions to answer (15 questions for each session) and each question had 3 answers associated with it. This meant that the participants had to collect up to 90 answers (minus the answers they already had) in the limited amount of time. I decided upon these particular quantities because I wanted a large enough number of answers that would prove to be challenging, and maybe impossible, for the participants to collect in 20 minutes. This has enabled me to make statistically significant numerical comparisons of the participants' number of collected answers without much worry that these counts are saturated at the maximum 90. The questions and their answers were designed not to be difficult and are listed in Appendix 2.

4.1.2. Participants and their tasks

At the start of each exercise, a non-repeated set of 15 questions were chosen and all 45 answers were randomly assigned to all participants such that each of them had at least one (or more) of the answers. The participants then had to collect answers to the questions by interacting with one another in their pre-assigned social network. There were exactly 3 *unique* answers to every question and the participants are instructed to *only* find the answers from other participants in the network. This simplified the design of the tool and eliminated the situation of people answering the questions by themselves, since this simulation was intended to replicate the dynamics of seeking information from others in a social network.

Simulating some characteristics of a knowledge network environment, all participants had a visual representation of their network and, through this, were aware of who knows who and who has knowledge of what *categories of answers* (not who has knowledge of which *answers*, otherwise the exercise might not be challenging enough and participants might not engage with the exercise for the full duration). The extent to which the participants knew who had certain knowledge was controlled and set to all of the participants for the low uncertainty scenarios (LU) and to none of the participants for the high uncertainty scenarios (HU).

Similar to the Suri & Watts (2011) experiment in network behavior, the design of this experiment had participants engaging with *k* neighbors and thus the relationship between their neighbors changed as a function of the network. When the network became fully connected the group size N would be $k + 1$. So N had to be at least equal to (if not preferably larger than) $k + 1$. In order to effectively identify the different network topologies that the design required, and also meet practical experimental parameters, I picked N to be nominally 10, although I had some instances where I was obliged to work with smaller N in some of the experiment's runs because of absent participants. This choice of N is similar to the network sizes studied in other observational or experimental network dynamics studies (Cassar, 2007; Doerfel & Taylor, 2004; Suri & Watts, 2011; Wang, Suri, & Watts, 2012).

As Table 1 illustrates, there were four scenarios listed: two scenarios that used a SF network (called A and B, which respectively use high-uncertainty vs. low-uncertainty set-ups) and two scenarios that used a SH network (called C and D, which, again, respectively use high-uncertainty vs. low-uncertainty set-ups). Every run in my experiment used the same network (SF or SH) in the high vs. low uncertainty scenario (so some runs were scenarios A and B, while others were scenarios C and D). In order to negate any possible learning effects (Wang et al., 2012), I ran some of the experiments as A → B (start with scenario A, then run scenario B with the same participants), while I ran others as B → A. Equivalently, some of the runs were C → D, while others were D → C. I originally planned to have at least 4 runs in total, each with 10 participants, so that my experiment would have at least 40 participants in order to have enough data points to analyze and get statistically significant results. The final total number of participants in my experiments was 46 because I managed to execute an extra couple of runs. These are shown in Table 4.

Table 4: Experimental runs

Run Number	Number of Participants	Run Type
1	9	B → A
2	7	D → C
3	10	C → D
4	10	A → B
5	10	D → C
TOTAL:	***46***	

I also gave each participant pre- and post-surveys, that is both before and after the completion of the exercise, in order to collect information on their demographics, experiences with information seeking (especially in social networks), and their experience with both the tool and the networks they interacted with in regards to how well (or not) they helped the participants complete their tasks. Both surveys are shown in Appendix 3: Surveys.

4.1.3. Creating a data collection tool

The data collection tool that I designed and used for these experiments is called a Web-based SIMulated social-computational Platform with a SOcial Network environment (or SIMPSON for short). This was created mostly in the PHP computer language using the Laravel technology platform. SIMPSON can capture participants' activities in regards to their information search behavior in such an environment. The participants were recruited and invited to interact with one another through SIMPSON with instructions to find information in the form of "answers" to given "questions". All participants were asked to take part of the experiment for about 45 to 60 minutes, including listening to an introduction to the experiment, watching a demonstration video of SIMPSON in use, filling out the pre- and post-surveys for about 20 minutes, using SIMPSON for another 20 minutes (10 minutes for each scenario run), and any time used in-between (getting seated, filling out paperwork, etc.). The participants, while using SIMPSON, gathered as many “answers” as they could while interacting with the other participants in their group.

4.1.4. Experiment protocol

This is the experiment protocol that I executed:

1. I asked the participants to fill out a consent form.
2. I directed the participants to sit down at a computer workstation. This was conducted in a computer lab at UCSB that held 10 computer terminals at workstation desks. The computers were all running Windows 10 and the Firefox Web browsing program. Firefox was already configured for SIMPSON use (SIMPSON resided on a Web server and Firefox was already directed at the login page at http://simpson.kevinalbs.com/auth/login).
3. I then directed the participants to take the pre-survey (see Appendix 3). The survey was conducted with paper and pen provided to the participants.
4. I then gave the participants a short introduction to the experiment and had them click on an icon on their computer in order to view a pre-recorded video demonstration on their computers of SIMPSON in use that includes several examples of what they might encounter. The video is just under 3 minutes long.

5. Each participant was then asked to log into SIMPSON with a unique participant ID number that they were each given.

6. A Web page instructed the participants as follows:

   "GOAL OF THE EXERCISES:

   You are now going to do 2 exercises, one after the other. For each exercise, you will be given a list of 15 questions, each with 3 unique answers to them. You have to find as many answers as you can in 10 minutes. The answers are distributed amongst all the participants (including yourself).

   OTHER PEOPLE DOING THE EXERCISE WITH YOU:

   In each exercise, you will see all the other participants in a visual representation – they will be the circles on the graph. They may be connected to someone else (including yourself) or they may not be connected to anyone at all. If they are connected to someone else, you will see a line drawn between their two circles. The visual representation can sometimes show you what kind of answers some other people in the exercise may have. This can vary from one exercise to the next.

   HOW TO MAKE CONNECTIONS WITH OTHERS:

   You will have to find the answers you are looking for by making connections with the other participants and asking them if they have the answer to a specific question. To connect with someone, or to ask them if they have an answer that you are looking for, you must click on their circle and follow the instructions to make a connection.

You can connect with someone new directly, or indirectly through someone you are both connected to (like a go-between or an intermediary). It is 'cheaper' for you to connect through an intermediary, if you can.

HOW TO GET ANSWERS FROM OTHERS:

Once you connect to someone, you can ask them if they have an answer to a specific question. Again, you should click on their circle and follow the instructions to ask for an answer that you need to collect.

FINISHING THE EXERCISES:

You have 10 minutes to complete each of the 2 exercises. The exercises will each run for exactly 10 minutes. If you complete your exercise before the 10 minutes are over, you must wait for the remainder of those 10 minutes to begin the next exercise. Your instructions and your goals are the same for each exercise.

SCORES:

As you go through the exercises, you will collect points, shown as SCORE on the top right hand side of the screen. You can see other people's SCORES too, when you click on their circle in the visual representation.

When you are finished with the exercises, you will be asked to fill out one last survey.

Thank you very much for your participation!"

7. Once everyone had logged into SIMPSON, the first run was enabled by myself via remote instruction. Right away, all participants begun with this first run of the experiment together. The timer on SIMPSON began counting down 10 minutes. If

certain participants completed their exercise goals before the 10 minutes were over, they had to wait for the next exercise to begin.

8. Once the timer ran out for the first run, the second run was enabled by myself via remote instruction. Again, right away, all participants begun with this second run of the experiment together. Once again, the timer on SIMPSON began counting down 10 minutes. If certain participants completed their exercise goals before the 10 minutes were over, they had to wait for the next step.
9. When the timer ran out for the second run, the participants had then completed the 2 exercises/runs.
10. SIMPSON then terminated the session and the participants were thanked.
11. I then directed the participants to take the post-survey (see Appendix 3). The survey was conducted with paper and pen provided to the participants.
12. When the participants handed in their surveys, I thanked them and posted their participation on the automated UCSB student-research recruitment system (called SONA – see Section 4.4 for more details on recruitment and SONA).

4.2. Metrics

SIMPSON tracked a metric called SCORE which it shared with the participants. SCORE is made up of 3 other internal metrics which reflect the participants' activities. The purpose of SCORE is primarily to encourage participation through motivation of "playing the game". The 3 sub-metrics are a "link number", a "network capital" metric and an "information capital". The "link number" (L) is simply a measure of the number of links that the participants have with other participants in their network (i.e. their degree centrality). The “network capital” (NC) metric is the accumulation of cost and payoff points that a participant gather while moving about the network, building or denying ties, and gathering the required “answers”. NC is based on a set of specific actions that participants are limited in taking, where each action is worth a certain number of points, some constant in value, others dependent on certain network characteristics. All participants start off with the same initial amount of NC points ($NC_0 = 1000$) but they are **not aware** of what points each action nets them. The “information capital” (IC) metric reflects the accumulation of the answers that each participant must undertake. SIMPSON showed the aggregate score of the sum of all 3 metrics (SCORE = L + NC + IC) and also POSITION, the ordered position of the participant. Both SCORE and POSITION were displayed prominently on the screen (on the top right corner) and were updated in real-time.

The participants in this study needed to connect with others in the exercise to get information in the form of answers and they could visually see their network. In scenarios of low-uncertainty (LU), as in real knowledge networks, all the participants were aware of who knows what *topical category* of at least one of the answers they had. In the cases of high-uncertainty (HU), no participants were aware of who knows what. The level of uncertainty, once established, is constant throughout each scenario.

As explained by social exchange theory, social relationships need some rules of exchange and some exchange of resources (Cropanzano & Mitchell, 2005). In order to encourage participation, I created a scoring system that depended on the participants’ interactions. The points that make up SCORE were gained or lost as the exercise unfolded, but the underlying mechanisms were not dynamically made visible to the participants. This is because these mechanisms had to reflect a "rational scenario" and thus revealing them to the participants could influence their behavior.

SCORE informed the metric POSITION, which was the overall ordered position of the participant in that run (i.e. 1st, 2nd, 3rd, etc…). These points were used as a passive motivation technique for the participants.

4.2.1. Permitted participant interactions

All participants were directed to undertake their choice of specific interactions, taken one at a time, in any order the participants wished. These interactions were based on 9 discrete and exclusive actions, **a** = $[a_1, \ldots, a_9]$, that the participants could take. These actions, in certain combinations, thereby defined a set of three distinct interactions, **i** = $[i_1, \ldots, i_3]$, as depicted in Table 5, and could be undertaken in all four scenarios.

Table 5: Description of permissible interactions

i_n	Description of interaction	Associated actions
i_1	Form a connection with someone new.	a_1, a_2, a_3
i_2	Form a connection/interaction with someone new through an intermediary.	a_4, a_5, a_6, a_7 a_8
i_3	Request an answer from one person and receive it (if they have it).	a_9

Each action had a net number of cost or payoff points associated with them for both the person who took action (*action taker, or AT*) and the person to whom the action was directed to (*action recipient, or AR*). These points were sometimes static and other times dynamically linked to certain network characteristics, such as the number of links the action recipient has, or the number of people in the network of the action taker. Participants were aware of the "point value" of these actions while they interacted in this simulation. The set of 9 permissible actions and their associated taker/recipient net points are listed out in Table 6.

4.2.2. Setting points for costs and payoffs of direct connection requests

Each action taken by the participants when they interacted with one another was associated with a *benefit point* (P_B) to both the action taker (AT) and the action recipient (AR). These benefit points were calculated internally (i.e. they are not revealed to the participants) and were meant to convey how rational actors might behave. The benefit points are positive contributors to the SCORE component, NC (where $NC = NC_0 + \Sigma P_B$).

Per the concept set forth by Hanaki et al. (2007) and described in my theoretical framework, if participant *i*, who has a degree centrality (or total number of relationships) equal to L_i wanted to connect to participant *j*, who had a degree centrality equal to L_j, then the cost of making that dyadic connection was:

$$\mathbf{CDC}_{i,j} \leq \mathbf{Min\ E}_{i,j} = \mathbf{2L}_{\mathbf{i,j}} - \mathbf{1}.$$

For the purposes of this design, the cost to connect with another participant was therefore $\mathbf{2L}_{\mathbf{i,j}} - \mathbf{2}$ if $L_{i,j}$ is larger than 1 and be 1 if $L_{i,j}$ is 0 or 1.

The payoff of successfully making that dyadic connection had to be:

$$\mathbf{PDC}_{i,j} \geq \mathbf{Min\ E}_{i,j}.$$

To make the payoff minimally higher than the cost (i.e. $\mathbf{PDC}_{i,j} = \mathbf{2L}_{i,j}$), then the net benefit, which is the difference between the payoff and the cost is thus:

$$\mathbf{BDC}_{\mathbf{i,j}} = \mathbf{PDC}_{\mathbf{i,j}} - \mathbf{CDC}_{\mathbf{i,j}} = \mathbf{1}.$$

This meant that, assuming a *successful* connection was made, both parties gained a token benefit, equal to 1.

In the case of an *unsuccessful* connection attempt initiated by participant *i* and targeting participant *j*, then *i* would still incur the minimal cost of connection without any payoff, meaning:

$$\mathbf{BDC_i \leq 2L_i - 1}.$$

For the purposes of this design, the benefit to participant *i* was therefore $\mathbf{BDC_{i,j} = -2}$ for any value of L_i (-2 being less than or equal to $2L_{i,j} - 1$ for any positive number $L_{i,j}$). Participant *j*, however, would not incur either a cost or a payoff, meaning $\mathbf{BDC_j = 0}$.

In summary, to make a connection between AT (action taker) and AR (action recipient), the benefits were set as follows:

$$Benefit\ to\ AT = \begin{cases} \mathbf{1}\ if\ successful \\ \mathbf{-2}\ if\ unsuccessful \end{cases}$$

$$Benefit\ to\ AR = \begin{cases} \mathbf{1}\ if\ successful \\ \mathbf{0}\ if\ unsuccessful \end{cases}$$

4.2.3. Setting points for costs and payoffs of indirect connection requests

If AT requests a connection to AR, through an intermediary AI, then according to the concept of transitivity, the connection costs for AT should be less than they are in a direct connection request, but the costs of failure are similarly enlarged. Ergo, the benefits should be more positive, if successful, but more negative if unsuccessful. Additionally, if the connection is successful, then both AR and AI should stand to benefit from this action, albeit not differently than the case of a direct connection request.

Therefore, to make a connection between AT and AR via AI, the benefits were set as follows:

$$Benefit\ to\ AT = \begin{cases} \mathbf{2}\ if\ successful \\ \mathbf{-3}\ if\ unsuccessful \end{cases}$$

$$Benefit\ to\ AR = \begin{cases} \mathbf{1}\ if\ successful \\ \mathbf{0}\ if\ unsuccessful \end{cases}$$

$$Benefit\ to\ AI = \begin{cases} \mathbf{1}\ if\ successful \\ \mathbf{0}\ if\ unsuccessful \end{cases}$$

4.2.4. Setting points for costs and payoffs of asking for an answer from someone

When an action taken necessitates a cost, then the action-taker loses points. Likewise, when an action taken necessitates a payoff, then the action-taker gains points. The design requires a cost to acquiring a resource, so if AT asked AR for an answer to a specific question, then a sum-zero trade should happen since AT and AR both got something in the trade that should at least be of equal value. In this case, I chose an arbitrary number for the cost and payoff points to be 2 points. It thus follows that AT's net benefit would be -2 and AR's net benefit +2, if the result was successful (that is, AT acquired the answer). If the result was unsuccessful, then the net benefits for both should be zero.

Table 6: Description of permissible actions and their net benefits

i_n	a_n	Description of action	AT net benefit	AR net benefit
i_1	a_1	Ask to connect directly with someone new	-	-
	a_2	Connection is accepted	1	1
	a_3	Connection is refused	-2	0
i_2	a_4	Ask to connect/interact with someone new via an intermediary (one-step removed).	-	-
	a_5	Connection is accepted by intermediary	-	1 to AI
	a_6	Connection is refused by intermediary	-3	0 to AI
	a_7	Connection is accepted by end node	2	1 to AR
	a_8	Connection is rejected by end node	-3	0 to AR
i_3	a_9	Ask for an "answer" from one person and automatically receive it (if they have it)	-2 if successful 0 if not	+2 if successful 0 if not

L*: Number of people directly connected to the action taker.*

4.3. Design of the front page

The front page of the tool, SIMPSON, looks slightly different with each of the four scenarios (SH-LU, SF-LU, SH-HU, and SF-HU) to accommodate the two types of initial network structures (SH vs. SF) and the two levels of information seeker uncertainty (HU vs. LU).

4.3.1. Scenario SH-LU

The front page of the tool that participants in Scenario SH-LU saw exhibited the following information (see Figure 4 for a screenshot of SIMPSON):

a) A scoreboard in the top right corner of the screen where the participants could keep track in real-time of:
    a. Their instrumental metrics (SCORE, POSITION).
    b. What questions they fully answered.
    c. The time remaining (from an initial 10 minutes) in the exercise.

b) A real-time updated visualization of the topology of the participant's social network and to whom he or she was connected to. Participants were anonymized (names showed up as "user1", "user2", etc…) In this instantiation (SH-LU), the network was one with 2 structural holes. This visualization served to make the participants aware of their social network. When participants looked at this, they saw:
    a. The network in real-time, and
    b. A label representing the topical category of answers that the others knew (unique to LU scenarios).

   When the participants clicked on a node in this visual network, they saw:
    a. A radio-button menu of options of interactions to take:
        i. Connect with this person directly.
        ii. Connect with this person indirectly (thru someone else).

iii. Ask this person if they have an answer (if connection is established).

c) A list of other participants' actions that have directly impacted the participant, specifically (as they apply):

   a. The acceptance or rejection of a request to connect from the participant.

   b. That they gave the participant an answer.

d) A list of additional actions that the participant could take in response to other participants' actions that directly impacted the participant, specifically:

   a. Accepting or rejecting another participant's requests to connect.

4.3.2. Scenario SH-HU

The design was the same as the one for scenario **SH-LU** described in 4.3.1 with the following exception: item (b)(b) was absent. That is, the participants could not see any information on what any other participant knew. All other features and visualizations were the same.

4.3.3. Scenario SF-LU

The design was the same as the one for scenario **SH-LU** described in Section 4.3.1 with the following exception: the real-time updated visualization of the topology of the participant's social network first appeared as a *scale-free network*.

4.3.4. Scenario SF-WU

The design was the same as the one for scenario **SF-LU** described in Section 4.3.3 with the following exception: item (b)(b) was absent. That is, the participants could not see any information on what any other participant knew. All other features and visualizations were the same.

4.4. Target population and recruitment

The participants were recruited from the body of undergraduate students at the University of California, Santa Barbara (UCSB). This is considered a convenience sample given that I am located there for my current job, as of this writing. I recruited UCSB students via a local university online system called SONA that matched potential student participants with a number of research projects by University researchers. The students could get credits on SONA that can be applied towards extra credit in a number of undergraduate classes. Multiple instructors at UCSB utilize the SONA system as a standard way for students to fulfill “research participation” requirements in their classes. The student participant compensation scheme was clarified in the consent form and on the SONA sign-up system.

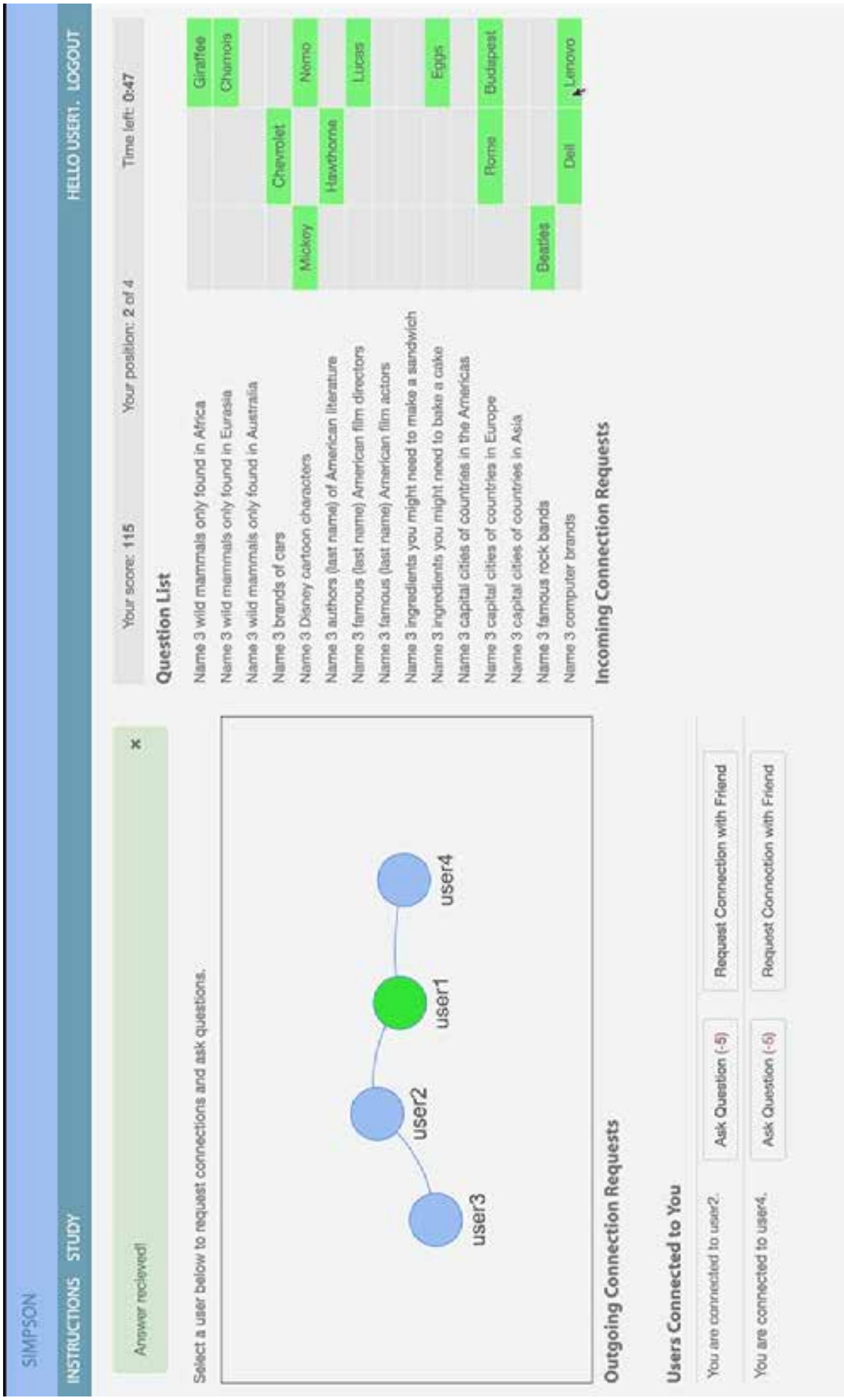

Figure 4: Front page of SIMPSON (screenshot)

4.5. Log data from SIMPSON

The networked environment simulator and data collection tool, SIMPSON, recorded all the participants' interactions with one another as they executed the search for the information tasks given to them. This data was collected in several tables in a database on the Web server. Note that all users had anonymized and unique user ID numbers. The data included the following:

a) Anonymized and unique user ID number.

b) Unique project (i.e. experimental run) ID number.

c) A time-stamped indication of acquiring an answer (all answers had unique ID numbers).

d) A time-stamped indication of initiator, recipient, and intermediary users when a connection was requested, created or rejected.

e) A time-stamped indication of initiator and recipient users when an answer was requested, answered or not answered.

f) A time-stamped indication of change to the SCORE metric per user.

4.6. Data analysis

An essential feature of log data is that it captures actual user behavior and is not subject to participant memory recall (as it might be on a survey) or subjective impressions of interactions. Log data does have disadvantages however, including the absence of annotations that record participants' motivations, successes, or satisfactions. Logs are very good at telling us *what* people are doing, but they don't tell us a lot about *why* they are doing something and whether or not they are satisfied (Dumais, Jeffries, Russell, Tang, & Teevan, 2014).

The log file from SIMPSON detailed everyone's activity on the tool over time. This gave me a time-series of evolving data on all participants, illustrating the presence or absence of ties that the participants have with one another and how these ties change over time in all permutations of network type (SF vs SH) and uncertainty (LU vs HU).

Additionally, the SIMPSON log file yielded data on the participants' choices of *how to connect with others* (did they interact with others to whom they were already connected? Did they connect directly to new nodes? Did they connect indirectly to new nodes?). The log file also allowed me to measure the frequency of these choices, as well as how often an individual connected or accessed someone else in their network, and how many total answers they collected in their runs and how long it took to collect these answers. The log file also provided me with the captured SCORE metric that the participants were made aware of in their exercise runs.

4.7. Hypotheses

RQ1: *How do these two different social network topologies (one topology with structural holes and one scale-free topology) influence the effectiveness of information seeking and gathering activities of individuals who experience different levels of information seeking uncertainty in their task?*

In a well-structured network (that is, one that follows a common social network topology like scale-free and/or contains at least one structural hole), we should see more benefits to information seekers, overall, and to those acting as structural holes, especially. Therefore:

HYPOTHESIS 1 (H1): If I were to observe two types of networks: one with structural holes (**SH**) and one with a scale-free topology (**SF**), I should see people being effective with their information seeking and gathering activities, but I should see **higher effectiveness** in a SH network. If these networks are further distinguished with characteristics that help reduce uncertainty (specifically, if the users have some knowledge of who-knows-what), then we should see *more* effective behavior in these low-uncertainty (**LU**) conditions than in those networks that give no information about who-knows-what (high-uncertainty, or **HU**). The combination of those two parameters (SH vs SF, HU vs LU) should be cumulative.

In other words, **H1** stated as a null-hypothesis, that is a hypothesis assumed to be true until evidence indicates otherwise, is:

$$\mathbf{H1_0}: \mu E(SH, LU) < \mu E(SF, LU) < \mu E(SH, HU) < \mu E(SF, HU).$$

$\mu E(N, U)$ is median effectiveness, which is measured per the two IV factors under discussion. Table 7 summarizes **H1**:

Table 7: Expected outcomes for H1

EXPECTED OUTCOMES FOR THESE FOUR PERMUTATIONS (SCENARIOS)	Network Type A **SH type**	Network Type B **SF type**
Higher uncertainty (HU) No information on who knows what	**μE(SH, HU)** *Low effectiveness in gathering information.*	**μE(SF, HU)** ***Lowest*** *effectiveness in gathering information.*
Lower uncertainty (LU) Some information on who knows what	**μE(SH, LU)** ***Highest*** *effectiveness in gathering information, especially from the person in the structural hole position.*	**μE(SF, LU)** *High effectiveness in gathering information, especially from the person in the structural hole position.*

RQ2: *How do information seekers' states of uncertainty influence what strategies they employ in order to get answers to their questions in these two different well-structured social network topologies?*

To get to the answers they need in a social network, individuals will employ and/or create links with others that they think will help them achieve that goal. I will make two hypotheses here.

HYPOTHESIS 2 (H2): In both **SH** and **SF** networks, the level of uncertainty that the information seekers have will **significantly determine** if they employ more steps (if they have high uncertainty) or fewer steps, likely more strategic ones (if they have low uncertainty) to get their information. Specifically, information seekers with higher uncertainty (**HU**) will take more steps to achieve their information goals than those with lower uncertainty (**LU**). **H2** stated as a null-hypothesis is: **$H2_0$**: μS(LU) < μS(HU).

μS(U) is median steps-taken, which is measured per the Uncertainty IV.

HYPOTHESIS 3 (H3): In the same vein as H2, in **SH** networks, individuals with high uncertainty (**HU**) will connect with those occupying structural holes *more often* than individuals with low uncertainty (**LU**), to get their information. **H3** stated as a null-hypothesis is: $\mathbf{H3_0}$: $\mu A(LU) < \mu A(HU)$.

$\mu A(U)$ is median number of times a node is accessed, which is measured per the Uncertainty IV. Table 8 summarizes both H2 and H3:

Table 8: Expected outcomes for H2 and H3

EXPECTED OUTCOMES FOR THESE FOUR PERMUTATIONS (SCENARIOS)	Network Type A **SF type**	Network Type B **SH type**
Higher uncertainty (HU) No information on who knows what	*Individuals will take* ***a greater number of steps ($\mu S(HU)$)*** *to achieve information goal (H2).* *In* ***SH****, structural holes will be* ***accessed ($\mu A(HU)$) more often*** *(H3).*	
Lower uncertainty (LU) Some information on who knows what	*Individuals will take* ***fewer steps ($\mu S(LU)$)*** *to achieve information goal (H2).* *In* ***SH****, structural holes will be* ***accessed ($\mu A(LU)$) less often*** *(H3).*	

4.8. Variables required

For H1, I had to measure the *effectiveness* of the information seekers in their quest to find answers to their questions. I define this as the number of gathered answers in the allotted time of a single exercise which is, at most, 10 minutes, although of course some participants finished their exercises before then. This variable was collected for every participant and at multiple times in an exercise in order to understand how it changed over time. This variable, EFFECTIVENESS, is an array of 3 dimensions and defined as $\text{EFFECTIVENESS}_{P,E,t}$, where *P* is the individual participant ID number (P = {P1, P2, P3,

… P480}), *E* is the exercise number (E = {S1, S2, S3, S4}), and *t* is the discrete time value when the measurement was taken.

For H2 and for H3, I had to measure the *total number of steps taken to achieve goal* and the *total number of times a node (participant) is accessed*. These variables were collected for every participant. The variables, respectively, STEPS and ACCESSED are each arrays of 2 dimensions and defined as $STEPS_{P,E}$ and $ACCESSED_{P,E}$ where *P* is the individual participant ID number and *E* is the exercise number.

4.9. Amount of data gathered

I aimed to use 10 participants, nominally, for each run. However, a few of the recruited participants did not turn up for the experiment. Out of 50 recruits, 46 showed up. There were 5 runs in total and those are shown in Table 4. A "run" went through 2 exercises/scenarios (either A→B and B→A, or C→D and D→C), in a specific sequence, each running for 10 minutes while the participants gathered answers to questions that were given to them. The total number of participants I had was 46.

The database recorded 57,643 entries across 9 tables (2 of them were administrative data). The total number of individual data points was 525,638.

CHAPTER 5: FINDINGS

This chapter provides the details of the findings of the study. The overview of the data analysis was conducted in the two phases: (1) the pre- and post-surveys and (2) the log data from the Web-based SIMulated social-computational Platform with a SOcial Network environment (or SIMPSON for short) are presented in Sections 5.1 and 5.2, respectively. These include detailed backgrounds of the participants in the survey and the characteristics of the selected participants in the log data collection.

In Section 5.3, the findings and analyses that address the first hypothesis (H1) regarding the effectiveness of achieving information goals in the 2x2 factorial-design experiment are presented. Sections 5.4 and 5.5 similarly address the second (H2) and third (H3) hypotheses, respectively, regarding the number of steps taken to achieve information goals and the role of network structure therein.

5.1. Overview of the survey data

A total of 54 participants took the pre-survey and post-surveys in this study. The questions therein were designed to collect information on the participants' demographics, their experiences with information seeking (especially in social networks), and their experiences with both the SIMPSON tool and the networks they interacted with in regards to how well (or not) they helped the participants complete their tasks. Both surveys are shown in Appendix 3.

5.1.1. General characteristics of the participants

The participants were recruited from an available pool of undergraduate students at the University of California, Santa Barbara (UCSB), mostly from the Department of Communication. This was a convenience sample and hence cannot be considered to be fully representative of the general population.

I recruited these participants using an online participant management software system called SONA that the Department of Communication at UCSB employs to recruit students (and others) to participate in their various and ongoing research projects. The available pool of undergraduate students who had access to this recruitment system was approximately 800 to 1000 students. Students were compensated with "extra-credit" grades in their various Communication classes for participating in this study.

My study had 46 participants in total, consisting of 28 females (60.9%) and 18 males (39.1%), ranging in age from 18 to 24 years old (M = 20, SD = 1.22), and all of whom had anytime access to a desktop, laptop or tablet computer, or Internet-capable smart-phone. Only 3 of the students (5.6%) did not live the majority of their lives in the USA or Canada.

5.1.2. Pre-Survey: General social media habits

The findings shown here are all outcomes of closed-ended questions, except where noted.

When asked how regularly (i.e. with what frequency) did they access any type of social media, the participants, when given a choice of 6 answers (pre-survey Question #1), overwhelmingly said "More than once a day" (96.2%).

When asked which social media sites did they generally use and how often they used them, the participants indicated that the top 3 sites they frequented on a daily basis were Instagram (79%), Snapchat (70%), and Facebook (58%).

Table 9: Count of participants who answered the question: "Which social media sites do you use and how often do you use them?" (Q# 2, N shown per categorical answer). Green cells indicate where an overwhelming majority (over 50%) of participants chose that particular answer).

	N	More than once a day	Once a day	Once to a few times a week	Less than once a week	I do not use this SM site
Facebook	53	58.5%	11.3%	13.2%	5.7%	11.3%
Twitter	53	37.7%	11.3%	5.7%	5.7%	39.6%
Reddit	53	5.7%	1.9%	5.7%	7.5%	79.2%
Tumblr	53	1.9%	1.9%	3.8%	15.1%	77.4%
YouTube	53	34.0%	17.0%	34.0%	15.1%	0.0%
Instagram	53	79.2%	7.5%	3.8%	0.0%	9.4%
Pinterest	53	3.8%	0.0%	13.2%	24.5%	58.5%
Snapchat	52	71.2%	11.5%	3.8%	1.9%	11.5%
Yahoo! Answers	52	0.0%	0.0%	3.8%	28.8%	67.3%
Quora	43	0.0%	0.0%	2.3%	7.0%	90.7%
Other	9	44.4%	11.1%	11.1%	0.0%	33.3%

5.1.3. Pre-Survey: Asking questions on social media

The participants were then asked to "think about the last time you checked into a social media site for the purpose of asking someone a question (any type of question)." When asked if they had ever asked someone a question on social media, 35.2% said they had not. For those that had replied "yes" to that question, the survey then asked 4 follow-up open-ended questions. The first of these follow-up questions asked them to describe what instigated that question-asking session on social media. Content analysis on the answer yielded 7 categories of which the most common was "for general social communication information needs" (43.2% of the answers). This included answers such as, "I saw a picture online of them on vacation and I commented below the picture asking when they would return…" and "finding out more information about how my friend was doing". Some questions sprang from specific work-related or school-related needs for information (27.3% of the answers), as evidenced by such answers as "(I had some) confusion on homework" or "I was shopping and I needed advice on store locations but did not have an individual's phone number". See the frequency distribution of all 7 categories for this question in Table 10.

Table 10: Results of content analysis of the open-ended question: "What instigated that session? (i.e. asking a question on social media) (Q# 3b, N = 44).

	Count	Freq.
general social communication information need	19	43.2%
work/school related information need	12	27.3%
entertainment/shopping/travel information need	4	9.1%
asking for contact information	4	9.1%
asking for general information	2	4.5%
asking for urgent information	1	2.3%
other	2	4.5%

The second follow-up question asked (open-endedly) what websites or services did they visit when they asked a question online? There was no majority answer, but the largest group said Facebook (37.5%), followed by Instagram (16.7%), and Snapchat (12.5%). Interestingly, only 6.3% cited Yahoo! Answers and only 4.2% cited Reddit, which are two commonly accessed question and answer online services.

Table 11: Results of content analysis of the open-ended question: "What website(s)/service(s) did you visit?" (Q# 3c, N = 48).

	Count	Freq.
Facebook	18	37.5%
Instagram	8	16.7%
Snapchat	6	12.5%
Facebook Messenger	5	10.4%
Yahoo Answers	3	6.3%
Yahoo	2	4.2%
Twitter	2	4.2%
Reddit	2	4.2%
Chegg	1	2.1%
YouTube	1	2.1%

The third follow-up question asked (open-endedly) how much time did that visit approximately last. Almost a third of the respondents said it took them 2 minutes or less (29.7%) and the majority indicated that it took them 5 minutes or less (64.9%).

Table 12: Results of content analysis of the open-ended question: "How much time did that visit last approximately?" (Q# 3d, N = 37).

	Count	Freq.
< 1 minute	1	2.7%
1 minute	4	10.8%
2 minutes	6	16.2%
3 minutes	2	5.4%
5 minutes	11	29.7%
10 minutes	9	24.3%
30 minutes	1	2.7%
60 minutes	2	5.4%
> 60 minutes	1	2.7%

The fourth and last follow-up question asked (open-endedly) what was the nature of the question they asked. Content analysis on the answer yielded 3 general categories: types of questions that asked for factual information about things and situations – "what" types of questions (56.8%), types that asked for factual and specific information about a person or event (25.0%) – “who” or “when” types of questions, and types that asked about social support vis-à-vis their online network (whether offering it or asking for it) (13.6%). See the frequency distribution of all 7 categories for this question in Table 2Table 13.

Table 13: Results of content analysis of the open-ended question: “What was the nature of the question you asked?” (Q# 3e, N = 44)

		Count	Freq.	
“What”	Getting information (school/work)	11	25.0%	
	Getting information (commercial/money)	7	15.9%	
	Getting information (general)	4	9.1%	
	Getting information (social situations)	3	6.8%	***56.8%***
“Who” or “When”	About specific person(s) (where he/she is/are)	7	15.9%	
	About specific event (when something is)	4	9.1%	***25.0%***
Social Support	About giving/asking for social support	5	11.4%	
	Relaying a "like"/"dislike"	1	2.3%	***13.7%***
Other	Other	2	4.5%	

The next question asked the participants, “Still thinking about the last time you checked into a social media site for the purpose of asking someone a question (any type of question), who did you ask your question to?”. The question was posed as close-ended and the answers were allowed to be non-exclusive (i.e. more than one answer was accepted). Out of 47 responses, the largest group said “a friend” (46.8%), an almost equal number of answers were given for “a stranger” (23.4%) and “an acquaintance” (21.3%). Only 4 participants (8.5%) said “a family member”.

Table 14: Results of the closed-ended question: “Who did you ask your question to?” (Q# 4, N = 47).

	Count	Freq.
A family member	4	8.5%
A friend	22	46.8%
An acquaintance	10	21.3%
A stranger	11	23.4%

The next question asked the participants, “Which one of these statements best describes the person you asked a question to?”. The question was posed as close-ended and the answers were allowed to be non-exclusive (i.e. more than one answer was accepted). Out of 59 responses, the largest group said that the person they reached out to was simply available to answer their question (33.9%). Almost equally answered were that they considered the person to be a subject-matter expert (28.8%) and/or a trustworthy person (23.7%).

Table 15: Results of the closed-ended question: “Which one of these statements best describes the person you asked a question to?” (Q# 5, N = 59).

	Count	Freq.
He/she is an expert on the subject of my question	17	28.8%
He/she is not an expert on the subject of my question	6	10.2%
He/she is a trustworthy person	14	23.7%
He/she was available to answer my question	20	33.9%
None of the above statements.	2	3.4%

Finally, the pre-survey asked participants “In general, which of these online venues do you use to look for information and how often?” and gave them 14 choices of the most popular social media sites. Unsurprisingly, the search engine site Google was, by far, the most popular choice (98.1% of respondents said they used it “often” and 0.0% said they “never” used it), but popular social media sites Instagram, YouTube, Snapchat and Facebook also got a substantial amount of respondents who said they used them “often” to look for information (in the 43.4 to 35.8% range). Wikipedia, YouTube, and Yahoo! Answers were most often cited as used “sometimes” (in the 49.1 to 50.9% range).

Table 16: Results of the closed-ended question: "In general, which of these online venues do you use to look for information and how often?" (Q# 6, N shown per categorical answer). Green cells indicate where an overwhelming majority (over 50%) of participants chose that particular answer).

	N	Often	Sometimes	Never
Google	54	98.1%	1.9%	0.0%
Yahoo	53	7.5%	47.2%	45.3%
Bing	52	3.8%	13.5%	82.7%
Wikipedia	53	39.6%	50.9%	9.4%
Facebook	53	35.8%	35.8%	28.3%
Twitter	53	34.0%	20.8%	45.3%
Instagram	53	43.4%	28.3%	28.3%
Reddit	53	7.5%	18.9%	73.6%
Tumblr	53	1.9%	11.3%	86.8%
Pinterest	53	5.7%	17.0%	77.4%
YouTube	54	37.0%	51.9%	11.1%
Snapchat	53	35.8%	26.4%	37.7%
Yahoo! Answers	53	9.4%	49.1%	41.5%
Quora	35	0.0%	11.4%	88.6%

5.1.4. Post-Survey: On the participants' use of the SIMPSON tool

The 14 post-survey questions asked participants in the study questions on their user-experiences with the SIMPSON tool and self-reported behaviors regarding their interactions with the other participants in the study.

5.1.4.1. User experience and tool design feedback

The first 5 questions had yes/no answers and were asked to evaluate the participants' understanding on how to use the tool, given its multiple "moving parts". The feedback here is useful in assessing if the given explanations of the tool and its design were helpful or not to the participants. Results are shown in Table 17.

Table 17: Results of the first 5 closed-ended questions of the post-survey (Q# 1 – 5, N = 54).

	Yes	No
1. Were the instructions on **how to connect** to SIMPSON clear to you before you started?	77.8%	22.2%
2. Were the instructions on **how the compensation worked** clear to you before you started?	61.1%	38.9%
3. Were the instructions on the **main landing page** clear in regards to how you should connect with others, how you can reject connection offers, and ask for questions?	81.5%	18.5%
4. Once you started the exercise, there was a s**core and a position number** displayed on the top right. Were they clear to you what they were about?	75.9%	24.1%
5. Were the **score and the position number helpful** to you as you went through the exercises?	68.5%	31.5%

One of the questions posed to the participants was an open-ended one asking: "Please describe your general (or any specific) observations, comments, thoughts, etc. regarding your experience with these exercises". A content analysis of the responses given was done (N = 47) and the results are shown in Table 18.

Table 18: Results of content analysis of the open-ended question: "Please describe your general (or any specific) observations, comments, thoughts, etc. regarding your experience with these exercises" (Q# 14, N = 47).

	Count	Percent.
Negative Sentiments (38.3%)		
frustrated at least once	3	6.4%
confused at least once	8	17.0%
network diagram hard to read	5	10.6%
hard to manage multiple connections	1	2.1%
devolved to random action	1	2.1%
Positive Sentiments (55.3%)		
figured out by doing	7	14.9%
fun / interesting experience	13	27.7%
got insights into social networks	6	12.8%
Other	3	6.4%

5.1.4.2. User self-reported behavior vis-à-vis connecting with others

When asked if they actively refused connections with someone (which they could do by ignoring requests for connections from other users), 96.3% of respondents said “no” (Question #6 in post-survey). When asked in an open-ended follow-up question to “explain why or why not”, the most common answer was that they accepted connections from other participants *for their own personal gains in the exercise* such as gaining more points or more answers (60.9% of respondents). The full results for this are shown in Table 19.

Table 19: Results of content analysis of the open-ended question: “Explain why or why not, regarding your answer to the above question #6?” (Q# 7, N = 46).

<table>
<tr><th></th><th></th><th>Count</th><th>Percent.</th></tr>
<tr><td rowspan="5">Why they accepted</td><td>accepted to help others (social support, norms)</td><td>7</td><td>15.2%</td></tr>
<tr><td>accepted for personal gain: hoped for reciprocation</td><td rowspan="3">28</td><td rowspan="3">60.9%</td></tr>
<tr><td>accepted for personal gain: to gain more points</td></tr>
<tr><td>accepted for personal gain: to gain more connections/answers</td></tr>
<tr><td>accepted for other reasons</td><td>2</td><td>4.3%</td></tr>
<tr><td rowspan="3">Why they refused</td><td>refused because was unaware of the request</td><td>5</td><td>10.9%</td></tr>
<tr><td>refused for personal gain</td><td>1</td><td>2.2%</td></tr>
<tr><td>refused for other reasons</td><td>1</td><td>2.2%</td></tr>
<tr><td>Other</td><td>other reason given</td><td>2</td><td>4.3%</td></tr>
</table>

5.1.4.3. User self-reported behavior vis-à-vis uncertainty

When asked, in a closed-ended question, what difference it made for them when conducting the exercises if they knew something about what topics the other participants knew, most respondents said that it made it easier to reach their information goals and conclude the exercise (63.0% of respondents). The full results are shown in Table 20.

Table 20: Results of closed-ended question: "What difference did it make for you when conducting the exercises if you knew something about what topics the others knew?" (Q# 11, N = 54).

	Count	Freq.
It made it much easier to reach my goals and conclude the exercise.	17	31.5%
It made it somewhat easier to reach my goals and conclude the exercise.	17	31.5%
It made no difference.	18	33.3%
It made it somewhat more difficult to reach my goals and conclude the exercise.	1	1.9%
It made it much more difficult to reach my goals and conclude the exercise	1	1.9%

5.1.4.4. User self-reported behavior vis-à-vis information seeking strategies

When asked in an open-ended question to explain what strategies, if any, they had developed to help maximize their scores in the exercise, a majority of respondents (75.9%) said that they did *have specific strategies*, while only 20.4% of them said they had *no specific strategy*. Content analysis of the participants' answers revealed 7 distinct categories of specific strategies used, as shown in Table 21.

Table 21: Results of content analysis of the open-ended question: "By the last exercise, do you think you had developed a strategy to help maximize your score?" (Q# 13, N = 54).

	Count	Percent.
Specific Strategies Used (75.9%)		
actively seeking acquisition of connections (i.e. ask everyone)	12	22.2%
emphasis on orderly acquisition of connections (i.e. ask in order)	9	16.7%
passive acquisition of connections (i.e. wait for others to acquire answers first)	6	11.1%
actively seek subject-matter experts (i.e. look for who-knows-what)	12	22.2%
actively seek well-connected users	2	3.7%
No Specific Strategy Used	11	20.4%
Other	2	3.7%

5.2. General notes on analysis of the SIMPSON log data

The networked environment simulator and data collection tool, SIMPSON, recorded all the participants' interactions with one another as they executed the search for the information tasks given to them. The data was stored in a relational database and retrieved using a Structured Query Language (SQL). The data comprised of 57,643 entries which represented 525,638 individual data points. The data was cleaned of non-relevant entries: these were test entries, administrative data, and erroneous runs of the SIMPSON tool. This was done for the purpose of addressing the 3 hypotheses of this study.

Once I collected the log data from SIMPSON, the first step was for me clean it up. The point of data cleaning is validation, that is, to understand any errors or noise in the data, and to transform it in a way that preserves the meaning and the information it wants to yield (Dumais et al., 2014). I normalized and synchronized some of the relevant data that I knew beforehand I would need before examining further, which included plotting several of the variables over a time axis and looking for explanations on how the data varied over time. A subsequent step was to generate descriptive statistics on all the variables to give me a clearer high-level picture of what interesting information the data might convey.

All of my analysis on SIMPSON log data was done using the statistical computational tool, R (R Core Team, 2016). The R-language scripts I used for my analyses are found in Appendix 4.

5.3. H1: Effectiveness of meeting information need goals

The first hypothesis **(H1)** states that if I were to observe two types of networks: one with structural holes (**SH**) and one with a scale-free topology (**SF**), I should see people being effective with their information seeking and gathering activities, but I should see **higher effectiveness** in a SH network. If these networks are further distinguished with characteristics that help reduce uncertainty (specifically, if the users have some knowledge of who-knows-what), then we should see *more* effective behavior in these low-uncertainty (**LU**) conditions than in those networks that give no information about who-knows-what (high-uncertainty, or **HU**). The combination of those two parameters (SH vs SF, HU vs LU) should be cumulative.

As Table 2 illustrates, this hypothesis assumes a 2x2 factorial design, where the 2 IVs are type of network (SF vs. SH) and level of uncertainty (HU vs. LU).

5.3.1. H1: Descriptive statistics

The data for H1 gathered from SIMPSON includes 86 participant runs (that is, 43 participants who each did 2 runs). Each entry shows a **total number of answers gathered** (maximum of 45) and a **time needed to acquire said answers** (maximum of 10 minutes). The data also includes 2 categorical types of data fields: network type (SH vs. SF) and uncertainty condition of the run (H vs. L). The **effectiveness** calculation was done by dividing the total number of answers gathered by the time needed to acquire said answers. Table 22 summarizes the mean, median, and standard deviation of these data points and Figure 5 shows the distribution of the Effectiveness variable in the data.

Table 22: Descriptive statistics for H1 (N = 87).

	Total Answers Gathered	Time (mins)	Effectiveness for ALL	Effectiveness for HU	Effectiveness for LU	Effectiveness for SH	Effectiveness for SF
MEAN	30	7.66	4.11	4.87	3.28	4.32	3.81
MEDIAN	30	8.00	3.67	4.86	3.16	4.12	2.90
STDEV	11.80	1.80	1.92	2.23	1.02	1.77	2.09

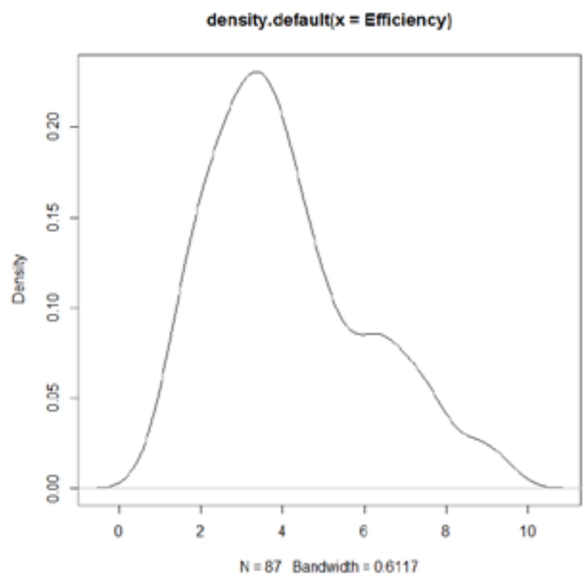

Figure 5: Density (distribution) plot for Effectiveness variable for H1 (N = 87).

5.3.2. H1: Plotting the different effectiveness measures under the 2 IVs

Examining the different effectiveness metrics under the 2 IVs yields the plots shown in Figure 6 and Figure 7. These show higher effectiveness under high-uncertainty conditions (median of 4.86 vs. 3.16 for low-uncertainty) and under SH types of networks (median of 4.12 vs. 2.90 for SF types).

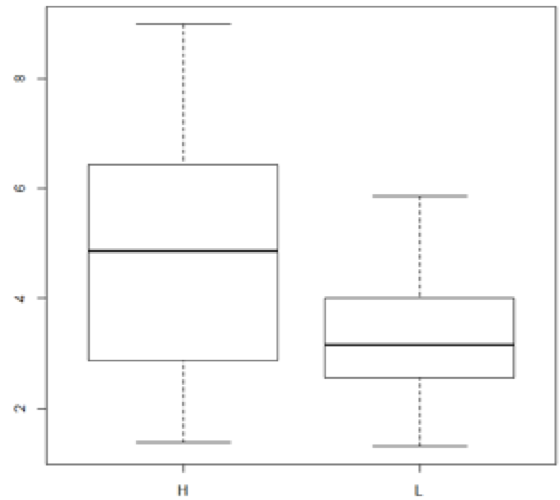

Figure 6: Box plots for effectiveness measures under high and low uncertainty conditions.

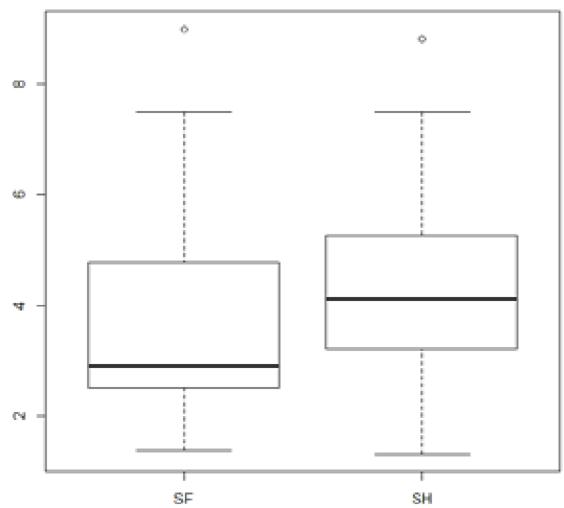

Figure 7: Box plots for effectiveness measures under SF and SH network type conditions.

5.3.3. H1: Analysis of variance, linear modeling, and correlations

In order to ascertain the significance of these differences of means, however, I ran multiple analyses. These included an analysis of variance (ANOVA), as well as a linear regression (linear coefficient modeling) analysis, and a polyserial correlation between the

IVs and DVs. The 2 IVs are NetworkType and Uncertainty. The DV is Effectiveness, which is calculated from the SIMPSON data.

When running these analyses, I took into consideration the interaction effects between the IVs (NetworkType and Uncertainty). In R, this is done by defining the variable relationships as (note the use of * operator):

Effectiveness ~ NetworkType * Uncertainty

The factorial ANOVA results, summarized in Table 23, demonstrate the following (significance for $p < 0.05$):

a) That the difference in the effectiveness results between SF and SH network types **were not significant**.

b) That the difference in the effectiveness results between LU and HU uncertainty settings **were significant**.

c) That there is **no significant interaction** between the 2 IVs: network types (SF / SH) and uncertainty levels (LU / HU) (see Figure 9).

Table 23: ANOVA for H1.

Response: Effectiveness	**Df**	**Sum Sq**	**Mean Sq**	**F value**	**p value**
NetworkType	1	5.541	5.541	1.7973	0.1837
Uncertainty	1	54.469	54.469	17.6678	**6.597e-05***
NetworkType:Uncertainty	1	0.062	0.062	0.0200	0.8878
Residuals	83	255.885	3.038		

The linear coefficient modeling analysis (linear regression analysis) revealed similar conclusions, as is shown in Table 24. That is to say, when modeling the relationship as a linear one, one can model the Uncertainty IV as a significant coefficient of the outcome variable, Effectiveness, but one cannot do that with the NetworkType IV. Additionally, this analysis reveals a weak linear model (adjusted R-squared of 0.1609).

Table 24: Linear coefficient modeling (regression analysis) results for H1 data (N = 84)

	Estimate	**Std. Err.**	**t value**	**Pr(>|t|)**
(Intercept)	4.5520	0.4028	11.300	< 2e-16*
NetSH	0.5524	0.5299	1.042	0.300
UncL	-1.5215	0.5775	-2.635	0.010*
NetSH:UncL	-0.1079	0.7619	-0.142	0.888

*Indicates significance p < 0.05
Residual standard error: 1.756 on 83 degrees of freedom
Multiple R-squared: 0.1901,   Adjusted R-squared: 0.1609
F-statistic: 6.495 on 3 and 83 DF, p-value: 0.0005303

Finally, I ran polyserial correlations between the quantitative DV (Effectiveness) and the bi-level categorical IVs (NetworkType and Uncertainty). Again, the correlations were in-line with the findings of the other 2 analyses I ran, meaning that there is a significant and somewhat strong correlation between Effectiveness and Uncertainty (-0.5293), but not with Effectiveness and Network Type. Furthermore, the correlation between Effectiveness and Uncertainty is negative: i.e. Effectiveness is higher when Uncertainty is LU. These findings are illustrated in Table 25.

Table 25: Polyserial correlation results for H1 data (N = 84)

polyserial(Effectiveness, Uncertainty, std.err=TRUE): Polyserial Correlation, 2-step est. = -0.5293 (0.1007) Test of bivariate normality: Chisquare = 15.19, df = 5, p = 0.009565
polyserial(Effectiveness, NetworkType, std.err=TRUE) Polyserial Correlation, 2-step est. = 0.1645 (0.1304) Test of bivariate normality: Chisquare = 10.45, df = 5, p = 0.0634

5.3.4. H1: Exploring possible mediating effects

Given that the results showed a significant effect for Uncertainty and a non-significant one for NetworkType, I decided to examine the effects of mediators in path models. The analysis results are shown in Table 26 and mediation plot is shown in Figure 8. The analysis shows no mediating effect from NetworkType further underlining the irrelevance of that variable.

Table 26: Mediation/Moderation Analysis for H1 data (N = 84)

The DV (Y) was Effectiveness. The IV (X) was Uncertainty. The mediating variable(s) = NetworkType.
Total effect(c) of Uncertainty on Effectiveness = 1.59 S.E. = 0.38 t = 4.22 df= 84 with p = 6.2e-05
Direct effect (c') of Uncertainty on Effectiveness removing NetworkType = 1.58 S.E. = 0.37 t = 4.23 df= 84 with p = 6e-05
Indirect effect (ab) of Uncertainty on Effectiveness through NetworkType = 0
Mean bootstrapped indirect effect = 0 with standard error = 0.07 Lower CI = -0.14 Upper CI = 0.15 R = 0.44 R2 = 0.19 F = 9.85 on 2 and 84 DF **p-value: 0.000144***

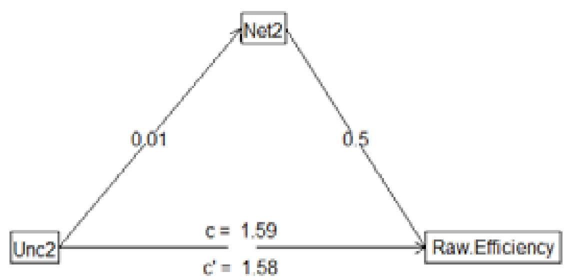

Figure 8: Mediation plot Effectiveness vs. Uncertainty with NetworkType as mediating factor

5.3.5. H1: Plotting the interaction effects between the 2 IVs

The ANOVA analysis showed that the 2 IVs did not interact with one another. An interaction plot of the factorial IVs, shown in Figure 9, bears this out.

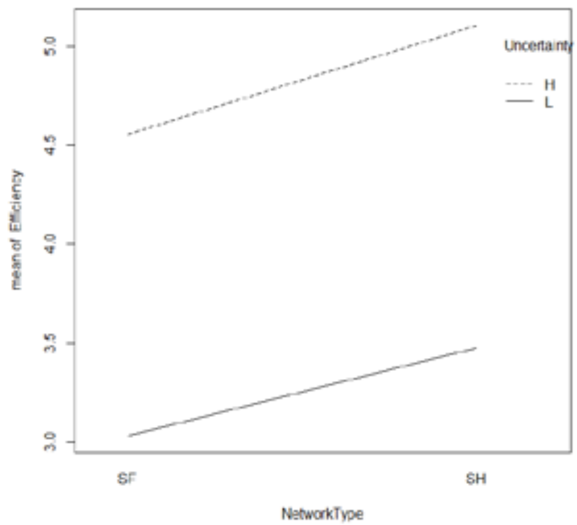

Figure 9: Effectiveness interaction plot for network type vs. uncertainty level

5.3.6. H1: Plotting the time-series

I plotted a variety of time-series graphs of the number of accumulated answers versus time. The time axis is measured in minutes and has discrete (i.e. whole number) values ranging from time = 0 to time = 10. The y-axis shows a range from 0 to 45, which is the maximum number of answers a participant could collect. I had 2,588 total number of data points, representing 86 participant-runs (i.e. 43 participants doing 2 runs).

The graph in Figure 10 shows the time-series graphs for all LU vs HU and all SF vs SH. Of note is the difference between LU and HU graphs which shows a steeper slope for HU. The graph in Figure 11 looks at all 4 IV factors: LU-SH, HU-SH, LU-SF and HU-SF.

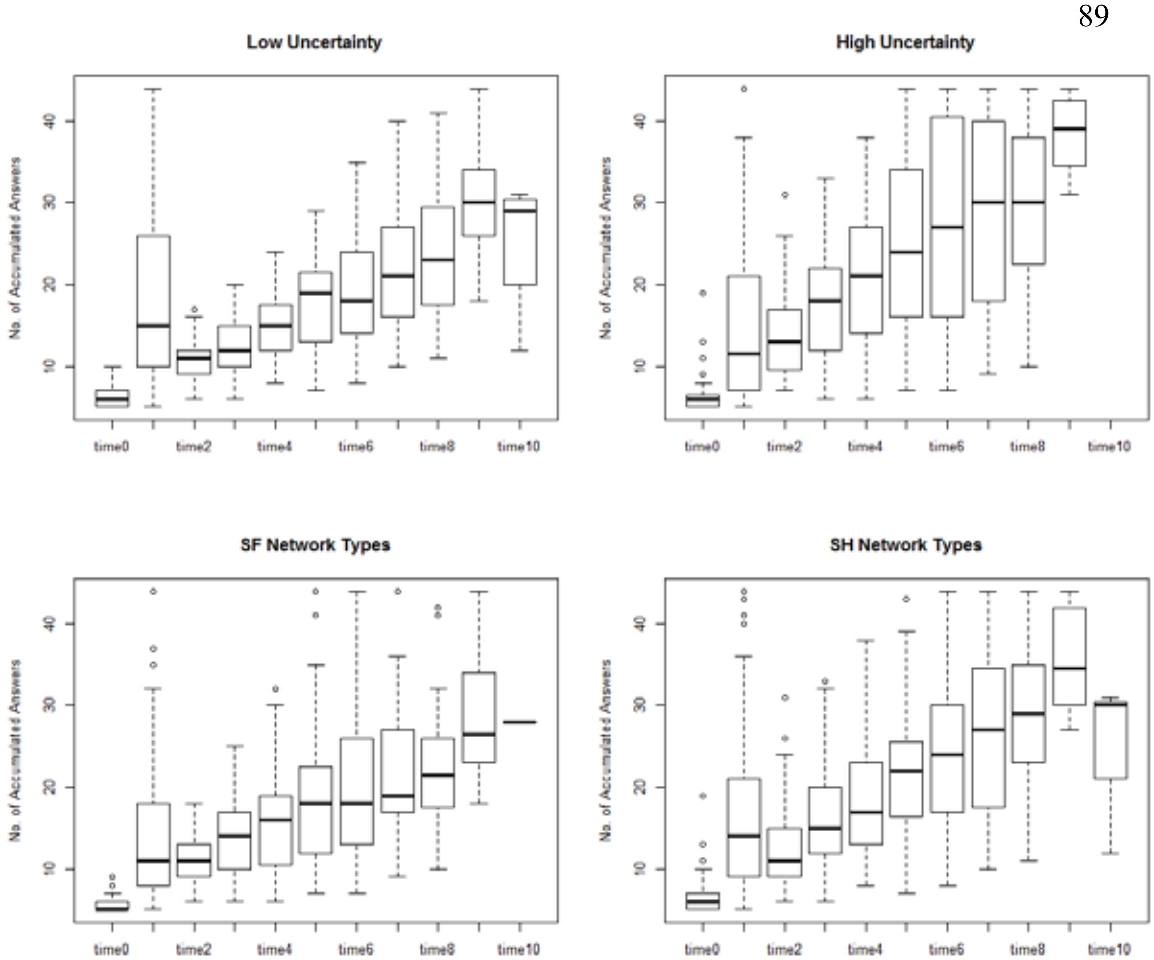

Figure 10: Time-series box plots of all low uncertainty cases (N = 1096), high uncertainty cases (N = 1492), SF network type cases (N = 1010), and SH network type cases (N = 1578).

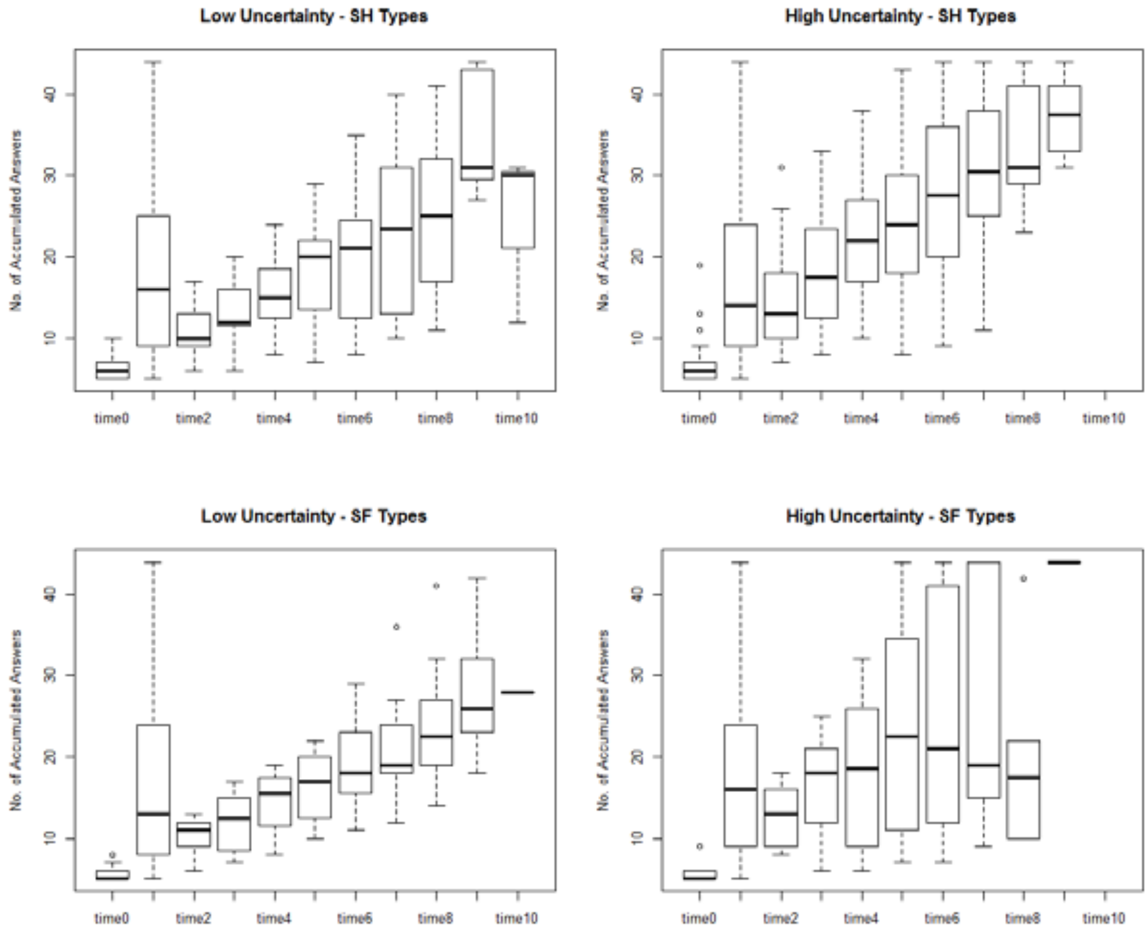

Figure 11: Time-series box plots of all LU-SH cases (N = 643), HU-SH cases (N = 936), LU-SF cases (N = 454), and HU-SF cases (N = 557).

5.4. H2: Number of steps taken to achieve information needs

The second hypothesis **(H2)** states that in both SH and SF networks, the level of uncertainty that the information seekers have will significantly determine if they employ more steps (if they have high uncertainty) or fewer steps (if they have low uncertainty) to get their information. Specifically, H2 claims that information seekers with higher uncertainty (HU) will take more steps to achieve their information goals than those with lower uncertainty (LU).

5.4.1. H2: Descriptive statistics

The data for H2 gathered from SIMPSON includes 92 entries. Each entry shows a total number of steps taken (that is, how many total actions were taken by the user) and a categorical data to indicate a low uncertainty or a high uncertainty scenario. Table 27 summarizes the mean, median, and standard deviation of these data points and Figure 12 shows the distribution of the Effectiveness variable in the data.

Table 27: Descriptive statistics for H2 (N = 92)

	Steps Taken for All	Steps Taken for LU	Steps Taken for HU
AVERAGE	108.54	97.20	119.89
MEDIAN	103.50	93.00	118.50
STDEV	50.07	47.69	50.32

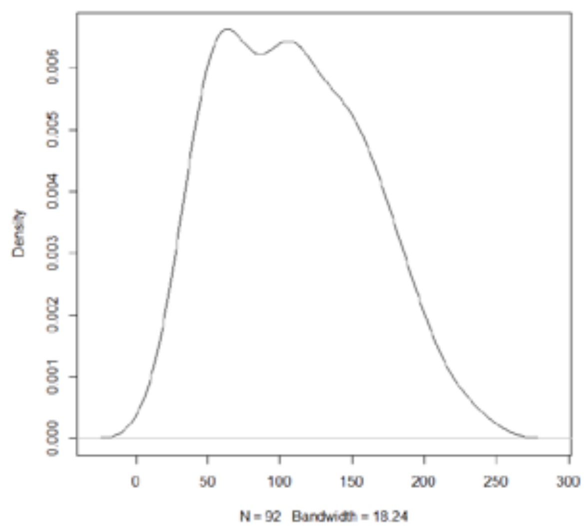

Figure 12: Distribution plot for the variable Steps for H2 (N = 92).

5.4.2. H2: Plotting the number of steps under the 2 uncertainty conditions

The plot in Figure 13 clarifies the difference in number of steps taken to achieve the user's information goal under the low and high uncertainty scenarios. The plot and statistical data show that high uncertainty situations necessitated more steps, on average, to be taken than low uncertainty ones (means of 119.89 for HU vs. 97.20 for LU).

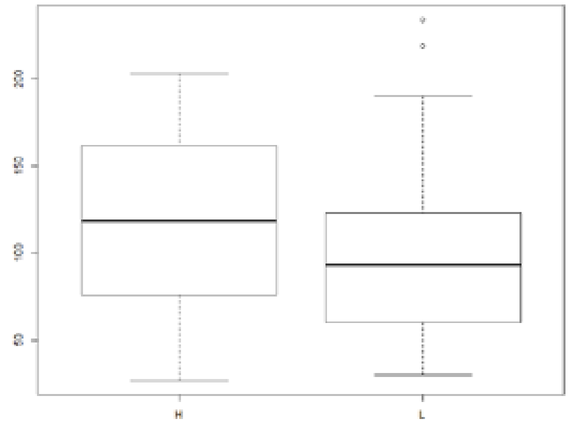

Figure 13: Box plots for number of steps measured under high vs low uncertainty conditions.

5.4.3. H2: Analysis of variance, linear modeling, and correlations

In order to ascertain the significance of these differences of means, however, I ran multiple analyses. These included an analysis of variance (ANOVA), as well as a linear regression (linear coefficient modeling) analysis, and a polyserial correlation between the IV and DVs. The IV is Uncertainty and the DV is Steps, which is calculated from the SIMPSON data.

The factorial ANOVA results, summarized in Table 28, demonstrate (with significance testing set for $p < 0.05$) that the difference between steps taken to achieve the user's information goal under the low versus the high uncertainty scenarios **is significant**.

Table 28: ANOVA for H2

Response: Steps	Df	Sum Sq	Mean Sq	F value	p value
Uncertainty	1	11847	11847.1	4.9292	**0.02892 ***
Residuals	90	216310	2403.4		

A linear coefficient modeling analysis and a post-hoc mean separation test for the main effect factor (Uncertainty) variable analysis revealed similar conclusions. However, this analysis also reveals a weak linear model (adjusted R-squared of 0.04139). Results are shown in Table 29 and Table 30.

Table 29: Linear coefficient modeling (regression analysis) results for H2 data (N = 92)

	Estimate	**Std. Err.**	**t value**	**Pr(>|t|)**
(Intercept)	119.891	7.228	16.59	**<2e-16 ***
uncL	-22.696	10.222	-2.22	**0.0289 ***
*Indicates significance p < 0.05 Residual standard error: 49.02 on 90 degrees of freedom Multiple R-squared: 0.05193, Adjusted R-squared: 0.04139 F-statistic: 4.929 on 1 and 90 DF, p-value: 0.02892				

Table 30: Post-hoc mean separation test for Uncertainty variable (pairwise Tukey) for H2 data (N = 92)

Uncertainty	**lsmean**	**SE**	**df**	**Lower.CL**	**Upper.CL**
H	119.89130	7.228328	90	105.53097	134.2516
L	97.19565	7.228328	90	82.83532	111.5560

Confidence level used: 0.95

Contrast	**Estimate**	**SE**	**df**	**t.ratio**	**p.value**
H – L	22.69565	10.2224	90	2.22	0.0289

Finally, I ran a polyserial correlation analysis between the quantitative DV (Steps) and the bi-level categorical IV (Uncertainty). This analysis revealed that the correlation between Steps and Uncertainty is -0.2796 (meaning Steps increased as Uncertainty was LU), but that it was not significant. This finding is illustrated in Table 31.

Table 31: Polyserial correlation results for H2 data (N = 92)

polyserial(Steps, Uncertainty, std.err=TRUE): Polyserial Correlation, 2-step est. = -0.2796 (0.1183) Test of bivariate normality: Chisquare = 6.095, df = 5, p = 0.2971

5.5. H3: Number of steps taken to achieve information needs

The third hypothesis **(H3)**, in the same vein as H2, states that users in SH networks in high uncertainty (HU) circumstances will connect with those occupying *structural holes* more often than individuals with low uncertainty (LU), to get their information. The number of times a node is connected with, or accessed, is represented by the variable **Accessed**.

In order to get a larger picture, I compiled my data for H3 to include, not just the **Uncertainty** variable, but also another categorical variable, called **NodeType**, that classified all the nodes as one of three types: nodes that are structural holes (SHC), nodes that are highly-centric in scale-free networks, that is, hubs (SFC), and neither of the two (NEITHER). Additionally, I included a quantitative variable, called **Degree**, that represented the degree centrality of each node at the initialization of the exercise (i.e. the initial value of the degree centrality of the node).

The statistical analysis that I did looks at the data from 2 points of view: one that considers only the type of node involved (using NodeType) and one that considers only the initial degree centrality of the node (using Degree).

5.5.1. H3: Descriptive statistics

The data for H3 gathered from SIMPSON includes 92 entries.

It is noteworthy to point out that of the three categories of nodes in NodeType, SHC comprised of only 8 nodes/entries, SFC comprised of only 4 nodes/entries, and the NEITHER category held the lion's share of the node/entries with 80.

Each entry shows a total number of times each node in the network was accessed (that is, the number of times another node requested access to it, whether directly or via a third-party), the type of node it was (SHC, SFC, or NEITHER), and a categorical data to indicate a low uncertainty or a high uncertainty scenario. Table 32 summarizes the mean, median, and standard deviation of these data points.

Table 32: Descriptive statistics for H3 (N = 92)

	Number of times accessed for All	Number of times accessed for LU	Number of times accessed for HU	Number of times accessed for SFC	Number of times accessed for SHC	Number of times accessed for NEITHER
AVERAGE	3.97	3.98	3.96	6.00	3.38	3.93
MEDIAN	4.00	4.00	4.00	5.50	3.50	4.00
STDEV	2.17	2.20	2.17	1.41	2.56	2.13
N	92	46	46	4	8	80

The distribution of the quantitative variable Accessed is shown in Figure 14.

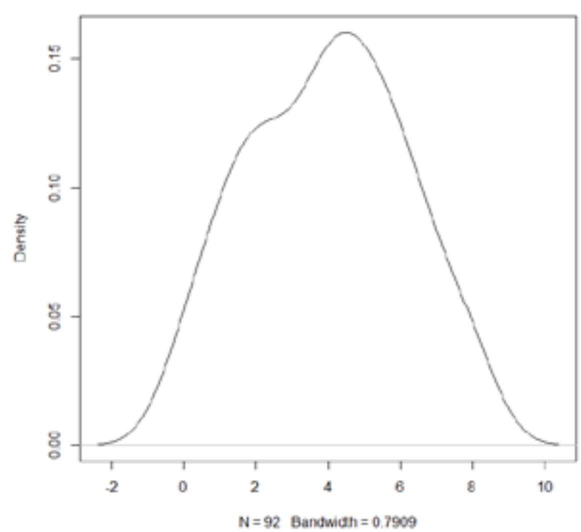

Figure 14: Distribution plot for the variable Accessed for H3 (N = 92)

5.5.2. H3: Plotting the number of times accessed

The plots in Figure 15, Figure 16, and Figure 17 show the difference in number of times a node was accessed under the low and high uncertainty scenarios (LU vs HU), according to node type (SHC vs SFC vs NEITHER), and according to node degree centrality, respectively. The plots and statistical data show that highly central nodes in the SF types of networks were accessed more often (6.00 vs. 3.39 and 3.93) than the other 2 types, regardless of uncertainty level. The plots also illustrate that the mean times a node was accessed in LU (3.98) vs. HU (3.96) was very similar. Additionally, the plots indicate that nodes with degree centrality 4 were more often accessed than other nodes with different degree centralities.

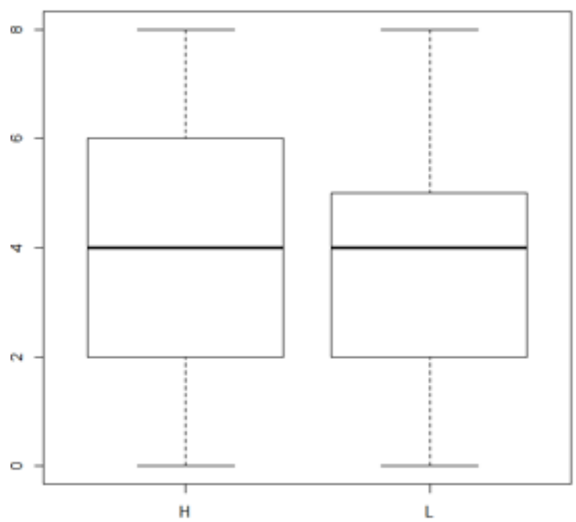

Figure 15: Box plots for number of times nodes were accessed in high vs low uncertainty conditions.

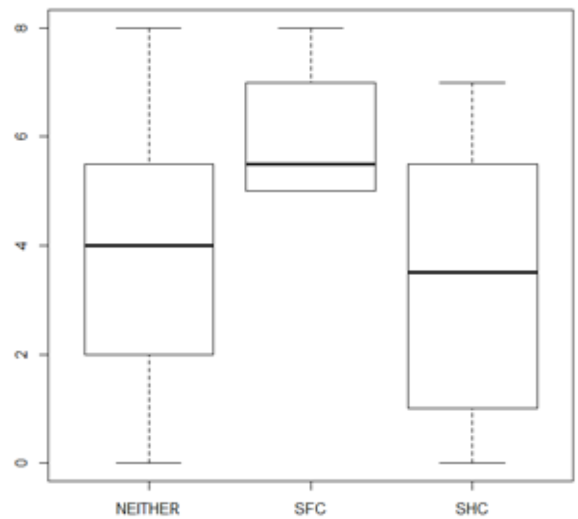

Figure 16: Box plots for number of times nodes were accessed by node types.

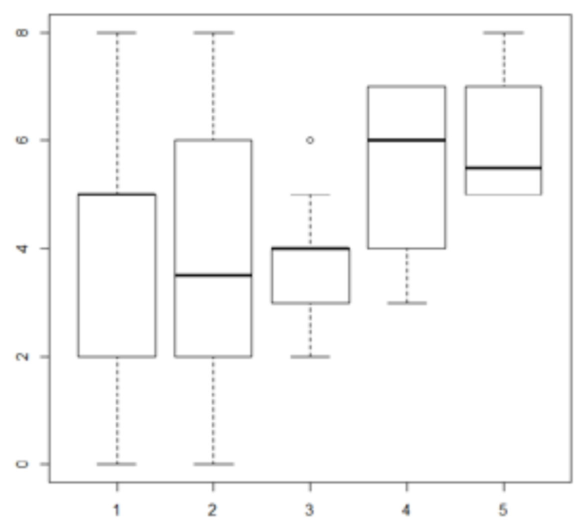

Figure 17: Box plots for number of times nodes were accessed by node degree centrality.

5.5.3. H3: Analysis of variance, correlations, and linear/non-linear modeling

In order to ascertain the significance of these differences of means, however, I ran an analysis of variance (ANOVA) with significance testing set for $p < 0.05$.

When running this analysis, I took into consideration the interaction effects between the IVs (NodeType and Uncertainty). In R, this is done by defining the variable relationships as (note the use of * operator):

Accessed ~ NodeType * Uncertainty

The ANOVA results, seen inTable 33, demonstrate that the differences between the number of times nodes were accessed according to their high-centrality types and under the low and high uncertainty scenarios **are not significant**. The same was found regarding the differences between the number of times nodes were accessed and the 3-level NodeType variable.

Table 33: ANOVA for H3 using NodeType and Uncertainty as IVs

Response: Accessed	Df	Sum Sq	Mean Sq	F value	p value
NodeType	2	0.3043	0.152165	2.0515	0.1348
Uncertainty	1	0.0002	0.00017	0.0023	0.9619
NodeType:Uncertainty	2	0.0182	0.009095	0.1226	0.8848
Residuals	86	6.3789	0.074173		

Given the skewed number of data points of the categories of NodeType, I decided to run an analysis of variance with the quantitative variable Degree instead of NodeType as IV. This, in R, this was done by re-defining the variable relationships as:

Accessed ~ Degree * Uncertainty

The ANOVA demonstrates that the differences between the number of times nodes were accessed according to their low and high uncertainty scenarios **are not significant**. However, the Degree variable showed a **significant** p-value.

Table 34: ANOVA for H3 using Degree and Uncertainty as IVs

Response: Accessed	**Df**	**Sum Sq**	**Mean Sq**	**F value**	**p value**
Degree	1	21.39	21.3943	4.6585	**0.03362***
Uncertainty	1	0.01	0.0109	0.0024	0.96131
Degree:Uncertainty	1	3.36	3.3569	0.7310	0.39489
Residuals	88	404.14	4.5925		

A Pearson's correlation of Accessed and Degree variables reveals that it is 0.223 and significant ($t = 2.173$, $df = 90$, p-value = 0.03235*). In order to examine the linearity of this relationship, I ran a linear coefficient modeling analysis as shown in Table 35. None of the coefficients came back as significant and the analysis also revealed a weak linear model (adjusted R-squared of 0.02561, i.e. only 2.56% of variance was explained).

Inspired by the plot in Figure 17, I also examined possible non-linear (polynomial) relationships. A quadratic model (Accessed ~ Degree^2 + Degree * Uncertainty) revealed a better adjusted R-squared (0.04865), but it was still very low (see Table 36). The quadratic term (Degree^2) was also not statistically significant at the $p < 0.05$ threshold, but it was significant for a $p < 0.1$. An examination of a cubic model (Accessed ~ Degree^3 + Degree^2 + Degree * Uncertainty) revealed adjusted R-squared of 0.03829 (i.e. a worse fit than the quadratic model) and no significant findings (see Table 37).

Table 35: Linear coefficient modeling (regression analysis) results for H3 data (N = 92)

	Estimate	**Std. Err.**	**t value**	**Pr(>|t|)**
(Intercept)	3.3043	0.7750	4.263	**5.05e-05***
Degree	0.2885	0.3130	0.922	0.359
UncertaintyL	-0.8340	1.0961	-0.761	0.449
Degree:UncertaintyL	0.3785	0.4427	0.855	0.395

*Indicates significance p < 0.05
Residual standard error: 2.143 on 88 degrees of freedom
Multiple R-squared: 0.05773, Adjusted R-squared: 0.02561
F-statistic: 1.797 on 3 and 88 DF, p-value: 0.1535

Table 36: Non-Linear (quadratic) coefficient modeling (regression analysis) results for H3 data (N = 92)

	Estimate	**Std. Err.**	**t value**	**Pr(>|t|)**
(Intercept)	5.1784	1.3070	3.962	**0.000152 ***
Degree2	0.2958	0.1672	1.770	**0.080314 #**
Degree	-1.3425	0.9722	-1.381	0.170867
UncertaintyL	-0.8340	1.0830	-0.770	0.443379
Degree:UncertaintyL	0.3785	0.4374	0.865	0.389279

*Indicates significance p < 0.05 #Indicates significance p < 0.1
Residual standard error: 2.118 on 87 degrees of freedom
Multiple R-squared: 0.09047, Adjusted R-squared: 0.04865
F-statistic: 2.163 on 4 and 87 DF, p-value: 0.07978

Table 37: Non-Linear (cubic) coefficient modeling (regression analysis) results for H3 data (N = 92)

	Estimate	**Std. Err.**	**t value**	**Pr(>|t|)**
(Intercept)	5.69500	2.44341	2.331	0.0221
Degree3	-0.03686	0.14696	-0.251	0.8026
Degree2	0.60307	1.23672	0.488	0.6270
Degree	-2.08666	3.12415	-0.668	0.5060
UncertaintyL	-0.83395	1.08892	-0.766	0.4459
Degree:UncertaintyL	0.37848	0.43979	0.861	0.38919

*Indicates significance p < 0.05 #Indicates significance p < 0.1
Residual standard error: 2.129 on 86 degrees of freedom
Multiple R-squared: 0.09113, Adjusted R-squared: 0.03829
F-statistic: 1.725 on 5 and 86 DF, p-value: 0.1375

CHAPTER 6: DISCUSSION

In this chapter, I will discuss my findings and examine whether or not my hypotheses were supported by the data or not. This chapter is conducted in four phases that entail discussing the findings of the pre- and post-survey analysis in Sections 6.1 and 6.2, followed by a discussion of the findings of hypotheses tests based on the log data from the Web-based SIMulated social-computational Platform with a SOcial Network environment (or SIMPSON for short) in Sections 6.3, 6.4, and 6.5. What follows then is a discussion of theoretical and practical implications of these findings in Sections 6.6 and 6.7. Finally, the chapter concludes with additional limitations of the research along with my recommendations for future studies in Section 6.8.

6.1. Pre-survey findings

A total of 54 participants took the survey both before and after running the experiment exercise on SIMPSON. The pre-survey questions were asked as to collect information on the participants' demographics, their experiences with information seeking (especially in social networks). The post-survey questions were asked as to collect information on the subjects' experiences with both the SIMPSON tool and the networks they interacted with in regards to how well (or not) they helped the participants complete their tasks. Both surveys are shown in Appendix 3.

The participant demographics broke down to 36 females (66.7%) and 18 males (33.3%), ranging in age from 18 to 24 years old ($M = 20$, $SD = 1.22$), and all of them had anytime access to a desktop, laptop or tablet computer, or Internet-capable smart-phone in their daily lives. Their general social media use was, as expected, very habitual, with 96.2% of them saying that they accessed any type of social media more than once a day.

When asked which social media sites they generally used and how often they used them, the participants indicated that the top 3 sites they frequented on a daily basis were Instagram (79%), Snapchat (70%), and Facebook (58%). This ordering by use is consistent with research findings from the Pew Internet & American Life Project on general use of social media services by adults ages 18 to 29 (Anderson & Jiang, 2018). When asked further about which online venues the participants usually used to look for information and how often, the vast majority (98.1%), unsurprisingly cited the search engine site Google. Furthermore, 0% said they "never" used Google. Interestingly, popular social media sites Instagram, YouTube, Snapchat, and Facebook also got a substantial amount of respondents who said they used them "often" to look for information (in the 43.4 to 35.8% range). Wikipedia, YouTube, and Yahoo! Answers were most often cited as used "sometimes" (in the 49.1 to 50.9% range). Surprisingly, the search engine site Bing – one of the main competitors to Google – was one of the most cited as "never" used for this purpose (82.7% of respondents).

Based on their last use of social media to ask a question, when the subjects who had participated in this activity (N = 44) were asked *what* instigated that session, the majority of them said that they had an information need of either a social nature (43.2%) or a work/school nature (27.3%). The majority of answers indicated that these sessions had lasted between 5 and 10 minutes (54.0%) and that the *type* of questions were mostly about getting factual information, that is, "what", "who", or "when" types of questions

(81.8%). The people they were asking questions to were described mostly as “friends” (46.8%) and a minority were described as “family” (8.5%). These answerers were further described as either perceived subject-matter experts (28.8%), trustworthy people (23.7%), or simply available to answer their questions (33.9%).

There were essentially no surprises in the answers given by my participants given available data we have from research from nation-wide polling efforts from Pew Research and others. We know that social media’s extremely pervasive presence in people’s online activities is also a large contributor to their motivations of everyday information seeking behaviors (Matni & Shah, 2014) even if the quality of the content varies (Agichtein et al., 2008) and the pre-survey results underline this clearly. The pre-survey results also show that, as far the participants’ perceptions of the social media services’ utility, most people asked questions online to pursue very practical informational needs (the quest for facts, mostly) from very practical answerers (people they saw as experts, trustworthy, or simply available). This simplicity of asking a question on social media, coupled with the easy knowledge of who to ask questions to, may be better understood as something that social media affords to its users (Treem & Leonardi, 2012).

Many of the commonly used social media sites that these participants have said they normally use are great examples of knowledge networks. Knowledge networks are used to make communities’ social networks visible to their users and serve as repositories of “who knows what?” and also “who knows who knows what?” (Contractor et al., 1998). Classically, knowledge networks have been studied in organizational settings and can say a lot about how organizations benefit from using them well or suffer from the opposite (Hansen, 2002; Majchrzak et al., 2013; Monge & Contractor, 2003).

6.2. Post-survey findings

The 14 post-survey questions asked participants in the study questions on their user-experiences with the SIMPSON tool and self-reported behaviors regarding their interactions with the other participants in the study.

In terms of **user experience** and questions designed to elicit feedback on the design of SIMPSON, the first 5 questions had yes/no answers and were asked to evaluate the participants' understanding on how to use the tool, given its multiple features. The feedback here is useful in assessing if the given explanations of the tool and its design were helpful or not to the participants. The survey participants said that the instructions of use and the tool features were mostly clear and helpful as the "yes" scores ranged from 68.5% to 81.5% (M = 73.0%, SD = 8.1%).

In open-ended follow-up questions, when asked to "Please describe your general (or any specific) observations, comments, thoughts, etc. regarding your experience with these exercises", the content analysis of the responses given (N = 47) shows that the majority expressed *positive sentiments* (55.3%) as opposed to *negative ones* (38.3%) or *neutral* ones (only 3 respondents). The biggest complaints or negative observations were that the SIMPSON tool confused them at times (17.0%), that the network diagram on SIMPSON was hard to read (10.6%), or that they felt frustrated at least once during the exercise (6.4%). On the other hand, a sizeable portion of the participants thought the exercise was a fun or interesting experience (27.7%), while others enjoyed figuring out what to do by actually doing the exercise (14.9%) or enjoyed finding insights into how social networks behave/work (12.8%).

What the post-survey revealed about the participants' user experience underlined that the majority of them found the tool useful for the task they were told to do and easy to navigate and use its multiple features. However, a substantial minority of them found the initial instructions unclear (22%, but then dropped to 18% after more instructions were given on the main landing page), the score and position measures unclear (24.1%), and expressed general negative sentiments overall, such as frustration or confusion, after running the exercises (38.3%). This is essential feedback to help in future re-design of the experiment and the instructions given to the subjects/participants. More specifically, I would recommend reviewing the clarity of the instructions. I would also recommend reviewing the data gathering tool's (i.e. SIMPSON or its next iteration) feature set, although I would caution against removing any features from the tool given the limitations already inherent to a software tool like SIMPSON.

The next 2 questions asked for **self-reported behavior while using the tool**. When asked if they actively refused connections with someone (which they could do by ignoring requests for connections from other users), the vast majority said "no" (96.3%). It is noteworthy to point out that the design feature of SIMPSON in regards to ignoring a connection request is as follows: users have to explicitly click on a "reject" button in order to reject an incoming request to connect, however, the request is also *timed-out* after 10 seconds resulting in a *de-facto rejection*. In other words, despite given the "easy way out" to reject a connection request, the vast majority of the subjects *actively* clicked the “accept” button when asked for a connection. When asked in an open-ended follow-up question to “explain why or why not”, the most common answer was that they accepted connections from other participants for their own personal gains in the exercise, such as gaining more points or more answers (60.9% of respondents). About 15.2% of the respondents said that they accepted connections because it fulfilled a perceived social norm of support of others (i.e. because it was the “good” thing to do and that it would be “bad” to refuse a connection request). Interestingly, there was only one respondent (out of 46) who cited their own personal gains in the exercise as a reason to *refuse* requests of connections with others. The most common answer to why the respondents said they refused connections with other participants was not for any strategic reason, but rather because they were unaware of the request for connections (10.9% of respondents), that is to say, because they did not fully understand the SIMPSON tool feature involved in this type of activity.

In terms of participants’ behavior while using the tool, more specifically their behavior around navigating and constructing their networks by accepting or rejecting requests from others to connect, it is interesting that the vast majority of the subjects actively accepted requests for connection. This indicates that, despite any misgivings about the clarity of instructions of use of SIMPSON, or any frustrations felt towards the tool, this “need” to connect with others was quickly understood by the users of the tool to be important for meeting their end information seeking/gathering goals.

The next question asked for the participants’ **experience with the different uncertainty scenarios** that they encountered. When asked, in a closed-ended question, what difference it made for them when conducting the exercises if they knew something about what topics the other participants knew, most respondents said that it *did make it easier* to reach their information seeking/gathering goals and to satisfactorily conclude the exercise (63.0% of respondents). However, a large minority of respondents also said that it *made no difference* to them (33.3% of respondents). Very few respondents (only 2 out of 54) said it made it more difficult to reach their information goals and conclude the exercise.

While the majority of the users said that the different LU and HU scenarios that they experienced made a difference to them in terms of seeking their goals, a third of them said the scenarios did not make a difference. This was the outcome of a closed-ended question with a Likert scale and thus making the answers given unambiguous. This finding is a little disconnected from some of the SIMPSON data findings that I analyzed which suggest that the LU vs HU settings were *statistically significant factors* in outcomes like the effectiveness of gathering information and the number of steps taken by the users to meet their information seeking/gathering goals in the exercises. This could indicate a disconnect between what users consciously think they are doing (i.e. seeing different uncertainty scenarios and not thinking that there was a difference) and what they might unconsciously be doing (i.e. experiencing different uncertainty scenarios and *acting* differently in each one). This conscious vs unconscious mind behavior is something that has been observed since the days of Freud's work in psychology in the early 20th century and has continued to be studied in social cognition research over the past few decades. This research suggests that many aspects of our decision-making, thoughts, and behaviors are, in fact, strongly influenced by unconscious processes (Vrabel & Zeigler-Hill, 2017). This observation of my subjects is not, in and of itself, very revealing beyond the fact that they acted differently than what about 33.3% of them claimed they did. It would be interesting to see if this discrepancy still shows up in future studies that have re-designed the data-gathering tool and re-written the instructions in an effort to lower the percentage of participants who felt that certain instructions or features were unclear.

The final question in the post-survey asked for the participants' self-reported **information seeking strategies** that they employed while doing the exercise. When asked in an open-ended question to explain what strategies, if any, they had developed to help maximize their scores in the exercise, a majority of respondents (75.9%) said that they *did utilize specific strategies*, while only 20.4% of them said they had *no specific strategy* (that they were at least self-aware of). It is worth noting that this is a question asked "after the fact" and it is possible that it yielded biased answers.

6.3. Hypotheses testing findings for H1: The influence of uncertainty and network types on effectiveness of information seeking/gathering

6.3.1. Restatement of H1

My initial hypothesis (**H1**) claims that if I were to observe two types of networks – one with structural holes (**SH**) and one with a scale-free topology (**SF**) – then I should see people being *more effective* with their information seeking and gathering activities in a SH network. If these networks are further distinguished with characteristics that help reduce uncertainty (specifically, if the users have some knowledge of who-knows-what), then we should see *more* effective behavior in these low-uncertainty (**LU**) conditions than in those networks that give no information about who-knows-what (high-uncertainty, or **HU**). The hypothesis is tested using a 2x2 factorial design, as illustrated in Table 1, and H1 expectations are summarized in Table 7.

The outcome variable (DV) sought was **Effectiveness**. This was calculated from SIMPSON data by dividing the total number of answers gathered by the time needed to acquire these answers. The 2 IVs were the categorical factors and called **NetworkType** (SF, SH) and **Uncertainty** (LU, HU).

6.3.2. Summary of statistical findings of H1

My analysis of the findings indicates that the Effectiveness variable was highest, on average, under the Uncertainty = HU factor (mean of 4.86 vs. 3.16 for LU), and under the NetworkType = SH factor (mean of 4.12 vs. 2.90 for SF). However, only the first finding (i.e. Uncertainty) was statistically significant ($p < 0.05$). This finding was consistent as observed outcomes of multiple statistical methods of inquiry and analysis that I used, including ANOVA, linear regression, polyserial correlations, and mediation effects.

6.3.3. Discussion of H1 findings

H1 was not supported by the data. I expected that the highest effectiveness in gathering information would occur in the SH network and under conditions of lower uncertainty. The data actually found that the SH network ***did*** show more effective data gathering behavior, but that finding was not statistically significant and may suggest that this variable was not very relevant to understanding effectiveness. The data also found, with statistical significance, that ***higher*** uncertainty conditions brought out higher effectiveness of data gathering behavior from my subjects.

H1 claimed that higher uncertainty would be associated with a lower effectiveness of gathering data based on the Information Search Process (ISP) model which describes uncertainty as a present factor in almost all information seeking tasks, especially at the start of such processes. Furthermore, the ISP model describes a drop in uncertainty as the information seekers get closer to their goals.

My data suggests, however, that a higher uncertainty environment made my subjects be *more effective* with their data gathering than the low uncertainty environments did. As described in earlier chapters, the experiment ensured that ordering effects (of LU vs HU environment runs) were mitigated by having some subjects start off with a LU environment run (then follow up with a HU run), while others had their runs ordered in reverse of that.

Kuhlthau (1991, 1993) has explored the role of uncertainty in great detail in her information search process (ISP) model. Her model shows that uncertainty levels in users searching for information drops as their search progresses, being highest at the beginning (or what she calls the Initiation stage) and lowest at the end (the Presentation stage). True to the complexities of actual human behavior, Kuhlthau's ISP model also accounts for a change in this otherwise linear relationship of uncertainty level versus time: more specifically, that the introduction of new information can increase uncertainty, especially in the earlier stages of the search process. However, overall, the effect of a low uncertainty scenario in an information search process is expected to be more beneficial to the users in terms of achieving their information goals, than being in a high uncertainty scenario. This is an important reason why H1 was formulated as it was.

The data, however, indicates the opposite effect in play: higher uncertainty scenarios begat better effectiveness measures in meeting the information seeking goals of the participants in my experiment. The graphs illustrated in Figure 10 show the *accumulated* number of answers that participants achieved over time (all LU vs HU and all SF vs SH). This is analogous to my effectiveness measures for H1. Of note is the difference between LU and HU graphs which shows a steeper slope for HU, indicating that participants collected their answers *faster* under this HU condition than the LU condition. Interestingly, the LU condition, as illustrated by the graph, shows a higher accumulation in the beginning of the task (i.e. within the first 2 minutes), but then the rate drops precipitously – this can be interpreted as a representation of the phenomena described by Kuhlthau wherein more information in the early stages of the search can create more uncertainty. Otherwise the graphs of LU vs. HU show more-or-less linear behavior, with the obvious difference being a slower rising trend in the LU scenarios. It is further interesting that the box-plot graphs here illustrate that, under the HU condition, there are participants who got to their maximum accumulation (45 answers) as early as in minute 5, while those under the LU condition got to that same metric much later in minute 9.

I posit that higher uncertainty environments motivated my subjects to be more careful about their actions, especially given that their actions were tied to points that ultimately were a decisive factor in how my subjects chose to behave. The subjects in my experiments did not have "free reign" to take any action they wanted – on the contrary, their actions came at a "price" – and if they "spent" all their points without some consideration, their information seeking goals would be compromised either by placing them behind others in the exercise or even by stopping the exercise if they completely ran out of points. This pushed the subjects of the experiment to weigh their actions carefully, and ever more so when they were in highly uncertain environments when trying to achieve their information seeking goals.

By contrast, the lower uncertainty environment may have lulled my experimental subjects into being a little more "lazy" in their pursuit of their information seeking/gathering goals and hence be less effective by comparison with the HU set of subjects.

We know from prior and important information science research that an information seeker's problem is usually not topical, but rather cognitive and needs to be understood within the larger situation of tasks and goals. In an effort to better understand my findings, especially since they were not explicitly explained by the ISP model, I looked at the literature in Cognitive Psychology to help clarify these findings for the H1 results. The prospect theory (Tversky & Kahneman, 1992) is a descriptive model of decision making under uncertainty that describes the way people choose between probabilistic alternatives (read as varying levels of uncertainty) that involve risk. Prospect theory states that people make decisions based on the potential value of losses and gains rather than the final outcome, and that people's behavior around that is not always rational or about pure utilitarianism (i.e. it is often irrational). One key aspect of this theory is that, when evaluating gains versus losses, the losses outweigh the gains in the minds of most people. In other words, people are usually averse to losing something of value (money, perceived power or influence, points in a game, etc…) in the same amount that they might gain it instead: or in other words, the hate of a loss is greater than the love of an equivalently valued gain.

I believe that this could explain my findings vis-à-vis the observed effectiveness in a high-uncertainty (HU) setting compared to a low-uncertainty (LU) setting. If we accept that a HU setting is likely to heighten the subjects' sense of aversion to potential losses, then we can understand how this might make the subjects more likely to make more of an effort to secure their information seeking/gathering gains, as observed via the Effectiveness variable difference of means in an HU vs. LU setting.

6.4. Hypotheses testing findings for H2: Number of steps taken

6.4.1. Restatement of H2

My second hypothesis (**H2**) claims that, in both SH and SF networks, the level of uncertainty that the information seekers have will significantly determine if they employ more steps (if they have high uncertainty) or fewer steps (if they have low uncertainty) to get their information. Specifically, H2 claims that information seekers with higher uncertainty (HU) will take more steps to achieve their information goals than those with lower uncertainty (LU).

The outcome variable (DV) sought was **Steps**. This was obtained from the SIMPSON data. The IV was the categorical factor **Uncertainty** (LU, HU).

6.4.2. Summary of statistical findings of H2

My analysis of the findings indicates that the Steps variable was highest, on average, under the Uncertainty = HU factor (mean of 119.89 steps taken vs. 97.20 for LU). This finding was consistent and statistically significant ($p < 0.05$) as observed outcomes of multiple statistical methods of inquiry and analysis that I used, including ANOVA, linear regression, and polyserial correlations.

6.4.3. Discussion of H2 findings

H2 was supported by the data. My findings backed up my expectations that people would take *a greater number of steps* in gathering information under conditions of *higher* uncertainty.

An interesting aspect of this H2 test finding is when we compare it with findings for the H1 test. The latter found better effectiveness in HU situations, meaning that, on average, the subjects took less time in HU situations to meet their information seeking/gathering goals. It turns out that the subjects also took fewer steps, on average, in LU situations. Again, prospect theory (Tversky & Kahneman, 1992) can again shed light on these two findings, as explained with the H1 findings discussion. Since people abhor losing something of value more than they delight in gaining it, and if we agree that a HU setting is likely to heighten this sense of aversion to losses, then we can understand how this might make the subjects more likely to make more of an effort to secure their information seeking/gathering gains in LU settings, as observed via the lower Steps mean for LU vs. HU setting.

6.5. Hypotheses testing findings for H3: Nodal access

6.5.1. Restatement of H3

The third hypothesis (**H3**) states that users in SH networks in high uncertainty (HU) circumstances will connect with those occupying structural holes more often than individuals with low uncertainty (LU), to get their information. The number of times a node is connected with, or accessed, is represented by the variable **Accessed** (DV). This was obtained from the SIMPSON data. The H3 test utilized three IVs in the test: the categorical factors **NodeType** (SFC, SHC, NEITHER) and **Uncertainty** (LU, HU) and the quantitative interval variable **Degree**. NodeType categorized if the node was a SF hub (SFC), a structural hold in a SH network type (SHC), or neither of the two (NEITHER). Degree represented the degree centrality of the nodes in the data set.

6.5.2. Summary of statistical findings of H3

My analysis of the findings indicates that the Accessed variable was highest in SFC type nodes, although that finding was not significant. Accessed was also undifferentiated per Uncertainty setting (and this factor was not significant either).

However, the variable Degree proved to be a significant IV and this finding was consistent and statistically significant ($p < 0.05$) as observed outcomes of ANOVA and Pearson correlation statistical methods. However, linear and non-linear (polynomial) regressions showed that Degree was not a significant linear/non-linear component to the DV.

6.5.3. Discussion of H3 findings

H3 was not supported by the data. My findings could not substantiate my expectations that people would access structural hole nodes more often (regardless of uncertainty setting). This is why I ventured further into the H3 testing to include a new variable, Degree (representing node degree centrality). Ultimately, the variable Degree showed a positive relationship with the variable Access ($r = 0.223$, p-value = 0.03235), indicating that the higher the node degree centrality, the more that node got accessed, regardless of network type (SF vs SH).

One of the biggest detriments with this testing, I believe, is the lack of enough data points. When trying to ascertain if structural hole types of nodes (or indeed if scale-free hub nodes) got accessed more often than others, I had data on too few of such nodes to examine ($N = 8$ for SFC types and $N = 4$ for SHC types of nodes).

6.6. Theoretical implications

This dissertation shed some light on our understanding of the dynamics between human behavior and their environmental structures. The research combines concepts and theories of both information seeking and social networks, two fields that should have more overlapping theories in common than they currently do, given the growing importance of the role of online social networks in helping people find information.

Most scholars in information science research would likely agree that information seeking comes from realizing a need for information and that cognitive perspectives involving communication, sensing, or thinking play a big part therein (Belkin et al., 1982; Dervin, 1992; Ingwersen, 1996; Kuhlthau, 1993). In utilizing information science's understanding of the role of uncertainty in information seeking, I sought to prove a hypothesis of how people met their information seeking/gathering goals (as represented by how effectively they did so) under low versus high uncertainty conditions. Key information science research suggests that low uncertainty conditions in information seeking are better for meeting those goals, but my data strongly supports the idea that, while low uncertainty conditions are significantly tied to information seekers taking *fewer steps* to achieve their information seeking/gathering goals, it is actually *high* uncertainty conditions that seem to engender *better effectiveness* of meeting their goals. This is a counter-intuitive finding that seems to belie the rational model of decision making.

There are hints to this in earlier important information science research. Saracevic and Kantor (1988b) claim that, in their experiments with TREC data, the number of search terms, preparation time, and total time used in a search had a *negative effect* on their effectiveness measures on relevance odds. Increasing the number of search terms, or the amount of preparation time, or the total time used in a search are all akin to creating a *less uncertain environment* for information seekers. This is a parallel finding to mine, that is, that effectiveness measurements are *better* in a high uncertainty scenario (see H1 test results). Saracevic and Kantor further point out that this particular finding of theirs is "a challenge to many accepted (but untested) models of searching" and that more research is needed (p. 201). My review of the relevant literature leads me to believe that that last recommendation had remained disappointingly unfulfilled and that the work in this dissertation is a strong step in that direction.

The role of non-rationality in decision making information seeking or information retrieval processes is not unexplored in information science. For instance, affective processes and how they play a role in people's decisions when facing complex informational tasks has been researched by HIB scholars who have modeled such behavior in general information seeking processes (Kuhlthau, 1991, 1993). Others have examined the role of affective states in specific areas of information science research like collaborative information seeking (Shah & González-Ibáñez, 2010) or information processing strategies (González-Ibáñez & Shah, 2016). However, while many descriptive theories and models are found in information science that pertain to the role of uncertainty in information seeking, there could be more research to help describe the role of irrationality in decision making in uncertainty conditions. I thus propose that HIB scholarship widen its reach further and integrate other theories of non-rational human behavior like the prospect theory (Tversky & Kahneman, 1992) or others to help explain certain phenomena in human-information behavior. I have elaborated on this theory and its ties to my hypotheses results in the earlier Sections 6.3.3 and 6.4.3 of this chapter.

Having made that recommendation, I must point out that prospect theory is not ignored or unknown in information science or social networks scholarship. For example, it is mentioned in research about decision making and information use by public safety teams responding to major incidents with heightened uncertainty conditions (Mishra, Allen, & Pearman, 2013), and how people make decisions when they are active in online social networks in risky situations (Askew & Coovert, 2013), but it has been mostly relegated to the background and has not been used as a main explanatory concept in human information behavior (HIB) scholarship, per se. Given its power in explaining certain human behavior outside of the classical rational model, I believe its inclusion in the compendium of HIB-related theories would be useful. Of course, more studies are needed in HIB, in general, and in the area of information seeking, in particular, that can utilize prospect theory to explain information seeking behavior, especially in uncertain environments. This dissertation, given its investigations and findings, can be used as a guideline for such future studies.

6.7. Practical implications

From a practical viewpoint, this dissertation contributes to both information seeking models and theories around networked environments. However, some of the results of the experiment in this dissertation suffer from low external validation and low generalizability. I will expand on this in the next section of this chapter. Should future similar studies on the role of uncertainty in information seeking behavior come to similar conclusions, then the information science scholarly community could be more certain of how high uncertainty scenarios play a role in improving information seeing processes.

The effect of network type was not very pronounced in the hypotheses test results in that any differences of means were decidedly not statistically significant. The implications of this are unfortunately vague in that they can be interpreted in multiple ways, such as: that different network types are not as influential on information seeking behavior as we previously thought, or that scale-free networks and networks with pronounced structural holes influence information seeking behavior much too similarly to one another (i.e. there is not enough differentiation).

The effect of uncertainty level, however, was pronounced in the hypotheses test results (H1 and H2 in this case) and statistically significant. An implication of this, and especially for future experimental studies, is that high and low uncertainty situations can be used as predictable settings for how well information seekers will perform in their tasks in simulated environments like the SIMPSON tool.

SIMPSON can be (and should be) further adapted to help academics and others research human information and communication behaviors in social networks. SIMPSON has been modeled with certain real-world online social networking sites that lend themselves well to people seeking information from others they are linked (or can link) to, such as certain community-oriented sites like Reddit, but on a smaller scale (smaller networks) and on a more limited basis (less functionality and choices of connection dynamics than most online social networking sites). SIMPSON has also been modeled with Web-based knowledge sharing tools, where users are made aware of “who knows what” and are therefore guided to certain individuals in their social network in order to get answers or gain knowledge. The design of the tool was very useful in helping gather the needed data for this dissertation. Features of SIMPSON, such as the type of networks in the environment, or the number or quality of questions posed to the users, or the initial scores that users begin the exercises with, are very easily modified with a preparatory input data file (in simple JSON format that does not require any computer programming, per se). Adding or modifying other features, like changing the “look and feel” of the tool, or incorporating built-in survey tools into SIMPSON, and so forth, can also be done relatively straightforwardly, albeit requiring more than just an input data file since some computer programming would be required.

6.8. Additional limitations of the research and suggestions for future research

While this dissertation has revealed certain interesting aspects of the role that network structure and level of uncertainty play in securing certain information seeking goals, it has also revealed certain limitations.

To begin with, there is an issue of low external validity of the findings that I alluded to earlier. While the pre- and post-surveys were conducted with a semi-random selection of undergraduate students at one university (University of California, Santa Barbara), this population could not be construed as a representative sample of a general population of undergraduates anywhere, let alone a general population of information seekers anywhere. This restricts the generalization (external validity) of the findings to a very specific population. Similarly, the significant findings stemming from the experiment run on SIMPSON, while having high internal validity, also has a problem of low external validity. This is a common weakness of most laboratory-based experiments. To alleviate this particular weakness of low external validity of the findings, I would recommend (a) running the surveys independently and to a larger target population selected via randomized selection methods, and (b) running another experiment that can be implemented "in the field" yet still in a somewhat controlled manner (although that has its own limitations in terms of internal validity and may have to be classified as a quasi-experiment instead).

Secondly, some of the experimental results lacked statistical significance. More specifically, this was true for part of H1 and for H3 testing. For H1 testing, network differences (between SF and SH types) did not show up as significant one way or another. As mentioned earlier, this could mean that there is not enough differentiation in effect on information seeking behavior between scale-free networks and networks with pronounced structural holes. It may also be that re-conducting this experiment with more participants than (nominally) 10 people may make a difference with the results obtained. As such, another recommendation for future research aiming at replicating this dissertation's experiments is to construct the networks with substantially more than 10 people per network (20-30 people may be a good target). The obvious difficulty with this is that recruitment of experimental subjects becomes more challenging, since the participants have to engage with each other in real-time and at the same time.

For H3 testing, it would have been more ideal, from a quantitative and statistical point of view, to have much more than 8 nodes specified as structural holes to study. Future studies examining the roles of important network nodes, such as structural holes or scale-free hubs, in information access should consider having at least 5 to 10 times more such nodes in their data. This, of course, would mean re-running these experiments with as many as 500 subjects, instead of this study's 43, which is a daunting task. However, this research has underlined my conviction that further research must push forward on examining how information seeking and social network structures influence one another. To avoid some of the pitfalls I came across here, I would suggest that future studies look at other topologies than just scale-free vs. structural hole types. This dissertation controlled for average node centrality and network densities in the experimental runs. Future studies might want to vary those control variables: in other words, have the network types all be one type (I recommend scale-free topologies) but with varying sizes, node centralities, and network densities. My suggestion for the use of scale-free topologies is made for two reasons: because scale-free networks best represent actual social networks found online or offline and because this dissertation has shown that the structural hole type of topology did not yield significantly different results from the scale-free topology.

CHAPTER 7: CONCLUSION

Subjectivity and socialization are the common elements in how we contextualize reality, experience life, and build up and use knowledge (Schutz & Luckmann, 1973). As seen with this lens, human behavior is best understood in a larger interconnected system. Who we are and what we know are intricately enmeshed. In this dissertation, I have examined the interplay between human behavior in and around seeking and gathering information and the environmental social structures they find themselves in and found interesting and useful results.

I have posed two general questions that the information science and social network literature do not answer: How do certain structures of social networks influence information seeking behavior? And how do information seekers' states of uncertainty influence what strategies they employ in order to get answers to their questions in a social network? This dissertation formed these into two general research questions:

RQ1*: How do these two different social network topologies influence the effectiveness of information seeking and gathering activities of individuals who experience different levels of information seeking uncertainty in their task?*

RQ2*: How do information seekers' states of uncertainty influence what strategies they employ in order to get answers to their questions in different well-structured social network topologies?*

With regards to RQ1, the dissertation showed that the effect of network type was not very pronounced, when we control for network density and average degree centrality of the nodes. The two social network structures used in the experiment proved not to be significant factors in terms of their differentiated influence on the effectiveness of information seeking/gathering activities of the subjects. There is room to further explore different social network topologies, or alternatively, different settings of network characteristics in the service of better understanding their influence on information seeking behavior.

As for RQ2, the dissertation did show that information seekers' states of uncertainty did indeed influence what strategies they employed. For one thing, it became clear that high uncertainty situations were tied significantly and positively to how effectively these individuals pursued their information seeking/gathering goals – that is to say, high uncertainty situations were the ones where information seekers achieved their goals in a shorter amount of time, even though they took more steps to achieve these goals.

This research sheds light on our understanding of the dynamic between human behavior and their environmental structures. We know that people are influenced by the social structures they find themselves in and what this dissertation has further shown is that the level of uncertainty in these situations plays an important role in the activities of people's information seeking/gathering. The dissertation has thus contributed to both information seeking models and theories around networked environments. This research combines concepts and theories of both information seeking and social networks, two fields that should have more overlapping theories in common than they currently do, given the growing importance of the role of online social networks in helping people find information.

Furthermore, the Web-based tool that I designed for use in this dissertation, called the SIMulated social-computational Platform with a SOcial Network environment (or SIMPSON for short), has proven itself quite useful in its data gathering capabilities of experiments of information seeking behavior in networked environments. SIMPSON can be further adapted to help academics and others research human information and communication behaviors in social networks with little to moderate effort.

Although social media and its use in information seeking, searching, and retrieval is a relatively new phenomenon, it has become virtually omnipresent and shows no signs of continuing to change and adapt itself in people's daily information-seeking lives. The combination of the ubiquity and ease-of-access of social media, its inherent richness of information, and its natural facilitation of social networking make it a powerful emerging way to seek and share information.

There is a continuing need for further research in this cross-sectional area of information science and social networks. Not least of which is research that continues to refine models of information seeking and the processes involved. I believe that bringing in ideas, concepts, and theories from other scholarly areas, especially behavioral psychology and/or economics, as evidenced by the usefulness of the prospect theory to explain the results of certain hypotheses presented here.

As a result of this work, I further put forward the idea that situations of high-uncertainty in networked environments, because of the added anxiety or stress that they can engender in people, can predictably bring out motivations of greater effort of reaching individuals' information seeking or information gathering goals under certain network topology characteristics. This idea needs more research to bear out a clearer understanding and maybe even predictive models.

In addition to providing initial evidence for this idea, this dissertation has contributed through extensive literature review toward research in this cross-disciplinary area. Moreover, the research framework and the methodology and tool design introduced in Chapters 3 and 4, respectively, provide a valuable foundation, approach and tools to address future research in this topic despite the limitations discussed in the prior chapter.

REFERENCES

Agichtein, E., Brill, E., & Dumais, S. (2006). Improving web search ranking by incorporating user behavior information. *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '06*, 19. https://doi.org/10.1145/1148170.1148177

Agichtein, E., Castillo, C., Donato, D., Gionis, A., & Mishne, G. (2008). Finding high-quality content in social media. In *ACM International Conference on Web Search and Data Minig (WSDM)* (p. 183). New York: ACM Press. https://doi.org/10.1145/1341531.1341557

Anderson, M., & Jiang, J. (2018). *Teens, Social Media, and Technology 2018*. Washington, D.C.

Aral, S., & Walker, D. (2012). Identifying influential and susceptible members of social networks. *Science*, *337*(6092), 337–341. https://doi.org/10.1126/science.1215842

Askew, K., & Coovert, M. D. (2013). Online decision-making. *The Social Net: Understanding Our Online Behavior*, 99.

Baeza-Yates, R., & Ribeiro-Neto, B. (2011). *Modern information retrieval*. New York: ACM Press.

Barabasi, A.-L. (2009). Scale-Free Networks: A Decade and Beyond. *Science*, *325*(5939), 412–413. https://doi.org/10.1126/science.1173299

Barabási, A.-L., & Albert, R. (1999). Emergence of Scaling in Random Networks.

Science, *286*(5439), 509–512. https://doi.org/10.1126/science.286.5439.509

Barabási, A.-L., & Bonabeau, E. (2003). Scale-Free Networks. *Scientific American*, 60–69.

Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Information Review*, *13*, 407–424.

Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982). ASK for information retrieval, Part I: Background and theory. *Journal of Documentation*, *38*(2), 61–71.

Bilal, D. (2000). Children's use of the Yahooligans! Web search engine: I. Cognitive, physical, and affective behaviors on fact□based search tasks. *Journal of the American Society for Information Science*, *51*(7), 646–665.

Björneborn, L. (2006). 'Mini small worlds' of shortest link paths crossing domain boundaries in an academic Web space. *Scientometrics*, *68*(3), 395–414. Retrieved from http://link.springer.com/article/10.1007/s11192-006-0119-8

Björneborn, L., & Ingwersen, P. (2004). Toward a basic framework for webometrics. *Journal of the American Society for Information Science and Technology*, *55*(14), 1216–1227. https://doi.org/10.1002/asi.20077

Borgatti, S. P. (2005). Centrality and network flow. *Social Networks*, *27*(1), 55–71. https://doi.org/10.1016/j.socnet.2004.11.008

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2013). *Analyzing Social Networks*. Thousand Oaks, CA: SAGE Publications.

Borgatti, S. P., & Foster, P. C. (2003). The Network Paradigm in Organizational Research: A Review and Typology. *Journal of Management*, *29*(6), 991–1013. https://doi.org/10.1016/S0149-2063

Boulding, K. E. (1956). *The image: Knowledge in life and society*. Ann Arbor, MI: University of Michigan Press.

Bourdieu, P. (1984). *Distinction. A social critique of the judgement of taste*. London: Routledge. Retrieved from https://monoskop.org/images/e/e0/Pierre_Bourdieu_Distinction_A_Social_Critique_of_the_Judgement_of_Taste_1984.pdf

Brookes, B. C. (1973). Numerical Methods of Bibliographic Analysis. *Library Trends*, 18–43.

Burger, P. L., & Luckmann, T. (1966). *The social construction of reality: A treatise in the sociology of knowledge*. *Allen lane*. Garden City, NY: Doubleday & Company, Inc.

Burke, M., Kraut, R. E., & Marlow, C. (2011). Social capital on Facebook: Differentiating uses and users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)* (pp. 571–580). New York.

Burt, R. S. (1992). The Social Structure of Competition. In *Networks and organizations: Structure, form, and action* (pp. 8–49). Cambridge, MA: Harvard Univeristy Press. Retrieved from http://172.218.224.236:8080/pdfs/Burt (1992) - Structural Holes The Social Structure Of Competition.pdf

Burt, R. S. (1997). The contingent value of social capital. *Administrative Science*

Quarterly, *42*(2), 339–365. Retrieved from http://www.jstor.org/stable/10.2307/2393923

Burt, R. S. (2001). Structural Holes Versus Network Closure as Social Capital. *Social Capital: Theory and Research*, 31–56. Retrieved from http://homes.chass.utoronto.ca/~wellman/gradnet05/burt - STRUCTURAL HOLES vs NETWORK CLOSURE.pdf

Byström, K., & Järvelin, K. (1995). Task complexity affects information seeking and use. *Information Processing & Management*, *31*(2), 191–213.

Case, D. O. (2016). *Looking for Information*. (D. O. Case & L. M. Given, Eds.) (4th ed.). Emerald Group Publishing.

Cassar, A. (2007). Coordination and cooperation in local, random and small world networks: Experimental evidence. *Games and Economic Behavior*, (December), 415–422. Retrieved from http://www.sciencedirect.com/science/article/pii/S089982560600042X

Choo, C. W., Detlor, B., & Turnbull, D. (2000). Information Seeking on the Web: An Integrated Model of Browsing and Searching. *First Monday*, *5*, 2–7. https://doi.org/10.5210/fm.v5i2.729

Coleman, J. S. (1988). Social Capital in the Creation of Human Capital. *American Journal of Sociology*, *94*(1988), S95–S120. Retrieved from http://www.jstor.org/stable/10.2307/2780243

Contractor, N. S., Zink, D., & Chan, M. (1998). IKNOW: A tool to assist and study the creation, maintenance, and dissolution of knowledge networks. In *Community computing and support systems* (pp. 201–217). Springer Berlin Heidelberg.

Cool, C. (2001). The concept of situation in information science. In *Annual review of information science and technology* (pp. 5–42).

Courtright, C. (2007). Context in Information Behavior Research. *Annual Review of Information Science and Technology*, *41*(1), 273–306.

Cropanzano, R., & Mitchell, M. S. (2005). Social Exchange Theory : An Interdisciplinary Review Social Exchange Theory : An Interdisciplinary Review. *Journal of Management*, *31*(December), 874–900. https://doi.org/10.1177/0149206305279602

Dervin, B. (1992). From the mind's eye of the user: The sense-making qualitative-quantitative methodology. *Qualitative Research in Information Management*.

DiMicco, J., Millen, D. R., Geyer, W., Dugan, C., Brownholtz, B., & Muller, M. (2008). Motivations for social networking at work. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work (CSCW)* (pp. 711–720). Chicago: ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1460674

Doerfel, M. L., & Taylor, M. (2004). Network dynamics of interorganizational cooperation: The Croatian civil society movement. *Communication Monographs*, *71*(4), 373–394. https://doi.org/10.1080/0363452042000307470

Dumais, S., Jeffries, R., Russell, D. M., Tang, D., & Teevan, J. (2014). Understanding User Behavior Through Log Data and Analysis. In *Ways of Knowing in HCI* (pp. 349–372). New York: Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4939-0378-8_14

Ellis, D. (1989). A behavioural approach to information retrieval system design. *Journal of Documentation*, *45*(3), 171–212.

Ellison, N. B., & Boyd, D. (2013). Sociality through Social Network Sites. In *The Oxford Handbook of Internet Studies*. Oxford, UK.

Evans, B. M., & Chi, E. H. (2008). Towards a model of understanding social search. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work (CSCW)*. ACM.

Evans, B. M., & Chi, E. H. (2010). An elaborated model of social search. *Information Processing & Management*, *46*(6), 656–678.

Evans, B. M., Kairam, S., & Pirolli, P. (2010). Do your friends make you smarter?: An analysis of social strategies in online information seeking. *Information Processing & Management*, *46*(6), 679–692. https://doi.org/10.1016/j.ipm.2009.12.001

Faris, R., & Felmlee, D. (2011). Status Struggles: Network Centrality and Gender Segregation in Same- and Cross-Gender Aggression. *American Sociological Review*, *76*(1), 48–73.

Foa, U. G., & Foa, E. B. (1980). Resource theory: Interpersonal behavior as exchange. In K. J. Gergen, M. S. Greenberg, & R. H.Willis (Eds.), *Social exchange: Advances in theory and research*. New York: Plenum.

Fulton, C. (2009). The pleasure principle: the power of positive affect in information seeking. *Aslib Proceedings*, *61*(3), 245–261. https://doi.org/10.1108/00012530910959808

Garton, L., Haythornthwaite, C., & Wellman, B. (1997). Studying online social networks. *Journal of Computer-Mediated Communication*, *3*(June), 1–32.

Ghosh, S., Viswanath, B., Kooti, F., Naveen, S. K., Korlam, G., Benevenuto, F., … Gummadi, K. P. (2012). Understanding and Combating Link Farming in the Twitter Social Network. In *Proceedings of the 21st International Conference on the World Wide Web (WWW)*.

Gibbs, J. L., Rozaidi, N. A., & Eisenberg, J. (2013). Overcoming the “Ideology of Openness”: Probing the Affordances of Social Media for Organizational Knowledge Sharing. *Journal of Computer-Mediated Communication*, *19*(1), 102–120. https://doi.org/10.1111/jcc4.12034

Gilbert, E., & Karahalios, K. (2009). Predicting tie strength with social media. In *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09* (pp. 211–221). New York, New York, USA: ACM Press. https://doi.org/10.1145/1518701.1518736

González-Ibáñez, R., & Shah, C. (2016). Using affective signals as implicit indicators of information relevance and information processing strategies. *Proceedings of the Association for Information Science and Technology*, *53*(1), 1–10. https://doi.org/10.1002/pra2.2016.14505301050

Gould, R. V. (1989). Power and social structure in community elites. *Social Forces*, *68*(2), 531–552.

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, *78*(6), 1360–1380.

Granovetter, M. S. (1974). *Getting a Job: A Study of Contacts and Careers*. University of Chicago Press.

Granovetter, M. S. (1983). The strength of weak ties: A network theory revisited. *Sociological Theory*, *1*, 201–233.

Gruzd, A., Wellman, B., & Takhteyev, Y. (2011). Imagining Twitter as an Imagined Community. *American Behavioral Scientist*, *55*(10), 1294–1318. https://doi.org/10.1177/0002764211409378

Hampton, K. N., & Wellman, B. (2003). Neighboring in Netville : How the Internet supports community and social capital in a wired suburb. *City & Community*, (December), 277–311.

Hanaki, N., Peterhansl, A., Dodds, P. S., & Watts, D. J. (2007). Cooperation in evolving social networks. *Management Science*, *53*(7), 1036–1050. https://doi.org/10.1287/mnsc.1060.0625

Hansen, M. T. (2002). Knowledge Networks : Explaining Effective Knowledge Sharing in Multiunit Companies. *Organization Science*, *13*(3), 232–248.

Hartel, J. (2010). Managing documents at home for serious leisure: a case study of the hobby of gourmet cooking. *Journal of Documentation*, *66*(6), 847–874.

https://doi.org/10.1108/00220411011087841

Haythornthwaite, C. (1996). Social network analysis: An approach and technique for the study of information exchange. *Library & Information Science Research*, *18*(4), 323–342.

Haythornthwaite, C. (2002). Strong, Weak, and Latent Ties and the Impact of New Media. *The Information Society*, *18*(5), 385–401. https://doi.org/10.1080/01972240290108195

Hearst, M. (2011). User interfaces for search. In R. Baeza-Yates & B. Ribeiro-Neto (Eds.), *Modern Information Retrieval: The concepts and technology behind search* (2nd ed., pp. 21–55). Addison-Wesley.

Hirsh, S., & Dinkelacker, J. (2004). Seeking information in order to produce information: An empirical study at Hewlett Packard Labs. *Journal of the American Society for Information Science and Technology*, *55*(9), 807–817.

Ho, C.-I., Lin, M.-H., & Chen, H.-M. (2012). Web users' behavioural patterns of tourism information search: From online to offline. *Tourism Management*, *33*(6), 1468–1482. https://doi.org/10.1016/j.tourman.2012.01.016

Huberman, B. A., Pirolli, P., Pitkow, J. E., & Lukose, R. M. (1998). Strong Regularities in World Wide Web Surfing. *Science*, *280*(5360), 95–97. https://doi.org/10.1126/science.280.5360.95

Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory. *Journal of Documentation*, *52*(1), 3–50.

Ingwersen, P. (1998). The calculation of Web impact factors. *Journal of Documentation*, *54*(2), 236–243. Retrieved from http://www.emeraldinsight.com/journals.htm?articleid=864074&show=abstract

Jones, W. (2007). Personal Information Management. In *Annual Review of Information Science and Technology* (Vol. 41, pp. 453–504). https://doi.org/10.1002/aris.2007.1440410117

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, *53*(1), 59–68. https://doi.org/10.1016/j.bushor.2009.09.003

Katz, N., Lazer, D., Arrow, H., & Contractor, N. S. (2004). Network theory and small groups. *Small Group Research*, *35*(3), 307–332. https://doi.org/10.1177/1046496404264941

Killworth, P. D., McCarty, C., Bernard, H. R., & House, M. (2006). The accuracy of small world chains in social networks. *Social Networks*, *28*(1), 85–96. https://doi.org/10.1016/j.socnet.2005.06.001

Kleinberg, J. (2000). The small-world phenomenon: An algorithmic perspective. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing* (pp. 163–170).

Krackhardt, D. (1992). The strength of strong ties: The importance of Philos in organizations. *Networks and Organizations: Structure, Form, and Action*, *216*(239).

Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science and Technology*, *42*(5), 361–371.

Kuhlthau, C. C. (1993). A principle of uncertainty for information seeking. *Journal of Documentation*, *49*(4), 339–355.

Lampe, C., Vitak, J., & Ellison, N. B. (2013). Users and Nonusers: Interactions between Levels of Facebook Adoption and Social Capital. In *Proceedings of the 2013 ACM conference on Computer Supported Cooperative Work (CSCW)* (pp. 809–819). ACM.

Lampe, C., Vitak, J., Gray, R., & Ellison, N. B. (2012). Perceptions of Facebook's value as an information source. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3195–3204). ACM.

Levin, D. Z., Walter, J., & Murnighan, J. K. (2011). Dormant Ties: The Value Of Reconnecting. *Organization Science*, *22*(4), 923–939. https://doi.org/10.1287/orsc.1100.0576

Lin, N. (1982). *Social resources and instrumental action*. State University of New York, Department of Sociology.

Lin, N. (1999). Building a network theory of social capital. *Connections*, *22*(1), 28–51.

Lin, N., & Dumin, M. (1986). Access to occupations through social ties. *Social Networks*, *8*(4), 365–385.

Majchrzak, A., Faraj, S., Kane, G. C., & Azad, B. (2013). The Contradictory Influence of

Social Media Affordances on Online Communal Knowledge Sharing. *Journal of Computer-Mediated Communication*, *19*(1), 38–55. https://doi.org/10.1111/jcc4.12030

Marchionini, G. (1997). *Information seeking in electronic environments*. Cambridge University Press.

Marchionini, G. (2010). The many meanings of information. In *Synthesis Lectures on Information Concepts, Retrieval, and Services* (pp. 1–9). Morgan & Claypool.

Marin, A., & Wellman, B. (2010). Social Network Analysis: An Introduction. In P. Carrington & J. Scott (Eds.), *Handbook of Social Network Analysis*. London: SAGE Publications.

Matni, Z., & Shah, C. (2014). For the Love of Information: Motivations and Affective Dynamics of Surfing the Web for Pleasure. In *Proceedings of the 77th ASIS&T Annual Meeting*. Seattle: ASIS&T.

McLure Wasko, M., & Faraj, S. (2005). Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice. *MIS Quarterly*, *29*(1), 35–57.

Meeker, B. F. (1971). Decisions and exchange. *American Sociological Review*, (36), 485–495.

Milgram, S. (1967). The Small World Problem. *Psychology Today*, *1*(1), 61–67.

Mishra, J. L., Allen, D. K., & Pearman, A. D. (2013). Information use, support and decision making in complex, uncertain environments. *Proceedings of the American Society for Information Science and Technology*, *50*(1), 1–10. https://doi.org/10.1002/meet.14505001045

Monge, P. R., & Contractor, N. S. (2003). *Theories of communication networks*. New York: Oxford University Press.

Morris, M. R., Teevan, J., & Panovich, K. (2010a). A Comparison of Information Seeking Using Search Engines and Social Networks. In *ICWSM* (pp. 2–5).

Morris, M. R., Teevan, J., & Panovich, K. (2010b). What do people ask their social networks, and why? A survey study of status message Q&A behavior. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1739–1748).

Naaman, M., Boase, J., & Lai, C. (2010). Is it Really About Me ? Message Content in Social Awareness Streams. In *CSCW '10: Proceedings of the 2010 ACM Conference of Computer Supported Cooperative Work* (pp. 189–192). ACM.

Nicolaisen, J. (1981). Citation analysis. *Annual Review of Information Science and Technology*, *41*(1), 609–641. Retrieved from https://www.ideals.illinois.edu/bitstream/handle/2142/7189/librarytrendsv30i1_opt.pdf?sequence=3#page=88

Panovich, K., Miller, R. C., & Karger, D. R. (2012). Tie Strength in Question & Answer on Social Network Sites. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work - CSCW '12* (pp. 1057–1066). ACM.

Park, H. W., & Thelwall, M. (2003). Hyperlink Analyses of the World Wide Web: A Review. *Journal of Computer-Mediated Communication*, *8*(July), 1–33.

Pathak, N., Mane, S., Srivastava, J., & Contractor, N. S. (2006). Knowledge perception analysis in a social network. In *Proceedings of the 6th SIAM–International Conference on Data Mining (SDM 06)* (pp. 1–10). Retrieved from http://www.siam.org/meetings/sdm06/workproceed/Link Analysis/10.pdf

Pirolli, P., & Card, S. (1995). Information foraging in information access environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 51–58). ACM.

R Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Rainie, L., & Wellman, B. (2012). *Networked: The new social operating system*. MIT Press. Retrieved from http://www.ceeol.com/aspx/getdocument.aspx?logid=5&id=db571d60-7bfa-4ca2-9cb7-30a490953d0f

Robinson, M. A. (2010). An Empirical Analysis of Engineers' Information Behaviors. *Journal of the American Society for Information Science and Technology*, *61*(4), 640–658. https://doi.org/10.1002/asi

Saracevic, T., & Kantor, P. B. (1988a). A Study of Information Seeking and Retrieving. II. Users, Questions, and Effectiveness. *Journal of the American Society for*

Information Science, *39*(3), 177–196.

Saracevic, T., & Kantor, P. B. (1988b). A Study of Information Seeking and Retrieving. III. Searchers, Searches, and Overlap. *Journal of the American Society for Information Science*, *39*(3), 197–216. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.87.5184&rep=rep1&type=pdf

Savolainen, R. (1995). Everyday life information seeking: Approaching information seeking in the context of "way of life." *Library & Information Science Research*, *17*, 259–294.

Savolainen, R., & Kari, J. (2004). Conceptions of the Internet in Everyday Life Information Seeking. *Journal of Information Science*, *30*(3), 219–226. https://doi.org/10.1177/0165551504044667

Schutz, A., & Luckmann, T. (1973). *The structures of the life-world*. Evanston, IL: Northwestern University Press.

Shah, C., & González□Ibáñez, R. (2010). Exploring Information Seeking Processes in Collaborative Search Tasks. In *Proceedings of the American Society for Information Science and Technology*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/meet.14504701211/full

Shah, C., & Kitzie, V. (2012). Social Q & A and Virtual Reference — Comparing Apples and Oranges with the Help of Experts and Users. *Journal of the American Society for Information Science and Technology*, *63*(10), 2020–2036.

https://doi.org/10.1002/asi

Shah, C., & Pomerantz, J. (2010). Evaluating and predicting answer quality in community QA. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval* (pp. 411–418).

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Chicago: Illinois Press.

Simmel, G. (2011). *On Individuality and Social Forms*. (D. N. Levine, Ed.). Chicago: University of Chicago Press.

Suri, S., & Watts, D. J. (2011). Cooperation and contagion in web-based, networked public goods experiments. *PLoS ONE*, *6*(3). https://doi.org/10.1371/journal.pone.0016836

Taylor, R. S. (1968). Question negotiation and information seeking in libraries. *College and Research Libraries*, *28*, 178–194.

Teevan, J., Alvarado, C., Ackerman, M. S., & Karger, D. R. (2004). The perfect search engine is not enough: A study of orienteering behavior in directed search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vol. 6, pp. 415–422).

Thelwall, M., Vaughan, L., & Björneborn, L. (2005). Webometrics. *Annual Review of Information Science and Technology*, *39*(1), 81–135.

Treem, J. W., & Leonardi, P. M. (2012). Social Media Use in Organizations: Exploring the Affordances of Visibility, Editability, Persistence, and Association. *Communication Yearbook*, *36*.

Tversky, A., & Kahneman, D. (1992). Advances in Prospect-Theory - Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323. https://doi.org/Doi 10.1007/Bf00122574

Vakkari, P. (2001). A theory of the task-based information retrieval process: a summary and generalisation of a longitudinal study. *Journal of Documentation*, *57*(1), 44–60.

Vrabel, J., & Zeigler-Hill, V. (2017). Conscious vs. Unconscious Determinants of Behavior. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of Personality and Individual Differences* (pp. 1–4). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-28099-8_1124-1

Wang, J., Suri, S., & Watts, D. J. (2012). Cooperation and assortativity with dynamic partner updating. *Proceedings of the National Academy of Sciences*, *109*(36), 14363–14368. https://doi.org/10.1073/pnas.1120867109

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge, MA: Cambridge University Press.

Watts, D. J. (2003). *Six degrees: The science of a connected world*. New York: WW Norton & Company.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of “small-world” networks. *Nature*, *393*(6684), 440–442. https://doi.org/10.1038/30918

Weber, M. S., & Monge, P. R. (2013). *Industries in Turmoil: Driving Transformation during Periods of Disruption*.

Weiler, A. (2005). Information-Seeking Behavior in Generation Y Students: Motivation, Critical Thinking, and Learning Theory. *The Journal of Academic Librarianship*, *31*(1), 46–53.

Wellman, B., & Wortley, S. (1990). Different Strokes from Different Folks: Community ties and social support. *American Journal of Sociology*, *9*(3), 558–588.

Wilson, T. D. (1981). On user studies and information needs. *Journal of Documentation*, *37*(1), 3–15.

Wilson, T. D. (1999). Models in information behavior research. *Journal of Documentation*, *55*(3), 249–270.

Wilson, T. D., Ford, N., Ellis, D., Foster, A., & Spink, A. (2002). Information seeking and mediated searching: Part 2. Uncertainty and its correlates. *Journal of the American Society for Information Science and Technology*, *53*(9), 704–715. https://doi.org/10.1002/asi.10082

Zhou, D., Lawless, S., & Wade, V. (2012). Improving search via personalized query expansion using social media. *Information Retrieval*, *15*(3–4), 218–242. https://doi.org/10.1007/s10791-012-9191-2

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. New York: Addison-Wesley.

Appendix 1: IRB Material

IRB was approved by Rutgers University per IRB Protocol Number 15-766 and by University of California, Santa Barbara per Protocol Number 1-17-0460.

Appendix 2: Questions and Answers for the Exercises

There are 4 types of scenarios/exercises. Each exercise nominally has 10 participants who have to find answers to a set of 15 questions. Each question will have 3 UNIQUE answers that the participants must find. These are the questions and their answers.

Set 1

Q#	Topical Category	Question	Answer 1	Answer 2	Answer 3
Q01	Animals	Name 3 wild animals only found in Africa	Gorilla	Zebra	Giraffe
Q02	Animals	Name 3 wild animals only found in Eurasia	Panda	Tiger	Chamois
Q03	Animals	Name 3 wild animals only found in Australia	Koala	Kangaroo	Wallaby
Q04	Cars	Name 3 brands of cars	Toyota	Chevrolet	Fiat
Q05	Arts	Name 3 Disney cartoon characters	Mickey	Minnie	Nemo
Q06	Literature	Name 3 authors (last name) of American literature	Poe	Hawthorne	Twain
Q07	Arts	Name 3 famous (last name) American film directors	Tarantino	Spielberg	Lucas
Q08	Arts	Name 3 famous (last name) American film actors	Clooney	Hepburn	Streep
Q09	Food	Name 3 ingredients you might need to make a sandwich	Bread	Cheese	Tomato
Q10	Food	Name 3 ingredients you might need to bake a cake	Flour	Sugar	Eggs
Q11	Geography	Name 3 capital cities of countries in the Americas	Brazilia	Ottawa	Caracas
Q12	Geography	Name 3 capital cities of countries in Europe	Paris	Rome	Budapest
Q13	Geography	Name 3 capital cities of countries in Asia	Tokyo	Bangkok	Beirut
Q14	Arts	Name 3 famous rock bands	Beatles	Pearl Jam	Maroon 5
Q15	Computers	Name 3 computer brands	Apple	Dell	Lenovo

Set 2

Q#	Topical Category	Question	Answer 1	Answer 2	Answer 3
Q01	Sports	Name 3 US baseball teams	Cardinals	Mets	Angels
Q02	Sports	Name 3 US football teams	Patriots	49ers	Seahawks
Q03	Homes	Name 3 kitchen appliances	Dishwasher	Freezer	Oven
Q04	Geography	Name 3 California cities	Los Angeles	San Diego	San Francisco
Q05	Geography	Name 3 Indian cities	Mumbai	New Delhi	Hyderabad
Q06	People	Name 3 famous psychologists	Jung	Freud	Skinner
Q07	Arts	Name 3 American films from the 1990s	Pulp Fiction	Star Wars: Episode 1	American Pie
Q08	Nature	Name 3 types of flowers	Lily	Rose	Sunflower
Q09	Cars	Name 3 brands of cars	Honda	Dodge	BMW
Q10	Food	Name 3 ingredients you might need to cook an omelet	Eggs	Cheese	Onions
Q11	Geography	Name 3 countries in Africa	Egypt	Nigeria	Kenya
Q12	Geography	Name 3 countries in Europe	Romania	Sweden	Greece
Q13	Geography	Name 3 countries in Asia	Nepal	Mongolia	Iran
Q14	Arts	Name 3 famous Western classical music composers	Bach	Beethoven	Mozart
Q15	Computers	Name 3 pieces of computer equipment	Mouse	Keyboard	CPU

Set 3

Q#	Topical Category	Question	Answer 1	Answer 2	Answer 3
Q01	Colors	Name 3 colors	Blue	Yellow	Orange
Q02	Food	Name 3 kinds of fruit	Banana	Apple	Strawberry
Q03	Food	Name 3 common ice cream flavors	Chocolate	Vanilla	Pistachio
Q04	Arts	Name 3 famous female singers from the US (last name)	Swift	Houston	Carey
Q05	Animals	Name 3 animals that are kept as pets	Cat	Dog	Mouse
Q06	Travel	Name 3 international airlines	United	Singapore	Alitalia
Q07	Geography	Name 3 Canadian provinces	Quebec	Ontario	Manitoba
Q08	Cars	Name 3 brands of cars	Nissan	Renault	Ford
Q09	Food	Name 3 Japanese dishes	Sushi	Yakisoba	Tempura
Q10	Leisure	Name 3 things you might take to the beach	Towel	Swimming suit	Umbrella
Q11	Food	Name 3 countries that produce wine	France	Italy	USA
Q12	Colors	Name 3 shades of red	Burgundy	Cherry	Scarlet
Q13	Arts	Name 3 TV shows (American)	That 70s Show	House	Friends
Q14	Arts	Name 3 famous rock bands	Beatles	Pearl Jam	Maroon 5
Q15	Geography	Name 3 countries in South America	Paraguay	Bolivia	Colombia

Set 4

Q#	Topical Category	Question	Answer 1	Answer 2	Answer 3
Q01	Cars	Name 3 brands of cars	Suzuki	Rolls Royce	Lada
Q02	Food	Name 3 foods that have added sugar in them	Chocolate	Candy	Soda
Q03	Animals	Name 3 animals found in almost all continents	Dogs	Rats	Cats
Q04	Sports	Name 3 sports regularly played in the Olympics	Swimming	Field Hockey	Rugby
Q05	Literature	Name 3 authors of horror books (last name)	King	Poe	Lovecraft
Q06	Arts	Name 3 famous Western painters	Da Vinci	Picasso	Whistler
Q07	Geography	Name 3 famous lakes	Victoria	Michigan	Titicaca
Q08	Geography	Name 3 famous rivers	Nile	Amazon	Indus
Q09	Leisure	Name 3 things you might take skiing	Skis	Poles	Hat
Q10	Arts	Name 3 famous Western operas	Tosca	Carmen	Don Giovanni
Q11	Food	Name 3 things you might put in your coffee	Milk	Sugar	Water
Q12	Sports	Name 3 sports that require a ball	Tennis	Football	Volleyball
Q13	Insects	Name 3 types of beetles	Ladybug	Cockroach	Weevil
Q14	Computers	Name 3 brands of computers	Asus	Sony	HP
Q15	Computers	Name 3 computer operating systems	Windows	Android	Linux

Appendix 3: Surveys

A3.1 Pre-survey

The pre-survey is designed to collect information on the participants' habits and familiarity with social media, in general, and online social networks in particular. The survey also seeks to get information on participants' habits of (or lack thereof) asking questions to people in online social network settings. The survey is also designed to collect general demographic data from the participants.

Introduction:

Pre-survey

Hello, and thank you for participating in this academic survey. We are investigating how people usually use social media in general, and in particular online social networks.

Please answer honestly and feel free to answer in whichever way feels right to you. The answers you provide are strictly confidential.

Questions:

A. YOUR SOCIAL MEDIA USE HABITS

1. How regularly do you access ***any type*** of social media?
    a. More than once a day
    b. Once a day
    c. A few times a week
    d. Once a week
    e. Less than once a week
    f. Not at all

2. Which **social media sites** do you use and how often do you use them?

SITE	More than once a day	Once a day	Once to a few times a week	Less than once a week	I do not use this SM site
Facebook					
Twitter					
Reddit					
Tumblr					
YouTube					
Instagram					
Pinterest					
Snapchat					
Yahoo Answers					
Quora					
Other:________					
Other:________					
Other:________					

3. Think about the last time you checked into a social media **site for the purpose of asking someone a question** (any type of question). Please briefly describe the following:
    a. If you have **never** asked a question of someone on social media, please check this box and ignore the rest of question 3: ☐
    b. What instigated that session?
    c. What website(s) did you visit?
    d. How long did that visit last (an approximation is fine)?
    e. What question did you ask (or what was the nature of the question you asked)?

4. Still thinking about the last time you checked into a social media site **for the purpose of asking someone a question** (any type of question). Who did you ask your question to (circle the appropriate letter)?
   a. A family member
   b. A friend
   c. An acquaintance
   d. A stranger
   e. I have never asked a question of someone on social media.

5. Still thinking about the last time you checked into a social media site **for the purpose of asking someone a question** (any type of question). Which one of these statements **best describes the person you asked a question to** (more than one answer is acceptable)?
   a. He/she is an expert on the subject of my question
   b. He/she is not an expert on the subject of my question
   c. He/she is a trustworthy person
   d. He/she was available to answer my question
   e. None of the above statements.
   f. I have never asked a question of someone on social media.

6. Which of these online venues do you use **to look for information** and **how often**?

SITE	**Often**	**Sometimes**	**Never**
Google			
Yahoo			
Bing			
Wikipedia			
Facebook			
Twitter			
Instagram			
Reddit			
Tumblr			
Pinterest			
YouTube			
Snapchat			
Yahoo Answers			
Quora			
Other: __________			

Other: __________			

B. CLOSE-ENDED QUESTIONS

7. How old are you?

8. What is your gender?

9. What is the highest educational degree that you have obtained? (choose one)
    a. Some high school
    b. High school degree
    c. Post-high school and currently pursuing a university degree
    d. University Bachelor's degree
    e. University graduate degree (Masters or higher)

10. Which of these statements is true for you? (choose all that apply)
    a. I own or have anytime access to a desktop computer
    b. I own or have anytime access to laptop computer
    c. I own or have anytime access to tablet computer (e.g. iPad)
    d. I own or have anytime access to a smart, Internet capable, phone (e.g. iPhone)
    e. None of the above applies to me

11. What country have you lived in for the majority of your life?

STOP RIGHT HERE! ☺

DO NOT CONTINUE THE SURVEY UNTIL AFTER YOU FINISH THE ONLINE EXERCISES FOR SIMPSON!

AWAIT FURTHER INSTRUCTIONS!

A3.2 Post-survey

The post-survey is designed to collect information on the participants' user experience with SIMPSON.

Post-survey

Hello, and thank you for participating in this academic survey. We want to know about your experiences with the exercises you ran on SIMPSON and your experiences with the social networks of the other participants that you interacted with.

Please answer honestly and feel free to answer in whichever way feels right to you. The answers you provide are *strictly confidential*.

Questions:

1. Were the instructions on how to connect to SIMPSON clear to you before you started?

   YES **NO**

2. Were the instructions how the compensation worked clear to you before you started?

   YES **NO**

3. Were the instructions on the main landing page clear in regards to how you should connect with others, how you can reject connection offers, and ask for questions?

   YES **NO**

4. Once you started the exercise, there was a score and a position number displayed on the top right. Were they clear to you what they were about?

   YES **NO**

5. Were the score and the position number helpful to you as you went through the exercises?

   YES **NO**

6. Did you actively refuse connections with someone?

   YES **NO**

7. Explain why or why not, regarding your answer to the above question #6 (short description).

8. How difficult was it to collect the answers that you did collect?
    a. Difficult
    b. Neither difficult, nor easy
    c. Easy

9. Did you run out of score points at any time?

   YES **NO**

10. If you answered Yes to #9, explain why you think this happened (short description).

11. What difference did it make for you when conducting the exercises if you knew something about what topics the others knew?
    a. It made it *much* easier to reach my goals and conclude the exercise.
    b. It made it *somewhat* easier to reach my goals and conclude the exercise.
    c. It made no difference.
    d. It made it *somewhat* more difficult to reach my goals and conclude the exercise.
    e. It made it *much* more difficult to reach my goals and conclude the exercise.

12. In your exercise, you saw 2 different types of networks that connected you with other participants. Which network type made reaching your goals easier or quicker?

    a. First run
    b. Second run
    c. Neither one (they had the same effect)
    d. I’m not sure

13. By the last exercise, do you think you had developed a strategy to help maximize your score? Please explain (short description)

14. Please describe your general (or any specific) observations, comments, thoughts, etc. regarding your experience with these exercises.

Appendix 4: Statistical Analysis (R Scripts Used)

A4.1 Effectiveness Testing (H1)

```
library(psych)

# Define data table and variables from source file
df <- read.table("effectiveness_R.csv", header = TRUE, sep = ",")
NetworkType <- df$Net
Uncertainty <- df$Unc
Effectiveness <- df$Raw.Effectiveness

# Define linear model for Effectiveness as DV and show regression
analysis
results <- lm(Effectiveness ~ NetworkType * Uncertainty)
summary(results)

# ANOVA for linear model
anova(results)

# Polyserial Correlation
polyserial(Effectiveness, Uncertainty, std.err=T)
polyserial(Effectiveness, NetworkType, std.err=T)

# Mediating/Moderating Analysis
# Net2 and Unc2 are NetworkType and Uncertainty variables with
values transposed to 1 as 0
mod <- mediate(Raw.Effectiveness~(Net2)+Unc2, Unc2, data = dfH1)
mod

# Create relevant plots
plot(density(Effectiveness))
plot(Uncertainty, Effectiveness)
plot(NetworkType, Effectiveness)
interaction.plot(NetworkType, Uncertainty, Effectiveness)
```

A4.2 Step Analysis (H2)

```
# Define data table and variables from source file
df <- read.table("steps_R.csv", header = TRUE, sep = ",")
Steps <-df$steps
Uncertainty<-df$unc

# Define linear model for Effectiveness as DV and show regression
analysis
results <- lm(Steps ~ Uncertainty)
summary(results)

# ANOVA for linear model
anova(results)

# Post-Hoc Tesing (Bonferroni and TukeyHSD)
a1 <- aov(Steps ~ Uncertainty)
pairwise.t.test(Accessed, Degree, p.adj="bonf")
TukeyHSD(a1)

# Polyserial Correlation
polyserial(Steps, Uncertainty, std.err=T)

# Create relevant plots
plot(density(Steps))
plot(Uncertainty, Steps)
```

A4.3 Access Analysis (H3)

```
# Define data table and variables from source file
df <- read.table("access_R.csv", header = TRUE, sep = ",")
Accessed <- df$total_accessed
Uncertainty <- df$unc
Degree <-df$degree
NodeType <- df$NodeType

# Pearson Correlation with significance testing
cor.test(Accessed,Degree)

# Define linear model for Effectiveness as DV and show regression
analysis and ANOVA - Run 1
LinearMod <- lm(Accessed ~ NodeType * Uncertainty)
summary(LinearMod)
anova(LinearMod)

# Define linear model for Effectiveness as DV and show regression
analysis and ANOVA - Run 2
LinearMod <- lm(Accessed ~ Degree * Uncertainty)
summary(LinearMod)
anova(LinearMod)

# Non-Linear (Polynomial) Modeling
Degree2 <- Degree^2
Degree3 <- Degree^3

NLM_Quad <- lm(Accessed ~ Degree2 + Degree * Uncertainty)
summary(NLM_Quad)

NLM_Cube <- lm(Accessed ~ Degree3 + Degree2 + Degree *
Uncertainty)
summary(NLM_Cube)

# Create relevant plots
plot(density(Accessed))
plot(Uncertainty, Accessed)
plot(NodeType, Accessed)
plot(Degree, Accessed)
plot(as.factor(Degree), Accessed)
interaction.plot(NodeType, Uncertainty, Accessed)
interaction.plot(Degree, Uncertainty, Accessed)
```