August 2018

Online Social Networks' Investigations of Individuals' Healthy and Unhealthy Lifestyle Behaviors and Social Factors Influencing Them —Three Essays

Mahyar Sharif Vaghefi
University of Wisconsin-Milwaukee

ONLINE SOCIAL NETWORKS' INVESTIGATIONS OF INDIVIDUALS' HEALTHY AND UNHEALTHY LIFESTYLE BEHAVIORS AND SOCIAL FACTORS INFLUENCING THEM —THREE ESSAYS

by

Mahyar Sharif Vaghefi

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Management Science

at

The University of Wisconsin-Milwaukee

August 2018

ABSTRACT

ONLINE SOCIAL NETWORKS' INVESTIGATIONS OF INDIVIDUALS' HEALTHY AND UNHEALTHY LIFESTYLE BEHAVIORS AND SOCIAL FACTORS INFLUENCING THEM —THREE ESSAYS

by

Mahyar Sharif Vaghefi

The University of Wisconsin Milwaukee, 2018
Under the Supervision of Professor Fatemeh (Mariam) Zahedi

More than half of U.S. adults suffer from one or more chronic diseases, which account for 86% of total U.S. healthcare costs. Major contributors to chronic diseases are unhealthy lifestyle behaviors, which include lack of physical activity, poor nutrition, tobacco use, and drinking too much alcohol. A reduction in the prevalence of health-risk behaviors could improve individuals' longevity and quality of life and may halt the exponential growth of healthcare costs. Prior studies in the field have acknowledged that a comprehensive understanding of health behaviors requires the examination of individual' behaviors in supra-dyadic social networks. In recent years, the growth of online social networks and popularity of location-based services have opened new research opportunities for observational studies on individuals' healthy and unhealthy lifestyle behaviors. The goal of this three-essay dissertation is to examine the effect of various social factors, shared images, and communities of interest on healthy and unhealthy lifestyle behaviors of individuals. This dissertation makes novel contributions in terms of theoretical implications, data collection and analysis methods, and policy implications for promoting healthy lifestyle behaviors and inhibiting unhealthy behaviors.

Essay 1 draws on a synthesis of social cognitive and social network theories to conceptualize a causal model for healthy and unhealthy behaviors. To test the conceptualized model, we developed a new method—dynamic sequential data extraction and integration—to collect and integrate data over time from Twitter and Foursquare. The captured dataset was then combined with relevant data from the U.S. Census Bureau. The final dataset has more than 32,000 individuals from all states in the United States. Using this dataset, we derived variables to measure healthy and unhealthy lifestyle behaviors and metrics for factors representing individuals' social support, social influence, and homophily, as well as the socioeconomic status of the communities where they live. To capture the impacts of social factors, we collected individuals' behaviors in two separate time periods. We used zero-inflated negative binomial regression method for data analysis. The results of this study uncover factors that have significant impacts on healthy and unhealthy lifestyle behaviors.

Essay 2 focuses on embedded images in self-disclosed posts related to healthy and unhealthy lifestyle behaviors. While online photo-sharing has become widely popular, and neuroscience has reported the influence of images in brain activities, to our knowledge, there is no published research on the impacts of shared photos on health-related lifestyle behaviors. This study addresses this gap and examines the moderating role of shared images and the direct impacts of their contents. We relied on social learning and multimodality theories to argue that images can attract individuals' attention and enhance the process of observational learning in online social networks. We developed a novel method for image analysis that involves the extraction, processing, dimensionality reduction, and categorization of images. The results show that the presence of photos in self-disclosed unhealthy lifestyle behaviors positively moderates friends' social influence. Moreover, the results indicate that the contents of shared photos influence individuals' health-related behaviors.

Essay 3 focuses on the role of personal interests in individuals' health-related lifestyle behaviors. Prior studies have demonstrated that health promotional programs can benefit from targeting individuals based on their interests. Specifically, prior studies have emphasized the role of interests as a factor influencing behaviors. However, current literature suffers from two major gaps. First, there is no systematic and comprehensive approach to capture individuals' interests in online social networks. Second, to our knowledge, the role of interests in individuals' healthy and unhealthy lifestyle behaviors as disclosed online has not been investigated. To address these gaps, we examine the role of individuals' interests in their health-related behaviors. The theoretical foundation of this study is a synthesis of homophily and self-determination theories. We developed a novel method—the homophily-based interest detection method—that involves network simplification, network clustering, cluster labeling, and interest metrics. This method was applied to social networks of individuals in Essay 1 to measure individuals' interests. The results show that health-related interests are associated with individuals' healthy and unhealthy lifestyle behaviors. Our findings indicate that other forms of interest, such as music taste and political views, also play a role. Moreover, our results show that belonging to healthy (unhealthy) communities of interest has an inhibitive role that prevents postings of unhealthy (healthy) behaviors.

To

my parents,

my wife,

my brother,

and especially my daughter

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Dr. Fatemeh (Mariam) Zahedi, my professor and research advisor. She patiently guided me throughout my graduate career. She has made a remarkable difference in my life as she has helped me see that a career in academia could be my future reality. I absolutely could not have completed this dissertation without her. I would also like to extend my thanks to my committee members: Dr. Sanjoy Ghose, Dr. Derek Nazareth, and Dr. Huimin Zhao. I sincerely appreciate their invaluable feedbacks and insights during my dissertation and the seminar courses I took with them.

I want to express my genuine appreciation and thanks to Mr. and Mrs. Lubar for their generosity in providing financial support for my research. I would like to extend my thanks to all the faculty and staff at Lubar School of Business for supporting my work. Especially, I want to thank Dr. Patrick Doran for his continued support, Dawn Koerten for all the support and guidance that she has given me, Keith Kunkel for providing technical support for my research.

Last but not least, I wish to express my gratitude to my lovely wife Neshat Beheshti for her unwavering support and understanding, to my brother Mahrad for his constructive feedbacks on my research, and to my parents who have always supported me and been there for me when I needed them.

CHAPTER 1

Introduction

In recent years, chronic diseases are increasingly becoming prevalent in different countries. Individuals' unhealthy lifestyle behaviors including lack of physical activities, poor nutrition, tobacco use, and drinking too much alcohol have been considered as the major causes of such diseases (CDC 2015). Thus, a reduction in the prevalence of health-risk behaviors could improve individuals' longevity and considerably reduce the cost burdens on health care systems. Developing a comprehensive understanding of social factors contributing to individual' healthy and unhealthy lifestyle behaviors is a big step towards controlling unhealthy lifestyle behaviors and promoting healthy lifestyle behaviors. Individuals' interests and preferences, and friends' social support and social influence are factors that can be highly influential in formation of lifestyle behaviors. This dissertation plans to study the effect of these social factors within online social networks.

In 2015, it was estimated that people spend an average of 1.7 hours daily on online social networks.[1] This time was reported to be 9 hours for teens.[2] Such pervasive reliance on online social networks, particularly for the younger generation, calls for a deeper understanding of how online social factors influence individuals' health-related lifestyle choices. In our studies, we rely

---

[1] http://www.globalwebindex.net
[2] http://www.cnn.com/2015/11/03/health/teens-tweens-media-screen-use-report/

on individuals location-based check-ins in online social networks. It was argued that the shared-location of individuals can represent individuals' type of activities (Cramer et al. 201). Thus, we consider the check-ins of individuals at health-related venues (gym and fitness center, bar, and fast food restaurants) as the proxy for capturing their healthy and unhealthy lifestyle behaviors. Throughout the chapters, we develop several tools, frameworks, and models to capture the effect of social factors on individuals' health-related lifestyle behaviors as observed in online social networks.

This dissertation advances the theory and techniques for analyzing individuals' behaviors as observed in online social networks. It consists of three research essays. The first essay offers a dynamic sequential approach for capturing, extracting and integrating data from online social networks and introduces the Health-related Lifestyle Behavior (HLB) model for analyzing individuals' health-related lifestyle behaviors as observed in online social networks. The second essay expands the introduced HLB model offered in the first essay by considering the role of friends' posted images – from healthy and unhealthy places – in formation of individuals' health-related lifestyle behaviors. The third essay offers a theory-based method for detection of individuals' interests and preferences from online social networks. Our analysis shows that detected interests and preferences can explain individuals' disclosed healthy and unhealthy lifestyle behaviors in online social networks.

Essay 1: Impact of Social Factors on People's Health-related Lifestyle Behaviors: A National Observational Study in Online Social Networks.

Prior studies have captured the influential power of online social networks in formation of political mobilization (Bond et al. 2012), and adoption of paid services (Bapna and Umyarov

2015). However, no observational studies have examined the role of online social networks in formation of health-related lifestyle behaviors. Considering the huge imposed cost of unhealthy lifestyle behaviors to healthcare systems, examining the role of online social networks in changing health-related lifestyle behaviors could demonstrate the power of online social networks and help in formulating policies to harness this power to promote healthy lifestyle behaviors. Therefore, we pose the following research questions: (i) How can we observe individuals' health-related lifestyle behaviors on online social networks? (ii) What are the online social factors that contribute to individuals' health-related lifestyle as observed on online social networks?

To answer these questions, we develop a new dynamic sequential approach to collect data from public online social networks and integrate that with data from U.S. census bureau. We also rely on Berkman framework (Berkman et al. 2000) to build a contextualized model for studying the individuals' health-related lifestyle behaviors within online social networks. Our findings show that individuals' health-related lifestyle behaviors are significantly influenced by their friends' behaviors. This study contributes to both theory and practice and provides great insights for health practitioners and policy makers.

Essay 2: The Moderating Impact of Friends' Posted Images on Observed Healthy and Unhealthy Lifestyle Behaviors of Individuals in Online Social Networks

Shared images in online social networks consists of personal recommendations that make them influential (Eftekhar et al. 2014). A recent experimental study shows that adolescents tend to post pictures of unhealthy foods in their online social network pages (Holmberg et al. 2016). This raises the concerns about the effects of shared photos in online social networks on individuals' health-related lifestyle behaviors. Thus, we build on top of the HLB model to investigate about the role of posted images on individuals' healthy and unhealthy lifestyle behaviors. In this study, we seek to answer the following research questions: (i) Does the presence of photos moderate impact of friends' healthy and unhealthy lifestyle behaviors? (ii) Do contents of posted photos contribute to friends' healthy and unhealthy lifestyle behaviors?

In answering these research questions, we collect the images posted along with individuals' healthy and unhealthy lifestyle behaviors and analyze them through a novel image processing framework. The results of this study show that presence of images along with posted unhealthy lifestyle behaviors can increase the social influence of reported unhealthy lifestyle behaviors within online social networks. Our findings also indicate that the content of images can be influential on observed individuals' behaviors.

Essay 3: Communities of Interest in Online Social Networks: Detection Method and its Application in Explaining Self-Disclosed Lifestyle Behaviors

Research has demonstrated that online social network platforms can be used for health promotional purposes (Valle et al. 2013, Pechmann et al. 2015, Ramo et al. 2015)**.** There are programs to promote healthy lifestyle behaviors and prevent diseases, disability, and premature death. Examples are VERB and TRUTH—programs by non-profit organizations to increase the level of physical activities and reduce smoking among adolescents (Huhman et al. 2004, Evans 2006). For such programs to succeed, there is a need identify individuals' interests, preferences, and values. Online social networks have become valuable sources to gain understanding of individuals' behaviors through their online social environments. This requires discovering their communities of interest and investigating the role of such communities in individuals' health-related lifestyles. Thus, we pose the following research questions (i) how individuals' interests and preferences can be detected within online social networks? and, (ii) how self-disclosed health-related lifestyle behaviors of individuals in online social networks are associated with their observed interests and preferences?

In answering our research questions, we develop a Homophily-based Interest Detection (HID) method that rely on the structure of individuals' social network for detection on their interest-based attributes. Our method could detect wide variety of individuals' interests such as their music taste, and political view. Then, we use the detected interests of individuals to investigate about the association between individuals' healthy and unhealthy lifestyle behaviors and their observed health-related interests. Our findings show that type of individuals' interests and existing norms within communities of interest can explain the observed individuals' lifestyle behaviors.

CHAPTER 2

Essay 1 - Impact of Social Factors on Peoples' Health-related Lifestyle Behaviors:
A National Observational Study in Online Social Networks

2.1. Introduction

More than half of adults in the U.S. suffer from one or more chronic diseases, which account for 86% of total healthcare costs (CDC 2015). Major contributors to chronic diseases are unhealthy lifestyle behaviors that include the "lack of exercise or physical activity, poor nutrition, tobacco use, and drinking too much alcohol" (CDC 2015). A reduction in the prevalence of health-risk behaviors could improve individuals' longevity and quality of life and may halt the exponential growth of healthcare costs (CDC 2015). Moreover, an increase in healthy lifestyle behaviors reduces the substantial economic burdens associated with chronic diseases (Scarborough et al. 2011). For example, compared with medical costs of normal-weight people, obese people pay 42% more to deal with their medical issues (Finkelstein et al. 2009).[3]

Healthy behavior is an ongoing process that has multiple social and personal dimensions. McNeill et al. (2006) have found that advising individuals to have physical activities without considering social norms and environmental factors is unlikely to lead to behavioral changes.

[3] A 2008 estimate shows that the medical cost for obesity has increased to $147 billion per year, comprising about 10 percent of all medical spending (Finkelstein et al. 2009).

Literature reports the influence of peer groups in studies of specific behavior such as smoking, substance abuse, drinking, or physical activity (Eisenberg et al. 2005, Trogdon et al. 2008, Chen et al. 2001, Kaplan et al. 2001, Clark and Lohéac 2007, Leung et al. 2014, Andrews et al. 2002, Lundborg 2006). However, such studies have limitations. First, offline health-related studies have relied mostly on relationships of people in schools or family settings. Such settings have geographical, cultural and political limitations in terms of development of relationships with others, thus limiting observations of lifestyle behaviors to a few number of friends in a given period and not accounting for individuals' entire social networks. By studying individuals' online social networks, we address this limitation. With pervasive use of mobile technologies and online social networks, the structure and sphere of human relationships have expanded. Online social networks allow people to observe and follow all friends' lifestyle behaviors daily, even hourly, or in some cases in real time, particularly in location-based social networks. This facilitates the process of observational learning for individuals. Thus, studying health-related lifestyle behaviors in online social networks allows us to account for the broader social forces influencing individuals' health-related lifestyle choices.

Second, another limitation is the focus on dyadic relations, thus ignoring the compounding effects of social influence emanating from different types of relationships and multiple groups to which an individual may belong. Research has already acknowledged the need for using supra-dyadic social networks to examine how health-related lifestyle behaviors can spread across social networks (Smith and Christakis 2008). The focus on online social networks provides a suitable lens that accounts for the spread of health-related lifestyle behaviors across social networks.

Third, samples in prior studies are limited to one small segment of a population and focused on a specific disorder or activity, limiting the generalizability of the findings to the entire population. We address this limitation by collecting representative samples from across the United States to investigate multiple health-related lifestyle behaviors.

Fourth, social networks affect individuals' health related life-style behaviors through multiple factors, including social influence, social support, and socioeconomic status. Prior studies have focused on a single factor, thus limiting the generalizability of results. We address this limitation by relying on the framework developed by Berkman et al. (2000) to develop a model that captures multiple pathways through which online social networks influence individuals' health-related lifestyle behaviors.

Fifth, data collection in the prior studies relies on participants' self-reported behaviors. Although valuable in understanding individuals' perceptions, self-reported data in this context could be biased due to the subjects' under-reporting unhealthy behaviors or over-reporting healthy behaviors. Moreover, the Hawthorne effect could also introduce bias in self-reported data (Adair 1984). The Hawthorne study has shown that participants' awareness of being the subject of research could modify and influence their behaviors. The extant literature has shown that the Hawthorne effect exists in self-reported and directly observed experimental data, leading to a call for a new approach to data collection in health-related behavioral research (McCambridge et al. 2014). Our work addresses this need since our dataset is collected and assembled from observing individuals as they post about their health-related lifestyle behaviors.

Online social networks have opened new research opportunities (Kane et al. 2014, Utz 2015). There is an emerging body of literature using online social networks on health-related topics, investigating willingness to disclose health information (Anderson and Agarwal 2011), sentiment about healthy and unhealthy foods (Widener and Li 2014) and vaccination (Salathé and Khandelwal 2011), diffusion of various diseases (Achrekar et al. 2011, Culotta 2010), registration in online health forums (Centola 2010), urban-rural health disparity (Goh et al. 2016), and geographical analysis through tweets (Chen and Yang 2014, Widener and Li 2014, Ghosh and Guha 2013). Compared to offline social networks, online social networks provide a lower level of social presence and information richness for individuals (Chan and Cheng 2004). In recent years, location-based social networks have gained popularity. These mobile-based social networks provide location services and allow people to share their lifestyle behaviors with their friends through posting their location information.

Posts on location-based social networks open a window for observing individuals' lifestyle behaviors as they take place. Research has argued that while privacy concerns negatively impact intention to disclose location-related information, perceived benefits have a stronger positive influence (Zhao et al. 2012). While research has reported online social networks can influence political voting (Bond et al. 2012), and adoption of paid services (Bapna and Umyarov 2015), no observational studies have examined the role of online social networks in health-related lifestyle behaviors. With the skyrocketing cost of healthcare, examining the role of online social networks in changing health-related lifestyle behaviors could demonstrate the power of online social networks and help in formulating policies to harness this power to promote behaviors that improve individuals' health.

In addressing these gaps, we use data from open online social network platforms to answer the following research questions: (i) How can we observe individuals' health-related lifestyle behaviors on online social networks? (ii) What are the online social factors that contribute to individuals' health-related lifestyle as observed on online social networks? We define health-related lifestyle behaviors as lifestyle choices that people pursue in their daily life that could have positive or negative health consequences. These consequences may not necessarily be the main goals of such behaviors (Ingledew et al. 1996).

To answer the first research question, we captured and integrated data from two popular public online social platforms—Twitter and Foursquare (a location-based social network)—as well as data from the U.S. Census. We developed a novel method to extract, integrate, and interpret individuals' health-related lifestyle behaviors by observing their self-disclosed locational check-ins related to fitness, alcohol & smoking, and fast food diets. We established friendship networks using Twitter and observed location-based health-related lifestyle behaviors that have been shared through Foursquare in Twitter. Using this method resulted in a dataset for more than 32,000 individuals in all 50 U.S. states plus the District of Columbia over a twenty-week period.

To answer the second research question, we draw on the Berkman framework (Berkman et al. 2000) and social learning theory as well as social-network metrics to conceptualize a model that identifies salient factors associated with health-related lifestyle behaviors. We applied a new technique, the zero-inflated negative binomial method, to estimate the model. Our results show that the online social network of friends, online social support, the strength of friendship ties, homophily (gender similarity and geographical proximity) have significant impacts on individuals' health-related lifestyle behaviors. The results also show the role of socioeconomic status in such behaviors.

To the best of our knowledge, this paper is the first national study to integrate friendship networks from online social networks with real-time posts of individuals' location-based check-ins. This work is also the first observational study of how online friends' health-related lifestyle behaviors change individuals' health-related lifestyle behaviors. This paper makes a number of important contributions to theory, practice, and policymaking. It provides an integrated multi-period model for the study of people's health-related lifestyle behaviors and the factors that change such behaviors. This model expands the nature of online friendship to include the strength of friendships, social support, and two types of homophily, thus contributing to the theoretical treatment of online friendships. The practical implications of this work highlight the importance of considering these factors when developing policies and incentives that promote healthy lifestyle behaviors and counter chronic diseases caused by unhealthy lifestyles.

2.2. Literature Review and Theoretical Foundation

2.2.1. Literature on Health-related Lifestyle Behaviors.

We define health behaviors as behaviors that could have positive or negative health consequences, and could be goal directed or lifestyle related. Goal-directed behaviors are purposeful actions directed to accomplish a goal (Bühler 1957), and goal-directed health-related behavior is a healthy person's behavior to "prevent a disease or detect it in an asymptomatic stage." (Kasl and Cobb 1966, p. 246). Many health studies have focused on the goal-directed behaviors (Wit et al. 2011, Bayliss et al. 2014, Esposito et al. 2016).

Lifestyle behaviors are defined as "patterns of choices made from the alternatives that are available to people according to their socioeconomic circumstances and the ease with which they are able to choose certain ones over others." (Milio 1981, p. 76). Lifestyle behaviors are self-determined (Deci 1992) and discretionary (Wiley and Camacho 1980). We define health-related lifestyle behaviors as those that have direct health consequences, including physical activities, alcohol consumption & smoking, and dietary habits. Table 2.1 reports a selected set of studies about the health consequences of these behaviors.

A meta-analysis of 15 studies has shown that a combination of healthy lifestyle behaviors can reduce the risk of diseases (Loef and Walach 2012). However, with the exception of Djoussé et al. (2009), health studies focus on a single health-related lifestyle behavior—physical activities, alcohol & smoking, or dietary habits (Fielding 1985, Hung et al. 2004, Room 2005, Haskell et al. 2009). Our study addresses all the three sets of behaviors.

Table 2.1 Overview of Literature on Health-Related Lifestyle Behaviors

Context	Study	Finding
Physical Activity	Haskell et al. 2007	To promote and maintain health, people should engage in moderate physical activity for half an hour on five days each week or vigorous physical activity for 20 minutes on three days each week.
	Haskell et al. 2009	The importance of physical activities and the health risks of inactivity.
	Powell et al. 2011	Even a light level of physical activity can provide positive health consequences. They showed that there is a dose-response relation between physical activity and health risks.
	Lee et al. 2012a	Inactivity can cause premature mortality and is the main cause for around 10% of type 2 diabetes, breast cancer and colon cancer. They also estimated that a 25% decrease in physical inactivity could prevent more than 1.3 million deaths worldwide.
Dietary Habits	Currie et al. 2010	Proximity to fast-food restaurants increases the risk of obesity in young teens and pregnant women.
	McEvoy et al. 2012	A review of 80 papers showed that vegetarian diets and low meat plant-based diets reduce the risk of disease such as cardiovascular diseases and type 2 diabetes.
Alcohol Consumption & Smoking	Sesso et al. 2008	Heavy alcohol consumption increases the risk of high blood pressure. The results showed that light-to moderate levels of alcohol consumption decrease hypertension risk in women but increase the same risk in men.
	Rehm et al. 2010	Moderate alcohol consumption can increase the likelihood of major diseases in individuals. There is a dose-response relationship between alcohol consumption and health risks—an increase in the alcohol consumption leads to increase in level of risk.
	Bulloch et al. 2012	Alcohol dependence can increase the risk of major depressive episodes.
	Wong et al. 2007	Smoking has a negative effect on bone density and contributes to osteoporotic fractures.
	Pope et al. 2009	Light smoking has the same risk for cardiovascular disease as daily smoking.
	Jha et al. 2013	Quitting smoking increases the life expectancy of people in various age ranges.

Social groups facilitate the process of learning and influence lifestyle and health behaviors (Bruhn 1988). Studies of health-related lifestyle behaviors in large social groups are scarce with the exception of the works by Christakis and colleagues that investigated the spread of smoking, alcohol consumption and obesity in an offline social network and reported on the significance of social ties (Christakis and Fowler 2007, 2008, Rosenquist et al. 2010). This scarcity is partly because identifying and collecting data from all groups to which an individual may belong is a formidable task. Online social networks have brought to light groups of friends with whom individuals regularly interact through online platforms and follow their lifestyle behaviors on a daily basis. While online social networks, compared to offline social networks, expand the level of connections among individuals, they involve a lower level of social presence and a lesser degree of information richness in different social contexts (Chan and Cheng 2004). Therefore, it is not clear whether online social networks influence health-related lifestyle behaviors, and if so, what pathways are involved. This study addresses this gap by focusing on the three main behaviors (physical activities, alcohol consumption & smoking, and dietary habits) to develop a nationwide self-disclosed observational dataset from public online social networks and to conceptualize a theoretical model for identifying factors that impact such behaviors.

2.2.2 Literature on Theories

There is an abundance of theories in health studies that conceptualize the health effects of individuals' social environments (Bowlby 1969, Link and Phelan 1995, Berkman et al. 2000). We rely on an integrative framework by Berkman et al. (2000)–henceforth called the Berkman framework–to conceptualize our model. Berkman et al. (2000) developed their framework using a synthesis of social psychology and network theories.

The Berkman framework includes a comprehensive set of social factors (at macro- and micro-levels) impacting individuals' health through three pathways—health behavioral, psychological, and physiological (Berkman et al. 2000). In this study, we focus on health-related lifestyle behaviors, which form a subset of health behavioral pathways. Two factors influence health-related lifestyle behaviors: social support and social influence. Individuals' social network structure and characteristics provide the setting for these factors to influence health-related lifestyle behaviors. This is in line with the social network theory, which argues that the types of ties and structures of social networks facilitate the impact of social influence and social support (Borgatti et al. 2009). Research has argued that sphere of social influence and social support extend beyond immediate family and close spatial proximity to include individuals' social networks that surpass such boundaries (Barnes 1954, Bott 1957).

The Berkman framework also includes social factors at a macro level—cultural, political and socioeconomic factors. Prior studies in health-related lifestyle behaviors emphasize the role of socioeconomic status at this level (Lynch et al. 1997, Hanson et al. 2012). We therefore include socioeconomic status in our analysis. In sum, this study focuses on social support and social influence at the micro level and socioeconomic status at the macro level as factors that could influence individuals' lifestyle behaviors.

2.3. Model Conceptualization

The conceptualized model is presented in Figure 2.1 and discussed below.

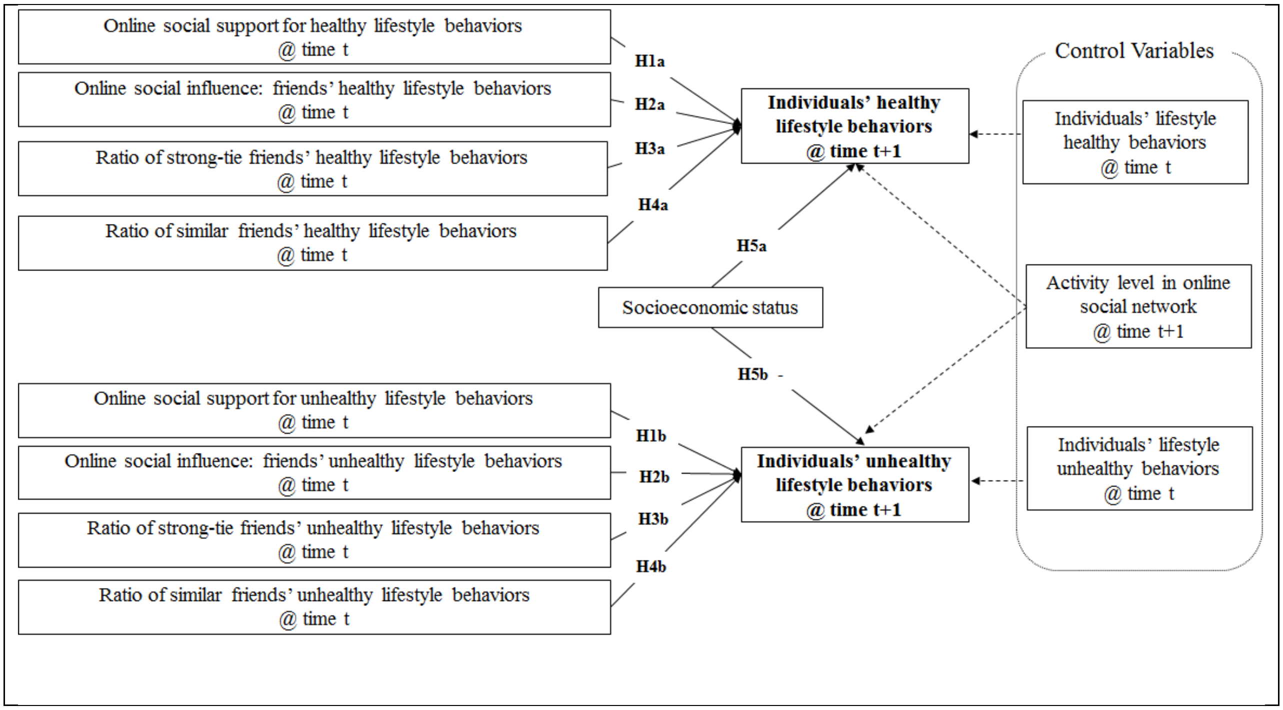


Figure 2.1. The Health-related Lifestyle Behavior (HLB) Model

2.3.1. Social Support

The Berkman framework argues that individuals' social network opens two main pathways for psychosocial mechanisms to influence individuals' health behaviors: (i) social support (ii) social influence (Berkman et al. 2000). Social support is a process that forms through human interactions (Rutter 1987) and constitutes one of the most important aspects of relationships in social networks (Appel et al. 2014). Social support is defined as "social resources that persons perceive to be available or that are actually provided to them by nonprofessionals in the context of both formal support groups and informal helping relationships" (Cohen el al. 2000, p. 4). Social support positively impacts individuals' health (Berkman et al. 2000, Bloom 1990, Fiori et al. 2006, Uchino 2006, Vandervoort 1999), and enhances individuals' ability to deal with personal health issues (Gallant 2003).

Research shows gaining social support requires connectedness (Langford et al. 1997), interactions (Barrera 1986, Vangelisti 2009), and feedback mechanisms (Caplan 1974, Barrera and Ainlay 1983). Focusing on these three elements has led researchers to study the concept of social support from sociological, communicational, and psychological perspectives (Vangelisti 2009, Goldsmith 2004, MacGeorge et al. 2011). The sociological perspective focuses on the availability of social support, and measures it by the level of individuals' connectedness or embeddedness into different social groups (Langford et al. 1997). The communicational perspective relies on interactions to conceptualize social support as the type of activity that others perform to support the recipient (Barrera 1986, Vangelisti 2009). The psychological perspective focuses on the functional aspects, including emotional (express emotion), informational (provide knowledge), instrumental (practical help), companionship (availability to participate in activities), and feedback (evaluate the appropriateness of behaviors) (Cohen et al. 2000, Vaux 1988, Wills 1991). These perspectives are at play in online social networks, particularly through the feedback mechanisms that provide functional social support.

In online social networks, people are connected to their family and friends (Ellison et al. 2007, Lampe et al. 2006) to inform them about their own activities (Hampton et al. 2011, Hampton 2016)—a type of self-disclosure that anticipates feedback from recipients (Lu and Hampton 2016). Personal feedback provides valuable information for the recipients to evaluate the appropriateness or normativeness of their behaviors (Cohen et al. 2000, DiClemente et al. 2001) and can be a source of motivation and inspiration for them (DiClemente et al. 2001). Research has shown that feedback received by individuals can play a major role in changing their health-related lifestyle behaviors (DiClemente et al. 2001, Kreuter et. al 1999), including alcohol consumption and smoking or dietary behaviors (DiClemente et al. 2001). A recent study on the association of online social networking and maternal well-being showed that sharing successful parenting experiences, receiving feedback from family and friends, and learning from others' experiences can enhance the perception of social support (McDaniel et al. 2012).

Feedback can range from generic to personalized types (DiClemente et al. 2001, Kreuter et. al 1999). In the most generic format, individuals receive general information that can be valid for a whole population. As feedback becomes more personalized, people receive information based on their own characteristics such as age, location, gender, ethnicity, or even based on the assessment of their own behaviors. The main advantage of personalized feedback is that people find it more relevant to themselves (DiClemente et al. 2001). Online social network platforms such as Facebook, Twitter and Foursquare make personalized feedback an easy process, enabling individuals to develop relationships with others, share various information about themselves and their activities, and receive positive feedback, affirmation and support in the form of "likes" and "favorites" for their behavior. The positive feedback can promote behavior continuance. Hence,

> **Hypothesis 1**: *Individuals who have received a higher level of online positive feedback in online social networks for their (a) healthy lifestyle behaviors at time t, are more likely to engage in healthy lifestyle behaviors at time t+1 (b) unhealthy lifestyle behaviors at time t, are more likely to engage in unhealthy lifestyle behaviors at time t+1.*

2.3.2. Social Influence

The Berkman framework argues for the role of social influence as a pathway from social network to health behavior (Berkman et al. 2000). Social influence is considered an important factor in the development of individuals' personality, physical characteristics and behavioral tendencies (Coleman 1980, Epstein 1989). Social influence is present when the likelihood of performing a particular action depends on engagement of the individual's peers in the same action (Agarwal et al. 2009, Aral 2011). Theoretically, social influence has roots in social learning, network externality, or pressure from the reference groups (Agarwal et al. 2009). While the latter two mechanisms emphasize the role of pressure from social groups for adoption of a specific behavior, the social learning mechanism argues that people learn new behaviors not only from their own personal experiences, but also by observing others' behaviors and the consequences of those behaviors (Bandura 1969, Latané 1981, Bandura 1986, Duflo and Saez 2002, Munshi 2004). In doing so, they compare and then align their behavior with their reference groups (Marsden and Friedkin 1993, Friedkin 2001). We focus on social learning as the source of social influence in the online social networks.

Research in individuals' different behaviors has argued that social influence is an important factor in propagation of most human behaviors through social networks (Christakis and Fowler 2009, Smith and Christakis 2008). Prior studies found three major challenges in detection of social influence in human behavior. First, it is difficult to distinguish between the endogenous effects (existence of social influence), and correlated effects (unobserved common characteristics) (Manski 1993). Second, behavior modification of social influence involves time-dependent factors (Van den Bulte and Lilien 2001, Risselada et al. 2014). Third, people tend to interact with similar others more frequently (McPherson et al. 2001) and it is expected that they behave similarly ("birds of a feather flock together."). We address these concerns in two ways (1) by using a dynamic approach in which we study the impact of friends' behaviors at time t on individuals' behaviors at time $t+1$ (Agarwal et al. 2009), (2) by separating social influence from the influence of individuals' similarities with their friends.

Studies in specific health issues have repeatedly demonstrated the significance of person-to-person social influence in the spread of obesity (Christakis and Fowler 2007), alcohol consumption & smoking (Christakis and Fowler 2008, Rosenquist et al. 2010), dietary behavior (Hutchinson and Rapee 2007, Cruwys et al. 2015), and poor fitness (Carrell et al. 2011). However, it is not clear whether social influence operates in the same manner in online social networks. Although online social networks provide a lower level of information richness for individuals (Chan and Cheng 2004), individuals are exposed to all online friends' health-related lifestyle behaviors daily or hourly as they interact with friends online and check their status. In the case of location-based social networks (the focus of our study), posted behaviors explicitly and concretely show the activities individuals' friends are doing, almost in real time. This extensive exposure provides a rich environment for the influence of online social networks in changing individuals' health-related behaviors since they provide more opportunities for what is called observational learning (Kwon et al. 2014).

Thus, we argue that social influence in terms of friends' health-related lifestyle behaviors could impact individuals' disclosed online health-related lifestyle behaviors. Hence,

> **Hypothesis 2**: *Individuals' (a) healthy lifestyle behaviors at time t+1, are positively influenced by their online friends' healthy lifestyle behaviors at time t (b) unhealthy lifestyle behaviors at time t+1 are positively influenced by their online friends' unhealthy lifestyle behaviors at time t.*

While social influence forms one of the main pathways from social networks to individuals' health behavior in the Berkman framework, friendship in social networks has different levels of strength. The strength of ties demonstrates the intensity and tightness of a friendship (Risselada et al. 2014, Van den Bulte and Wuyts 2007). Strong ties increase friends' influences due to a higher level of trust and more interactions (Bapna et al. 2017, Iyengar et al. 2011, Coleman 1988), and can be measured in multiple ways (Bapna et al. 2017, Aral and Walker 2014). In offline social networks, the strength of ties is a perceptual concept in which individuals may face difficulties in judging the directionality and strength of their friendship (Almaatouq et al. 2016). Whereas the capability to view and traverse network connections is one of the main features of online social networks that distinguishes them from offline social networks (Kane et al. 2014).

This feature provides direct observational information for the assessment of strength of friendship in networks. Reciprocity and embeddedness are considered manifestations of friendship strength in online social networks (Bapna et al. 2017, Aral and Walker 2014). Reciprocity in online social networks is defined as bidirectional friendship of two individuals. Embeddedness is defined by the number of common friends between two individuals with reciprocated friendship (Aral and Walker 2014, Easley and Kleinberg 2010). Research has demonstrated that a higher number of common friends in embedded relationships increases the level of trust between individuals, and exerts greater social influence (Aral and Walker 2014). Recent studies on online social networks have argued that two-way relationships are stronger than one-way relationships (Shi et al. 2014, Kwak et al. 2010) and are instrumental for spreading online behaviors (Bond et al. 2012). Applied to health-related lifestyle behaviors, we posit that the higher ratio of strong-tie friends' healthy behaviors at time *t* should positively influence individuals' healthy behaviors at time *t+1*. Similarly, the higher ratio of strong-tie friends' unhealthy behaviors at time *t* should positively impact individuals' unhealthy behaviors at time *t+1*.

> **Hypothesis 3**: *The higher ratio of strong-tie friends' (a) healthy lifestyle behaviors at time t positively influence individuals' healthy lifestyle behaviors at time t+1, (b) unhealthy lifestyle behaviors at time t positively influence individuals' unhealthy lifestyle behaviors at time t+1*

Homophily is another source of influence. It is a measure of similarity of friends in social networks that can lead to similar behaviors. The mechanisms of social influence and homophily are not mutually exclusive (Bapna and Umyarov 2015). Research shows that individuals are more likely to trust and endorse others who are similar to them (Feick and Higie 1992), and that can increase the level of social influence in social relationships (Risselada et al. 2014, Nitzan and

Libai 2011, Choi et al. 2010). In the context of online social networks, research has reported that geographical proximity (Dubois and Gaffney 2014) and demographic similarities (Nahon and Hemsley 2014) impact the magnitude of social influence.

Geographical proximity is considered one of the main sources of similarity in the social network theory (Borgatti et al. 2009) and social influence literature (Choi et al. 2010, Agarwal et al. 2009). Prior studies of online social networks considered geographical location as one of the main sources of homophily (Yuan and Gay 2006, Choudhury 2011, Pelechrinis and Krishnamurthy 2012). This factor has also been widely used in the assessment of friendships and behaviors (Choi et al. 2010, Agarwal et al. 2009, Wang et al. 2011, Back et al. 2008). Research in the role of geographical proximity has reported that geographical proximity can exert a strong level of social influence on people (Wang et al. 2011). Geographical proximity represents a shared environment in which individuals can have physical interactions with friends (Agarwal et al. 2009) and influence friends' short term decisions (Choi et al. 2010). Location similarity can affect individuals' online behavior (Tang et al. 2015). Prior studies of location-based social networks emphasize on the importance of geographical proximity in the formation of social influence among friends (Zhang et al. 2012).

The second source of similarity is gender. Studies show that similarity can go beyond geographical proximity (Fischer 1978, Van Alstyne and Brynjolfsson 2005) and can be defined by other demographic characteristics. Studies have reported the importance of gender similarity in social settings (Lewis et al. 2011, Linden-Andersen et al. 2008). The tendency to have same-gender friendships has a long history (Lewis et al. 2011). In ancestral environments, same-gender friendships helped men in hunting, warfare and related skills. Same-gender friendships helped women gain knowledge about food, pregnancy, nursing and childcare. It is argued that in modern societies, people still benefit more from their same-gender friendships as a source of assistance in social and emotional issues (Lewis et al. 2011). Research has shown that gender similarity has a role in online behavior (Aral and Walker 2012, Tang et al. 2015).

Therefore, in this study homophily entails geographical proximity and gender similarity. We argue that the higher ratio of similar friends' healthy behaviors at time *t* should positively influence individuals' healthy behaviors at time *t+1*. Similarly, the higher ratio of similar friends' unhealthy behaviors at time *t* should positively impact individuals' unhealthy behaviors at time t+1.

> **Hypothesis 4**. *The higher ratio of similar friends'(a) healthy lifestyle behaviors at time t positively influence individuals' healthy lifestyle behaviors at time t+1, (b) unhealthy lifestyle behaviors at time t positively influence individuals' unhealthy lifestyle behaviors at time t+1*

2.3.3. Socioeconomic Status

There are opposing views related to health-related lifestyle choices (Lynch et al. 1997). One view considers lifestyle behaviors with health consequences as "intra-individual" resulting from individuals' lifestyle choices. The second view argues that while individuals are responsible for their choices, their socioeconomic status limit their available options. In this view, individuals' socioeconomic status influence health behaviors in a society—a view supported by the Berkman framework. Following the Berkman framework, we argue that individuals' health-related lifestyle behaviors are associated with their socioeconomic status (SES). Studies have shown that social inequalities in income, opportunities, resources and social status are factors associated both with health and healthy behaviors (Naidoo and Wills 2009) and with mortality rates (Phelan 2004). Individuals from low SES groups have unhealthier lifestyle behaviors, such as physical inactivity and poor diet (Lynch et al. 1997, Hanson et al. 2012), and suffer from poor health

(Pickett and Pearl 2001).

The reason is that socioeconomic status shapes individuals' physical environments, social environments, psychological patterns and health-related lifestyle behaviors (Adler et al. 1994). A higher crime rate and a lack of safety can decrease the level of physical activity (Foster and Giles-Corti 2008). Moreover, neighborhood studies have shown the health consequences of individuals' lack of adequate access to food stores (Moore and Diez Roux 2006, Zenk et. al 2005). For instance, a study of neighborhoods in a metropolitan area in Michigan observes that inadequate access to supermarkets leads to less-nutritious diets, leading to chronic, diet-related diseases (Zenk et. al. 2005). In Berkman framework, SES is been considered a macro-level factor that should be taken into account when studying health-related behaviors. Furthermore, it is shown that social and physical attributes of communities and neighborhoods can be good predictors of individuals' health-related lifestyle behavior (Diez Roux and Mair 2010). Since, individuals' online health-related check-ins are posts about actual behaviors in their physical and social environments, health-related lifestyle behaviors should be associated with their SES. We focus on association (rather than causality) because SES does not change in the short term from one period to the next. Thus, based on the Berkman framework and empirical studies supporting the Berkman framework, we posit that:

> **Hypothesis 5**: *SES is (a) positively associated with observed healthy lifestyle behaviors of individuals in online social networks (b) negatively associated with observed unhealthy lifestyle behaviors of individuals in online social networks.*

2.4. Data Collection

We relied on three data sources in assembling a panel of data: Twitter[4], Foursquare[5] and US census data (American Community Survey 5-year data 2013). Twitter is a popular online social network platform that can be used for real-time information sharing. Users of this platform are willing to share information about different aspects of their life such as their activities and locations. In order to facilitate this process, Twitter allows their users to integrate their accounts with other social network applications such as Foursquare. In Foursquare, members identify the location of venues they visit—called check-ins—and share their check-ins with friends. The integration capability of Twitter allows users to share their check-in information in their Twitter account right from the Foursquare platform. We took advantage of this integration and developed an extensive set of tools to capture, match, extract, and analyze information downloaded from Twitter, Foursquare and U.S. Census Bureau APIs. The data collection was conducted in three stages: user identification, data collection at times *t* and *t+1*, and complementary data collection. At the first stage, we captured check-ins of users in Twitter for a twelve-week period (January 28 – April 22 2014). In this period, we selected users in the U.S. who post at least one check-in every two weeks after their initial captured check-in. Of our collected data, 32,700 unique individuals met this requirement. During the first stage, the selected users posted on average 3.8 check-ins each week. At the second stage, we captured user check-ins at time *t* for a four-week period and then for time *t+1* for another four-week period. At the third stage, we captured

---

[4] Twitter is a social networking service that provides microblog features in which users can post 140-character messages on their own page and simultaneously keep in touch with their friends and followers.

[5] Foursquare is a location-based social networking service that provides location searching and sharing capabilities for their users. In May of 2014, Foursquare decided to split its mobile application into two parts: (1) Foursquare (2) Swarm. In the new plan, Foursquare (the main app) only provides information about locations and helps people to discover their desired place. Swarm (the new app) handles the social check-ins to help people share their location. As this migration took place during our data collection, we used both “Swarm” and “Foursquare” keywords to collect data from Twitter API.

complementary information about the pattern of social network connections, the socioeconomic status of users, the foursquare venue categories from which users shared their check-ins, the number of Favorites that users received for their posted check-ins, and the demographic information. Our approach involved a series of steps to identify individuals from their tweets and capture and integrate their publicly available data. We developed eleven tools with a variety of complexity to capture, extract, integrate, and compute data as shown in Figure 2.2.

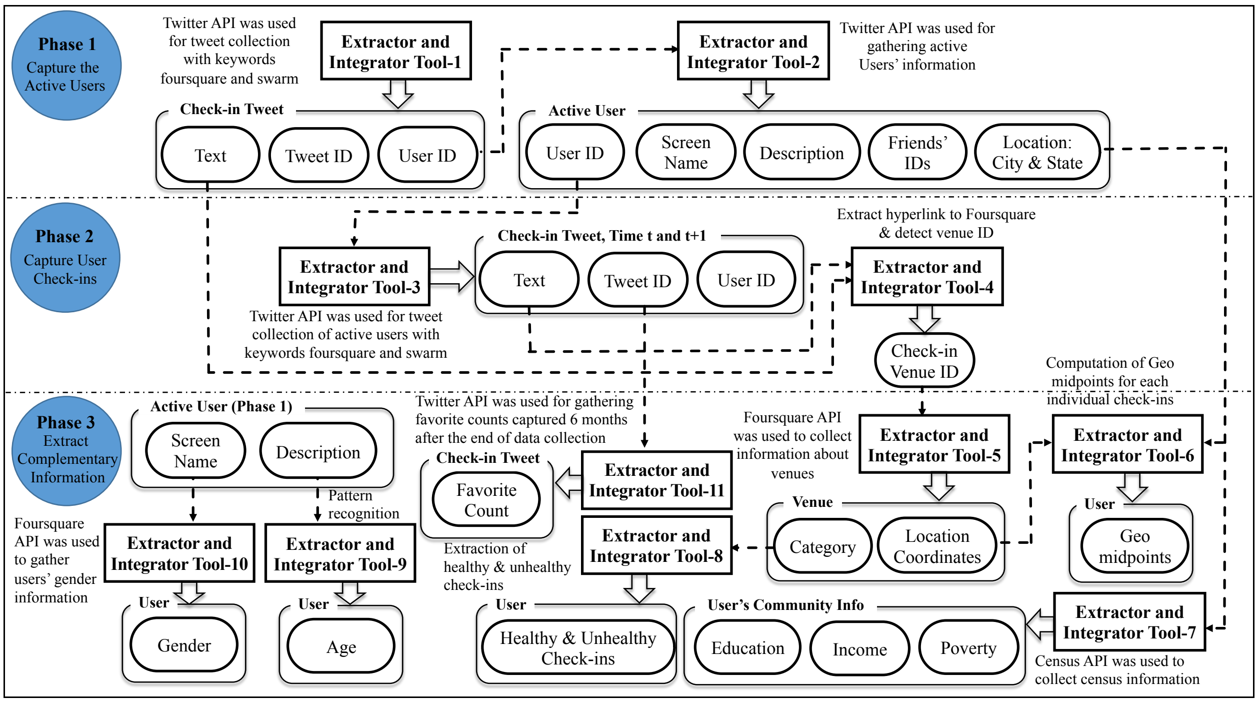


Figure 2.2. The Process of Data Capture, Extraction, Integration, and Computation

The tool numbers in Figure 2.2 indicate the sequence of data extraction and integration. The tools were coded in Python, R, and MySQL. We captured more than 5 million check-in tweets, 1,127,420 distinct venues in the U.S. and 259,255 unique individual users before reducing the sample to active users.

2.5. Behavior Observations and Representations

2.5.1. Representing Health-Related Lifestyle Behaviors

In this study, individuals' behaviors are observed based on what they have posted online as the places visited—type of venues in their Foursquare check-ins. We first discuss how we identified the venue types, and then present the arguments why visiting given venue types represents health-related lifestyle behaviors.

In order to capture the venue types, we used two steps. First, we extracted the Venue IDs from the hyperlinks in check-ins' text. Second, we used the Foursquare API to collect data about venues. There are various venues on Foursquare, such as restaurants, shopping, movie theaters and others. Our dataset has a total of 1,127,420 venues. Foursquare categorizes venues, and had 599 categories of venues at the time of our data collection. We used these categories to identify healthy lifestyle and unhealthy lifestyle check-ins. In doing so, we examined the types of establishment that fall under each category of venues for a better understanding of the categories.

Per Table 2.1 (in the Literature Review section), physical activities are considered as healthy lifestyle behaviors, whereas alcohol consumption & smoking and unhealthy diet are unhealthy lifestyle behaviors. Hence, we combined the salient categories in Foursquare to identify three types of venues associated with health-related lifestyle behaviors: *fitness center & gym*, *bar*, and *fast food restaurant*. Table 2.2 lists the Foursquare categories and number of venues in each type.

Table 2.2. List of Categories

Venue Type	Foursquare Categories	# of Venues
Fitness Center & Gym	Badminton Court, Baseball Field, Basketball Court, Boxing Gym, Climbing Gym, College Basketball Court, College Cricket Pitch, College Football Field, College Gym, College Hockey Rink, College Soccer Field, College Tennis Court, Cricket Ground, Gym, Gym / Fitness Center, Gym Pool, Gymnastics Gym, Hockey Field, Paintball Field, Rock Climbing Spot, Roller Rink, Rugby Pitch, Skate Park, Skating Rink, Soccer Field, Sports Club, Squash Court, Swim School, Tennis Court, Volleyball Court, Yoga Studio	36,047
Bar	Apres Ski Bar, Bar, Beach Bar, Beer Garden, Beer Store, Champagne Bar, Cocktail Bar, Dive Bar, Gastropub, Gay Bar, Hookah Bar, Hotel Bar, Irish Pub, Karaoke Bar, Piano Bar, Pub, Sake Bar, Sports Bar, Whisky Bar, Wine Bar	66,687
Fast Food Restaurant	BBQ Joint, Fast Food Restaurant, Food Court, Fried Chicken Joint, Hot Dog Joint, Mac & Cheese Joint, Pizza Place, Wings Joint	109,575

2.5.2. Representing Healthy Lifestyle Behaviors

We argue that going to fitness center & gym venues represents healthy lifestyle behaviors. In the selection of venues for this type, we distinguished between venues where people go to engage in physical activities and venues where people watch sports. Accordingly, we omitted all venues labeled as stadium. Fitness center & gym venues provide facilities for various physical activities and exercises. Individuals pay membership dues to utilize the machines, tools, trainers, classes, pools and other facilities these venues offer. The primary reasons for individuals going to such venues is to engage in physical activities and exercises—healthy lifestyle behaviors.

In order to examine the behavior focus of people at fitness center & gym venues, we collected and analyzed Foursquare highlighted keywords for these venues. Foursquare analyzes the tips (short reviews) by people who go to the venues and highlights the repeated words that represent the nature of the venue. Figure 2.3 shows examples of users' reviews and boldface-highlighted keywords that were repeated by users. We analyzed all the highlighted keywords based on venue type (Figure A.1. in **Appendix A**). Figure A.2. (Appendix A) reports top ten keywords with highest frequencies for fitness center & gym venues. These keywords are all focused on physical activities and exercise, thus providing a strong support for our argument that going to fitness center & gym venues represents engagement in physical activities, thus healthy lifestyle behaviors.

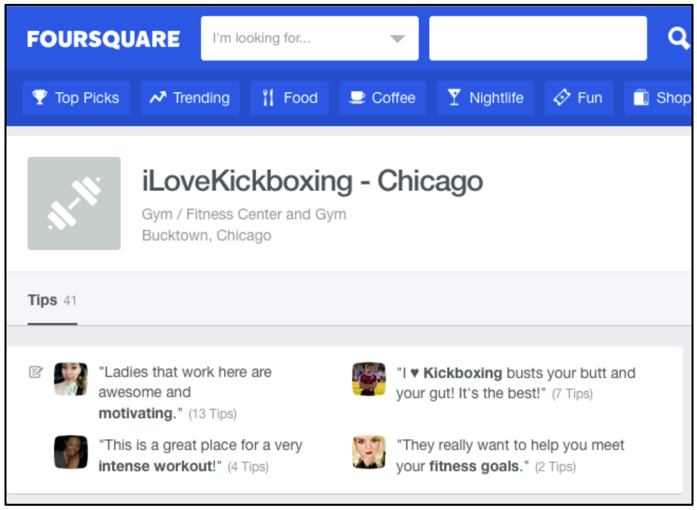


Figure 2.3. Example of Foursquare Boldface-Highlighted Keywords

2.5.3. Representing Unhealthy Lifestyle Behaviors

We argue that going to bars and fast food restaurants represent two types of unhealthy lifestyle behaviors. The main purpose of bars is to sell alcoholic beverages. Some bars offer Hookah smoking as well. Moreover, although smoking is banned in most U.S. bars, twenty states still allow cigarette/cigar smoking in some towns.[6] Even though many go to bars to socialize, the main activities in bars are drinking (and in some cases smoking), and socialization in such venues involves drinking. Alcohol increases appetite, making people crave food (Caton et al. 2004). However, bar foods are not healthy. Research has reported food quality is not an important factor for people who drink beers; the preference is for foods such as pizzas and fried food (Pettigrew and Charters 2006) that complement drinks, facilitate more drinking (Pettigrew and Charters 2006) and are associated with obesity (Arruda et al. 2016).

---

[6] There are 1208 Hookah bar venues in our dataset; There are 20 states in the U.S. that do not follow the complete ban of smoking in bars, namely, Alabama, Alaska, Arkansas, Florida, Georgia, Idaho, Indiana, Kentucky, Louisiana, Mississippi, Montana, Nevada, Oklahoma, Pennsylvania, South Carolina, Tennessee, Texas, Virginia, West Virginia, and Wyoming. More information is available on http://www.no-smoke.org/goingsmokefree.php?id=519

In order to examine the validity of our argument, we analyzed the Foursquare highlighted keywords for bar venues. As reported in Figure A.3. (**Appendix A**), the ten most frequent keywords describing this venue type refer to types of drinks and bar foods. The results indicate that those who go to bars primarily engage in drinking and in some cases (such as Hookah bars) in smoking as well. Going to bars represents an unhealthy type of lifestyle behavior.

The second type of venue that represents unhealthy lifestyle behavior is going to fast food restaurants. People who choose such venues look for quick, convenient and inexpensive foods (Rydell et al. 2008).[7] Fast food tends to be high in fat, energy dense, poor in micronutrients, low in fiber, high in glycemic load and excessive in portion size, which provides more energy than required for daily activities (Isganaitis and Lustig 2005, Rosenheck 2008). A systematic review on fast food consumption studies found that eating fast food is positively associated with gaining weight (Rosenheck 2008). Hence, going to fast food restaurants represents another type of unhealthy behavior.

To validate this argument, we analyzed the Foursquare highlighted keywords for fast food restaurant venues. As reported in Figure A.4. (**Appendix A**), the ten most frequent keywords describing this type of venue emphasize food names, supporting our argument that those who visit this type of venue engage in eating fast foods.

2.6. Variable Measurements

An individual's healthy and unhealthy lifestyle behaviors are measured as the number of days

[7] In prior studies, vegetarian, vegan, and low meat diets were considered healthy (Currie et al. 2010, McEvoy et al. 2012). Since the check-in data for vegetarian and vegan venues were too few and low-meat venues were not identifiable, healthy diet venues were not included in this analysis.

that each individual posted check-ins from venues within each type (fitness center & gym for healthy and bar and fast food restaurant for unhealthy lifestyle behaviors). The reason for this measure was that some users had repeated check-ins for a given type repeatedly on a given day. This causes bias in the data. Considering at most one check-in for each category per day removes this bias. Variable measurements are reported in Table 2.3 and discussed below.

Table 2.3. Variable Measurements at Individual Level

Model Variable	Definition	Metric and Computation
Dependent Variables		
Individual's healthy lifestyle behavior at time $t+1$	lifestyle behaviors that promote health	Individual' total number of days with check-ins at fitness center & gym type of venues at time $t+1$
Individual's unhealthy lifestyle behavior at time $t+1$	lifestyle behaviors that inhibit health	Individuals' total number days with check-ins at time $t+1$ measured for two venue type separately: 1. Bar. 2. Fast food restaurant.
Independent Variables, all lagged to measure impacts		
Online social support healthy (or unhealthy) lifestyle behaviors at time t	The support provided via feedback in online social networks for individuals' healthy (or unhealthy) lifestyle behaviors	1. For healthy lifestyle: Average number of Favorites an individual receives for check-ins at fitness center & gym venues at t, computed as the sum all Favorite counts received for fitness center & gym check-ins divided by number of days with fitness center & gym check-ins. 2. For unhealthy lifestyle: the fitness counts are replaced once by bar counts and again by fast food restaurant counts in the above computation. All computed at time t.
Social influence of friends' healthy (unhealthy) lifestyle behaviors at time t	The influence of friends' engagement in the same healthy (unhealthy) lifestyle behaviors	1. For healthy lifestyle: the average of friends' number of days with fitness center & gym check-ins at time t, computed as: sum of all friends' number of days with fitness center & gym check-ins divided by number of friends. 2. For unhealthy lifestyle: the fitness check-ins are replaced once by bar check-ins and again by fast food restaurant check-ins in the above computation. All computed at time t.
Social influence of strong ties' healthy (unhealthy) lifestyle behaviors at time t	The impact of strong friendship ties in social influences of friends' engagement in the same healthy (unhealthy) lifestyle behaviors	1. For healthy lifestyle: the ratio of weighted average strong ties' number of days with check-ins at fitness center & gym venues divided by non-weighted average of friends' number of days with check-ins at fitness center & gym venues. This ratio is computed for the first measurement period. The weights for strong ties are computed as follows: (i) Non-reciprocated friends' lifestyle behavior gets no weight, (ii) Reciprocated friends' lifestyle behavior gets weight proportional to the number of common friends with focal individual. 2. For unhealthy lifestyle: the fitness check-ins are replaced once by bar check-ins and again by fast food restaurant check-ins in the above computation. All computed at time t.

Social influence of similar friends' healthy (unhealthy) lifestyle behaviors at time *t*	The impact of similarity with friends/homophily in social influences of friends' engagement in the same healthy (unhealthy) lifestyle behaviors	1. For healthy lifestyle: The ratio of weighted average of similar friends' number of days with check-ins at fitness center & gym venues divided by non-weighted average of friends' number of days with check-ins at fitness center & gym venues. This ratio is computed for the first measurement period. The weights for similar friends are computed as follows: (i) Friends get .5 similarity score if they reside in 0-10 miles of the focal individual (ii) Friends get .5 similarity score if they have similar gender as focal individual (iii) Friends' lifestyle behavior gets weight proportional to the final calculated similarity score. 2. For unhealthy lifestyle: the fitness check-ins are replaced once by bar check-ins and again by fast food restaurant check-ins in the above computation. All computed at time *t*.
Socioeconomic Status		
Socioeconomic status (SES)	Socioeconomic status of the community in which each individual resides	A factor loading derived from the exploratory factor analysis of income, education and poverty levels of the city/town in which each individual resides.
***Control Variable*s**		
Activity level in social network at time *t+1*	Activity level in online social network at time *t+1*	Individuals total number of check-ins in online social network at time *t+1*
Individuals' healthy (unhealthy) lifestyle behavior at time *t*	Individuals' healthy (unhealthy) lifestyle behavior at time *t*	Individuals' healthy (unhealthy) lifestyle behavior at time *t*

2.6.1. Online Social Support

One reason online social networks are so popular is the opportunity such media provide to its members to receive social support from their friends and other network members. Literature has shown that a number of forces influence people's reflective/reasoned behaviors, including other people's opinions, which may have positive or negative consequences (Thaler and Sunstein 2008). Social support through people's opinions could have emotional, informational, instrumental, companionship, and feedback forms (Cohen et al. 2000). In online social networks, social support is expressed in feedback form. We argue that social support in feedback form is expressed in reactions such as "like" or "comment" in Facebook and "favorite" or "retweet" in Twitter. When individuals tweet Foursquare check-ins, their friends get a chance to designate them as a favorite, indicating positive support for the check-ins. Since check-in tweets are rarely retweeted in Twitter, we use favorite counts for each check-in as a measure of social support in the form of positive feedback. Therefore, we measured social support for healthy and unhealthy lifestyle check-ins by the average number of favorites that each individual received for his/her fitness center & gym, bar, and fast food restaurant check-ins at time t. Following Bray (2012), who reported on the time it takes for all favorites to be posted, the check-in favorites were collected six months after the completion of data collection to make sure all favorites were captured.

2.6.2. Social Influence: Friends' Healthy and Unhealthy Lifestyle Behaviors

Friends' healthy and unhealthy lifestyle behaviors were computed by the same method as individuals' healthy and unhealthy lifestyle behaviors. Friends were identified in an egocentric network of individuals. In social network theory, an egocentric network is defined as the network of a single individual (ego) together with his or her friends (alters). In these networks, the relationship between individuals can be directional or un-directional. Facebook is an example of an un-directional social network, in which a friendship link forms only when both of the individuals consent to create the relationship. In contrast, Twitter allows directional relationships in its platform in which individuals can follow each other without permission. In a directional social network, individuals can only see the activities of people who they directly follow. In Twitter terminology, the person who follows others is called "follower" and the person who is followed by others is called "friend". As individuals can only see the behavior of their friends and not their followers, in the first step, we identified all the friends of each individual in our dataset. Then, for each friend, the number of days with check-ins in fitness center & gym, bar, and fast food restaurant venues in the first measurement period were counted. The sum of fitness center & gym check-ins of friends divided by the total number of friends with reported Foursquare check-ins in Twitter measured the healthy lifestyle behaviors of the individual's friends. Taking the average accounted for individuals' differences in the number of friends. Similarly, the sum of bar check-ins of friends divided by the total number of friends as well as the sum of fast food restaurant check-ins divided by the total number of friends were used as two measures of the unhealthy lifestyle behaviors of the individual's friends.

2.6.3. Social Influence: Ratio of Strong Ties' Healthy and Unhealthy Lifestyle Behaviors. In online social networks, one-way relationships are more fragile than two-way relationships (Shi et al. 2014). According to Kwak et al. (2010) the likelihood of breaking the relationships (unfollow) is twice as high in one-way relationships as in two-way relationships. It is argued that reciprocated friendships can indicate emotional closeness for both users. Moreover, the number of common friends between two individuals demonstrate how these two individuals are embedded inside the egocentric networks of each other. Higher embeddedness in a social network increases trust between individuals (Uzzi 1997). Thus, to measure the role of strong ties, we computed the following: (1) we identified reciprocated friends (two-way relationships) in the online social network, (2) for each individual, we computed the weighted average of reciprocated friends' healthy lifestyle behavior, in which each reciprocated friend's behavior got weight proportional to his/her number of common friends with the ego (focal individual), and (3) we divided this weighted average of the strong ties' healthy lifestyle behavior by the non-weighted average of all friends' healthy lifestyle behavior. For unhealthy lifestyle behaviors, this computation was repeated for bar check-ins and fast food restaurant check-ins to measure the ratio of strong ties' unhealthy behaviors.

2.6.4. Social Influence: Ratio of Similar Friends' Healthy and Unhealthy Behaviors.

To analyze the effect of homophily on social influence, we measured homophily based on geographical proximity and gender similarity. To estimate the geographical proximity of individuals we applied a two-step approach. (1) We computed the center point of each user's check-ins in his or her state. **Appendix B** provides a short description of the computation. (2) We computed the Euclidean distance of center points of each individual and his or her friends. The distances of check-in center points for individuals and their friends ranged from 0.05 to 6034.4 miles. Figure 2.4 shows the probability distribution graph of distances of check-in center points. To identify close proximity, we partitioned the distance range into four bins (near, moderate, long and far away) with equal probability (p = 0.25), resulting in ranges in miles for Bin1=(0-10.3), Bin2=(10.3-216.4), Bin3=(216.4-1324.4), and Bin4=(1324.4-6034.4) miles. **Appendix C** provides detailed information about the distance frequency in each bin. We used near (Bin1) distance for measuring geographical proximity among individuals.

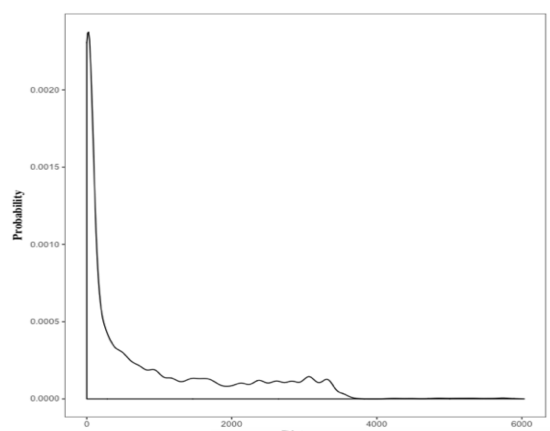


Figure 2.4. Probability Distribution of Distance in Miles

In order to combine the two homophily metrics, we computed the homophily score for each relationship between two individuals in the network, in which all similarities have equal weights (Nitzan and Libai 2011, Risselada et al. 2014). In our case, geographical proximity and gender similarity received 0.5 each. Thus, if two individuals were similar in proximity and gender, their homophily score was 1, for no similarity the score was zero, and for one similarity the score was 0.5. The homophily score was used as the weight for the connection between two individuals in the computation of the healthy and unhealthy lifestyle behaviors of an individuals' friends in the network.

To measure the role of homophily of friends, we computed the ratio of similar friends' healthy lifestyle behaviors as follows: (1) for each individual, we computed the weighted average (where homophily scores are used as weights) of friends' healthy lifestyle behavior, (2) we divided this weighted average by the non-weighted average of all friends' healthy lifestyle behavior. This gave us the ratio of similar friends' healthy lifestyle behaviors. Likewise, we repeated this computation for unhealthy lifestyle behaviors by replacing for bar and fast food restaurant check-ins in the above steps.

2.6.5. Socioeconomic Status

To measure the socioeconomic status of individuals, we used the data extracted from American Community Survey 5-year data (2013) to extract associated individuals' income, education and poverty levels at the city/town level, as identified from their Twitter profiles. These variables together represent the socioeconomic status of the community and are highly correlated. We used the explanatory factor analysis (EFA) to combine these factors and use one single representative factor for socioeconomic status. The load factors for income, education and poverty were 0.86, 0.62, and -0.94, respectively. This indicated an acceptable level of load to represent socioeconomic status.

2.6.6. Control variables

In order to infer causality of different factors, we controlled for past behaviors in the model. Accordingly, we included individuals' healthy and unhealthy lifestyle behaviors at time *t* as a control variable at time t+1. Moreover, individuals have different levels of activity on online social networks, impacting the frequency of their posts in the online social networks. To control for this variability, we included activity level in online social networks as a control variable in our model.

2.7. Data Analysis and Model Estimation

The correlation matrices for the behavior groups are reported in Tables D.1-D.3 in **Appendix D**.

2.7.1. Checking for Selection Bias

We checked for selection bias in a number of ways. First, our dataset covers all U.S. states, as reported in Figure 2.5.

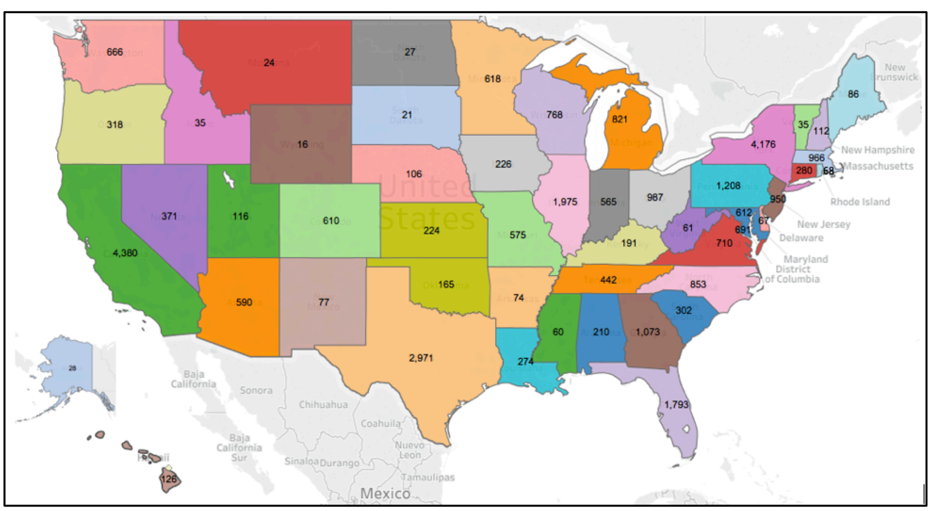


Figure 2.5. Number of Individuals in Each State

Second, using data from the latest available Census Bureau report on individuals' Internet usage by state in 2012[8], we computed the ratio of internet usage in each state (state usage divided by the country total)—called the *state ratio*. We also computed the ratio of individuals in each state in our sample, called the *sample ratio*. Figure 2.6 reports the deviation of the two ratios (*sample ratio – state ratio*) for each state.

Figure 2.6 shows that the sample and state ratios are relatively close. The under-representation in our sample is quite small (maximum underrepresentation is -0.008). Four states with large metropolitan cities are slightly over-represented in our sample: New York, Illinois, Texas, and California (maximum over-representation is 0.066). This could have been caused by a higher level of Internet growth in large cities since 2012. Overall, Figure 2.6 indicates that our sample is a good representation of the U.S. States.

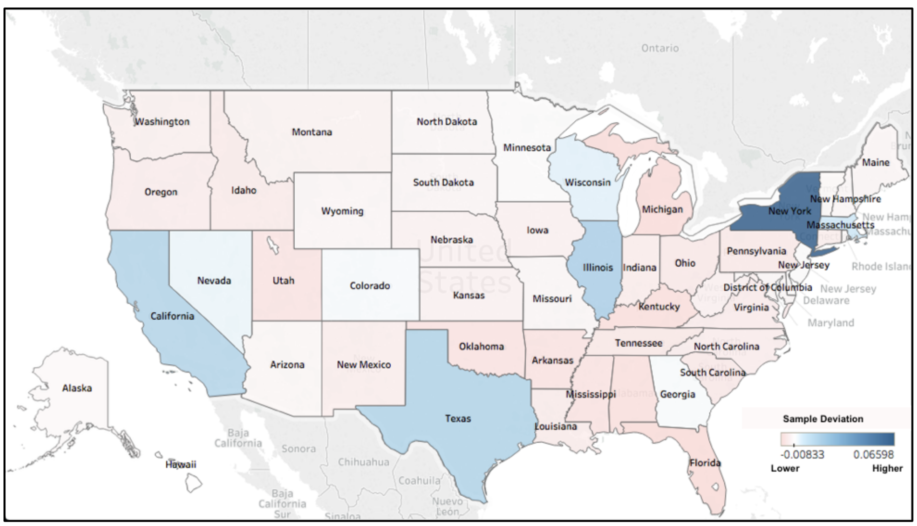


Figure 2.6. Deviation of Sample Ratio from State Ratio

Third, we computed the gender distribution of Twitter users based on the data published in the

---

[8] https://www.census.gov/data/tables/2012/demo/computer-internet/computer-use-2012.html

Pew report[9] and Census Bureau for 2014[10], which showed that 52% of U.S. Twitter users are male and 48% are female. In our dataset, 60% of users are male and 40% are female, indicating a relatively close match.

Fourth, using the same sources, we compared the average age of Twitter users in 2014, which was 37 years, with an approximation of the average age of users in our dataset. While we do not have access to the ages of all users in our dataset, we used a text analysis approach to collect the age of a sample of users who mentioned their age in their profile description. The average age of users in this sample is 31 years. Considering older people are less willing to post their ages, the average age in our sample is relatively close to the average of the 2014 Twitter user population. These checks indicated that selection bias did not pose a serious threat in our dataset.

2.7.2. Model Estimation

The distributions of individual check-ins in the second measurement period (**Appendix E**) indicated the presence of over-dispersion—greater variability than expected in data—which could be caused by high occurrences of zero values (Lee et al. 2012b). We tested for over-dispersion using the alpha test (Cameron and Trivedi 1990), which showed the presence of over-dispersion. In our case, zeros resulted from either not having such check-ins or not reporting them online.

When the data are over-dispersed, negative binomial estimators are the preferred estimation methods (Cameron and Trivedi 2013). When over-dispersion is accompanied with

---

[9] http://www.pewinternet.org/data-trend/internet-use/latest-stats/
and http://www.pewinternet.org/2015/01/09/demographics-of-key-social-networking-platforms-2/
[10] http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk

excessive zeros, then the zero-inflated negative binomial (ZINB) is a suitable estimation method. In order to check the appropriateness of ZINB for our analysis, we compared ZINB estimation results with the equivalent models of standard negative binomial regression using the Vuong test (Vuong 1989). A Vuong statistic greater than 1.96 indicates a preference to use the ZINB method (Long 1997). For our models, the lowest Vuong statistic was 26.17, indicating the superiority of ZINB as the estimation method in our case (Faraj et al. 2015).

The ZINB model identifies two latent groups who could generate zero value for the dependent variable (Y). (In our case, we have three dependent variables: individuals' posts about one healthy and two unhealthy lifestyle behaviors.) The first latent group consists of individuals who post about a given health-related lifestyle behavior—Group A. For a given *Y*, for each individual in Group A, $Y \geq 0$, depending on the count of individual's posts about check-ins of a venue type. For a given *Y*, the latent Group B consists of individuals who do not post about their healthy (or unhealthy) lifestyle behaviors in online social networks. For those in Group B, $Y=0$ by definition—inflating the zero values of *Y*. Zero values in Group A represent the lack of individuals' check-ins for the given behaviors, whereas zeros in Group B indicate individuals' unwillingness to post their behaviors in online social networks.

Each observation has a probability of belonging to either Group A or Group B. The result of a Bernoulli trial determines which process occurs. For each observation i, φ_{it+1} is the occurrence probability of Group B at time t+1, and $1-\varphi_{it+1}$ is the occurrence probability of Group A at time t+1, which has a count generated from a process that has negative binomial distribution $f(y_{it+1}|X_{it})$:

$$y_{i,t+1} \sim \begin{cases} 0 & with\ probability\ of\ \varphi_{i,t+1} \\ f(y_{i,t+1}|X_{i,t}) & with\ probability\ of\ 1-\varphi_{i,t+1} \end{cases}$$

The probability of the dependent variable is:

$$P(Y_i = y_{i,t+1}|X_{i,t}, Z_{i,t}) = \begin{cases} \varphi(\gamma Z_{i,t}) + [1-\varphi(\gamma Z_{i,t})]f(0|X_{i,t}) & if\ y_{i,t+1} = 0 \\ [1-\varphi(\gamma Z_{i,t})]f(y_{i,t+1}|X_{i,t}) & if\ y_{i,t+1} > 0 \end{cases},$$

where $X_{i,t}$ is the vector of independent variables for individual i at time t, $Z_{i,t}$ is a vector of covariates that contribute to not adopting the target behavior (zero-inflated part) by individual i at time t; γ is the vector of estimated zero-inflated coefficients, $\varphi_{i,t+1}$ is a function (φ) of $\gamma Z_{i,t}$, $\varphi(\gamma Z_{i,t})$ represents the probability of not engaging in the behavior, and $[1-\varphi(\gamma Z_{i,t})]f(0|X_{i,t})$ is the probability of engaging in the target behavior but not posting. Z is a vector of independent variables that are associated with not engaging in the target behavior. We investigated the role of all independent variables in not engaging in the target behavior. We used R for model estimation (R Development Core Team 2016).

2.7.3 Estimation Results

Tables 2.4, 2.5, and 2.6 report the estimations of the HLB model for fitness center & gym, bar, and fast food restaurant.[11] The estimation method was ZINB. The HLB model is estimated in Group A—testing the hypotheses about individuals who post their check-ins on online social networks. Group B is the zero-inflated estimation that provides additional information about the

[11] In order to capture the effect of strong ties and homophily on the individuals' healthy and unhealthy lifestyle behaviors, we had to consider only those individuals who have friends with non-zero average value for the same health-related lifestyle behavior. That reduced the number of observations in each model, as reported in Tables 2.4-2.6. We used the same number of observations in the estimations reported in each table. We also ran Models 1, 2, and 4 in each table using the full 32, 700 observations and got similar results.

role of factors in reducing or increasing the inhibition of individuals who do not post.

Table 2.4. Estimated HLB Model: Healthy Lifestyle Behaviors (Fitness Center & Gym)

	Model 1	Model 2	Model 3	Model 4
Group A/Count				
Control Variable: Activity level in online social network @ t+1	**.002*****	**.002*****	**.002*****	**.022*****
Control Variable: Individual's healthy lifestyle behavior @ t	**.185*****	**.185*****	**.185*****	
H1a.Online social support: healthy lifestyle behavior @ t	**.140*****	**.145*****	**.142*****	
Social Influence				
H2a.Friends' healthy lifestyle behavior @ t		**.025*****	**.026*****	
H3a.Ratio of strong ties' healthy lifestyle behavior @ t			.016	
H4a.Ratio of similar friends' healthy lifestyle behavior @ t			**.048****	
H5a.Socioeconomic status: SES				**.339*****
Constant	**-.052***	**-.095 *****	**-.160*****	**-.502*****
Group B/Zero Inflated.				
Control Variable: Activity level in online social network @ t+1	**-.082*****	**-.084*****	**-.083*****	**-.803*****
Control Variable: Individual's healthy lifestyle behavior @ t	**-1.405*****	**-1.399*****	**-1.398*****	
Online social support: healthy lifestyle behaviors @ t	0.005	-.0006	-.0001	
Social Influence				
Friends' healthy lifestyle behavior @ t		**-.037****	**-.039****	
Ratio of strong ties' healthy lifestyle behavior @ t			**-.081****	
Ratio of similar friends' healthy lifestyle behaviors @ t			-.015	
Socioeconomic status: SES				.323
Constant	**1.889*****	**1.948*****	**2.039*****	**2.443*****
Log Likelihood	-21,521	-21507	-21,499	-24,165
Wald x^2	2,345***	2,364***	2,374***	463***

N=22,423; *p<.05; **p<.01; ***p<.001.

Table 2.5. Estimated HLB Model: Unhealthy Lifestyle Behaviors (Bar)

	Model 1	Model 2	Model 3	Model 4
Group A/Count				
Control Variable: Activity level in online social network @ t+1	**.005*****	**.005*****	**.005*****	**.015*****
Control Variable: Individual's unhealthy lifestyle behavior @ t	**.180*****	**.176*****	**.175*****	
H1b.Online social support: unhealthy lifestyle behaviors @ t	**.102*****	**.106*****	**.104*****	
Social Influence				
H2b.Friends' unhealthy lifestyle behavior @ t		**.039*****	**.039*****	
H3b.Ratio of strong ties' unhealthy lifestyle behavior @ t			**.052*****	
H4b.Ratio of similar friends' unhealthy lifestyle behavior @ t			.020	
H5b.Socioeconomic status: SES				**-.411*****
Constant	.007	**-.069*****	**-.133*****	**.267*****
Group B/Zero Inflated				
Control Variable: Activity level in online social network @ t+1	**-1.079*****	**-1.072*****	**-1.072*****	**-1.251*****
Control Variable: Individual's unhealthy lifestyle behavior @ t	**-.106*****	**-.102*****	**-.101*****	
Online social support: unhealthy lifestyle behavior @ t	**-.163***	**-.173***	**-.168***	
Social Influence				
Friends' unhealthy lifestyle behavior @ t		**-.058****	**-.059****	
Ratio of strong ties' unhealthy lifestyle behavior @ t			.000	
Ratio of similar friends' unhealthy lifestyle behaviors @ t			-.092	
Socioeconomic status: SES				-.300
Constant	**3.064*****	**3.182*****	**3.270*****	**3.196*****
Log Likelihood	-39,416	-39,359	-39,347	-42,241
Wald x^2	6,837***	6,959***	6,990***	1,590***

N= 28594; *p<.05; **p<.01; ***p<.001

Table 2.6. Estimated HLB Model: Unhealthy Lifestyle Behaviors (Fast Food Restaurant)

	Model 1	Model 2	Model 3	Model 4
Group A/Count				
Control Variable: Activity level in online social network @ t+1	**.009*** **	**.009*** **	**.009*** **	**.019*** **
Control Variable: Individual's unhealthy lifestyle behavior @ t	**.156*** **	**.156** **	**.154** **	
H1b.Online Social support: unhealthy lifestyle behaviors @ t	**.050** **	**.051** **	**.051** **	
Social Influence				
H2b.Friends' unhealthy lifestyle behavior @ t		**.016*** **	**.017*** **	
H3b.Ratio of strong ties' unhealthy lifestyle behavior @ t			-.011	
H4b.Ratio of similar friends' unhealthy lifestyle behavior @ t			**.028***	
H5b.Socioeconomic status: SES				**-.116** **
Constant	**-.131*** **	**-.152*** **	**-.168*** **	-.011
Group B/Zero Inflated				
Control Variable: Activity level in online social network @ t+1	**-.606*** **	**-.606*** **	**-.606*** **	**-.680*** **
Control Variable: Individual's unhealthy lifestyle behavior @ t	**-.086*** **	**-.081*** **	**-.081*** **	
Online social support: unhealthy lifestyle behaviors @ t	-.033	-.042	-.042	
Social Influence				
Friends' unhealthy lifestyle behavior @ t		**-.100*** **	**-.098*** **	
Ratio of strong ties' unhealthy lifestyle behaviors @ t			-.021	
Ratio of similar friends' unhealthy lifestyle behaviors @ t			.052	
Socioeconomic status: SES				-.338
Constant	**2.6018*** **	**2.745*** **	**2.714*** **	**2.765*** **
Log Likelihood	-32,516	-32,503	-32,501	-33,752
Wald x^2	4,641***	4,647***	4,652***	2,394***

N=27,253; *p<.05; **p<.01; ***p<.001

In each table, variables were progressively added as Models 1-3. The increases in log likelihood and Wald x^2 values (comparing each model to the base model with no variables) as the level of the estimation increases indicate the improvement in the fit as the factors are added to the model. Model 4 tests the association of behaviors with SES and is reported separately.[12]

Group A/Count: In H1(a/b), we hypothesized that online social support in terms of positive feedback on healthy/unhealthy lifestyle behaviors at time *t* has positive impact on individuals' healthy/unhealthy behaviors at time *t+1*. H1a was supported in Models 1-3 for healthy lifestyle behaviors (fitness center & gym check-ins). Similarly, H1b was supported for unhealthy lifestyle behaviors (bar at p<0.001 and fast food restaurant at p<0.01) in Models 1-3.

In H2(a/b), we hypothesized that social influence is at work on healthy/unhealthy

[12] The reasons for omitting other variables in this model are twofold: First, while the SES variable can define individuals' health related lifestyle behaviors both at time *t* and time *t+1*, then individuals' lagged behavior cannot be used as the control variable. Second, the coefficient of other variables become meaningless without the individuals' lagged behavior variable.

lifestyle behaviors and argued that friends' healthy/unhealthy lifestyle behaviors at time *t* have a positive effect on individuals' healthy/unhealthy lifestyle behaviors at time t+1. These hypotheses were supported for both healthy and unhealthy lifestyle behaviors (fitness center & gym, bar, fast food restaurant check-ins) at $p<0.001$ in Models 2-3. Furthermore, in H3(a/b) we hypothesized that the strength of friendships in online social networks exerts additional social influence on individuals' health-related lifestyle behaviors. These hypotheses were supported for bar at $p<0.001$, but not for fast food restaurant and healthy lifestyle behaviors (fitness center & gym).

In H4(a/b), we hypothesized the additional influence of similarity (homophily) as measured by geographical proximity and gender similarity of individuals. These hypotheses were supported for healthy lifestyle behaviors (fitness center & gym) at $p<0.01$ and unhealthy lifestyle behaviors (fast food restaurant) at $p<0.05$ in Model 3, but not for bar.

Thus, the results showed a mixed support for strength of social ties (H3) and friends' similarity, in that the strength of social ties is important for promoting healthy behaviors whereas similarity of friends plays a role in some unhealthy behaviors but not others. This is an important finding since it shows that while social influence in online friendship is critical in promoting all types of health-related behaviors, the impacts of friendship attributes in terms of strength and similarity depend on the context of the behaviors. This is an unexpected finding in that it introduces context-dependency in the study of social influence.

In H5, we hypothesized socioeconomic status (SES) has positive association with healthy lifestyle behaviors (H5a) and negative association with unhealthy lifestyle behaviors (H5b). These hypotheses were supported for healthy lifestyle behaviors (fitness center & gym) at $p<0.001$ and unhealthy lifestyle behaviors (bar at $p<0.001$ and fast food restaurant at $p<0.01$) in

Model 4 in Tables 2.4-2.6. The control variables—individuals' behaviors at time *t* and activity level at time t+1—were significant in all estimated models. Table 2.7 summarizes estimation results for each hypothesis.

Table 2.7. Supported Hypotheses

Healthy Lifestyle Behavior	**H1a**	**H2a**	**H3a**	**H4a**	**H5a**
Fitness Center & Gym	yes	yes	no	yes	yes
Unhealthy Lifestyle Behavior	**H1b**	**H2b**	**H3b**	**H4b**	**H5b**
Bar	yes	yes	yes	no	yes
Fast Food Restaurant	yes	yes	no	yes	yes

Group B/Zero-inflated. The ZINB estimations provide additional insights regarding the impacts of social factors on inhibiting individuals to post about health-related lifestyle behaviors. The zero-inflated parts of Tables 2.4-2.6 report these impacts. A significant negative coefficient for a factor in the Zero Inflated part of tables indicates that the factor reduces individuals' inhibition to post about specific health-related behaviors, thus reducing the probability of their memberships in Group B.

Per Part B/Zero Inflated of Table 2.4, the social influence of friends' check-ins of fitness center & gym at time *t* significantly reduces individuals' inhibition to post about their check-ins at time *t+1* (coefficient -0.039, $p < 0.01$). Similarly, social influence of the ratio of friends with strong ties who post their check-ins of fitness center & gym at time *t* significantly reduces inhibitions about posting check-ins in such venues at time *t+1* (coefficient -0.052, $p < 0.01$). Social support and similarity of friends have no impact on willingness to post check-ins of fitness center & gym.

Per Part B/Zero Inflated of Table 2.5, social support of friends who post their check-ins of bar venues at time *t* significantly reduces the inhibition of individuals about posting their bar check-ins at time *t+1* (coefficient -0.168, $p<0.05$). This finding is interesting in that social support related to bar venues increases individuals' inclinations to post their check-ins of bar venues. In other words, social support increases individuals' incentives to reveal their unhealthy behaviors online. Moreover, social influence of friends' bar check-ins at time *t* also reduces individuals' inhibition about posting their check-ins at time *t+1* (coefficient -0.059, $p<0.01$). The strength and similarity of friendship do not have any impact on individuals' inhibition about posting bar check-ins.

Per Part B/Zero Inflated of Table 2.6, the social influence of friends' posting their check-ins of fast-food restaurants venues reduce inhibition about posting about check-ins of this venue type (coefficient 0.098, $p< 0.001$)—hence increasing individuals' willingness to reveal their unhealthy behaviors online. Strength and similarity of friendship, and social support do not play a role here.

Part B/Zero Inflated results in Tables 2.4-2.6 uniformly show that online social network activities at time *t+1* and higher records of check-ins about health-related lifestyle behaviors at time *t* reduce the inhibition to post health-related lifestyle check-ins at time *t+1*. Finally, we found that socioeconomic status is not associated with individuals' willingness to post health-related lifestyle behaviors online.

2.8. Discussions

This study's first research question was whether it was possible to observe individuals' health-related lifestyle behaviors in online social networks. We answered this question by our dynamic sequential approach to data capture, extraction, and integration from posts on Twitter, Foursquare location-based check-ins and integration with Census Bureau community data. This data collection process was guided by the literature on health-related lifestyle behaviors—

physical activities, alcohol & smoking and diet. Our approach made it possible to assemble the first large nationwide online observational dataset for individuals' healthy and unhealthy lifestyle behaviors

The second research question in this study was to identify significant social factors contributing to individuals' healthy and unhealthy lifestyle behaviors. Guided by the Berkman framework and associated theories, we developed the Health-related Lifestyle Behavior (HLB) model, which covered the main social pathways in the Berkman framework and identified the social factors contributing to individuals' health-related lifestyle behaviors. The estimation of the HLB model revealed the high potential of the online social network ecosystem to influence individuals' healthy and unhealthy lifestyle behaviors.

First, the empirical results of estimating the HLB model uncovered the way friends' healthy lifestyle behaviors could influence individuals' healthy choices. We found strong impacts of friends' online social support through Twitter's favorites on individuals' healthy lifestyle behaviors. This is a novel finding, documenting the significant soft power of online nudging by friends. In their seminal work, Thaler and Sunstein (2008) refer to nudging as soft persuasion in human decision making for health and other critical choices without compulsion. Our findings show that online social support for healthy choices acts as the nudge that could steer individuals toward healthy lifestyle behaviors or encourage them to maintain healthy lifestyle choices.

Second, online social support on unhealthy choices could be just as effective. In unhealthy lifestyle behaviors related to bars and fast food restaurants, online social support in the form of Twitter's favorites positively and significantly contributes to the adoption and reporting of these unhealthy behaviors. Our results indicate that in bar venues, online social support could reduce individuals' inhibition to share their bar-going behaviors. The soft power of social-network nudging works both ways. This is another novel finding that shows the critical role of online social support in promoting or accentuating health-related lifestyle choices.

Third, online social influence exerts its impact though the friends' visible healthy and unhealthy choices in online social networks. This influence is uniformly significant for healthy and unhealthy lifestyle behaviors. An interesting insight gained from the Zero-Inflated results was that for both healthy and unhealthy behaviors, friends' online social influence encourages people to share about their healthy and unhealthy behaviors in online social networks.

Fourth, our study revealed that the friendship attributes (strength and similarity/homophily) play different roles depending on the context and nature of behaviors. This is an unexpected and novel finding. The strength of friendship as measured by reciprocity and embeddedness can significantly increase the social influence of friends' unhealthy lifestyle behaviors related to bars. This is in line with prior studies on social influence of alcohol consumption within offline social networks in which Rosenquist et al. (2010) found that increase in social distance significantly reduces the social influence of friends. In the case of healthy lifestyle behaviors (fitness center & gym), it is homophily/similarity in friendship that exerts additional influence on people's healthy behaviors. This is in line with prior studies that found homophily significantly improves the adoption of healthy behaviors (Centola 2011). Our work shows homophily increases social influence in fast food restaurants but not bars. Thus, our findings introduce the perspective of context-dependency in studying the influence of homophily and the strength of social ties.

Fifth, our results showed that the socioeconomic status of the communities in which individuals reside has significant positive association with healthy lifestyle (fitness center &

gym) choices and negative association with unhealthy lifestyle (bar and fast food restaurant) choices. This confirmed offline study reports that environmental and socioeconomic factors influence health-related lifestyle behaviors (Phelan et al. 2004, Zenk et al. 2005, Foster and Giles-Corti 2008). Finally, the significant coefficients of activity level in online social network with almost identical coefficients in Models 2-4 and for all three choices—fitness center & gym, bar and fast food restaurant—shows that this factor contributes to the variability of online posts and must be taken into account in modeling online behaviors.

2.9. Implications

Most behavior studies in IS have focused on psychological and perceptual studies involving information technology. Our study is the first to focus on a national observational study of individuals' behaviors that have health consequences as observed on location-based social networks with the following theoretical and practical implications. Our study showed how online friends' healthy and unhealthy lifestyle behaviors cause significant changes in individuals' health-related behaviors.

2.9.1. Theoretical Contributions

This paper makes a number of novel contributions to theory and research. First, our dynamic sequential approach through capturing, extracting and integrating online social network public data and the derivation of healthy and unhealthy lifestyle behaviors opens a new avenue in observational study of health-related lifestyle behaviors at national, regional, and global levels. It demonstrates the great potential of online social networks for large-scale health studies.

Second, our theory-based Health-related Lifestyle Behavior (HLB) model provides a conceptual framework for studying online self-disclosed behaviors and the social factors that could influence them. This first attempt could motivate researchers to build on it to create a comprehensive theory of self-disclosed online health-related lifestyle behaviors. Therefore, the HLB model makes a significant contribution to theory in this respect.

Third, this work makes novel contributions in measuring social factors that are salient to online health-related lifestyle behaviors. Our measurement of favorites as representative of online social support showed that online appraisals have a nudging effect that could steer individuals equally to healthy and unhealthy lifestyle choices.

Fourth, this study has a broad implication studying strength of social ties in health-related lifestyle behaviors and by extension in other types of lifestyle behaviors. Research has reported that different types of ties and relationships are the main factors that distinguish social networks from other forms of network (Borgatti et al. 2009), and both strong and weak ties facilitate the process of information dissemination (Granovetter 1973). Our work adds to this body of work by showing that the causal influence of strength of ties depends on the context of lifestyle behaviors that have health consequences. It seems that strength of ties plays a significant role in unhealthy lifestyle behaviors that are less socially acceptable (bars as compared to fast food venues). Our work adds to the growing body of literature on the importance of context in the study of individuals' behaviors (Chen and Zahedi 2016, Hong et al. 2014).

Fifth, we addressed the divide between social influence and homophily and showed that such a divide is unwarranted. Indeed, homophily in terms of gender similarity and geographical proximity could increase the level of social influence on health-related lifestyle choices associated with fitness center & gym and fast food restaurants. Another novel research implication of this work is that the extent of such influences depends on the context and outcome clarity of behavior choices and not their healthy and unhealthy nature.

Sixth, our study showed that the online social network ecosystem does reflect the realities of socioeconomic divides. Therefore, it is possible to study different strata of people within this eco-system and investigate the forces that operate on and exacerbate such divides through negative social reinforcements online.

2.9.2. Practical and Policy Implications

The results of our study offer important practical implications. In recent years, mobile technologies and online social networks have become an inseparable part of daily life in which people share a great amount of information about their lifestyle behaviors. In 2015, it was estimated that people spend an average of 1.7 hours daily on online social networks.[13] This figure was reported to be 9 hours for teens.[14] Such pervasive reliance on online social networks, particularly for the younger generation, calls for a deeper understanding of how online social factors positively and negatively influence health-related lifestyle choices. It is, therefore, important to study whether and how online social networks could influence health-related lifestyle behaviors. Such studies require data extraction and integration approaches that go beyond the simple one-time download of posts in one social network. Observational studies of publicly accessible activities on online social networks require a well-planned, dynamic, and sequential data capture, extraction, and integration. Our study provides a first example of such an approach and provides evidence for the significant effect of online social factors in changing

[13] http://www.globalwebindex.net
[14] http://www.cnn.com/2015/11/03/health/teens-tweens-media-screen-use-report/

individuals’ health-related lifestyle choices. We showed how multiple factors of social influence and social support could alter such behaviors.

Our work presented metrics that capture social factors based on a sound theoretical foundation. These metrics could be used for the prediction of positive and negative impacts of policy initiatives. Awareness of the influence of online social factors provides personal coaches, school psychologists and government policy-making bodies with the tools to promote personalized strategies and public policies that positively influence such factors and reduce their negative roles.

Our findings showed that close friends exert additional influence in the selection of bars for lifestyle activities. Furthermore, friends’ posting of their check-ins in bars encourages individuals to do the same, hence accelerating the promotion of alcohol consumption across online social networks. Therefore, any policy for helping individuals to address alcohol abuse needs to consider the online social networks with which the individual interacts. This work emphasizes the possible role of close online friends’ behaviors in other unhealthy behaviors, such as drug abuse.

Finally, online social networks go beyond physical and cultural boundaries. Our work shows the wide reach of online social friends in changing individuals’ health-related behaviors. Online social networks could be an important channel when developing policies to deal with unhealthy behaviors or promoting healthy behaviors. Social support in the context of face-to-face support groups produces positive results. Our work shows that online social support and influences have similar consequences, which could supplement and reinforce face-to-face counseling.

2.10. Limitation and Future Research Directions

While our population of interest for this study is individuals who post location-based activities on their online social network pages, our dataset does not cover those who are not active on Twitter and Foursquare at the same time. Therefore, interpretations of our results are limited to the population who self-disclose their lifestyle behaviors on Twitter and Foursquare. Second, although location-based check-ins provide the opportunity for a nationwide data collection, caution should be exercised in using our results since check-ins are only a surrogate for actual health-related lifestyle activities and individuals may also differ in their willingness to share their location from specific types of venues leading to self-selection bias. Third, individuals' friends within the social networks were captured once. However, such networks are dynamic in nature and friends change over time. Future studies are needed to collect friends' information over time to gain a deeper understanding of the social impacts of online friendships.

In this study, we investigated social factors in relation to egocentric networks. Future studies need to investigate different types of online communities and memberships in such communities. Furthermore, our work can be extended to studying the role of online social networks in changing other behaviors, such as disclosing personal information or engaging addictive behaviors in online platforms. Furthermore, the future extension of our work could also involve investigating the structure of the network and the positions of people in the network as additional social factors. Finally, our work opens new avenues for comparative studies of peoples' behaviors across different regions, countries, and cultures. Such comparative studies could provide insights about health issues and the ways to deal with them by taking advantage of unique features of online social networks and communities.

CHAPTER 3

Essay 2: The Moderating Impact of Friends' Posted Images on Observed Healthy and Unhealthy Lifestyle Behaviors of Individuals in Online Social Networks

3.1. Introduction

Over the past few years, photo-sharing services have gain popularity. Users of Facebook alone share hundreds of millions of photos every single day.[15] The number of active users in Instagram reached 300 million users in 2016,[16] and looking at photos has replaced listening to music as the first entertainment activity on the Internet (Dutton 2013). It is shown that social interaction and self-expression are strong motives for photo sharing (Lee et al. 2015). Research argues that photographs are good for impression management and have credibility that text lacks (Marwick 2015). Photo-sharing provides a quick method for people to share their preferences, lifestyles, and behaviors with their social circles.

A recent experimental study found that most adolescent users tend to share photos that contain food items and in a majority of cases, depict foods that do not have nutrientional value (Holmberg et al. 2016). Pictures of shared photos could influence people's friends or relatives who view them. Research in neuroscience has found that food images affect appetite-related brain activities (Beaver et al. 2006) and can provoke reactions from people (Mejova et al. 2015).

---

[15] https://newsroom.fb.com/news/2016/06/introducing-360-photos-on-facebook/
[16] http://blog.business.instagram.com/post/146255227588/500m-instagram

This phenomena raises the question: can sharing photos with friends and family in online social networks influence people's behaviors? Of particular importance is the examination of its impacts on people's healthy and unhealthy lifestyle behaviors. To our knowledge, no prior study has investigated health-related behavioral impacts of shared photos in online social networks.

To address this gap, we extend the work in Essay 1 to answer the following research questions within the context of self-disclosed behaviors in online social networks, (i) Does the presence of photos moderate the impact of friends' healthy and unhealthy lifestyle behaviors? (ii) How do the contents of posted photos contribute to friends' healthy and unhealthy lifestyle behaviors?

To answer to first research question, we rely on multimodality and observational learning as the theoretical basis of our study. We argue that adding images to texts in self-disclosed posts enhances the effectiveness of communication among individuals and facilitates the process of learning from others. We modify the dynamic sequential data extraction and integration method discussed in Essay 1 to capture the photos posted along with location-based check-ins at gyms and fitness centers, bars, and fast food restaurants. We explore the effects of both posted photos and individuals' disclosed health-related lifestyle behaviors. Our results indicate that the presence of photos—regardless of content—in self-disclosed check-ins at bars and fast food restaurants increases friends' social influence over unhealthy lifestyle behaviors.

To answer the second research question, we develop a novel approach in image analysis to identify image contents. This approach combines analysis tools offered by Amazon Web Services (AWS)[17] with our newly developed tools to capture and categorize image contents within the three health-related contexts—gym and fitness center, bar, and fast food restaurant.

---

[17] https://aws.amazon.com

This approach allows us to examine the effects of various types of image content on individuals' healthy and unhealthy lifestyle behaviors. Our results show that photo content types related to each context have significant effect either on the frequency of engaging in health-related lifestyle behaviors or on individuals' decision to disclose their health-related lifestyle activities.

To our knowledge this is the first study that captures the impact of individuals' posted photos in online social networks on their friends' health-related lifestyle behaviors. This paper provides insights about the role of online social networks in formation of health behaviors and makes several important contributions. First, we offered a new approach to image analysis in identifying and categorizing images' contents. Second, we capture the additive effect of visual contents in online social networks. Another contribution is our novel method of collecting a unique dataset from Twitter and Foursquare that contains both the visual and non-visual contents of individuals' self-disclosed health-related lifestyle behaviors. Third, the results of our work uncover different pathways by which shared photos in online platforms can impact individuals' behaviors. Fourth, we add to the literature of observational learning by considering the effect of image content on the process of learning from each other.

3.2. Literature Review

3.2.1. Health-related Lifestyle Behaviors and Social Environment

The impact of social environment on individuals' health behaviors is well established in health literature (Zenk et. al 2005, Moore and Diez Roux 2006, Christakis and Fowler 2007, 2008, Naidoo and Wills 2009, Rosenquist et al. 2010). Social environment is defined as "the immediate physical surroundings, social relationships and cultural milieus in which defined groups of people function and interact" (Barnett and Casper 2001 p. 465) and it can impact individuals' health at two levels (macro- and micro- levels)[18] and from three different pathways (health behavioral, psychological, and physiological) (Berkman et al. 2000). The health-related lifestyle behavior of individuals is a prominent area that has been affected by the social environment. Health-related lifestyle behavior is a subset of the health behavioral pathway. Reports show that unhealthy lifestyle behaviors are the major contributors to the chronic diseases that pose a huge cost burden to healthcare systems (CDC 2015). In Essay 1 we defined health-related lifestyle behavior as a pattern of choices made by people from a set of available alternatives with health-consequences. Physical activity, alcohol consumption, smoking and unhealthy food diets are prominent forms of health-related lifestyle behaviors.

Social influence is a micro-level factor of social environment that may have significant impact on individuals' health-related lifestyle behaviors. In a large study of offline friends, Christakis and colleagues have observed that unhealthy behaviors, such as smoking, alcohol consumption, and obesity can spread across the social network through interaction among friends (Christakis and Fowler 2007, 2008, Rosenquist et al. 2010), leading to the argument that friends observe and mimic one another's lifestyle behaviors. Online social networks provide another type of environment for the social influence of friends. The mechanisms of social influence in online social networks differ from those in offline social networks.

Compared to offline social networks, online social networks expand the level of connections but provide lower levels of social presence and information richness for individuals (Chan and Cheng 2004). Social presence refers to "the degree of salience of the other person in an interaction" (Short et al. 1976, p. 65), and information richness is defined as "the ability of

[18] The macro level consists of cultural, political and socioeconomic factors and the micro level is formed by the psychosocial mechanisms underpinning human relationships.

information to change understanding within a time interval" (Daft and Lengel 1986, p. 560). Lower levels of social presence and information richness can negatively affect social influence in online social networks, which could differ depending on the type and nature of online platforms. Social presence and information richness in online social networks depend on how extensively platforms mediate individuals' interaction and also the structure and contents of individuals' posts (Strekalova and Krieger 2017, Bateman et al. 2017). For instance, visual contents can provide a higher level of social presence and can convey meaning faster than texts (Barry 1997). Content also plays a role in the effectiveness of the health messages. Research shows that different formats of health-related messages can affect individuals' choices and behaviors differently (Gallagher and Updegraff 2012, Rothman et al. 2006).

To our knowledge no prior study has analyzed the influence of images and their contents on individuals' lifestyle behaviors. This study addresses this gap by focusing on photos that are posted along with self-disclosed health-related lifestyle behavior in online social networks and distinguishes the effects of photos from the text-based self-disclosures studied in Essay 1.

3.2.2. Photo Sharing in Online Social Networks

"Visual imagery is central to how individuals represent themselves, make meaning, create identities, and communicate with the rest of the world" (Kenix 2013 p. 1). Research has found that the structure of visual communication is different from linguistic communication (Kress and van Leeuwen 2010). Visual contents are attention grabbing (Powell et al. 2015), reproduce reality (Messaris and Abraham 2001), heighten emotional experience (Iyer and Oldmeadow 2006), and can be more memorable (Lutz and Lutz 1977, Powell 2015). Moreover, individuals have higher levels of trust for what they can see over what they can just read about (Sundar 2008).

Over the past few years, with the growth of online photo-sharing platforms, scholars have investigated the role of images in changing the pattern of individuals' interaction with online social networks. Research has shown that posted contents with images get more feedbacks (like, shares, and comments) than those without image content (Corliss 2012). Additionally, findings show that the presence of images can increase the likelihood of clicking on the provided content within online social network sites (Ulloa et al. 2015). A large portion of literature on online photo sharing has focused on individuals' motivational factors. Research shows that multiple factors could motivate individuals to share a photo, including fulfillment of intrinsic and extrinsic needs (Nov et al. 2010), self-disclosure and self-presentation (Rui and Stefanone 2013, Sheldon and Bryant 2016), surveillance (Sheldon and Bryant 2016), impression management (Lee et al. 2015), documentation and archiving (Sheldon and Bryant 2016, Lee et al. 2015), and enjoyment (Nov et al. 2010, Nightingale 2007).

In online social networks, photo sharing serves as a personal recommendation to others—a capability that makes photos influential (Eftekhar et al. 2014). Insights about the power of images in online social networks are limited and come primarily from marketing literature. Research in marketing has reported the effect of photos on users' engagement (Shin et al. 2017), click-through rate (Jalali and Papatla 2016), and purchase intention (Kim and Lennon 2008). However, these studies have not examined the social influence of photos in the formation of individuals' lifestyle behaviors. In other words, it is not clear whether the images individuals post in online social networks have any impact on their friends' lifestyle behaviors. We aim to fill this gap in the context of healthy and unhealthy lifestyle behaviors.

3.3. Theoretical Background and Hypothesis Development

In this study, we rely on multimodality and social learning theories to build on the Health-related Lifestyle Behavior (HLB) model. Multimodality is a theory of communication which discusses the effectiveness of using combinatory modes in communication (Kress and Van Leeuwen 2001). Modes of communication consist of but are not limited to written text, gesture, posture, gaze, photo, and video. Multimodality uses different modes to generate meaning beyond the capacity of either alone (Geise and Baden 2015, O'Halloran and Smith 2012). Online social-network sites (similar to traditional media such as newspapers and TV) benefit from multimodality. However, in contrast to traditional media, online social networks are not formed based on a one-to-many principle of mass communication platforms but on a network of connections among peers in which the roles of producer and consumer constantly change (Bateman et al. 2017). In such platforms, individuals can observe each other's self-disclosed multimodal posts revealing their behaviors over time.

According to social learning theory (Bandura 1969), behavior is learned through the process of observing others' behaviors in social environments. In this process, people pay attention to what others do and try to imitate those behaviors. If people find the imitation process is rewarding, they will continue repeating the behavior. In online social networks, textual words and visual contents are the elements that can be independently used for disclosure of behaviors. However, multimodality is a factor that can contribute to the process of learning from others. According to *multimedia principle*, "people learn more deeply from words and pictures than from words alone" (Mayer 2005, p. 3). This assertion is in line with the information processing theory, which argues that the level of elaboration on a concept affects how well information is processed and learned (Craik and Lockhart 1972). We therefore add multimodality to our model of health-related lifestyle behavior in Essay 1 and analyze the effect of photos in diffusion of health-related lifestyle behaviors across the online social networks.

Research has shown that media and its type of content play an important role in attracting individuals' attention (Bucher and Schumacher 2006). Eye-tracking studies emphasized the role of visual cues and found that visual cues command a higher level of visual attention (Geise 2011, Yantis 2005). However, people have limited level of attention (Kahneman 1973). The first step in media reception is to grab the attention of observers to a stimulus (Bucher and Schumacher 2006). The study of online news readership shows that individuals' visual attention is first drawn to photos (Bucher and Schumacher 2006). Moreover, individuals spend longer duration of time looking at posted images than looking at the same size area of textual content in online social network sites (Ulloa et al. 2015).

Prior research shows that individuals' visual attention influences the level of observational learning (Yussen 1974). It has also been established that attention-grabbing contents are more likely to influence individuals (Barber and Odean 2007). Accordingly, we argue that images increase individuals' attention to friends' self-disclosed lifestyle behaviors within online social networks, thus positively moderating the observed social influence of friends in such platforms. Hence,

> **Hypothesis 1.** *Presence of images in self-disclosed lifestyle behaviors (a) increases the influence of friends' healthy lifestyle behaviors at time t on the individuals' healthy lifestyle behavior at time t+1 (b) increases the influence of friends' unhealthy lifestyle behaviors at time t on the individuals' unhealthy lifestyle behavior at time t+1*

The effect of images is not just limited to grabbing individuals' attention, but it can also play direct roles in the process of observational learning. Once the visual attention is attracted, the human brain simultaneously processes the incoming stimulus including image-based and text-based information from different channels (LaBerge and Samuels 1974). Research shows that receiving information from multiple channels improves learning and memory (Paivio 1991). In this phase, summation of cues between channels helps people to remember the information better (Severin 1967). In other words, combining of text with related visual contents provides the greatest gain in learning. A recent study found that high congruency of image and text in media play a major role in the process of learning (Powell et al. 2015). Accordingly, we expect that the content of disclosed images from health-related venues be another source of social influence in online social networks, in which congruency between posted images and type of disclosed behavior positively impact friends' health-related lifestyle behaviors over time. By congruence, we refer to whether the shared images match the textual content describing the healthy and unhealthy lifestyle behaviors of individuals. Research has shown that repeating the same message over different channels helps recipients to better understanding the message (Lane 2000). Hence,

> **Hypothesis 2.** *Individuals' (a) healthy lifestyle behaviors at time t+1, are positively influenced by context-congruent images posted along with friends' disclosed healthy lifestyle behaviors at time t (b) unhealthy lifestyle behaviors at time t+1, are positively influenced by context-congruent images posted along with friends' disclosed unhealthy lifestyle behaviors at time t.*

3.4. Data Collection

To collect both individuals' health-related lifestyle behavior and any posted photos, we used the dynamic sequential data extraction and integration method described in Essay 1 with some modifications to meet the data requirements of this study. We modified the three-phase procedure in the dynamic sequential data extraction and integration method. In Phase 1, we identified active users in U.S. who have posted at least one check-in every two weeks after their first check-in in the time period of January 28–April 22, 2014. Of the total collected data, 32,700 unique individuals met this requirement. In Phase 2, we continued the procedure of data collection from active users for two sequential four-week time periods (time *t* and *t+1*). In this phase, in addition to general check-in information, we collected image URLs posted along with tweets using Twitter API. Finally, in Phase 3, we collected complementary information such as the number of received favorite counts and information about venues. The final captured dataset contains more than 5 million check-in tweets, more than 100 thousand images, and 1,127,420 distinct U.S. venues.

In order to analyze the effect of posted images on individuals health-related lifestyle behaviors, we focused on check-ins that represent individuals' health-related lifestyle behaviors at time period *t* (April 22 to May 20, 2014) and time *t+1* (May 20 to June 17, 2014). In doing so, we relied on venue type as the proxy and combined the salient categories in Foursquare to identify three types of venues associated with health-related lifestyle behaviors: *fitness center & gym*, *bar*, and *fast food restaurant*. Table 3.1 lists the Foursquare categories and number of venues in each type.

Table 3.1. List of Categories

Venue Type	Foursquare Categories	# of Venues
Fitness Center & Gym	Badminton Court, Baseball Field, Basketball Court, Boxing Gym, Climbing Gym, College Basketball Court, College Cricket Pitch, College Football Field, College Gym, College Hockey Rink, College Soccer Field, College Tennis Court, Cricket Ground, Gym, Gym / Fitness Center, Gym Pool, Gymnastics Gym, Hockey Field, Paintball Field, Rock Climbing Spot, Roller Rink, Rugby Pitch, Skate Park, Skating Rink, Soccer Field, Sports Club, Squash Court, Swim School, Tennis Court, Volleyball Court, Yoga Studio	36,047
Bar	Apres Ski Bar, Bar, Beach Bar, Beer Garden, Beer Store, Champagne Bar, Cocktail Bar, Dive Bar, Gastropub, Gay Bar, Hookah Bar, Hotel Bar, Irish Pub, Karaoke Bar, Piano Bar, Pub, Sake Bar, Sports Bar, Whisky Bar, Wine Bar	66,687
Fast Food Restaurant	BBQ Joint, Fast Food Restaurant, Food Court, Fried Chicken Joint, Hot Dog Joint, Mac & Cheese Joint, Pizza Place, Wings Joint	109,575

3.5. Image Analysis: Extraction, Processing, Dimensionality Reduction, and Categorization

To capture the effect of posted images on friends' healthy and unhealthy lifestyle behaviors, we extracted, processed and categorized posted images at time period *t* using a number of tools in four distinct phases. At Phase 1 (image extraction), posted images along with health-related lifestyle check-ins at time period *t* have been extracted through image URLs. From the total number of check-ins, 5.1% of check-ins at fitness center & gym venues, 8.3% of check-ins at bars, and 7.1% of check-ins at fast food restaurants had images in addition to textual content.

At Phase 2 (image processing), we used services inside the AWS (Amazon Web Services) to process the images extracted at the first phase. AWS is a cloud-based platform that offers various computational analysis, data management and web development services. In order to perform the image processing task, we initially transferred images into Buckets inside the AWS platform. Buckets are cloud-based logical units of storage that can be used for analysis of data using available services at AWS. Then, we used Rekognition API to perform the image processing task. Rekognition is a deep learning technology developed by Amazon's computer vision scientists to analyze images and videos. The API detects objects inside visual objects and reports them with associated confidence values. A confidence value shows the probability that an object exists inside a photo. Figure 3.1 shows an example of detected objects for a sample photo in our dataset.

Object Label	Confidence
Fries	98.78 %
Food	98.78 %
Ketchup	94.20 %
Seasoning	94.20 %
Aluminium	78.26 %
Tin	78.26 %
Can	78.26 %
Meal	73.53 %
Plate	60.65 %
Dish	60.65 %
Bowl	58.10 %
Taco	54.95 %
Beverage	51.59 %
Drink	51.59 %
Salad	51.14 %
Meat Loaf	50.58 %

Figure 3.1. An Example of Analyzed Photo by Rekognition API

Phase 3 (dimensionality reduction) reduced the dimensionality of detected labels by identifying the granularity of objects in images, and then clustering and categorizing objects within images. We rely on Rekognition API, which provides labels at different levels of granularity. A high-granular label defines what an object exactly is, and a low-granular label shows the type of an object without exactly specifying that object (i.e. “Fries” is a high-granular label, and “Food” is a low-granular label). Each label at a high level of granularity is associated with one or more labels at lower levels of granularity and always comes together in the list of detected objects (i.e. “Fries” is associated with “Food” and “Meal”).

This phase involves a two-step procedure: (Step 1) label granularity identification and filtering, and (Step 2) label clustering. At Step 1 (of Phase 3), we identified the level of granularity by applying association rule mining techniques (Agrawal et al. 1993). Association rule mining is a popular technique to study transactional data where identified rules can show how two different item sets are associated with each other in large number of transactions.

Accordingly, each rule is composed by two different sets of items and can be represented in the format of $A \rightarrow B$, in which A is the antecedent item and B is the consequence item. In such rules, existence of items of set A inside a transaction list increases the chance having items of set B inside the same transaction list. In our context, we use the rule mining techniques to capture the pairwise association of labels inside images. Association rules can be at different levels of accuracy. Thus, in order to find meaningful rules, we used Confidence and Support constraints as two main factors in the selection of association rules (Klemettinen et al. 1994). These two constraints are computed as follows:

$$Confidence(A \rightarrow B) = \frac{|A \cap B|}{|A|}$$

$$Support(A, B) = \frac{|A \cap B|}{M}$$

Where *A* and *B* are two different object labels, $|A|$ is the number of images with label A, $|A \cap B|$ is the number of images with *A* and *B*, and *M* is the total number of images in a dataset. Confidence and Support are two values between 0 and 1. Confidence shows the association of two labels, where 0 shows no association and 1 indicates the highest level of association between two labels. Support is another metric that shows how frequently two labels appear in images with each other. In this study, we consider association rules that have a Confidence value of 1 and a value more than 0.05 for its Support constraint. The confidence value of 1 indicates that label A (label with higher granularity) is always be presented by label B (label with lower granularity). Support value has also been used to eliminate the rules with low frequency values. **Appendix F** shows labels at different levels of granularity and the pattern of association among labels for each health-related lifestyle behavior separately. At the end, we reduced the assigned labels of images by limiting the labels to those at the lowest level of granularity.

At Step 2 (of Phase 3), we aggregated labels at the lowest level of granularity using their similarity in describing images. In doing that, we applied a three-step sequential method: (Step 2.1) computing of labels' similarity matrix, (Step 2.2) applying a modularity-based clustering method, and (Step 2.3) assigning descriptive labels to clusters. At Step 2.1, in order to compute the similarity matrix, we measured similarity of labels by using the Jaccard index. Jaccard index is a pairwise similarity method that can be computed as follows:

$$Jaccard(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

Where A and B are two different object labels at the lowest level of granularity that describe a series of images, $|A \cap B|$ is the number of images that can be described by both A and B, $|A \cup B|$ is the number of images that can be described by either A or B, $|A|$ is the number of images with label A, and $|B|$ is the number of images with label B. The Jaccard index always returns a value between 0 and 1. Computation of a similarity index across labels helps us to form similarity matrices of labels–one similarity matrix for each health-related lifestyle behavior. Figure 3.2 illustrates the process of generating a similarity matrix.

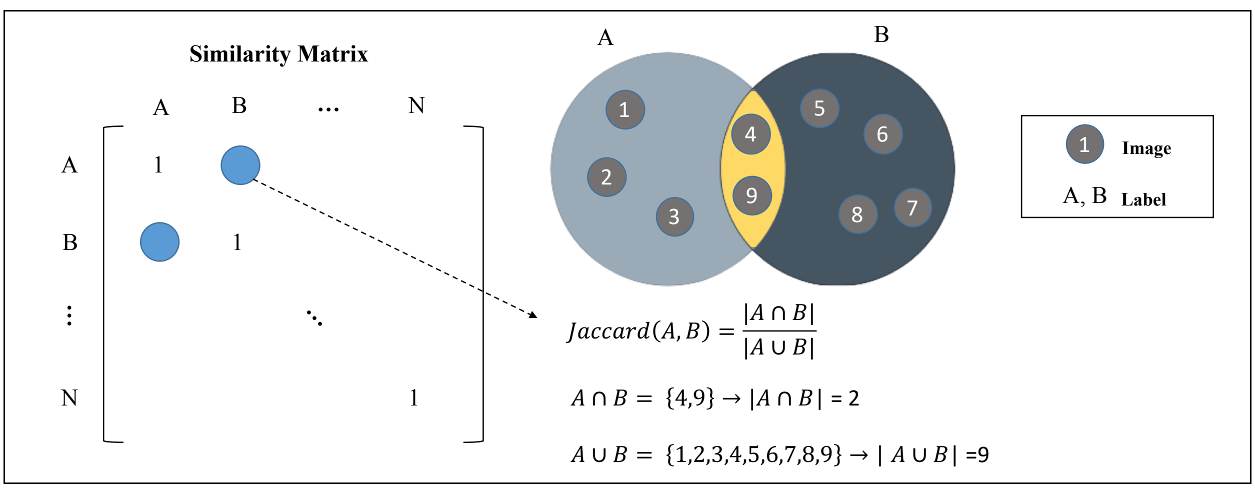


Figure 3.2. Formation of Similarity Matrix

At Step 2.2, we applied the Louvain clustering method (Blondel et al. 2008) to put labels into different clusters. Louvian is a modularity-based optimization method that identifies clusters by computing the deviation of a similarity matrix from a randomly generated similarity matrix. This process put the labels for Fitness Center & Gym types of venue into seven clusters, labels for Bar venues into nine clusters, and labels for Fast Food Restaurant into eight clusters. At Step 2.3, we assigned a unique label to each cluster. Table 3.2 shows detected clusters along with their assigned labels for all the three types of health-related lifestyle behaviors.

Table 3.2. List of Object Clusters

Venue Type	Clusters	Labels at the Lowest Level of Granularity
Fitness Center & Gym	Flora	Flora, Jar, Pottery
	Food & Beverage	Beverage, Bowl, Food
	Human	Clothing, Glasses, Head, Human
	Indoors	Electronics, Flooring, Furniture, Indoors, Screen, Wood
	Outdoors	Animal, Asphalt, Bench, Billboard, Building, Field, Lighting, Machine, Nature, Outdoors, Park, Path, Road, Soil, Terminal, Transportation, Urban, Water
	Sport	Sport
	Text	Emblem, Paper, Poster, Sign, Text, Trademark, Word
Bar	Beverage	Aluminium, Beverage, Bottle, Cup, Glass
	Club	Club, Leisure Activities, Light, Lighting, Night Life, Stage
	Decor	Accessories, Art, Home Decor, Ornament
	Flora	Flora, Jar, Pottery
	Food	Bowl, Food
	Human	Clothing, Hair, Head, Human
	Indoors	Bench, Crypt, Electrical Device, Electronics, Furniture, Indoors, Market, Pub, Restaurant, Screen, Shelf, Shop, Wood
	Outdoors	Animal, Asphalt, Billboard, Brick, Building, Canopy, Nature, Outdoors, Path, Pier, Road, Soil, Terminal, Transportation, Urban, Water
	Text	Blackboard, Book, Emblem, Paper, Poster, Text, Trademark, Word
Fast Food Restaurant	Beverage	Aluminium, Beverage, Bottle, Cup, Glass
	Decor	Art, Home Decor, Ornament
	Flora	Flora, Jar, Pottery
	Food	Bowl, Food
	Human	Accessories, Clothing, Glasses, Hair, Head, Human, Leisure Activities
	Indoors	Bench, Electrical Device, Electronics, Furniture, Indoors, Lighting, Market, Night Life, Pub, Restaurant, Screen, Shelf, Shop, Wood
	Outdoors	Animal, Asphalt, Brick, Building, Canopy, Machine, Nature, Outdoors, Parking Lot, Path, Road, Terminal, Transportation, Urban
	Text	Book, Emblem, Paper, Poster, Sign, Text, Trademark, Word

At phase 4 (image categorization), each image got one or more cluster labels based on the types of object label originally identified by the Rekognition API. For instance, the depicted image in Figure 3.1 gets two cluster labels of Food and Beverage. Figure 3.3 shows the process of image analysis in detail.

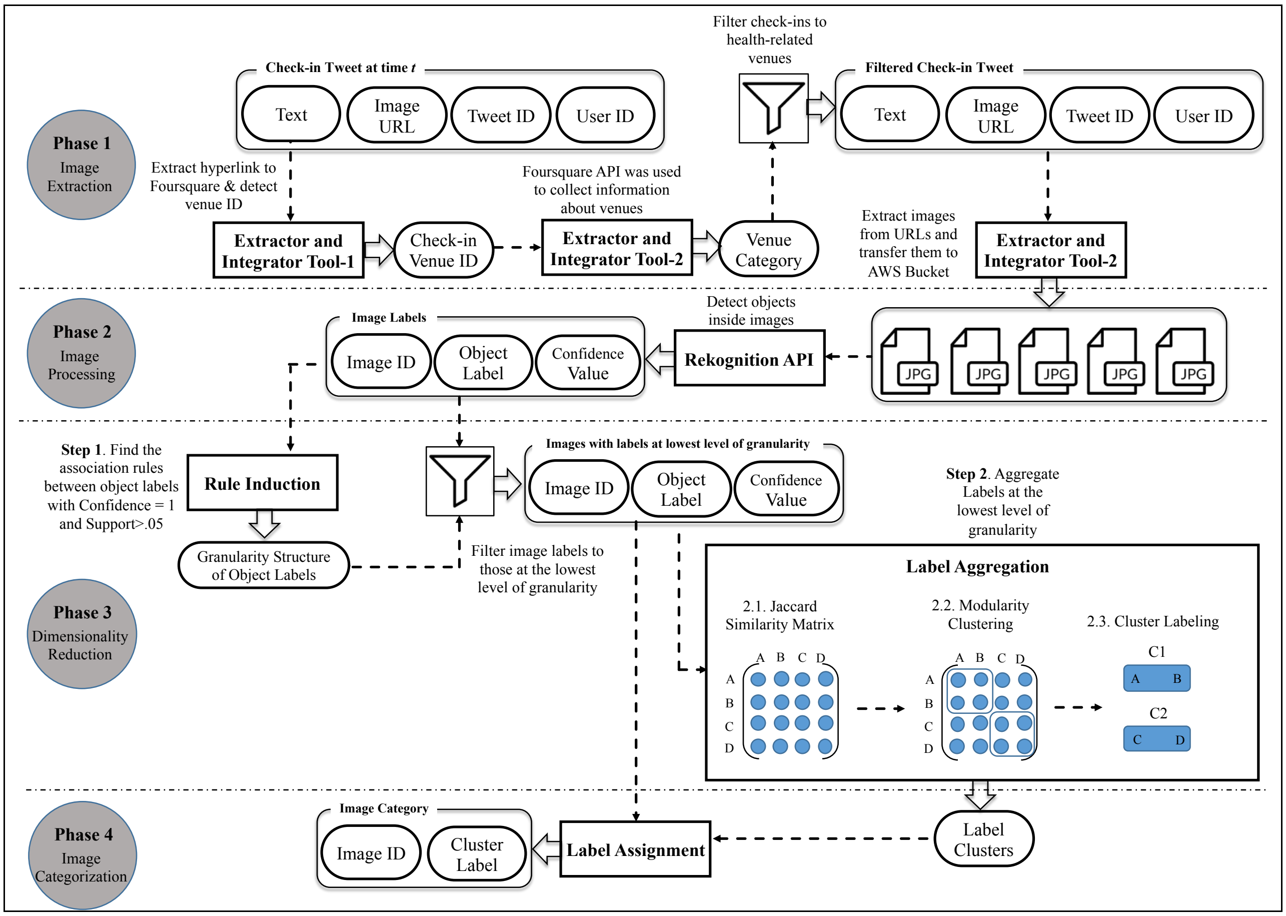


Figure 3.3. Image Analysis Process

To check the accuracy of assigned categories, we randomly selected 100 images and manually labeled them using the clustered labels. The comparison of assigned categories and manual labeled categories shows an accuracy of 95 percent in our categorization of images.

3.6. Variable Measurements

In this study, we followed the measurement method described in Essay 1 and measured individual's healthy and unhealthy lifestyle behaviors by the number of days that each individual had check-ins within each type of health-related venue (fitness center & gym for healthy and bar and fast food restaurant for unhealthy lifestyle behaviors) in two sequential four-week time periods. We also measured social support and social influence variables (variables in the original HLB model) using the same procedures described in Essay 1. In the following sections, we describe how we measured the social influence variables related to images.

3.6.1. Social Influence Moderation: Image Presence

To capture the effect of friends' posted images on social influence, we computed the number of images posted along with friends' health-related lifestyle check-ins at time period t. To compute this value in the fitness center & gym context, we considered the directional egocentric network of users inside Twitter. The egocentric network demonstrates the friendship network of a single individual (ego) within a large social network. In a directional egocentric network, individuals can only observe activities of people who are directly followed by them. Thus, for each individual we counted the number of observable images that are posted along with friends' check-ins at fitness center & gym venues at time period t. Figure 3.4 illustrates this concept. We independently repeated this process for bar and fast food restaurant venues. Later, in our model estimation, we consider the interaction of the number of images variable and the social influence variable in each health-related lifestyle context to capture the moderation effect of posted images.

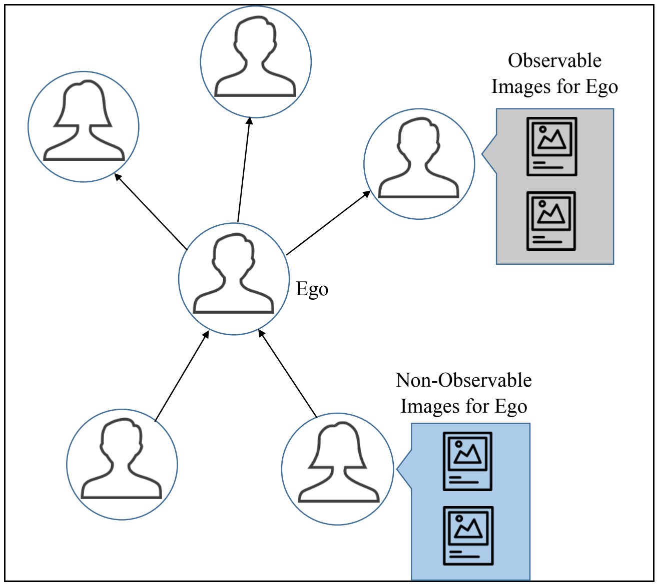


Figure 3.4. Egocentric Network

3.6.2. Social Influence: Image Types

To analyze the social influence of congruent images posted along with friends' disclosed health-related lifestyle behaviors, we relied on the cluster labels identified in our image analysis process. Those labels capture repeated types of objects within posted images in different health-related lifestyle contexts. Accordingly, in the context of fitness center & gym venues, we computed the number of friends' observable images at time period *t* that had been labeled by each of the Flora, Food & Beverage, Human, Indoors, Outdoors, Sport, and Text labels separately. This computation provided seven unique values representing the number of images posted by friends with specified labels. We repeated this computation for bar and fast food restaurant venues using their own detected cluster labels.

3.6.3. Control Variables

In this study we control for the number of friends who have posted images at time period *t* since the effect of observed images on individuals could be different when they have been observed through the posts of few versus large numbers of friends. To measure this variable in the context of fitness center & gym, we considered each individual's directional egocentric network and computed the number of friends who have posted images along with their check-in at fitness center & gym type of venues. We also repeated this computation for bar and fast food restaurant venues. In our model, we also controlled for the two control variables in original HLB model. Table 3.3 summarizes the measurement of the variables in our model – both original HLB variables and variables in this study.

Table 3.3 Variable Measurements at Individual Level

Model Variable	Definition	Metric and Computation
Dependent Variables		
Individual's healthy lifestyle behavior at time *t+1*	Lifestyle behaviors that promote health	Individual's total number of days with check-ins at fitness center & gym venues at time *t+1*
Individual's unhealthy lifestyle behavior at time *t+1*	Lifestyle behaviors that inhibit health	Individuals' total number days with check-ins at time *t+1* measured for two venue types separately: 1. Bar. 2. Fast food restaurant.
Original HLB Model Independent Variables, all lagged to measure impacts		
Online social support healthy (or unhealthy) lifestyle behaviors at time *t*	The support provided via feedback in online social networks for individuals' healthy (or unhealthy) lifestyle behaviors	1. For healthy lifestyle: Average number of Favorites an individual receives for check-ins at fitness center & gym venues at t, computed as the sum all Favorite counts received for fitness center & gym check-ins divided by number of days with fitness center & gym check-ins. 2. For unhealthy lifestyle: the fitness counts are replaced once by bar counts and again by fast food restaurant counts in the above computation. All computed at time *t*.
Social influence of friends' healthy (unhealthy) lifestyle behaviors at time *t*	The influence of friends' engagement in the same healthy (unhealthy) lifestyle behaviors	1. For healthy lifestyle: the average of friends' number of days with fitness center & gym check-ins at time t, computed as: sum of all friends' number of days with fitness center & gym check-ins divided by number of friends. 2. For unhealthy lifestyle: the fitness check-ins are replaced once by bar check-ins and again by fast food restaurant check-ins in the above computation. All computed at time *t*.
Social influence of strong ties' healthy (unhealthy) lifestyle behaviors at time *t*	The impact of strong friendship ties in social influences of friends' engagement in the same healthy (unhealthy) lifestyle behaviors	1. For healthy lifestyle: the ratio of weighted average strong ties' number of days with check-ins at fitness center & gym venues divided by non-weighted average of friends' number of days with check-ins at fitness center & gym venues. This ratio is computed for the first measurement period. The weights for strong ties are computed as follows: (i) Non-reciprocated friends' lifestyle behavior gets no weight, (ii) Reciprocated friends' lifestyle behavior gets weight proportional to the number of common friends with focal individual. 2. For unhealthy lifestyle: the fitness check-ins are replaced once by bar check-ins and again by fast food restaurant check-ins in the above computation. All computed at time *t*.
Social influence of similar friends' healthy (unhealthy) lifestyle behaviors at	The impact of similarity with friends/homophily in social influences of	1. For healthy lifestyle: The ratio of weighted average of similar friends' number of days with check-ins at fitness center & gym venues divided by non-weighted average of friends' number of days with check-ins at fitness center & gym venues. This ratio is

time *t*	friends' engagement in the same healthy (unhealthy) lifestyle behaviors	computed for the first measurement period. The weights for similar friends are computed as follows: (i) Friends get .5 similarity score if they reside in 0-10 miles of the focal individual (ii) Friends get .5 similarity score if they have similar gender as focal individual (iii) Friends' lifestyle behavior gets weight proportional to the final calculated similarity score. 2. For unhealthy lifestyle: the fitness check-ins are replaced once by bar check-ins and again by fast food restaurant check-ins in the above computation. All computed at time *t*.
Independent Variables for The Present Study, all lagged to measure impacts		
Moderation effect of images on social influence at time *t*	The moderation effect of image presence in friends' disclosed healthy (unhealthy) lifestyle behaviors on the level of social influence	1. For healthy lifestyle: Number of images posted by friends along with their check-ins at fitness center & gym venues at time *t* multiplied by the healthy social influence variable measured at time *t*. 2. For unhealthy lifestyle: the fitness check-ins are replaced once by bar check-ins and multiplied by unhealthy(bar) social influence variable and again by fast food restaurant check-ins and multiplied by unhealthy (fast food restaurant) social influence variable. All computed at time *t*.
Social influence of images posted along with friends' healthy (unhealthy) lifestyle behaviors at time *t*	The influence of different type of images posted along with friends' healthy (unhealthy) lifestyle behaviors.	1. For healthy lifestyle: first, posted images at time *t* have been labeled by one or more of the following labels: Flora, Food & Beverage, Human, Indoors, Outdoors, Sport, and Text. Second, the number of friends' posted images containing each of the above-mentioned labels has been counted separately. That generates 7 different variables representing friends' posted image types at time *t*. 2. For unhealthy lifestyle: the labels were replaced with associated unhealthy lifestyle behaviors which are Beverage, Club, Décor, Flora, Food, Human, Indoors, Outdoors, and Text for bar venues and Beverage, Décor, Flora, Food, Human, Indoors, Outdoors, and Text for fast food restaurant venues. Then we repeat the above computation for bar and fast food restaurant separately. That generates 9 different variables representing friends' type of posted images at bar and 8 different variables representing friends' type of posted images at fast food restaurants. All computed at time *t*.
***Control Variable*s**		
Activity level in social network at time *t+1*	Activity level in online social network at time *t+1*	Individuals total number of check-ins in online social network at time *t+1*
Individuals' healthy (unhealthy) lifestyle behavior at time *t*	Individuals' healthy (unhealthy) lifestyle behavior at time *t*	Individuals' healthy (unhealthy) lifestyle behavior at time *t*
Number of friends posted images along with healthy lifestyle behavior at time *t*	Number of friends who have posted images along with healthy (unhealthy) lifestyle behavior at time *t*	1. For healthy lifestyle: Number of friends who have posted at least one image with their check-ins at fitness center & gym venues at time *t* 2. For unhealthy lifestyle: the fitness check-ins are replaced once by bar check-ins and again by fast food restaurant check-ins in the above computation. All computed at time *t*

3.7. Data Analysis and Model Estimation

The distributions of individual check-ins at different health-related venues in the second measurement time period (*t+1*) are presented in **Appendix E**. The distributions of check-ins show over-dispersion with high occurrence of zeros at this time period. We confirmed over-dispersion using the alpha test (Cameron and Trivedi 1990). Over-dispersion occurs when the variance in data is greater than the mean. As described in Essay 1, zero-inflated negative binomial regression (ZINB) is a suitable method for estimation of over-dispersed count data with a high occurrence of zeros.

ZINB undertakes two distinct processes in formation of data which both can lead to observation of zeros. In the first process, individuals – Group A – disclose their health-related lifestyle behaviors as they occur in their real life. For these individuals, the dependent variable – number of check-ins at time period *t+1*–is greater than or equal to zero ($y_i \geq 0$). For this group of individuals, zero values indicate that individuals have not gone to the captured type of venue. ZINB assumes negative binomial distribution in the first process. The second process relates to people – Group B – who have gone to the captured type of venue but have not reported it in their online social networks. The dependent variable for this group of people are naturally equal to zero ($y_i = 0$). ZINB considers a probability value (φ) to distinguish between each different process by which the data is generated. The following equation represents this concept:

$$y_{i,t+1} \sim \begin{cases} 0 & with\, probability\, of\; \varphi_{i,t+1} \\ f\left(y_{i,t+1} \middle| X_{i,t}\right) & with\, probability\, of\; 1 - \varphi_{i,t+1} \end{cases}$$

Accordingly, the probability of y_i number of check-ins at the captured venues is equal to:

$$P(Y_i = y_{i,t+1} | X_{i,t}, Z_{i,t}) = \begin{cases} \varphi(\gamma Z_{i,t}) + [1 - \varphi(\gamma Z_{i,t})] f(0 | X_{i,t}) & if \ y_{i,t+1} = 0 \\ [1 - \varphi(\gamma Z_{i,t})] f(y_{i,t+1} | X_{i,t}) & if \ y_{i,t+1} > 0 \end{cases}$$

where $X_{i,t}$ is the vector of independent variables for individual i at time period t, $Z_{i,t}$ is a vector of covariates measured at time period t that contribute to not reporting corresponding health-related behavior at time period t+1, and γ is the vector of estimated zero-inflated coefficients. We used R for the model estimation (R Development Core Team 2016).

3.7.1. Estimation Results

We test our hypothesizes using the ZINB method. The models were independently estimated for each health-related lifestyle behavior. The result of estimation models is presented in Tables 3.4-3.6.[19] The Group A/Count part of the tables shows the estimated coefficient for the group of people who disclose their captured lifestyle behavior in online social networks. We use the estimated coefficients in these parts of the tables to test our hypothesizes. The Group B/Zero Inflated part of the tables provides additional insight about people who have not reported the captured health-related lifestyle behaviors in online social networks. The estimated coefficients in this part represent factors that can contribute to the inhibition against disclosing health-related lifestyle behaviors.

[19] To capture the effect of strong ties and homophily on the individuals' healthy and unhealthy lifestyle behaviors in original theHLB model, we had to consider only those individuals who have friends with non-zero average value for the same health-related lifestyle behavior. That reduced the number of observations in each model, as reported in Tables 3.4-3.6.

Table 3.4. Estimated Model for Healthy Lifestyle Behaviors (Fitness Center & Gym)

	Model 1	**Model 2**	**Model 3**
Group A/Count			
Control Variables			
Activity level in online social network @ t+1	**.002*****	**.002*****	**.002*****
Individual's healthy lifestyle behavior @ t	**.185*****	**.185*****	**.185*****
Number of friends posted images along with healthy lifestyle behavior @ *t*		-.010	**-.048****
Online social support: healthy lifestyle behavior @ t	**.142*****	**.144****	**.142*****
Social Influence – Friends' Observed Behavior			
Friends' healthy lifestyle behavior @ t	**.026*****	**.025*****	**.026*****
Ratio of strong ties' healthy lifestyle behavior @ t	.016	.017	.015
Ratio of similar friends' healthy lifestyle behavior @ t	**.048****	**.049****	**.048****
Social Influence Moderation– Friends' Posted Images			
Friends' healthy lifestyle behavior @ t × Number of friends' posted images @ t		-.001	
Social Influence – Image Types			
'Flora' images posted along with friends' healthy lifestyle behaviors @ t			.025
'Food&Beverage' images posted along with friends' healthy lifestyle behaviors @ t			.101
'Human' images posted along with friends' healthy lifestyle behaviors @ t			-.009
'Indoors' images posted along with friends' healthy lifestyle behaviors @ t			.051
'Outdoors' images posted along with friends' healthy lifestyle behaviors @ t			.006
'Sport' images posted along with friends' healthy lifestyle behaviors @ t			**.063***
'Text' images posted along with friends' healthy lifestyle behaviors @ t			-.003
Constant	**-.160*****	**-.153*****	**-.148*****
Group B/Zero Inflated			
Control Variables			
Activity level in online social network @ t+1	**-.083*****	**-.083*****	**-.083*****
Individual's healthy lifestyle behavior @ t	**-1.398*****	**-1.391*****	**-1.397*****
Number of friends posted images along with healthy lifestyle behavior @ *t*		.045	-.011
Online social support: healthy lifestyle behaviors @ t	-.0001	-.011	-.013
Social Influence – Friends' Observed Behavior			
Friends' healthy lifestyle behavior @ t	**-.039****	-.028	**-.037****
Ratio of strong ties' healthy lifestyle behavior @ t	**-.081****	**-.079****	**-.082****
Ratio of similar friends' healthy lifestyle behaviors @ t	-.015	-.011	-.014
Social Influence Moderation– Friends' Posted Images			
Friends' healthy lifestyle behavior @ t × Number of friends' posted images @ t		**-.035****	
Social Influence – Image Types			
'Flora' images posted along with friends' healthy lifestyle behaviors @ t			-.194
'Food&Beverage' images posted along with friends' healthy lifestyle behaviors @ t			.138
'Human' images posted along with friends' healthy lifestyle behaviors @ t			.029
'Indoors' images posted along with friends' healthy lifestyle behaviors @ t			-.096
'Outdoors' images posted along with friends' healthy lifestyle behaviors @ t			-.043
'Sport' images posted along with friends' healthy lifestyle behaviors @ t			.066
'Text' images posted along with friends' healthy lifestyle behaviors @ t			-.102
Constant	**2.039*****	**2.027*****	**2.054*****
Log Likelihood	-21,499	-21,496	-21,487
Wald x^2	2,374***	2,390***	2,398***

N=22,423; *p<.05; **p<.01; ***p<.001.

Table 3.5. Estimated Model for Unhealthy Lifestyle Behaviors (Bar)

	Model 1	Model 2	Model 3
Group A/Count			
Control Variables			
Activity level in online social network @ t+1	**.005*****	**.005*****	**.005*****
Individual's unhealthy lifestyle behavior @ t	**.175*****	**.175*****	**.175*****
Number of friends posted images along with unhealthy lifestyle behavior @ *t*		**-.009*****	-.005
Online social support: unhealthy lifestyle behavior @ t	**.104*****	**.104*****	**.105*****
Social Influence – Friends' Observed Behavior			
Friends' unhealthy lifestyle behavior @ t	**.039*****	**.034*****	**.038*****
Ratio of strong ties' unhealthy lifestyle behavior @ t	**.052*****	**.050*****	**.050*****
Ratio of similar friends' unhealthy lifestyle behavior @ t	.020	0.019	.020
Social Influence Moderation– Friends' Posted Images			
Friends' unhealthy lifestyle behavior @ t × Number of friends' posted images @ t		**0.003*****	
Social Influence – Image Types			
'Beverage' images posted along with friends' unhealthy lifestyle behaviors @ t			.008
'Club' images posted along with friends' unhealthy lifestyle behaviors @ t			.017
'Decor' images posted along with friends' unhealthy lifestyle behaviors @ t			.045
'Flora' images posted along with friends' unhealthy lifestyle behaviors @ t			-.021
'Food' images posted along with friends' unhealthy lifestyle behaviors @ t			-.015
'Human' images posted along with friends' unhealthy lifestyle behaviors @ t			.000
'Indoors' images posted along with friends' unhealthy lifestyle behaviors @ t			-.002
'Outdoors' images posted along with friends' unhealthy lifestyle behaviors @ t			.009
'Text' images posted along with friends' unhealthy lifestyle behaviors @ t			.006
Constant	**-.133*****	**-.122*****	**-.130*****
Group B/Zero Inflated			
Control Variables			
Activity level in online social network @ t+1	**-1.072*****	**-1.070*****	**-1.070*****
Individual's unhealthy lifestyle behavior @ t	**-.101*****	**-.101*****	**-.101*****
Number of friends posted images along with unhealthy lifestyle behavior @ *t*		.035	.011
Online social support: unhealthy lifestyle behaviors @ t	**-.168***	**-.175***	**-.172***
Social Influence – Friends' Observed Behavior			
Friends' unhealthy lifestyle behavior @ t	**-.059****	**-.050***	**-.054****
Ratio of strong ties' unhealthy lifestyle behavior @ t	.000	.001	.000
Ratio of similar friends' unhealthy lifestyle behaviors @ t	-.092	-.097	-.095
Social Influence Moderation– Friends' Posted Images			
Friends' unhealthy lifestyle behavior @ t × Number of friends' posted images @ t		-.006	
Social Influence – Image Types			
'Beverage' images posted along with friends' unhealthy lifestyle behaviors @ t			**-.093****
'Club' images posted along with friends' unhealthy lifestyle behaviors @ t			-.114
'Decor' images posted along with friends' unhealthy lifestyle behaviors @ t			.249
'Flora' images posted along with friends' unhealthy lifestyle behaviors @ t			-.020
'Food' images posted along with friends' unhealthy lifestyle behaviors @ t			.044
'Human' images posted along with friends' unhealthy lifestyle behaviors @ t			**.152*****
'Indoors' images posted along with friends' unhealthy lifestyle behaviors @ t			.006
'Outdoors' images posted along with friends' unhealthy lifestyle behaviors @ t			-.044
'Text' images posted along with friends' unhealthy lifestyle behaviors @ t			**-.197****
Constant	**3.270*****	**3.243*****	**3.244*****
Log Likelihood	-39,347	-39,339	-39,330
Wald x^2	6,990***	7,007***	7,017***

N= 28594; *p<.05; **p<.01; ***p<.001

Table 3.6. Estimated Model for Unhealthy Lifestyle Behaviors (Fast Food Restaurant)

	Model 1	Model 2	Model 3
Group A/Count			
Control Variables			
Activity level in online social network @ t+1	**.009******	**.009******	**.009******
Individual's unhealthy lifestyle behavior @ t	**.154****	**.154****	**.154****
Number of friends posted images along with unhealthy lifestyle behavior @ *t*		**-.016******	**-.040******
Online social support: unhealthy lifestyle behavior @ t	**.051****	**.056******	**.051****
Social Influence – Friends' Observed Behavior			
Friends' unhealthy lifestyle behavior @ t	**.017******	**.010****	**.013****
Ratio of strong ties' unhealthy lifestyle behavior @ t	-.011	-.010	-.010
Ratio of similar friends' unhealthy lifestyle behavior @ t	**.028****	**.029****	**.029****
Social Influence Moderation– Friends' Posted Images			
Friends' unhealthy lifestyle behavior @ t × Number of friends' posted images @ t		**.004******	
Social Influence – Image Types			
'Beverage' images posted along with friends' unhealthy lifestyle behaviors @ t			.018
'Decor' images posted along with friends' unhealthy lifestyle behaviors @ t			-.008
'Flora' images posted along with friends' unhealthy lifestyle behaviors @ t			.011
'Food' images posted along with friends' unhealthy lifestyle behaviors @ t			**.042******
'Human' images posted along with friends' unhealthy lifestyle behaviors @ t			-.001
'Indoors' images posted along with friends' unhealthy lifestyle behaviors @ t			.010
'Outdoors' images posted along with friends' unhealthy lifestyle behaviors @ t			.017
'Text' images posted along with friends' unhealthy lifestyle behaviors @ t			-.019
Constant	**-.168******	**-.154******	**-.159******
Group B/Zero Inflated			
Control Variables			
Activity level in online social network @ t+1	**-.606******	**-.605******	**-.615******
Individual's unhealthy lifestyle behavior @ t	**-.081******	**-.081******	**-.081******
Number of friends posted images along with unhealthy lifestyle behavior @ *t*		.010	-.064
Online social support: unhealthy lifestyle behaviors @ t	-.042	-.045	-.050
Social Influence – Friends' Observed Behavior			
Friends' unhealthy lifestyle behavior @ t	**-.098******	**-.100******	**-.102******
Ratio of strong ties' unhealthy lifestyle behavior @ t	-.021	-.023	.062
Ratio of similar friends' unhealthy lifestyle behaviors @ t	.052	.050	.074
Social Influence Moderation– Friends' Posted Images			
Friends' unhealthy lifestyle behavior @ t × Number of friends' posted images @ t		-.000	
Social Influence – Image Types			
'Beverage' images posted along with friends' unhealthy lifestyle behaviors @ t			.036
'Decor' images posted along with friends' unhealthy lifestyle behaviors @ t			.088
'Flora' images posted along with friends' unhealthy lifestyle behaviors @ t			**.319****
'Food' images posted along with friends' unhealthy lifestyle behaviors @ t			.030
'Human' images posted along with friends' unhealthy lifestyle behaviors @ t			-.043
'Indoors' images posted along with friends' unhealthy lifestyle behaviors @ t			.118
'Outdoors' images posted along with friends' unhealthy lifestyle behaviors @ t			-.086
'Text' images posted along with friends' unhealthy lifestyle behaviors @ t			-.040
Constant	**2.714******	**2.711******	**2.732******
Log Likelihood	-32,501	-32,492	-32,478
Wald x^2	4,652***	4,668***	4,695***

N=27,253; *p<.05; **p<.01; ***p<.001

In Tables 3.4-3.6, Model 1 represents the coefficients for the original HLB model as it was captured in Essay 1. Model 2 tests the moderation effect of image presence on the social influence. Finally, Model 3 was used to capture the effect of different image types and tests for the influence of congruent images.

Group A/Count: In H1(a/b), we hypothesized that presence of images in self-disclosed lifestyle behaviors increases the influence of friends' healthy/unhealthy lifestyle behaviors at time *t* on the individuals' healthy/unhealthy lifestyle behavior at time *t+1*. H1a was not supported in Model 2 for healthy lifestyle behaviors (fitness center & gym check-ins). However, H1b was supported for unhealthy lifestyle behaviors (bar and fast food restaurant both check-ins, at $p<0.001$).

In H2(a/b) we hypothesized that individuals' healthy/unhealthy lifestyle behaviors at time t+1, are positively influenced by congruent images posted along with friends' disclosed healthy/unhealthy lifestyle behaviors at time t. For the healthy lifestyle behaviors, we could only find a significant positive coefficient (at $p<.05$) in Model 3 for images containing Sport-related objects. This finding provides support for H2a as the images with a Sport label are congruent with the context of healthy lifestyle behaviors. For unhealthy lifestyle behaviors related to bars, we could not find any significant coefficient for different types of images. Thus, we could not find support for H2b in the context of unhealthy lifestyle behaviors associated with bars. However, as we will discuss in the next part (Group B/Zero inflated), higher numbers of images containing Beverage- or Text-related objects posted along with friends' disclosed unhealthy lifestyle behaviors (bar) at time *t*, significantly reduces individuals' inhibition to post about their check-ins in bar places at time *t+1*. For the unhealthy lifestyle behaviors related to fast food restaurants, we found positive significant coefficient (at $p<.001$) for images containing Food items. This result supports H2b, since Food is highly congruent with the context of unhealthy lifestyle behaviors related to fast food restaurants. As we discussed, our findings showed a mixed support for the influential power of congruent images in the context of unhealthy lifestyle behaviors (H2b). This result provides more evidence confirming our findings in Essay 1 showing that social influence is a context-dependent factor. Table 3.7 summarizes estimation results for each hypothesis.

Table 3.7. Supported Hypotheses

Healthy Lifestyle Behavior	**H1a**	**H2a**
Fitness Center & Gym	no	yes
Unhealthy Lifestyle Behavior	**H1b**	**H2b**
Bar	yes	no
Fast Food Restaurant	yes	Yes

Group B/Zero inflated. As discussed in Essay 1, the estimation coefficients in the Zero Inflated part of ZINB models provides additional insights regarding factors that inhibit individuals from disclosing their health-related lifestyle behaviors in online social networks. A significant negative coefficient for a factor in the Zero-Inflated part of tables indicates that the factor reduces individuals' inhibition to post about specific health-related lifestyle behaviors.

Per Part B/Zero Inflated of Table 3.4, presence of images in friends' check-ins at fitness places at time *t* significantly moderates (at $p<.01$) the social influence of friends on individuals in posting their healthy lifestyle behaviors at time *t+1* in online social networks. This is an interesting finding, showing that a higher number of images posted by friends from healthy places can increase individuals' incentive to share their healthy lifestyle behaviors in online social networks. We could not find significant impact in the Zero Inflated part of the table for types of images.

Per Part B/Zero Inflated of Table 3.5, images containing one of the labels of Beverage, Text, or Human can impact the likelihood of reporting unhealthy lifestyle behaviors related to bars in online social networks. The coefficients for Beverage and Text labels are both negative and significant at $p<.01$. This result indicates that the existence of Beverage and Text objects in posted images along with unhealthy lifestyle check-ins at bars decreases the likelihood of inhibition to disclose similar behaviors in online social networks. An interesting finding relates to images with Human labels. Our result reveals that a higher number of images containing Human objects posted by friends from bar places can increase the inhibition to share similar behaviors by individuals in online social networks. We could not find significant result for the moderation role of image presence on social influence for inhibition of behavior disclosing related to bar venues.

Per Part B/Zero Inflated of Table 3.6, friends' images containing Flora objects can decrease the chance of reporting unhealthy lifestyle behaviors related to fast food restaurants by individuals in online social networks. We could not find significant results for other types of images. We also did not find a significant coefficient for the moderation role of image presence in inhibition behavior of individuals to report their check-ins at fast food restaurants.

3.8. Discussions

These days, online social network sites increasingly play a prominent role in our lives and can influence both our online and offline behaviors. Indeed, prior studies observed the impact of online social networks on behaviors such as adoption of paid services (Bapna and Umyarov 2015), political mobilization (Bond et al. 2012), and physical activity (Althoff et al. 2016). In light of these findings, understanding the factors that can influence individuals' behaviors in online social networks becomes paramount.

In this study, we focused on individuals' health-related lifestyle behavior and built from the model in Essay 1 to investigate the influential power of images. In Essay 1, we have established that disclosed healthy and unhealthy lifestyle behaviors of individuals in online social networks influence others' health-related lifestyle behaviors. Our content investigation has shown that images are a prominent part of revealed lifestyle behaviors in online social networks. Images grab individuals' attention (Yantis 2005, Bucher and Schumacher 2006, Geise 2011) and play roles in the process of observational learning (Yussen 1974). Thus, this study's first research question: Does the presence of photos moderate the impact of friends' healthy and unhealthy lifestyle behaviors? We answered this question by collecting image content posted along with lifestyle check-ins and computed the number of observable images posted on egocentric network of individuals. Later, we estimated the interaction effect of average number of friends' posted lifestyle behaviors and the number of observable images in different health-related contexts.

The second research question in this study: How do the contents of posted photos contribute to friends' healthy and unhealthy lifestyle behaviors? We answered this question by capturing content of images. We used Rekognition API in our study and analyzed images in three different phases. Our approach helped us to find major type of objects inside posted images in different contexts and label each of them by one or more labels. Later, we aggregated the number of observable images of each type and captured the influence they had on individuals' healthy and unhealthy lifestyle behaviors over time. Our estimated models in this study revealed the potential power of shared photos in formation of individuals' healthy and unhealthy lifestyle behaviors.

First, our empirical results uncovered effect of images in grabbing individuals' attention and the moderation effect of that on friends' social influence over healthy and unhealthy lifestyle behaviors. We found that while the presence of images can increase the social influence of observed friends' unhealthy lifestyle behaviors, it cannot moderate the effect of observed friends' healthy lifestyle behaviors. In another insight gained from the Zero Inflated part of our analysis, we found that images posted from gym and fitness center places can moderate the effect of friends' posted check-ins on incentive of individuals to share their healthy lifestyle behaviors. These novel findings show that posted images moderate the social influence of friends over healthy and unhealthy lifestyle behaviors differently.

Second, our study revealed that congruent images posted along with health-related lifestyle behaviors in online social networks can be the source of social influence and impact individuals' healthy and unhealthy lifestyle behaviors. Particularly, we found that sport-related images posted along with friends' check-ins at gym and fitness centers can lead to higher numbers of disclosed healthy lifestyle behaviors for individuals over time. Similarly, we found that food-related images posted along with friends' check-ins at fast food restaurants can increase the number of observed individuals' unhealthy lifestyle behaviors at fast food restaurants. These are important findings showing that relevance of images and disclosed lifestyle behaviors can reinforce the social influence power of friends in online social networks.

3.9. Implications

Online photo-sharing has been considered an identity-construction tool (Marcus 2015, Eftekhar et al. 2014) that helps online users to form their implicit identities by showing rather than telling (Zhao et al. 2008). Research has also shown that shared photos in online platforms contain personal recommendations that can make them influential (Eftekhar et al. 2014). This study is the first to investigate about the social influence power of images on individuals' healthy and unhealthy lifestyle behaviors. Our findings showed that the influential power of images is extended to health-related lifestyle behaviors of individuals as observed in online social networks. This study also makes several contributions to theory and practice.

3.9.1. Theoretical Contributions

This study makes a number of novel contributions to theory and research. First, to our knowledge, this is the first study that investigates the role of visual contents in the mechanism of peers' influence in the context of health behaviors. This research can provide a solid foundation for future studies to build from this work and study the other roles that visual contents can play in formation of individuals' lifestyle behaviors.

Second, we contribute to observational learning theory by expanding its application to analysis of image content. Our study suggests that the effect of visual content in grabbing individuals' attention should be separated from their direct effect on individuals' behaviors. In some contexts, visual contents can only play the role of moderators for the effects of textual information and in some others, they can be the main source of influence on human behavior.

Third, this study provides a framework for analysis of images. The introduced image analysis procedure in this study can be widely used by other studies and can help them to identify the prominent types of objects and categorize images in different contexts.

3.9.2. Practical and Policy Implications

This study also offers important practical implications. In recent years, the concern over lack of physical activity and unhealthy lifestyle behavior of individuals has grown. We found that online social network platforms can play significant roles in formation of individuals healthy and unhealthy lifestyle behaviors. It is, therefore, important for us to make better use of these platforms and promote our healthy behaviors rather than unhealthy behaviors as they can have broad impact on the long-term wellbeing of individuals in societies.

Our work showed that visual content can moderate the imposed social influence of friends over unhealthy lifestyle behaviors and can also be the direct source of influence on individuals' behavior. Thus, it is important for health policy makers to increase individuals' awareness about the consequences of unhealthy lifestyle behaviors. Health practitioners can also use proper visual contents to promote healthy lifestyle behaviors and target those individuals who have suffered from unhealthy lifestyle behaviors the most.

Finally, the results of our study suggest that the health practitioners should pay additional attention to online photo-sharing platforms like Instagram, as these platforms have facilitated the process of sharing photos for adolescents and can be widely used for sharing health-related lifestyle activities.

3.10. Limitation and Future Studies

Since this study has built on the HLB model that was developed in the Essay 1, it suffers from similar limitations including lack of access to a full list of active users in Foursquare, the possibility of self-selection bias in report of healthy and unhealthy lifestyle behaviors by individuals, and the static nature of captured social networks in our study. Additionally, in this study we have only captured the impact of friends' posted images on individuals' lifestyle behaviors in a relatively long-term period. Therefore, our study might not be able to capture the full influential power of images, as images can also have short-term impact on humans' behavior. Future studies can fill this gap by capturing larger numbers of observation in a short-term period and provide additional insights about the short-term impact of images.

In this study, we have only considered the role of image presence and different types of content inside images as the sources of influence. Our work can be extended by considering other attributes of images such as their visual appeal rating and the structure of images. Furthermore, future studies might be able to use the content of images to estimate the food calories and more accurately investigate the social impact of unhealthy images.

CHAPTER 4

Essay 3: Communities of Interest in Online Social Networks: Detection Method and its Application in Explaining Self-Disclosed Lifestyle Behaviors

4.1. Introduction

Unhealthy lifestyle behavior is one of the major causes of disease and death in the U.S. (CDC 2015). Approximately one third of American adults suffer from cardiovascular diseases (Rosamond et al. 2008) and more than half of them struggle with one or more type of chronic diseases (CDC 2015). Poor dietary habits, obesity, physical inactivity, and excessive alcohol drinking are considered the main contributors (CDC 2015, Artinian 2010). In the U.S. these unhealthy lifestyle behaviors impose a huge cost burden on the healthcare system. The healthcare cost for inadequate levels of physical activity alone has been estimated to be $117 billion per year (Carlson et al. 2015). Despite all the evidences supporting the benefits of healthy lifestyle behaviors, American adults are increasingly burdened with the consequences of unhealthy lifestyle behaviors.

Research has demonstrated that online social network platforms can be used for health promotional purposes such as advocating physical activity (Valle et al. 2013), smoking cessation (Pechmann et al. 2015, Ramo et al. 2015), and weight loss (Waring et al. 2016, Napolitano et al. 2013). There are programs in online social networks to promote healthy lifestyle behaviors and prevent diseases, disability, and premature death. Examples are VERB and TRUTH—programs by non-profit organizations to increase physical activity and reduce smoking among adolescents (Huhman et al. 2004, Evans 2006). These programs offer a practical framework for understanding how individuals' lifestyle behaviors can lead to health problems and then suggests a set of motivational alternatives to improve individuals' and societies' health status (Aceves-Martins 2016, Hastings and Haywood 1991). For such programs to succeed, there is a need to identify vulnerable individuals in online platforms and recognize the motivational factors underpinning their healthy and unhealthy lifestyle behaviors. Hence, such programs need information about individuals' lifestyle behaviors. Self-disclosure in OSNs provides an opportunity to capture the needed information.

Self-disclosure is the "process of making the self known to others" (Jourard & Lasakow, 1958, p. 91) and can fulfill basic social needs of individuals for belonging and connectedness (Bazarova and Choi 2014). While self-disclosure can happen in both offline and online social environments, research shows online social network sites have unique features that facilitate the process of self-disclosure (Lee et al. 2013, Nguyen et al. 2012). There is a continuum of modes for self-disclosure within digital platforms. These modes range from explicit to implicit (Zhao et al. 2008). In explicit presentation, people have the opportunity to self-disclose themselves in a narrative format. However, in implicit presentation, people present themselves through shared activities, interests, and preferences without actually telling them. Research shows that most people prefer to present themselves online by disclosing implicit information (Zhao et al. 2008).

Study of self-disclosed information in online social networks for analysis of individuals' health-related behaviors is limited to only a few studies (Essay 1). Research has demonstrated that interests and preferences are sources of intrinsic motivations to perform activities (Deci 1992, Ryan and Deci 2000) and can play important roles in the formation of lifestyle behaviors (Deci 1992, Sagiv et al. 2011, Schwartz 2015). Thus, having access to individuals' interests helps health practitioners to not only understand reasons for the behaviors but also provide opportunities for them to develop and suggest alternatives that are in line with the interests and preferences of targeted individuals. However, to our knowledge, no prior study has investigated the relationship of individuals' observed preferences and interest in online social networks with their health-related lifestyle behavior.

To fill this gap, in this study we plan to investigate (i) how individuals' interests and preferences can be detected within online social networks and, (ii) how self-disclosed health-related lifestyle behaviors of individuals in online social networks are associated with their observed interests and preferences.

To answer these questions, we rely on the structure of online social networks and develop a theory-based community detection algorithm that can detect various communities of interest. Later, we use a dataset of more than 32,000 active users in Twitter and Foursquare to conduct our study. Our algorithm has successfully identified 43 different communities of interest representing individuals' interest and preferences. Our statistical models also show that such interests have direct relationships with observed healthy and unhealthy lifestyle behaviors of individuals in online social networks.

Our study makes several contributions. First, we developed a theory-based community detection algorithm that can capture a wide variety of individuals' interests and preferences. Second, we found that the interests that individuals reveal in online social networks are associated with individuals' healthy and unhealthy lifestyle behaviors. Third, our results uncover and distinguish the online social communities associated with healthy behaviors as opposed to unhealthy behaviors. Fourth, we add to the literature of disclosure in online communities by showing that individuals' self-disclosure follows the norm of disclosure on those communities. Finally, the results of our work could be used in targeting individuals for health promotion programs, for understanding how unhealthy programs get propagated, for projects and programs to promote healthy lifestyle, and for comparative analysis of how social networks in various countries create communities of interest and how such communities promote or inhibit healthy and unhealthy behaviors.

4.2. Literature Review

4.2.1. Literature on Self-Disclosure Behavior

Self-disclosure is an important part of our lives. It refers to "the act of revealing personal information to others" (Jourard, 1971, p. 2). The act of self-disclosure is intrinsically rewarding (Tamir and Mitchell 2012) and can fulfill human's social needs such as sense of belonging (Bazarova and Choi 2014). Research demonstrates that individuals tend to reveal their offline attributes and lifestyle activities in online environments through posted contents (Rahman 2016, Moore and McElroy 2012, Amichai-Hamburger and Vinitzky 2010, Barkhuus et al. 2008). This tendency makes individuals self-reporting objects (Mitrou et al. 2014) who reveal information about themselves over time.

There are five goals involved in self-disclosure: social validation, self-expression, relational development, identity clarification, and social control (Bazarova and Choi 2014). Self-disclosure is considered one of the key elements of online social networks (Kaplan and Haenlein 2010) that has a direct impact on the success of these platforms (Wang et al 2016). Research shows that individuals are more satisfied with online platforms that promote self-disclosure (Special and Li-Barber 2012) and tend to use those platforms more often (Trepte and Reinecke 2013). Prior studies also compared the interactions of individuals in online platforms with face-to-face relationships and found that the unique features of online platforms promote a higher level of self-disclosure (Lee et al. 2013, Nguyen et al. 2012). Accordingly, we have witnessed that millions of users share their thoughts, beliefs, experiences and lifestyle behaviors on a daily basis with their friends and followers in online social networks. It was argued that self-disclosure in online social networks fosters feelings of connectedness (Utz 2015) and increases the level of intimacy among users (Park et al. 2011). In fact, revealing personal information helps individuals to draw attention and be liked by others, which facilitates the process of social interactions (Sheldon 2009, Posey et al. 2010)

However, people generally tend to present an idealized version of themselves in their disclosures (Goffman 1959). This practice allows them to manage the impression they make (Wang et al 2016). Hogan (2010) discussed that, similar to real-life situations, online social networks consist of "front stage" and "back stage" settings. In the front stage, people are trying to present the idealized version of themselves according to their social role in society. But back stage is the place where people do the real work to keep up the appearances (Hogan 2010). This difference causes doubts about the representativeness of self-disclosed behaviors. Despite the importance of this issue in understanding self-disclosed behaviors, insights into the nature of self-disclosure in online platforms remain scarce. In this work, we study the relationship between individuals' self-disclosed lifestyle behaviors and their observed interests to investigate how disclosed behaviors can represent the lifestyle behaviors of individuals.

4.2.2. Literature on Location-Sharing Behavior

With ubiquitous access of individuals to internet and mobile services and the increasing growth of location-based social networks, people have started to share their location with their friends and followers in online platforms. Research shows that while privacy concerns negatively impact intention to share location-related information, perceived benefits have a stronger positive influence (Zhao et al. 2012). A recent study suggests that in the context of online location-sharing, the concept of "location has changed from being something you *have* (a property or state) to something you *do* (an action)" (Cramer et al. 2011, p. 65). This indicates that individuals' shared location in online social networks is not about expressing the current geographical coordinates but the type of their current activities.

Prior studies find several motivations behind location sharing include gaming, sending signals to friends, and self-impression (Patil et al. 2012, Lindqvist et al. 2011). Location sharing is less about showing physical presence and more about to achieve socially oriented goals (Rahman 2016). Indeed, there is a distinction between location as a "space" and location as a "place" (Dourish 2006). Location as a space describes a geometrical arrangement that can be helpful for activities such as movement, and location as a place refers to recognizable and persistent social meaning (Dourish 2006). It is important to understand this distinction and distinguish between purpose-driven and social-driven location sharing. In a purpose-driven location-sharing, the main goal is to share "space" and perform activities like coordination or planning. However, in a social-driven location-sharing people share their "place" through online social networks to attract attention and do self-presentation (Tang et al. 2010). In social-driven location-sharing the main focus in on the semantic aspects of the location (Tang et al. 2010).

4.3. Communities of Interest

Online social networks create platforms that connect people with similar interests and values (Boyd and Ellison 2008). Research shows that individuals tend to reveal their attributes in online environments through posted contents and their pattern of relationships (Moore and McElroy 2012, Amichai-Hamburger and Vinitzky 2010). Accordingly, prior studies used text-based and network analysis approaches to detect individuals' common interests. Table 4.1 lists a selected number of recent studies that offered relevant techniques for capturing individuals' attributes in online platforms.

Table 4.1. Review of Attribute Detection Methods

Study	Method	Summary
Trusov et al. (2016)	Text Mining	This study used a rich dataset from a leading global information company to develop an extension of the Correlated Topic Model in Natural Language Processing. Their model captures individual's roles (i.e. Information Seeker, Online Shopper) during their web surfing time periods.
Argamon et al. (2009)	Text Mining	This study used full sets of blog posts for more than 19,000 users to train a supervised learning model that could be used for prediction of age, gender, language and personality of individuals.
Li et al. (2014)	Network Analysis	This study considered networks around individuals as the source for attribute detection. It argued that each ego individual has several social circles in her/his ego network and she/he only takes the attributes that are common in each social circle. The main goal of the study was to find the optimal number of social circles in ego-networks that can represents individual attributes.
Palsetia et al. (2012)	Network Analysis	This study used Jaccard index to compute similarity among set of preselected accounts in Facebook and Twitter. The similarity index was computed based on users' pattern of interaction with the accounts. Later, a hierarchical approach was used to partition the accounts into different communities of interest.
Mislove et al. (2010)	Network Analysis	This study considered social networks of users on Facebook. It captured individuals' attributes and formed communities of individuals based on commonality of their attributes. This study found that these communities generate significant values for the community index (modularity). Accordingly, they conclude that community detection can be used for attribute detection.
Ikeda et al. (2013)	Hybrid Model	This study offered a three-layer framework (i) Extracting demographic features from users' tweets (ii) Applying community detection techniques to put users in different communities (iii) Estimating demographic information for community members using the extracted features
Pennacchiotti et al. (2011)	Hybrid Model	This study used LDA model to detect linguistic features for different classes and then applied sentiment analysis to capture individuals' sentiment toward the features in each class. These features along with some captured features in online social network have been used to train a supervised machine learning algorithm.

Proposed methods generally suffer from one or more of the following limitations: (1) they cannot represent actual interests of individuals and suffer from the social desirability element of self-disclosure (2) they can only capture a few attributes (3) while the silent users (lurkers) make up the majority of online communities (Gong et al. 2015, Nielsen 2006), the proposed algorithms only capture the attributes of active users (4) capturing an individual's attributes requires advanced knowledge about the personal attributes of the individual's friends (5) they limit the relationships of users to inside community relationships and ignore interaction of them with outside users. Our study addresses these limitations by introducing a novel theory-based clustering model for capturing individuals' community of interest in online social networks.

We develop our algorithm based on the concept of homophily (McPherson et al. 2001, Lazarsfeld and Merton 1954) in social science. Homophily refers to the strong tendency of individuals with similar attributes to interact with each other rather than with people with dissimilar attributes (McPherson et al. 2001, Lazarsfeld and Merton 1954). Studies have shown that homophilous relationships can promote the spread of similar behavior among individuals (McPherson et al. 2001, Rogers 1995). Homophily has also been considered as one of the main dimensions of social structure in healthy lifestyle theory, where it plays a role in the formation of health-related lifestyle behaviors (Cockerham 2005).

Homophily has roots in demographic or psychographic factors (Gu et al. 2014). These factors can change over time (i.e. marital status) or can be constant attributes (i.e. race) (Li et al. 2013). Researchers articulated several explanations for observed homophilous relationships among individuals. Gu et al. (2014) argued that individuals develop relationships with similar others because: (1) it increases the chance of being liked by others and (2) it is easier to get confirmation from similar others. Kossinets and Watts (2009) considered the role of trust and solidarity in creation of homophilous ties and noted that the ongoing cost of maintaining relationships with similar others is lower than with dissimilar ones. They also emphasized a prominent fact of social life: that individuals' choices of relationship are primarily constrained by other factors such as geographical locations, neighborhoods, working places, and schools. These constraints in nature provide homogenous choices of relationships for individuals.

4.3.1. Homophily-based Interest Detection (HID)

Users of online social networks come from different social and cultural backgrounds and possess different interests and preferences. According to selective exposure theory, people have a tendency to expose themselves to those mass communication channels which reinforce their own views and are in agreement with their own interests and type of thinking (Sears and Freedman 1967, Zillmann and Bryant 1985, Zillmann 1988). Therefore, it is expected that in interaction with online social networks, individuals follow the social pages[20] that promote their own views and are in line with their interests and preferences. Relying on this assumption, and by applying clustering techniques in network studies, we offer a Homophily-based Interest Detection (HID) method for capturing communities of interest in online social networks.

HID is an algorithm that can be applied to extended bipartite graphs within online social networks and is composed of four sequential steps: (1) Network Simplification, (2) Network Clustering, (3) Cluster Labeling, and (4) Measurement of Interest. An extended bipartite graph in online social networks consists of two separate networks: the social network of individuals and the bipartite network of individuals and social pages. A social network of individuals refers to a graph in which a node represents an individual and an edge indicates the existence of reciprocated relationship between two individuals. A bipartite network of individuals and social pages is a graph that has two types of node (individuals and social pages), and edges represent

[20] A social page refers to an account in online social networks related to an organization, a brand, a celebrity, a program, a news agency or any other popular entity that can attract individuals' interests.

the pattern of following social pages by individuals. Figure 4.1 shows a sample of an extended bipartite graph in online social networks.

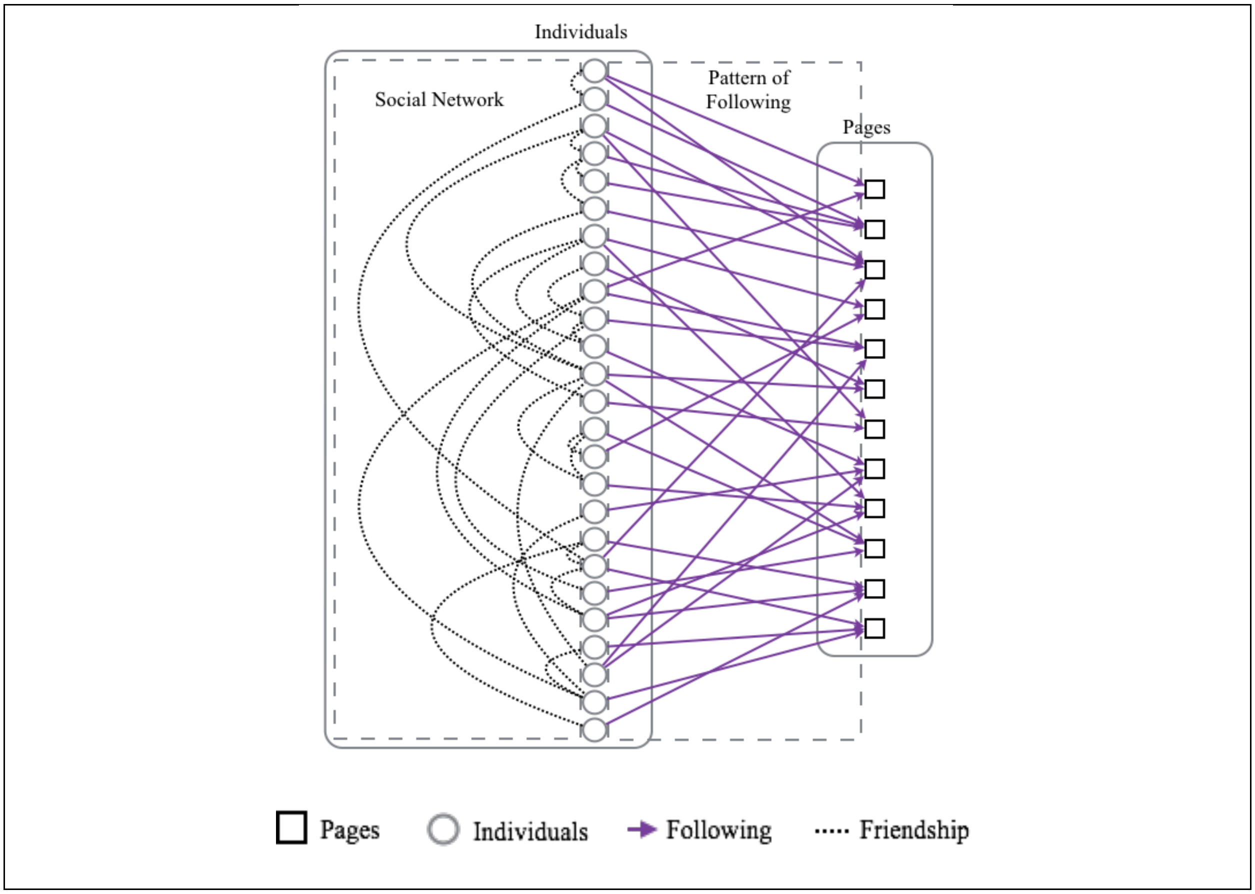


Figure 4.1. Extended Bipartite Graph

Step 1: **Network Simplification.** The main purpose of this step is to convert an extended bipartite graph into a weighted graph of social pages, where weights represent the similarity of the social pages. To achieve this goal, in the first stage, each social page is mapped to the graph of its followers and their relationships. In the second stage we compute the similarity of the social pages using the mapped graphs. Similarity is computed based on two main assumptions: (1) individuals who are friends with each other are more likely to have similar interests and follow similar social pages in online social networks. This criterion is in line with the concept of homophily (McPherson et al. 2001, Lazarsfeld and Merton 1954) (2) social pages with higher numbers of common followers are more likely to be similar to each other. This assumption is in line with the concept of embeddedness (Aral and Walker 2014, Easley and Kleinberg 2010) in social network studies. According to this concept a higher number of common friends is an indicator of a strong relationship between two individuals. Accordingly, two mapped graphs have a higher level of similarity when (1) the number of edges among common nodes in two mapped graphs be higher than the number of edges among the rest of the nodes in the graph (2) the number of common nodes in the two mapped graphs is higher than the number of uncommon nodes in the graphs. To satisfy those needs, we use the following similarity function in graph theories (Johnson 1985):

$$Sim(G_A, G_B) = \frac{(|V(G_A, G_B)| + |E(G_A, G_B)|)^2}{(|V(G_A)| + |E(G_A)|) \,.\, (|V(G_B)| + |E(G_B)|)}$$

Where G_A, G_B represent the mapped graphs of social page A and social page B. $|V(G)|$ returns the number of nodes (follower) and $|E(G)|$ returns the number of links (relationships) in graph G . $|V(G_A, G_B)|$ and $|E(G_A, G_B)|$ are the number of common nodes and links between G_A and G_B respectively. This formula considers both the numbers of common followers and their relationship in computation of similarities and return a value between 0 and 1. The output of this similarity function can be used to weight the simplified network of social pages in online social networks.[21] Figure 4.2 depicts the process of converting an extended bipartite graph to a simplified weighted graph.

---

[21] We ignore the similarity of each social page with itself as it produces self-loop connections for all the nodes in the graph which cause bias in the graph clustering process.

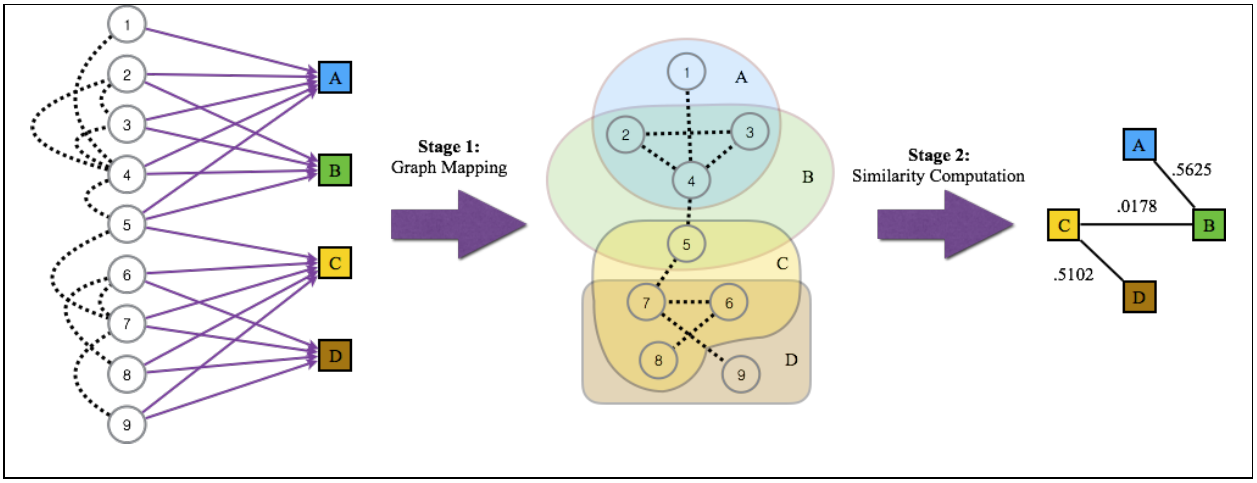


Figure 4.2. Mapped Network of Social Pages

Step 2: **Network Clustering.** In this step we use optimization techniques for graph clustering. The method can be used to find the best value of modularity in weighted graphs (Newman 2004). Modularity (Newman and Girvan 2004) is a quantity that measures how the structure of communities is different from a random graph. The following formula can be used to compute the modularity in weighted graphs (Newman 2004):

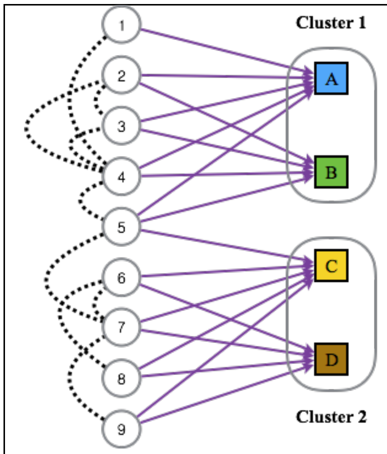
$$Q = \frac{1}{2m}\sum_{ij}[A_{ij} - \frac{k_i k_j}{2m}]\delta(c_i, c_j)$$

Where A_{ij} is the weighted adjacency matrix of a graph, m is the total sum of the weights in the graph and computed as $m = \frac{1}{2}\sum_{ij} A_{ij}$, k_i is the sum of the edges' weight that have one end in node i and computed as $k_i = \sum_j A_{ij}$, c_i represents the cluster that node i is assigned to it. Finally, the function $\delta(c_i, c_j)$ returns 1 if node i and j are assigned to the same cluster and 0 otherwise. The main purpose of modularity-based graph clustering algorithms is to assign nodes to various clusters in order to maximize the value of modularity. This process leads to detection of graph clusters that has meaningful difference with random graphs. In this study we use the modularity optimization method that was proposed by Blondel et al. (2008). This method is one of the best modularity optimization methods in terms of performance and speed (Fortunato 2010). This method uses a greedy approach that iteratively optimizes the value of modularity by assigning the nodes to different clusters. Applying the algorithm to the graph in Figure 4.2 leads to a modularity of .48 and two clusters (C1 = {A, B}, C2 = {C, D}). By definition a nonzero value for the modularity indicates the existence of graph clusters that deviate from random graphs. It was argued that real world communities have a modularity value between 0.3 and 0.7 (Clauset et al. 2004).

However, Modularity-based algorithms suffer from the resolution problem in large graphs (Fortunato and Barthelemy 2007, Kumpula et al. 2007). This means that modularity-based algorithms cannot capture small clusters within the large set of clusters. In order to solve this problem, we use the hierarchical clustering approach in HID. In the hierarchical clustering approach, we continuously apply our clustering algorithm to newly detected clusters in order to divide them to the smaller cluster. In this process, we check the value of modularity in each level to make sure that it satisfies the minimum threshold of real world communities.

Step 3: Cluster Labeling. After detection of page clusters in online social networks, we have to consider the functional property of the clusters. The functional property can be captured by considering common attributes of social pages in each cluster. Detection of functional property for clusters confirms the validity of the clustering approach (Yang and Leskovec 2015). We later assign the functional property of each cluster as its label.

Step 4: Measurement of Interest. Each cluster of social pages represents a community of interest with a set of homogenous social pages. Accordingly, the final step in the HID model is to measure the connectedness of users to their communities of interest. In order to measure this quantity, the normalized number of social pages that each user has followed in each community should be computed. The captured values demonstrate the strength of interest of users to different communities of interest in online social networks. Figure 4.3 demonstrates this process.



Stage 1: Number of paged followed in each cluster

	Cluster1	Cluster2
User 1	1	0
User 2	2	0
User 3	2	0
User 4	2	0
User 5	2	1
User 6	0	2
User 7	0	2
User 8	0	2
User 9	0	2

Stage 2: Normalized the values

	Cluster1	Cluster2
User 1	0.5	0
User 2	1	0
User 3	1	0
User 4	1	0
User 5	1	0.5
User 6	0	1
User 7	0	1
User 8	0	1
User 9	0	1

Figure 4.3. Measurement of Users' Interests

4.4. Theoretical Foundation and Hypothesis Development

This section discusses the theoretical foundation and hypothesis development related to communities of interest.

4.4.1. Communities of Interest and Motivational Factors

According to the functional theory of disclosure, understanding individuals' self-disclosed behaviors requires identification and measurement of major sources of interests and values for individuals (Derlega and Grzelak 1979). Research demonstrates that personal attributes such as interests, values, and preferences are not just the reasons behind self-disclosure of behavior, but also form motivational factors behind performing different activities and lifestyle behaviors (Deci 1992, Sagiv et al. 2011, Schwartz 2015).

Self-determination theory (SDT) is a motivational theory in psychology that captures the role of personal attributes in formation of self-regulated behaviors (Ryan and Deci 2000, Deci and Ryan 2011). It distinguishes between the motivational basis of self-determined behaviors and those activities that are instrumental for some forms of reward (Ryan and Deci 2000). SDT argues that developing a sense of autonomy (being the origin of our own behavior), competence (control over the desired outcome) and relatedness (having interaction with others and being understood and cared for by them) are critical in the formation of self-motivated behavior (Ryan and Deci 2000). Prior studies in the field found that the internalized values, interests and preferences of individuals play important roles in the process of developing senses of autonomy, competence, and relatedness in individuals (Deci 1992, Ryan et al. 2008, Waterman et al. 2003). In fact, interest stimulates effort (Dewey 1913) and is associated with engagement in activities (Ainley et al. 2002).

Accordingly, we expect that interests and preferences of individuals play a role in the formation of their health-related lifestyle behaviors. Thus, we argue that health-related interests of individuals are associated with their health-related lifestyle behaviors. Hence,

> **Hypothesis 1**: *Disclosed healthy lifestyle behaviors are (a) positively associated with observed individuals' healthy interests. (b) negatively associated with observed individuals' unhealthy interests.*

> **Hypothesis 2**: *Disclosed unhealthy lifestyle behaviors are (a) positively associated with observed individuals' unhealthy interests. (b) negatively associated with observed individuals' healthy interests.*

4.4.2. Inhibition Role of Communities of Interest

The self is constructed through "a process of social interactions with various communities, physical structures, environments, as well as with other humans and objects" (Morie et al, 2008, p. 367). According to the social identity model of deindividuation (SIDE), the notion of self consists of two identities: (1) personal identity (2) social identity (Reicher et al. 1995). Personal identity refers to unique personal attributes of individuals and social identity refers to different groups that the individuals belong to. Research indicates that social identities play important roles in demarcating the accepted behaviors and norms within social groups (Pegg et al. 2018, Pugh 1997, Erikson 1994). Social identity is also at work in online social networks.

Communities of interest not only represent the strength of individuals' interest in different topics but also reveal the social group that individuals interact with, in which they have their own defined norms. Accordingly, self-disclosure in online social networks follow the perceived norms of self-disclosure within these groups (Nguyen et al. 2012, Cramer et al. 2011).

However, not all individuals' social identities are compatible (Farnham and Churchill 2011). This means that while performing one form of lifestyle behavior could match with existing norms in one social group, it may not be appropriate in the other group. We argue that in these situations, people who participate in conflicting lifestyle behaviors are less likely to disclose their behavior in online social networks. Accordingly, we propose our last set of hypotheses as:

> **Hypothesis 3:** *Individuals who belong to unhealthy communities of interest are less likely to disclose their healthy lifestyle behaviors in online social networks.*
>
> **Hypothesis 4:** *Individuals who belong to healthy communities of interest are less likely to disclose their unhealthy lifestyle behaviors in online social networks.*

4.4.3. Control Variables

In our study, we control for three factor groups that can impact individuals' self-disclosure in online social network. First, we control for other non-health-related interests of individuals. This helps to separate the effects of health-related interests from non-health-related interests. Moreover, capturing the effect of non-health-related interests provides additional insights for interpretation of relationships between communities of interest and individuals' disclosed health-related lifestyle behaviors. Second, we control for the individuals' social network size. Wang et al. (2016) showed that a higher number of friends in Facebook is negatively associated with the individual's level of self-disclosure in that platform. Finally, we control for the effects of demographic factors. A recent study argued that gender can play an important role in making decisions for self-disclosure (Wang et al. 2016). Socio-economic status (SES) is another important demographic factor. In Essay 1, we have demonstrated that SES is positively associated with observed individuals' healthy lifestyle behaviors and negatively associated with their unhealthy lifestyle behaviors.

4.5. Data Collection and Measurement

Data for this study were collected from two popular online social network sites, Twitter and Foursquare. Twitter is a platform that allows users to share their opinions and activities in real time with others. Foursquare is a location-based social network that gives individuals the opportunity to share their lifestyle activities and can be integrated with other social networks such as Twitter. In this study, we took advantage of this integration and followed the proposed data collection approach in Essay 1. We captured tweets of users inside the U.S. who have connected their Foursquare account to their Twitter account and shared their lifestyle activities from Foursquare in Twitter. The data collection period was from January 28 to June 17, 2014. Data collection was conducted in three stages: user identification, health-related lifestyle behavior observation, and complementary data collection. At the first stage, we captured check-ins of users in Twitter for a twelve-week period (January 28 – April 22, 2014). In this period, we selected users who post at-least one check-in every two weeks after their initial captured check-in. Of our collected data, 32,700 unique individuals met this requirement with average posted check-ins of 3.8 in each week. At the second stage, we captured health-related lifestyle behaviors of users for an eight-week time period (April 22 – June 17, 2014) from their check-ins in different places. We followed the Essay 1's ideology and considered the types of location from which people check-in as the proxy for inferring their health-related lifestyle behaviors. At the third stage, we captured the social network of individuals using Twitter API. In the data collection period, we captured more than 5 million check-in tweets, and 1,127,420 distinct venues in the U.S.

4.5.1. Individuals' Health-related Lifestyle Behaviors

As it was discussed, we captured individuals' health-related lifestyle behavior based on the shared locations of individuals. We combined the salient categories in Foursquare that represent health-related lifestyle behaviors: *fitness center & gym*, *bar*, and *fast food restaurant*. Table 4.2 lists the Foursquare categories and number of venues in each type.

Table 4.2. List of Categories

Venue Type	Foursquare Categories	# of Venues
Fitness Center & Gym	Badminton Court, Baseball Field, Basketball Court, Boxing Gym, Climbing Gym, College Basketball Court, College Cricket Pitch, College Football Field, College Gym, College Hockey Rink, College Soccer Field, College Tennis Court, Cricket Ground, Gym, Gym / Fitness Center, Gym Pool, Gymnastics Gym, Hockey Field, Paintball Field, Rock Climbing Spot, Roller Rink, Rugby Pitch, Skate Park, Skating Rink, Soccer Field, Sports Club, Squash Court, Swim School, Tennis Court, Volleyball Court, Yoga Studio	36,047
Bar	Apres Ski Bar, Bar, Beach Bar, Beer Garden, Beer Store, Champagne Bar, Cocktail Bar, Dive Bar, Gastropub, Gay Bar, Hookah Bar, Hotel Bar, Irish Pub, Karaoke Bar, Piano Bar, Pub, Sake Bar, Sports Bar, Whisky Bar, Wine Bar	66,687
Fast Food Restaurant	BBQ Joint, Fast Food Restaurant, Food Court, Fried Chicken Joint, Hot Dog Joint, Mac & Cheese Joint, Pizza Place, Wings Joint	109,575

Later, we measured individual's health-related lifestyle behaviors as the number of days that each individual posted check-ins from venues within each venue type at the second stage of data collection for an eight-week time period. Therefore, for each individual, we computed one healthy lifestyle score (Fitness Center & Gym) and two unhealthy lifestyle scores (Bar and Fast Food Restaurant).

4.5.2. Individuals' Interest Metrics

To capture individuals' interests, we used the proposed HID model in this study. We used the captured social network data in the third stage of data collection to form both the social network of individuals and the bipartite network between individuals and social pages. The social network comprises 32,700 nodes (individuals) and 100,898 non-directed edges (reciprocated relationships) among them. The bipartite graph includes 32,700 individual nodes, 4,893 social page nodes[22], and 3,978,613 directed edges that show the pattern of following social pages by individuals. Applying HID in a hierarchical structure with a minimum modularity threshold of 0.3 puts the social pages into 43 distinct communities of interest.[23]

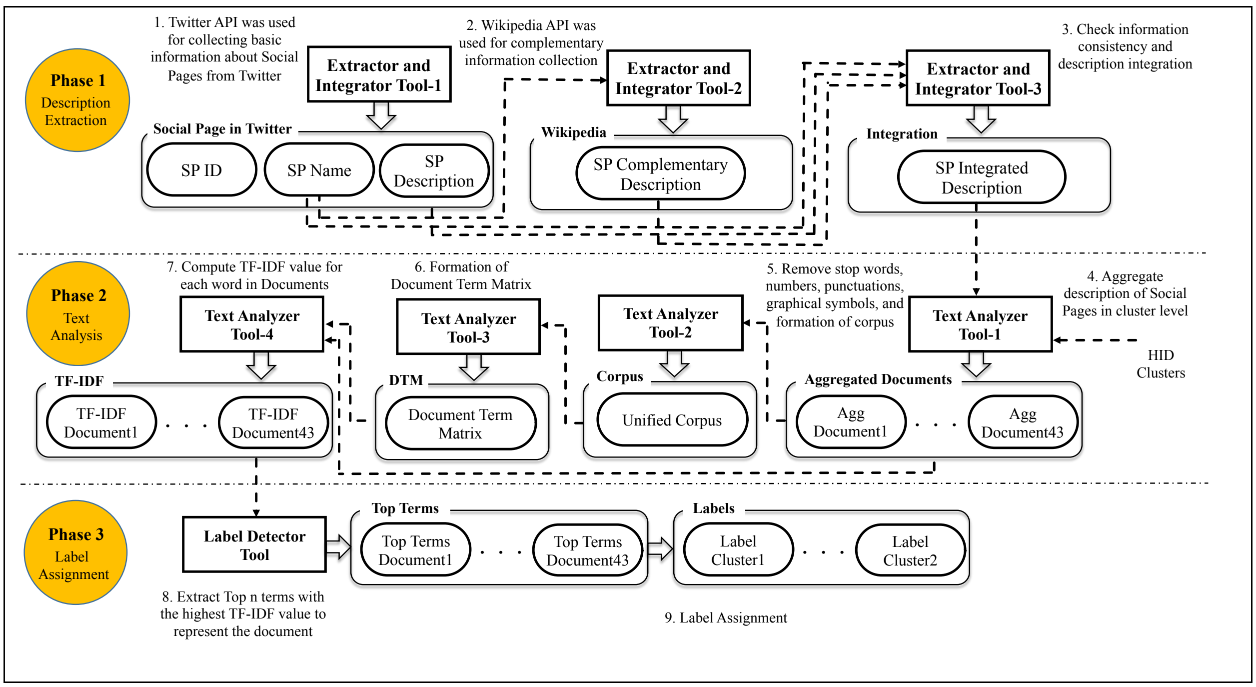


Figure 4.4. Cluster Labeling Process

Next, in the cluster labeling step, we developed several information extractor and analyzer tools to assign labels to clusters based on extracted descriptions for social pages within each cluster. Figure 4.4 illustrates the cluster labeling process in detail.

This process was conducted in three stages: description extraction, text analysis, and label assignment. In the description extraction stage, Twitter and Wikipedia APIs were used to extract

[22] In selection of social pages, we considered the top most popular social pages that have been followed by at least 1% of individuals in the social network side of the network.
[23] It was argued that real world communities have a modularity value between 0.3 and 0.7 (Clauset et al. 2004).

summary descriptions of the social pages. In the second stage, we aggregated the descriptions of social pages at the community level and formed an aggregated document for each community. Later, we applied the TF-IDF approach to find representative terms for each document (cluster). TF-IDF is a standard tool in information retrieval that represents each document by a weighted vector in the size of the overall vocabulary $(w_1, w_2, \ldots, w_n)$. Where w_i is calculated as:

$$w_i = TF_i \times log\ (IDF_i)$$
$$TF_i = The\ term\ frequency\ of\ term\ i\ in\ document\ D$$
$$IDF_i = \frac{Total\ number\ of\ documents}{Number\ of\ documents\ contain\ term\ i}$$

In the final stage of cluster labeling, we considered the terms with the highest TF-IDF weight in each document as the representatives of the associated community and assigned a label to each community accordingly. Figure 4.5 shows an example of captured terms for a Brewery community of interest.

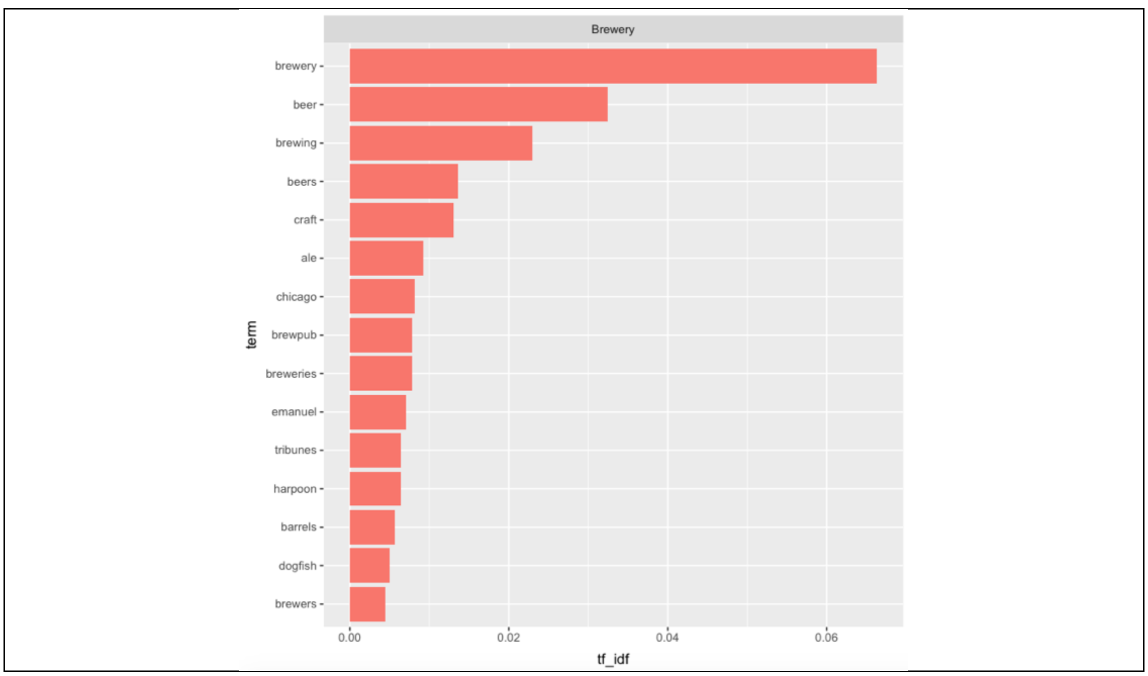


Figure 4.5. Terms with High TF-IDF Weight in Brewery Community of Interest

Figure 4.6 depicts the hierarchical structure of communities of interest as they were captured by the HID method.

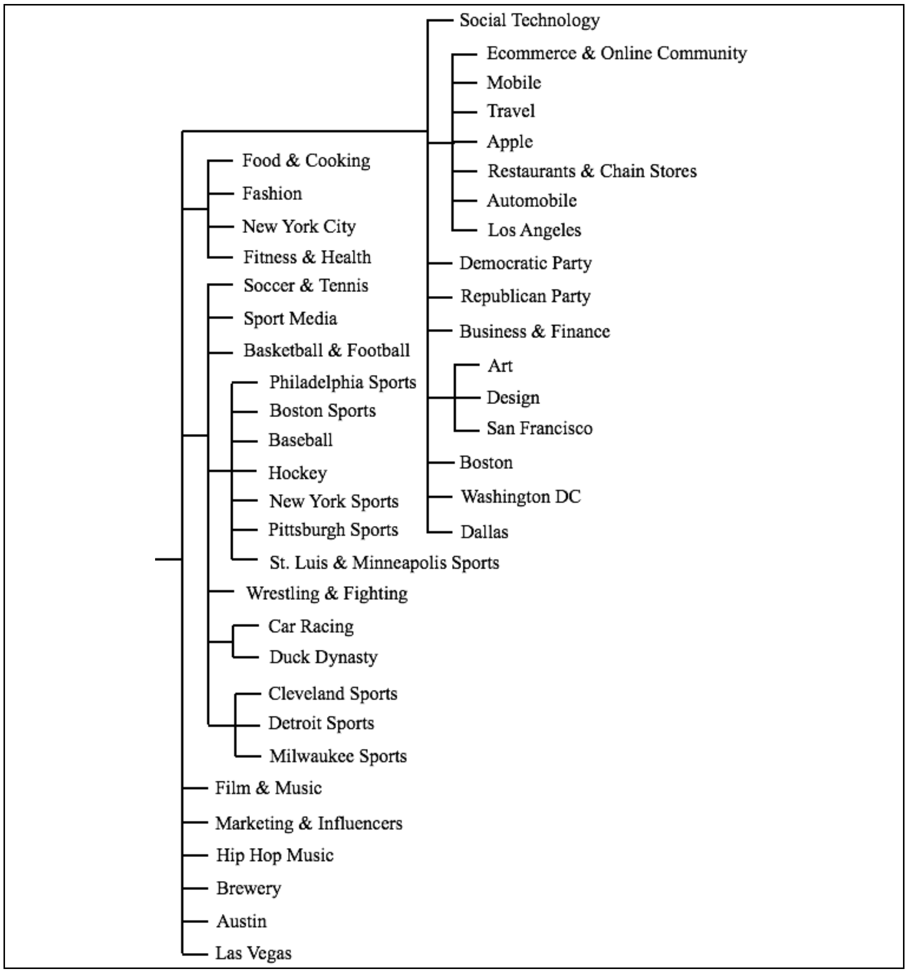


Figure 4.6. Hierarchical Structure of Communities of Interest

Figure 4.7 illustrates an example of a cluster that has been divided into four separated sub-clusters. We used the ForceAtlas 2 layout (Jacomy et al. 2014) in Gephi (Bastian et al. 2009) for visualization of clusters. ForceAtlas 2 is a force-directed layout where nodes repulse each other, and edges attract the nodes they are connected to toward each other (Jacomy et al. 2014).

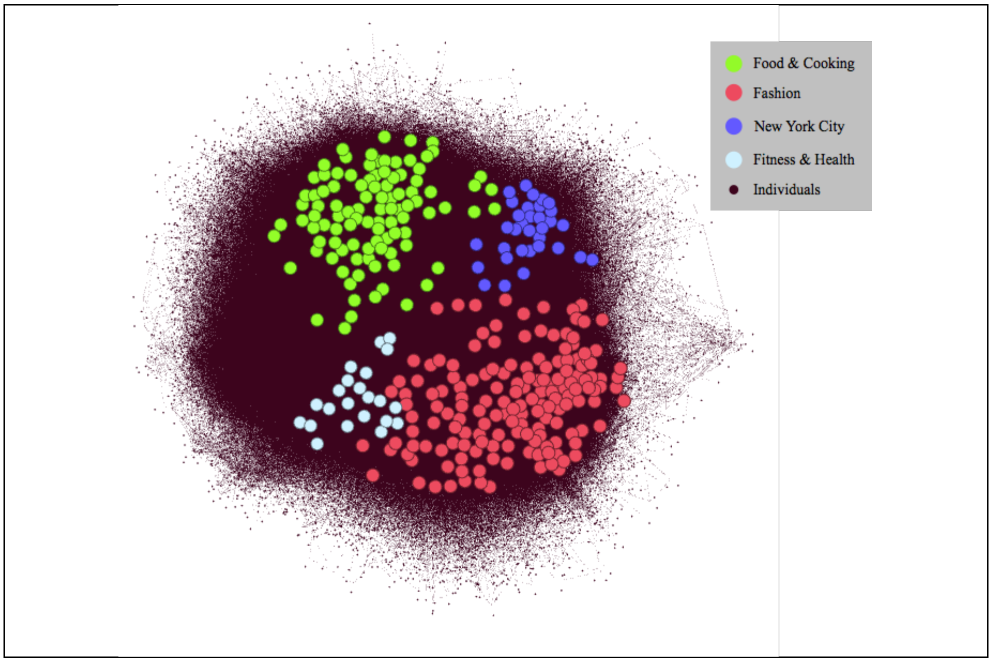


Figure 4.7. Cluster Visualization

We classified communities of interest into nine different categories (two of them are health-related). Table 4.3 shows these categories along with the communities of interest associated with them.

Table 4.3. List of Detected Communities of Interest

Food Communities of Interest	Travel & Entertainment Communities of Interest	Technology Communities of Interest
Brewery (89) Food & Cooking (101) Restaurants & Chain Stores (81)	Duck Dynasty (10) Film & Music (1538) Hip Hop Music (342) Sport Media (189) Travel (55)	Apple (25) Automobile (9) Social Technology (352) Mobile Technology (58)
Business Communities of Interest	**Fashion & Art Communities of Interest**	**Politics Communities of Interest**
Business & Finance (33) Ecommerce & Online Com. (125) Influencers (596)	Art (11) Design (12) Fashion (159)	Democratic Party (402) Republican Party (79)
Location Communities of Interest	**Sports Team Communities of Interest**	**Sport Communities of Interest**
Austin (12) Boston (10) Dallas (5) Las Vegas (87) Los Angles (8) New York City (33) San Francisco (9) Washington DC (10)	Boston Sports (16) Cleveland Sports (3) Detroit Sports (11) New York Sports (7) Milwaukee Sports (8) Philadelphia Sports (13) Pittsburgh Sports (4) St. Luis & Minneapolis Sports (6)	Baseball (42) Basketball & Football (179) Car Racing (14) Fitness & Health (22) Hockey (33) Soccer & Tennis (31) Wrestling & Fighting (54)

* Number of social pages is in parenthesis

Finally, in the last step of HID, we measured interests of individuals and assigned a normalized value based on the level of connectedness to each of the captured communities of interest.[24]

4.5.3. Control Variables

Per our discussion in the previous section, in addition to individuals' interests, we control for individuals' social network size and demographic factors. In order to control the effect of social network size, we consider individuals' number of followers in Twitter. In Twitter, followers are those users who can observe individuals' disclosed behaviors. In order to be consistent with other factors in our study, we normalized this value by dividing the number of followers by the maximum number of followers that users in our study had. We also controlled for two demographic attributes of individuals: gender, and socioeconomic status (SES). Gender has been captured through the Foursquare profile of individuals using Foursquare API. We use a dummy variable to show individuals' gender (0 = female, 1=male) in our study. For measuring SES, we captured the residence of individuals from their Twitter profile and used Census Bureau API to extract associated individuals' income, education and poverty at the city/town level from American Community Survey 5-year data (2013). While the extracted values together represent SES, they are highly correlated. We used the explanatory factor analysis (EFA) to combine these factors and use one single representative factor for SES. The load factors for income, education and poverty were 0.86, 0.62, and -0.94, respectively.

---

[24] We test for possible correlation between the size of communities of interest (number of social pages in each community) and the average interest score of individuals on those communities. Our results show no significant correlation between these two factors.

4.6. Data Analysis and Model Estimation

4.6.1. Checking for Collinearity

To check for the possibility of a collinearity problem among independent variables, we first examined the conditioning index and variance proportion associated with independent variables. Per Belsley et al. (2005), a conditioning index greater than 30 could be an indicator of acollinearity problem. In our case the highest value for the conditioning index was 9.12. As the second test, we used the Variance Inflation Factor (VIF). Existence of any variable with a VIF score larger than 10 in a model can significantly influence the stability of the estimated parameters (Dielman, 2001). The highest value of VIF in our models was 3.16. Therefore, we conclude that our estimated models do not suffer from the collinearity problem.

4.6.2. Model Estimation

The distributions of individuals' check-ins in the second stage of data collection (**Appendix G**) were over-dispersed[25]. The high number of zeroes in the data (our dependent variable) makes the zero-inflated negative binomial (ZINB) a suitable method for estimation. Zeroes can have two sources: (1) individuals did not go to the captured type of venue and so the number of check-ins for them are naturally equal to zero (2) individuals went to the captured type of venue but they did not report it online.

The ZINB offers a probability model that distinguishes between these two sources of zeros by considering two latent groups. Group A represents individuals who reported their captured health-related lifestyle behavior as it happened (will be used for testing hypotheses 1 &

[25] We also test the presence of over-dispersion by the recommended alpha test method (Cameron and Trivedi 1990)

2), and Group B consists of people who did not report that behavior (will be used for testing hypotheses 3 & 4). The following equation represent this concept:

$$y_i \sim \begin{cases} 0 & \textit{with probability of } \varphi \quad (\textit{Group } B) \\ f(y_i|X_i) & \textit{with probability of } 1-\varphi \ (\textit{Group } A) \end{cases}$$

The φ indicates the probability that individuals come from the latent Group A or Group B. Accordingly, the probability of y_i number of check-ins for the captured health-related lifestyle behavior is equal to:

$$P(Y_i = y_i|X_i, Z_i) = \begin{cases} \varphi(\gamma Z_i) + [1-\varphi(\gamma Z_i)]f(0|X_i) & \textit{if } y_i = 0 \\ [1-\varphi(\gamma Z_i)]f(y_i|X_i) & \textit{if } y_i > 0 \end{cases}$$

where X_i is the vector of independent variables for individual i, and Z_i is a vector of covariates that contribute to the generation of zeros by not reporting the behavior – in our case X_i , and Z_i refer to the same vector of individuals' interests; γ is the vector of estimated zero-inflated coefficients.

4.6.3. Estimation Result

We used R for model estimation (R Development Core Team 2016). Tables 4.4 – 4.6 report the relationships of individuals' interests with their online observed behavior associated with fitness center & gym, bar, and fast food restaurant venues respectively. The coefficients in the count section of the tables show the significance and importance of each of individuals' interests in explaining their observed health-related lifestyle behavior. The coefficients in zero-inflated part of the tables determine the odds of being the member of latent Group B.

Table 4.4. Estimated Model: Healthy Lifestyle Behaviors (Fitness Center & Gym)

Variables		Group A/Count		Group B/Zero Inflated	
		Model 1	Model 2	Model 1	Model 2
Control Variables	Network Size	**-2.875****	-2.247	-50.525	-12.400
	Gender	**0.071****	**0.113*****	252.725	0.580
	SES	**0.071*****	**0.080*****	-0.201	0.024
Travel & Entertainment	Duck Dynasty		-0.068		-1.435
	Film & Music		**-1.373*****		**10.110*****
	Hip Hop Music		**-0.648****		**-30.760****
	Sport Media		**-1.544*****		**-21.560****
	Travel		**-0.411***		0.787
Technology	Apple		-0.161		0.970
	Automobile		0.095		-3.785
	Social Technology		**-3.386*****		3.123
	Mobile Technology		**-0.608****		-0.442
Business	Business & Finance		**1.019*****		-3.110
	Ecommerce & Online Com		**2.526*****		1.922
	Marketing & Influencers		**0.593***		**-14.770****
Fashion & Art	Art		**-0.684*****		**-5.137****
	Design		0.200		2.114
	Fashion		0.135		**-20.190***
Politics	Democratic Party		0.243		2.226
	Republican Party		0.092		**2.341****
Location	Austin		**-0.727*****		-0.120
	Boston		**-0.283***		1.020
	Dallas		**-0.288****		-0.046
	Las Vegas		0.281		0.879
	Los Angeles		-0.268		-0.483
	New York City		0.038		2.018
	San Francisco		-0.214		0.638
	Washington DC		-0.192		-0.210
Sports Team	Boston Sports		0.149		1.062
	Cleveland Sports		-0.281		0.259
	Detroit Sports		0.010		-11.640
	New York Sports		-0.021		-0.064
	Milwaukee Sports		0.161		0.102
	Philadelphia Sports		-0.157		1.175
	Pittsburgh Sports		0.039		1.345
	St. Louis & Minneapolis Sports		-0.232		-11.800
Food	Brewery		**-1.830*****		**2.490*****
	Food & Cooking		**-0.426****		0.748
	Restaurants & Chain Stores		**-0.983*****		1.943
Sport	Baseball		0.007		**-13.950***
	Basketball & Football		**2.776*****		**-59.600*****
	Car Racing		**-0.332****		0.642
	Fitness & Health		**2.718*****		-1585.000
	Hockey		-0.019		-4.015
	Soccer & Tennis		-0.001		**-13.800****
	Wrestling & Fighting		**0.431****		-2.148
	Constant	**0.720*****	**0.777*****	-255.592	**-1.157*****
	Log likelihood			**-52873**	**-52292**
	Wald x^2			**35.68**	**660.5**

N=32,700; *p<.05; **p<.01; ***p<.001

Table 4.5. Estimated Model: Unhealthy Lifestyle Behaviors (Bar)

Variables		Group A/Count		Group B/Zero Inflated	
		Model 1	Model 2	Model 1	Model 2
Control Variables	Network Size	**-3.336*** **	-1.202	-6.380	-100.612
	Gender	**0.040** **	**0.091*** **	**148.671***	0.296
	SES	**-0.089*** **	**-0.085*** **	**-5.606*** **	**-0.301** **
Travel & Entertainment	Duck Dynasty		**-0.276*** **		**2.102*** **
	Film & Music		**2.277*** **		**-58.682*** **
	Hip Hop Music		**-0.441*** **		-1.986
	Sport Media		**0.739*** **		-1.760
	Travel		**0.535*** **		**-25.747*** **
Technology	Apple		**-1.175*** **		**11.949*** **
	Automobile		-0.008		**3.857*** **
	Social Technology		**-0.712*** **		**-46.682*** **
	Mobile Technology		**-0.286***		2.499
Business	Business & Finance		0.264		**8.377*** **
	Ecommerce & Online Com		**-1.307*** **		**16.103*** **
	Marketing & Influencers		**-0.764*** **		5.042
Fashion & Art	Art		-0.129		-27.556
	Design		-0.006		2.673
	Fashion		**0.318** **		**-316.63*** **
Politics	Democratic Party		**-0.834*** **		2.680
	Republican Party		**-0.785*** **		**2.594** **
Location	Austin		0.118		-362.716
	Boston		**0.236*** **		-1.254
	Dallas		0.060		**2.110** **
	Las Vegas		**1.019*** **		**-336.800***
	Los Angeles		-0.135		-2.817
	New York City		**1.075*** **		-233.873
	San Francisco		**0.310*** **		-4.383
	Washington DC		**0.627*** **		-324.034
Sports Team	Boston Sports		0.025		-12.765
	Cleveland Sports		**0.262*** **		0.374
	Detroit Sports		**0.168***		-0.663
	New York Sports		**0.156***		-280.880
	Milwaukee Sports		**0.396*** **		**-12.862** **
	Philadelphia Sports		**0.477*** **		-3.442
	Pittsburgh Sports		0.018		**-8.012** **
	St. Louis & Minneapolis Sports		**0.379*** **		-0.529
Food	Brewery		**1.432*** **		**-1318.073***
	Food & Cooking		**0.337*** **		**-41.411*** **
	Restaurants & Chain Stores		**-0.697*** **		**19.498*** **
Sport	Baseball		**-0.572*** **		**-8.116** **
	Basketball & Football		**-0.777*** **		**9.398** **
	Car Racing		**-0.341*** **		-1.681
	Fitness & Health		**-0.684*** **		**8.619*** **
	Hockey		**0.225***		0.159
	Soccer & Tennis		0.012		-2.410
	Wrestling & Fighting		**-0.261** **		**3.084** **
	Constant	**1.062*** **	**1.045*** **	**-168.803** **	**-2.662*** **
	Log likelihood			**-73636**	**-72804**
	Wald x^2			**135.38**	**1305.8**

N=32,700; *p<.05; **p<.01; ***p<.001

Table 4.6. Estimated Model: Unhealthy Lifestyle Behaviors (Fast Food Restaurant)

Variables		Group A/Count		Group B/Zero Inflated	
		Model 1	Model 2	Model 1	Model 2
Control Variables	Network Size	**-2.588*****	-1.066	-750.820	274.382
	Gender	**0.207*****	**0.179*****	98.120	52.942
	SES	0.002	0.001	-239.65	-12.627
Travel & Entertainment	Duck Dynasty		-0.030		-56.610
	Film & Music		**1.186*****		-182.126
	Hip Hop Music		**0.246***		73.936
	Sport Media		-0.325		-614.073
	Travel		-0.111		**113.672***
Technology	Apple		-0.201		9.686
	Automobile		**-0.507*****		-105.411
	Social Technology		**-3.230*****		**173.115***
	Mobile Technology		**0.843*****		**-615.188***
Business	Business & Finance		0.050		**129.457***
	Ecommerce & Online Com		**0.598****		379.308
	Marketing & Influencers		0.291		**180.681***
Fashion & Art	Art		**-0.231****		-33.359
	Design		0.187		-313.567
	Fashion		**-0.742*****		**158.326***
Politics	Democratic Party		**-0.679*****		16.152
	Republican Party		**0.397*****		-314.702
Location	Austin		**-0.178***		-38.053
	Boston		**-0.241****		-60.095
	Dallas		0.067		-101.723
	Las Vegas		**-0.188***		**101.085***
	Los Angeles		0.060		-151.885
	New York City		0.015		-248.500
	San Francisco		-0.059		-179.696
	Washington DC		**-0.417*****		11.733
Sports Team	Boston Sports		**-0.307*****		**92.095***
	Cleveland Sports		**0.259*****		-181.091
	Detroit Sports		0.027		56.271
	New York Sports		-0.038		47.009
	Milwaukee Sports		-0.059		33.140
	Philadelphia Sports		-0.012		65.653
	Pittsburgh Sports		-0.116		**-66.968***
	St. Louis & Minneapolis Sports		0.097		-229.603
Food	Brewery		**-0.266****		-377.953
	Food & Cooking		**0.217***		-301.590
	Restaurants & Chain Stores		**3.047*****		-1189.716
Sport	Baseball		**0.314****		-184.443
	Basketball & Football		-0.217		175.210
	Car Racing		**0.402*****		**34.313***
	Fitness & Health		**-1.010*****		**182.866***
	Hockey		**0.346*****		**157.067***
	Soccer & Tennis		-0.128		**146.533***
	Wrestling & Fighting		**0.348*****		-36.352
	Constant	**0.744*****	**0.730*****	-750.820	**-156.476***
	Log likelihood			**-66877**	**-66202**
	Wald x^2			**210.09**	**1378.2**

N=32,700; *p<.05; **p<.01; ***p<.001

Healthy Lifestyle Behaviors (Fitness Center & Gym). The results in Group A/Count section of Table 4.4 show that having interest in three communities of interest in the sport category (healthy interest) have positive significant association (Basketball & Football, Fitness & Health at $p<.001$, and Wrestling & Fighting at $p<.05$) with the self-disclosed numbers of check-ins at fitness center & gym venues. Thus, these three healthy interests provide support for H1a. However, we could not find a significant association between self-disclosed healthy lifestyle behaviors and healthy interests related to Baseball, Hockey, or Soccer & Tennis. We also find a significant negative association (at $p<.01$) between Car Racing and the number of check-ins at fitness center and gym. This result can be due to the different nature of an interest in Car Racing and interests in other types of sport. Additionally, results in Zero Inflated data imply that people who have interests in Baseball, Soccer & Tennis, and Basketball & Football have significantly higher likelihood (Baseball at $p<.05$, Soccer & Tennis at $p<.01$, Basketball & Football at $p<.001$) to share their healthy lifestyle behavior.[26] This finding shows that sharing check-ins in fitness center and gym is not against the norms of sport communities of interest.

The second category of interest that we developed a hypothesis on is Food. The results in Group A/Count section of Table 4.4 show that all communities of interest in the Food category have a negative significant association with the reported number of check-ins in fitness center and gym (Brewery, Restaurants & Chains Stores at $p<.001$ and Food & Cooking at $p<.01$). This result provides support for H1b, where we argued that individuals' disclosed-healthy lifestyle behaviors are negatively associated with observed healthy interests of individuals. We also found there is a significantly ($p<.001$) lower probability of sharing check-ins inside fitness centers &

[26] A significant negative coefficient for a factor in the Zero Inflated part of tables indicates that the larger value of the factor decreases the probability of inhibition to report the associated behavior (increases the probability of reporting the behavior) in online social networks.

gyms for people who have a higher level of interest in alcohol beverages (Brewery).[27] This effect provides partial support for H3, where we argue that sharing healthy lifestyle behaviors in online social networks is against the norm for people who belong to unhealthy communities of interest.

The remaining categories of interest communities represent control variables. The first category is Travel & Entertainment. Communities of interest within the Travel & Entertainment category have significant negative association with healthy lifestyle behaviors (except for Duck Dynasty). This finding indicates that a higher level of interest in Film & Music (at $p<.001$), Hip Hop Music (at $p<.01$), Sport Media (at $p<.001$), and Travel (at $p<.05$) is associated with lower levels of disclosed check-ins in fitness center & gym venues within online social networks. Zero Inflated part of the table also reveals that people who followed social pages related to Sport Media and Hip-Hop Music have a significantly higher (at $p<.01$) tendency to share their check-ins within fitness center & gym venues (if they visit to those places). Moreover, we found that a high interest to Film & Music can significantly ($p<.001$) reduce the chance of reporting activities within fitness center & gym venues. This fact indicates that reporting healthy lifestyle activities is not a norm for people in the Film & Music community of interest.

The second controlled category of interest is Technology. Results for technology-based interests reveal that individuals' interests in mobile and social technologies have negative significant association (Social Technology at $p<.001$, and Mobile Technology at $p<.01$) with numbers of check-ins at fitness center & gym venues. Zero Inflated part of the table does not show significant results for communities of interest in this category.

---

[27] A significant positive coefficient for a factor in the Zero Inflated part of tables indicates that a larger value of the factor increases the probability of inhibition to report the associated behavior (reduces the probability of reporting the behavior) in online social networks.

The other controlled interest category is dedicated to business communities of interest. Results in Table 4.4 show that having interests in business related social pages is positively associated (Business & Finance, E-commerce and Online communities at $p<.001$ and Marketing & Influencers at $p<.05$) with the number of times that individuals have participated in healthy lifestyle activities within fitness center & gym venues. Zero Inflated part of Table 4.4 also indicates that having interests in social pages within the Marketing & Influencers community of interest significantly increases ($p<.01$) the likelihood of sharing healthy lifestyle behaviors in online social networks.

The fourth category of interest is Fashion & Art. We found that higher level of interest to social pages within a Art community of interest is negatively and significantly associated ($p<.001$) with the number of check-ins in fitness centers & gyms. Zero Inflated results also show that people who have developed an interest in Fashion and Art have significantly (Fashion at $p<.05$, and Art at $p<.01$) higher tendency to share their healthy lifestyle behaviors related to fitness center and gym.

Another set of results is about Politics. The results reveal that having interests to politics is not associated with level of healthy lifestyle behaviors. However, the Zero Inflated part of Model 2 indicates that individuals with higher tendency toward republican party have lower inclination to share their healthy behaviors in online social networks.

The next sets of interest categories are related to location and sport teams. The first category represents communities of interest related to several large cities inside the United States. Our results indicate that interests in two prominent cities in Texas (Dallas and Austin) are negatively associated (at $p<.01$) with healthy lifestyle behaviors. We find similar results for Boston, where findings show that having interest in social pages representing the City of Boston is negatively associated with lower number of check-ins at fitness center & gym venues. We could not find any significant result in Zero Inflated part for this category. Finally, results indicate that having interest in different sport teams does not have association with neither the number of check-ins nor with tendency of them to report their healthy behaviors.

The other set of control variables are demographics and network size. Our results indicate that men report significantly higher (at $p<.001$) numbers of check-ins at fitness center & gym venues. We also found that SES is positively associated (at $p<.001$) with self-reported healthy behaviors in online social networks. We could not find significant results for the network size variable.

Unhealthy Lifestyle Behaviors (Bar). Table 4.5 shows the relationship between communities of interest and unhealthy lifestyle behaviors associated with bar venues. The first category is Food. Group A/Count results show that interest to Brewery companies is significantly (at $p<.001$) associated with higher numbers of check-ins in bar venues. Our findings show similar significant positive association at $p<.001$ for the Food & Cooking community of interest. However, in the case of Restaurants & Chain Stores, we surprisingly found that a higher level of interests in this community of interest can lead to a significantly lower level of check-ins in bar venues at the level of $p<.001$. The final community of interest in this category is Travel. Results show that people who have a higher level of interest in social pages related to traveling are significantly (at $p<0.001$) more inclined to go to bar venues. Group B/Zero Inflated results only show a significant coefficient (negative significant at $p<0.001$) for Brewery in this category. Thus, higher levels of interest in Brewery companies in online social networks is associated with a higher level of willingness to share check-ins from bar venues. This result supports our H2a hypothesis and shows that interests in alcohol beverages is positively associated with a higher number of observed unhealthy behaviors related to Bar. Group B/Zero Inflated results also indicate that people with higher levels of interest to Brewery and Food & Cooking have a significantly higher (Brewery at $p<.05$, and Food & Cooking at $p<.001$) tendency to self-disclose their check-ins in bar venues. Similar to the Count result, we observed that the Restaurant & Chain Stores category shows distinct results from the other two communities of interest in this category. The result show that people who have a higher level of interest Restaurants & Chain Stores have a lower inclination to self-report their unhealthy behaviors related to bars in online social networks. This indicates that posting about unhealthy behaviors related to bars is not a norm for people within Restaurant & Chain Stores community of interest.

The second category is sport. The Count section of Table 4.5 shows that all sport communities of interest – except form Hockey and Soccer &Tennis – have a significant negative association with individuals' unhealthy behaviors related to bar venues (Baseball, Basketball & Football, Car Racing, Fitness & Health at $p<.001$ and Wrestling & Fighting at $p<.01$). For the other two communities of interest, we found a marginal positive and significant coefficient for Hockey (at $p<.05$) and an insignificant coefficient for Soccer & Tennis. This finding provides partial support for H2b. Zero Inflated results also indicate that only people who have a higher interest in Baseball have significantly higher tendency to share their bar check-ins in online social networks. Other people with interests to Basketball & Football, Fitness & Health, and Wrestling and Fighting have a significantly lower tendency to share their bar-related unhealthy behaviors (Basketball & Football, Wrestling & Fighting at $p<.01$, and Fitness & Health at $p<.001$). This result shows that posting unhealthy behaviors related to a bar is not a norm for people with healthy interests related to above discussed communities of interest. This can be considered as partial support for H4.

The next category is Travel & Entertainment. This represents a controlled interest category. In this category, we obtained two distinct sets of results. The first set of results relates to Film & Music, Sport Media and Travel. Having interests in all three communities of interest have a positive and significant association (at $p<0.001$) with number of check-ins at bar venues. The Zero Inflated result also shows that interest in Travel and Film & Music is associated with a significantly higher (at $p<.001$) tendency to self-disclosed unhealthy behaviors related to bar. The second sets of result are dedicated to Duck Dynasty (a TV series that has characters with conservative Protestant Christian views) and Hip Hop Music. We found that individuals who have higher levels of interest in these two communities of interest have a significantly (at $p<0.001$) lower number of check-ins at bar venues. The results in Zero Inflated section of the table for this set also indicate that interest in the Duck Dynasty TV series is associated with (at $p<0.01$) with a lower tendency to share unhealthy behaviors at bar venues. In other words, as \ expected, sharing check-ins at bar venues is not a norm in the Duck Dynasty community of interest.

The second controlled category is Technology. Results show that people who have followed social pages related to Apple, Social technologies, and Mobile technologies have significantly lower (Apple, Social technologies at $p<.001$ and Mobile technologies at $p<.05$) numbers of check-ins at bar venues. Zero Inflated result also indicates that while interests in social technologies are associated with a significantly higher inclination to self-disclosed unhealthy lifestyle behaviors related to bars, interests in Apple and Automobile technologies are associated with a lower tendency to self-disclose check-ins at bar venues.

Business and Politics categories of interest show similar patterns. In the Business category, two communities of interest (Ecommerce & Online Community and Marketing & Influencers) have a negative significant association (at $p<.001$) with the number of check-ins at bars. Similarly, individuals' interests in both Republican and Democratic parties have a significant negative association (at $p<.001$) with reported numbers of check-ins in bar venues. The results also show similar patterns in Zero Inflated part, where people with higher levels of interest in Business & Finance, Ecommerce & Online Community and Republican Party have a significantly lower tendency (Business & Finance, Ecommerce & Online Community at $p<.001$ and Republican Party at $p<.01$) to reveal their behavior in online social networks.

In the Fashion & Art category, we could only find a significant result for the Fashion community of interest. Results in the Count part of the table in this category show that individuals' interest in Fashion has a positive significant relationship at $p<0.001$ with numbers of check-ins at bar venues. Zero Inflated data also shows that interest in Fashion significantly increases the tendency of individuals (at $p<.001$) to reveal their unhealthy check-ins at bar venues in online social networks.

The location category of interest also contains communities related to major cities in the United States. Results show that individuals who have followed a higher number of social pages related to most of these large cities (Boston, Las Vegas, New York City, San Francisco, Washington DC) have a significantly (for all at $p<.001$) higher level of check-ins in bar venues. Zero Inflated results also indicate that people with a higher level of interest in the Las Vegas community of interest have a significantly higher tendency (at $p<.05$) to share their behavior within bar venues. Our results also indicate that people who are following social pages related to Dallas have significantly lower tendency (at $p<.01$) to self-disclose their places in online social networks when they are in bars.

The final set of results about communities of interest relates to Sports Teams. We found that fans of sports teams have reported a significantly higher number of check-ins (Cleveland, Milwaukee, Philadelphia, St. Louis & Minneapolis at $p<0.001$, and Detroit and New York at $p<0.05$) at bar venues. Zero Inflated findings also indicate that there is a significantly higher likelihood for supporters of sports teams in Pittsburgh and Milwaukee (both at $p<0.01$) to self-report unhealthy lifestyle behaviors related to bar venues.

As with healthy behaviors, we could not find significant results for association between network size and self-disclosed unhealthy behaviors related to bar venues. We also found that men share a significantly higher number of check-ins (at $p<.001$) at bar venues. Our findings also show that SES is negatively associated (at $p<.001$) with unhealthy lifestyle behaviors related to bar venues.

Unhealthy Lifestyle Behaviors (Fast Food Restaurant). Table 4.6 shows the result of analysis for relationships between communities of interest and lifestyle behaviors related to fast food restaurants. The first category that we discuss in this section is Food. Results indicate that there are positive and significant associations between two food-related communities of interest (Food & Cooking at $p<0.01$ and Restaurants & Chain Stores at $p<0.001$) and the number of check-ins at fast food restaurants. We also found a significant negative coefficient (at $p<.01$) for interest in Brewery. This result indicates that having interest in alcoholic beverages has direct a negative association with unhealthy behaviors related to fast food restaurants.

The second category is Sport. In contrast with our expectation, results show that fans of most sports have a significantly higher number of check-ins (Baseball, Car Racing, Hockey, and Wrestling & Fighting at $p<.001$) at fast food restaurants. Fitness & Health is the only community of interest in this category with a significant negative association (at $p<.001$) with unhealthy lifestyle behaviors related to fast food restaurants. Therefore, we could only find marginal support for H2b in unhealthy lifestyle behaviors related to fast food restaurants. On the other hand, Zero Inflated result indicates that people who have an interest in sport-related communities have a significantly lower tendency (Car Racing, Fitness & Health, Hockey, and Soccer & Tennis at $p<.05$) to share their check-ins in fast food restaurants within online social networks. This finding indicates that self-disclosure of unhealthy lifestyle behaviors related to fast food restaurants is not a norm.

Next, we discuss Travel & Entertainment. The Group A/Count shows that Film & Music and Hip Hop Music are two communities of interest that have significant positive relationships (Film & Music at $p<.001$, Hip Hop Music at $p<.05$) with the number of check-ins at fast food restaurants. The Zero-Inflated result also indicates that there is a significantly lower likelihood ($p<0.05$) for people who have an interest in Travel to share their unhealthy lifestyle behavior related to fast food restaurants in online social networks.

The second controlled category is Technology. We found that interests in Automobiles and Social Technology are two factors that are negatively and significantly (Automobiles at $p<0.01$ and Social Technology at $p<0.001$) associated with the number of check-ins at fast food restaurants. In contrast, results indicate that individuals with higher levels of interest in Mobile Technology are going significantly more (at $p<0.001$) to fast food restaurants. Zero Inflated result also indicates that individuals with interest in Social Technology have a significantly lower tendency (at $p<.05$) and those with interest in Mobile Technology have a significantly higher tendency (at $p<.05$) to share their check-ins inside fast food restaurants with their peers in online social networks.

The third controlled category of communities of interest is Business. Among communities of interest in this category only Ecommerce & Online Communities have a significant coefficient (at $p<.01$). This finding indicates that there is a positive association between having an interest in Ecommerce & Online Communities and going to fast food restaurants. However, Zero Inflated result shows that there is a significantly lower chance ($p<0.05$) of sharing check-ins from fast food places into online social networks for people with interest in Business & Finance and Marketing & Influencers.

The next category is Fashion & Art. Except for Design, communities of interest in this category have a significant negative association with the number of check-ins at fast food restaurants (Fashion at $p<.001$ and Art at $p<.01$). Zero Inflated result also shows that people who have a higher level of interest in social pages related to Fashion have a significantly lower (at $p<.05$) desire to share their check-ins at fast food restaurants.

The other controlled category of interest is Politics. In contrast to other last two discussed lifestyle behaviors that belonging to communities of interest related to both Democratic Party and Republican party had consistent outcomes, we get different results for Fast Food lifestyle behaviors. On one hand, people with higher levels of interest in Democratic Party have a significantly lower (at $p<.001$) number of check-ins at fast food restaurants. On the other hand, interest in Republican Party is significantly associated with a higher (at $p<.001$) number of check-ins at fast food restaurants. We could not find any significant result in the Zero Inflated part of the table for Politics.

The sixth controlled category is Location. Our results indicate that individuals who are interested in social pages about Austin, Boston, Las Vegas and Washington DC have a significantly lower number of check-ins (Austin and Las Vegas at $p<.05$, Boston at $p<.01$, and Washington DC at $p<.001$) at fast food restaurants. Zero-inflated result also indicates that individuals with interest in Las Vegas have a significantly lower (at $p<.05$) tendency to share their fast food restaurant check-ins in online social networks.

The final category is Sports Team. The Count result shows that fans of sports teams in Boston have a significantly lower number of check-ins (at $p<.001$) and fans of sports teams in Cleveland have a higher number of check-ins (at $p<.001$) at fast food restaurants. We could not find significant results for other sports teams. Zero Inflated result also shows that fans of Boston sports teams have a significantly lower (at $p<.05$) inclination to share their fast food check-ins in online social networks. In contrast, findings show that fans of Pittsburgh sports teams have a significantly higher (at $p<.05$) desire to share their fast food check-ins in online social networks.

For the demographic and social network size control variables, we could only find a significant result for gender. We found that as with the other two lifestyle behaviors, men have a significantly higher number of check-ins at fast food restaurants. Table 4.7 summarizes tests of our hypotheses.

Table 4.7 – Supported Hypotheses

	Healthy Behavior		**Unhealthy Behavior**			
Healthy Interest	**Gym & Fitness Center**		**Bar**		**Fast Food Restaurant**	
	H1a		**H2b**	**H4**	**H2b**	**H4**
Baseball	No		Yes	No	No	No
Basketball & Football	Yes		Yes	Yes	No	No
Car Racing	No		Yes	No	No	Yes
Fitness & Health	Yes		Yes	Yes	Yes	Yes
Hockey	No		No	No	No	Yes
Soccer & Tennis	No		No	No	No	Yes
Wrestling & Fighting	Yes		Yes	Yes	No	No
Unhealthy Interest	**Gym & Fitness Center**		**Bar**		**Fast Food Restaurant**	
	H1b	**H3**	**H2a**		**H2a**	
Brewery	Yes	Yes	Yes		No	
Restaurants & Chain Stores	Yes	No	No		Yes	

4.7. Discussions

This study's first objective was to introduce an approach to detect communities of interest that reflects individuals' personal interests. We achieve this goal by developing a homophily-based Interest Detection method (HID). Our method relies on selective exposure theory and the concept of homophily in social science as its theoretical bases. It also takes advantage of structural features of online social networks and detects individuals' interests by applying a community detection algorithm to an extended bipartite graph of individuals' relationships in online social networks. Applying HID model to an extended bipartite graph of individuals in Twitter containing 32,700 individuals and 4,893 social pages shows that this method can successfully capture communities of interest within the online platform. The second objective of this study was to investigate the relationship between individuals' observed interests and their disclosed health-related lifestyle behaviors within online social networks. We captured individuals' disclosed health-related lifestyle behaviors based on their location-based check-ins within online social networks. Our estimation models show that self-disclosed healthy and unhealthy lifestyle behaviors of individuals in online social networks are consistent with their observed interests and have direct significant relationships with them. Our study had several important findings.

First, we found that in addition to the nature of interests which can determine the level of individuals' disclosure about their health-related lifestyle behaviors, the existing norms in communities of interest also play an important role in individuals' decision to share health-related lifestyle behaviors within online social networks. This is a novel finding, documenting the importance of communities of interest in formation of individuals' self-disclosure behaviors in online social networks.

Second, the empirical results of our study show that Fitness & Health community of interest as the best representative of individuals' healthy interest is positively associated with individuals' healthy lifestyle behaviors and negatively with both unhealthy lifestyle behaviors (bar and fast food restaurants) of them. This important and novel finding shows that Fitness & Health social pages can play critical roles in the promotion of healthy behaviors among online users. Moreover, our estimations show that while people belonging to a Fitness & Health community may participate in unhealthy lifestyle activities, disclosing those behaviors is against norms in those communities. Considering the impact of social influence in propagation of unhealthy behaviors (Essay 1), healthy communities of interest can indirectly help to control diffusion of unhealthy lifestyle behaviors among users of online social networks.

Third, our study distinguishes between interests of individuals who self-disclosed their bar check-ins and those who disclosed their fast food restaurant check-ins. Estimated models show that interest in brewery social pages is positively associated with the number of check-ins at bars and negatively associated with the number of check-ins at fast food restaurants. We found similar result for interest in fast food restaurants, where people with higher interests in social pages related to fast food restaurants had a higher number of check-ins at fast-food restaurants and a lower number of check-ins at bars. This indicates that while lifestyle behaviors related to fast food restaurants and bars both represent unhealthy lifestyle behaviors, they have different natures that can even be contradictory.

Fourth, results revealed that interest in location-based communities for large U.S. cities are associated with a lower level of self-reported healthy lifestyle behaviors (Austin, Boston and Dallas) and a higher level of self-disclosed behaviors related to bar venues (Boston, Las Vegas, New York City, San Francisco, and Washington DC). This result indicates that healthy lifestyle behaviors in large cities require additional attention. Thus, it is important that health promotional programs target individuals who live in large cities.

Fifth, our findings indicate that followers of sports teams have a significantly higher number of check-ins at bar venues. This is another important finding indicating that sports fans are more vulnerable to alcohol drinking problems. Therefore, it is essential that healthy lifestyle promoters consider the interest of individuals and offer appropriate alternative choices to them.

Sixth, the results of the estimated model for healthy lifestyle behavior reveal that interest in Social Technology communities of interest is the most important factor in reducing the number of check-ins at fitness center & gym venues. This is another novel finding indicating that excessive interest in Social Technology reduces individuals' levels of physical activity.

Seventh, our results indicate that people belonging to business-related communities of interest have healthier life style behaviors in comparison with most other interest communities. Estimated models show that all three communities of interest in this category have positive relationships with the number of check-ins at fitness centers and gyms. People in these communities (E-commerce & Online community, and Marketing & Influencers) also have a lower number of check-ins at bar venues. The results also indicate that revealing check-ins about unhealthy lifestyle behaviors is generally not a norm in these communities.

Finally, the empirical results of our study show that men are more likely than women to disclose their lifestyle behaviors within online social networks. This finding is in contrast with the findings of Wang et al. (2016), who found that men have a lower tendency for self-disclosure.

4.8. Implications

4.8.1. Theoretical Contributions

This study makes several contributions to theory and research. First, our Homophily-based interest detection method (HID) introduces a theory-based community detection approach that can be used for collection of individuals' interests form online social networks. To the best of our knowledge this is the first study that offers such a comprehensive method for detection of individuals' personal interests solely based on the structure of their social networks. This demonstrates the great potential of online social networks for studying the effect of individuals' personal interests in different contexts.

Second, our study is the first to focus on observed disclosed behaviors within online social networks. Prior research used interviews and surveys to conduct their studies. While those studies are helpful in understanding the nature of self-disclosure in online social networks, they suffer from elements of social desirability and hawthorn effects. Our observational study eliminates those sources of bias in data and confirms that disclosed behaviors of individuals could be a good representative of their observed interests.

Third, our study shows that the nature of individual's interests and existing norms within online communities of interest can play important roles in disclosure of behaviors within online social networks. Additionally, the methodological approach introduced in this study offers an analytical method to distinguish between the effects of the interest communities and presented norms within those communities. This has a broad implication in future studies of individuals' self-disclosure in online social networks.

4.8.2. Practical and Policy Implications

The results of our study offer a number of important implications for practitioners and policy makers. Massive growth and pervasive use of online social networks make these platforms ideal for targeting individuals to promote different products or behaviors. In recent years, social marketing programs such as VERB and TRUTH have developed health promotion programs using a traditional marketing approach to target individuals in online social networks. The main goals of these programs are to promote healthy lifestyle behaviors and encourage people to stop their unhealthy lifestyle behaviors by offering alternative healthy lifestyle choices. However, these programs cannot easily find their target audience and have limited information about the motivational background of targeted users. Our study provides a practical approach for detection of communities of interest associated with healthy and unhealthy lifestyle of people. This helps health practitioners to not only find vulnerable people inside online social networks, but also understand their main interests. This approach provides good sources to them to develop alternative lifestyle choices.

Our work also shows that social pages in online social networks can play significant roles in the formation of individuals' lifestyle behaviors. Specifically, our findings show that people who have followed social pages promoting alcoholic beverages have a higher tendency to go to bar venues, and individuals who have followed fast food restaurant pages show a higher inclination to go to fast food restaurants, and those who have showed interest in health and fitness social pages had higher numbers of check-ins at fitness center and gym venues. This is another reason for the importance of attention to online social networks for promoting health behaviors. In fact, policy makers can create regulations to control the activity of social pages that promote unhealthy lifestyle behaviors, and also provide facilities for healthy social pages to expand their activities within online social networks.

Finally, the proposed approach for detection of interest communities can be widely applied to different contexts such as marketing and psychology. Practitioners in those fields can use communities of interest to evaluate individuals' behaviors based on specific interests and the norms of these communities of interest.

4.9. Limitations and Future Research Direction

Similar to other empirical studies, our research has some limitations. First our dataset is limited to self-disclosed behaviors of users who disclosed their location-based activities from Foursquare into their Twitter accounts. Therefore, interpretation of our results was limited to the captured population in our study. Future studies might be able to collect data from multiple locations based on social networks and further investigate the topics discussed in this paper. Second, in the development of the HID model we only considered relationships between strong ties (two-way relationships). Future studies can also consider the role of one-way relationships in formation of communities of interest within online social networks. Third, in this study we referred to Clauset et al. (2004) and limit the granularity of modularity to 0.3. The future extension of this work can consider higher levels of granularity and investigate the role of larger numbers of communities of interest in the self-disclosed behavior of individuals.

CHAPTER 5

Conclusion

This dissertation makes a number of significant contributions to theory and practice. First, it offers a dynamic sequential approach for capturing, extracting and integrating online social network public data and the derivation of healthy and unhealthy lifestyle behaviors. This approach provides a new venue for studying individuals health-related lifestyle behaviors in a large scale. Second, it offers a new theory-based Health-related Lifestyle Behavior (HLB) model which provides a conceptual framework for studying online self-disclosed behaviors and the social factors that could influence them. Third, it provides new insights about the role of images on formation of individuals' health-related lifestyle behaviors as observed in online social networks. Fourth, this work formulates a sequential approach for analysis and categorization of images in different contexts. Fifth, it develops a Homophily-based interest detection (HID) method that can be used for detection of wide variety of individuals' interests and preferences within online social networks. Sixth, the results of this study show that established norms within online communities of interest, and the nature of individuals' interests and preferences can play important roles in disclosure of health-related lifestyle behaviors within online social networks.

REFERENCES

Aceves-Martins M, Llauradó E, Tarro L, Moreno-García CF, Escobar TG, Solà R, Giralt M (2016) Effectiveness of social marketing strategies to reduce youth obesity in European school-based interventions: a systematic review and meta-analysis. *Nutrition reviews* **74**(5) 337-351.

Achrekar H, Gandhe A, Lazarus R, Yu SH, Liu B (2011) Predicting flu trends using Twitter data. In *Proceedings of the Computer Communications Workshops 2011 IEEE Conference* 702-707.

Adair JG (1984) The Hawthorne effect: A reconsideration of the methodological artifact. *Journal of applied psychology*, **69**(2), 334-345.

Adler, NE., Boyce T, Chesney MA, Cohen S, Folkman S, Kahn RL, Syme SL (1994) Socioeconomic status and health: The challenge of the gradient. *American Psychologist* **49**(1) 15-24.

Agarwal R, Animesh A, Prasad K (2009) Research note-social interactions and the "digital divide": Explaining variations in Internet use. *Information Systems Research* **20**(2) 277-294.

Agrawal R, Imieliński T, Swami A (1993) Mining association rules between sets of items in large databases. In proceedings of SIGMOD **22**(2) 207-216. ACM.

Ainley M, Hidi S, Berndorff D (2002). Interest, learning, and the psychological processes that mediate their relationship. *Journal of Educational Psychology* **94** 545–561.

Almaatouq A, Radaelli L, Pentland A, Shmueli E (2016) Are you your friends' friend? Poor perception of friendship ties limits the ability to promote behavioral change. *PloS one* **11**(3) p.e0151588.

Althoff T, White RW, Horvitz E (2016) Influence of Pokémon Go on physical activity: study and implications. *Journal of medical Internet research* **18**(12) e315.

Amichai-Hamburger Y, Vinitzky G (2010) Social network use and personality. *Computers in human behavior* **26**(6) 1289-1295.

Anderson CL, Agarwal R (2011) The digitization of healthcare: boundary risks, emotion, and consumer willingness to disclose personal health information. *Information Systems Research* **22**(3) 469–49.

Andrews JA, Tildesley E, Hops H, Li F (2002) The influence of peers on young adult substance use. *Health Psychology* **21**(4) 349-357.

Appel L, Dadlani P, Dwyer M, Hampton K, Kitzie V, Matni ZA, Moore P, Teodoro R (2014) Testing the validity of social capital measures in the study of information and communication technologies. *Information, Communication & Society* **17**(4) 398-416.

Aral S (2011) Commentary-identifying social influence: A comment on opinion leadership and social contagion in new product diffusion. *Marketing Science* **30**(2) 217-223.

Aral, S. and Walker (2012) Identifying influential and susceptible members of social networks. *Science* **337**(6092) 337-341.

Aral S, Walker D. (2014) Tie strength, embeddedness, and social influence: A large-scale networked experiment. *Management Science* **60**(6) 1352-1370.

Argamon S, Koppel M, Pennebaker JW, Schler J (2009) Automatically profiling the author of an anonymous text. *Communications of the ACM* **52**(2) 119-123.

Arruda, S.P.M., da Silva, A.A.M., Kac, G., Vilela, A.A.F., Goldani, M., Bettiol, H. and Barbieri, M.A., (2016) Dietary patterns are associated with excess weight and abdominal obesity in a cohort of young Brazilian adults. *European Journal of Nutrition* **55**(6), 2081-2091.

Artinian NT, Fletcher GF, Mozaffarian D, Kris-Etherton P, Van Horn L, Lichtenstein AH, Kumanyika S, Kraus WE, Fleg JL, Redeker NS, Meininger JC (2010) Interventions to promote physical activity and dietary lifestyle changes for cardiovascular risk factor reduction in adults. A scientific statement from the American Heart Association. *Circulation* **122** 406-441.

Back MD, Schmukle SC, Egloff B (2008) Becoming friends by chance. *Psychological Science* **19**(5) 439-440.

Bandura A (1969) Social-learning theory of identificatory processes. *Handbook of socialization theory and research* 213-262.

Bandura A (1986) *Social Foundations of Thought and Action: A Social Cognitive Theory*. Prentice-Hall.

Bapna R, Umyarov A (2015) Do your online friends make you pay? A randomized field experiment on peer influence in online social networks. *Management Science* **61**(8) 1902-1920.

Bapna R, Gupta A, Rice S, Sundararajan A (2017) Trust and the strength of ties in online social networks: An exploratory field experiment, *MIS Quarterly* **41**(1) 115-130.

Barber BM, Odean T (2007) All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The review of financial studies* **21**(2) 785-818.

Bargh JA, McKenna KY, Fitzsimons GM (2002). Can you see the real me? Activation and expression of the “true self” on the Internet. *Journal of social issues* **58**(1) 33-48.

Barkhuus L, Brown B, Bell M, Hall M, Sherwood S, Chalmers M (2008). From Awareness to Repartee : Sharing Location within Social Groups. In *proceedings of the SIGCHI conference on human factors in computing systems* 497-506

Barnes JA (1954) Class and committees in a Norwegian island parish. *Human Relations* **7** 39-58.

Barnett E, Casper M (2001) A definition of 'social environment'. *American Journal of Public Health* **91**(3) 465.

Barrera M, Ainlay SL (1983) The structure of social support: A conceptual and empirical analysis. *Journal of Community Psychology* **11**(2) 133-143.

Barrera Jr M (1986) Distinctions between social support concepts, measures, and models. *American Journal of Community Psychology* **14**(4) 413-445.

Barry, A. M. (1997). Visual intelligence: Perception, image, and manipulation in visual communication. Albany: State University of New York Press.

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

Bateman J, Wildfeuer J, Hiippala T (2017) Multimodality: Foundations, Research and Analysis– A Problem-Oriented Introduction. Walter de Gruyter GmbH & Co KG

Bazarova NN, Choi YH (2014) Self-disclosure in social media: Extending the functional approach to disclosure motivations and characteristics on social network sites. *Journal of Communication* **64**(4) 635-657.

Beaver JD, Lawrence AD, van Ditzhuijzen J, Davis MH, Woods A, Calder AJ (2006). Individual differences in reward drive predict neural responses to images of food. *Journal of Neuroscience* **26**(19) 5160-5166.

Berkman LF, Glass T, Brissette I, Seeman TE (2000) From social integration to health: Durkheim in the new millennium. *Social Science & Medicine* **51**(6) 843–857.

Bayliss EA, Bonds DE, Boyd CM, Davis MM, Finke B, Fox MH, Glasgow RE, Goodman RA, Heurtin-Roberts S, Lachenmayr S, Lind C (2014) Understanding the context of health for persons with multiple chronic conditions: Moving from what is the matter to what matters. *The Annals of Family Medicine* **12**(3) 260-269.

Blondel VD, Guillaume JL, Lambiotte R, Lefebvre (2008) Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* 2008(10) P10008.

Bloom JR (1990) The relationship of social support and health. *Social Sci. & Medicine*, **30**(5), 635-637.

Bond RM, Fariss CJ, Jones JJ, Kramer AD, Marlow C, Settle JE, Fowler JH (2012) A 61-

million-person experiment in social influence and political mobilization. *Nature* **489**(7415) 295-298.

Borgatti SP, Mehra A, Brass DJ, Labianca G (2009) Network analysis in the social sciences. *Science* **323**(5916) 892-895.

Bott E (1957) *Family and Social Network*. Tavistock Press, London.

Bowlby J (1969) *Attachment and Loss: Attachment. V*. Basic Books.

Boyd D (2007) Why youth (heart) social network sites: The role of networked publics in teenage

Bray P (2012) When is my tweet's prime of life? A brief statistical interlude. The MOZ Blog. https://moz.com/blog/when-is-my-tweets-prime-of-life

Bruhn JG (1988) Life-style and health behavior. *Health behavior* 71-86.

Bucher HJ, Schumacher P (2006) The relevance of attention for selecting news content. An eye-tracking study on attention patterns in the reception of print and online media. *Communications* **31**(3) 347-368.

Bühler C (1957) About the goal-structure of human life. Theoretical considerations. *Dialectica* **11**(1-2) 187-205.

Bulloch A, Lavorato D, Williams J, Patten S (2012) Alcohol consumption and major depression in the general population: The critical importance of dependence. *Depression and Anxiety* **29**(12) 1058-1064.

Cameron AC, Trivedi PK (1990) Regression-based tests for overdispersion in the Poisson model. *Journal of Econometrics* **46**(3) 347-364.

Cameron AC, Trivedi PK (2013) Regression analysis of count data. **53** Cambridge University Press.

Caplan G (1974) *Support Systems and Community Mental Health: Lectures on Concept Development.* Behavioral Publications.

Carlson SA, Fulton JE, Pratt M, Yang Z, Adams EK (2015). Inadequate physical activity and health care expenditures in the United States. *Progress in cardiovascular diseases* **57**(4) 315-323.

Carrell SE, Hoekstra M, West JE (2011) Is poor fitness contagious? Evidence from randomly assigned friends. *Journal of public Economics* **95**(7) 657-663.

Caton, S. J., Ball, M., Ahern, A., Hetherington, M. M. (2004) Dose-dependent effects of alcohol on appetite and food intake. Physiology & Behavior. **81**(1) 51-58.

CDC (2015) www.cdc.gov/chronicdisease/overview.

Centola D (2010) The spread of behavior in an online social network experiment. *Science*, **329**(5996) 1194-1197.

Centola D (2011) An experimental study of homophily in the adoption of health behavior. *Science*, **334**(6060) 1269-1272.

Chan DKS, Cheng GHL (2004). A comparison of offline and online friendship qualities at different stages of relationship development. *Journal of Social and Personal Relationships* **21**(3) 305-320.

Chen PH, White HR, Pandina RJ (2001) Predictors of smoking cessation from adolescence into young adulthood. *Addictive Behaviors* **26**(4) 517–529.

Chen X, Yang X (2014) Does food environment influence food choices? A geographical analysis through "tweets". *Applied Geography* **51** 82-89.

Chen, Y. and Zahedi, F. (2016) Individuals' Internet security perceptions and behaviors: Polycontextual contrasts between the United States and china. *MIS Quarterly* **40**(1) 205-222.

Choi J, Hui SK, Bell DR (2010) Spatiotemporal analysis of imitation behavior across new buyers at an online grocery retailer. *Journal of Marketing Research* **47**(1) 75-89.

Choudhury DM (2011) Tie formation on twitter: Homophily and structure of egocentric networks. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom)* 465-470

Christakis NA, Fowler JH (2007) The spread of obesity in a large social network over 32 years. *New England Journal of Medicine* **357**(4) 370–379

Christakis NA, Fowler JH (2008) The collective dynamics of smoking in a large social network. *New England Journal of Medicine*, **358**(21) 2249-2258.

Christakis NA, Fowler JH (2009) Connected: The surprising power of our social networks and how they shape our live*s*. Little, Brown.

Clark AE, Lohéac Y (2007) "it wasn't me, it was them!" social influence in risky behavior by adolescents. *Journal of Health Economics* **26**(4) 763–784.

Clauset A, Newman ME, Moore C (2004). Finding community structure in very large networks. Physical review E **70**(6) 066111

Cockerham, WC (2005) Health Lifestyle Theory and the Convergence of Agency and Structure. *Journal of health and social behavior* **46**(1) 51-67.

Cohen S, Underwood LG, Gottlieb BH, editors (2000). *Social support measurement and* *intervention: A guide for health and social scientists*. Oxford University Press.

Coleman JC (1980) Friendship and the peer group in adolescence. *Handbook of Adolescent Psychology* 408-431.

Coleman JS (1988) Social capital in the creation of human capital. *American Journal of Sociology* **94** 95-120.

Corliss R (2012). Photos on Facebook generate 53% more likes than the average post [NEW DATA]. *Hubspot. Available at: http://blog. hubspot. com/blog/tabid/6307/bid/33800/Photos-on-FacebookGenerate-53-More-Likes-Than-the-Average-Post-NEW-DATA. aspx*

Craik FI, Lockhart RS (1972) Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior* **11**(6) 671-684.

Cramer H, Rost M, Holmquist LE (2011) Performing a check-in: emerging practices, norms and'conflicts' in location-sharing using foursquare. In *Proceedings of the 13th international conference on human computer interaction with mobile devices and services* 57-66

Cruwys T, Bevelander KE, Hermans RC (2015) Social modeling of eating: A review of when and why social influence affects food intake and choice. *Appetite* **86** 3-18.

Culotta A (2010) Towards detecting influenza epidemics by analyzing Twitter messages. In *Proceedings of the first workshop on social media analytics* 115-122.

Currie J, DellaVigna S, Moretti E, Pathania V (2010) The effect of fast food restaurants on obesity and weight gain. *American Economic Journal: Economic Policy* **2** 32-63.

Daft RL, Lengel RH (1986) Organizational information requirements, media richness and structural design. *Management science* **32**(5) 554-571.

Deci EL (1992) The relation of interest to the motivation of behavior: A self-determination theory perspective.

Deci EL, Ryan RM (2011) Self-determination theory. *Handbook of theories of social psychology* **1** 416-433.

Derlega VJ, Grzelak J (1979) Appropriateness of self-disclosure. In G. J. Chelune (Ed.), Self-disclosure: Origins, patterns, and implications of openness in interpersonal relationships 151–176. San Francisco, CA: Jossey-Bass.

Dewey, J. (1913). Interest and effort in education. Boston: Riverside Press.

DiClemente CC, Marinilli AS, Singh M, Bellino LE (2001) The role of feedback in the process of health behavior change. *American Journal of Health Behavior* **25**(3) 217-227.

Diez Roux AV, Mair C (2010) Neighborhoods and health. *Annals of the New York Academy of*

Sciences **1186**(1) 125–145.

Djoussé L, Driver JA, Gaziano JM (2009). Relation between modifiable lifestyle factors and lifetime risk of heart failure. *Jama* **302**(4) 394-400.

Dourish P (2006) Re-space-ing place: place and space ten years on. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work* 299-308.

Dubois E, Gaffney D (2014) The multiple facets of influence identifying political influentials and opinion leaders on Twitter. *American Behavioral Scientist*, **58**(10) 1260-1277.

Duflo E, Saez E (2002) Participation and investment decisions in a retirement plan: The influence of colleagues' choices. *Journal of public Economics* **85**(1) 121-148.

Dutton WH, Blank G, Groselj D (2013) Cultures of the internet: The internet in Britain. Oxford Internet Survey 2013. Oxford Internet Institute, University of Oxford.

Easley D, Kleinberg J (2010) *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press.

Eftekhar A, Fullwood C, Morris N (2014) Capturing personality from Facebook photos and photo-related activities: How much exposure do you need? *Computers in Human Behavior* **37** 162-170.

Eisenberg ME, Neumark-Sztainer D, Story M, Perry C (2005) The role of social norms and friends? influences on unhealthy weight-control behaviors among adolescent girls. *Social Science & Medicine* **60**(6) 1165–1173.

Ellison NB, Steinfield C, Lampe C (2007) The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. *Journal of Computer‐Mediated Communication* **12**(4) 1143-1168.

Epstein JL (1989) The selection of friends: Changes across the grades and in different school environments. *Peer Relationships in Child Development* 158–187.Esposito G, van Bavel R, Baranowski T, Duch-Brown N (2016). Applying the model of goal-directed behavior, including descriptive norms, to physical activity intentions: a contribution to improving the theory of planned behavior. *Psychological reports* **119(**1) 5-26.

Erikson EH (1994) *Identity: Youth and crisis.* No. 7, WW Norton & Company.

Evans WD (2006). How social marketing works in health care. *British Medical Journal* **332**(7551) 1207-1210.

Faraj S, Kudaravalli S, Wasko M (2015) Leading collaboration in online communities. *MIS Quarterly* **39**(2) 393-412.

Farnham SD, Churchill EF (2011) Faceted identity, faceted lives: social and technical issues with being yourself online. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work* 359-368

Feick L, Higie RA (1992) The effects of preference heterogeneity and source characteristics on ad processing and judgements about endorsers. *Journal of Advertising* **21**(2) 9-24.

Fielding JE (1985) Smoking: Health effects and control. *New England J. of Medicine* **313**(8) 491–498.

Finkelstein EA, Trogdon JG, Cohen JW, Dietz W (2009) Annual medical spending attributable to obesity: Payer-and service-specific estimates. *Health Affairs* **28**(5) w822–w831.

Fiori KL, Antonucci TC, Cortina KS (2006) Social network typologies and mental health among older adults. *The Journals of Gerontology Series B: Psychological Sci. and Social Sciences* **61**(1) 25-32.

Fischer CS (1978) Urban-to-rural diffusion of opinions in contemporary America. *American Journal of Sociology* **84**(1) 151-159.

Fortunato S, Barthelemy M (2007) Resolution limit in community detection. *Proceedings of the National Academy of Sciences* **104**(1) 36-41.

Fortunato S (2010) Community detection in graphs. *Physics reports* **486**(3) 75-174.

Foster S, Giles-Corti B (2008) The built environment, neighborhood crime and constrained physical activity: An exploration of inconsistent findings. *Preventive Medicine* **47**(3) 241–251.

Friedkin NE (2001) Norm formation in social influence networks. *Social Networks* **23**(3) 167-189.

Gallagher KM, Updegraff JA (2011). Health message framing effects on attitudes, intentions, and behavior: a meta-analytic review. *Annals of behavioral medicine* **43**(1) 101-116.

Gallant MP (2003) The influence of social support on chronic illness self-management: A review and directions for research. *Health Education & Behavior* **30**(2) 170-195.

Geise S (2011) Vision that matters. Die Wirkungs- und Funktionslogik Visueller Politischer Kommunikation am Beispiel des Wahlplakats. Wiesbaden, Germany: VS.

Geise S. Baden C (2015) Putting the image back into the frame: Modeling the linkage between visual communication and frame-processing theory. *Communication Theory* **25**(1) 46-69.

Ghosh D, Guha R (2013) What are we ‘tweeting’ about obesity? Mapping tweets with topic modeling and Geographic Information System. *Cartography and geographic information science* **40**(2) 90-102.

Goffman E (1959). The presentation of self in everyday life. New York: Doubleday Anchor Books.

Goh JM, Gao GG, Agarwal R (2016) The creation of social value: Can an online health community reduce rural–urban health disparities? *MIS Quarterly* **40**(1) 247-263.

Goldsmith DJ (2004) *Communicating social support*. New York: Cambridge University Press.

Gong W, Lim EP, Zhu F (2015) Characterizing Silent Users in Social Media Communities. In *Proceedings of the Ninth International AAAI Conference on Web and Social Media* 140-149.

Gosling SD, Augustine AA, Vazire S, Holtzman N, Gaddis S (2011). Manifestations of personality in online social networks: Self-reported Facebook-related behaviors and observable profile information. *Cyberpsychology, Behavior, and Social Networking* **14**(9) 483-488.

Granovetter MS (1973) The strength of weak ties. *American journal of sociology* **78**(6) 1360-1380.

Gu B, Konana P, Raghunathan R, Chen HM (2014) Research Note—The Allure of Homophily in Social Media: Evidence from Investor Responses on Virtual Communities. *Information Systems Research* **25**(3) 604-617.

Hampton KN, Lee CJ, Her EJ (2011) How new media affords network diversity: Direct and mediated access to social capital through participation in local social settings. *New Media & Society* **13**(7) 1031-1049.

Hampton KN (2016) Persistent and pervasive community new communication technologies and the future of community. *American Behavioral Scientist* **60**(1) 101-124.

Hanson MA, Gluckman PD, Ma RCW, Matzen P, Biesma RG (2012) Early life opportunities for prevention of diabetes in low and middle income countries. *BMC Public Health* **12**(1) 1-9.

Haskell WL, Lee IM, Pate RR, Powell KE, Blair SN, Franklin BA, Macera CA, Heath GW, Thompson PD, Bauman A (2007) Physical activity and public health: Updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Circulation* **116**(9) 1081-1093.

Haskell WL, Blair SN, Hill JO (2009) Physical activity: Health outcomes and importance for public health policy. *Preventive Medicine* **49**(4) 280-282.

Hastings G, Haywood A (1991) Social marketing and communication in health promotion. *Health Promotion International* **6**(2) 135-145.

Hogan B (2010) The presentation of self in the age of social media: Distinguishing performances and exhibitions online. *Bulletin of Science, Technology & Society* **30**(6) 377-386.

Holmberg C, Chaplin JE, Hillman T, Berg C (2016). Adolescents' presentation of food in social media: An explorative study. *Appetite* **99** 121-129.

Hong, W, Chan, FK, Thong, JY, Chasalow, LC, Dhillon, G. (2013) A framework and guidelines for context-specific theorizing in information systems research. *Information Systems Research* ***25***(1) 111-136.

Huhman M, Heitzler C, Wong F (2004) The VERB™ campaign logic model: a tool for planning and evaluation. *Preventing chronic disease* **1**(3) 1-6.

Hung HC, Joshipura KJ, Jiang R, Hu FB, Hunter D, Smith-Warner SA, Colditz GA, Rosner B, Spiegelman D, Willett WC (2004) Fruit and vegetable intake and risk of major chronic disease. *Journal of the National Cancer Institute* **96**(21) 1577–1584.

Hutchinson DM, Rapee RM (2007) Do friends share similar body image and eating problems? The role of social networks and peer influences in early adolescence. *Behaviour Research and Therapy* **45**(7) 1557-1577.

Ikeda K, Hattori G, Ono C, Asoh H, Higashino T (2013) Twitter user profiling based on text and community mining for market analysis. *Knowledge-Based Systems* **51** 35-47.

Ingledew DK, Hardy L, Cooper CL, Jemal H (1996) Health behaviours reported as coping strategies: A factor analytical study. *British Journal of Health Psychology* **1**(3) 263–281.

Isganaitis E, Lustig RH (2005) Fast food, central nervous system insulin resistance, and obesity. *Arteriosclerosis, Thrombosis, and Vascular Biology* **25**(12) 2451-2462.

Iyengar R, Van den Bulte C, Valente TW (2011) Opinion leadership and social contagion in new product diffusion. *Marketing Science* **30**(2) 195-212.

Iyer A, Oldmeadow J (2006) Picture this: Emotional and political responses to photographs of the Kenneth Bigley kidnapping. *European Journal of Social Psychology* **36**(5) 635-647.

Jacomy M, Venturini T, Heymann S, Bastian M (2014) ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PloS one* **9**(6) e98679.

Jalali NY, Papatla P (2016) The palette that stands out: Color compositions of online curated visual UGC that attracts higher consumer interaction. *Quantitative Marketing and Economics* **14**(4) 353-384.

Jha P, Ramasundarahettige C, Landsman V, Rostron B, Thun M, Anderson RN, McAfee T, Peto R (2013) 21st-century hazards of smoking and benefits of cessation in the United States. *New England Journal of Medicine* **368**(4) 341-35.

Johnson M (1985) Relating metrics, lines and variables defined on graphs to problems in medicinal chemistry. In *Graph theory with applications to algorithms and computer science*

457-470 John Wiley & Sons, Inc.

Jourard, S. M. (1971). Self-disclosure: An experimental analysis of the transparent self. New York, NY: Wiley-Interscience.

Kahneman D (1973). Attention and Effort. Prentice Hall, Eaglewood Cliffs, NJ

Kane GC, Alavi M, Labianca GJ, Borgatti S (2014) What's different about social media networks? a framework and research agenda. *MIS Quarterly* **38**(1) 275-304.

Kaplan CP, Nápoles-Springer A, Stewart SL, Pérez-Stable EJ (2001) Smoking acquisition among adolescents and young Latinas: The role of socioenvironmental and personal factors. *Addictive Behaviors* **26**(4) 531–55.

Kaplan AM, Haenlein M (2010) Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons* **53**(1) 59-68.

Kasl SV, Cobb S (1966) Health behavior, illness behavior and sick role behavior: I. Health and illness behavior. *Archives of Environmental Health: An International Journal* **12**(2) 246-266.

Kenix LJ (2013) A converging image? Commercialism and the visual identity of alternative and mainstream news websites. *Journalism Studies* **14**(6) 835-856.

Kim M, Lennon S (2008) The effects of visual and verbal information on attitudes and purchase intentions in internet shopping *Psychology & Marketing* **25**(2) 146-178.

Klemettinen M, Mannila H, Ronkainen P, Toivonen H, Verkamo A I (1994) Finding interesting rules from large sets of discovered association rules. In *Proceedings of the third international conference on Information and knowledge management* 401-407. ACM.

Kossinets G, Watts DJ (2009) Origins of homophily in an evolving social network 1. *American journal of sociology* **115**(2) 405-450.

Kress G, Van Leeuwen TV (2001) Multimodal discourse: The modes and media of contemporary communication.

Kress G van Leeuwen T (2010) *Reading Images. e Grammar of Visual Design*. London, New York: Routledge

Kreuter MW, Strecher VJ, Glassman B (1999) One size does not fit all: The case for tailoring print materials. Annals of behavioral medicine. **21**(4) 276-283.

Kumpula JM, Saramäki J, Kaski K, Kertész J (2007) Limited resolution in complex network community detection with Potts model approach. *The European Physical Journal B* **56**(1) 41-45.

Kwak H, Lee C, Park H, Moon S (2010) What is Twitter, a social network or a news media? In

Proceedings of the 19th International Conference on World Wide Web 591-600 ACM.

Kwon KH, Stefanone MA, Barnett GA (2014) Social network influence on online behavioral choices exploring group formation on social network sites. *American Behavioral Scientist* **58**(10) 1345-136.

LaBerge D, Samuels SJ (1974) Toward a theory of automatic information processing in reading. *Cognitive psychology* **6**(2) 293-323.

Lampe C, Ellison N, Steinfield C (2006) A Face (book) in the crowd: Social searching vs. social browsing. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work* 167-170 ACM.

Lane VR (2000) The impact of ad repetition and ad content on consumer perceptions of incongruent extensions *Journal of Marketing* **64**(2) 80-91.

Langford CP, Bowsher J, Maloney JP, Lillis PP (1997) Social support: A conceptual analysis. *Journal of Advanced Nursing* **25**(1) 95-10.

Latané B (1981) The psychology of social impact. *American Psychologist* **36**(4) 343-356.

Lazarsfeld PF, Merton RK (1954) Friendship as a social process: A substantive and methodological analysis. *Freedom and control in modern society* **18**(1) 18-66.

Lee IM, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT, Lancet Physical Activity Series Working Group (2012a) Effect of physical inactivity on major non-communicable diseases worldwide: An analysis of burden of disease and life expectancy. *Lancet* **380**(9838) 219-229.

Lee JH, Han G, Fulp WJ, Giuliano AR (2012b) Analysis of overdispersed count data: Application to the human papillomavirus infection in men (HIM) study. *Epidemiol Infect* **140**(06), 1087-1094.

Lee KT, Noh MJ, Koo DM (2013) Lonely people are no longer lonely on social networking sites: The mediating role of self-disclosure and social support. *Cyberpsychology, Behavior, and Social Networking* **16**(6) 413-418.

Lee E, Lee JA, Moon JH, Sung Y (2015). Pictures speak louder than words: Motivations for using Instagram. *Cyberpsychology, Behavior, and Social Networking*, **18**(9) 552-556.

Leung RK, Toumbourou JW, Hemphill SA (2014) The effect of peer influence and selection processes on adolescent alcohol use: A systematic review of longitudinal studies. *Health Psychology Review* **8**(4) 426–457.

Lewis DM, Conroy-Beam D, Al-Shawaf L, Raja A, DeKay T, Buss DM (2011) Friends with benefits: The evolved psychology of same-and opposite-sex friendship. *Evolutionary Psychology* **9**(4) 543-563.

Li Y, Wu C, Luo P, Zhang W (2013) Exploring the characteristics of innovation adoption in social networks: structure, homophily, and strategy. *Entropy* **15**(7) 2662-2678.

Li R, Wang C, Chang KC (2014) User profiling in an ego network: co-profiling attributes and relationships. In *Proceedings of the 23rd international conference on World wide web* 819-830.

Linden-Andersen S, Markiewicz D, Doyle AB (2008) Perceived similarity among adolescent friends: The role of reciprocity, friendship quality, and gender. *The Journal of Early Adolescence* **29**(5) 617-637.

Lindqvist J, Cranshaw J, Wiese J, Hong J, Zimmerman J (2011). I'm the mayor of my house: examining why people use foursquare-a social-driven location sharing application. In *Proceedings of the SIGCHI conference on human factors in computing systems* 2409-2418

Link BG, Phelan J (1995) Social conditions as fundamental causes of disease. *Journal of Health and Social Behavior* (extra issue) 80-94.

Loef M, Walach H (2012) The combined effects of healthy lifestyle behaviors on all cause mortality: A systematic review and meta-analysis. *Preventive Medicine* **55**(3) 163–17.

Long JS (1997) *Regression Models for Categorical and Limited Dependent Variables.* Thousand Oaks, CA.

Lu W, Hampton KN (2016) Beyond the power of networks: Differentiating network structure from social media affordances for perceived social support. *New Media & Society* 1-19

Lundborg P (2006) Having the wrong friends? Peer effects in adolescent substance use. *Journal of Health Economics* **25**(2) 214–233.

Lutz KA, Lutz RJ (1977) Effects of interactive imagery on learning: Application to advertising. *Journal of Applied Psychology* **62**(4) 493-498.

Lynch JW, Kaplan GA, Salonen JT (1997) Why do poor people behave poorly? Variation in adult health behaviours and psychosocial characteristics by stages of the socioeconomic lifecourse. *Social Science & Medicine* **44**(6) 809-819.

MacGeorge EL, Feng B and Burleson BR (2011) Supportive communication. In Knapp M and Daly J Handbook of interpersonal communication. 4th ed. Thousand Oaks, CA 317-354.

Manski CF (1993) Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies* **60**(3) 531-542.

Marcus SR (2015) Picturing' ourselves into being: assessing identity, sociality and visuality on Instagram. In *international communication association conference*. Puerto Rico: San Juan.

Marriott TC and Buchanan T (2014) The true self online: Personality correlates of preference for self-expression online, and observer ratings of personality online and offline. *Computers in Human Behavior* **32** 171-177.

Marsden PV, Friedkin NE (1993) Network studies of social influence. *Sociological Methods & Research* **22**(1) 127-151.

Marwick, AE (2015) Instafame: Luxury selfies in the attention economy. *Public culture*, **27**(75), 137-160.

Mayer RE (2005). Introduction to multimedia learning. In R.E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 1-16). New York: Cambridge University Press.

Mejova Y, Haddadi H, Noulas A, Weber I (2015) # foodporn: Obesity patterns in culinary interactions. *In Proceedings of the 5th International Conference on Digital Health 2015* 51-58

Messaris P, Abraham L (2001) The role of images in framing news stories. *Framing public life: Perspectives on media and our understanding of the social world* 215-226.

McCambridge J, Witton J, Elbourne DR (2014) Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. *J Clinical Epidemiology* **67**(3) 267-277.

McDaniel BT, Coyne SM, Holmes EK (2012) New mothers and media use: Associations between blogging, social networking, and maternal well-being. *Maternal and Child Health Journal* **16**(7) 1509-1517.

McEvoy CT, Temple N, Woodside JV (2012) Vegetarian diets, low-meat diets and health: A review. *Public Health Nutrition* **15**(12) 2287-2294.

McNeill LH, Kreuter MW, Subramanian SV (2006) Social environment and physical activity: A review of concepts and evidence. *Social Science & Medicine* **63**(4) 1011–1022.

McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: Homophily in social networks. *Annual Review of Sociology* 415-444.

Mitrou L, Kandias M, Stavrou V, Gritzalis D (2014) Social media profiling: A Panopticon or Omniopticon tool? In *Proc. of the 6th Conference of the Surveillance Studies Network*.

Milio N (1981) Promoting health through public policy. F. A. Davies, Philadelphia.

Mislove A, Viswanath B, Gummadi KP, Druschel P (2010) You are who you know: inferring user profiles in online social networks. In *Proceedings of the third ACM international conference on Web search and data mining* 251-260.

Moore LV, Diez Roux AV (2006) Associations of neighborhood characteristics with the location and type of food stores. *American Journal of Public Health* **96**(2) 325-331.

Morie JF, Verhulsdonck G (2008). Body/persona/action!: Emerging non-anthropomorphic communication and interaction in virtual worlds. In Proceedings of the 2008 International Conference on Advances in Computer Entertainment Technology 365-372

Munshi K (2004) Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution. *Journal of development Economics* **73**(1) 185-213.

Naidoo J, Wills J (2009) Foundations for health promotion. *Elsevier Health Sciences*.

Nahon K, Hemsley J (2014) Homophily in the Guise of Cross-Linking Political Blogs and Content. *American Behavioral Scientist* **58**(10) 1294-1313.

Napolitano MA, Hayes S, Bennett GG, Ives AK, Foster GD (2013) Using Facebook and text messaging to deliver a weight loss program to college students. *Obesity* **21**(1) 25-31.

Newman ME (2004) Analysis of weighted networks. *Physical review E* **70**(5) 056131.

Newman ME, Girvan M (2004) Finding and evaluating community structure in networks. *Physical review E* **69**(2) 026113.

Nguyen M, Bin YS, Campbell A (2012) Comparing online and offline self-disclosure: A systematic review. *Cyberpsychology, Behavior, and Social Networking* **15**(2) 103-111.

Nielsen J (2006) Participation inequality: lurkers vs. contributors in internet communities. *Jakob Nielsen's Alertbox*.

Nightingale V (2007) The cameraphone and online image sharing. *Continuum* **21**(2) 289-301.

Nitzan I, Libai B (2011) Social effects on customer retention. *Journal of Marketing* **75**(6) 24-38.

Nov O, Naaman M, Ye C (2010) Analysis of participation in an online photo-sharing community: A multidimensional perspective. *Journal of the Association for Information Science and Technology* **61**(3) 555-566.

O'Halloran K, Smith BA (Eds.) (2012). Multimodal studies: Exploring issues and domains. New York, NY: Routledge.

Paivio A (1991) Images in mind. New York: Harvester Wheatsheaf.

Park N, Jin B, Jin SAA (2011) Effects of self-disclosure on relational intimacy in Facebook. *Computers in Human Behavior* **27**(5) 1974-1983.

Patil S, Norcie G, Kapadia A, Lee AJ (2012). Reasons, rewards, regrets: privacy considerations in location sharing as an interactive practice. In *Proceedings of the Eighth Symposium on Usable Privacy and Security* **5.**

Pechmann C, Pan L, Delucchi K, Lakon CM, Prochaska JJ (2015) Development of a Twitter-based intervention for smoking cessation that encourages high-quality social media interactions via automessages. *Journal of medical Internet research* **17**(2) e50.

Pegg KJ, O'Donnell AW, Lala G, Barber BL (2018). The role of online social identity in the relationship between alcohol-related content on social networking sites and adolescent alcohol use. *Cyberpsychology, Behavior, and Social Networking* **21**(1) 50-55

Pelechrinis K, Krishnamurthy P (2012) Location affiliation networks: bonding social and spatial information. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* 531-547.

Pempek TA, Yermolayeva YA, Calvert SL (2009). College students' social networking experiences on Facebook. Journal of applied developmental psychology **30**(3) 227-238.

Pennacchiotti M, Popescu AM (2011) Democrats, republicans and starbucks afficionados: user classification in twitter. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* 430-438.

Pettigrew S, Charters S (2006) Consumers' expectations of food and alcohol pairing. *British food journal* **108**(3) 169-180.

Phelan JC, Link BG, Diez-Roux A, Kawachi I, Levin B (2004) "Fundamental causes" of social inequalities in mortality: A test of the theory. *Journal of Health and Social Behavior* **45**(3) 265–285.

Pickett KE, Pearl M (2001) Multilevel analyses of neighbourhood socioeconomic context and health outcomes: A critical review. *Journal of Epidemiology and Community Health* **55**(2) 111–122.

Pope, CA, Burnett RT, Krewski D, Jerrett M, Shi Y, Calle EE, Thun MJ (2009) Cardiovascular mortality and exposure to airborne fine particulate matter and cigarette smoke shape of the exposure-response relationship. *Circulation* **120**(11) 941-948.

Posey C, Lowry PB, Roberts TL, Ellis TS (2010). Proposing the online community self-disclosure model: the case of working professionals in France and the UK who use online communities. *European Journal of Information Systems* **19**(2) 181-195.

Powell KE, Paluch AE, Blair SN (2011) Physical activity for health: What kind? How much? How intense? On top of what? *Public Health* **32**(1) 349-365.

Powell TE, Boomgaarden HG, De Swert K, de Vreese CH (2015) A clearer picture: The contribution of visuals and text to framing effects. *Journal of Communication* **65**(6) 997-1017.

Pugh MJV, Hart D (1999) Identity development and peer group participation. *New directions for child and adolescent development* **1999**(84) 55-70.

R Development Core Team (2016) *R*: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http: //www.R-project.org.

Rahman A (2016) The influence of social identity when digitally sharing location (Doctoral dissertation, University of Nottingham).

Ramo DE, Thrul J, Chavez K, Delucchi KL, Prochaska JJ (2015) Feasibility and quit rates of the Tobacco Status Project: A Facebook smoking cessation intervention for young adults. *Journal of medical Internet research* **17**(12) e291.

Rehm J, Baliunas D, Borges GL, Graham K, Irving H, Kehoe T, Parry CD, Patra J, Popova S, Poznyak V, Roerecke M (2010) The relation between different dimensions of alcohol consumption and burden of disease: An overview. *Addiction* **105**(5) 817-843.

Reicher SD, Spears R, Postmes T (1995) A social identity model of deindividuation phenomena. *European review of social psychology* **6**(1) 161-198.

Risselada H, Verhoef PC, Bijmolt TH (2014) Dynamic effects of social influence and direct marketing on the adoption of high-technology products. *Journal of Marketing* **78**(2) 52-68.

Rogers EM (1995) Diffusion of innovations. *New York*, 12.

Room R, Babor T, Rehm J (2005) Alcohol and public health. *The lancet* **365**(9458) 519–53.

Rosamond W, Flegal K, Furie K, Go A, Greenlund K, Haase N, Hailpern SM, Ho M, Howard V, Kissela B, Kittner S, Lloyd-Jones D, McDermott M, Meigs J, Moy C, Nichol G, O'Donnell C, Roger V, Sorlie P, Steinberger J, Thom T, Wilson M, Hong Y (2008) American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics—2008 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation* **117** e25–e146.

Rosenheck R (2008) Fast food consumption and increased caloric intake: a systematic review of a trajectory towards weight gain and obesity risk. *Obesity Reviews* **9**(6) 535-547.

Rosenquist JN, Murabito J, Fowler JH, and Christakis NA (2010) The spread of alcohol consumption behavior in a large social network. *Annals of Internal Medicine* **152**(7) 426-433.

Rothman AJ, Bartels RD, Wlaschin J, Salovey P (2006). The strategic use of gain-and loss-framed messages to promote healthy behavior: How theory can inform practice. *Journal of communication* **56**(1) S202-S220.

Rosenquist JN, Murabito J, Fowler JH, and Christakis NA (2010) The spread of alcohol consumption behavior in a large social network. *Annals of Internal Medicine* **152**(7) 426-433.

Rui JR, Stefanone MA (2013) Strategic image management online: Self-presentation, self-esteem and social network perspectives. *Information, Communication & Society* **16**(8) 1286-

1305.

Rutter M (1987) Psychosocial resilience and protective mechanisms. *American journal of orthopsychiatry* **57**(3) 316-331.

Ryan RM, Deci EL (2000) Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist* **55**(1) 68-78.

Rydell SA, Harnack LJ, Oakes JM, Story M, Jeffery RW, French SA (2008) Why eat at fast-food restaurants: reported reasons among frequent consumers. *Journal of the American Dietetic Association* **108**(12) 2066-2070.

Sagiv L, Sverdlik N, Schwarz N (2011) To compete or to cooperate? Values' impact on perception and action in social dilemma games. *European Journal of Social Psychology*, **41**(1) 64-77.

Salathé M, Khandelwal S (2011) Assessing vaccination sentiments with online social media: Implications for infectious disease dynamics and control. *Computational. Biology* **7**(10) e1002199.

Scarborough P, Bhatnagar P, Wickramasinghe KK, Allender S, Foster C, Rayner M (2011) The economic burden of ill health due to diet, physical inactivity, smoking, alcohol and obesity in the UK: An update to 2006–07 NHS costs. *Journal of Public Health* **33**(4) 527–535.

Schwartz SH (2015) Basic individual values: Sources and consequences. *Handbook of value: Perspectives from economics, neuroscience, philosophy, psychology and sociology* 63-84.

Sears DO, Freedman JL (1967) Selective exposure to information: A critical review. *Public Opinion Quarterly* **31**(2) 194-213.

Sesso HD, Cook NR, Buring JE, Manson JE, Gaziano JM (2008) Alcohol consumption and the risk of hypertension in women and men. *Hypertension* **51**(4) 1080-1087.

Severin W (1967) The effectiveness of relevant pictures in multiple-channel communications. *AV Communication Review* **15**(4) 386-401.

Sheldon P (2009) " I'll poke you. You'll poke me!" Self-disclosure, social attraction, predictability and trust as important predictors of Facebook relationships. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* **3**(2) 5-15.

Sheldon P, Bryant K (2016) Instagram: Motives for its use and relationship to narcissism and contextual age. *Computers in Human Behavior* **58** 89-97.

Short J, Williams E, Christie B (1976) The social psychology of telecommunications.

Shi Z, Rui H, Whinston AB (2014). Content sharing in a social broadcasting environment: Evidence from Twitter. *MIS Quarterly* **38**(1) 123-142.

Shin D, He S, Lee GM, Whinston AB, Cetintas S, Lee KC (2017) Enhancing Social Media Analysis with Visual Analytics: A Deep Learning Approach.

Smith, KP, Christakis NA (2008) Social networks and health. *Annual Review of Sociology* **34** 405–429.

Special WP, Li-Barber KT (2012) Self-disclosure and student satisfaction with Facebook. *Computers in Human Behavior* **28**(2) 624-630.

Strekalova YA, Krieger JL (2017) A picture really is worth a thousand words: Public engagement with the National Cancer Institute on social media. *Journal of Cancer Education* **32**(1) 155-157.

Subrahmanyam K, Reich SM, Waechter N, Espinoza G (2008). Online and offline social networks: Use of social networking sites by emerging adults. *Journal of applied developmental psychology* **29**(6) 420-433.

Sundar SS (2008) The MAIN model: A heuristic approach to understanding technology effects on credibility. *Digital media, youth, and credibility*, *73100*.

Tang KP, Lin J, Hong JI, Siewiorek DP, Sadeh N (2010) Rethinking location sharing: exploring the implications of social-driven vs. purpose-driven location sharing. In *Proceedings of the 12th ACM international conference on Ubiquitous computing* 85-94

Tang X, Miao Q, Quan Y, Tang J, Deng K (2015) Predicting individual retweet behavior by user similarity: A multi-task learning approach. *Knowledge-Based Systems* **89** 681-688.

Tamir DI, Mitchell JP (2012) Disclosing information about the self is intrinsically rewarding. In *Proceedings of the National Academy of Sciences* **109**(21) 8038-8043.

Thaler RH, Sunstein CR (2008) *Nudge: Improving Decisions about Health, Wealth, and Happiness*, Penguin Books.

Tidwell LC, Walther JB (2002) Computer-mediated communication effects on disclosure, impressions, and interpersonal evaluations: Getting to know one another a bit at a time. *Human communication research* **28**(3) 317-348.

Trepte S, Reinecke L (2013) The reciprocal effects of social network site use and the disposition for self-disclosure: A longitudinal study. *Computers in human behavior* **29**(3) 1102-1112.

Trogdon JG, Nonnemaker J, Pais J (2008) Peer effects in adolescent overweight. *Journal of Health Economics* **27**(5) 1388–1399.

Trusov M, Ma L, Jamal Z (2016) Crumbs of the cookie: User profiling in customer-base analysis and behavioral targeting. *Marketing Science* **35**(3) 405-426.

Uchino BN (2006) Social support and health: A review of physiological processes potentially underlying links to disease outcomes. *Journal of Behavioral Medicine* **29**(4) 377–387.

Ulloa LC, Mora MC, Pros RC, Tarrida AC (2015). News photography for Facebook: Effect of images on the visual behavior of readers in three simulated newspaper formats. *Information Research* **20**(1) 315–333.

Utz S (2015) The function of self-disclosure on social network sites: Not only intimate, but also positive and entertaining self-disclosures increase the feeling of connection. *Computers in Human Behaviors.* **45** 1–1.

Uzzi B (1997) Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative science quarterly* **42**(1) 35-67.

Valle CG, Tate DF, Mayer DK, Allicock M, Cai J (2013) A randomized trial of a Facebook-based physical activity intervention for young adult cancer survivors. *Journal of Cancer Survivorship* **7**(3) 355-368.

Van Alstyne M, Brynjolfsson E (2005) Global village or cyber-balkans? Modeling and measuring the integration of electronic communities. *Management Science* **51**(6) 851-868.

Van den Bulte C, Lilien GL (2001) Medical innovation revisited: Social contagion versus marketing effort. *American Journal of Sociology* **106**(5) 1409-1435.

Van den Bulte C, Wuyts SHK (2007) Social networks in marketing. Cambridge, MA: Marketing Science Institute.

Vandervoort D (1999) Quality of social support in mental and physical health. *Current Psychology* **18**(2) 205-221.

Vangelisti, A L (2009) Challenges in conceptualizing social support. *Journal of Social and Personal Relationships* **26**(1) 39-51.

Vaux, A (1988) *Social support: Theory, research, and intervention*. Praeger publishers.

Vuong QH (1989) Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society* **57**(2) 307-333.

Wang H, Chin A, Wang H (2011) Interplay between social selection and social influence on physical proximity in friendship formation. In *SRS 2011 Workshop*.

Wang YC, Burke M, and Kraut R (2016) Modeling self-disclosure in social networking sites. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* 74-85.

Waring ME, Schneider KL, Appelhans BM, Simas TA, Xiao RS, Whited MC, Busch AM, Evans MM, Pagoto SL (2016) Interest in a Twitter-delivered weight loss program among women of childbearing age. *Translational behavioral medicine* **6**(2) 277-284.

Waterman AS, Schwartz SJ, Goldbacher E, Green H, Miller C, Philip S (2003) Predicting the subjective experience of intrinsic motivation: The roles of self-determination, the balance of challenges and skills, and self-realization values. *Personality and Social Psychology Bulletin* **29**(11) 1447-1458.

Wiley JA, Camacho TC (1980) Life-style and future health: evidence from the Alameda County Study. *Preventive medicine* **9**(1) 1-21.

Widener MJ, Li W (2014) Using geolocated Twitter data to monitor the prevalence of healthy and unhealthy food references across the US. *Applied Geography* **54** 189-97.

Wills T A (1991) Social support and interpersonal relationships. *Review of personality and social psychology **12*** 265-289.

Wit SD, Barker RA, Dickinson AD, Cools R (2011) Habitual versus goal-directed action control in Parkinson disease. *Journal of Cognitive Neuroscience* **23**(5) 1218-1229.

Wong PK, Christie JJ, Wark JD (2007) The effects of smoking on bone health. *Clinical Science* **113**(5) 233-241.

Yang J, Leskovec J (2015) Defining and Evaluating Network Communities Based on Ground-Truth. *Knowledge and Information Systems* **42**(1) 181-213.

Yantis S (2005) How visual salience wins the battle for awareness. *Nature Neuroscience* **8**(8) 975–977.

Yuan YC, Gay G (2006) Homophily of network ties and bonding and bridging social capital in computer-mediated distributed teams. *Journal of Computer-Mediated Communication* **11**(4) 1062-1084.

Yussen SR (1974) Determinants of visual attention and recall in observational learning by preschoolers and second graders. *Developmental Psychology* **10**(1) 93-100.

Zenk SN, Schulz AJ, Israel BA, James SA, Bao S, Wilson ML (2005) Neighborhood racial composition, neighborhood poverty, and the spatial accessibility of supermarkets in metropolitan Detroit. *American Journal of Public Health* **95**(4) 660–667.

Zhang C, Shou L, Chen K, Chen G, Bei Y (2012) Evaluating geo-social influence in location-based social networks. *In Proceedings of the 21st ACM international conference on Information and knowledge management* 1442-1451.

Zhao S, Grasmuck S, Martin J (2008) Identity construction on Facebook: Digital empowerment in anchored relationships. *Computers in human behavior* **24**(5) 1816-1836.

Zhao L, Lu Y, Gupta S (2012) Disclosure intention of location-related information in location-based social network services. *International Journal of Electronic Commerce* **16**(4) 53–9.

Zillmann D, Bryant J (1985) Affect, mood, and emotion as determinants of selective exposure. In D. Zillmann & J. Bryant (Eds.), Selective exposure to communication 157–190 Hillsdale, NJ: Lawrence Erlbaum Associates.

Zillmann D (1988) Mood management through communication choices. *American Behavioral Scientist* **31**(3) 327–341.

Appendix A:

Analysis of Highlighted Keywords of Venue Types

Foursquare boldface-highlighted keywords were collected at venue levels. We collected keywords for a total of 114,125 venues (9,054 fitness center & gym venues; 40,172 bar venues; 64,899 fast food restaurant venues). One unique corpus was created for each type of health-related venue (fitness center & gym, bar, fast food restaurant). High frequency terms in each of the corpuses represent the main characteristics of the venues. To compute the term frequency matrices, we applied text preparation algorithms to each corpus. Figure A1 shows the steps for text processing of keywords. First, we removed all the stop words and punctuations from the texts, thus removing useless words with high levels of frequency. Second, we converted all the words to lowercase formats, making it possible to count keywords with different capitalization such as "Gym" and "gym". Third, we stemmed all the keywords to their roots, making it possible to count keywords with same root. For example, "plays", "played", and "playing" was stemmed to "play" and counted as the same word.

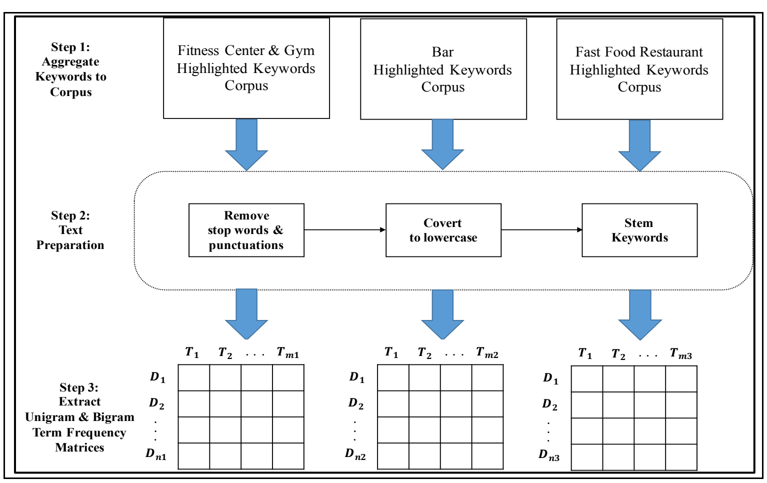


Figure A1. Analysis of Highlighted Keywords

We then created the unigram and bigram term frequency matrices. A term frequency matrix shows the frequency of each term in a corpus. Unigram involves single words whereas bigram involves two words that show up together in the term frequency matrix. Figures A2-A4 shows the 10 most repeated keywords that characterize each venue type based on users' reviews.[28]

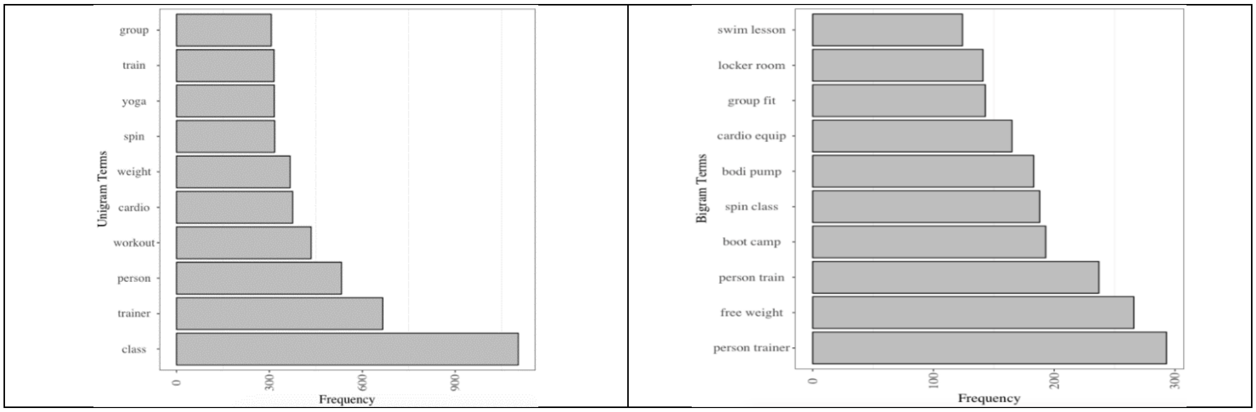


Figure A2. High Frequency Keywords Fitness Center & Gym

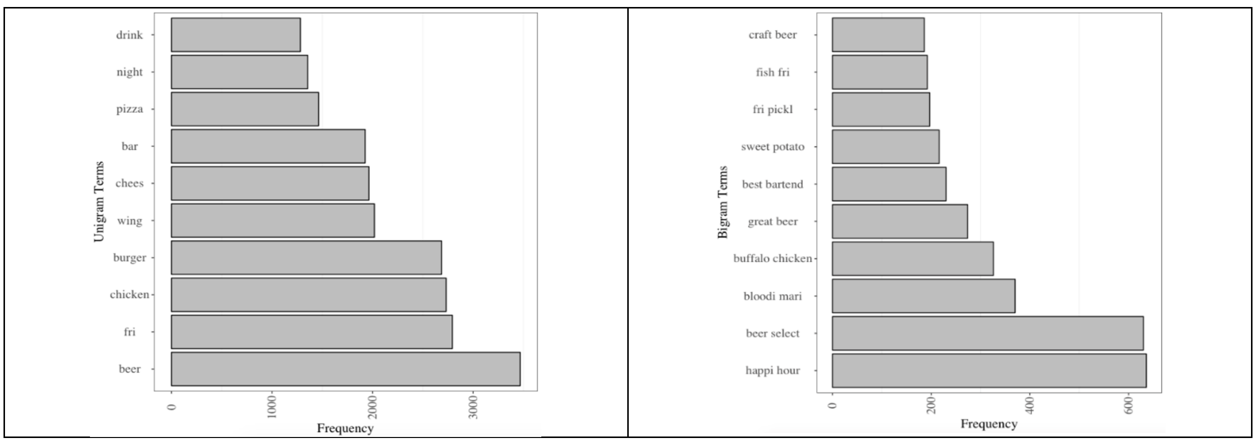


Figure A3. High Frequency Keywords Bar

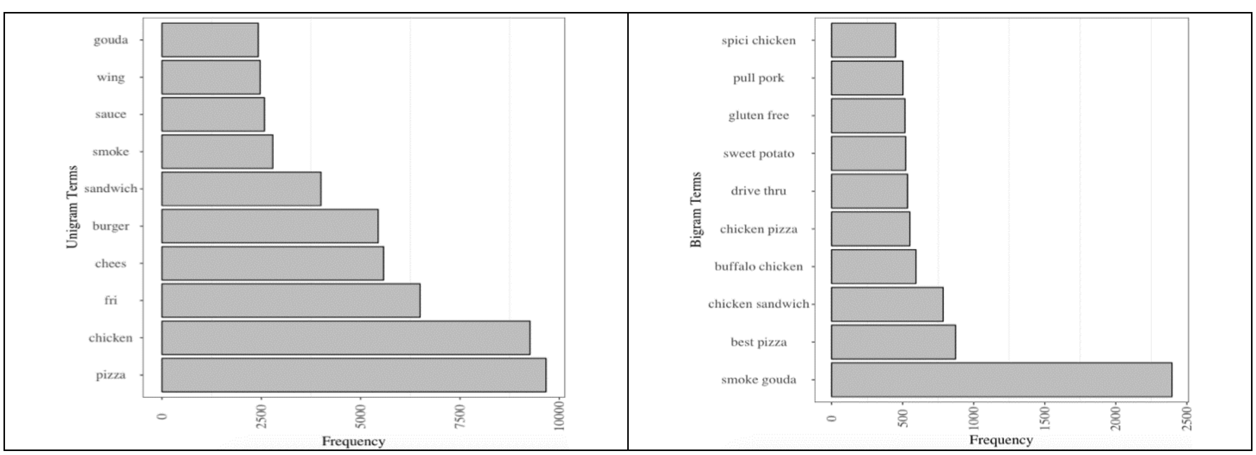


Figure A4. High Frequency Keywords Fast Food Restaurant

[28] We omitted adjectives such as great and best in our unigram term visualization since they only provide qualitative information about venues.

Appendix B:

Computation of Geo Points as Individuals' Center-of-Activity Locations

For each individual's check-ins, we computed the center of gravity associated with the check-ins in venue locations. This was needed for the computation of individuals' distances from their friends. We applied the geographic midpoint for this purpose (Zi-xia and Wei 2010, geomidpoint.com). To compute geo midpoints, we used the venues that were located in the state where the individual resided because geo points are proxies for individuals' centers-of-activity.

The latitude and longitude for each location was converted into Cartesian coordinates. All venue locations had equal weights in the computation. The computed x, y, and z coordinates for each location were then added together and divided by the total number of check-ins. A line can be drawn from the center of the earth out to this new x, y, z coordinate, and the point where the line intersects the surface of the earth is the geographic midpoint. This surface point was converted into the latitude and longitude for the midpoint. The pseudo code for our algorithm is listed below.

```
Venues = findUserCheckinsInTheState(UserId);
Counter = 0;
SumX = 0; SumY = 0; SumZ = 0;
for Venue in Venues{ Counter++;
        Lat = convertDegreeToRadian(Venue.latitude);
        Lng = convertDegreeToRadian(Venue.longitude);
        SumX = SumX + cos(Lat) * cos(Lng);
        SumY = SumY + cos(Lat) * sin(Lng);
        SumZ = SumZ + sin(Lat); }
AvgX = SumX / Counter;
AvgY = SumY / Counter;
AvgZ = SumZ / Counter;
midLng = convertRadianToDegree(arctan2(AvgY,AvgX));
hyp=sqrt(AvgX^2+ AvgY^2);
midLat =  convertRadianToDegree(arctan2(AvgZ,hyp));
```

References for Appendix B

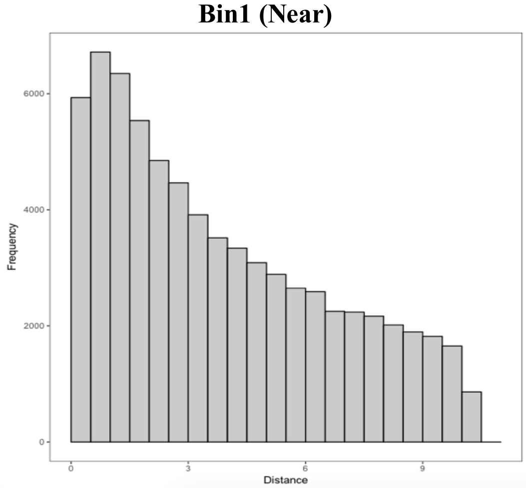
Geomidpoint website. URL: http://www.geomidpoint.com/calculation.html

Zi-xia C, Wei H. (2010) Study and application of Center-of-Gravity on the location selection of distribution center. In *Logistics Systems and Intelligent Management, 2010 International Conference* 2 981-984.

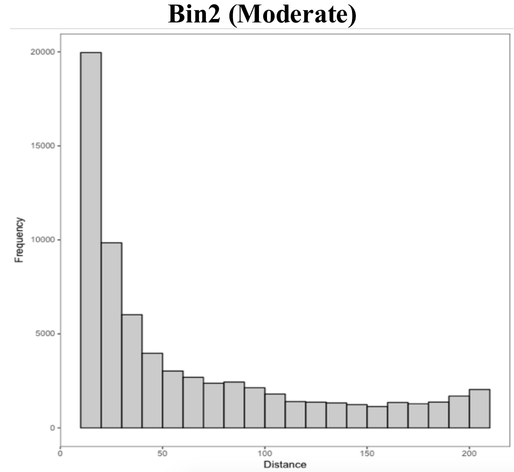
Appendix C:

Frequency of Distances in Bins

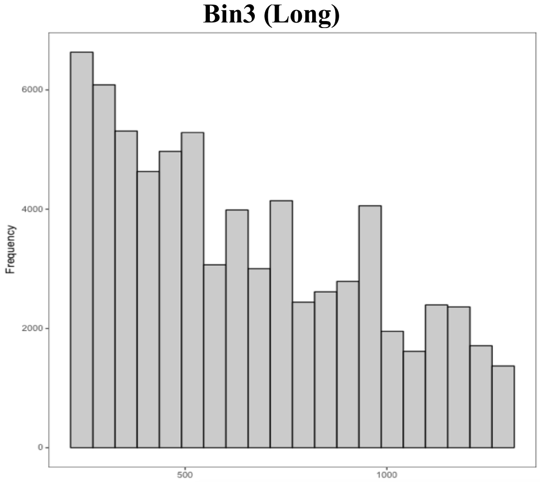
Bin1 (Near)



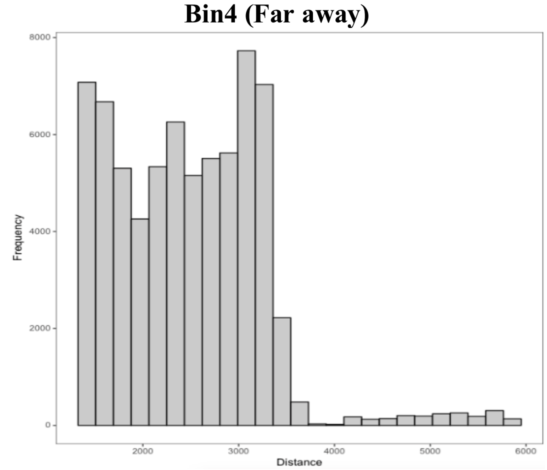
Bin2 (Moderate)



Bin3 (Long)



Bin4 (Far away)



Appendix D:

Spearman Correlations

Table D1- Correlations for Fitness Check-ins, N=22,423

	Variable	1	2	3	4	5	6	7
1	Individuals' healthy behavior: Fitness @ *t+1*							
2	Activity level in social network @ *t+1*	.373						
3	Individuals' healthy behavior: Fitness @ *t*	.548	.222					
4	Socioeconomic status	.018	.008	.029				
5	Social support for healthy behavior: Fitness @ *t*	.299	.115	.472	.002			
6	Friends' healthy behavior: Fitness @ t	.092	.019	.098	-.005	.039		
7	Ratio of strong ties' healthy behaviors: Fitness @ *t*	.034	.018	.035	-.011	.035	.030	
8	Ratio of similar friends' healthy behaviors: Fitness @ *t*	.020	.011	.018	.006	.014	-.045	.089

Table D2- Correlations for Alcohol & Smoking Check-ins, N=28,594

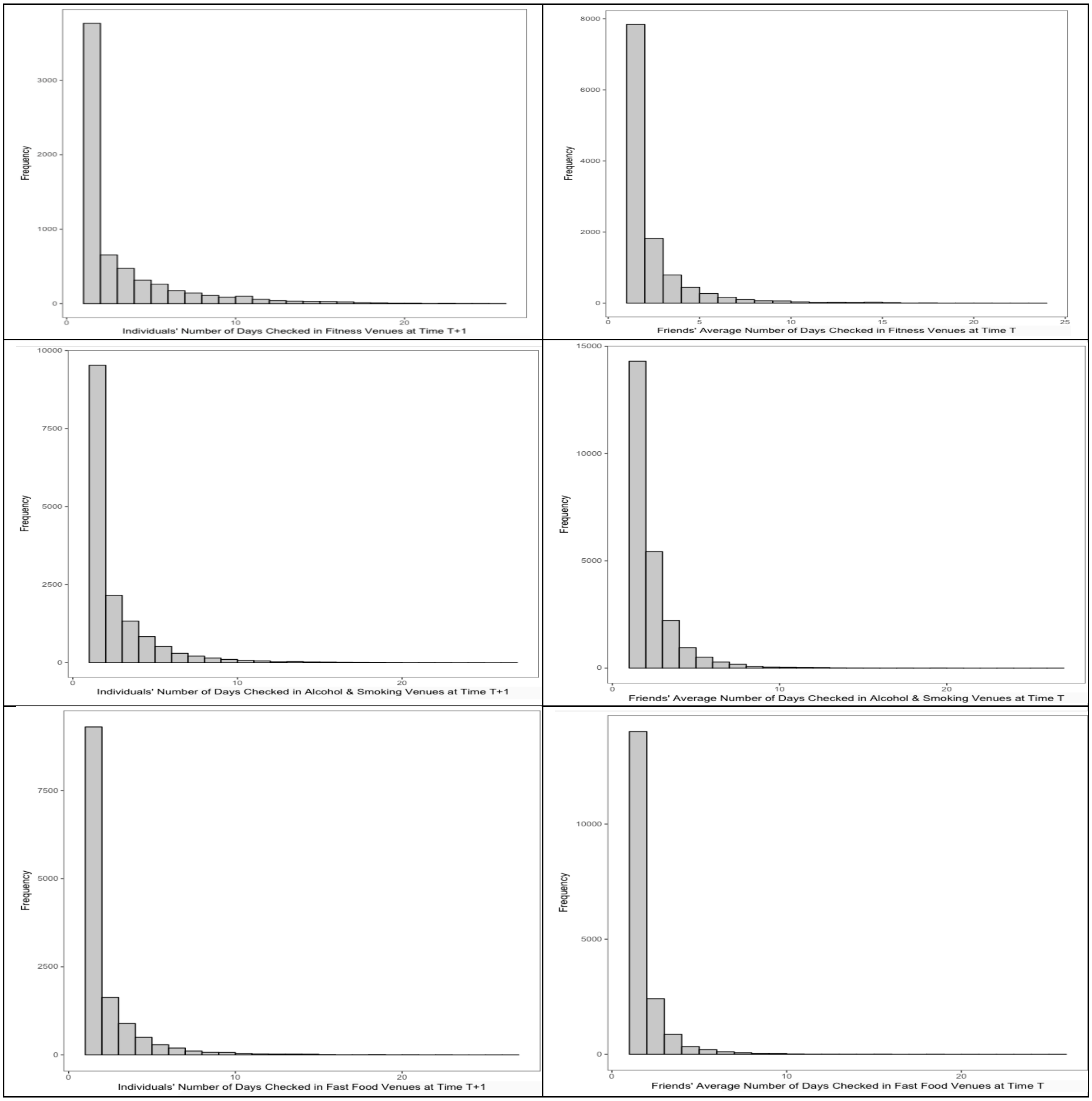
	Variable	1	2	3	4	5	6	7
1	Individuals' unhealthy behavior: A&S @ *t+1*							
2	Activity level in social network @ *t+1*	.547						
3	Individuals' unhealthy behavior: A&S @ *t*	.456	.274					
4	Socioeconomic status	-.036	.008	-.032				
5	Social support for unhealthy behavior: A&S @ *t*	.214	.098	.464	-.029			
6	Friends' unhealthy behavior: A&S @ *t*	.110	-.022	.137	-.034	.079		
7	Ratio of strong ties' unhealthy behaviors: A&S @ *t*	.059	.018	.075	-.019	.054	.044	
8	Ratio of similar friends' un healthy behaviors: A&S @ *t*	.027	.004	.037	-.017	.042	.027	.116

Table D3- Correlations for Fast Food Check-ins, N=27,253

	Variable	1	2	3	4	5	6	7
1	Individuals' unhealthy behavior: FF @ *t+1*							
2	Activity level in social network @ *t+1*	.621						
3	Individuals' unhealthy behavior: FF @ t	.395	.360					
4	Socioeconomic status	-.007	.005	-.002				
5	Social support for unhealthy behavior: FF @ *t*	.150	.122	.391	-.008			
6	Friends' unhealthy behavior: FF @ t	.083	.040	.090	-.006	.017		
7	Ratio of strong ties' unhealthy behaviors: FF @ t	.010	.018	.007	-.012	.018	-.004	
8	Ratio of similar friends' unhealthy behaviors: FF @ t	.013	.006	.016	.005	.025	-.031	.108

Appendix E:

Distributions of Check-ins of Individuals and Their Friends



Appendix F:

Label Granularity Levels

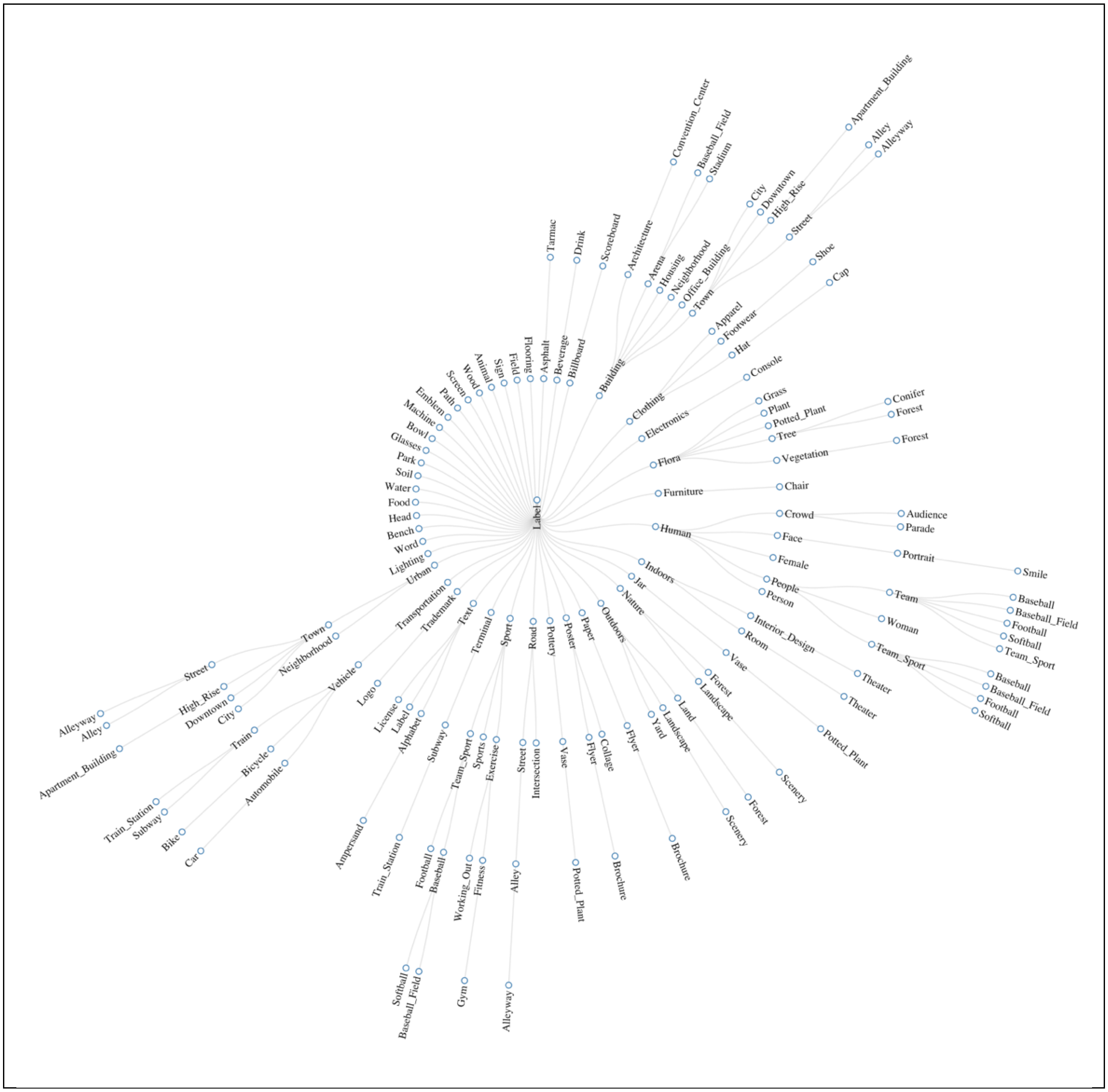


Figure F1. Label Structure for Fitness Center & Gym

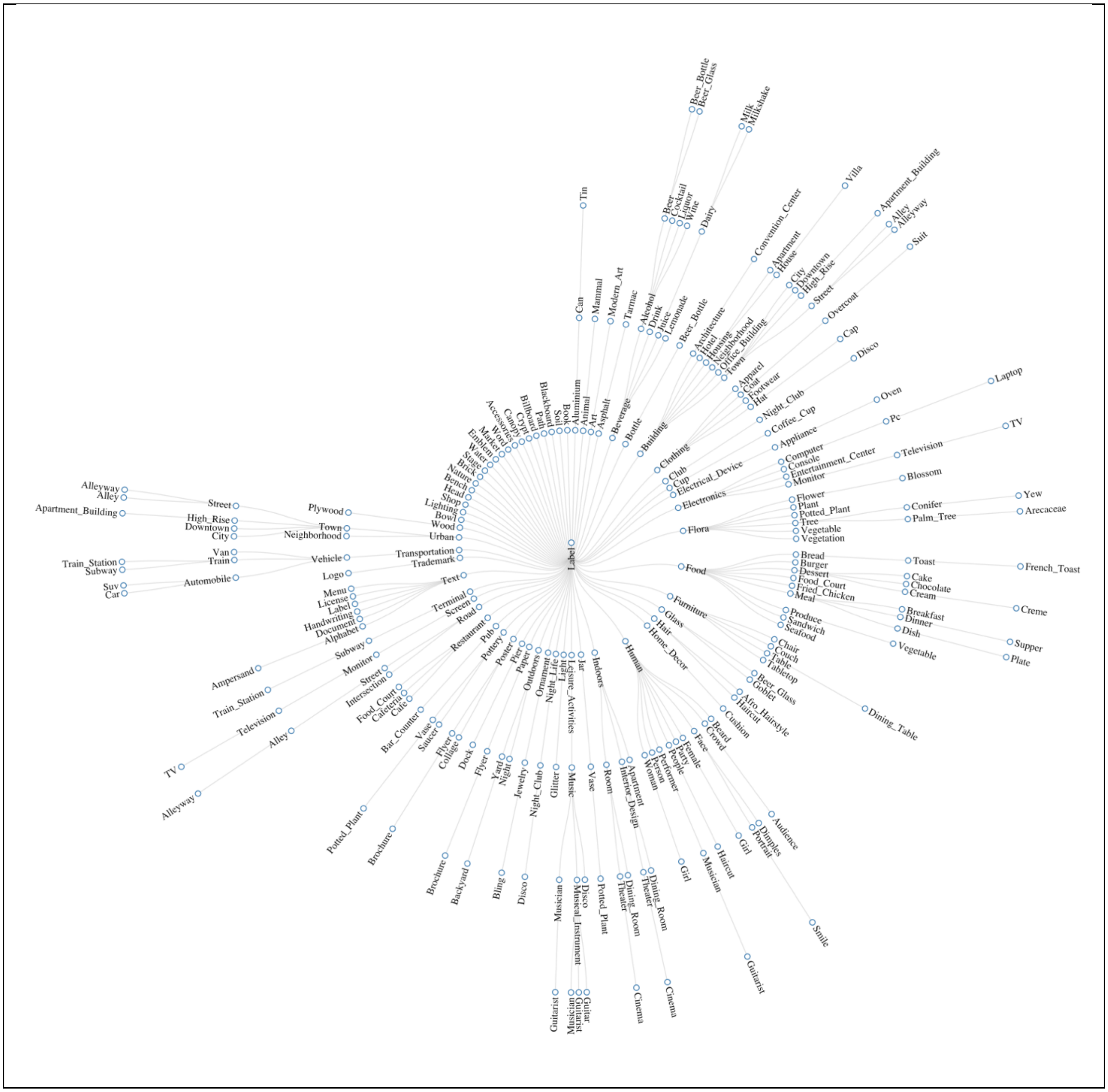


Figure F2. Label Structure for Bar

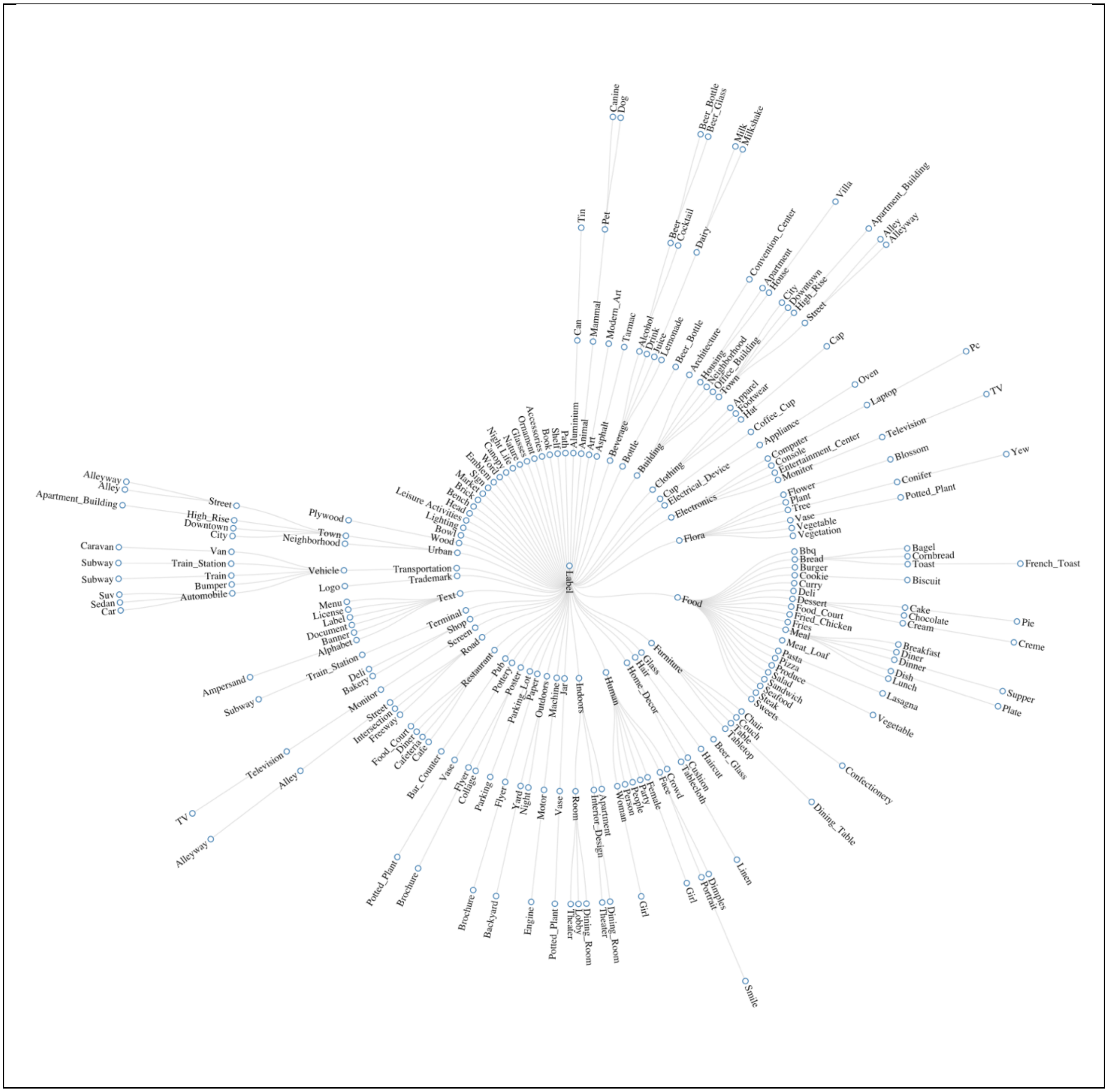
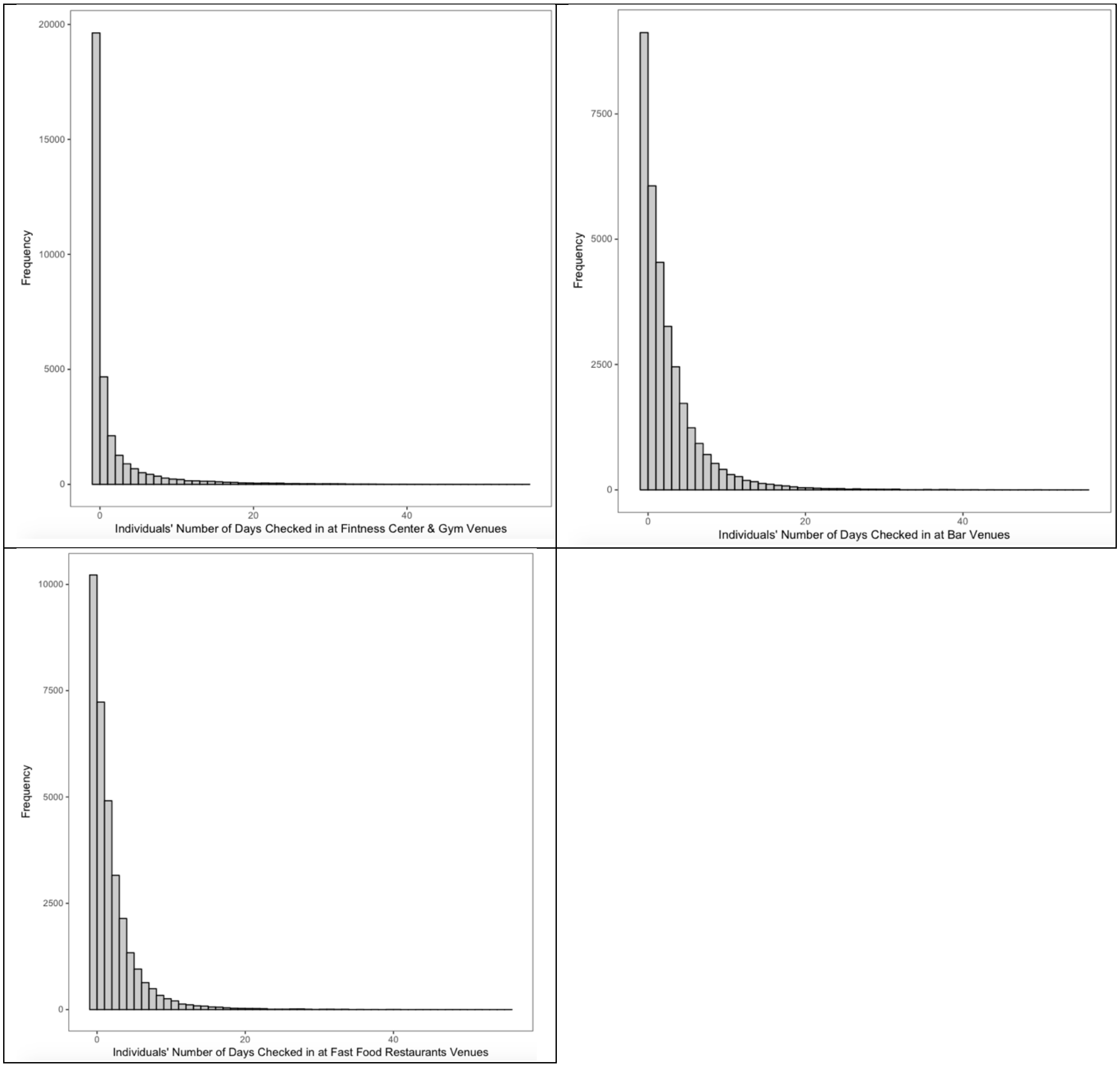


Figure F3. Label Structure for Fast Food Restaurant

Appendix G:

Distribution of Individuals' Check-ins



MAHYAR SHARIF VAGHEFI

EDUCATION

2013-2018 **Ph.D. Candidate: Information Systems Management**
University of Wisconsin – Milwaukee (Lubar School of Business)
- Recipient of Fitzsimonds Doctoral Scholarship (2016-2017)
- Recipient of Distinguished Graduate Student Fellowship (2015-2016)
- Recipient of Sheldon B. Lubar Doctoral Scholarship (2014-2105)

2009-2011 **Master of Science: Information Systems Management**
K.N.Toosi University of Technology – Tehran
- Recipient of Distinguished Graduate Student Mention Award

2001-2006 **Bachelor of Science: Computer Software Engineering**
Azad University – Tehran

WORK HISTORY

02/2003 to 07/2003 **Internship on Computer Networks**
Karafarin Bank – Iran-Tehran
- Managing Local Area Networks
- Computer Maintenance
- Remote Management of Software Packages
- Network Monitoring

09/2008 to 05/2012 **High School Teacher**
Dino-Danesh High School – Iran-Tehran
- Teaching Mathematics
- Teaching Web Programming
- Teaching Computer Networks

09/2013 to Current **Graduate Assistant**
University of Wisconsin-Milwaukee
- Research Assistant
- Teaching Assistant for Introduction to Information Technology
- Course Instructor for Introduction to eBusiness
- Course Instructor for Information Technology Infrastructure

Publications

Refereed Journals

France S, **Vaghefi MS**, Batchelder WH (2018) FlexCCT: A Methodological Framework and Software for Ratings Analysis and Wisdom of the Crowd Applications. *Computational Social Systems* 1-13.

France S, **Vaghefi MS**, Zhao H (2016) Characterizing Viral Videos: Methodology and Applications. *Electronic Research and Application* 19 19-32.

Vaghefi MS, Sharifvaghefi M, Beheshti N (2014) A Pricing Model for Group-Buying Auction Based on Customers' Waiting-time. *Marketing Letters* 25(4) 425-434

Momtaz NJ, Alizadeh S, **Vaghefi MS** (2013) A New Model for Assessment Fast Food Customer Behavior Case Study: An Iranian Fast Food Restaurant. British Food Journal 115(4) 601-613.

Conference Papers

Vaghefi MS, Nazareth D (2016) Detecting Communities of Interest in Social Media Platforms using Genetic Algorithms. Proceedings of 22nd *Americas Conference On Information Systems.* San Diego

France S, **Vaghefi MS**, Batchelder W (2013) FlexCCT: Software for Continuous CCT, The 6th *International Conference on Educational Data Mining*. Memphis