NETWORK STABILITY AND SOCIAL CONTAGION ON THE MOBILE INTERNET

Completed Research Paper

Anindya Ghose
Stern School of Business
New York University
44 West 4th Street, New York, NY USA
aghose@stern.nyu.edu

Sang Pil Han
College of Business
City University of Hong Kong
83 Tat Chee Ave, KowloonTong, Hong Kong
sangphan@cityu.edu.hk

Raghuram Iyengar
Wharton School
University of Pennsylvania
756 Jon M. Huntsman Hall
3730 Walnut Street Philadelphia, PA USA
riyengar@wharton.upenn.edu

Abstract

We study the dual roles of the stability of an individual's social network and social contagion on individual behavior in the mobile Internet setting. We use a panel dataset containing all mobile records for a sample of 3G mobile subscribers. Our data includes information about their frequency of mobile Internet usage, and their communication patterns across voice calls and messages, which allow us to map any dynamics in their social network. We find three main results. First, users with high network stability have a low intrinsic tendency to engage in content usage and generation on the mobile Internet. Second, the extent of positive social contagion effect is mitigated for users with stable networks. Third, we find that network stability is a significant predictor for individual behavior even after controlling for network closure. We discuss the implications of these findings for social network theory, social contagion and managerial practice.

Keywords: Network Stability, Social Contagion, Mobile Internet

Introduction

Marketers have become very keen on using on using individual-level data to better understand how peers affect individual behavior, especially in adoption and use of new services. Recent research has focused on documenting the presence of social contagion after controlling for potential confounds such as marketing effort, customer heterogeneity, and various contextual factors (Ghose and Han 2011a, Hill et al. 2006, Libai et al. 2010b, Oestricher-Singer and Sundararajan 2011, Van den Bulte and Lilien 2001). While such work provides evidence that peer effects are important for individual behavior, it typically assumes that the ties among individuals (i.e., who interacts with whom) are static or unchanging (e.g., Coleman et al. 1966, Iyengar et al. 2011a). Such an assumption may be reasonable for some contexts (e.g., networks among firms in business to business settings) but may not be for many others. For instance, Aral et al. (2009) find that communication patterns among customers of a mobile service application are quite dynamic. Recognizing the importance of network dynamics, sociologists have been giving increasing attention to how structural properties, such as reciprocity, of networks change over time (Doreian and Stokeman 1997, Huisman and Snijders 2003, Snijders and Doreian 2010, Snijders et al. 2010).

The stability of an individual's social network is the extent of overlap in their contacts over time. At the tie level, network stability captures how often ties reappear in a network (Bien et al. 1991, Broese van Groneau et al. 1990). Thus, people with unstable networks may contact a diverse range of people over time and are likely to obtain information from non-redundant sources. Past research shows that such information acquisition from non-redundant sources can impact individual behavior (Tucker 2011). Our current research investigates the role of network stability on individual behavior by addressing two key questions. First, how does network stability affect individual behavior in managerially relevant contexts? Second, does it moderate the impact of peer effects on individual behavior? Understanding the role of stability is important not only theoretically but also managerially because which customers to target and which ties to activate using what message depends on it (Van den Bulte and Wuyts 2007).

Network stability can impact individual behavior based on its relationship with another key structural property of networks - network closure. While network stability captures temporal changes to a social network, network closure describes cross-sectional relationships among network neighbors. In particular, network closure summarizes the extent to which two contacts within a network are reconnected by mutual third parties (Coleman 1990). According to Coleman (1988), networks with more closure – networks in which everyone is connected to each other – provide more benefits, such as access to information and trust within the network, compared to networks with less closure. For example, Aral and Van Alstyne (2011) emphasize that cohesive networks tend to have greater channel bandwidth, such that an individual can access more detailed information. Similarly, Coleman (1990, p. 306) notes that economic institutions such as rotating-credit associations operate only in environments with a high degree of trustworthiness, which is a result of high network closure among the members.

While high closure in a network leads to trustworthiness, it also results in a high degree of redundancy in ties from a perspective of information transfer among members (Burt 1992, 2000). Similarly, Reagans and Zuckerman (2008) note that as the ties in a network become non-redundant, individuals can be expected to become more knowledgeable and more quickly. Whether trust versus non-redundancy in ties within a network is valued depends on the context. In contexts where there is significant perceived risk, high network closure, and the resulting high level of trust among members, is useful (Allcott et al. 2007). The earlier example of a rotating credit association is one such context. In contrast, in a context in which there is little risk, non-redundancy of ties and low network closure may be preferable as it provides individuals with access to more network neighbors. In sum, network closure plays an important role in how peers may impact individual behavior and how it does so depends on the application context.

The relationship between network closure and network stability is not immediately obvious. Past research indicates that network closure is positively related to tie strength (typically based on volume of communications) among members - networks with high volume (or strong) ties in general have a high level of closure (Granovetter 1973, Onnela et al. 2007). This captures the notion that "my close friends are friends of one another". Network stability, on the other hand, is related to incidence of communications, which can be thought of the rate at which members with whom one communicates changes over time. Thus, the relationship between network closure and network stability depends on whether stable ties (with high incidence) are also strong ties (high volume ties). The extent and directional nature of the impact of network stability on social contagion then is an empirical question and depends on its relationship with network closure.

While the concept of network stability and its impact on individual behavior is an intriguing question to investigate, it would be difficult to do so without having access to data that reliably captures the relationship among individuals in large scale realistic settings. Typically, social network analysis uses self-report data from surveys to construct ties between network members (e.g., Bramoullé et al. 2009, Coleman et al. 1966, Conley and Udry 2010, Iyengar et al. 2011a). While a main benefit is that questions in a survey can be tailored such that they capture the type of tie that is of theoretical interest, self-reports can be limiting in terms of the breadth of the network being investigated. As a result, typically small, well-bounded populations are examined in detail (Christakis and Fowler 2011).

To address our research questions, we secured the cooperation of a telecommunications company that gave us access to all mobile records of a large sample of 180,000 subscribers over a five week period. For all users, our data includes information about their weekly frequency of mobile Internet usage, and their communication patterns across three modes, voice calls, short messaging service (SMS) and multimedia messaging (MMS), which allow us to map dynamics in their social network. We develop an econometric model to investigate whether customer-level mobile Internet usage can be explained in part due to their dynamic communication patterns and Internet usage of their network neighbors. The institutional details of our research context are such that there is unlikely to be any perceived risk for members.

We find three main results. First, users with high network stability have a low intrinsic tendency to engage in content usage and generation on the mobile Internet. Second, the extent of positive social contagion effect is mitigated for users with stable networks. Speculatively, both these results are due to our context involving little perceived risk as well as stable networks showing significantly more network closure. Members within a more clustered social network are likely to receive more redundant information from their peers, and this is less likely to lead to any changes in individual behavior. This suggests that, all else being equal, network neighbors that one less frequently communicates with, and hence shares less information that is in common, are more likely to affect one's usage decision more so than network neighbors that one more frequently communicates with in the network. Third, we find that network stability is a significant predictor for individual behavior even after controlling for network closure. We speculate that this is due to network instability leading individuals to be uncertain about their future communication patterns.

Theory and Literature Review

Numerous studies have documented evidence of positive social contagion in adoption and use of new services and products (Van den Bulte and Lilien 2001, Hill et al. 2006, Aral et al. 2009, Trusov et al. 2009, Libai et al. 2010, Oestricher-Singer and Sundararajan 2011, Ghose and Han 2011), the switching from an existing service provider (Dasgupta et al. 2008), and the diffusion of content in online social network space (Susarla et al. 2010, Katona et al. 2011). More recent work has also begun investigating *whether* contagion is really at work to *why* it occurs (e.g., Iyengar et al. 2011b). Social contagion may occur for several reasons. In the context of new product adoption, these can be organized in five categories. The process may operate through (i) spreading awareness and interest, (ii) social learning leading one to change one's beliefs about the product's risks and benefits, (iii) social-normative influence increasing the legitimacy of the new product, (iv) concerns that not adopting may result in a competitive or status disadvantage, or (v) direct and indirect "network" or installed base effects (Van den Bulte and Lilien 2001).

Understanding the mechanisms driving contagion is important not only theoretically but also managerially, because which customers to target as seeding points and which ties to activate using what message or appeal depends on what mechanism is at work (Van den Bulte and Wuyts 2007). For instance, if contagion works through spreading awareness and interest, then viral campaigns could be made more effective by making the message more "buzz-worthy" by focusing on unusual or otherwise remarkable content. Also, for such campaigns, reaching far into the network is the main objective so seeding people with great indirect coverage should be effective. If contagion operates through social learning, in contrast, the campaign should be designed quite differently. Mobilizing the expertise embedded in the network should be a key campaign objective. So, a marketer should then focus on experts and favor people with many direct ties as opposed to people with great indirect coverage.

This distinction in the mechanisms behind contagion also helps in reconciling the seemingly opposing ways in which some individual characteristics moderate social contagion. For example, in the context of the adoption of a risky drug, Iyengar et al. (2011a) find that the amount of social contagion is moderated by the volume of product usage of the contagion sources. Godes and Mayzlin (2009) focus on the adoption of a relatively safe product (a brewery chain) and, in contrast, find that light users may be more likely to generate extra adopters. Whether usage volume enhances or depresses the amount of social contagion exerted is likely to depend on whether contagion operates by boosting awareness or evaluation, the two key stages in the adoption process (e.g., Lin and Burt 1975). For products that do not benefit from marketing communication and that present little perceived risk such that no additional information is required in the evaluation stage, light users will be very effective sources of influence to spread awareness as heavy users tend to be connected mostly to people already predisposed to be early adopters. For products that are supported by a fair amount of standard marketing communication but pose significant perceived risk, in contrast, contagion operates at the evaluation stage rather than at the awareness stage, so heavy users are likely to be more effective sources of influence.

The distinction between underlying mechanisms for contagion is also relevant for how network characteristics (e.g., network closure) may function as its moderators. For instance, in contexts where contagion spreads through awareness, low network closure is useful as it allows individuals with a wider indirect coverage. With low network closure, network ties are less redundant and information can spread faster (Burt 2000, Reagans and Zuckerman 2008). On the other hand, in more risky contexts where contagion operating at the evaluation stage, a dense network with high closure is more important (Allcott et al. 2007). In such situations, redundancy in the information from contacts may actually be desirable, since it is both reassuring and reaffirming one's beliefs (e.g., Iyengar et al. 2011a). In certain instances, multiple doses of influence may actually even be necessary to change consumer behavior. For instance, Centola and Macy (2007) describe “complex contagions” as those that require multiple doses of influence and contrast it with “simple contagions” where a single source of information is enough for information to spread (e.g., rumors). In a similar vein, Aral and Van Alstyne (2011) contrast “wide bridges” with “thick bridges” and examine whether it is more important to have multiple reinforcing doses of influence from infrequent weak ties (wide bridges á la Centola and Macy) or rather few doses of influence from a few trusted strong tie that reside in a distant network neighborhood (thick bridges á la Aral and Van Alstyne).

Understanding how network closure moderates contagion depending on the context is important for explaining the role of network stability. Past research indicates that tie strength among members is related to network closure - strong ties among individuals in general lead to a high level of closure in their network (Granovetter 1973, Onnela et al. 2007). If others that a focal customer calls regularly (stable ties) are those that he/she interacts most with as well (high volume ties), then network stability should moderate social contagion in a manner similar to network closure. Further, network stability may have an impact on individual behavior even after controlling for network closure.

Research Setting

To provide valid answers to our research questions the research setting should ideally satisfy several conditions. First, context should have characteristics making it theoretically justified to expect contagion to be at work. Second, one must have data on who can influence whom. Third, one must have data not only for each person whose behavior is analyzed, but also on the behavior on usage of others in their network. Fourth, any marketing efforts deployed must be observed or otherwise controlled for.

We secured the cooperation of a telecommunications company to meet these conditions. The company, like many others in its industry, was interested in understanding how its customers use data-related services such as access to the Internet. In our context, although users can access to the entire Internet, there are two broad categories of websites that users access through their mobile phones: regular social networking websites and mobile carrier's portals. Subscribers can either generate content and upload it on these websites or download multimedia content from these sites on their mobile phone. The former is termed as uploads while the latter are downloads. These websites are community-oriented sites that allow users to download and upload (to share with others) multimedia content like photos, music, videos, apps, etc. Further, three modes of mobile communication – voice calls, short messaging service and multimedia messaging– allow consumers to interact with each other and become aware of the behavior of their network neighbors. We consider these channels as being proxies for social relationships and users can discuss their mobile content they download or upload through all manner of communication channels, but that connected users are more likely to do that than those who are not connected via these networks (i.e., voice, SMS, and MMS). In addition, the firm did not do any targeted marketing during our sampling period to promote their data related services.

Our research setting allows us to model awareness based contagion operating through different members that a focal customer interacts with. As explained in greater detail in the subsequent sections, this is modeled as volume-weighted contagion operating over the three modes of communication. The institutional details of the research setting are such that other contagion mechanisms are very unlikely to be at work. Since there is little perceived risk in the usage of Internet over a mobile phone, contagion is unlikely to take place at the evaluation stage. There are also no competing standards for technology and hence the use of the Internet by individual member is independent of the size of the installed base. Finally, there is no reason to expect competitive contagion as the company does no targeted advertising.

Data

Our dataset contains mobile records for a sample of 180,000 3G mobile subscribers who used the services of this company between May 10, 2008 and June 15, 2008 (5 weeks). For these users, we have information on (i) weekly frequency of Internet content generation and usage (2.34 million records), (ii) voice calls with the sender and receiver identification numbers, which allow us to map their social networks (7 million records) and (iii) SMS and MMS communication data containing the sender and receiver identification numbers, which allow us to build alternative social networks (24 million SMS records and 0.5 million MMS records). Compared to the prior analysis by Ghose and Han (2011a) where only voice data was used to construct social networks, data item (iii) is new and makes it possible to construct alternative social networks including a different group of people a user may communicate with when using SMS or MMS services. Social network variables in our sample correspond to the level of content activity (i.e., frequencies of content uploads and downloads) of network neighbors for each user in the sample. Table 1 shows the summary statistics for the key variables. As shown in Table 1, user content downloading in our data occurs more frequently than does user content uploading.

Covariates

Network Stability: We measure the ego-centric network stability for each user using the extent of overlap of their network neighbors between time *t*-1 and time *t* (Broese van Groneau et al. 1990, Bien et al. 1991, Ferligoj and Hlebec 1999). For a given user *i* at time *t*,

$$\text{Network Stability}_{i,t} = \frac{\text{number of people user i contacted at both time t} - 1 \text{ and time t}}{\text{number of people user i contacted at time t} - 1}.$$

For example, if user i calls user A, B, and C at time t-1, and calls user A, B, D, and E at time t, then the network stability of user i at time t is computed as 2/3. Note that the metric of network stability only considers whether there is contact between users and ignores the volume of their interactions. Also we only consider a contact to be those a user called or sent messages to (not those a user received calls or messages from). We capture the strength of communication among users with social network weights, which are described later.

Network Closure: Network closure measures the extent to which a network is directly or indirectly concentrated in a single contact (Burt 1992). Network closure/constraint is low for large networks of disconnected contacts and high in a small network of contacts densely connected to one another (Burt 2000). Network closure for user i at time t, c_{it} measures the extent of redundancy in a network neighborhood of user i as follows:

$$C_{i,t} = (p_{ijt} + \sum_q p_{iqt} p_{qjt})^2,$$

where $q \neq i, j$ and p_{ijt} is the proportion of user i calls made to user j during week t.

Table 1: Summary Statistics

Variable	Observations	Mean	Std. dev.
Weekly, User-Specific Content Activity Data			
Mobile Internet Session Activation (1: Yes, 0: No)	900,000	0.27	0.44
Number of Uploads	243,449	0.25	3.39
Number of Downloads	243,449	22.31	88.41
Weekly, User-Specific Communication Data			
Number of Voice Calls Made	900,000	11.88	16.32
Voice Call Duration (Hours)	900,000	2.61	5.68
Number of SMS made	900,000	40.20	97.10
Number of MMS made	900,000	3.24	6.33
Network Characteristics			
Network Stability by Voice Call	567,411	0.66	0.33
Network Closure by Voice Call	622,153	0.69	0.30
Network Stability by SMS	560,658	0.38	0.33
Network Stability by MMS	159,049	0.17	0.31
User Characteristics			
Gender (1: Male, 0: Female)	180,000	0.53	0.50
User Age (Years)	180,000	30.13	5.91
Handset Age (Months)	180,000	9.63	3.97

Notes: Content generation and usage data is observed only when a user starts mobile Internet sessions; thus the number of uploading and downloading observations is lower than the number of session activations. SMS refers to short-messaging service, and MMS refers to multimedia-messaging service.

Social Contagion: We operationalize the exposure to neighbors through network ties using lagged endogenous autocorrelation terms. The extent to which user i is exposed at time t-1 to neighbors is captured through the term $\Sigma_j\, w_{ijt-1}\, z_{jt-1}$ where w_{ijt-1} captures how relevant each user j is to i at time t-1 where z_{jt-1} is a variable capturing the behavior of j at time t-1. To address potential endogeneity issues, similar to past literature, we use a lagged social network variable (Nair et al. 2010) and a lagged social network weight (Ghose and Han 2011a). The lagged social network weight w_{ijt-1} can be constructed in various ways (e.g., Valente 1995). We use two metrics, call frequency and call duration, to capture the strength of the communication between a user and his/her network neighbors. We operationalize a normalized social network weight in the following manner: Based on call frequency, the lagged social network weight w_{ijt-1} is determined as a fraction of call numbers from user i to user j in week t-1 with respect to the total call numbers originating from user i in week t-1. Similarly, based on call duration, the social network weight w_{ijt-1} is determined as a fraction of call duration from user i to user j in week t-1 with respect to the total call duration originating from user i in week t-1. In addition, we capture the behavior of fellow users by using their level of Activity, which are either Uploads or Downloads.[1] Having operationalized both the social network weight w_{ijt-1} and behavior z_{jt-1}, we then calculate the extent of social network exposure that user i is experiencing at time t and create the following variable:

$$Social\ Network\ Activity_{i,t} = \sum_{m \in n_{t-1}(i)} \left(w_{i,m,t-1} * Activity_{m,t-1} \right)$$

where *Activity* is either upload or download (see Table 2 for detailed description of notations) and $n_{t-1}(i)$ is user i's social network neighbors based on voice call (or SMS or MMS) records in week t-1. To assess the

[1] We also tested a "cross-network effect" in which one's upload (download) activity is affected by network neighbors' download (upload) activity. This can arise if there is social reciprocity or if there is some local social network supply and demand such that the content a focal user uploads is what is available to be consumed by their friends and the content they upload is consumed by the focal user. We find qualitatively the same results with and without this control. Details are available upon request.

extent to which network stability moderates the effect of this contagion variable, we also create the necessary interaction term (i.e., *Network Stability*$_{it}$ x *Social Network Activity*$_{it}$).

Table 2: Notations and Variable Descriptions

$\text{Upload}_{i,t}$	Number of times user *i* uploaded in week *t*
$\text{SN\_Upload}_{i,t-1}$	Weighted number of times user *i*'s network neighbors uploaded in week *t*-1
$\text{Download}_{i,t}$	Number of times user *i* downloaded in week *t*
$\text{SN\_Download}_{i,t-1}$	Weighted number of times user *i*'s network neighbors downloaded in week *t*-1
$\text{n}_{\text{t-1}}(i)$	User *i*'s network neighbors based on voice call (or SMS or MMS) records in week *t*-1
$w_{i,m,t-1}$	Normalized number of calls user *i* made to user *m* in week *t*-1, that is, $w_{i,m,t-1}$ is a fraction of voice calls from user *i* to user *m* in week *t*-1 with respect to the total voice calls originated from user *i* in week *t*-1
$\text{Stability}_{i,t}$	The extent of network stability of user *i* at week *t*
$\text{Outdegree}_{i,t}$	The number of call (or message) receivers of user *i* at week *t*. The size of $\text{n}_{\text{t-1}}(i)$.
$\text{Selection}_{i,t}$	Selection correction term for user *i* at week *t*
$g_{-i,t}$, $h_{-i,t}$	Mean uploading and downloading frequencies of all other users in user *i*'s billing (or calling) zip code area in week *t*, respectively
τ_t , φ_t	Time-period dummies for download and upload equations, respectively
$\varepsilon_{i,t}$, $\nu_{i,t}$	Unobservable, user-specific, time-specific effect for download and upload equations, respectively, $\nu_{i,t} \sim IIN\left(0,\sigma_\nu^2\right)$ and $\varepsilon_{i,t} \sim IIN\left(0,\sigma_\varepsilon^2\right)$

Other Communication Networks: Since a user may communicate with a different group of people when using SMS or MMS as opposed to voice calls, these two additional modes of communication allow us to capture a user's alternative social networks. We use these alternative networks to check for the robustness of our main results based on voice call data. As shown in Table 1 Summary Statistics, the weekly average number SMS's of a user are relatively higher than that for voice calls (i.e., 40.20 > 11.88). However, the weekly average number MMS's of a user is relatively lower than that for voice calls (i.e., 3.24 < 11.88). Further, the mean for the network stability based on SMS and MMS data is 0.36 and 0.16, respectively. Note that the mean for network stability based on voice calls is 0.66. Hence, our data suggests that on average, a voice network is relatively more stable than either SMS or MMS networks.

Control Variables: To distinguish this spurious effect from the impact of social contagion, similar to Ghose and Han (2011a), we incorporate additional controls – (i) time-period fixed effects, (ii) location fixed effects, and (iii) time and location fixed effects. Time period fixed effects control for common factors or shocks to all individuals at a given time (e.g., the mobile company's nationwide mobile marketing campaigns, such as free trial downloads of content). Location- fixed effects (i.e., zip code dummies) control for time-invariant, spatially correlated unobservables. For example, users in urban areas may be more tech-savvy and more prone to engaging in mobile Internet activities, compared to users in rural areas. Finally, time and location fixed effects (i.e., spatio-temporal effects) control for unobservables that correlate at the level of zip code and time. For example, celebrity sightings, social events, unusual street incidents, etc., may give people the opportunity to capture and share such moments with friends and families via their mobile phones. Hence, we include time- and location-specific mean content upload and download frequencies of all other users in the zip code area of user *i*, as denoted by g_{-it} and h_{-it} in the main equations, respectively. We also apply both billing zip code-based and calling zip code-based content activity variables. The former captures unobservables that correlate to a user's time-invariant billing address, while the latter captures unobservables that correlate to a user's time-varying travel locations. We also include a time-varying control for the outdegree of a user (i.e., consistent with the lagged endogenous variable, the number of call receivers or message receivers at time *t*-1). To account for any potential selection bias in who uploads and downloads content, we include selection correction terms.

Model

We estimate an individual-level, simultaneous equations panel data model, using three-stage, least-squares (3SLS) estimation. In the mobile Internet space where users generate and use content with their phones, users face a two-step decision-making process (Ghose and Han 2011): In Step 1, users decide whether to initiate a mobile Internet session by touching the mobile phone screen. In Step 2, once users have initiated a mobile Internet session, they determine how much data to upload (if any) and how much to download (if any). They can also engage in both uploading and downloading activities multiple times in a single given mobile Internet session. To incorporate the two-step decision-making process users undertake, we explicitly specify our econometric model by extending Verbeek and Nijman's (1996) two-step method. In Step 1, as related to the user's decision made in Step 1, we run a random effect dynamic probit model for the user's binary decision to initiate a mobile internet session or not initiate one during a given week. Estimates from Step 1 are then used to obtain a Heckman's (1979) selection correction term. In Step 2, we insert that correction term into content usage and content generation equations, respectively, and estimate the two equations simultaneously, using the three-stage least-squares (3SLS) method. Our model consists of selection equations and main equations. Details about selection equations and the estimates of the selection equations are not reported only due to brevity. Notations and variable descriptions are in Table 2.

Main Equations

We include network stability in our main equations. To assess the extent to which ego-network stability moderates the social contagion effect, we also include interaction terms. We include the number of call receivers to control for the user's range of social network. Finally we incorporate control variables, including user-specific dummies, time-period dummies, and time-period and location specific fixed effects at the user level to control for endogeneity from using a social network variable as a regressor.

We take the logarithm for variables to control for their right-skewed nature. We implement 3SLS estimation on the first-differenced equations of log-transformed content usage and content generation frequencies. The simultaneous estimation method allows for efficiency gain as compared to the single equation estimation methods by taking into account the cross-equation error correlation. The first-differencing transformation on each variable in the model alleviates the potential bias from the fixed-effect model (Wooldridge 2002) and difference out both observed and unobserved user-specific, time-invariant variables (e.g., age, gender, job characteristics, prior mobile Internet experience). Although we find there is no serial correlation in the error term from the selection equation as well as in the error term from each of the main equations separately, we control for any potential serial correlation in the main simultaneous equations of content generation and usage by using the robust variance matrix estimator (Wooldridge 2002). The robust variance matrix estimator (Arellano 1987) is valid in the presence of serial correlation in error terms (Wooldridge 2002). To be specific, the first-differenced content usage frequency and generation frequency equations that we estimate are specified as follows, for t = 2, 3…T:

$$\begin{aligned}\Delta\log\left(Download_{i,t}\right) = {} & \gamma_1\Delta\log\left(SN\_Download_{i,t}\right) + \gamma_2\Delta Stability_{i,t} + \gamma_3\Delta\log\left(SN\_Download_{i,t}\right) * \Delta Stability_{i,t} \\ & + \gamma_4\Delta\text{Outdegree}_{i,t} + \gamma_5\Delta\log\left(h_{-i,t}\right) + \gamma_6\Delta Selection_{i,t} + \Delta\tau_t + \Delta\varepsilon_{i,t}\end{aligned} \tag{1}$$

$$\begin{aligned}\Delta\log\left(Upload_{i,t}\right) = {} & \beta_1\Delta\log\left(SN\_Upload_{i,t}\right) + \beta_2\Delta Stability_{i,t} + \beta_3\Delta\log\left(SN\_Upload_{i,t}\right) * \Delta Stability_{i,t} \\ & + \beta_4\Delta\text{Outdegree}_{i,t} + \beta_5\Delta\log\left(g_{-i,t}\right) + \beta_6\Delta Selection_{i,t} + \Delta\varphi_t + \Delta\nu_{i,t}\end{aligned} \tag{2}$$

Identification

We discuss two issues in the identification of our model: identifying the main equations and identifying the social contagion effect. For the main equations, we impose an exclusion restriction by including variables in the selection equations that are not included in the main equations. Such restrictions make the identification cleaner (Puhani 2000). We include a few time-invariant variables (e.g., age and gender) as well as a time-varying variable (e.g., mobile Internet session initiation by social network) only in the selection equation, but exclude these variables in the main equations. In addition, in the absence of better data, we use time-series-based instruments for identification in the main equations. Similar to Ghose and

Han (2011), content download and upload frequency variables in a given week are taken to be endogenous to the system of equations, whereas all other variables in the system are treated as exogenous to the system or predetermined. For example, in the content download equation, variables like lagged network stability and lagged social network are exogenous or predetermined. As a robustness check, we also include a non time-series-based variable as an instrument. In particular, we include the handset age variable as an additional instrument in the content generation equation, and exclude it from the content usage equation. The key assumption behind this argument is that the age of the handset is more likely to impede users from uploading multimedia content than downloading it. The qualitative nature of all results remains the same even with the inclusion of this variable as an instrument.

Identifying peer influence is by itself difficult and the subject of a lot of scientific debate and criticism (Aral and Walker 2011). Previous research shows that correlated unobservables between a user and his or her network neighbors may impact individual behavior and will be confounded with the true contagion effect (e.g., Nair et al. 2010). Shalizi and Thomas (2011) claim that the identification of social influence is generically confounded without experiments. That said, to address the endogeneity issue of the social contagion effect, we adopt the identification strategy and modeling approach in accordance with prior work (e.g., Manski 1993, Hartman et al. 2008). We specifically control for each of three sources of spurious correlation: (1) endogenous group formation, (2) correlated unobservables, and (3) simultaneity. First, regarding endogenous group formation, the observed correlation in the behavior of an individual and other individuals in the social network could arise from omitted individual characteristics that correlate within the group. We include a user-specific random effect in the selection equations (see Appendix) and difference out a user-specific fixed effect in the main equations. Second, regarding correlated unobservables, some unobservables could impact the behavior of a focal individual and other individuals in their social network. We incorporate aforementioned three controls – time-period fixed effects, location fixed effects, and time and location fixed effects. Third, simultaneity can arise if network neighbors affect the user and the user, in turn, also affects them simultaneously. We use a lagged social network variable and a lagged social network weight. Nonetheless, we would like to be cautious in our interpretation on social contagion effects. In the absence of controlled variation using natural or field experiments during our sampling period, we interpret the social contagion effect in our paper as being one that establishes an upper bound on the causal effect of the social contagion.

Results

Correlation Results

We find a correlation of 0.09 ($p < 0.001$) between network stability and network closure. The correlation is low but significant and positive. Given the positive correlation, network stability should moderate contagion in a manner similar to network closure. In particular, given the low perceived risk in our context and that contagion will operate at the awareness stage, low network stability should be better as it allows individuals with a wider indirect coverage. Note that the low correlation suggests that stability may still be an important predictor of individual behavior even after controlling for network closure. We further corroborate the relationship between network stability and information redundancy by considering its correlations with volume of interactions with network members. We find significant correlations of 0.19 ($p < 0.001$) and 0.12 ($p < 0.001$) between network stability and frequency and volume of communication with other members, respectively. Recall that strong ties are those characterized typically by high volume of interactions (Granovetter 1973, 1982). Thus, the positive correlations indicate that stable ties are also high volume ties. Put differently, high network instability leads to network diversity with weak tie relationships, which in turn provide non-redundant information.

Model Estimates

Table 3 shows the main results. The top (bottom) panel corresponds to the estimates for the download (upload) equation. In each panel, the first (last) two columns correspond to the estimates for frequency (duration) of a weighted social network. Further, the first and third (second and fourth) columns correspond to models with billing-based (calling-based) controls. Note that billing zip code-based controls capture unobservables that correlate to a user's time-invariant billing address, while calling zip code-based controls capture unobservables that correlate to a user's time-varying travel locations.

Our results in Table 3 show that the relationship between content usage and generation behavior of users and their social networks is positive and statistically significant. As an example, the coefficient estimates are 0.165 and 0.298 in the download and upload equations, respectively, for the model with frequency-based social contagion variables and calling-based spatio-temporal controls. We also find that our network stability metrics based on voice call data negatively associate with content usage and generation activities. For example, network stability associates with content usage and generation (the coefficient estimates are -0.076 and -0.019, respectively). This result suggests that users with high network stability have a lower intrinsic tendency to engage in content usage and generation. Interestingly, our results show that the interaction effect between network stability and social network variable is negative and statistically significant (for example, the coefficient estimates are -0.119 and -0.222, respectively). That is, ego-network stability negatively moderates the social contagion effect. Thus, the extent of positive social contagion effect is mitigated for those users with high network stability.

Table 3: Parameter Estimates for Main Model based on Voice Call Network Data

Independent Variable	Estimate			
	Frequency-Weighted Voice		Duration-Weighted Voice	
	Billing-based Control	Calling-based Control	Billing-based Control	Calling-based Control
Download Equation				
Log SN Downloading	0.165 (0.004)***	0.162 (0.004)***	0.164 (0.004)***	0.162 (0.004)***
Network Stability	-0.076 (0.010)***	-0.085 (0.010)***	-0.077 (0.010)***	-0.086 (0.010)***
Log SN Downloading * Network Stability	-0.119 (0.014)***	-0.125 (0.014)***	-0.118 (0.014)***	-0.124 (0.014)***
Log Out Degree	0.175 (0.010)***	0.171 (0.010)***	0.176 (0.010)***	0.171 (0.010)***
Billing-based Time and Location Effects	0.068 (0.002)***		0.068 (0.002)***	
Calling-based Time and Location Effects		0.183 (0.002)***		0.183 (0.002)***
Selection Correction Term	0.120 (0.041)***	0.120 (0.041)***	0.121 (0.041)***	0.121 (0.041)***
Chi-square	37648.35	44121.22	37625.08	44097.03
Upload Equation				
Log SN Uploading	0.298 (0.005)***	0.298 (0.005)***	0.292 (0.005)***	0.292 (0.005)***
Network Stability	-0.019 (0.002)***	-0.019 (0.002)***	-0.020 (0.002)***	-0.020 (0.002)***
Log SN Uploading * Network Stability	-0.222 (0.017)***	-0.223 (0.017)***	-0.206 (0.017)***	-0.208 (0.017)***
Log Outdegree Degree	0.004 (0.002)**	0.004 (0.002)**	0.004 (0.002)**	0.004 (0.002)**
Billing-based Time and Location Effects	0.029 (0.004)***		0.029 (0.004)***	
Calling-based Time and Location Effects		0.074 (0.004)***		0.074 (0.004)***
Selection Correction Term	0.020 (0.010)**	0.020 (0.010)**	0.020 (0.010)**	0.020 (0.010)**
Chi-square	4847.87	5213.73	4654.33	5019.72

Notes: The dependent variables are log-transformed frequency of content downloading and uploading, respectively. SN refers to social network. Estimates for individual-specific dummies, time dummies, and indicators for missing network stability metrics are omitted. *** significant at 0.01, ** significant at 0.05.

Table 4: Results on Main Model using Unnormalized Weights based on Voice Network Data

Independent Variable	Estimate			
	Frequency-Weighted Voice		Duration-Weighted Voice	
Download Equation				
Log SN Downloading	0.123 (0.004)***	0.122 (0.003)***	0.039 (0.001)***	0.038 (0.001)***
Network Stability	-0.076 (0.010)***	-0.086 (0.010)***	-0.078 (0.010)***	-0.088 (0.010)***
Log SN Downloading * Network Stability	-0.101 (0.010)***	-0.106 (0.010)***	-0.037 (0.004)***	-0.039 (0.004)***
Log Outdegree	0.177 (0.010)***	0.173 (0.010)***	0.184 (0.010)***	0.179 (0.010)***
Billing-based Time and Location Effects	0.068 (0.002)***		0.068 (0.002)***	
Calling-based Time and Location Effects		0.183 (0.002)***		0.183 (0.002)***
Selection Correction Term	0.121 (0.041)***	0.121 (0.041)***	0.121 (0.041)***	0.121 (0.041)***
Chi-square	37461.14	43930.23	36833.64	43261.25
Upload Equation				
Log SN Uploading	0.145 (0.003)***	0.144 (0.003)***	0.041 (0.001)***	0.041 (0.001)***
Network Stability	-0.020 (0.002)***	-0.020 (0.002)***	-0.020 (0.002)***	-0.020 (0.002)***
Log SN Uploading * Network Stability	-0.149 (0.010)***	-0.149 (0.010)***	-0.039 (0.003)***	-0.040 (0.003)***
Log Outdegree	0.003 (0.001)***	0.003 (0.001)***	0.003 (0.001)***	0.003 (0.001)***
Billing-based Time and Location Effects	0.029 (0.004)***		0.029 (0.004)***	
Calling-based Time and Location Effects		0.074 (0.004)***		0.074 (0.004)***
Selection Correction Term	0.020 (0.010)**	0.020 (0.010)**	0.020 (0.010)**	0.020 (0.010)**
Chi-square	3707.27	4070.85	3671.49	4032.77

Notes: The dependent variables are log-transformed. SN refers to social network. Estimates for individual-specific dummies, time dummies and indicators for missing network stability metrics are omitted. *** denotes significant at 0.01, ** denotes significant at 0.05.

Why is this happening? In our context, contagion most likely operates through awareness. Given the simple correlation result, individuals with networks that are stable also have networks that have high network closure. Within such clustered social networks, members are likely to receive more redundant information from their peers (Burt 2000). This redundancy in information, in turn, makes them less likely to be susceptible to their neighbors' actions. We find there are statistically significant estimates for our control variables like time-period dummies, mean uploading and downloading frequency of all other users in billing- or call-based zip code areas and missing indicators for network stability measures. We also find the relationship between the outdegree of a user and his/her content usage and generation behavior is positive and statistically significant. Finally, the estimates for selection correction terms are positive and statistically significant for both equations, indicating that controlling for sample selection bias is crucial in our setting. [2]

---

[2] When users are more active on their mobile phones, they upload or/and download more frequently and they communicate with more people and more different people in successive time intervals. To control for the extent of activity of a user, we implemented all models with the number of calls made (or number of SMS/MMS sent). In the main results, we removed the outdegree of a user because it is positively

Robustness Checks

Our results are quite robust to various changes in model structure. We present the results of several robustness checks using unnormalized social network weights, lagged dependent variable, count data estimation, a composite activity metric, and an alternative social contagion variable.

Table 5: Results on Alternative Social Contagion Variable using Voice Network Data

Independent Variable	Estimate			
	Frequency-Weighted Voice		Duration-Weighted Voice	
	Billing-based Control	Calling-based Control	Billing-based Control	Calling-based Control
Download Equation				
Log SN Downloading	0.116 (0.021)***	0.113 (0.021)***	0.110 (0.021)***	0.112 (0.021)***
Network Stability	-0.153 (0.010)***	-0.166 (0.010)***	-0.153 (0.010)**	-0.166 (0.010)***
Log SN Downloading * Network Stability	-0.165 (0.074)**	-0.130 (0.073)*	-0.159 (0.074)**	-0.128 (0.072)*
Log Outdegree (Voice)	0.051 (0.010)***	0.049 (0.010)***	0.051 (0.010)***	0.049 (0.010)***
Billing-based Time and Location Effects	0.066 (0.002)***		0.066 (0.002)***	
Calling-based Time and Location Effects		0.185 (0.002)***		0.185 (0.002)***
Selection Correction Term	0.120 (0.041)***	0.120 (0.041)***	0.121 (0.041)***	0.121 (0.041)***
Chi-square	35284.34	41694.86	35281.21	41694.19
Upload Equation				
Log SN Uploading	0.435 (0.013)***	0.429 (0.013)***	0.414 (0.013)***	0.408 (0.013)***
Network Stability	-0.022 (0.002)***	-0.023 (0.002)***	-0.023 (0.002)***	-0.023 (0.002)***
Log SN Uploading * Network Stability	-0.186 (0.047)***	-0.185 (0.047)***	-0.121 (0.017)***	-0.120 (0.017)***
Log Outdegree (Voice)	0.004 (0.002)**	0.004 (0.002)**	0.004 (0.002)**	0.004 (0.002)**
Billing-based Time and Location Effects	0.025 (0.004)***		0.026 (0.004)***	
Calling-based Time and Location Effects		0.070 (0.004)***		0.070 (0.004)***
Selection Correction Term	0.020 (0.010)**	0.020 (0.010)**	0.020 (0.010)**	0.020 (0.010)**
Chi-square	2481.42	2822.17	2389.17	2731.61

Notes: We consider contagion variables not based on volume of activity but whether there was activity or not. The dependent variables are log-transformed frequency of content downloading and uploading, respectively. SN refers to social network. Estimates for individual-specific dummies, time dummies and indicators for missing network stability metrics are omitted. *** significant at 0.01, ** significant at 0.05, * significant at 0.1.

correlated with the extent of activity variable. We find qualitatively the same result with and without the extent of activity. Details are available upon request.

Table 6: Dynamic Panel Data Model Estimation Result

Independent Variable	Estimate			
	Frequency-Weighted Voice		Duration-Weighted Voice	
	Billing-based Control	Calling-based Control	Billing-based Control	Calling-based Control
Download Equation				
Lagged Log Downloading	0.126 (0.006)***	0.122 (0.006)***	0.126 (0.006)***	0.122 (0.006)***
Log SN Downloading	0.089 (0.007)***	0.088 (0.007)***	0.089 (0.007)***	0.088 (0.007)***
Network Stability	-0.004 (0.001)***	-0.004 (0.001)***	-0.004 (0.001)***	-0.004 (0.001)***
Log SN Downloading * Network Stability	-0.005 (0.002)**	-0.004 (0.002)**	-0.006 (0.002)***	-0.005 (0.012)**
Log Outdegree	0.024 (0.011)**	0.023 (0.011)**	0.024 (0.011)**	0.024 (0.012)***
Billing-based Time and Location Effects	0.017 (0.007)**		0.017 (0.008)***	
Calling-based Time and Location Effects		0.069 (0.009)***		0.070 (0.009)***
Wald Chi-square	1695.43	1759.34	1690.46	1754.38
Upload Equation				
Lagged Log Uploading	0.107 (0.004)***	0.105 (0.004)***	0.107 (0.004)***	0.106 (0.004)***
Log SN Uploading	0.246 (0.013)***	0.245 (0.013)***	0.244 (0.013)***	0.243 (0.013)***
Network Stability	-0.004 (0.002)**	-0.004 (0.002)**	-0.004 (0.002)**	-0.004 (0.002)**
Log SN Uploading * Network Stability	-0.069 (0.022)***	-0.069 (0.022)***	-0.077 (0.022)***	-0.077 (0.022)***
Log Outdegree	0.003 (0.001)***	0.003 (0.001)***	0.003 (0.001)***	0.003 (0.001)***
Billing-based Time and Location Effects	0.058 (0.015)***		0.058 (0.015)***	
Calling-based Time and Location Effects		0.173 (0.019)***		0.174 (0.019)***
Wald Chi-square	1640.46	1818.01	1586.31	1764.00

Notes: The dependent variables are log-transformed frequency of content downloading and uploading, respectively. SN refers to social network. Estimates for individual-specific dummies, time dummies and indicators for missing network stability metrics are omitted. *** significant at 0.01, ** significant at 0.05.

An Unnormalized Social Network Weight Model: The social network weight can be constructed in various ways (e.g., Valente 1995). We used a normalized social network weight for the main model. We define an unnormalized lagged social network weight w_{ijt-1}, simply determined as the number of calls (frequency based) or the duration of calls (duration based) from user *i* to user *j* in week *t*-1. The estimates based on unnormalized social network weights are shown in Table 4. These results are qualitatively similar to the key results from our main model.

The Alternative Social Contagion Variable: We also consider another contagion variable, based on whether fellow users show an activity or not, rather than their volume of activity. Table 5 shows these results. While the key results are qualitatively the same as from our main simultaneous equation model, this model has a slightly worse fit. This result suggests that the quantity of data network neighbors upload or download is more important than just the incidence.

The Dynamic Panel Data Model: We also incorporate dynamics for user content activities. Due to the evidence of positive state dependence from the selection equation results, we conduct tests to check the robustness of the results by estimating the main equations separately with a lagged dependent variable to control for state dependence. To be specific, we specify a linear dynamic panel-data model for each equation (i.e., download and upload) and perform a generalized method of moments (GMM)-based estimation. Similar to our main model estimation, in the GMM-based dynamic panel data model estimation, we use exogenous and predetermined variables as instruments. We test the validity of the instruments using the Sargan Test and find that our instruments are valid. Table 6 shows that the results qualitatively remain the same as in our main equation estimation.

Table 7: Results on Main Model based on SMS and MMS Network Data

Independent Variable	Estimate	
	Frequency-Weighted SMS	Frequency-Weighted MMS
Download Equation		
Log SN Downloading	0.099 (0.006)***	0.248 (0.009)***
Network Stability	-0.581 (0.013)***	-0.616 (0.020)***
Log SN Downloading * Network Stability	-0.066 (0.024)**	-0.012 (0.005)**
Log Outdegree	-0.005 (0.006)	0.017 (0.015)
Billing-based Time and Location Effects	0.011 (0.003)***	0.068 (0.002)***
Selection Correction Term	0.120 (0.041)***	0.121 (0.041)***
Chi-square	34414.06	29909.32
Upload Equation		
Log SN Downloading	0.245 (0.009)***	0.622 (0.012)***
Network Stability	-0.032 (0.003)***	-0.022 (0.004)***
Log SN Downloading * Network Stability	-0.255 (0.032)***	-0.027 (0.015)*
Log Outdegree	-0.002 (0.002)	-0.004 (0.003)
Billing-based Time and Location Effects	0.026 (0.004)***	0.031 (0.004)***
Selection Correction Term	0.021 (0.010)**	0.020 (0.010)**
Chi-square	1629.21	4420.35

Notes: Dependent variables are log-transformed. Estimates for individual-specific dummies, time dummies and indicators for missing network stability metrics are omitted. *** significant at 0.01, ** significant at 0.05.

Other Communication Modes: We examine network stability metrics from other communications network data, such as SMS and MMS network data. If different types of relations are structured by different types of networks, then the multiplexity of relations among individuals may create systematic, important patterns of cross-cutting social circles (McPherson et al. 2001). All three network stability metrics from voice call, SMS, and MMS data are positively correlated. To be specific, the voice-call-based stability metric positively correlate with the SMS-based and MMS-based metrics (corr. = 0.18, p-value <0.001; corr. = 0.10, p-value < 0.001), respectively. The SMS-based and the MMS-based stability metrics also positively correlate (corr. = 0.15, p-value < 0.001). Estimates based on SMS and MMS social network data appear in Table 7. The estimates show that our key results are robust. For example, we find that network stability metrics negatively associate with content usage and generation activities (the coefficient estimates being -0.581 and -0.0616 in the content download equation and based on SMS and MMS data, respectively). Recall that the coefficient estimate in the content download equation based on voice data is -0.076. Hence, our results suggest that, on average, the magnitude of the impact of network stability is higher in SMS or MMS networks than in the voice network. Interestingly, our data also shows that on average SMS or MMS networks are less stable than are voice networks. Hence, these results further emphasize that users with less stable networks have a high intrinsic tendency to engage in content usage and generation in a mobile Internet setting.

Additional Analyses

Additional analysis results also suggest that network stability has an impact on individual behavior. We believe that these effects of network stability are a result of its relationship with network closure. Note, however, that while network stability is positively correlated with network closure, the correlation is low, which suggests that stability may still be an important predictor of individual behavior even after controlling for network closure. We test this here. Table 8 shows the results from a model that includes only the simple effect of network closure. The results show that even after controlling for network closure, stability is still an important predictor of individual behavior. We also extend the model by including the interaction of network closure with contagion. Table 9 shows the results. The results corroborate the importance of stability even after explicitly controlling for closure.

Table 8: Results for Main Model with only Closure based on Voice Call Network Data

Independent Variable	Estimate			
	Frequency-Weighted		Duration-Weighted	
	Billing-based Control	Calling-based Control	Billing-based Control	Calling-based Control
Download Equation				
Log SN Downloading	0.200 (0.004)***	0.202 (0.004)***	0.199 (0.004)***	0.202 (0.004)***
Network Stability	-0.128 (0.012)***	-0.136 (0.012)***	-0.129 (0.012)***	-0.136 (0.012)***
Log SN Downloading * Network Stability	-0.074 (0.017)***	-0.078 (0.017)***	-0.075 (0.017)***	-0.079 (0.017)***
Network Closure	-0.080 (0.023)***	-0.112 (0.022)***	-0.080 (0.022)***	-0.112 (0.022)***
Log Out Degree	0.230 (0.012)***	0.238 (0.012)***	0.233 (0.012)***	0.238 (0.012)***
Billing-based Time and Location Effects	0.011 (0.002)***		0.011 (0.002)***	
Calling-based Time and Location Effects		0.137 (0.003)***		0.136 (0.003)***
Selection Correction Term	0.122 (0.041)***	0.122 (0.041)***	0.122 (0.041)***	0.122 (0.041)***
Chi-square	35540.10	38817.13	35509.72	38786.15
Upload Equation				
Log SN Uploading	0.309 (0.005)***	0.309 (0.005)***	0.302 (0.005)***	0.302 (0.005)***
Network Stability	-0.016 (0.002)***	-0.016 (0.002)***	-0.016 (0.002)***	-0.017 (0.002)***
Log SN Uploading * Network Stability	-0.206 (0.022)***	-0.208 (0.022)***	-0.191 (0.022)***	-0.192 (0.022)***
Network Closure	-0.008 (0.004)**	-0.009 (0.004)**	-0.008 (0.004)**	-0.009 (0.004)**
Log Outdegree Degree	0.004 (0.002)**	0.005 (0.002)**	0.005 (0.002)**	0.005 (0.002)**
Billing-based Time and Location Effects	0.023 (0.004)***		0.023 (0.004)***	
Calling-based Time and Location Effects		0.071 (0.004)***		0.004 (0.0005)***
Selection Correction Term	0.020 (0.010)**	0.020 (0.010)**	0.020 (0.010)**	0.020 (0.010)**
Chi-square	4302.48	4632.94	4116.26	4446.41

Notes: Dependent variables are log-transformed. Estimates for individual-specific dummies, time dummies, and indicators for missing network stability metrics are omitted. *** significant at 0.01, ** significant at 0.05.

Table 9: Results for Main Model with Closure and Contagion based on Voice Network Data

Independent Variable	Estimate			
	Frequency-Weighted		Duration-Weighted	
	Billing-based Control	Calling-based Control	Billing-based Control	Calling-based Control
Download Equation				
Log SN Downloading	0.201 (0.004)***	0.205 (0.004)***	0.200 (0.004)***	0.204 (0.004)***
Network Stability	-0.126 (0.012)***	-0.133 (0.012)***	-0.127 (0.012)***	-0.134 (0.012)***
Log SN Downloading * Network Stability	-0.083 (0.018)***	-0.090 (0.017)***	-0.084 (0.018)***	-0.090 (0.017)***
Network Closure	-0.095 (0.023)***	-0.132 (0.022)***	-0.095 (0.023)***	-0.132 (0.023)***
Log SN Downloading * Network Closure	-0.065 (0.025)***	-0.084 (0.025)***	-0.065 (0.025)***	-0.084 (0.025)***
Log Out Degree	0.230 (0.012)***	0.238 (0.012)***	0.233 (0.012)***	0.238 (0.012)***
Billing-based Time and Location Effects	0.011 (0.002)***		0.011 (0.002)***	
Calling-based Time and Location Effects		0.137 (0.003)***		0.138 (0.003)***
Selection Correction Term	0.122 (0.040)***	0.122 (0.040)***	0.122 (0.041)***	0.122 (0.041)***
Chi-square	35547.42	38831.03	35516.28	38799.07
Upload Equation				
Log SN Uploading	0.309 (0.005)***	0.309 (0.005)***	0.302 (0.005)***	0.302 (0.005)***
Network Stability	-0.016 (0.002)***	-0.016 (0.002)***	-0.016 (0.002)***	-0.017 (0.002)***
Log SN Uploading * Network Stability	-0.207 (0.022)***	-0.209 (0.022)***	-0.191 (0.022)***	-0.193 (0.022)***
Network Closure	-0.008 (0.004)**	-0.009 (0.004)**	-0.008 (0.004)**	-0.009 (0.004)**
Log SN Uploading * Network Closure	-0.006 (0.009)	-0.006 (0.009)	-0.006 (0.009)	-0.007 (0.009)
Log Outdegree Degree	0.004 (0.002)**	0.005 (0.002)**	0.005 (0.002)**	0.005 (0.002)**
Billing-based Time and Location Effects	0.023 (0.004)***		0.023 (0.004)***	
Calling-based Time and Location Effects		0.071 (0.004)***		0.070 (0.004)***
Selection Correction Term	0.020 (0.010)**	0.020 (0.010)**	0.020 (0.010)**	0.020 (0.010)**
Chi-square	4304.10	4634.98	4118.83	4448.39

Notes: The dependent variables are log-transformed frequency of content downloading and uploading, respectively. SN refers to social network. Estimates for individual-specific dummies, time dummies, and indicators for missing network stability metrics are omitted. *** significant at 0.01, ** significant at 0.05, * significant at 0.1.

Discussion and Implications

We conducted a detailed study to understand the impact of network stability on individual usage of mobile Internet. In contrast to other studies that typically assume that ties among individuals are static or unchanging (e.g., Coleman et al. 1966), we find evidence that network dynamics are an important predictor of individual behavior. In particular, we find three main results: (1) users with high network stability have a low intrinsic tendency to engage in content usage and generation in the mobile internet setting; (2) the extent of positive social contagion effect is mitigated for those users with high network stability and (3) stability impacts individual behavior even after controlling for closure. Collectively, these results indicate that users within a more constrained/clustered social network are more likely to receive redundant information from each other, hence, this will less likely to lead to individual behavior.

The insights from this study have managerial implications. For example, the mitigating effect of network stability on social contagion can provide companies with further insight into how to influence user behavior in an interaction-intensive setting. To be specific, this result implies that less stable ties intensify the social contagion effect, whereas stable ties mitigate it. This finding complements past research, which suggests that quantity of influence or exposure is important in terms of network-based influence (Iyengar et al. 2011a). Our results suggest that it is important to consider the quality of influence (information or influence from non-redundant sources). Our research adds to the past work on tie strength (strong vs. weak ties) and its role in the diffusion of information within a network (Godes and Mayzlin 2009).

Our research also generates a number of questions for future research. An interesting issue is to systematically investigate the effect of network stability, network closure and redundancy of information on consumer behavior in different contexts. For example, in contexts with reasonable perceived risk, redundancy in the information from contacts may actually be desirable, since it is both reassuring and reaffirms one's beliefs (e.g., Iyengar et al. 2011a). In certain instances, multiple doses of influence may actually be necessary to change consumer behavior. For instance, Centola and Macy (2007) describe "complex contagions" as those that require multiple doses of influence and contrast it with "simple contagions" where a single source of information is enough for information to spread (e.g., rumors). It is plausible that more risky contexts are associated with complex contagions. In the context of less risky user decisions (e.g., in our mobile Internet setting), however, there is less need for redundant information. For example, arguably there is less risk in downloading or uploading information on a mobile phone, as opposed to investing in stocks based on new information from others.

Another question regards the timeliness of information and how that aspect relates to the redundancy of contacts. The timeliness of information is important in a mobile Internet setting, as people like to obtain timely information using their mobile phones. No one wants to watch a viral video two weeks after it became popular. For instance, Buskens and Yamaguchi (1999) find that non-redundancy increases the speed at which an ego acquires information. A third question is to evaluate the kind of information being transferred with what type of communication channel (e.g., voice, SMS, MMS) and how that information impacts social contagion. For instance, Berger and Iyengar (2011) find that the drivers of word-of-mouth are different in online versus offline channels. While we did not have information about the specific type of content uploaded or downloaded, as more researchers do gain access to this type of data (e.g., Aral and Van Alstyne 2011), these questions can be investigated. A fourth question is to identify social contagion using more direct methods. For instance, Bramoullé et al. (2009) provide an identification strategy of peer effects through social networks. They use characteristics of network structure to identify contagion and influence. Also Aral et al. (2009) use dynamic matched sample estimation. Finally, we found that network stability in a significant predictor for individual behavior even after controlling for network closure. We speculate that this is the case as instability leads to uncertainty among individuals about their future interactions. Future studies can corroborate our hypothesis and also investigate other mechanisms at work. We hope that our study will generate increased interest in the emerging literature on social contagion and more broadly, on how network structure plays a key role in moderating individual behavior.

References

Allcott, H., D. Karlan, M. Möbius, T. Rosenblat, A. Szeidl. 2007. Community size and network closure. *American Economic Review*, **97**(2), 80-85.

Aral, S., M. Van Alstyne. 2011. Networks, information and brokerage: The diversity-bandwidth tradeoff. *American Journal of Sociology*, In Press.

Aral, S., D. Walker, D. 2011. Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Science*, In Press.

Aral, S., L. Muchnik, A. Sundararajan. 2009. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of National Academy of Science*, **106**(51): 21544-21549.

Arellano, M. 1987. Computing robust standard errors for within-groups estimators. *Oxford Bulletin of Economics and Statistics*, **49**(4), 431-34.

Bateman, P., P. Gray, B. Butler. 2010. The impact of community commitment on participation in online communities. *Information Systems Research*, Forthcoming.

Berger, J., R. Iyengar. 2011. How interest shapes word-of-mouth over different channels, Working Paper, University of Pennsylvania.

Bien, W., J. Marbach, F. Neyer. 1991. Using ego-centered networks in survey research. A methodological preview on an application of social network analysis in the area of family research. *Social Networks*, **13**, 75-90.

Borgatti, S.P., P.C. Foster. 2003. The network paradigm in organizational research: A review and typology. *Journal of Management,* **29**(6), 991–1013.

Bramoullé, Y., H. Djebbari, B. Fortin. 2009. Identification of peer effects through social networks. *Journal of Econometrics*, **150**(1), 41-55.

Broese van Groneau, M., E. van Sonderen, J. Omel. 1990. Test-retest reliability of personal network delineation, in Antonucci, T., Knipscheer, C. (Eds), *Social Network Research: Methodological Questions and Substantive Issues*. Swets & Zeitlinger, Rockland, Amsterdam.

Burt, R. 1983. Range. In Burt, R. and Minor, M. (Eds), *Applied Network Analysis*, CA: Sage, 176–194.

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

Burt, R. 2000. Structural holes versus network closure as social capital, in Lin, N., Cook, N. and Burt, R. (Eds), *Social Capital: Theory and Research*, Aldine De Gruyter, New York, NY, 31-56.

Buskens, V., K. Yamaguchi. 1999. A new model for information. *Sociological Methodology,* **29**(1), 281-325.

Centola, D., M. Macy. 2007. Complex contagions and the weakness of long ties. *American Journal of Sociology*, **113**(3), 702-34.

Coleman, J. 1988. Social capital in the creation of human capital. *American Journal of Sociology*, 94, 95-120.

Coleman, J. 1990. Foundations of Social Theory. Cambridge, MA: Harvard Business School Press.

Christakis, N.A., J.H. Fowler. 2010. Contagion in prescribing behavior among networks of doctors. *Marketing Science*, **30**, 2, 213-216.

Dasgupta, K., R. Singh, B. Viswanathan, D. Chakraborty, S. Mukherjea, A. Nanavati, A. Joshi. 2008. Social ties and their relevance to churn in mobile telecom networks. *Proceedings of the 11th International Conference on Extending Database Technology Conference*, Nantes, France.

Doreian, P., F. Stokeman. 1997. *Evolution of Social Networks*. Overseas Publishers Association. Amsterdam.

Faraj, S., S. Johnson. 2010. Network exchange patterns in online communities. *Organization Science*, In Press.

Ferligoj, J.,V. Hlebec.1999.Evaluations of social network measurement instruments. *Social Networks*,**21**,111-30.

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

Granovetter, M. 1982. The strength of weak ties: A network theory revisited. In Marsden, P. and Lin, N. (Eds), *Social Structure and Network Analyses*, New York: John Wiley and Sons, 105-130.

Ghose, A., S. Han. 2011. An empirical analysis of user content generation and usage behavior on the mobile Internet. *Management Science*, In Press.

Goldenberg, J., L. Barak, M. Eitan, S. Stremerch. 2010. The evolving social network of marketing scholars. *Marketing Science*, **29**(3), 561–567.

Godes, D., D. Mayzlin. 2009, Firm-created word-of-mouth communication: Evidence from a field study. *Marketing Science*, **28**(4), 721-739.

Hartmann, W.R., P. Manchanda, H. Nair, M. Bothner, P. Dodds, D. Godes, K. Hosanagar, C. Tucker. 2008. Modeling social interactions: Identification, empirical methods and policy implications. *Marketing Lett*ers, **19**(3), 287-304.

Heckman, J.J. 1979. Sample selection bias a specification error. *Econometrica*, **47**(1), 153-161.

Hill, S., F. Provost, C. Volinsky. 2006. Network-based marketing: Identifying likely adopters via consumer networks. *Statistical Science*, **21**(2), 256-276.

Huisman, M., T.A.B. Snijders. 2003. Statistical analysis of longitudinal network data with changing composition. *Sociological Methods and Research*, **32**(2), 253-287.

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

Iyengar, R., C.Van den Bulte, J. Choi. 2011b. Distinguishing between multiple mechanisms of social contagion: Social learning versus normative legitimation in new product adoption, Working Paper, U. of Pennsylvania.

Katona, Z., P.P. Zubcsek, M. Sarvary. 2011. Network effects and personal influences: The diffusion of an online social network. *Journal of Marketing Research*, **48**(3), 518-531.

Lancaster, T. 2000. The incidental parameters problem since 1948. *Journal of Econometrics*, **95**(2), 391-414.

Libai, B., E. Muller, R. Peres. 2010. Sources of social value in word-of-mouth programs. Working Paper. Marketing Science Institute.

Lin, N., R.S. Burt. 1975. Differential effects of information channels in the process of innovation diffusion. *Soc. Forces* **54** 256-74.

McPherson, M., L. Smith-Lovin, J.M. Cook. 2001. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, **27**, 415-444.

Nair, H., P. Manchanda, T. Bhatia. 2010. Asymmetric social interactions in physician prescription behavior: The role of opinion leaders. *Journal of Marketing Research*, Forthcoming.

Oestreicher-Singer, G., A. Sundararajan. 2011. The visible hand of social networks in electronic markets. *Management Information Systems Quarterly*, Forthcoming.

Onnela, J.P., J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, A.L. Barabási. 2007. Structure and tie strengths in mobile communication networks. *PNAS*, **104**(18): 7332-7336.

Pearson, M.A., L. Michell. 2000. Smoke rings: Social network analysis of friendship groups, smoking and drug-taking. *Drugs: Education, Prevention and Policy,* **7**(1), 21–37.

Puhani, P. 2000. The Heckman correction for sample selection and its critique. *Journal of Economic Surveys*, **14**(1), 53-68.

Reagans, R.E., E.W. Zuckerman. 2008. Why knowledge does not equal power: the network redundancy trade-off. *Industrial and Corporate Change*, **17**(5), 903-944.

Shalizi, C., A.C. Thomas. 2011. Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods and Research*, **40**(2), 211-239.

Snijders T.A.B., P. Doreian. 2010. Introduction to the special issue on network dynamics. *Social Networks*, **32**(1), 1-3.

Stephen, A., O. Toubia. 2009. Explaining the power-law degree distribution in a social commerce network. *Social Networks*, **31**(4), 262-270.

Susarla, A., J. Oh, Y. Tan. 2010. Social networks and the diffusion of user-generated content: Evidence from YouTube. *Information Systems Research*, Forthcoming.

Trusov, M., R.E. Bucklin, K. Pauwels. 2009. Effects of word-of-mouth versus traditional marketing: Findings from an internet social networking site. *Journal of Marketing*, **73**(5), 90-102.

Tucker, C. 2011. Network stability, network externalities and technology adoption. NET Institute Working Paper No. 08-30. Available at SSRN: http://ssrn.com/abstract=1287437.

Valente, T.W. 1995. *Network Models of the Diffusion of Innovations*. Hampton Press, Cresskill, NJ.

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

Verbeek, M., T. Nijman. 1996. Incomplete panels and selection bias, in Mátyás, L., P. Sevestre (Eds.). *The Econometrics of Panel Data: A Handbook of the Theory with Applications*. Kluwer Academic Publishers.

Wooldridge, J.M. 2002. Econometric Analysis of Cross section and Panel Data, MIT Press, MA.