This is the Pre-Published Version

Uncovering the Importance of Relationship Characteristics in Social Networks: Implications for Seeding Strategies

Xi Chen[1], Ralf van der Lans[2] and Tuan Q. Phan[3]

[1]Rotterdam School of Management
Erasmus University Rotterdam, 3062 PA Rotterdam
chen@rsm.nl
[2]School of Business and Management
Hong Kong University of Science and Technology, Hong Kong SAR
rlans@ust.hk
[3] Department of Information Systems
National University of Singapore, Singapore
tphan@comp.nus.edu.sg

May 17, 2016

Accepted for publication in the Journal of Marketing Research

Acknowledgements:
The authors thank Xueming Luo for his valuable suggestions and for his help in obtaining data for this project. The second author gratefully acknowledges financial support from the Hong Kong Research Grant Council (GRF645511), and the third author gratefully acknowledges support from the National University of Singapore Research Grant R-253-000-110-112.

Uncovering the Importance of Relationship Characteristics in Social Networks: Implications for Seeding Strategies

ABSTRACT

Seeding influential social network members is crucial for the success of a viral marketing campaign and product diffusion. Based on the assumption that connections between customers in social networks are binary (either present or absent), previous research mostly recommended seeding network members that are well-connected. However, the importance of connections between customers varies substantially depending on the characteristics of the relationships, such as its type (i.e., friend, colleague, and acquaintance), as well as its duration and interaction intensity. This research introduces a new Bayesian methodology to identify influential network members and takes into account the relative influence of different relationship characteristics on product diffusion. Two applications of the proposed methodology – the launch of a microfinance program across 43 Indian villages and information propagation in a large online social network – demonstrate the importance of weighting connections in social networks. Compared with traditional seeding strategies, the proposed methodology recommends substantially different sets of seeds that increased the reach by up to 10 percent in the first empirical application and up to 92 percent in the second.

Keywords: product diffusion; social network analysis; seeding strategy; Bayesian method; multigraph network

INTRODUCTION

An individual's decision to adopt a new product or service often depends on behaviors of friends, colleagues and acquaintances. As a consequence, the structure of a social network and the position of influencers in a network are important drivers of the spread of information and the diffusion of products (Van den Bulte and Wuyts 2007). To capitalize on these social effects, companies such as Philips, HP and Microsoft have adopted seeding strategies that target influential customers in social networks to launch new products (Libai et al. 2013). However, due to complexity of social networks, determining which customers are most influential is a non-trivial question that has triggered much theoretical interest (Godes and Mayzlin 2009; Goldenberg et al. 2006; Iyengar et al. 2011). Recently, due to the increasing availability of social network data, this question has received a growing amount of attention by both practitioners and academics. These efforts have led to a more detailed understanding of how networks stimulate diffusion and which individuals possess influential positions (Goldenberg et al. 2009; Hinz et al. 2011; Hu and Van den Bulte 2014; Iyengar et al. 2011; Katona et al. 2011; Tucker 2008; Yoganarasimhan 2012). For instance, Goldenberg et al. (2009) and Yoganarasimhan (2012) demonstrated that the number of connections of an individual has a positive effect on the diffusion process. Tucker (2008) found that, in addition to the number of connections, the effect of betweenness (how many times an individual lies on the shortest path between two members of the social network) on adoption may be even stronger. Based on this, previous research mostly recommended using network metrics to pinpoint influential customers in a social network (Hinz et al. 2011; Iyengar et al. 2011).

Although previous studies have compared the effects of many different network measures on diffusion, they mostly assumed a binary network structure with connections that are either absent or present (0 vs. 1). This is a strong assumption, given ample evidence in marketing and sociology that consumers are connected through relationships with different characteristics of varying importance (Ansari et al. 2011; Brown and Reingen 1987; De Bruyn and Lilien 2008; Granovetter 1973; Trusov et al. 2010). Furthermore, relationships often vary in their importance depending on the type of product in the diffusion process (Schulze et al. 2014). Consequently, researchers do not observe the importance of relationship characteristics in social networks and uncovering it is a primary question in marketing and sociology (Aral and Walker 2014). Dover et al. (2012) recognized the limitation of assuming 0-1 connections between customers in diffusion processes, and called for future research that incorporates information on the strength of connections. The present study follows up on this call and introduces a novel approach to identify influential customers in a social network. Our proposed multi-network methodology uncovers 1) the relative importance of different characteristics of relationships in a social network, and 2) which individuals in the subsequent weighted social network should be seeded in order to generate the highest impact on the diffusion of products.

We demonstrate the effectiveness of our multi-network methodology in two empirical applications. The first application involves the diffusion of a microfinance program rollout in 43 Indian villages. The second application focuses on information propagation in a large online social network consisting of more than 42,000 users. In both applications, we found that taking into account the importance of relationship characteristics is crucial to identify influential network members. Moreover, our proposed methodology was able to increase the reach of seeding strategies by up to 10 percent in the first empirical application and up to 92 percent in the second.

The rest of the paper is organized as follows. We first discuss social networks and the strength of connections as a function of relationship characteristics in Section 2. In Section 3, we introduce our multi-network methodology, which we illustrate in two empirical applications in Sections 4 and 5. Section 6 concludes with a discussion of our findings, implications and directions for future research.

CONNECTIONS IN SOCIAL NETWORKS

Strength of Connections

Granovetter (1973) argued that connections in networks vary in strength and that this has major implications for the diffusion of information. Although Granovetter categorized connections between people into strong and weak ones, he recognized that the underlying strength of a connection is a continuous variable. As argued by Granovetter, the continuous variable for the strength of a connection is a function of duration, emotional intensity, intimacy, and exchange of services between people. The strength of a connection, thus, not only depends on the characteristics of the relationship, but also the type of information or service exchanged. Several empirical studies in marketing support these observations. For instance, Moschis and Moore (1979) found in a study among young adolescents that friend relationships were most important for the adoption of products where peer acceptance played a role (e.g., sunglasses), while parent relationships were essential for products with a higher perceived risk in terms of price and performance (e.g., hair dryer). Similarly, among sorority members at a university, Reingen et al. (1984) found that social influence on brand choice heavily depended on the type of product and characteristics of the relationship (e.g., roommate, friend, neighbor, or study partner). More recently in online settings, De Bruyn and Lilien (2008) found that emails received from stronger relationships, such as friends and family, are more likely to be opened. Aral and Walker (2014), in a large-scale field experiment on Facebook, found that the recency of a relationship was a strong predictor for friends to adopt a recommended application. In contrast, Godes and Mayzlin (2009) discovered that word of mouth through weak relationships (acquaintances) had stronger impact on sales than that through strong relationships (friends and relatives). Schulze et al. (2014) demonstrated that these effects were moderated by the type of product, and apps for hedonic products are shared more effectively by strong relationships (friends) than weak relationships (strangers).

Despite the evidence that strengths of connections between individuals vary systematically, previous research mostly treated the social network as given and *a priori* determined the strength of connections between individuals (e.g., Banerjee et al. 2013; Goldenberg et al. 2009; Hinz et al. 2011; Iyengar et al. 2011; Katona et al. 2011; Tucker 2008; Yoganarasimhan 2012). Specifically, previous research used the following three approaches to deal with different characteristics of relationships. First, many studies ignored the differential influence of relationship characteristics, and designed one binary network (e.g., Banerjee et al. 2013; Hinz et al. 2011; Katona et al. 2011; Tucker 2008; Yoganarasimhan 2012). Second, other studies recognized the limitations of one binary network and related multiple binary networks, each corresponding to a different relationship characteristic, to diffusion (e.g., Aral and Walker 2012; Hu and Van den Bulte 2014). Third, some studies determined the importance of different relationship characteristics *a priori*, and subsequently integrated them to obtain a weighted network (e.g., Newman 2001; Rothenberg et al. 1995). In a weighted network, a connection between individuals receives a positive value corresponding to its strength (see Van den Bulte and Wuyts 2007). Finally, others used a combination of these methods to test the robustness of their results against these different approaches (e.g., Iyengar et al. 2011; Van den Bulte and Lilien 2001). All three approaches determine the importance of relationship characteristics *a priori,* despite the fact that their importance varies. Such treatment could lead to misspecified networks, which in turn may lead to biased estimates of social influence and a suboptimal selection of seeds, as discussed by both sociologists and econometricians (Leenders 2002; LeSage and Pace 2014; Páez et al. 2008). Next we discuss how weighted networks may capture diffusion in complex social networks.

Social Relationships in Weighted Networks

A natural way to capture strengths of connections between individuals is to represent them in a weighted network. In a weighted network, each connection between two individuals receives a continuous value representing its strength (Van den Bulte and Wuyts 2007). High positive values correspond to strong connections between individuals, and low positive values to weak. Although weighted networks are more difficult to analyze, several researchers have developed network measures that extend their dichotomous network counterparts to weighted networks (e.g., Bonacich and Lloyd 2001; Newman 2004; Opsahl et al. 2010).

The analysis of weighted networks has received much attention in the literature, especially for neural networks, transportation networks, and food webs in biology (e.g., Luczkovich et al. 2003; Opsahl et al. 2008; Watts and Strogatz 1998). In such networks, assigning weights to connections is relatively straightforward as strengths between connections are observed. For instance, in a food web, weights are characterized by carbon flows between species, while in transportation networks travel time or the number of vehicles commuting between locations can be used to evaluate the weight of a connection. Assigning weights to connections in social networks, however, is more difficult. Strengths of connections are not observed and depend on the characteristics of underlying relationships and the information exchanged. Some attempts have been made, though. For instance, to study the role of social network structure on disease transmission, Rothenberg et al. (1995) weighted connections based on their riskiness, with weights of .5 for injectable drug use, .3 for sexual contacts, and .1 for non-injectable drug use. Similarly, Newman (2001) weighted the connections in a social network of co-authors with the inverse of the number of authors on a paper. Recently in a marketing context, Ansari et al. (2011) assigned downloads between artists as the weight for connections between artists in an online social network. Iyengar et al. (2011), in a study on the prescription behavior of physicians, used the number of different types of relationships (discussion or referral) between physicians as the weight in the “total” network.

An important commonality of these empirical applications is that weights are assigned *a priori*. Some researchers recognized this problem and compared different weights based on statistical fit (Leenders 2002; Páez et al. 2008), or robustness of results (Iyengar et al. 2011; Van den Bulte and Lilien 2001). A disadvantage of these approaches is that they are only able to compare a limited number of possible pre-determined weights. In reality, this number is infinite as the weight of a connection is a continuous variable. As discussed above, the weight of a connection not only depends on the characteristics of the relationships it consists of, but also on the type of information exchanged. Thus, ideally, weights of connections should be inferred from the actual diffusion taking place on the network, rather than assigned *a priori* by the researcher. To the best of our knowledge, such a methodology does not exist. The only research in marketing that estimates weights in social networks is the study by Trusov et al. (2010). In an online social network, they estimated how susceptible an individual was towards social influence and which connections influenced an individual’s login behavior. However, their model was developed specifically to infer influence and susceptibility from repeated login decisions in an online social network and focused on egocentric networks only. Hence, their approach cannot be applied to seeding decisions in diffusion processes in which consumers make only one adoption decision. In the next section, we introduce a new approach that infers weights of connections based on relationship characteristics and the actual diffusion process.

THE MULTI-NETWORK APPROACH FOR SEEDING DECISIONS

Consider the diffusion of a product or service on a social network consisting of *N* individuals that are connected through *K* different networks, each representing a different relationship characteristic. For instance, a network could represent only friend relationships, while another network could represent the duration of relationships. A network $k \in K$ is represented by an $N \times N$ adjacency matrix A_k with elements a_{ij}^k representing the relationship characteristic *k* between sender *i* and receiver *j*. These elements could be dichotomous (0 vs. 1), indicating whether a relationship characteristic is absent or present, or continuous representing a numeric property of a relationship *k*, such as its duration. Furthermore, we permit $a_{ij}^k \neq a_{ji}^k$ to allow both directed and undirected relationship characteristics. For instance, $a_{ij}^k = 1$ and $a_{ji}^k = 0$ in a network of advisors and students in which individual *i* is an advisor and individual *j* is a student and *k* represents advisor-advisee relationship.

To initiate the diffusion process, we assume that *S* individuals on the social network are seeded by the company, with *S* a subset of *N*. For each seed $s \in S$, we count the number of individuals y_s (which may include seed *s* itself) that adopt a product due to an information cascade that is initiated by seed *s*. Thus, y_s represents the reach of seed *s*, and $\sum_{s=1}^{S} y_s$ the total number of adopters in the diffusion process. Note that the definition of y_s could be more general, and could include a time index *t* to indicate the number of adopters in period *t* due to seed *s*. Moreover, the number of seeds *s* may also change over time to reflect situations in which a company seeds additional individuals at different moments in time (van der Lans et al. 2010). The goal of our multi-network methodology is to construct a weighted social network that optimally links the weighted network measures of seed *s* to the observed reach y_s . To do so, our approach consists of three components. The first component creates the weighted network, which is a function of multiple networks representing different relationship characteristics. The second component computes network measures of each seed *s* from the constructed weighted network. Finally, the third component relates these derived network measures, in combination with possible other control variables, to the obtained reach of each seed. Figure 1 summarizes the three components of our multi-network procedure and how they are related schematically. In the following subsections we will discuss the specification of the three model components, our Bayesian estimation strategy and model identification.

---- Insert Figure 1 about here ----

Constructing the Weighted Social Network

Given the *K* adjacency matrices A_k , we define the weighted network construction function $g(.)$, which generates $(N \times N)$ weighted adjacency matrix *W* with elements w_{ij} as below:

(1) $$w_{ij} = \begin{cases} 0 & \text{if } a_{ij}^k = 0, \forall k = 1,..,K \\ g\left(a_{ij};\beta\right) & \text{otherwise} \end{cases}.$$

In (1), $w_{ij} \in [0,\infty)$ represents the weight of the connection between individuals *i* and *j* given the diffusion process. Vector $a_{ij} = \left[a_{ij}^1, a_{ij}^2, .., a_{ij}^K \right]$ represents the set of all *K* relationship characteristics between *i* and *j*, and β is a vector of parameters. Function $g(\cdot)$ is allowed to represent any (non)linear function, as long as the outcome is nonnegative. In this research, similar to Ansari et al. (2011), we choose an exponential function for $g(\cdot)$, such that $g(a_{ij};\beta) = \exp\left(\beta_0 + \sum_{k=1}^{K} \beta_k a_{ij}^k \right)$ is nonnegative[1]. If none of the *K* relationship characteristics is present between individuals *i* and *j* (i.e., $a_{ij}^k = 0, \forall k = 1,..,K$), we set the corresponding weight w_{ij} equal to zero.

Extracting Network Measures

Given the weighted social network *W*, the researcher needs to determine for each seed *s* which network measures x_s are important for diffusion. As discussed earlier, previous researches in marketing proposed several network measures that are important for diffusion, such as degree centrality and betweenness centrality. Although these measures are all derived from unweighted networks, previous research has developed corresponding measures for weighted networks (e.g., Bonacich and Lloyd 2001; Newman 2004; Opsahl et al. 2010). To derive network measures x_s from weighted social network *W*, we introduce the deterministic function $h(.)$ that is defined as below:

(2) $$x_s = h(W).$$

In (2), x_s could contain *J* network measures that are expected to explain the diffusion process y_s. Examples of such network measures are weighted degree centrality, eigenvector centrality and

---

[1] In the robustness checks in our empirical applications, we also tested different specifications of $g(\cdot)$.

weighted betweenness centrality.

Relating Network Measures to Diffusion

In the final component of our proposed methodology, we relate network measures *x*, in addition to possible control variables *z*, to the $S \times 1$ vector *y* summarizing the reach of the diffusion process. As discussed above, outcome variable y_s may represent for each seed $s = 1,..,S$ and time period $t = 1,..,T$ the number of adopters reached by *s*. In that case, *y* equals a $(S \times T)$ matrix with element y_{st} representing the reach of seed *s* at time *t*. Note that the outcome variable can be adapted to incorporate diffusions of multiple products or diffusions on multiple social networks.

To identify how the weighted network measures of seed *s* predict the number of adopters, we propose a diffusion equation $f(.)$ that relates both control variables *z* and network measures *x* to reach y:

(3) $$y = f(x, z; \theta).$$

In (3), θ represents model parameters that relate weighted network measures *x* and control variables *z* to diffusion. Note that function (3) is flexible and may represent any (non-) linear relationship between reach *y* and network measures *x* and control variables *z.* For instance, in our first empirical application we chose a linear specification for $f(.)$ and in our second we used a generalized linear model with Poisson link function.

Model Estimation and Identification

Combining equations (1) to (3) results in the following model likelihood:

(4) $$y = f\left(h\left(g\left(A; \beta\right)\right), z; \theta\right),$$

which is, due to the possible nonlinearity of $g(.)$ and $h(.)$, highly nonlinear. However, as explained above and illustrated in Figure 1, the model parameters β and θ can be naturally decomposed in separate blocks. Conditional on β, estimation of θ is relatively straightforward depending only on the diffusion equation (3). The Bayesian estimation framework, which estimates model parameters iteratively conditional on the value of other parameters, is therefore a natural tool to estimate the model parameters. Conditional on β, one can use standard Bayesian estimation procedures, such as the Gibbs sampler, to draw θ. Furthermore, conditional on θ, a Metropolis-Hastings step can be used to draw β as the posterior distribution usually has an unfamiliar form. In sum, our proposed multi-network methodology for diffusion processes is flexible and optimally integrates multiple networks of relationship characteristics into a weighted network. Our approach allows identifying which seeds are most influential given their weighted network position and possible other control variables.

To apply the Bayesian estimation procedure, we need to make sure that the model is identified. Identification of θ is similar to a standard model without estimation of β, and identification thus depends on the diffusion equation (3) given β. However, identification of β is less straightforward as network measures computed by $h(.)$ in (2) are oftentimes relative measures, such that $h(W)=h(\alpha W)$, with $\alpha > 0$. To control for this scaling issue, we need to fix one of the elements of parameter vector β for identification. For instance, if $g(a_{ij};\beta)=\exp\left(\beta_0+\sum_{k=1}^{K}\beta_k a_{ij}^k\right)$, it is sufficient to set the constant to zero ($\beta_0 = 0$).

Multi-network Approach for Seeding Decisions

Seeding decisions are executed at the start of the diffusion process. To be able to apply the multi-network approach for optimal seed selection, marketers need to know the parameter estimates β. There are three ways to obtain these estimates at the launch of a campaign. First, it is possible to estimate the parameters on a small set of seeds, and use this information to optimally select the remaining set of seeds. This approach is popular in viral marketing campaigns, in which companies test a few seeding strategies on a smaller set of customers before starting the actual launch of the campaign (van der Lans et al. 2010). Our second empirical application adopts a holdout seeding procedure following this approach. Second, if a company has information on the diffusion process of comparable products on the same social network, it is possible to use this information to determine the importance of relationship characteristics. Such approach is similar to the "guessing by analogy" approach that is popular to predict sales before the product launch (Jiang et al. 2006; Lilien et al. 1981). Finally, marketers could also obtain information about the importance of relationship characteristics by sequentially launching the products in different, but comparable social networks. Such sequential market rollout strategy is common practice as it reduces the risk of a new product launch (Bronnenberg and Mela 2004; Kalish et al. 1995), and is also applied by the microfinance company in our first empirical application that we describe next.

EMPIRICAL APPLICATION I:
SEEDING A MICROFINANCE DIFFUSION PROGRAM IN INDIAN VILLAGES

We applied our multi-network methodology to a dataset depicting the diffusion of a program launched by a microfinance institution called Bharatha Swamukti Samsthe (BSS) across 43 Indian villages (Banerjee et al. 2013; Jackson et al. 2012). BSS is located in Bangalore (southwest India) and provides microcredit to households in small villages in southern India. To promote their microfinance program, BSS marketing strategy depended on word of mouth and seeded in each village an initial group of leaders, such as teachers, shop keepers, priests and social workers. These seeds were stimulated to spread the microfinance program among their contacts, and subsequently these contacts could influence their contacts. Because seeds were not selected based on their position in the network, this situation is a suitable setting to study the effect of network positions of seeds on total diffusion. Banerjee et al. (2013) investigated this research question and collected social network data in 43 Indian villages, as well as the final reach of the diffusion of the program. Using aggregate diffusion patterns and average network positions of seeds in each village, they found that adoption was higher for villages that were seeded with leaders that had on average a higher eigenvector centrality than degree centrality. However, similar to other researches, Banerjee et al. (2013) assumed a binary network in which all connections between households had the same strength.

The goal of this empirical application is threefold. First, we want to test whether different relationship characteristics affect the reach of the diffusion process differently. Second, we want to examine whether degree or eigenvector centrality captures social influence better, if we take into account the importance of relationship characteristics. Third, using a holdout sample of villages, we want to benchmark optimal seeding strategies based on our approach with traditional seeding strategies based on binary networks. Next, we describe the dataset, the definitions of model equations (1) to (3), the estimation results and implications for seeding strategies.

Data Description

To study the effects of the network position of seeds on the total reach of the diffusion process, Banerjee et al. (2013) collected relationship characteristics between households in 75 villages, six months before BSS launched their program. Because these villages are relatively isolated with a median distance of 46 km, there were no connections between households from different villages, which resulted in 75 independent social networks, each corresponding to a village. Because of operational difficulties, BSS finally launched the microfinance program in 43 of these villages. For each of these 43 villages BSS provided diffusion data, which described for each household whether or not the microfinance program was adopted.

Table 1 provides descriptive statistics of the social networks that we used to validate our methodology. On average the 43 villages contained 223.2 (SD = 56.2) households, of which on average 26.9 (SD = 9.2) were classified as seeds (in total, BSS seeded 1,157 households). Diffusion, as measured by the percentage of households that participated in the program, varied substantially across villages (average percentage: 18.5 percent, SD = 8.4 percent). Social relationships between households and their characteristics were collected using surveys, which is a common procedure in marketing as well as sociology to construct social networks (Van den Bulte and Wuyts 2007). In total, twelve types of relationships between households were collected that revealed considerable overlap due to their similarity. To identify the underlying relationship dimensions, we used categorical factor analysis on the tetrachoric correlation matrix between the twelve types of relationships (Parry and McArdle 1991). Based on Kaiser's MSA, two types of relationships were classified as independent (i.e., "are related to" MSA = .28 and "go to temple with" MSA = .13). The remaining ten types of relationships loaded on two factors with eigenvalues larger than one. Based on these results we categorized the twelve measured relationship characteristics in four underlying dimensions: 1) economic relationships (borrow money from, lend money to, borrow kerosene or rice from, lend kerosene or rice to), 2) social relationships (give advice to, help with a decision, obtain medical advice from, engage socially with, invite to one's home, visit in another's home), 3) religious relationships (go to temple with), and 4) family relationships (are related to). We thus obtained for each village four networks corresponding to the four types of relationships between households in a village. The elements of these adjacency matrices were either zero (no relationship) or one (a relationship is present). We coded a relationship as present between two households, if a household mentioned the other household in one of the survey questions. This is similar to Banerjee et al. (2013), who constructed only one adjacency matrix (we call it the Total Network in Table 1) with a relationship present if any of the twelve types of relationships existed. Table 1 presents for each adjacency matrix summary statistics of the two network measures that Banerjee et al. (2013) used in their study (i.e., degree centrality and eigenvector centrality). Next, we describe how we constructed for each seed *s* in each village *v* its obtained reach y_{vs}, weighted network measures x_{vs}, and control variables z_{vs}.

---- Insert Table 1 about here ----

The Reach of a Seed s in Village v (y_{vs}). We observed for each household in village *v* whether or not this household adopted the microfinance program. However, because we did not observe word of mouth communication in the network, we did not observe through which seed(s) a household was informed about the microfinance program. We assigned an adoption to seed *s* in village *v* if this seed had the shortest path to the adopted household in the unweighted (binary) social network of village *v*. If there were *m* seeds nearest to an adopter, we distributed this adopter equally among these seeds (i.e., we increased the reach of these seeds by $1/m$). This approach is similar to Van den Bulte and Lilien (2001) and Iyengar et al. (2011), who attributed adoptions of drugs by physicians equally across their adopted neighbors[2]. Table 2 reports summary statistics of the seeds' reach in our database. On average, each seed generated 1.45 adoptions (SD = 1.60).

[2] We also tried different rules to construct seeds' reach and found consistent estimation results (see the robustness checks in Web Appendix A.1 and A.2).

Network Measures (x_{vs}). For each seed *s* in village *v*, we computed two network measures: weighted degree centrality and eigenvector centrality. Weighted degree centrality of seed *s* corresponds to the weighted sum of connections of that seed. Weighted eigenvector centrality not only takes into account the weighted sum of connections of a seed, but also takes into account the centrality of these connections, with central neighbors contributing more to the centrality of a seed *s* (Bonacich and Lloyd 2001). According to Borgatti (2005), degree and eigenvector centrality are network measures at two extremes that are ideally suited to capture influence. Degree centrality assumes an underlying transmission process where only direct connections are involved. On the other hand, eigenvector centrality assumes an underlying transmission process that involves unrestricted walks (i.e., each network member may influence their neighbors, and neighbors may subsequently influence their neighbors etc.). It is an empirical question which of these two network measures best describes the underlying diffusion process. The formulae for these weighted centrality measures are specified as below:

(5) $$x_{vs1} = \sum_{j=1}^{N_v} w_{vsj} \text{ (Weighted Degree Centrality) },$$

(6) $$x_{vs2} = \left[x_{v2}\right]^{(s)} \text{ (Eigenvector Centrality) },$$

where x_{v2} is the solution to the following system of equations:

$$W_v x_{v2} = \lambda_{\max}^{W_v} x_{v2}.$$

In equation (5), w_{vsj} denotes the *sj*-th entry of weighted adjacency matrix W_v of village *v* and N_v represents the number of households in village v. In equation (6), operator $[.]^{(s)}$ denotes element *s* of a vector, and $\lambda_{\max}^{W_v}$ is the maximum eigenvalue of adjacency matrix W_v. Table 2 reports summary statistics of a seed's (unweighted) degree and eigenvector centrality using binary social networks. The high correlation between degree and eigenvector centrality and the reach of a seed (.58 and .53, respectively) serves as preliminary evidence that the position of a seed in the network affects its reach. We also find that degree and eigenvector centrality measures are highly correlated (r = .93), which prevents us from including them simultaneously in equation (3). In addition to these two network measures, Table 2 also reports summary statistics of the following control variables.

---- Insert Table 2 about here ----

Percentage of Seeds. We constructed this variable by dividing the number of seeds in a village by the total number of households (network size) in that village. To reduce possible collinearity, we mean centered this variable. This variable was used to control for saturation effects, because seeds in villages that contained many other seeds may generate fewer adopters.

Household Characteristics. These variables were constructed from various survey questions concerning different aspects of households. We first constructed four dummy variables to summarize roof materials of different households. If roof material was thatch, then the variable *Roof (Thatch)* equals to one, otherwise zero. The other three roof types were represented similarly: *Roof (Tile)*, *Roof (Stone)* and *Roof (Sheet)*. We constructed *No. of Rooms* from the survey question: how many rooms you have in your house? We also mean centered this variable to control for possible collinearity. The *Electricity* variable is a dummy variable with value one if a household privately owned electricity and zero otherwise. Similarly, we coded the variables *Latrine* and *House*, which equal one if a household privately owned a latrine or a house, respectively, and zero otherwise.

Model Specification and Estimation

To apply our multi-network methodology, we used equations (5) and (6) to extract the weighted network measures x (i.e., step 2 of the multi-network approach, see equation 2). To construct the weighted networks W (step 1 of the multi-network approach, see equation 1), we used the following specification:

(7) $$w_{vij} = \begin{cases} 0 & \text{if } a_{vij}^{k} = 0, \forall k = 1,..,4 \\ \exp\left(\sum_{k=1}^{4} \beta_k a_{vij}^{k}\right) & \text{otherwise} \end{cases}.$$

In equation (7), a_{vij}^{k} corresponds to relationship characteristic k between households i and j in village v, and β_k represents the importance of relationship characteristic k. Finally, we used the following equation for $f(.)$ in equation (3) to relate network measures x and control variables z to the reach of seed s in village v:

(8) $$y_{vs} = \theta_0 + \theta_1' z_{vs} + \theta_2 x_{vsm} + \varepsilon_{vs}\text{, with } \varepsilon_{vs} \sim N\left(0, \sigma^2\right).$$

Equation (8) regresses control variables z_{vs} and network measure x_{vsm} on the reach y_{vs} of seed s in village v, with $m \in \{1, 2\}$ corresponding to weighted degree and eigenvector centrality, respectively. We assumed ε_{vs} follows independent normal distributions with mean zero and variance σ^2.

For model estimation, we assumed diffuse priors. To estimate the parameters in equation (8), we used standard Gibbs steps in our Bayesian framework (see for instance Rossi et al. 2005). To estimate β in equation (7) we used a Metropolis-Hastings step, because the posterior distribution is non-standard. We estimated our model using a total of 20,000 iterations, with a burn-in period of 10,000 iterations, long before the Markov chain converged. Appendix A details our estimation procedure. Application of this procedure to synthetic data shows that our model recovered all parameters well. To determine model fit, we computed the log marginal density (LMD) using the method proposed by Chib and Jeliazkov (2001).

Estimation Results

Table 3 presents the estimation results of four models, two centrality measures (degree vs. eigenvector centrality) by two approaches (traditional vs. multi-network approach). In the traditional approach, we followed previous research and defined a binary adjacency matrix W in equation 7, with $w_{ij} = 1$ if $a_{ij}^{k} = 1$ for any $k = 1,..,4$, and $w_{ij} = 0$ otherwise. For both approaches, Model 1 uses degree centrality (x_{vs1}) as network measure to explain the reach of a seed, while Model 2 uses eigenvector centrality (x_{vs2}). As expected, for each of the four models, network centrality measures positively relate to the reach of a seed. The traditional approach suggests that eigenvector centrality is a stronger explanatory network measure for the reach of a seed than degree centrality (LMD Model 2: -1,642 vs. -1,647 for Model 1)[3]. In contrast, our multi-network approach shows that weighted degree centrality fits better than eigenvector centrality (LMD Model 1: -1,627 vs. -1,641 for Model 2). Moreover, these LMD measures also demonstrate that taking into account the importance of relationship characteristics improves model fit.

To interpret the importance of relationship characteristics, we focus on Model 1 of the multi-network approach, because this model best fits the data. Parameter estimates of the different types of relationships suggest that social relationships are the most important driver of the adoption of the microfinance program (β=.24, 97.9% of posterior draws are positive). Interestingly, economic relationships (β= -.23, 98.3% of posterior draws are negative) tend to be the least important for the adoption of the microfinance program. This result is in line with previous research that information received from stronger (social) relationships is more

[3] Note that Banerjee et al. (2013) estimated the model at the aggregate diffusion level within each village, and thus does not take into account heterogeneity across seeds.

influential than that from weaker (economic) relationships (Brown and Reingen 1987; De Bruyn and Lilien 2008).

Finally, the parameter estimates of the control variables are stable across all four models (see Table 3). As expected, the percentage of households that is seeded in a village has a negative effect on the total reach of a seed (θ= -11.74, all posterior draws are negative). In addition to the percentage of seeded households, the only other significant control variable is whether a household privately owns electricity (θ= -.35, all posterior draws are negative).

In sum, our estimation results demonstrate that 1) different relationship characteristics indeed have different influence on the diffusion process, and 2) ignoring the importance of different relationship characteristics may lead to a different conclusion of which network measure better captures social influence. Next, we demonstrate how the multi-network approach results in better seeding strategies than the traditional approach.

---- Insert Table 3 about here ----

Implications for Seed Selection

To validate whether the multi-network approach is a valuable tool for seeding decisions, we performed holdout seeding practices using the following procedure. First, we randomly split the data into four disjoint subsets of villages with approximately 10-11 villages in one subsample. Then we re-estimated the models on each subsample and used the remaining villages as holdout sample to determine forecasting accuracy and seeding performance. We computed in- and out-of-sample fit statistics, performed seeding analysis based on each of the four subsamples and averaged the results. Table 4 presents the results of the in- and out-of-sample fit statistics. Corroborating the results based on LMD (see Table 3), the proposed multi-network approach systematically outperforms the traditional approach on all fit statistics.

To illustrate whether the multi-network approach is a valuable tool for reseeding practices, we estimated our model on four subsamples of 290 seeds and ranked the remaining 867 seeds in holdout samples based on their predicted reach. Table 5 presents the average actual reach across the four holdout subsamples of different reseeding strategies. Subsequently, we computed the actual reach based on the traditional as well as the multi-network approach if BSS decided to seed the best n out of 867 seeds in the holdout sample, with $n = 1,..,867$. Table 5 presents the results and shows that the multi-network approach obtains a higher reach for both network measures. The multi-network approach does especially well if a relatively small proportion of seeds is selected, which is important as the goal of a seeding strategy is to select a relatively small number of customers for targeting. For instance, if BSS selects n = 100 seeds in the holdout sample, the multi-network approach with degree centrality (Model 1) obtains a reach that is 10.14 percent higher than its traditional counterpart, while the model with eigenvector centrality increases reach by 10.68 percent (Model 2).

----Insert Tables 4 and 5 about here----

Robustness Checks

To ensure that the estimates of the importance of different relationships and the superior holdout seeding performance of the multi-network approach are robust against different specifications and alternative benchmark models, we performed six robustness checks (see Web Appendices A.1 to A.6). First, we tested an alternative rule to assign adopting households to seeds. Instead of equally dividing an adopter among multiple seeds with shortest distance, we assigned the household to the seed with the largest number of shortest paths to that household[4]. Web Appendix A.1 presents the results of this assignment rule. Similar to our previous findings,

[4] If this rule resulted again in multiple seeds, then we looked at the number of paths with length equal to the shortest path plus one and repeated the process until each household was assigned to one unique seed.

the multi-network approach fits the data better than the traditional approach based on LMD. Moreover, we found that weighted degree centrality better explains the reach of a seed and that social relationships are the most important while economic relationships are the least important, corroborating our previous findings. Both in- and out-of-sample fit statistics also confirm the relative performance of different models as presented in Table A.1.2. Finally, in a holdout sample seeding strategy, the multi-network approach obtains a higher actual reach than the traditional approach as seen from Table A.1.3. One may also argue that the assignment of seeds should be based on the weighted network instead of the original binary one. To account for this, we performed another robustness check that constructs seeds' reach at each iteration of the MCMC sampler based on the draw of the importance of relationship characteristics θ. We then computed the shortest (weighted) path between a seed and an adopter and assigned the adopter to the nearest seed(s). If, in rare occasions, multiple nearest seeds exist, we distributed the adopter evenly among them. Web Appendix A.2 documents this robustness check. Notice that the holdout seeding practice is different as seeds' reach is not fixed, but computed based on the importance of relationship characteristics. Despite these differences, our results are stable and the multi-network approach outperforms the traditional approach (see Table A.2.2 and Table A.2.3).

Second, we also tested whether there were possible interaction effects between different relationships characteristics (see Web Appendix A.3). Due to multi-collinearity issues, we were only able to include three interaction terms. Our estimation results reveal that none of the interactions were significant and that adding these interactions neither substantially changed the estimation results, nor the holdout seeding performance.

Third, we further estimated a model in which we used a different functional form for the weighted network construction function $g(.)$ (see equation 1 and 7). While we selected the exponential function based on previous research, this function only assigns strictly positive weights to connections if a relationship exists. To allow connections with zero weights if a relationship exists, we applied a stochastic network construction function that follows a binomial distribution based on Trusov et al. (2010). Web Appendix A.4 presents the estimation results. The LMD indicates that weighted degree centrality fitted the data much better than eigenvector centrality, and that overall the exponential function fitted better. For weighted degree centrality, the holdout seeding performance using the stochastic network construction function still outperformed the traditional approach, although it does slightly worse than the multi-network approach with the exponential function. Consistent with LMD and other in- and out-of-sample fit statistics (see Table A.4.2 and Table A.4.3), eigenvector centrality did not perform well in the holdout seeding procedure.

Finally, as described in the Introduction, previous research mostly ignored relationship characteristics and assumed a binary network. However, some researchers estimated models with multiple network measures corresponding to each relationship characteristic (e.g. Aral and Walker 2012; Hu and Van den Bulte 2014) or assigned weights *a priori* to relationship characteristics (e.g., Newman 2001; Rothenberg et al. 1995). Web Appendices A.5 and A.6 compared our multi-network approach to, respectively, the procedure using multiple network measures and *a priori* assigning equal weights to different relationship characteristics. Again the LMD and other in- and out-of-sample fits statistics favored the multi-network approach in terms of model fit. More importantly, the holdout seeding performance demonstrates that estimating the importance of different relationship characteristics leads to better seeding decisions in terms of actual obtained reach.

This microfinance diffusion application illustrates that the multi-network approach not only improved in- and out-of-sample model fit, but also resulted in better reseeding strategies. Although our results are robust for several assumptions, as shown by robustness checks, a limitation of this data is that we did not observe actual cascades and therefore needed to make assumptions to construct the reach of seeds. In addition, the size of social networks in microfinance diffusion is relatively small (on average 223.2 households in a village). Our second empirical application addresses these concerns, and involves a large online social network in which we observe information propagation among network members. In this application, we aim to further test the robustness of our proposed method in a very different empirical setting (online vs. offline, large vs. small networks, and microfinance product adoption vs. information dissemination) and showcase the generalizability of our results.

EMPIRICAL APPLICATION II:
INFORMATION PROPOGATION IN AN ONINE SOCIAL NETWORK

Data Description

Our dataset contains anonymized detailed records of information transmission in a large online social network platform. The social network consists of 42,858 enrolled undergraduate students of a major university in the US. The detailed dataset contains background information of each student (age, gender, and date of joining the online social network) and detailed information about a student's online behavior, including one-to-one messaging, public posting, and commenting. For each message, we observed the sender, receiver, time of the message, as well as the content.

To study the effect of a seed's position on the spread of information, we focused on information cascades that were generated during and after the 2010 Superbowl event. We selected the Superbowl event, because many brands launched new ads that were not announced in advance during this event. Moreover, these ads tended to be of high quality and were, therefore, often mentioned in social media. The Superbowl event, thus, provides an external shock to the diffusion cascades of ads posted on social media, which allows us to validate our method. Triggered by this event, 1,620 seed students initiated messages about advertisements used during the Superbowl (see Table 6). Subsequently, we observed many instances in which friends of these posters forwarded or posted messages about the same advertisements to their friends. For each student that initiated a message about a Superbowl advertisement, we were able to identify the subsequent actual cascades (who influenced whom on what and when). Therefore, we were able to observe actual reach of seeds. Note that each cascade involved a message about only one advertisement, but that different cascades could involve different ads. This is similar to previous research that investigated the effects of network structure on the spread of different YouTube videos (Yoganarasimhan 2012). The cascades involving Superbowl advertisements reflect the complexities observed in other large-scale social network studies (Hinz et al. 2011), such that there are many short diffusion paths and much less frequent long paths (i.e., the average reach of a seed is 1.66, but the highest reach of a seed is 22, and cascade lengths are on average 1.12 with longest ones of length 4, equal to the diameter[5] of the network; see Table 7 for more details).

To uncover the importance of different relationship characteristics in the social network, we used two different sources of information. First, we used the number of messages exchanged between students in the two month period before the Superbowl event. Second, we obtained the duration of each dyadic link in the network at the moment of the Superbowl. While the former reflects the effect of interaction frequency, the latter captures possible recency effects. As found

[5] The diameter of a network refers to the longest shortest distance between two individuals in the network.

in Aral and Walker (2014), frequency and recency are strong predictors of tie strength. This setting, thus, allows us to validate whether our methodology is in line with previous research by uncovering a positive weight for the number of messages exchanged and a negative weight for relationship duration (implying a positive recency effect). Next we discuss how we constructed the variables to apply the multi-network approach to predict the reach of different seeds in the social network.

----Insert Tables 6 and 7 about here----

The Reach of a Seed (y_s). The reach of a seed is constructed by counting the number of users involved in the cascades initiated by the corresponding seed[6]. Table 7 reports summary statistics for the reach of seeds in our dataset. On average, each seed generated 1.66 adoptions or messages (SD = 2.07). In addition, we observed many seeds influencing only one adopter (about 65% in our data) and thus generating cascades of length one. However, some seeds generated a very high reach of up to 22 messages posted, and cascades of length four, which equals the diameter of this network.

Network Measures (x_s). Similar to the previous application, we used weighted degree and eigenvector centrality to predict the reach of seeds (see equation 5 and 6). Table 7 provides initial support that seeds with higher degree and eigenvector centralities obtain a higher reach (correlation coefficient equals .12 for both measures). We again find a high correlation between these two centrality measures (correlation coefficient .80), which prevents us from including both in one regression.

[6] In only 5.64 percent of the cascades, it was not possible to uniquely assign a user to a seed, because a user received messages about the same ad from multiple sources that were initiated by different seeds. In these cases, we assigned the user to the seed that initiated the earliest message. We also estimated our model by assigning the user to the seed that initiated the latest message, which resulted in almost identical results, as the two dependent variables are highly correlated (correlation coefficient was .984).

Control Variables (z_s). For each seed, we included age, gender, and how long the seed was a member of the social platform at the moment of the Superbowl event. Age and membership duration were standardized across all students in the network and gender was dummy coded, with one for males and zero for females. In addition, because we expect that early messages about the Superbowl event are more influential, we also included a timing variable that indicated the elapsed time (log of seconds) between the start of the event and sending time of the message. Table 7 presents descriptive statistics of these variables. It indicates that younger students were more likely to initiate discussions about Superbowl ads than older members (average standardized age = -.49), but correlations to reach are weak. In addition, seeding times are shown to be negatively correlated to seeds' reach (correlation coefficient equals to -.14), indicating that early seeds are possibly more influential.

Model Specification and Estimation

To apply our multi-network methodology, we used degree and eigenvector centrality to extract the weighted network measures *x*, similar to Empirical Application I (equations 5 and 6). To construct the weighted networks *W* (step 1 of the multi-network approach, see equation 1), we used the following specification:

(9)
$$w_{ij} = \begin{cases} 0 & \text{if } a_{vij}^{k} = 0, \forall k = 1, 2 \\ \exp\left(\sum_{k=1}^{2} \beta_k a_{ij}^{k}\right) & \text{otherwise} \end{cases} .$$

In equation (9), a_{ij}^{k} corresponds to the number of messages exchanged (*k*=1) or relationship duration (*k*=2) between user *i* and *j*, and β_k represents the importance. Finally, because the reach y_s is a count variable, we used a generalized linear model with Poisson link function to specify the diffusion equation $f(.)$ (equation 3). The expectation of the Poisson distribution is as follows:

(10) $$\mathrm{E}\left(y_s \,|\, x_s, z_s\right) = \exp\left(\theta_0 + \theta_1' z_s + \theta_2 x_{sm}\right).$$

Equation (10) regresses control variables z_s and network measure x_{sm} on the reach y_s of seed *s*, with $m \in \{1, 2\}$ corresponding to weighted degree and eigenvector centrality, respectively.

We estimated our model using a total of 20,000 iterations, with a burn-in period of 10,000 iterations, long before the Markov chain converged (see Appendix B for the details of MCMC sampler).

Estimation Results

Table 8 presents the estimation results of four models, two centrality measures (degree vs. eigenvector centrality) by two approaches (traditional vs. multi-network approach), with the traditional approach defined as in Empirical Application I. Similar to the first empirical application, the multi-network approach outperforms the traditional approach for both models based on LMD (Model 1: -2498 vs. -2623; Model 2: -2535 vs. -2624, respectively for the multi-network and traditional approach). Interestingly, similar to the first empirical application, the multi-network approach suggests that weighted degree centrality is a better predictor of the reach of a seed than eigenvector centrality (LMDs: -2498 vs. -2535). This is consistent with the descriptive results that most cascades are short, with few longer ones. Because both models produce similar results and Model 1 fits better, we focus on Model 1 to interpret the estimation results and uncovered importance of different relationship characteristics.

As expected, the number of messages exchanged is positively related to social influence in the network ($\beta = 4.76$, all posterior draws positive). Moreover, longer relationships are associated with weaker social influence ($\beta = -.44$, all posterior draws negative), confirming recent findings of frequency and recency effects on tie strength (Aral and Walker 2014). These results not only demonstrate that the importance of relationship characteristics vary in the social network, but also validate the uncovered importance of relationship characteristics by the proposed methodology[7]. Finally, the parameter estimates of control variables are stable across the four models, with significant effects for *age* and *seeding time* (age: $\theta = -.02$, 97.7% posterior draws negative; seeding time: $\theta = -.07$, with all posterior draws negative).

----Insert Table 8 about here----

Implications for Seed Selection

Similar to the previous application, we assessed the holdout reseeding performance of our approach and compared it to the traditional benchmark models. To do so, we randomly divided the seeds into four disjoint subsamples of 405 seeds that we used as estimation sample and the remaining 1,215 seeds as holdout sample. Table 9 presents the prediction accuracy for both in- and out-of-sample, averaged across the four samples. Corroborating the model selection criterion based on LMD, models of the multi-network approach outperform their traditional counterparts by all criteria (MAD, MAPE, RMSE and R-squared), both in the estimation as well as in the holdout sample. Moreover, the results also validate that weighted degree centrality is a better predictor of the reach of a seed, compared with eigenvector centrality.

Table 10 presents the results of the reseeding strategies based on the four different models. The multi-network approach again outperformed the traditional approach and did particularly well when a relatively small number of seeds were selected. Compared with the previous empirical application, the improvements were much larger. If 50 seeds were selected (4.12% of the holdout sample), the multi-network approach obtained reaches that were 91.91 and 31.29 percent higher than the traditional approach, respectively for degree and eigenvector centrality.

[7] We further validated the multi-network approach using the number of messages exchanged and relationship duration in Web Appendix B.

----Insert Tables 9 and 10 about here----

Robustness Checks

Similar to the first empirical application, we performed two robustness checks to assure the stability of our results (see Web Appendix C.1 to C.2). Web Appendix C.1 compared the proposed multi-network approach to the use of multiple network measures, while Web Appendix C.2 compared it to *a priori* weighted networks. Both robustness checks demonstrate the superiority of the proposed multi-network approach over these alternatives (see Tables C.1.2, C.1.3, C.2.2 and C.2.3 for more details).

DISCUSSION

Connections between consumers in social networks vary in strength, depending on the characteristics of relationships and the information exchanged. While different relationship characteristics may have different impact on diffusion processes, previous research mostly ignored this information and treated all relationship characteristics as equal. In this research, we developed a new methodology that uncovers the importance of relationship characteristics based on the observed diffusion process of a product, service or message. Our approach results in a model that can be decomposed into three components that are conditional on each other. The Bayesian estimation procedure, consisting of a Gibbs sampler nested with a Metropolis-Hastings step is therefore a natural tool to estimate the model parameters. Our Bayesian approach is efficient and can easily be applied to large scale social networks, as we demonstrated in our second empirical application. The major computational challenge for large scale social networks may be the computation of network measures (step 2 in Figure 1), such as eigenvector centrality that requires calculating the inverse of large matrices. In extreme situations where tens of millions of users are involved, it may be useful to approximate these network measures (Brandes and Pich 2007; Brin and Page 1998).

In two empirical applications that varied substantially in terms of network size (small vs. large), context (online vs. offline) and diffusion process (a microfinance product vs. an online message about an advertisement), the proposed multi-network approach demonstrated that the importance of relationship characteristics substantially varied. In both applications, recognizing these differences not only resulted in a better statistical fit, but also led to better seeding strategies. In the first application, the multi-network approach proposed seeding strategies that increased the number of actual adoptions of a microfinance program by up to 10 percent. In the second empirical application, our approach was able to increase the actual reach in a holdout sample by up to 92 percent compared with a benchmark that ignored the importance of different relationship characteristics. Interestingly, in both empirical applications we found that weighted degree centrality was a better criterion for seed selection than eigenvector centrality. However, as explained above, degree and eigenvector centrality capture two extremes of transmission processes (Borgatti 2005). It is possible that a measure in between these two extremes better captures the diffusion process. Banerjee et al. (2013) proposes such a measure – diffusion centrality. In Web Appendix D we compared different versions of this measure with degree and eigenvector centrality using the multi-network approach. Interestingly, degree centrality remains optimal in both empirical applications, although diffusion centrality obtained similar performance in the second empirical application. This finding is also in line with Hinz et al. (2011), who reported that targeting seeds with high degree centrality resulted in a higher reach than high global betweenness centrality.

For future research, it would be useful to investigate whether specific campaign characteristics, such as the type of product or service and characteristics of consumers in the network, are able to explain the importance of relationship characteristics. Such an analysis would allow marketers to predict the importance of relationship characteristics in a social network before the start of the diffusion process. Another interesting area for future research is to allow for multiple weighted social networks. In our application, we constructed only one weighted social network. However, it is possible that in other applications diffusion is better explained by two or more weighted social networks. For instance, previous research in organizational networks of employees distinguished between formal and informal relationships (Soda and Zaheer 2012; Tucker 2008). As argued by Soda and Zaheer (2012), formal relationships (consisting of workflows and organizational structures) transmit different types of information, compared with informal relationships (consisting of social relationships such as friendships). The former are more likely to transmit approvals and task-related information, while the latter are more likely to transmit advice and affect. These types of information could have a different impact on the adoption decisions of employees, such as the adoption of a new technology. In addition, it would be interesting to investigate the importance of different relationship characteristics on diffusion measures other than reach, such as the speed of diffusion or repeated purchases.

Future research could also extend our methodology in several ways. First, in our first application, we found through categorical factor analysis that different types of relationships loaded on the same underlying relationship dimensions. Future research could extend our methodology to model such similarities directly. Second, in our model, the strength of connections between consumers in a social network only varied with the characteristics of relationships. Future research could incorporate additional heterogeneity through a hierarchical Bayesian approach that allows strengths to be individual or dyad specific. This would require a dataset containing diffusion processes with multiple adoption observations for individuals. Third, we assumed all strengths between connections of consumers in the social network to be positive. It would be interesting to study whether some characteristics of relationships are negative and thus inhibit the diffusion process. For instance, Chandrashekaran et al. (2010) showed that non-repeat buyers of shareware inhibited the diffusion process through spreading negative word-of-mouth, while repeat buyers facilitated diffusion through positive word-of-mouth. Finally, the weighted network construction function in equation (1) assumes that two customers are not connected if there is no observed relationship between those customers. In reality, such customers may still influence each other and it would be interesting for future research to extend our approach to allow for such interactions. A possible direction would be to incorporate network formation models in the network construction function (Braun and Bonfrer 2011).

In conclusion, in this study we proposed a methodology that derives the strengths of connections as a function of relationship characteristics. Our approach is demonstrated to be flexible and suitable for any diffusion process where social network data is available. We believe that our methodology may be a valuable tool for managers to optimize seeding strategies to facilitate the diffusion of their products and services.

REFERENCES

Ansari, Asim, Oded Koenigsberg, and Florian Stahl (2011), "Modeling Multiple Relationships in Social Networks," *Journal of Marketing Research*, 48 (4), 713-28.

Aral, Sinan and Dylan Walker (2012), "Identifying Influential and Susceptible Members of Social Networks," *Science*, 337 (6092), 337-41.

---- (2014), "Tie strength, embeddedness, and social influence: A large-scale networked experiment," *Management Science*, 60 (6), 1352-70.

Banerjee, Abhijit, Arun G. Chandrasekhar, Esther Duflo, and Matthew O. Jackson (2013), "The Diffusion of Microfinance," *Science*, 341 (6144), 362-64.

Bonacich, Phillip and Paulette Lloyd (2001), "Eigenvector-like measures of centrality for asymmetric relations," *Social Networks*, 23 (3), 191-201.

Borgatti, Stephen P (2005), "Centrality and network flow," *Social networks*, 27 (1), 55-71.

Brandes, Ulrik and Christian Pich (2007), "Centrality estimation in large networks," *International Journal of Bifurcation and Chaos*, 17 (7), 2303-18.

Braun, Michael and André Bonfrer (2011), "Scalable inference of customer similarities from interactions data using Dirichlet processes," *Marketing Science*, 30 (3), 513-31.

Brin, Sergey and Lawrence Page (1998), "The anatomy of a large-scale hypertextual Web search engine," *Computer networks and ISDN systems*, 30 (1), 107-17.

Bronnenberg, Bart J and Carl F Mela (2004), "Market Roll-Out and Retailer Adoption for New Brands," *Marketing Science*, 23 (4), 500-18.

Brown, Jacqueline Johnson and Peter H. Reingen (1987), "Social Ties and Word-of-Mouth Referral Behavior," *Journal of Consumer Research*, 14 (3), 350-62.

Browne, William J. and David Draper (2000), "Implementation and performance issues in the Bayesian and likelihood fitting of multilevel models," *Computational Statistics*, 3 (15), 391-420.

Chandrashekaran, Murali, Rajdeep Grewal, and Raj Mehta (2010), "Estimating Contagion on the Internet: Evidence from the Diffusion of Digital/Information Products," *Journal of Interactive Marketing*, 24 (1), 1-13.

Chib, Siddhartha and Ivan Jeliazkov (2001), "Marginal Likelihood From the Metropolis-Hastings Output," *Journal of the American Statistical Association*, 96 (453), 270-81.

De Bruyn, Arnaud and Gary L. Lilien (2008), "A Multi-Stage Model of Word of Mouth Influence Through Viral Marketing," *International Journal of Research in Marketing*, 25 (3), 151-63.

Dover, Yaniv, Jacob Goldenberg, and Daniel Shapira (2012), "Network Traces on Penetration: Uncovering Degree Distribution from Adoption Data," *Marketing Science*, 31 (4), 689-712.

Godes, David and Dina Mayzlin (2009), "Firm-Created Word-of-Mouth Communication: Evidence from a Field Test," *Marketing Science*, 28 (4), 721-39.

Goldenberg, Jacob, Sangman Han, Donald R. Lehmann, and Jae Weon Hong (2009), "The Role of Hubs in the Adoption Process," *Journal of Marketing*, 73 (2), 1-13.

Goldenberg, Jacob, Donald R Lehmann, Daniella Shidlovski, and Michal Master Barak (2006), "The role of expert versus social opinion leaders in new product adoption," *Cambridge, Mass.: Marketing Science Institute, Report* (06-124).

Granovetter, Mark S. (1973), "The Strength of Weak Ties," *American Journal of Sociology*, 78 (6), 1360-80.

Hinz, Oliver, Bernd Skiera, Christian Barrot, and Jan U. Becker (2011), "Seeding Strategies for Viral Marketing: An Empirical Comparison," *Journal of Marketing*, 75 (6), 55-71.

Hu, Yasong and Christophe Van den Bulte (2014), "Nonmonotonic Status Effects in New Product Adoption," *Marketing Science*, 33 (4), 509-33.

Iyengar, Raghuram, Christophe Van den Bulte, and Thomas W. Valente (2011), "Opinion Leadership and Social Contagion in New Product Diffusion," *Marketing Science*, 30 (2), 195-212.

Jackson, Matthew O., Thomas Rodriguez-Barraquer, and Xu Tan (2012), "Social Capital and Social Quilts: Network Patterns of Favor Exchange," *American Economic Review*, 102 (5), 1857-97.

Jiang, Zhengrui, Frank M Bass, and Portia Isaacson Bass (2006), "Virtual Bass model and the left-hand data-truncation bias in diffusion of innovation studies," *International Journal of Research in Marketing*, 23 (1), 93-106.

Kalish, Shlomo, Vijay Mahajan, and Eitan Muller (1995), "Waterfall and sprinkler new-product strategies in competitive global markets," *International Journal of Research in Marketing*, 12 (2), 105-19.

Katona, Zsolt, Peter Pal Zubcsek, and Miklos Sarvary (2011), "Network Effects and Personal Influences: The Diffusion of an Online Social Network," *Journal of Marketing Research*, 48 (3), 425-43.

Leenders, Roger Th. A. J. (2002), "Modeling Social Network through Network Autocorrelation: Constructing the Weighting Matrix," *Social Networks*, 24 (1), 21-47.

LeSage, James P. and R. Kelley Pace (2014), "The biggest myth in spatial econometrics," *Econometrics*, 2 (4), 217-49.

Libai, Barak, Eitan Muller, and Renana Peres (2013), "Decomposing the Value of Word-of-Mouth Seeding Programs: Acceleration Versus Expansion," *Journal of Marketing Research*, 50 (2), 161-76.

Lilien, Gary L, Ambar G Rao, and Shlomo Kalish (1981), "Bayesian Estimation and Control of Detailing Effort in a Repeat Purchase Diffusion Environment," *Management Science*, 27 (5), 493-506.

Luczkovich, Joseph J., Stephen P. Borgatti, Jeffrey C. Johnson, and Martin G. Everett (2003), "Defining and measuring trophic role similarity in food webs using regular equivalence," *Journal of Theoretical Biology*, 220 (3), 303-21.

Moschis, George P. and Roy L. Moore (1979), "Decision Making Among the Young: A Socialization Perspective," *Journal of Consumer Research*, 6 (2), 101-12.

Newman, Mark E.J. (2004), "Analysis of Weighted Networks," *Physical Review E*, 70 (5), 1-9.

---- (2001), "Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality," *Physical Review E*, 64 (1), 016132-1-7.

Opsahl, Tore, Filip Agneessens, and John Skvoretz (2010), "Node centrality in Weighted Networks: Generalizing Degree and Shortest Paths," *Social Networks*, 32 (3), 245-51.

Opsahl, Tore, Vittoria Colizza, Pietro Panzarasa, and José J. Ramasco (2008), "Prominence and control: The weighted rich-club effect," *Physical Review Letters*, 101 (16), 1687021-4.

Páez, Antonio, Darren M. Scott, and Erik Volz (2008), "Weight Matrices for Social Influence Analysis: An Investigation of Measurement Errors and Their Effect on Model Identification and Estimation Quality," *Social Networks*, 30 (4), 309-17.

Parry, Charles DH and JJ McArdle (1991), "An applied comparison of methods for least-squares factor analysis of dichotomous variables," *Applied Psychological Measurement*, 15 (1), 35-46.

Reingen, Peter H., Brian L. Foster, Jacqueline Johnson Brown, and Stephen B. Seidman (1984), "Brand Congruence in Interpersonal Relations: A Social Network Analysis," *Journal of Consumer Research*, 11 (3), 771-83.

Rossi, Peter E., Greg M. Allenby, and Robert McCulloch (2005), Bayesian Statistics and Marketing. Chichester, UK: Wiley.

Rothenberg, Richard B., John J. Potterat, Donald E. Woodhouse, William W. Darrow, Stephen Q. Muth, and Alden S. Klovdahl (1995), "Choosing a centrality measure: Epidemiologic correlates in the Colorado Springs study of social networks," *Social Networks*, 17 (3-4), 273-97.

Schulze, Christian, Lisa Schöler, and Bernd Skiera (2014), "Not All Fun and Games: Viral Marketing for Utilitarian Products," *Journal of Marketing*, 78 (1), 1-19.

Soda, Giuseppe and Akbar Zaheer (2012), "A Network Perspective on Organizational Architecture: Performance Effects of the Interplay of Formal and Informal Organization," *Strategic Management Journal*, 33 (6), 751-71.

Trusov, Michael, Anand V. Bodapati, and Randolph E. Bucklin (2010), "Determining Influential Users in Internet Social Networks," *Journal of Marketing Research*, 47 (4), 643-58.

Tucker, Catherine (2008), "Identifying Formal and Informal Influence in Technology Adoption with Network Externalities," *Management Science*, 54 (12), 2024-38.

Van den Bulte, Christophe and Gary L. Lilien (2001), "Medical Innovation Revisited: Social Contagion versus Marketing Effort," *American Journal of Sociology*, 106 (5), 1409-35.

Van den Bulte, Christophe and Stefan Wuyts (2007), Social Networks and Marketing. Cambridge, MA: Marketing Science Institute.

van der Lans, Ralf, Gerrit van Bruggen, Jehoshua Eliashberg, and Berend Wierenga (2010), "A Viral Branching Model for Predicting the Spread of Electronic Word of Mouth," *Marketing Science*, 29 (2), 348-65.

Watts, Duncan J. and Steven H. Strogatz (1998), "Collective dynamics of 'small-world' networks," *Nature*, 393 (6684), 440-42.

Yoganarasimhan, Hema (2012), "Impact of Social Network Structure on Content Propagation: A Study using YouTube Data," *Quantitative Marketing and Economics*, 10 (1), 111-50.

Table 1
MICROFINANCE DIFFUSION: DESCRIPTIVES OF SOCIAL NETWORKS

	All households mean (SD)	**Seeds mean (SD)**	**Non-seeds mean (SD)**
Network Size	223.2 (56.2)	26.9 (9.2)	196.3 (50.2)
Network Metrics			
Total Network			
Degree centrality	1.27 (1.00)	1.75 (1.21)	1.21 (.95)
Eigenvector centrality	.58 (.53)	.84 (.65)	.55 (.50)
Economic relations			
Degree centrality	1.21 (1.00)	1.67 (1.17)	1.15 (.96)
Eigenvector centrality	.50 (.57)	.72 (.70)	.46 (.54)
Social relations			
Degree centrality	1.25 (1.00)	1.74 (1.23)	1.18 (.95)
Eigenvector centrality	.55 (.54)	.81 (.69)	.52 (.51)
Religious relations			
Degree centrality	.46 (1.00)	.62 (1.19)	.44 (.97)
Eigenvector centrality	.07 (.36)	.12 (.50)	.06 (.34)
Family relations			
Degree centrality	1.12 (1.00)	1.34 (1.10)	1.09 (.99)
Eigenvector centrality	.25 (.61)	.29 (.66)	.24 (.61)

Note: Degree centrality is normalized for comparison across networks.

Table 2
MICROFINANCE DIFFUSION: DESCRIPTIVES STATISTICS OF ALL VARIABLES

Variables	Mean	SD.	Max	Min	Correlation matrix											
					1	2	3	4	5	6	7	8	9	10	11	12
Dependent variable																
1. Reach of a seed (y)	1.45	1.60	13.26	.00	1.00											
Network measures																
2. Degree Centrality	1.75	1.21	9.40	.00	.58	1.00										
3. Eigenvector Centrality	.84	.65	4.63	.00	.53	.93	1.00									
Control variables																
4. Percentage of Seeds	.00	.03	.07	-.06	-.22	-.02	-.01	1.00								
5. Roof (Thatch)	.28	.45	1.00	.00	.05	.00	-.01	-.01	1.00							
6. Roof (Tile)	.31	.46	1.00	.00	-.01	.01	.01	.09	-.42	1.00						
7. Roof (Stone)	.20	.40	1.00	.00	-.08	-.09	-.09	-.02	-.32	-.33	1.00					
8. Roof (Sheet)	.16	.37	1.00	.00	.04	.12	.11	-.07	-.28	-.29	-.21	1.00				
9. No. of Rooms	.00	1.58	15.25	-2.75	.09	.26	.26	-.02	-.12	-.03	-.05	.29	1.00			
10. Electricity	.73	.45	1.00	.00	-.04	.11	.12	-.07	-.13	-.01	.01	.19	.26	1.00		
11. Latrine	.39	.49	1.00	.00	.05	.15	.14	-.03	-.19	.01	-.03	.29	.33	.32	1.00	
12. House	.93	.26	1.00	.00	.01	.04	.03	-.02	.07	.03	-.09	-.02	.06	.04	.05	1.00

Table 3
MICROFINANCE DIFFUSION: ESTIMATION RESULTS

Variables	**Traditional Approach** Model 1	**Traditional Approach** Model 2	**Multi-network Approach** Model 1	**Multi-network Approach** Model 2
Construction of weighted network				
Economic			**-.23**	**.30**
			(-.38, -.05)	(.18, .44)
Social			**.24**	**.36**
			(.085, .385)	(.24, .47)
Religious			.48	.20
			(-.02, .93)	(-.01, .43)
Family			-.10	.04
			(-.20, .02)	(-.05, .21)
Diffusion Equation				
Constant	**.50**	**.67**	**.52**	**.63**
	(.07, .92)	(.25, 1.11)	(.094, .95)	(.20, 1.07)
Control variables				
Percentage of Seeds	**-11.69**	**-12.10**	**-11.74**	**-12.83**
	(-14.21, -9.77)	(-14.61, -9.35)	(-14.10, -9.19)	(-15.38, -10.35)
Roof_1 (Thatch)	-.06	.01	-.048	-.05
	(-.39, .28)	(-.35, .37)	(-.40, .29)	(-0.40, 0.31)
Roof_2 (Tile)	-.14	-.08	-.15	-.15
	(-.48, .19)	(-.43, .28)	(-.49, .18)	(-.50, .20)
Roof_3 (Stone)	-.24	-.18	-.24	-.23
	(-.58, .12)	(-.53, .17)	(-.58, .11)	(-.58, .12)
Roof_4 (Sheet)	-.22	-.14	-.23	-.21
	(-.59, .15)	(-.52, .25)	(-.59, .15)	(-.59, .17)
No of Rooms	-.04	-.03	-.04	-.03
	(-.10, .01)	(-.09, .02)	(-.09, .01)	(-.08, .03)
Electricity	**-.36**	**-.40**	**-.35**	**-.35**
	(-.54, -.19)	(-.58, -.21)	(-.53, -.18)	(-.53, -.17)
Latrine	.03	.06	.03	.07
	(-.14, .19)	(-.11, .23)	(-.14, .20)	(-.01, .23)
House	-.05	.04	-.08	-.03
	(-.33, .21)	(-.26, .33)	(-.35, .21)	(-.32, .25)
Network centrality measures				
Degree	**.81**		**.82**	
	(.74, .86)		(.75, .88)	
Eigenvector		**1.36**		**1.07**
		(1.24, 1.48)		(.94, 1.19)
Variance of Error	**1.56**	**1.69**	**1.55**	**1.61**
	(1.44, 1.71)	(1.56, 1.84)	(1.43, 1.68)	(1.48, 1.75)
LMD	-1647.19	-1642.28	-1627.13	-1640.61

Note: 95 percent posterior intervals are reported between brackets.

Table 4
MICROFINANCE DIFFUSION: IN- AND OUT-OF-SAMPLE FIT STATISTICS

Criterion	**In-Sample**				**Out-of-Sample**			
	Traditional Approach		Multi-network Approach		Traditional Approach		Multi-network Approach	
	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)
MAD	.99	.95	.87	.89	1.01	.97	.88	.90
MAPE	.68	.65	.60	.61	.69	.66	.60	.61
RMSE	1.36	1.31	1.24	1.26	1.39	1.33	1.26	1.28
R-squared	.39	.41	.54	.51	.38	.40	.53	.51

Notes: MAD – Mean Absolute Deviance; MAPE: Mean Absolute Percentage of Error; RMSE: Root Mean Squared Error.

Table 5
MICROFINANCE DIFFUSION: RESEEDING PERFORMANCE

No. of Selected Seeds (%)		**Traditional Approach**		**Multi-network Approach**		**% Improvement**	**% Improvement**
		Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)
50	(5.77%)	182.42	177.44	200.75	203.75	10.04%	14.83%
100	(11.53%)	343.84	319.17	378.72	353.25	10.14%	10.68%
150	(17.30%)	469.41	450.12	501.46	473.56	6.82%	5.12%
200	(23.07%)	576.95	556.89	598.55	576.96	3.74%	3.60%
250	(28.84%)	665.91	642.41	690.41	680.57	3.68%	5.94%
300	(34.60%)	756.29	717.80	780.66	768.93	3.22%	7.12%
400	(46.14%)	892.59	872.47	921.34	919.38	3.22%	5.38%
500	(57.67%)	1015.82	979.51	1041.66	1020.73	2.54%	4.21%
600	(69.20%)	1117.19	1099.47	1128.52	1122.53	1.01%	2.10%
700	(80.74%)	1189.04	1172.83	1198.49	1188.22	.80%	1.31%
800	(92.27%)	1237.21	1231.61	1242.11	1232.23	.40%	.05%
867	(100.00%)	1258.24	1258.24	1258.24	1258.24	.00%	.00%

Table 6
INFORMATION PROPAGATION: DESCRIPTIVES OF SOCIAL NETWORKS

	All households mean (SD)	Seeds mean (SD)	Non-seeds mean (SD)
Network Size	42,852	1,620	41,232
Network Metrics			
Total Network (Binary)			
Degree centrality	1.05 (1.00)	1.67 (1.16)	1.03 (.99)
Eigenvector centrality	.30 (.54)	.53 (.69)	.30 (.53)
Number of Messages Exchanged (Weighted)			
Degree centrality	.16 (1.00)	.76 (1.77)	.14 (.95)
Eigenvector centrality	.0009 (.06)	.002 (.11)	.0008 (.05)
Relationship Duration (Weighted)			
Degree centrality	1.05 (1.00)	1.66 (1.16)	1.03 (.99)
Eigenvector centrality	.30 (.54)	.52 (.69)	.30 (.53)

Note: Degree centrality is normalized for comparison across networks.

Table 7
INFORMATION PROPAGATION: DESCRIPTIVES OF ALL VARIABLES

Variables	Mean	SD.	Min	Max	Correlation Matrix						
					1	2	3	4	5	6	7
Dependent variable											
1. Reach of Seeds (y)	1.66	2.07	1.00	22.00	1.00						
Network measures											
2. Degree Centrality	1.67	1.16	.01	7.68	.12	1.00					
3. Eigenvector Centrality	.53	.69	.00	5.24	.12	.80	1.00				
Control variables											
4. Age	-.49	1.23	-2.06	2.94	-.04	-.07	-.01	1.00			
5. Gender	.32	.47	.00	1.00	.00	-.12	-.12	.16	1.00		
6. Membership Duration	.01	.19	-.79	.39	.03	.11	.14	.54	-.02	1.00	
7. Seeding Time	1.71	1.00	0.00	3.85	-.14	-.08	-.09	.05	.06	-.02	1.00

Table 8
INFORMATION PROPAGATION: ESTIMATION RESULTS

Variables	**Traditional Approach**		**Multi-network Approach**	
	Model 1	Model 2	Model 1	Model 2
Construction of Weighted Network				
Number of Messages Exchanged			**4.76**	**1.46**
			(1.91, 6.25)	(1.37, 1.56)
Relationship Duration			**-.44**	**-.23**
			(-.54, -.34)	(-.33, -.14)
Diffusion Equation				
Intercept	**.90**	**.93**	**.85**	**.91**
	(.85, .94)	(.89, .97)	(.80, .89)	(.86, .95)
Age	-.02	-.02	**-.02**	-.02
	(-.03, .00)	(-.04, .00)	(-.04, -.00)	(-.03, .00)
Gender	.00	.00	.00	.00
	(-.04, .04)	(-.04, .05)	(-.03, .04)	(-.04, .04)
Membership Duration	.10	.11	.06	.08
	(-.02, .22)	(-.01, .23)	(-.05, .17)	(-.03, .20)
Seeding Time	**-.06**	**-.06**	**-.07**	**-.06**
	(-.08, -.04)	(-.08, -.04)	(-.08, -.05)	(-.08, -.04)
Network Measures				
Degree	**.04**		**.06**	
	(.03, .06)		(.05, .07)	
Eigenvector		**.07**		**.10**
		(.04, .09)		(.08, .12)
LMD	-2623.46	-2623.97	-2497.71	-2535.25

Note: 95 percent posterior intervals are reported between brackets.

Table 9
INFORMATION PROPAGATION: IN- AND OUT-OF-SAMPLE FIT STATISTICS

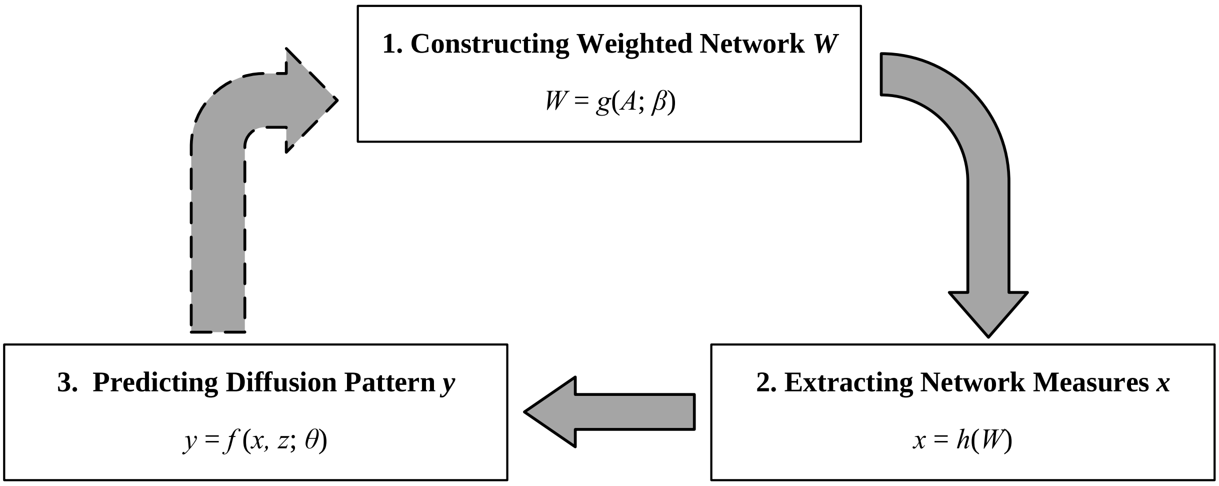
Criterion	**In-Sample**				**Out-of-Sample**			
	Traditional Approach		Multi-network Approach		Traditional Approach		Multi-network Approach	
	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)
MAD	.815	.817	.752	.788	.842	.844	.771	.810
MAPE	.383	.384	.358	.377	.388	.388	.361	.381
RMSE	2.009	2.009	1.742	1.791	2.161	2.162	1.785	1.836
R-squared	.026	.030	.090	.074	.023	.023	.082	.063

Notes: MAD – Mean Absolute Deviance; MAPE: Mean Absolute Percentage of Error; RMSE: Root Mean Squared Error; R-squared: McFadden's Pseudo R-squared.

Table 10
INFORMATION PROPAGATION: HOLDOUT SEEDING PERFORMANCE

No. of Selected Seeds		**Traditional Approach**		**Multi-network Approach**		**% of Improvement**	**% of Improvement**
		Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)
50	(4.12%)	111.25	126.25	213.50	165.75	91.91%	31.29%
100	(8.23%)	256.00	255.00	323.50	294.25	26.37%	15.39%
150	(12.35%)	384.50	402.00	434.00	425.25	12.87%	5.78%
200	(16.46%)	480.50	493.50	520.00	511.50	8.22%	3.65%
250	(20.58%)	571.00	569.00	615.50	609.75	7.79%	7.16%
450	(37.04%)	956.75	948.50	963.00	983.00	.65%	3.64%
650	(53.50%)	1260.25	1255.25	1296.50	1282.75	2.88%	2.19%
850	(69.96%)	1545.75	1532.00	1572.00	1572.75	1.70%	2.66%
1050	(86.42%)	1822.00	1810.50	1827.00	1812.75	.27%	.12%
1215	(100.00%)	2014.50	2014.50	2014.50	2014.50	.00%	.00%

Figure 1
CONCEPTUALIZATION OF MULTI-NETWORK APPROACH



Note: In Component 1, W is a weighted adjacency matrix and specified as a function of different characteristics of relationships, summarized in A and parameter vector β. In Component 2, weighted network measures x (such as weighted degree centrality, eigenvector centrality, etc.) are derived using function $h(\cdot)$ of the weighted network W. Finally, in Component 3, the reach of seeds y is a function of network measures x, control variables z and parameter vector θ. The dashed feedback arrow from Component 3 to Component 1 represents the feedback loop in the Bayesian iterative estimation procedure.

APPENDIX A

MCMC SAMPLER FOR MICROFINANCE DIFFUSION

Priors:

We used the following diffuse priors for the model parameters:

$\theta \sim N\left(0,100 I_{M\times M}\right)$ and $\beta \sim N\left(0,100 I_{4\times 4}\right)$, and $\sigma^2 \sim IG\left(\frac{r_0}{2},\frac{s_0}{2}\right)$, where *I*'s indicate identity matrices, and *IG* refers to the inverse gamma distribution. In our application, we set $r_0 = s_0 = 2$. This leads to the following MCMC procedure.

MCMC procedure:

In our MCMC sampler, we draw sequentially from the following distributions:

1) Metropolis-Hasting steps to draw $\beta_k, \forall k = 1,..,4$. Given the current value of β_k , draw a new value $\beta_k^{new} = \beta_k + \rho_k \varphi$, with $\varphi \sim N\left(0,1\right)$. ρ_k is dynamically tuned to insure appropriate level of acceptance rate between 25%-45% (see Browne and Draper 2000). The new proposed value is accepted with the following probability:

$$P_{accept} = \min\left\{1, \frac{\phi\left(\beta_k^{new} \middle| 0,100\right) L\left(y \mid z, h\left(g\left(A;\left[\beta_k^{new}, \beta_{-k}\right]\right)\right), \theta, \sigma^2\right)}{\phi\left(\beta_k \middle| 0,100\right) L\left(y \mid z, h\left(g\left(A;\beta\right)\right), \theta, \sigma^2\right)}\right\},$$

where $\phi\left(\cdot \middle| 0,100\right)$ represents the *p.d.f.* of a normal distribution with mean 0 and variance 100. Furthermore, $L\left(\cdot\right)$ is the likelihood function as obtained by combining equations (5) to (8).

2) $\theta \sim N\left(Q\left(\sigma^{-2} X'y\right), Q\right)$, with $X = \left[1 \quad z \quad x\right]$ and $Q = \left(\sigma^{-2} X'X + .01 I_{M\times M}\right)^{-1}$.

3) $\sigma^2 \sim IG\left(\frac{r_n}{2},\frac{s_n}{2}\right)$, with $r_n = r_0 + NS$ and $s_n = s_0 + \left(y - X\theta\right)'\left(y - X\theta\right)$, where *NS* denotes the number of seeds.

APPENDIX B

MCMC SAMPLER FOR ONLINE INFORMATION PROPAGATION

Priors:

We used the following diffuse priors for the model parameters:

$\theta \sim N\left(0,100 I_{M\times M}\right)$ and $\beta \sim N\left(0,100 I_{2\times 2}\right)$, where I's are identify matrices. This leads to the following MCMC procedure.

MCMC procedure:

In our MCMC sampler, we draw sequentially from the following distributions:

1) Because this step is similar to Step 1 of the MCMC sampler in the microfinance diffusion application (see Appendix A), we will not repeat here.
2) To draw the parameters in Poisson regression, we use an independent Metropolis-Hastings step, where the proposal distribution is assumed to be a multivariate normal distribution centered at the maximum likelihood estimates of the Poisson regression, with variance-covariance matrix set to the asymptotic covariance matrix (i.e. approximated by inverse of Hessian H of the log-likelihood). The new proposal value θ^{new} is accepted with the following probability,

$$P_{accept}\left(\theta^{new}\right)=\min\left\{1,\frac{\phi\left(\theta^{new}\middle|0,100I_K\right)L\left(y\,|\,z,h\left(g\left(A;\beta\right)\right),\theta^{new}\right)}{\phi\left(\theta\middle|0,100I_K\right)L\left(y\,|\,z,h\left(g\left(A;\beta\right)\right),\theta\right)}\times\frac{\phi\left(\theta\middle|\theta^{MLE},H^{-1}\right)}{\phi\left(\theta^{new}\middle|\theta^{MLE},H^{-1}\right)}\right\}.$$

WEB APPENDIX A
ROBUSTNESS CHECKS FOR MICROFINANCE DIFFUSION

ROBUSTNESS CHECK A.1:
RESULTS WITH DIFFERENT RULE TO ASSIGN ADOPTERS TO SEEDS

TABLE A.1.1
ESTIMATION RESULTS OF DIFFERENT MODELS

Variables	**Traditional Approach**		**Multi-network Approach**	
	Model 1	Model 2	Model 1	Model 2
Construction of weighted network				
Economic			**-.10**	**.33**
			(-.26, -.06)	(.20, .46)
Social			**.09**	**.27**
			(.00, .24)	(.17, .38)
Religious			.81	.23
			(-.16, 1.74)	(-.03, .64)
Family			-.11	.06
			(-.20, .01)	(-.03, .21)
Diffusion Equation				
Constant	-.46	-.19	-.40	-.26
	(-1.06, .18)	(-.85, .46)	(-1.02, .22)	(-.88, .36)
Control variables				
Percentage of Seeds	**-11.41**	**-11.72**	**-11.31**	**-13.00**
	(-14.92, -7.79)	(-15.61, -8.08)	(-14.77, -7.66)	(-16.65, -9.32)
Roof_1 (Thatch)	.19	.28	.18	.19
	(-.30, .68)	(-.26, .80)	(-.31, .67)	(-.31, .69)
Roof_2 (Tile)	.10	.17	.08	.08
	(-.41, .59)	(-.35, .69)	(-.41, .56)	(-.42, .57)
Roof_3 (Stone)	-.02	.06	-.03	-.01
	(-.52, .48)	(-.49, .59)	(-.52, .47)	(-.51, .51)
Roof_4 (Sheet)	.00	.11	-.02	.01
	(-.53, .55)	(-.46, .68)	(-.56, .52)	(-.54, .54)
No of Rooms	**-.07**	-.06	-.07	-.06
	(-.15, -.00)	(-.13, .02)	(-.15, .00)	(-.13, .02)
Electricity	**-.40**	**-.46**	**-.40**	**-.40**
	(-.67, -.15)	(-.73, -.20)	(-.65, -.16)	(-.66, -.15)
Latrine	-.043	.007	-.05	.01
	(-.29, .20)	(-.24, .25)	(-.29, .20)	(-.24, .25)
House	.05	.10	.01	.07
	(-.37, .46)	(-.33, .51)	(-.40, .42)	(-.35, .49)
Network centrality measures				
Degree	**1.19**		**1.21**	
	(1.10, 1.28)		(1.11, 1.31)	
Eigenvector		**2.04**		**1.68**
		(1.87, 2.22)		(1.49, 1.86)
Variance of Error	**3.31**	**3.59**	**3.31**	**3.37**
	(3.06, 3.61)	(3.28, 3.89)	(3.05, 3.59)	(3.14, 3.66)
LMD	**-1665.28**	**-1665.55**	**-1653.38**	**-1658.48**

Note: 95 percent posterior intervals are reported between brackets.

TABLE A.1.2
IN- AND OUT-OF-SAMPLE FIT STATISTICS

Criterion	In-Sample				Out-of-Sample			
	Traditional Approach		Multi-network Approach		Traditional Approach		Multi-network Approach	
	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)
MAD	1.35	1.42	1.28	1.30	1.47	1.48	1.30	1.31
MAPE	.91	.96	.88	.88	1.00	1.01	.88	.89
RMSE	2.57	2.58	1.97	2.07	2.89	2.89	2.33	2.39
R-squared	.45	.45	.52	.52	.39	.40	.47	.46

Notes: MAD – Mean Absolute Deviance; MAPE: Mean Absolute Percentage of Error; RMSE: Root Mean Squared Error.

TABLE A.1.3
HOLDOUT SEEDING PERFORMANCE

No. of Selected Seeds (%)		Traditional Approach		Multi-network Approach		% Improvement	% Improvement
		Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)
100	(11.53%)	533.75	572.38	597.57	595.17	11.96%	3.98%
200	(23.07%)	780.83	787.57	823.47	812.60	5.46%	5.46%
300	(34.60%)	913.64	918.08	955.01	930.26	4.53%	1.33%
400	(46.14%)	1035.45	1003.05	1052.34	1033.65	1.63%	3.05%
500	(57.67%)	1097.02	1078.53	1131.16	1121.69	3.11%	4.00%
600	(69.20%)	1167.69	1165.78	1186.79	1171.75	1.64%	.51%
700	(80.74%)	1224.86	1214.93	1233.31	1224.45	.68%	.78%
800	(92.27%)	1243.95	1233.13	1247.53	1246.08	.29%	1.05%
867	(100.00%)	1258.24	1258.24	1258.24	1258.24	.00%	.00%

ROBUSTNESS CHECK A.2:
RESULTS WITH WEIGHTED ASSIGNMENT RULE

TABLE A.2.1
ESTIMATION RESULTS OF DIFFERENT MODELS

Variables	**Traditional Approach**		**Multi-network Approach**	
	Model 1	Model 2	Model 1	Model 2
Construction of weighted network				
Economic			**-.21**	**.083**
			(-.22, -.19)	(.078, .086)
Social			**.11**	**.197**
			(.08, .11)	(.185, .203)
Religious			**.02**	**.029**
			(.02, .03)	(.022, .034)
Family			**.34**	**.202**
			(.32, .38)	(.187, .210)
Diffusion Equation				
Constant	**.64**	**.81**	**.57**	**.71**
	(.11, 1.17)	(.28, 1.34)	(0.04, 1.09)	(.18, 1.25)
Control variables				
Percentage of Seeds	**-13.03**	**-13.37**	**-13.02**	**-13.69**
	(-16.06, -10.01)	(-16.50, -10.23)	(-16.01, -9.99)	(-16.74, -10.56)
Roof_1 (Thatch)	.07	.14	.07	.11
	(-0.35, .50)	(-.29, .57)	(-.35, .49)	(-.32, .54)
Roof_2 (Tile)	.04	.13	.04	.08
	(-0.38, .45)	(-0.30, 0.55)	(-.38, .45)	(-.34, .51)
Roof_3 (Stone)	-.13	-.00	-.13	-.02
	(-.57, .29)	(-.44, .44)	(-.56, .30)	(-.46, .41)
Roof_4 (Sheet)	-.10	.01	-.08	-.02
	(-.55, .35)	(-.45, .48)	(-.55, .36)	(-.49, .43)
No of Rooms	-.04	-.01	-.05	-.01
	(-.10, .02)	(-.07, .06)	(-.11, .02)	(-.08, .05)
Electricity	**-.41**	**-.53**	**-.41**	**-.51**
	(-.62, -.19)	(-.75, -.31)	(-.63, -.20)	(-.72, -.29)
Latrine	-.02	-.03	-.04	-.04
	(-.22, .18)	(-.24, .18)	(-.24, .16)	(-.24, .17)
House	.04	.05	.05	.05
	(-.31, .38)	(-.30, .39)	(-.29, .38)	(-.30, .39)
Degree	**.60**		**.71**	
	(.52, .68)		(.63, .78)	
Eigenvector		**1.01**		**1.10**
		(.86, 1.15)		(.97, 1.22)
Variance of Error	**2.56**	**2.53**	**2.33**	**2.41**
	(2.37, 2.76)	(2.29, 2.70)	(2.15, 2.52)	(2.22, 2.62)
LMD	N.A.	N.A.	-2709.91	-2731.37

1. 95 percent posterior intervals are reported in the bracket.
2. LMD for traditional approach is not available as the "estimates" are aggregated over those regression coefficients obtained by regressing DV derived by the multi-network approach at each iteration of the sampler.

TABLE A.2.2
IN- AND OUT-OF-SAMPLE FIT STATISTICS

Criterion	**In-Sample**				**Out-of-Sample**			
	Traditional Approach		Multi-network Approach		Traditional Approach		Multi-network Approach	
	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)
MAD	1.11	1.11	.99	1.03	1.12	1.13	1.05	1.07
MAPE	.74	.74	.65	.68	.74	.75	.69	.71
RMSE	1.52	1.55	1.37	1.41	1.53	1.57	1.46	1.46
R-squared	.27	.26	.42	.37	.26	.24	.34	.32

Notes: MAD – Mean Absolute Deviance; MAPE: Mean Absolute Percentage of Error; RMSE: Root Mean Squared Error.

TABLE A.2.3
HOLDOUT SEEDING PERFORMANCE

No. of Selected Seeds (%)		**Traditional Approach**		**Multi-network Approach**		**% Improvement**	**% Improvement**
		Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)
100	(11.53%)	284.99	282.06	337.60	346.94	18.46%	23.00%
200	(23.07%)	490.90	488.20	569.49	564.32	16.01%	15.59%
300	(34.60%)	678.68	669.71	760.88	764.63	12.11%	14.17%
400	(46.14%)	837.16	836.99	924.30	937.88	10.41%	12.05%
500	(57.67%)	965.42	964.53	1061.60	1048.51	9.96%	8.71%
600	(69.20%)	1088.25	1081.80	1163.67	1152.57	6.93%	6.54%
700	(80.74%)	1179.84	1185.65	1237.49	1244.54	4.89%	4.97%
800	(92.27%)	1226.67	1233.32	1246.14	1251.46	1.59%	1.47%
867	(100.00%)	1258.24	1258.24	1258.24	1258.24	.00%	.00%

ROBUSTNESS CHECK A.3:
RESULTS WITH INTERACTIONS BETWEEN RELATIONSHIPS

TABLE A.3.1
ESTIMATION RESULTS OF DIFFERENT MODELS

Variables	**Traditional Approach**		**Multi-network Approach**	
	Model 1	Model 2	Model 1	Model 2
Construction of weighted network				
Economic			**-.25**	**.15**
			(-.41, -.02)	(.08, .21)
Social			**.24**	**.53**
			(.05, .48)	(.37, .70)
Religious			.72	.12
			(-.09, 1.53)	(-.05, .45)
Family			-.14	.10
			(-.26, .05)	(-.03, .31)
Economic*Social			.03	.07
			(-.22, .31)	(-.22, .39)
Economic*Religious			-.17	-.14
			(-.55, .14)	(-.42, .12)
Social*Religious			.05	-.40
			(-.24, .31)	(-1.12, .18)
Diffusion Equation				
Constant	**.50**	**.67**	**.51**	**.53**
	(.07, .92)	(.25, 1.11)	(.10, .93)	(.09, .95)
Control variables				
Percentage of Seeds	**-11.69**	**-12.10**	**-11.80**	**-12.16**
	(-14.21, -9.77)	(-14.61, -9.35)	(-14.21,-9.32)	(-14.68, -9.61)
Roof_1 (Thatch)	-.06	.01	-.05	-.07
	(-.39, .28)	(-.35, .37)	(-.39, .29)	(-.41, .29)
Roof_2 (Tile)	-.14	-.08	-.15	-.21
	(-.48, .19)	(-.43, .28)	(-.48, .19)	(-.54, .13)
Roof_3 (Stone)	-.24	-.18	-.24	-.30
	(-.58, .12)	(-.53, .17)	(-.58, .11)	(-.65, .07)
Roof_4 (Sheet)	-.22	-.14	-.22	-.23
	(-.59, .15)	(-.52, .25)	(-.58, .14)	(-.61, .14)
No of Rooms	-.04	-.03	-.04	-.04
	(-.10, .01)	(-.09, .02)	(-.09,.01)	(-.09, .01)
Electricity	**-.36**	**-.40**	**-.35**	**-.32**
	(-.54, -.19)	(-.58, -.21)	(-.53, -.18)	(-.50, -.15)
Latrine	.03	.06	.02	.02
	(-.14, .19)	(-.11, .23)	(-.14, .19)	(-.15, .19)
House	-.05	.04	-.07	-.04
	(-.33, .21)	(-.26, .33)	(-.34, .21)	(-.32, .25)
Network centrality measures				
Degree	**.81**		**.82**	
	(.74, .86)		(.75, .88)	
Eigenvector		**1.36**		**1.13**
		(1.24, 1.48)		(.99, 1.28)
Variance of Error	**1.56**	**1.69**	**1.55**	**1.57**
	(1.44, 1.71)	(1.56, 1.84)	(1.43, 1.68)	(1.45, 1.71)
LMD	**-1647.19**	**-1642.28**	**-1635.61**	**-1647.90**

Note: 95 percent posterior intervals are reported between brackets.

TABLE A.3.2
IN- AND OUT-OF-SAMPLE FIT STATISTICS

Criterion	In-Sample				Out-of-Sample			
	Traditional Approach		Multi-network Approach		Traditional Approach		Multi-network Approach	
	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)
MAD	.95	.99	.85	.88	.97	1.01	.87	.89
MAPE	.65	.68	.58	.60	.66	.69	.60	.61
RMSE	1.31	1.36	1.20	1.23	1.33	1.39	1.24	1.27
R-squared	.41	.39	.55	.52	.40	.38	.53	.51

Notes: MAD – Mean Absolute Deviance; MAPE: Mean Absolute Percentage of Error; RMSE: Root Mean Squared Error.

TABLE A.3.3
HOLDOUT SEEDING PERFORMANCE

No. of Selected Seeds (%)		Traditional Approach		Multi-network Approach		% Improvement	% Improvement
		Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)
100	(11.53%)	341.49	311.25	391.21	364.43	14.56%	17.09%
200	(23.07%)	575.44	553.28	597.20	584.45	3.78%	5.63%
300	(34.60%)	751.76	706.37	786.18	776.89	4.58%	9.98%
400	(46.14%)	893.33	867.93	943.24	935.78	5.59%	7.82%
500	(57.67%)	1016.37	971.49	1064.00	1055.76	4.69%	8.68%
600	(69.20%)	1126.38	1103.84	1153.59	1140.73	2.42%	3.34%
700	(80.74%)	1213.20	1195.64	1225.62	1219.01	1.02%	1.95%
800	(92.27%)	1243.19	1237.60	1246.44	1240.64	.26%	.25%
867	(100.00%)	1258.24	1258.24	1258.24	1258.24	.00%	.00%

ROBUSTNESS CHECK A.4:
RESULTS WITH DIFFERENT FUNCTIONAL FORM FOR WEIGHTED NETWORK CONSTRUCTION FUNCNTION $g(.)$

TABLE A.4.1
ESTIMATION RESULTS OF DIFFERENT MODELS

Variables	Multi-network Approach	
	Model 1	Model 2
Construction of weighted network		
Economic	.04	**1.05**
	(-.15, .23)	(.57, 1.55)
Social	**.30**	**-.65**
	(.15, .48)	(-.85, -.40)
Religious	-.01	.25
	(-.30, .29)	(-.26, .63)
Family	**-.37**	.53
	(-.62, -.10)	(-.29, .81)
Diffusion Equation		
Constant	**.45**	**1.64**
	(.05, .86)	(1.13, 2.15)
Control variables		
Percentage of Seeds	**-1.08**	**-12.05**
	(-12.33, -7.74)	(-15.08, -8.94)
Roof_1 (Thatch)	-.02	.09
	(-.34, .30)	(-.33, .51)
Roof_2 (Tile)	-.13	-.02
	(-.45, .18)	(-.45, .39)
Roof_3 (Stone)	-.26	-.29
	(-.59, .05)	(-.73, .15)
Roof_4 (Sheet)	-.19	-.05
	(-.54, .16)	(-.51, .41)
No of Rooms	**-.05**	.09
	(-.11, -.00)	(.03, .15)
Electricity	**-.36**	**-.30**
	(-.53, -.19)	(-.51, -.08)
Latrine	.10	.14
	(-.07, .26)	(-.07, .35)
House	.00	-.01
	(-.28, .28)	(-.35, .34)
Degree	**.23**	
	(.21, .25)	
Eigenvector		**3.24**
		(1.29, 5.20)
Variance of Error	**1.55**	**2.37**
	(1.42,1.69)	(2.19, 2.58)
LMD	**-1645.69**	**-2306.01**

Note: 95 percent posterior intervals are reported between brackets.

TABLE A.4.2
IN- AND OUT-OF-SAMPLE FIT STATISTICS

Criterion	In-Sample				Out-of-Sample			
	Traditional Approach		Multi-network Approach		Traditional Approach		Multi-network Approach	
	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)
MAD	.95	.99	.89	1.07	.97	1.01	.89	1.10
MAPE	.65	.68	.61	.75	.66	.69	.63	.76
RMSE	1.31	1.36	1.26	1.47	1.33	1.39	1.33	1.50
R-squared	.41	.39	.44	.07	.40	.38	.42	.06

Notes: MAD – Mean Absolute Deviance; MAPE: Mean Absolute Percentage of Error; RMSE: Root Mean Squared Error.

TABLE A.4.3
HOLDOUT SEEDING PERFORMANCE

No. of Selected Seeds (%)		Traditional Approach		Multi-network Approach		% Improvement	% Improvement
		Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)
100	(11.53%)	341.02	311.17	351.43	255.49	3.05%	-17.89%
200	(23.07%)	576.36	553.21	579.19	433.30	0.49%	-21.68%
300	(34.60%)	752.60	706.47	761.73	605.65	1.21%	-14.27%
400	(46.14%)	893.71	867.93	922.45	747.29	3.22%	-13.90%
500	(57.67%)	1016.34	971.03	1053.92	885.55	3.70%	-8.80%
600	(69.20%)	1126.24	1103.83	1151.91	1005.56	2.28%	-8.90%
700	(80.74%)	1212.76	1195.50	1219.83	1110.75	.58%	-7.09%
800	(92.27%)	1233.04	1227.41	1242.71	1168.50	.74%	-4.80%
867	(100.00%)	1258.24	1258.24	1258.24	1258.24	.00%	.00%

ROBUSTNESS CHECK A.5:
RESULTS WITH SEPARATE NETWORK APPROACH

TABLE A.5.1
ESTIMATION RESULTS OF DIFFERENT MODELS

Variables	**Separate Network Approach** Model 1	Model 2
Diffusion Equation		
Constant	**1.86**	**1.88**
	(1.44, 2.28)	(1.45, 2.30)
Control variables		
Percentage of Seeds	**-11.40**	**-11.67**
	(-13.90, -8.94)	(-14.19, -9.08)
Roof_1 (Thatch)	.00	.02
	(-.35, .36)	(-.33, .37)
Roof_2 (Tile)	-.11	-.10
	(-.47, .23)	(-.45, .25)
Roof_3 (Stone)	-.188	-.17
	(-.55, .16)	(-.53, .20)
Roof_4 (Sheet)	-.19	-.19
	(-.55, .20)	(-.56, .19)
No of Rooms	-.04	-.03
	(-.09, .02)	(-.08, .02)
Electricity	**-.35**	**-.37**
	(-.52, -.17)	(-.55, -.19)
Latrine	.04	.05
	(-.13, .21)	(-.12, .22)
House	-.08	-.09
	(-.35, .21)	(-.37, .19)
Network centrality measures		
Degree (Economic)	**.19**	
	(.05, .39)	
Degree (Social)	**.50**	
	(.36, .64)	
Degree (Religious)	-.02	
	(-.08, .05)	
Degree (Family)	.08	
	(-.10,.27)	
Eigenvector (Economic)		**.52**
		(.27, .78)
Eigenvector (Social)		**.86**
		(.62, 1.11)
Eigenvector (Religious)		-.12
		(-.30, .05)
Eigenvector (Family)		.16
		(-.27, .59)
Variance of Error	**1.62**	**1.56**
	(1.46, 1.72)	(1.51, 1.78)
LMD	**-1646.53**	**-1646.86**

Note: 95 percent posterior intervals are reported between brackets.

TABLE A.5.2
IN- AND OUT-OF-SAMPLE FIT STATISTICS

Criterion	In-Sample				Out-of-Sample			
	Separate Network Approach		Multi-network Approach		Separate Network Approach		Multi-network Approach	
	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)
MAD	.95	.99	.87	.89	.96	.99	.88	.90
MAPE	.64	.66	.60	.61	.65	.68	.60	.61
RMSE	1.31	1.34	1.24	1.26	1.32	1.35	1.26	1.28
R-squared	.46	.43	.54	.51	.43	.41	.53	.51

Notes: MAD – Mean Absolute Deviance; MAPE: Mean Absolute Percentage of Error; RMSE: Root Mean Squared Error.

TABLE A.5.3
HOLDOUT SEEDING PERFORMANCE

No. of Selected Seeds (%)		Separate Network		Multi-network		% Improvement	% Improvement
		Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)
100	(11.53%)	300.12	278.63	341.19	310.74	13.69%	11.52%
200	(23.07%)	547.58	518.43	575.65	553.87	5.13%	6.84%
300	(34.60%)	733.63	699.98	752.21	706.54	2.53%	0.94%
400	(46.14%)	884.52	839.37	893.78	868.54	1.05%	3.47%
500	(57.67%)	1010.95	970.90	1016.40	980.97	.54%	1.03%
600	(69.20%)	1126.05	1091.53	1127.01	1103.92	.09%	1.14%
700	(80.74%)	1212.92	1195.61	1213.87	1198.87	.08%	.27%
800	(92.27%)	1229.19	1227.65	1233.73	1233.01	.37%	.44%
867	(100.00%)	1258.24	1258.24	1258.24	1258.24	.00%	.00%

ROBUSTNESS CHECK A.6:
RESULTS WITH A PRIORI WEIGHTED NETWORK APPROACH (EQUAL WEIGHTS)

TABLE A.6.1
ESTIMATION RESULTS OF DIFFERENT MODELS

Variables	**A Priori Weighted Approach**	
	Model 1	Model 2
Diffusion Equation		
Constant	**.44**	**.47**
	(.00, .86)	(.04, .91)
Control variables		
Percentage of Seeds	**-11.67**	**-11.87**
	(-14.19, -9.19)	(-14.44, -9.23)
Roof_1 (Thatch)	-.01	.02
	(-.34, .33)	(-.34, .39)
Roof_2 (Tile)	-.114	-.07
	(-.45, .22)	(-.43, .29)
Roof_3 (Stone)	-.20	-.16
	(-.55, .16)	(-.52, .19)
Roof_4 (Sheet)	-.17	-.16
	(-.54, .20)	(-.54, .23)
No of Rooms	-.04	-.03
	(-.09, .01)	(-.089, .018)
Electricity	**-.35**	**-.37**
	(-.52, -.17)	(-.56, -.19)
Latrine	.04	.05
	(-.13,.20)	(-.125, .223)
House	-.043	-.050
	(-.321, .229)	(-.333, .244)
Degree	**.80**	
	(.73, .86)	
Eigenvector		**1.56**
		(1.43, 1.70)
Variance of Error	**1.54**	**1.68**
	(1.48, 1.74)	(1.54, 1.81)
LMD	**-1649.00**	**-1653.22**

Note: 95 percent posterior intervals are reported between brackets.

TABLE A.6.2
IN- AND OUT-OF-SAMPLE FIT STATISTICS

Criterion	In-Sample				Out-of-Sample			
	A Priori Weighted Approach		Multi-network Approach		A Priori Weighted Approach		Multi-network Approach	
	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)
MAD	.96	.99	.87	.89	.97	1.01	.88	.90
MAPE	.66	.69	.60	.61	.66	.70	.60	.61
RMSE	1.32	1.36	1.24	1.26	1.34	1.39	1.26	1.28
R-squared	.41	.39	.54	.51	.40	.38	.53	.51

Notes: MAD – Mean Absolute Deviance; MAPE: Mean Absolute Percentage of Error; RMSE: Root Mean Squared Error.

TABLE A.6.3
HOLDOUT SEEDING PERFORMANCE

No. of Selected Seeds (%)		A Priori Weighted		Multi-network		% Improvement	% Improvement
		Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)
100	(11.53%)	333.20	310.12	391.33	348.38	17.45%	12.34%
200	(23.07%)	564.77	559.14	598.09	569.56	5.90%	1.86%
300	(34.60%)	751.48	730.96	787.21	766.16	4.75%	4.82%
400	(46.14%)	895.17	867.73	944.68	925.04	5.53%	6.60%
500	(57.67%)	1025.91	998.40	1064.22	1039.92	3.73%	4.16%
600	(69.20%)	1127.58	1117.39	1153.95	1144.41	2.34%	2.42%
700	(80.74%)	1217.31	1203.91	1227.06	1214.49	.80%	.88%
800	(92.27%)	1231.03	1231.41	1240.16	1237.54	.74%	.50%
867	(100.00%)	1258.24	1258.24	1258.24	1258.24	.00%	.00%

WEB APPENDIX B
ALTERNATIVE VALIDATION OF RECOVERED IMPORTANCE OF RELATIONSHIP CHARACTERISTICS FOR APPLICATION II

In the second empirical application, we used 1) the number of messages exchanged, and 2) relationship duration as relationship characteristics to validate the multi-network approach. Corroborating the findings of Aral and Walker (2014), we found that the number of messages exchanged had a positive effect on tie strength, while relationship duration negative. In this validation, these two relationship characteristics were directly used as variables in our model. An alternative way to validate our approach is to determine the weights of connections using different relationship characteristics, and subsequently correlate these recovered weights with 1) the number of messages exchanged, and 2) relationship duration[8], as we describe next.

To execute this alternative validation, we constructed relationship characteristics from demographics and network measures in Empirical Application II. More specifically, we used gender, age and degree centrality of users to construct three relationship characteristics: 1) same gender (i.e., relationships in which two connected users have the same gender), 2) age difference (i.e., the absolute difference in age between two connected users), and 3) degree difference (i.e., the absolute difference in the number of friends between two connected users). Using these three relationship characteristics, we formulated the model following equations 9 and 10, and used the same estimation procedure.

Table B.1 presents the estimation results of the multi-network and the traditional approach for both degree (Model 1) and eigenvector centrality (Model 2). Corroborating our previous results, the multi-network approach better fits the data than the traditional approach (LMD Model 1: -2614.26 vs. -2623.46 and Model 2: -2617.48 vs. -2623.97). Moreover, the multi-network approach again recommends degree centrality over eigenvector centrality. More importantly, using the estimated weights of Model 1 in the multi-network approach, we found 1) a positive correlation with the number of messages exchanged (correlation coefficient: .109 and p-value $< .000$) and 2) a negative correlation with relationship duration (correlation coefficient: -.022 and p-value $< .000$). Similarly, using the estimated weights of Model 2 in the multi-network Approach, we found 1) a positive correlation with the number of messages exchanged (correlation coefficient: .093 and p-value $< .000$) and 2) a negative correlation with relationship

[8] We thank an anonymous reviewer for this valuable suggestion.

duration (correlation coefficient: -.023 and p-value $< .000$). These findings imply that the recovered weights indeed correspond to the frequency and recency effects (Aral and Walker 2014) and thus further validate the multi-network approach.

TABLE B.1

INFORMATION PROPAGATION: ESTIMATION RESULTS

Variables	**Traditional Approach**		**Multi-network Approach**	
	Model 1	Model 2	Model 1	Model 2
Construction of Weighted Network				
Same Gender			1.22	1.45
			(-1.75, 3.85)	(-1.35, 4.13)
Age Difference			.04	.02
			(-.47, .31)	(-.40, .44)
Degree Difference			**-.60**	**-.67**
			(-.80, -.37)	(-.92, -.32)
Diffusion Equation				
Intercept	**.90**	**.93**	**.85**	**.87**
	(.85, .94)	(.89, .97)	(.79, .91)	(.81, .93)
Age	-.02	-.02	-.01	-.01
	(-.03, .00)	(-.04, .00)	(-.03, .01)	(-.03, .01)
Gender	.00	.00	.01	.00
	(-.04, .04)	(-.04, .05)	(-.03, .05)	(-.04, .04)
Membership Duration	.10	.11	.09	.09
	(-.02, .22)	(-.01, .23)	(-.03, .21)	(-.03, .21)
Seeding Time	**-.06**	**-.06**	**-.06**	**-.06**
	(-.08, -.04)	(-.08, -.04)	(-.08, -.04)	(-.08, -.04)
Network Measures				
Degree	**.04**		**.06**	
	(.03, .06)		(.04, .08)	
Eigenvector		**.07**		**.07**
		(.04, .09)		(.04, .09)
LMD	-2623.46	-2623.97	-2614.26	-2617.48

Note: 95 percent posterior intervals are reported between brackets.

WEB APPENDIX C
ROBUSTNESS CHECKS FOR ONLINE INFORMATION PROPAGATION

ROBUSTNESS CHECK C.1:
RESULTS WITH SEPARATE NETWORK APPROACH

TABLE C.1.1
ESTIMATION RESULTS OF DIFFERENT MODELS

Variables	**Separate Network Approach**	
	Model 1	Model 2
Diffusion Equation		
Intercept	**.86**	**.93**
	(.81, .90)	(.88, .97)
Age	-.01	-.02
	(-.03, .01)	(-.03, .00)
Gender	.00	.00
	(-.04, .04)	(-.03, .05)
Membership Duration	.06	.10
	(-.06, .17)	(-.02, .22)
Seeding Time	**-.06**	**-.06**
	(-.08, -.04)	(-.08, -.04)
Network Measures		
Degree (No. of Messages Exchanged)	**.07**	
	(.06, .08)	
Degree (Relationship Duration)	**.03**	
	(.02, .05)	
Eigenvector (No. of Messages Exchanged)		**.36**
		(.19, .53)
Eigenvector (Relationship Duration)		**.07**
		(.04, .09)
LMD	-2576.32	-2615.90

Note: 95 percent posterior intervals are reported between brackets.

TABLE C.1.2
IN- AND OUT-OF-SAMPLE FIT STATISTICS

Criterion	In-Sample				Out-of-Sample			
	Separate Network Approach		Multi-network Approach		Separate Network Approach		Multi-network Approach	
	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)
MAD	.782	.811	.752	.788	.822	.826	.771	.810
MAPE	.367	.383	.358	.377	.381	.388	.361	.381
RMSE	1.905	2.003	1.742	1.791	2.055	2.151	1.785	1.836
R-squared	.041	.029	.090	.074	.037	.024	.082	.063

Notes: MAD – Mean Absolute Deviance; MAPE: Mean Absolute Percentage of Error; RMSE: Root Mean Squared Error; R-squared: McFadden's Pseudo R-squared.

TABLE C.1.3
HOLDOUT SEEDING PERFORMANCE

No. of Selected Seeds (%)		Separate Network		Multi-network		% of Improvement	% of Improvement
		Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)
50	(4.12%)	172.66	133.16	213.67	166.04	23.75%	24.69%
100	(8.23%)	286.28	260.02	323.89	294.68	13.14%	13.33%
150	(12.35%)	412.49	399.22	434.83	425.27	5.42%	6.52%
200	(16.46%)	506.92	491.90	520.80	512.48	2.74%	4.18%
250	(20.58%)	595.12	569.30	615.56	609.92	3.43%	7.13%
450	(37.04%)	950.35	949.70	963.40	983.11	1.37%	3.52%
650	(53.50%)	1270.37	1254.68	1297.03	1283.12	2.10%	2.27%
850	(69.96%)	1549.82	1535.33	1572.42	1572.95	1.46%	2.45%
1050	(86.42%)	1818.12	1810.58	1827.66	1813.24	.52%	.15%
1215	(100.00%)	2014.50	2014.50	2014.50	2014.50	.00%	.00%

ROBUSTNESS CHECK C.2:
RESULTS WITH A PRIORI WEIGHTED NETWORK APPROACH (EQUAL WEIGHTS)

TABLE C.2.1
ESTIMATION RESULTS OF DIFFERENT MODELS

Variables	**A Priori Weighted Approach**	
	Model 1	Model 2
Diffusion Equation		
Intercept	**.90**	**.93**
	(.85, .95)	(.89, .97)
Age	-.02	-.02
	(-.03, .00)	(-.04, .00)
Gender	.00	.00
	(-.04, .04)	(-.04, .05)
Membership Duration	.10	.11
	(-.02, .22)	(-.01, .23)
Seeding Time	**-.06**	**-.06**
	(-.08, -.04)	(-.08, -.04)
Network Measures		
Degree	**.04**	
	(.03, .06)	
Eigenvector		**.07**
		(.04, .09)
LMD	-2624.34	-2624.42

Note: 95 percent posterior intervals are reported between brackets.

TABLE C.2.2
IN- AND OUT-OF-SAMPLE FIT STATISTICS

Criterion	In-Sample				Out-of-Sample			
	A Priori Weighted Approach		Multi-network Approach		A Priori Weighted Approach		Multi-network Approach	
	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)
MAD	.815	.817	.752	.788	.832	.833	.771	.810
MAPE	.384	.385	.358	.377	.390	.392	.361	.381
RMSE	2.045	2.044	1.742	1.791	2.169	2.171	1.785	1.836
R-squared	.024	.024	.090	.074	.023	.023	.082	.063

Notes: MAD – Mean Absolute Deviance; MAPE: Mean Absolute Percentage of Error; RMSE: Root Mean Squared Error; R-squared: McFadden's Pseudo R-squared.

TABLE C.2.3
HOLDOUT SEEDING PERFORMANCE

No. of Selected Seeds (%)		A Priori Weighted		Multi-network		% of Improvement	% of Improvement
		Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 1 (Degree)	Model 2 (Eigenvector)
50	(4.12%)	105.95	125.80	213.53	166.39	101.54%	32.27%
100	(8.23%)	262.17	257.70	324.06	294.63	23.61%	14.33%
150	(12.35%)	381.30	399.17	434.88	425.44	14.05%	6.58%
200	(16.46%)	479.99	491.54	520.67	511.93	8.48%	4.15%
250	(20.58%)	568.77	570.45	615.69	610.23	8.25%	6.97%
450	(37.04%)	960.17	950.17	963.37	983.12	0.33%	3.47%
650	(53.50%)	1256.55	1254.93	1296.96	1283.34	3.22%	2.26%
850	(69.96%)	1547.69	1533.38	1572.98	1572.98	1.63%	2.58%
1050	(86.42%)	1821.67	1811.50	1827.16	1813.13	.30%	.09%
1215	(100.00%)	2014.50	2014.50	2014.50	2014.50	.00%	.00%

WEB APPENDIX D
TESTING THE MODERATION OF NUMBER OF STEPS WITH DIFFUSION CENTRALITY

In both Empirical Applications I and II, we found that models with degree centrality better predicted reach than those with eigenvector centrality. However, as argued by Borgatti (2005), degree and eigenvector centrality describe two extremes of the transmission process. Degree centrality assumes an underlying transmission process where only direct connections are involved, while eigenvector centrality walks of unrestricted lengths. Therefore, it is possible that the "best fitting" centrality measures lies somewhere in between[9]. To further explore this possibility, we estimated additional models for both applications using the diffusion centrality (DC) measure proposed by Banerjee et al. (2013). Diffusion centrality is defined as below:

(D1) $$DC = \left(\sum_{t=1}^{T}[pW]^{t}\right)\cdot\mathbf{1},$$

where p is the probability that a consumer propagates information to their neighbors, T captures the number of steps of the diffusion process, and **1** is a vector of ones. In (D1), T represents the assumption on the number of steps in the diffusion process, with $T = 1$ corresponding to degree centrality and $T \to \infty$ eigenvector centrality.

In order to apply diffusion centrality in Equation (D1), we needed to determine the probability p that consumers shared information. For Empirical Application I, we used the results from Banerjee et al. (2014, p. 19), which set $p = .20$. We then estimated two models for both the multi-network and traditional approach by setting the number of periods T to either 2 (Model 3) or 3 (Model 4) (see Table D.1). For Empirical Application II, we set the forwarding probability p to the average sharing probability of an adopter, which equaled .0258:

(D2) $$p = \frac{1}{N_a}\sum_{i=1}^{N_a}\frac{F_i}{N_i},$$

where N_a is the number of adopters, F_i the number of friends adopter i forwarded the information to, and N_i the total number of friends of adopter i. Similar to Application I, we set the number of periods T to 2 (Model 3) or 3 (Model 4) (see Table D.3).

For both empirical applications, the multi-network approach outperforms the traditional approach in terms of model fit (Empirical Application I: Model 3 LMD: -1635.06 vs. -1643.08;

[9] We thank an anonymous reviewer for this valuable suggestion.

Model 4 LMD: -1638.15 vs. -1642.98; Empirical Application II: Model 3 LMD: -2497.18 vs. -2622.92 and Model 4 LMD: -2497.86 vs. -2622.82, see Table D.1 and D.3, respectively). Moreover, for Empirical Application I, the multi-network approach finds that degree centrality best describes the underlying diffusion process (LMD degree centrality: -1627.13 vs. -1635.06 for $T = 2$, the best fitting alternative model). These results are further supported by the in-sample fit statistics presented in Table D.2. Regarding Empirical Application II, the LMD does not favor a specific network measure for T from 1 to 3 (i.e., LMD degree centrality or $T = 1$: -2497.71 vs. $T = 2$: -2497.18 vs. $T = 3$: -2497.86, see Table D.3)[10], while eigenvector centrality is significantly worse (LMD: -2535.25). This is also supported by the in-sample fit statistics presented in Table D.4, with models of T set from 1 to 3 almost identical.

In sum, these follow-up tests suggest that for these two empirical applications, weighted degree centrality describes the diffusion process better (Application I) or at least as good (Application II) as diffusion centrality measures with $T = 2$ and 3.

[10] Kass and Raftery (1995) recommend a difference of at least 1 to infer a strong difference. (Kass and Raftery (1995), “Bayes Factors,” *Journal of the American Statistical Association*, 90 (430), p. 773-795.

TABLE D.1
MICROFINANCE DIFFUSION: ESTIMATION RESULTS

Variables	**Traditional Approach**				**Multi-network Approach**			
	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
Construction of weighted network								
Economic					**-.23**	**.30**	**-.21**	**.10**
					(-.38, -.05)	(.18, .44)	(-.36, -.07)	(.06, .16)
Social					**.24**	**.36**	**.42**	**.46**
					(.085, .385)	(.24, .47)	(.30, .54)	(.34, .59)
Religious					.48	.20	.42	.39
					(-.02, .93)	(-.01, .43)	(-.09, .91)	(-.09, .86)
Family					-0.10	.04	-.01	-.02
					(-.20, .02)	(-.05, .21)	(-.10, .09)	(-.09, .11)
Diffusion Equation								
Constant	**.50**	**.67**	**.49**	**.57**	**.52**	**.63**	**.51**	**.54**
	(.07, .92)	(.25, 1.11)	(.06, .94)	(.15, 1.00)	(.094, .95)	(.20, 1.07)	(.09, .94)	(.11, .97)
Control variables								
Percentage of Seeds	**-11.69**	**-12.10**	**-12.18**	**-12.21**	**-11.74**	**-12.83**	**-12.09**	**-12.01**
	(-14.21, -9.77)	(-14.61, -9.35)	(-14.67, -9.68)	(-14.76, -9.60)	(-14.10, -9.19)	(-15.38, -10.35)	(-14.58, -9.60)	(-14.46, -9.53)
Roof_1 (Thatch)	-.06	.01	-.05	-.04	-.048	-.05	-.07	-.07
	(-.39, .28)	(-.35, .37)	(-.41, .29)	(-0.39, 0.31)	(-.40, .29)	(-0.40, 0.31)	(-.42, .27)	(-.41, .28)
Roof_2 (Tile)	-.14	-.08	-.12	-.10	-.15	-.15	-.17	-.17
	(-.48, .19)	(-.43, .28)	(-.46, .22)	(-.44, .25)	(-.49, .18)	(-.50, .20)	(-.51, .16)	(-.50, .18)
Roof_3 (Stone)	-.24	-.18	-.21	-.20	-.24	-.23	-.26	-.25
	(-.58, .12)	(-.53, .17)	(-.56, .14)	(-.54, .14)	(-.58, .11)	(-.58, .12)	(-.61, .08)	(-.60, .10)
Roof_4 (Sheet)	-.22	-.14	-.21	-.19	-.23	-.21	-.24	-.23
	(-.59, .15)	(-.52, .25)	(-.58, .17)	(-.57, .19)	(-.59, .15)	(-.59, .17)	(-.62, .13)	(-.61, .14)
No of Rooms	-.04	-.03	-.05	-.04	-.04	-.03	-.04	-.04
	(-.10, .01)	(-.09, .02)	(-.10, .00)	(-.10, .01)	(-.09, .01)	(-.08, .03)	(-.09, .01)	(-.09, .01)
Electricity	**-.36**	**-.40**	**-.39**	**-.39**	**-.35**	**-.35**	**-.35**	**-.35**
	(-.54, -.19)	(-.58, -.21)	(-.56, -.21)	(-.57, -.21)	(-.53, -.18)	(-.53, -.17)	(-.53, -.18)	(-.52, -.17)
Latrine	.03	.06	.03	.04	.03	.07	.02	.03
	(-.14, .19)	(-.11, .23)	(-.14, .19)	(-.13, .21)	(-.14, .20)	(-.01, .23)	(-.15, .19)	(-.13, .20)
House	-.05	.04	-.06	-.05	-.08	-.03	-.08	-.06
	(-.33, .21)	(-.26, .33)	(-.35, .23)	(-.33, .24)	(-.35, .21)	(-.32, .25)	(-.36, .19)	(-.34, .21)
Network centrality measures								
Degree ($T = 1$)	**.81**				**.82**			
	(.74, .86)				(.75, .88)			
Diffusion ($T = 2$)			**.78**				**.81**	
			(.72, .84)				(.75, .87)	
Diffusion ($T = 3$)				**.75**				**.80**
				(.69, .81)				(.74, .87)
Eigenvector ($T = \infty$)		**1.36**				**1.07**		
		(1.24, 1.48)				(.94, 1.19)		
Variance of Error	**1.56**	**1.69**	**1.59**	**1.63**	**1.55**	**1.61**	**1.55**	**1.56**
	(1.44, 1.71)	(1.56, 1.84)	(1.47, 1.73)	(1.50, 1.77)	(1.43, 1.68)	(1.48, 1.75)	(1.43, 1.68)	(1.44, 1.70)
LMD	-1647.19	-1642.28	-1643.08	-1642.98	-1627.13	-1640.61	-1635.06	-1638.15

TABLE D.2
MICROFINANCE DIFFUSION: IN-SAMPLE FIT STATISTICS

Criterion	**In-Sample**							
	Traditional Approach				Multi-network Approach			
	Model 1 (Degree)	Model 2 (Eigenvector)	Model 3 (Diffusion $T = 2$)	Model 4 (Diffusion $T = 3$)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 3 (Diffusion $T =$ 2)	Model 4 (Diffusion $T =$ 3)
MAD	.99	.95	.95	.95	.87	.89	.89	.89
MAPE	.68	.65	.65	.65	.60	.61	.61	.61
RMSE	1.36	1.31	1.32	1.31	1.24	1.26	1.25	1.25
R-squared	.39	.41	.41	.41	.54	.51	.52	.51

Notes: MAD – Mean Absolute Deviance; MAPE: Mean Absolute Percentage of Error; RMSE: Root Mean Squared Error; R-squared: McFadden's Pseudo R-squared.

TABLE D.3
INFORMATION PROPAGATION: ESTIMATION RESULTS

Variables	**Traditional Approach**				**Multi-network Approach**			
	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
Construction of Weighted Network								
Number of Messages Exchanged					**4.76**	**1.46**	**4.48**	**3.98**
					(1.91, 6.25)	(1.37, 1.56)	(1.44, 7.58)	(.99, 8.68)
Relationship Duration					**-.44**	**-.23**	**-.40**	**-.41**
					(-.54, -.34)	(-0.33, -.14)	(-.54, -.32)	(-.58, -.31)
Diffusion Equation								
Intercept	**.90**	**.93**	**.91**	**.91**	**.85**	**.91**	**.85**	**.85**
	(.85, .94)	(.89, .97)	(.86, .95)	(.87, .96)	(.80, .89)	(.86, .95)	(.80, .89)	(.80, .89)
Age	-.02	-.02	-.02	-.02	**-.02**	-.02	**-.02**	**-.02**
	(-.03, .00)	(-.04, .00)	(-.03, .00)	(-.04, .00)	(-.04, -.00)	(-.03, .00)	(-.04, -.00)	(-.04, -.00)
Gender	.00	.00	.00	.01	.00	.00	.00	.00
	(-.04, .04)	(-.04, .05)	(-.04, .05)	(-.03,.05)	(-.03, .04)	(-.04, .04)	(-.03, .04)	(-.03, .04)
Membership Duration	.10	.11	.10	.10	.06	.08	.06	.06
	(-.02, .22)	(-.01, .23)	(-.02, .22)	(-.02, .22)	(-.05, .17)	(-.03, .20)	(-.05, .18)	(-.05, .18)
Seeding Time	**-.06**	**-.06**	**-.06**	**-.06**	**-.07**	**-.06**	**-.07**	**-.07**
	(-.08, -.04)	(-.08, -.04)	(-.08, -.04)	(-.08, -.04)	(-.08, -.05)	(-.08, -.04)	(-.08, -.05)	(-.08, -.05)
Network Measures								
Degree ($T = 1$)	**.04**				**.06**			
	(.03, .06)				(.05, .07)			
Diffusion ($T = 2$)			**.04**				**.06**	
			(.03, .06)				(.05, .07)	
Diffusion ($T = 3$)				**.04**				**.06**
				(.02, .06)				(.05, .07)
Eigenvector ($T = \infty$)		**.07**				**.10**		
		(.04, .09)				(.08, .12)		
LMD	-2623.46	-2623.97	-2622.92	-2622.82	-2497.71	-2535.25	-2497.18	-2497.86

Note: 95 percent posterior intervals are reported between brackets.

TABLE D.4
INFORMATION PROPAGATION: IN-SAMPLE FIT STATISTICS

	In-Sample							
	Traditional Approach				Multi-network Approach			
Criterion	Model 1 (Degree)	Model 2 (Eigenvector)	Model 3 (Diffusion $T = 2$)	Model 4 (Diffusion $T = 3$)	Model 1 (Degree)	Model 2 (Eigenvector)	Model 3 (Diffusion $T = 2$)	Model 4 (Diffusion $T = 3$)
MAD	.815	.817	.814	.814	.752	.788	.751	.752
MAPE	.383	.384	.383	.383	.358	.377	.358	.358
RMSE	2.009	2.009	2.005	2.005	1.742	1.791	1.742	1.742
R-squared	.026	.030	.026	.026	.090	.074	.090	.090

Notes: MAD – Mean Absolute Deviance; MAPE: Mean Absolute Percentage of Error; RMSE: Root Mean Squared Error; R-squared: McFadden's Pseudo R-squared.