ASIM ANSARI, ODED KOENIGSBERG, and FLORIAN STAHL*

Firms are increasingly seeking to harness the potential of social networks for marketing purposes. Therefore, marketers are interested in understanding the antecedents and consequences of relationship formation within networks and in predicting interactivity among users. The authors develop an integrated statistical framework for simultaneously modeling the connectivity structure of multiple relationships of different types on a common set of actors. Their modeling approach incorporates several distinct facets to capture both the determinants of relationships and the structural characteristics of multiplex and sequential networks. They develop hierarchical Bayesian methods for estimation and illustrate their model with two applications: The first application uses a sequential network of communications among managers involved in new product development activities, and the second uses an online collaborative social network of musicians. The authors' applications demonstrate the benefits of modeling multiple relations jointly for both substantive and predictive purposes. They also illustrate how information in one relationship can be leveraged to predict connectivity in another relation.

Keywords: social networks, online networks, Bayesian, multiple relationships, sequential relationships

Modeling Multiple Relationships in Social Networks

The rapid growth of online social networks has led to a resurgence of interest in the marketing field in studying the structure and function of social networks. A better understanding of social networks can enable managers to comprehend and predict economic outcomes (Granovetter 1985) and, in particular, to interface with both external and internal actors. Online brand communities, which are composed of users interested in particular products or brands, allow such an external interface with customers. Such communities not only help firms interact with customers and prospects but also enable customers to communicate and exchange information with each other, consequently increasing the value that can be derived from a firm's products.

Similarly, firms forge alliances and enter into collaborative relationships with other firms for coproduction and social commerce (Stephen and Toubia 2010; Van den Bulte and Wuyts 2007) using interorganizational networks. Within the firm, intraorganizational networks of managers play a crucial role in cross-functional integration, as is the case with networks of marketing and organizational professionals engaged in new product development (Van den Bulte and Moenaert 1998).

As Van den Bulte and Wuyts (2007) point out, network structure has implications for power, knowledge dissemination, and innovation within firms and for contagion and diffusion among customers. Thus, understanding and predicting the patterns of interactions and relationships among network members is an important first step in using them effectively for marketing purposes. The focus of social network analysis is on (1) explaining the determinants of relationship formation; (2) identifying well connected actors; and (3) capturing structural characteristics of the network as described by reciprocity, clustering, transitivity, and other measures of local and global structure using a combination of assortative, relational, and proximity perspectives (Rivera, Soderstrom, and Uzzi 2010).

Actors belonging to a social network connect with each other using multiple relationships, possibly of different

*Asim Ansari is the William T. Dillard Professor of Marketing (e-mail: maa48@columbia.edu), and Oded Koenigsberg is Barbara & Meyer Feldberg Associate Professor (e-mail: ok2018@columbia.ed), Columbia Business School, Columbia University. Florian Stahl is Assistant Professor, Department of Business Economics, University of Zurich (e-mail: florian.stahl@uzh.ch). Christophe Van den Bulte served as associate editor for this article.

types. In this article, we develop statistical models of multiple relationships that yield an understanding of the drivers of multiplex relationships and predict the connectivity structure of such multiplex networks.

Multiplexity of relationships can arise from different modes of interaction or because of different roles people play within a network setting. For example, in many online networks, members can form explicit friendship and business relations, exchange content and communicate with one another. The relationships that connect a group of actors can differ not only in their substantive content but also in their directionality and intensity. For example, some relationships are symmetric in nature, whereas others can be directed. Some relationships involve the flow of resources, thus necessitating a focus on the intensity of such weighed connections. Finally, multivariate patterns of connections can also arise from viewing the same relationship at different points, as is the case of sequential networks.

An understanding of multiplex patterns in network structures is important for marketers. For example, Tuli, Bharadwaj, and Kohli (2010) find that in a business-to-business setting, increasing multiplexity in relationships leads to an increase in sales and to a decrease in sales volatility to a customer. Multiplexity contributes to the total strength of a tie and increases the number of ways one can reciprocate favors (Van den Bulte and Wuyts 2007). Thus, it is relevant for identifying influential actors such as opinion leaders in diffusion contexts and powerful executives within intraorganizational networks. In analyzing sequential relationships, a multivariate analysis can help investigate the impact of managerial interventions on the relationship structure of a network across points. Sequential dyadic interactions are also useful for understanding the dynamics of power and cooperation in intraorganizational networks and in modeling long-term relationships between buyers and sellers in business markets (Iacobucci and Hopkins 1992). Finally, when marketers are interested in predicting relationship patterns, multiplexity allows leveraging information from one relationship to predict connections on other relationships.

Researchers can obtain a substantive understanding of multiplex relationships by simultaneously analyzing the multiple connections among the network actors. They can investigate whether these multiple relationships exhibit multiplex patterns that are characterized by the flow of multiple relationships in the same direction or whether they represent patterns of generalized exchange in which a tie in one direction on one relationship is reciprocated with a connection in the other direction using different relationships. However, most models of social networks analyze a single relationship among network members. When the lens is trained on a single relationship, an incomplete understanding of the nature of linkages can result. For example, it is unclear whether people play a similar role across multiple relationships. A joint analysis can also help uncover common antecedents that affect relationships. Moreover, if some relationships exhibit unique patterns, such uniqueness can emerge only when multiple relationships are contrasted.

In this article, we develop an integrated latent-variable framework for modeling multiple relationships. We make a methodological contribution to the social networking literature in both marketing and the wider social sciences by offering a rich framework for modeling multiple relationships of different types. Our modeling framework has several novel features when compared with previously proposed models for multirelational network data. Specifically, our framework can (1) model multiple relationships of different types (i.e., weighted, unweighted, undirected, and directed), (2) model sequential relationships, (3) leverage partial information from one network (or relationship) to predict connectivity in another relationship, (4) accommodate missing data in a natural way, (5) capture sparseness in weighted relationships, (6) incorporate sources of dyadic dependence, (7) account for higher order effects such as triadic effects, and (8) include continuous covariates. Although previous models have incorporated some of these aspects, we do not know of any research in the social network analysis literature that simultaneously incorporates all of them.

We illustrate the benefits of our approach using two applications. Our first application involves sequential network data that studies network structure over two points. We reanalyze data from Van den Bulte and Moenaert (1998) involving a network of research-and-development (R&D), marketing, and operations managers who are engaged in new product development. The data contain communications among these managers both before and after collocation of R&D teams. The results show that substantive conclusions can be affected if the full generality of our framework is not utilized. We also show how our methods can be used to leverage information from one relationship to predict the connections in another relationship.

In our second application, we use data from an online social networking site involving the interactions among a set of musicians. We model friendship, communications, and music download relationships within this network to show how a combination of directed and undirected, and weighted and unweighted, relationships can be modeled jointly. We analyze the determinants of these relationships and assess the importance of our model components in capturing different facets of the network structure. Our results show that artists exhibit similar network roles across the three relationships and that these relationships are mostly influenced by common antecedents. We also show that when dealing with weighted relationships (e.g., music downloads), it is crucial to jointly model both the incidence and the intensity of such relationships, rather than simply focusing on the intensity; otherwise, prediction and recovery of structural characteristics suffers appreciably.

We organize the rest of the article as follows: The next section provides a brief review of the marketing and statistical literature on social networks. Then, we present the components of our modeling framework and describe inference and identification of model parameters. The following two sections describe the two applications. Finally, we conclude with a discussion of our contributions and model limitations and outline future research possibilities.

LITERATURE REVIEW

Social network data offer considerable opportunities for research in marketing, as Van den Bulte and Wuyts (2007) identify in their expansive survey of the role and importance of social networks in the marketing field. Most research in marketing on social networks falls into one of two streams. In the first stream, researchers explore the impact of word of mouth on the behavior of others and thus are primarily concerned about the role of social interactions and contagion (Iyengar, Van den Bulte, and Valente 2011; Nair, Manchanda, and Bhatia 2010; Trusov, Bodapati, and Bucklin 2010; Watts and Dodds 2007).

Research in the second stream focuses on modeling network structure. Iacobucci and Hopkins's (1992) work is a pioneering contribution in this area. The focus here is on understanding the antecedents of relationship formation and in studying how interventions influence future connectivity (Van den Bulte and Moenaert 1998). The current study adds to this second stream of research by offering a comprehensive framework for modeling multivariate or sequential relations.

There is a rich history of statistical modeling of network structure within sociology and statistics that spans more than 70 years. More recent approaches stem from the log-linear p_1 model developed by Holland and Leinhardt (1981), which assumes independent dyads. Because the p_1 model is incapable of representing many structural properties of the data, the literature has proposed two general ways of capturing the dependence among the relationships. The first approach uses exponential random graph models or p* models (Frank and Strauss 1986; Pattison and Wasserman 1999; Robins et al. 2007; Snijders et al. 2006; Wasserman and Pattison 1996) that capture the dependence pattern in the network using a set of statistics that embody important structural characteristics of the network. However, care is necessary when using these models because parameter estimation sometimes suffers from model degeneracy, and how to handle this degeneracy is an active area of research. The second approach handles the dependence among the dyads using correlated random effects and latent positions in a Euclidean space for the individual people (Handcock, Raftery, and Tantrum 2007; Hoff 2005; Hoff, Raftery, and Handcock, 2002). In addition to these two approaches, multiple regression quadratic assignment procedure (MRQAP) methods (Dekker, Krackhardt, and Snijders 2007) have also been used in network analysis to account for dependence among dyads.

Exponential random graph models describe the network using a set of summary statistics, such as the total number of ties, the number of triangles, and the degree of distribution, among others. This is good for describing "global" properties and in assessing particular hypotheses of substantive interest, such as the extent of reciprocity or clustering and triadic closure. In contrast, latent space models capture the "local" structure by estimating a latent variable for each node in the network, which describes a person's position in the network. These models are thus suitable when the focus is on understanding the determinants of connectivity using covariates and in identifying influential people. The latent variable framework is capable of recovering the structural characteristics using a small set of model parameters (similar to nuisance parameters) and thus can be parsimonious in some contexts.

When researchers are interested in specific substantive hypotheses and when all relationships are binary in nature, they may prefer exponential random graph models. However, the latent variable framework can accommodate multiplex relationships of different types, including weighted relationships, and can also handle missing data in a straightforward fashion using data augmentation. Thus, it is preferable when interest is in analyzing such complex multivariate data structures.

Whereas the preceding methods explicitly model the network structure, MRQAP methods offer a nonparametric alternative for conducting permutation tests to assess covariate effects using multiple regression while correcting for the dependency and autocorrelation present in network data. The MRQAP approach is useful for continuous data. It can be used for binary relations using a linear probability model, and to a certain extent for count data; however, its effectiveness for multivariate relations of different types is not clear.

Sequential data can also be modeled using two approaches: a multivariate approach such as ours and the conditional, continuous-time approach popularized by Snijders (2005). The multivariate approach models the network at each point in time and thus is useful when one is interested in assessing the impact of interventions that occur between these discrete times. In contrast, the continuous-time approach is inherently dynamic, focusing on either edge-oriented or node-oriented dynamics, and can model the evolution of the network one edge at a time. However, this approach is limited to binary relations, whereas the multivariate approach that we use can handle both weighted and binary relationships.

In contrast to the current study, most models of social network structure analyze a single relationship, and to the best of our knowledge, none have incorporated the entire constellation of desirable model characteristics that we outlined in the introduction. In particular, there has been no work on using the latent space framework for modeling multivariate relationships or sequential data.

Although some researchers have modeled multiple relationships, these models either assume independence across dyads, which is restrictive, or limit attention to binary relations (Fienberg, Meyer and Wasserman 1985; Iacobucci 1989; Iacobucci and Wasserman 1987; Pattison and Wasserman 1999; Van den Bulte and Moenaert 1998). Thus, there is a need for an integrative framework for modeling multiple relationships of different types (i.e., binary or weighted) in a flexible way. The latent space framework offers such flexibility, and using it, we develop an integrated approach for multiple relationships in the following section.

MODELING FRAMEWORK

We develop a modeling framework for the simultaneous analysis of multiple relationships among a set of network actors. Our framework accommodates multiple relationships of different types and also enables us to simultaneously model the determinants of the relationships as well as the structural characteristics such as the extent of reciprocity or transitivity within each relationship and across relationships. When analyzing multiple relationships, structural characteristics of interest include those that account for *multiplex* patterns (i.e., flow of multiple relationships in the same direction) and *exchange*, in which a flow in one direction on one relationship is reciprocated with a flow in the other direction using a different relationship. Similarly, patterns of transitivity that involve more than one relationship can also be investigated. When focusing on the determinants of relationships, we can infer how the attributes of the network actors influence the formation of relationships between them. Here the interest is in understanding whether actors exhibit similar popularity and expansiveness across different relationships, and whether homophily governs relationship formation.

The multiple relationships describing a common set of actors can vary along different facets, such as existence, intensity, and directionality. A relationship is directed if we can distinguish the sender and receiver of the tie. For example, a communication relationship typically has a sender and a receiver. In contrast, relationships could be undirected, such as a collaboration relationship. In modeling both directed and undirected relationships, the focus of the analysis could be on modeling the existence of a relationship (i.e., the presence or absence of a tie) or on the intensity of a weighted tie (e.g., the intensity of the flow of resources between a pair of people). Our objective is to show how such disparate relationships can be jointly modeled within a common framework. In the following section, we describe formally our model.

Model Description

We describe our model using two relationships. Although these two relationships could represent a single relationship observed over different time periods, for the sake of generality, we describe a model for two distinct, directed relationships of different types.[1] These two relationships are observed over the same set of n actors.

Directed binary relationship. The first relationship is directed and binary. Thus, we can distinguish between the sender and the receiver of the tie and the sociomatrix matrix, $\mathbf{X_1}$, which shows that the incidence of ties among actors can, therefore, be asymmetric. We use the ordered pair of binary dependent variables $\{X_{ij1}, X_{ji1}\}$ to represent the presence or absence of ties for a pair of actors i and j. The variable X_{ij1} specifies the existence of interaction in the direction from i to j (i.e., $i \rightarrow j$), whereas X_{ji1} represents the presence of a tie in the opposite direction from j to i (i.e., $i \leftarrow j$).

Directed weighted relationship. The second relationship is directed and weighted (i.e., valued). The entries in the associated matrix, $\mathbf{X_2}$, indicate the bidirectional intensity of interaction between the different pairs of people. Here, we assume a count variable for the intensity, because this is consistent with our second application presented in the section "Online Social Network." However, our model can be adapted for continuous or ordinal measures of intensity. An ordered pair of count variables (X_{ij2}, X_{ji2}) can represent the observed intensity of interaction in the dyad, where the variable X_{ij2} specifies the strength of the interaction from i to j (i.e., $i \rightarrow j$), and X_{ji2} specifies the intensity in the reverse direction.

In modeling this weighted relationship, we deviate from the previous literature on social networks by jointly modeling both the existence and the intensity of the relationship. This allows us to distinguish between the mechanisms that drive the incidence from those that affect the intensity of relationships. In addition, it also accommodates a preponderance of zeros due to sparseness of ties. We can then ascertain whether a specification that directly models the intensity (such as a Poisson specification; e.g., Hoff 2005) is sufficient for weighted relationships. Therefore, we use a multivariate correlated hurdle count specification to jointly model both the incidence of the relationship within a dyad and the intensity of the relationship conditional on the existence of the tie. We model the incidence using the ordered pair of binary variables $(X_{ij2,I}, X_{ji2,I})$. Then, the magnitude of the relationship, conditional on its existence in a given direction, can be modeled using the positively valued truncated count variables $X_{ij2,S}$ or $X_{ji2,S}$.

Dyadic multigraph. Bringing together the two relationships, we can then specify the ${}^{n}C_2$ dyads in the multirelational social network using the dyad-specific random variables:

$$\mathbf{D}_{ij} = \begin{pmatrix} X_{ij1}, & X_{ji1} \\ X_{ij2,I}, & X_{ji2,I} \\ X_{ij2,S}, & X_{ji2,S} \end{pmatrix}, \quad i < j.$$

The relationships can be further specified in terms of underlying latent variables. The latent variable specification enables us to model these random variables in terms of dyad- and actor-specific covariates.

Latent variable specification. We use underlying latent utilities u_{ij1} for modeling the existence of a tie in the direction $(i \rightarrow j)$ and u_{ji1} in the reverse direction, for the first relationship:

$$(1) \qquad X_{ij1} = \begin{cases} 1, & \text{if } u_{ij1} > 0, \\ 0, & \text{otherwise,} \end{cases}$$

$$X_{ji1} = \begin{cases} 1, & \text{if } u_{ji1} > 0, \\ 0, & \text{otherwise.} \end{cases}$$

For the second relationship, let u_{ij2} and u_{ji2} represent the underlying utilities that characterizes the existence of the relationship. Again, we assume that

$$(2) \qquad X_{ij2,I} = \begin{cases} 1, & \text{if } u_{ij1} > 0, \\ 0, & \text{otherwise,} \end{cases}$$

$$X_{ji2,I} = \begin{cases} 1, & \text{if } u_{ji2} > 0, \\ 0, & \text{otherwise.} \end{cases}$$

We model the truncated counts, conditional on a tie in a given direction, using a Poisson distribution truncated at zero; in other words,

$$(3) \qquad X_{ij2,S} \sim \text{tPoisson}(\lambda_{ij}) \text{ if } u_{ij2} > 0, \text{ and}$$

$$X_{ji2,S} \sim \text{tPoisson}(\lambda_{ji}) \text{ if } u_{ji2} > 0,$$

where the λ_{ij} and λ_{ji} are the rate parameters of the Poisson.

Covariates and homophily. Each latent utility is composed of a systematic part involving the observed covariates and a stochastic part that incorporates unobserved variables. We distinguish between dyad-specific covariates and individual-specific covariates. Let $\mathbf{x}^d_{ij1}$ and $\mathbf{x}^d_{ij2}$ be vectors of dyad-specific covariates that influence the two relationships. These allow for homophily, which implies that people who

[1]We limit our model description to two relationships for clarity of presentation. Our approach, however, can be extended readily to more relationships, including undirected ones, as we do in our second application.

share observable characteristics tend to form connections. We also use individual-specific covariates for modeling directed relationships. Let $\mathbf{x}_{i1}$, and $\mathbf{x}_{i2}$ be the vector of individual i's covariates, respectively, for the two relationships. The vector $\mathbf{x}_{ij1} = (\mathbf{x}^d_{ij1}, \mathbf{x}_{i1}, \mathbf{x}_{j1})$ then represents all the covariates that affect the tie in the direction $(i \rightarrow j)$ for the first relationship, and $\mathbf{x}_{ji1} = (\mathbf{x}^d_{ji1}, \mathbf{x}_{j1}, \mathbf{x}_{i1})$ represents the covariates for the reverse direction $(i \leftarrow j)$. The sender and receiver effects are important to model the asymmetry in the two directions. For the weighted relationship, we can further distinguish between the incidence and intensity components. Therefore, we use covariate vectors $\mathbf{x}_{ij2,I} = (\mathbf{x}^d_{ij2,I}, \mathbf{x}_{i2,I}, \mathbf{x}_{j2,I})$ and $\mathbf{x}_{ji2,I} = (\mathbf{x}^d_{ji2,I}, \mathbf{x}_{j2,I}, \mathbf{x}_{i2,I})$ for the binary component, where, for example, $\mathbf{x}^d_{ij2,I}$ contains a subset of the dyad-specific variables in $\mathbf{x}^d_{ij2}$ that affect intensity, and $\mathbf{x}_{i2,I}$ is similarly a subset of $\mathbf{x}_{i2}$. We use analogously defined vectors $\mathbf{x}_{ij2,S}$ and $\mathbf{x}_{ji2,S}$ to model the Poisson rate parameters λ_{ij} and λ_{ji}.

Heterogeneity. The dyads cannot be considered independently, because multiple dyads share a common actor either as a sender or receiver. Accounting for such dependence is important for obtaining proper inferences about substantive issues. Therefore, we use heterogeneous and correlated random effects to account for the dependence structure. Whereas for undirected relationships, a single random effect is needed, for directed relationships, we can use two distinct random effects per actor to distinguish between *expansiveness*, which is the propensity to "send" ties and popularity, or *attractiveness*, which is the propensity to "receive" ties. The expansiveness parameter α_i captures the outdegree, which is the number of connections emanating from an individual i, and the attractiveness parameter β_i captures the indegree, which is the number of connections impinging on an individual. Thus, for the directed binary relationship, we use random effects α_{i1} and β_{i1}. We similarly use $\alpha_{i2,I}$ and $\beta_{i2,I}$ for the incidence component and $\alpha_{i2,S}$ and $\beta_{i2,S}$ for the intensity equations of the weighted directed relationship.

Let $\boldsymbol{\theta}_i = \{\alpha_{i1}, \beta_{i1}, \alpha_{i2,I}, \beta_{i2,I}, \alpha_{i2,S}, \beta_{i2,S}\}$. We allow these random effects to be correlated across the relationships and assume that $\boldsymbol{\theta}_i$ is distributed multivariate normal[2] $N(0, \boldsymbol{\Sigma}_\theta)$, where $\boldsymbol{\Sigma}_\theta$ is an unrestricted covariance matrix consisting of the following submatrices:

$$\Sigma_\theta = \begin{pmatrix} \Sigma_{\theta,11} & \Sigma_{\theta,12} \\ \Sigma_{\theta,21} & \Sigma_{\theta,22} \end{pmatrix}.$$

The diagonal submatrices capture the within-relationship covariation in the random effects within a relationship. A positive correlation between the random effects for a relationship implies that popular individuals also tend to reach out more to others. The off-diagonal submatrices capture correlation across relationships and help determine whether individuals exhibit similar tendencies across relationships. If the off-diagonal submatrices in $\boldsymbol{\Sigma}_\theta$ indicate positive correlations, this could possibly be a result of a latent trait governing commonality in behavior. In contrast, if these submatrices are zero, the relationships can be modeled separately as the attractiveness and expansiveness parameters will be independent across the different relationships. Other patterns of correlations are also possible, and their meaning and relevance depend on the empirical context of a particular application.

Latent space. Social networks also exhibit patterns of higher-order dependence involving triads of actors. There is a potential for misleading inferences if such extradyadic effects are ignored. Hoff and his colleagues demonstrate how transitivity and other triad-specific structural characteristics such as balance and clusterability can be modeled using a latent space framework (Handcock, Raftery, and Tantrum 2007; Hoff 2005; Hoff, Raftery, and Handcock 2002). We employ a latent space for each relationship. We assume that individual i has a latent position $\mathbf{z}_{ir}$ in a Euclidean space associated with each relationship r, where $r \in \{1, 2\}$. The latent space framework stochastically models transitivity; if i is located close to individual j and if j is located close to individual k, then, because of the triangle inequality, i will also be close to k.

In the current study, we follow Hoff (2005) and use the inner-product kernel $z'_{ir}z_{jr}$.[3] In particular, for a two-relationship model, we use two kernels, one for each relationship, represented generically as $z'_{ir}z_{jr}$. The latent vectors for each relationship are assumed to come from a relationship-specific multivariate-normal distribution $\mathbf{z}_{ir} \sim N(0, \Sigma_{zr})$. The dimensionality of the latent space can be determined using a scree plot of the sum of the mean absolute prediction error of the entire triad census versus the dimensionality, as is usually done in the multidimensional scaling literature.[4]

Full model. Bringing together all the components of the model, we can write the latent utilities and response propensities as follows:

$$\begin{aligned}
u_{ij1} &= \mathbf{x}_{ij1}\boldsymbol{\mu}_1 + \alpha_{i1} + \beta_{j1} + \mathbf{z}'_{i1}\mathbf{z}_{j1} + e_{ij1},\\
u_{ji1} &= \mathbf{x}_{ji1}\boldsymbol{\mu}_1 + \alpha_{j1} + \beta_{i1} + \mathbf{z}'_{i1}\mathbf{z}_{j1} + e_{ji1},\\
u_{ij2} &= \mathbf{x}_{ij2,I}\boldsymbol{\mu}_{2,I} + \alpha_{i2,I} + \beta_{j2,I} + \mathbf{z}'_{i2}\mathbf{z}_{j2} + e_{ij2,I},\\
u_{ji2} &= \mathbf{x}_{ji2,I}\boldsymbol{\mu}_{2,I} + \alpha_{j2,I} + \beta_{i2,I} + \mathbf{z}'_{i2}\mathbf{z}_{j2} + e_{ji2,I},\\
\log\lambda_{ij} &= \mathbf{x}_{ij2,S}\boldsymbol{\mu}_{2,S} + \alpha_{i2,S} + \beta_{j2,S} + \mathbf{z}'_{i2}\mathbf{z}_{j2} + e_{ij2,S},\\
\log\lambda_{ji} &= \mathbf{x}_{ji2,S}\boldsymbol{\mu}_{2,S} + \alpha_{j2,S} + \beta_{i2,S} + \mathbf{z}'_{i2}\mathbf{z}_{j2} + e_{ji2,S}.
\end{aligned}$$

We assume that the vector of all errors $\mathbf{e}_{ij}$ is distributed multivariate normal $N(0, \boldsymbol{\Sigma})$. Also, $\boldsymbol{\theta}_i \sim N(0, \Sigma_\theta)$ and $\mathbf{z}_{ir} \sim N(0, \Sigma_{zr})$, $\forall r$.

Identification. Not all parameters of the model are identified. The error variance matrix $\boldsymbol{\Sigma}$ has a special structure

[2]Even though we assume a symmetric distribution for heterogeneity, when it is combined with the data from the individuals, the resulting posterior random effects can mimic skewed degree distributions that can arise from a preferential attachment mechanism. We verified this using a simulation that generated data from a preferential attachment mechanism and were able to recover highly skewed degree distributions. Details of this simulation are available on request.

[3]Other kernels such as those based on the Euclidean norm can also be used. We leave a detailed examination of the pros and cons of using different kernel forms for further research.

[4]The Bayes factor can also be used to determine dimensionality. However, it is difficult to compute in our model given the high dimensional numerical integration that is involved in obtaining the likelihood for each observation. Therefore, we opt for the predictive MAD criterion that focuses on triad census recovery.

because of scale restrictions on the binary utilities and because of exchangeability considerations stemming from the fact that the labels i and j are arbitrary within a pair. As the scale of the utilities of the binary responses cannot be determined from the data, the error variances associated with the binary components are set to 1. In addition, symmetry restrictions on the correlations stem from the exchangeability considerations, and the resulting variance matrix can be written as follows:

$$(5)\quad \Sigma = \begin{array}{c} \\ X_{ijl} \\ X_{jil} \\ X_{ij2,I} \\ X_{ji2,I} \\ X_{ij2,S} \\ X_{ji2,S} \end{array} \begin{array}{c} \begin{array}{cccccc} X_{ijl} & X_{jil} & X_{ij2,I} & X_{ji2,I} & X_{ij2,S} & X_{ji2,S} \end{array} \\ \begin{pmatrix} 1 & \rho_1 & \rho_2 & \rho_3 & \sigma\rho_4 & \sigma\rho_5 \\ & 1 & \rho_3 & \rho_2 & \sigma\rho_5 & \sigma\rho_4 \\ & & 1 & \rho_6 & \sigma\rho_7 & \sigma\rho_8 \\ & & & 1 & \sigma\rho_8 & \sigma\rho_7 \\ & & & & \sigma^2 & \sigma\rho_9 \\ & & & & & \sigma^2 \end{pmatrix} \end{array}.$$

The correlation parameter ρ_1 captures the impact of common unobserved variables affecting the binary relationship and also accounts for reciprocity. Similarly, ρ_6 and ρ_9 capture correlations for the weighted relationship and also account for reciprocity within this relationship. The correlation parameters ρ_7 and ρ_8 reflect common unobserved variables that influence both the incidence and intensity equations of the weighted component of the second relationship and are akin to selectivity parameters. Note that the intensity equations have a common variance.

The expansiveness and attractiveness random effects are individual specific and are thus separately identifiable from the equation errors that are dyad specific. The latent positions also are individual specific, but because they enter the equations as interactions, they can be separately identified from the random effects. However, because they appear in bilinear form, we can only identify these subject to rotation and reflection transformations. Finally, the covariance matrices Σ_{z1} and Σ_{z2} associated with the latent space parameters are restricted to be diagonal as their covariance terms are not identified. Moreover, each matrix has a common variance term across all the dimensions within a latent space.

Table 1 summarizes how the different model parameters can be related to substantive issues of interest. Given that the model components work in tandem, a parameter may also be related to other aspects apart from the one shown in the table. For example, ρ_1 is needed to capture reciprocity, but may also represent the impact of other shared unobservables.

Table 1

DESCRIPTION OF PARAMETERS

Variable	*Effects*
μ	Covariate effects, homophily, and heterophily
$\tilde{\alpha}_i$	Expansiveness, productivity
β_i	Attractiveness, popularity, or prestige
$\tilde{z}_j$	Transitivity, balance, and clusterability
ρ_1, ρ_6, ρ_9	Reciprocity
ρ_7, ρ_8	Selectivity
ρ_3, ρ_5	Generalized exchange
ρ_2, ρ_4	Multiplexity
Σ_θ	Heterogeneity

Bayesian Estimation

We now describe briefly our inference procedures. The likelihood for the model is computationally complex. Conditional on the random effects and latent positions, the dyad-specific likelihood requires numerical integration to obtain the multivariate normal cumulative distribution function. Moreover, the unconditional likelihood for the entire network requires additional multiple integration of very high dimensionality because of the crossed nature of the random effects. The dependency structure of our model is considerably more intricate than what is typically encountered in typical panel data settings in marketing, because we cannot assume independence across individuals or dyads for computing the unconditional likelihood. Therefore, we use Markov chain Monte Carlo (MCMC) methods involving a combination of data augmentation and the Metropolis–Hastings algorithm to handle the numerical integration. The data augmentation step allows us to leverage information from one relation to predict missing data on other relationships. The complexities involved in modeling multiple relationships and the identification restrictions on the covariance matrix $\mathbf{\Sigma}$ mean that the methods of inference for existing latent space models, as outlined, for example, in Hoff (2005) and Hoff, Raftery, and Handcock (2002), cannot be used directly for our model. Therefore, we provide a full derivation of the posterior full conditionals in Appendix A.

NEW PRODUCT DEVELOPMENT NETWORK

In the first application, we illustrate our modeling framework on sequential network data. We start with a special case of our general modeling framework and handle the simpler situation of a single directed binary relationship observed at two points in time. Therefore, we do not need the intensity component in this application. We use the same data as in Van den Bulte and Moenaert (1998; hereinafter, VdBM) on communications among members of different R&D teams and marketing and operations professionals who are all involved in new product development activities. Here, we briefly analyze this data set to revalidate the results in VdBM and to investigate whether our modeling framework (which differs significantly from that in VdBM) is better able to recover the structural characteristics of the network and whether it generates different conclusions or additional insights.

Data are available about communication patterns both before and after the R&D teams were collocated into a new facility. The data set used in VdBM and the one we reanalyze here come from a survey conducted in the Belgian subsidiary of a large U.S. telecommunication cooperation. The data consist of two 22×22 binary (who talks to whom) sociomatrices, $\mathbf{X}_1$ and $\mathbf{X}_2$, one for 1990 (before collocation) and one for 1992 (after the R&D teams were collocated in a separate facility). The actors in both years are the same, 13 R&D professionals spread over four teams and nine members of a steering committee consisting of seven positions in marketing and sales and two in operations. The characteristic element $x_{ij,t}$ in each of the two matrices is 1 if i reports to talk to j at least once a week in year t and 0 if otherwise. The specific area VdBM study is the impact of the collocation intervention on the communication and cooperation patterns among the different R&D teams and between R&D and marketing/operations by contrasting these patterns before and after collocation.

Barriers between R&D and marketing professionals resulting from differences in personality, training, or department culture imply that intrateam and intrafunctional communication will be more prevalent than cross-team and cross-functional communication. The idea of collocation was to foster communication among the different R&D teams. Of the five hypotheses VdBM propose, the first two test the communicative implications of the barriers between R&D and marketing and relate to team- and function-specific homophily effects. The third and fourth hypotheses focus on effects of collocating R&D teams to foster communication among these teams. Finally, VdBM posit that collocating R&D groups may not just foster between-group communication, but may even annihilate any difference between within- and between-group communication.

VdBM use Wasserman and Iacobucci's (1988) p_1 log-linear models to test their hypotheses. Our approach differs from that of these previous studies on several counts. First, in contrast to the p_1 model, we do not assume dyadic independence. In our model, the dyads are independent conditional on the random effects but are dependent unconditionally. Second, we allow for individual-specific expansiveness and attractiveness parameters in contrast to group-specific parameters to yield a richer specification of heterogeneity. Finally, we account for higher-order effects using a latent space. The added generality of our model is consistent with VdBM's (p. 16) call for "a new generation of models better able to handle triadic effects and other dependency issues."

Models and Variables

We estimated four models on the data set:

1. The full model involves all the components that form part of our modeling framework. These components include dyad-specific variables, attractiveness and expansiveness random effects that are correlated both within and across years, separate latent spaces for the two years, and correlated error terms for the utility equations of the four binary variables characterizing a dyad.
2. The Uncorr model is a restriction of the full model. It assumes that the random effects and the utility errors are correlated within a year but are uncorrelated across years. This is akin to having a separate model for each year, and this offers limited leeway in modeling muliplexity.
3. The NoZiZj model is a restriction of the full model such that the higher-order terms that characterize the latent space, (i.e., the $z_i'z_j$ terms are not included. We use this model to assess whether using the latent space results in better recovery of the triadic structure of the network and whether it substantively affects conclusions.
4. The team model closely mimics the VdBM article within our modeling framework. In this model, we restrict the expansiveness and attractiveness parameters to be the same for all individuals within a group and also do not include the latent space.

Variables

The following variables used in our investigation are the same as those VdBM use:

INTEAM_{ij} = 1 if i and j are R&D professionals on the same team and 0 if otherwise,

BETWTEAM_{ij} = 1 if i and j are R&D professionals but in different teams and 0 if otherwise,

INRD_{ij} = 1 if i and j are R&D professionals and 0 if otherwise, and

INMKTOPS_{ij} = 1 if i and j are both marketing or both operations executives and 0 if otherwise.

Results

We estimated the four models using MCMC methods. Each MCMC run is for 250,000 iterations, and the results are based on 200,000 iterations after discarding a burn-in of 50,000.

Recovery of structural characteristics. We begin by comparing the previously described models in their ability to recover the structural characteristics of the network. Given our interest in modeling sequential relationships, we focus on aspects of the network structure that pertain to the two relationships simultaneously. In particular, we compute statistics involving the dyadic as well as transitivity patterns of interactions that span both years.

We can describe dyadic relationships in each year as belonging to one of the following three types: mutual (M), asymmetric (A), and null (N). Observing across the two years, we can construct the following ten possible combinations (Fienberg, Meyer, and Wasserman 1985): NN, AN, NA, MN, NM, $A\bar{A}$, AA, AM, MA, and MM. The names are self-explanatory for most pairs. For example, NN refers to the number of dyads that are null in both years. Two patterns that require greater explanation are AA and $A\bar{A}$. The pair AA represents a dyad in which one actor is connected to the other in both years but neither relationship is reciprocated. The pair $A\bar{A}$ represents a dyad in which one actor initiates communication with the other in the first year and the other actor reciprocates by initiating in the second year, a kind of generalized exchange.

Table 2 reports the recovery of the sequential dyadic patterns. The columns report the absolute deviations between the actual frequencies and those predicted by the different models. The last row of the table reports the mean absolute deviations (MAD) across all the patterns for each model. It is clear that the uncorrelated model (Uncorr), which ignores cross-year linkages and models the two years separately, is significantly worse in recovering the cross-relationship dyadic patterns compared with the other models. All other models are roughly similar in their recovery, with the team model being the best. This indicates that it is important to model the

Table 2
APPLICATION 1: RECOVERY OF CROSS-RELATION DYADIC COUNTS

Pattern	*Full*	*NoZiZj*	*Uncorr*	*Team*
NN	7.51	6.44	12.67	7.04
AN	3.00	4.17	4.24	2.70
NA	4.25	4.90	4.82	5.50
MN	3.79	1.53	7.45	2.15
NM	.81	2.65	6.07	1.43
$A\bar{A}$	.92	1.89	1.89	.80
AA	1.97	.38	4.84	.02
AM	2.14	2.26	1.67	1.92
MA	.10	.80	.84	.77
MM	5.33	4.48	9.46	3.80
MAD	2.983	2.950	5.394	2.612

two relations jointly so that we can recover the cross-relation dyadic patterns, as all the models, and, except Uncorr, accommodate correlations across the two relationships.

We also investigate transitive patterns spanning both years to understand how well the model recovers extradyadic effects. This is necessary to assess whether adding the latent space is important. Eight such transitivity effects are possible: $\{X_{ij1}, X_{jk1}, X_{ik2}\}$, $\{X_{ij1}, X_{jk2}, X_{ik1}\}$, $\{X_{ij1}, X_{jk2}, X_{ik2}\}$, $\{X_{ij2}, X_{jk1}, X_{ik2}\}$, $\{X_{ij2}, X_{jk2}, X_{ik1}\}$, $\{X_{ij2}, X_{jk1}, X_{ik1}\}$, $\{X_{ij1}, X_{jk1}, X_{ik1}\}$, and $\{X_{ij2}, X_{jk2}, X_{ik2}\}$. For the sake of brevity, we do not include a full table of results (available on request). We find that the full model performs significantly better that all other models in recovering these transitive patterns (MAD = 34.43). The full and Uncorr models (MAD = 47.09), both of which include the latent space, perform significantly better than NoZiZj (MAD = 109.46) and team (MAD = 164.69), which do not include the higher-order effects.

In summary, observing across all dyadic and triadic measures, we find that the full model always does better than Uncorr, thus highlighting the need for joint modeling. The full model also performs better than all the other models in recovering the transitivity patterns, indicating that in this application, the latent space is important for handling extradyadic patterns. We can conclude that, on the whole, the full model recovers best the structural characteristics of the network.

Parameter Estimates and Hypotheses

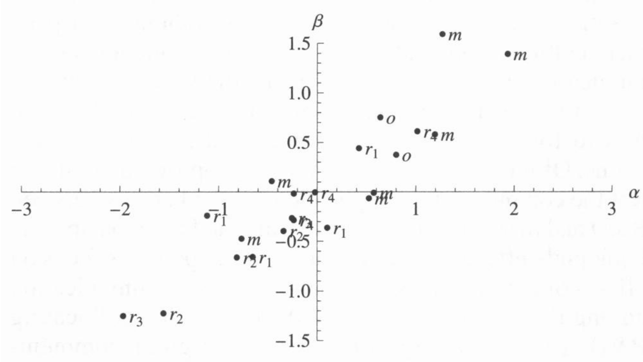
We begin by investigating whether the assumption of individual-level expansiveness and attractiveness parameters in our models has support. Figures 1 and 2 report the α and β values for the 22 managers for Years 1 and 2, respectively, for our full model. The labels for each point in the figures represent the team name. It is apparent from these figures that although some members of a group are clustered together, many groups exhibit considerable within-group heterogeneity. This is particularly noticeable for the marketing group (m) and the R&D groups (r_3, r_1, and r_4), which exhibit greater within-group variability. This suggests that the data support a richer characterization of heterogeneity than what is possible with group-specific effects.

Table 3 summarizes the parameter estimates. It is evident from the table that the parameters values differ substantially across the models in their magnitude and significance, indicating that the different model components influence substantive conclusions. Focusing on the bottom part of the table, we see that models that do not include the latent space yield higher estimates of the error correlations, possibly due to the confounding of variances across levels. We can use the coefficients to infer the degree of support for the different hypotheses studied in VdBM. Table 4 reports the extant of support for each hypothesis according to the different models. All other entries are computed from our reanalysis. Each entry for a particular model represents the probability that the corresponding hypothesis is true under that model. Several significant differences across the models are evident from this table.

The first two hypotheses pertain to within- and between-team homophily effects. Comparing the full model with VdBM indicates that our model supports both H_1 and H_2, whereas VdBM find mixed support for these. In particular, we find that H_{1b} has significant probability across all our models, indicating that R&D professionals tended to communicate predominantly with other R&D professionals before the move. The full, NoZiZj, and Uncorr models suggest strong support for H_{2a} and H_{2b} in contrast to team and VdBM. It seems that differences in support for this hypothe-

Figure 1

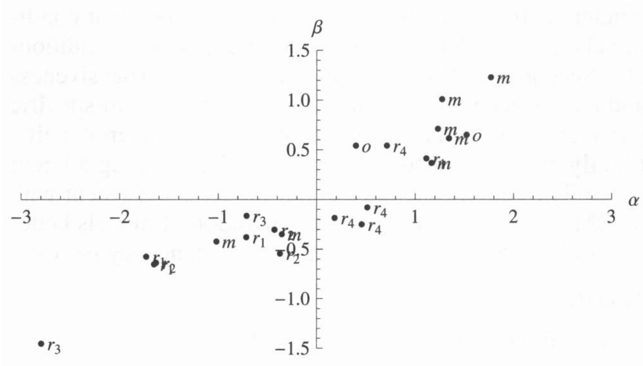
THE α AND β VALUES FOR THE MANAGERS IN YEAR 1



Notes: The point labels indicate team membership.

Figure 2

THE α AND β VALUES FOR THE MANAGERS IN YEAR 2



Notes: The point labels indicate team membership.

Table 3

APPLICATION 1: PARAMETER ESTIMATES FROM DIFFERENT MODELS

Pattern	*Full*	*NoZiZj*	*Uncorr*	*Team*
INTERCEPT1	**–1.48**	**–.87**	**–1.23**	.48
INTEAM1	**5.03**	**3.02**	**4.03**	**3.42**
BETWTEAM1	**1.60**	**.88**	.84	**1.42**
INMKTOPS1	1.03	**1.40**	1.02	.49
INTERCEPT2	**–1.18**	**–.99**	**–1.12**	.78
INTEAM2	**3.51**	**3.01**	**3.17**	**3.79**
BETWTEAM2	**2.15**	**1.73**	**1.80**	**2.71**
INMKTOPS2	**1.23**	**1.61**	1.02	.07
ρ_1	**.59**	**.83**	**.60**	**.79**
ρ_2	**.43**	**.64**	—	**.63**
ρ_3	**.26**	**.57**	—	**.53**
ρ_6	**.48**	**.72**	**.53**	**.67**

Notes: Bold indicates that the 95% posterior interval does not span 0.

Table 4

APPLICATION 1: HYPOTHESIS SUPPORTED BY DIFFERENT MODELS

Hypothesis		*Formal Test*	*Full*	*Team*	*VdBM*	*NoZiZj*	*Uncorr*
H_1: Both before and after collocation, R&D professionals tend to communicate with other R&D professionals rather than with marketing or operations executives.	H_{1a}	INTEAM1 > 0	.99	.99	.9	.99	.99
	H_{1b}	BETWTEAM1 > 0	.99	.99	n.s.	.99	.91
	H_{1c}	INTEAM2 > 0	.99	.99	.99	.99	.99
	H_{1d}	BETWTEAM2 > 0	.99	.99	.99	.99	.99
H_2: Both before and after collocation, marketing and operations executives tend to communicate with members of their own department.	H_{2a}	INMKTOPS1 > 0	.93	.85	n.s.	.99	.92
	H_{2b}	INMKTOPS2 > 0	.97	.55	n.s.	.99	.94
H_3: R&D professionals have a higher probability of communicating with members of other R&D teams after collocation than before.	H_3	BETWTEAM2 > BETWTEAM1	.82	.99	.99	.99	.86
H_4: When collocating R&D teams implies increasing the physical distance with other departments, R&D professionals' tendency to communicate with other R&D people rather than executives from other departments increases.	H_4	INRD2 > INRD1	.86	.99	.99	.99	.82
H_5: Before collocation, R&D professionals have a higher probability of communicating with members of their own team rather than with members of other R&D teams. After collocation, the tendency to communicate among R&D people is as strong between as within teams.	H_{5a}	INTEAM1 > BETWTEAM1	.99	.99	.99	.99	.99
	H_{5b}	INTEAM2 = BETWTEAM2	n.s.	n.s.	n.s.	n.s.	n.s.

Notes: Entries represent the probabilities of a hypothesis being true; n.s. indicates that the corresponding hypothesis is not supported.

sis are driven by the extent of within-group heterogeneity captured by the model. Recall that both team and VdBM assume that all members within a group share the same attractiveness and expansiveness parameters. Figures 1 and 2 show, however, that there is considerable heterogeneity in the recovered α and β parameters in the marketing group, and we find that failure to model this heterogeneity comprehensively can substantively affect conclusions.

We also find some differences in the support for the collocation hypotheses (H_3 and H_4) across the models. The full and Uncorr models have a lower probability associated with these hypotheses compared with VdBM, team, and NoZiZj. These differences can be explained by the presence or absence of the latent space for capturing higher-order effects and are also consistent with the strong support for H_{1b} in our model. Finally, all models yield no support for H_5.[5] These differences in supported hypotheses (H_{1b}, H_{2a}, H_{2b}, H_3, and H_4) across models demonstrate that the different model components can affect the theoretical and substantive conclusions and that it is important to account for dyadic-dependence and higher-order effects.

Leveraging Information Across Relationships

We now illustrate how our framework can be used to leverage information in one relationship to predict relationships in another. For example, we assume that the data involving the entire marketing groups are missing in the second year. In such a situation, we cannot readily use the log-linear modeling framework previous researchers have employed, because the group-specific parameters used in such models will not be available for the marketing group. However, for our models, the natural reliance on data augmentation to obtain the utilities and the individual-specific random effects when estimating parameters ensures that such missing data can be handled seamlessly. In particular, the covariance matrix of the random effects can be used to leverage information from Year 1 to Year 2 about these individual-specific parameters. Therefore, we estimate our full and Uncorr models on such a data set to determine whether incorporating cross-relationship linkages improves predictions in such situations. Note that in the full model, the data on Year 1 for the individuals in the marketing group can be leveraged to predict relationships in Year 2. This is not possible in the Uncorr model, in which the two years are modeled separately. Tables WA 1 and WA 2 of the Web Appendix (see http://www.marketingpower.com/jmraug11) report how well the cross-year dyadic and transitivity patterns are recovered on such a data set with missing values for the marketing group. These tables report the absolute deviations between the actual frequencies and those predicted by the full and Uncorr models and illustrate that the full model does significantly better in recovering these cross-year relationships.[6]

ONLINE SOCIAL NETWORK

In this application, our focus is on modeling relationships of different types. We use a combination of undirected and directed binary relationships and a directed weighted relationship. In particular, we show how it is important to model both the incidence and intensity of weighted relationships, because conclusions and predictions depend crucially on this distinction. The data for this application come from a Swiss online social network on which members can create a profile as either a user or an artist and can then connect with other registered members through friendship relationships. The social networking site offers different services to these two distinct user groups. Whereas both groups can publish user-generated content such as blogs, photos, or

[5]For H_{5b}, we found that the probability associated with INTEAM2 > BETWTEAM2 is .99 for all our models.

[6]We also investigated the role of demographics that were part of the data but did not find any significant impact. Our models can handle such continuous covariates, something that is not possible with a log-linear specification.

videos on their profiles, only artists can, in addition, publish up to a maximum of 30 songs on their profile.

Artists use the different services offered by the platform to promote their music and concerts and to seek collaboration with other musicians and bands. They establish friendship relationships with other artists, send personal messages, and write public comments on other artists' profiles. For entertainment and informational reasons, users, as well as artists, visit profiles of artists and download songs. Artists engage in active promotion and relationship effort in the hope that it will result in increased collaboration, communication, popularity, and song downloads. In summary, the online networking site offers a platform that combines social networking services with entertainment and communication services.

There are four components of the data set: member data, friendship data, communication data, and music download data. The member data contain information collected at registration and include stable variables such as the registration date, date of birth, gender, and city or, in the case of artists, their genre and information about their offline concerts and performances before joining the network. In addition, the data also contain information on the number of page views of each member's page on the network during a given time period. The other data components pertain to our three dependent variables and are described in greater detail in the following section.

Data and Structural Characteristics

Our sample consists of 230 artists who created a profile on the network between February 1, 2007, and March 31, 2007, provided information about their activities and characteristics, and uploaded at least one song on their profile. We model three types of relationships among these artists over the course of the six months between April 1, 2007, and September 30, 2007.[7] These relationships include friendship (f), communications (c) and music downloads (m). The data set thus contains three 230×230 matrices (Y_f, Y_c, and Y_m, respectively) for these relationships.

Structural characteristics. We now briefly describe the structural characteristics of the three relationships for our set of artists.

1. *Friendship:* The friendship relation y_{ij}^f is binary and undirected. It indicates whether a friendship is formed between the pair {i, j} before the end of our data period. The network has 3564 friendship relations among a maximum possible 26,335 connections, yielding a network density of 13.53%.
2. *Communications:* The communication relation y_{ij}^c is binary and directed and indicates whether artist i sent a communication (direct message or comment) to artist j within the time period of the data. We observed 4575 communication relations, yielding a density of 8.68%. The relation exhibits considerable reciprocity or mutuality (defined as the ratio of the number of pairs with bidirectional relations to the number of pairs having at least one tie) equal to 30.9%. Artists vary in their level of expansiveness (or outdegree), as measured by the number of artists they communicate with and their popularity or receptivity(i.e., the number of artists communicating with a given artist [indegree]). The indegree and outdegree distributions are highly skewed. The mean degree is 19.97. The maximum and minimum for the indegree distribution are 203 and 0, respectively, whereas the maximum and minimum for the outdegree distribution are 182 and 0, respectively.
3. *Music Downloads:* The music downloads represent a directed and weighted relationship. Each song download entry y_{ij}^m is a count of the number of times that artist j listens to a song on artist i's profile and may include multiple downloads of the same song. As discussed in the "Modeling Framework" section, we distinguish between the incidence and intensity of music downloads. Focusing first on incidence, we find that of the possible 52,670 ties, our data contain only 17,912 binary ties, implying a density of 34%. The reciprocity is 39.1%. The outdegree of an artist is the number of other artists who download from that artist, and the indegree is the number of other artists from whom the artist downloads music. For the binary component, the mean degree is 17.89, and the maximum indegree and outdegree are 213 and 135, respectively. Because this is a weighted relation, we can also study the intensity of connections. On average, each artist downloads songs on 59.98 occasions (including multiple song downloads). The maximum weighted indegree (i.e., the number of songs downloaded from) an artist is 2899, whereas the maximum weighted outdegree (i.e., the maximum number of times a single artist listens to songs is 612). We also find that the artist-specific degree statistics are highly correlated across the three relationships.

Model Specifications and Variables

We estimate several different variants of the full model (hereinafter, we refer to this as "full model") that we outlined in the "Modeling Framework" section. Our null models impose different restriction on the full model; we constructed them to investigate how crucial the different components of the full model are in capturing important aspects of the data generating process. The models are as follows:

- *Full Model*. The full model includes dyad-specific covariates ($\mathbf{x}_{ij}$) to accommodate homophily and heterophily; artist-specific covariates ($\mathbf{x}_i$) and ($\mathbf{x}_j$) to account for asymmetry in responses; artist- and relationship-specific sender and receiver parameters (θ_i), to model heterogeneity in expansiveness and receptivity; correlations in these random effects across relationships ($\boldsymbol{\Sigma}_\theta$); latent spaces of random locations for the three relations ($\mathbf{z}_{im}$, $\mathbf{z}_{ic}$, and $\mathbf{z}_{if}$) to capture higher order effects; and correlations in errors ($\boldsymbol{\Sigma}_m$, $\boldsymbol{\Sigma}_c$, and $\boldsymbol{\Sigma}_f$), to incorporate reciprocity within each relationship. In addition, the full model uses a correlated hurdle count model to account for the sparse nature of many network data sets.
- *Poisson*. The Poisson model uses a Poisson distribution for the music download relationships, rather than the correlated hurdle Poisson. Thus, the counts are modeled directly, and we do not distinguish between the incidence and intensity of counts. The model is otherwise identical to the full model in all other respects.
- *Uncorr*. In the Uncorr model, we treat the three relationships as independent. Thus, we assume the artist-specific random effects in $\boldsymbol{\theta}_i$ to be uncorrelated across the relationships, and therefore, $\boldsymbol{\Sigma}_\theta$ has a block-diagonal structure. This model is thus equivalent to running three separate models on the three relationships.
- *NoZiZj*. For the NoZiZj model, we do not include the higher order terms $\mathbf{z}_i'\mathbf{z}j$ that characterize the latent space.

Each artist can be described using the variables detailed in Table 5. We use these artist-specific variables to compute dyad-specific variables. We use different combinations of

[7]Our data are not entirely representative of the whole network, which consists of other artists who joined the network subsequent to our data period. In addition, we focus only on the subnetwork of connections involving artists, rather than also considering fans, because this is consistent with the primary focus of the network in offering a platform for artists to present themselves and seek collaborations.

the artist- and dyad-specific variables in explaining the three relationships. (Appendix B gives details of the covariate specifications for the three relationships.) Table 6 shows descriptive statistics associated with these covariates. The last two rows of Table 6 contain dyad-specific binary variables. The variable "common region" is equal to 1 when both artists in a dyad are from the same region; otherwise, it is 0. We code the last variable, "common genre," analogously.

Table 5
APPLICATION 2: DESCRIPTION OF VARIABLES

Variable	*Description*
Songs	The number of songs available on the artist's profile.
Band	Whether the artist belongs to a band. Equal to 1 if artist belongs to a band and 0 if otherwise.
Audience	The audience size of the largest concert by the artist. A median split yields 1 when the audience is > 700 and 0 otherwise.
Genre	Genre of the artist. We distinguish between rock genres and nonrock genres.
Region	The geographical region to which the artist belongs. Artists belong to one of the 26 cantons in Switzerland. These are aggregated into three regions: French, German, and Italian.
PageViews	The number of page views of the artist's profile during the data period. These page views originate from other registered members, including the fans, or from Internet users from outside the social network.
YActive	The number of years of activity of the artist in the music industry.

Table 6
APPLICATION 2: DESCRIPTIVE STATISTICS FOR ARTIST VARIABLES

Variable	*Observations*	*M*	*Mdn*	*SD*	*Min*	*Max*
Page views	230	781.2	467	1069.8	43	10,931
Songs	230	4.108	3	3.26	1	29
Audience	230	.482	0	.5	0	1
Band	230	.704	1	.457	0	1
Years active	230	7.24	6	5.82	1	27
Common region	26,335	.628	1	.483	0	1
Common genre	26,335	.564	1	.495	0	1

Results

We estimated the models using MCMC methods (described in Appendix A). Each MCMC run is for 250,000 iterations, and the results are based on a sample of 200,000 iterations after a burn-in of 50,000.

Model adequacy. Before discussing parameter estimates, we use a combination of predictive measures and posterior predictive model checking (Gelman, Meng, and Stern 1996) to compare the adequacy of our models. This involves generating G hypothetical replicated data sets from the model using the MCMC draws and comparing these data sets with the actual data set. These comparisons are made using various test quantities that represent different structural characteristics of the network. If the replicated data sets differ systematically from the actual data on a given test quantity, the model does not adequately mimic the structural characteristic that the test quantity represents. The discrepancy between the replicated data sets and the actual data can be assessed using posterior predictive p-value. This p-value is the proportion of the G replications in which the simulated test quantity exceeds the realized value of that quantity in the observed data. An extreme p-value, (either close to 0 or 1; i.e., $\leq .05$ or $\geq .95$) suggests inadequate recovery of the corresponding test quantity.

We use several test statistics associated with the weighted relationship to assess whether the distinction between incidence and intensity is crucial. Table 7 shows the model adequacy results for the music download relationship based on G = 10,000 MCMC draws. Column 2 of the table reports the value of the test statistics for the observed data, and the other columns report the posterior predictive mean and p-values for the different models. A few conclusions can be readily drawn from the table. First, Poisson, which models the intensity directly using a Poisson specification (as in Hoff 2005), does not recover any of the test statistics adequately, because almost all p-values in Column 6 are extreme. Furthermore, the posterior predictive mean values for the test statistics (Column 5) are appreciably different from their counterparts in the actual data. Second, we observe that *none* of the p-values associated with the other models are extreme, and this indicates that modeling inci-

Table 7
APPLICATION 2: POSTERIOR PREDICTIVE CHECKING FOR MUSIC DOWNLOADS FOR THE FOUR MODELS

		Full		*Uncorr*		*NoZiZj*		*Poisson*	
	Data	*M*	p-*Value*	*M*	p-*Value*	*M*	p-*Value*	*M*	p-*Value*
Binary									
Null dyads	23,376	23,420	.79	23,416	.77	23,420	.80	23,275	.03
Asymmetrical dyads	1802	1783	.35	1785	.36	1777	.30	2134	1
Mutual dyads	1157	1132	.24	1134	.26	1137	.28	925	0
Reciprocity	.39	.39	.44	.39	.45	.39	.51	.30	0
Transitive triads	51,692	50,523	.3	50,710	.32	49,534	.16	44,303	0
Intransitive triads	136,230	133,398	.27	133,321	.26	135,012	.39	117,494	0
Mean degree	17.89	17.59	.18	17.62	.2	17.62	.21	17.32	.04
Standard indegree	17.924	18.1	.65	18.07	.62	18.16	.69	16.61	0
Standard outdegree	31.61	31.06	.14	31.09	.16	31.07	.14	27.74	0
Degree correlation	.898	.893	.3	.89	.31	.893	.32	.896	.45
Quantity									
Mean strength	59.58	58.83	.25	58.69	.23	58.86	.26	209.4	1
Standard indegree	213.2	208.44	.34	206.34	.29	209.66	.38	1327.9	1
Standard outdegree	74.13	70.06	.15	69.47	.12	69.51	.13	751.1	1

dence and intensity separately is important for adequately capturing the structural network characteristics associated with weighted relations. Finally, because the NoZiZj model performs almost as well as the full model, we conclude that the latent space is not crucial for the recovery of structural characteristics in this application.

The preceding discussion focuses on model adequacy for music downloads to highlight the contribution of the correlated multivariate hurdle Poisson in modeling weighted relationships. The results from the other two relationships show that all models (including Poisson) are similar in their recovery of structural characteristics for these relationships. The model adequacy results are based on in-sample simulations. (We report the predictive performance of our model in the Web Appendix at http://www.marketingpower.com/jmraug11.) We find that the Full model outperforms other models on almost all measures. We also find that the Poisson model does very poorly in predicting future activity, indicating that there are significant gains in modeling the incidence and intensity separately.

Parameter Estimates

We now discuss the parameter estimates based on the entire sample of six months. We focus on the parameter estimates from the full model. Estimates from other models mostly yield similar qualitative conclusions, and we do not include these for the sake of brevity.

Covariate effects. Table 8 reports the posterior means and standard deviations of the coefficient estimates for the three relationships. In interpreting this table, recall that all variables associated with the friendship relation are dyad specific and binary, whereas, for the other relations, some variables are dyad specific and some are artist specific. Also, note that both components of the music relation have the same set of covariates. (The definitions for these covariates are available in Appendix B.)

For the sake of brevity, we synthesize the results across all the relationships. There is clear evidence of homophily and proximity: For all three relationships, the dyad-specific variables CRegion and CGenre have a positive and significant impact on the likelihood of forming connections. The positive coefficient for CRegion is consistent with the notion that a common language and geographical proximity can enhance the likelihood of collaborative effort. The positive coefficient for CGenre means that pairs of artists who produce music in the same genre have a higher propensity to form friendship connections, communicate with each other, and download music from each other. We also find that artists belonging to a band have a higher chance of forming friendships (BothBand) and a higher probability of sending and receiving communications (SBand and RBand).

We find that measures of online popularity that are based on the total number of page views for an artist (BothPopular, SPviews, DPviews, and PPviews) positively influence relationship formation. For example, the positive coefficients for DPviews in the music relation indicate that artists with greater online popularity in this network have a greater likelihood of downloading songs from other artists and that they tend to download more music. In contrast, most measures of prior and offline popularity or experience (indicating audience size of concerts or years of activity) do not seem to affect the formation of online relationships within the network.[8] Finally, the data indicate that for the music relationship, different coefficients influence the incidence and intensity components of the music relationship. Thus, we conclude that these two facets are not isomorphic and need to be modeled separately.

Covariance structure of the relationships. The covariance matrix Σ_θ captures the linkages among the expansiveness and attractiveness parameters across the relationships. Table 9 reports the elements of Σ_θ. A striking feature of Table 9 is that all the covariances are significantly positive. The positive correlation within a relationship implies that attractiveness goes hand-in-hand with expansiveness (i.e., artists who are sought by others also tend to be active in seeking relationships with others). Moreover, for the music relationship, the attractiveness and expansiveness parameters for the intensity equations are also positively correlated with the corresponding random effects for the incidence component. The positive correlations across relationships imply that an artist who is popular in one relationship is also likely to be both popular and productive in other relationships. Similarly, an artist who is productive in one relationship is also likely to be productive and popular in other relationships.

The utility errors for the two incidence equations of the music download relationship are positively correlated (.74), owing to shared unobservable influences and reciprocity. The intensity equations are also positively correlated (.497). Furthermore, the errors for the incidence equations are positively correlated with the errors for the intensity equations (.395 and .411) implying selectivity through shared unobserved factors driving both incidence and intensity. This corroborates the need for a multivariate correlated hurdle Poisson specification. Finally, the communication utilities also exhibit positive correlation (.511) driven by reciprocity and other shared unobservables.

CONCLUSIONS

As interest in social networks and brand communities grows, marketers are becoming increasingly focused on understanding and predicting the connectivity structure of such networks. In this article, we developed a methodological framework for jointly modeling multiple relationships that vary in their directionality and intensity. Our integrated approach for social network analysis is unique in that it weaves together several distinct model components needed for capturing multiplexity in networks.

We applied our framework to two distinct applications that showcased different benefits of our approach. In the first application, we investigated the impact of an organizational intervention (R&D collocation) on the patterns of communications among professionals involved in new product development activities. In this application, we specifically investigated the gains from modeling relationships jointly. Our results clearly indicate how the different components of our framework are needed for a clear assessment of substantive hypotheses. We found that the hetero-

[8]We thank an anonymous reviewer for pointing out that covariates relating to popularity could be potentially endogenous. However, given that we include both online and offline correlates of popularity, as well as actor-specific random effects that capture attractiveness, it is unclear whether additional unobserved variables relating to popularity are part of the utility error. However, caution is still needed in interpreting the results.

Table 8

APPLICATION 2: COEFFICIENT ESTIMATES FOR THE THREE RELATIONSHIPS FOR THE FULL MODEL

Parameter	*M*	*SD*
Friendship Relationship		
Intercept	**−2.732**	.167
CRegion	**.341**	.050
CGenre	**.150**	.033
BothPopular	**.163**	.063
BothNotPopular	.015	.087
BothBigSongs	.061	.061
BothSmallSongs	.016	.065
BothBand	**.640**	.153
BothNoBand	−.257	.163
BothBigAudience	.203	.143
BothSmallAud	−.134	.143
BothLongActive	−.036	.075
BothShortActive	.034	.071
Communications Relationship		
Intercept	**−3.446**	.270
CRegion	**.321**	.042
CGenre	**.144**	.028
SPviews	**.015**	.004
SSongs	.009	.010
SBand	**.321**	.106
SAudience	.135	.093
SYActive	−.003	.006
RPviews	.002	.003
RSongs	.013	.009
RBand	**.464**	.154
RAudience	.043	.138
RYActive	−.004	.006
Music Relationship: Incidence		
Intercept	**−3.007**	.272
CRegion	**.183**	.040
CGenre	**.118**	.031
PPviews	**.021**	.004
PSongs	−.006	.010
PBand	.106	.104
PAudience	.100	.091
PYActive	**−.013**	.006
DPviews	**.031**	.005
DSongs	.022	.014
DBand	.062	.141
DAudience	−.113	.120
DYActive	−.015	.009
Music Relationship: Intensity		
Intercept	**−2.024**	.269
CRegion	**.235**	.050
CGenre	**.152**	.038
PPviews	**.011**	.003
PSongs	**.023**	.009
PBand	.169	.086
PAudience	−.005	.074
PYActive	−.003	.006
DPviews	**.034**	.006
DSongs	.021	.017
DBand	−.055	.149
DAudience	−.157	.120
DYActive	−.012	.011

Notes: Bold indicates that the 95% posterior interval does not span 0.

Table 9

APPLICATION 2: COVARIATION IN THE RANDOM EFFECTS Σ_θ

V	α^m_{i1}	β^m_{i1}	α^m_{i2}	β^m_{i2}	α^i_c	β^i_c	α^f_i
α^m_{i1}	.397 (.051)	.378 (.056)	.234 (.037)	.222 (.049)	.268 (.040)	.500 (.064)	.499 (.065)
β^m_{i1}		.707 (.091)	.238 (.042)	.441 (.073)	.329 (.053)	.617 (.086)	.593 (.086)
α^m_{i2}			.192 (.034)	.146 (.038)	.195 (.033)	.315 (.050)	.319 (.052)
β^m_{i2}				.489 (.082)	.188 (.048)	.313 (.079)	.288 (.081)
α^i_c					.407 (.052)	.478 (.065)	.513 (.067)
β^i_c						1.000 (.109)	.988 (.104)
α^f_i							1.031 (.111)

Note: Posterior standard deviations are in parentheses.

geneity specification, the latent space, and the correlations across relationships can affect both substantive conclusions and the recovery of structural characteristics. Finally, as the Web Appendix shows (see http://www.marketingpower.com/jmraug11), our approach can be used to leverage information from one relation to predict connectivity in another.

In our second application, we focused on modeling multiple relationships of different types. In particular, we showed how it is critical to model both the incidence and intensity of weighted relationships such as music downloads; otherwise, recovery of structural characteristics and predictive performance suffers appreciably. On the substantive front, our results show that the friendship, communications, and music download relationships share common antecedents and exhibit homophily and reciprocity. We found that offline proximity is relevant for all online relationships, and this is consistent with our understanding that these connections are formed to forge collaborative relationships. We also found that the artists exhibit similar roles across relationships and that popular artists seem to be more productive regardless of the relationship being studied.

Across the two applications, we found mixed evidence regarding the benefits from incorporating a latent space. In the first application, the latent space improved the recovery of cross-relationship transitivity patterns and affected the parameter estimates and the substantive findings. However, higher-order effects did not seem to be important in the second application. The latter result is consistent with Faust's (2007) conclusion that much of the variation in the triad census across networks could be explained by simpler local structure measures. These results suggest that the impact of extradyadic effects could be application specific.

On the theoretical and substantive front, our framework facilitates a detailed description of antecedents of relationship formation and allows for theory testing taking into account systematic variations in degree arising from homophily and heterophily, local structuring, as well as temporal or cross-relationship carryover (Rivera, Soderstrom and Uzzi 2010). Our enquiry can be extended in many directions. Our applications involved small networks. Most online networks are much larger, and statistical methods cannot scale directly to the level of these large networks. However, recent research has shown that while online networks can have millions of members, communities within such networks are relatively small, with sizes in the vicinity of the 100–200 member range (Leskovec et al. 2008). This implies that these very large networks can be broken down into clusters of tightly knit communities, and when such communities are identified, our methodological framework can then be used on such communities to further understand the nature of linkages within these subcommunities. However, such a divide-and-conquer approach is unlikely to provide a complete picture of the nature of link formation in such large networks.

We focused on modeling static relationships or on sequential relationships observed over a few time periods. However, networks are dynamic entities in which connections are formed over time. Incorporating such dynamics would be interesting. We used a parametric framework based on the normal distribution for the latent variables and random effects, and this was sufficient for recovery of skewed degree distributions. However, in other situations, Bayesian nonparametrics (Sweeting 2007) may be more useful.

APPENDIX A: FULL CONDITIONAL DISTRIBUTIONS

1. The full conditional for precision matrix Σ_θ^{-1} of the actor-specific random effects is a Wishart distribution given by

$$\text{(A1)} \quad \left(p\Sigma_\theta^{-1}\middle|\{\theta_i\}\right) = \text{Wishart}\left[\rho_\theta + N, \left(\sum_{i=1}^{N}\theta_i\theta_i' + R_\theta^{-1}\right)^{-1}\right],$$

where the prior for Σ_θ^{-1} is Wishart(ρ_θ, R_θ). The quantities ρ_θ and R_θ refer to the scalar degree of freedom and the scale matrix for the Wishart, respectively, and N is the number of actors in the network.

2. The covariance matrices Σ_z^r, for relationship r are diagonal, because z_i^r is a p-dimensional vector of independent components. Let $\sigma_{z,r}^2$ denote the common variance for the components of z_i^r. The full conditional for $\sigma_{z,r}^2$ is an inverse gamma distribution given by

$$\text{(A2)} \quad p(\sigma_{z,r}^2 \mid \{z_i^4\}) = \text{IG}\left[a_r + pN, \left(\frac{1}{p}\sum_{i=1}^{N}\sum_{k=1}^{p} z_{irk}^2 + b_r^{-1}\right)^{-1}\right].$$

3. The full conditional for the coefficients μ is multivariate normal because we have a seemingly unrelated regression system of equations conditional on knowing the latent variables. Form the adjusted utilities (e.g., $\tilde{u}_{ij1} = u_{ij1} - \alpha_{i1} - \beta_{j1} - \mathbf{z}_{i1}'\mathbf{z}_{j1}$) and adjusted log-rate parameters by subtracting terms that do not involve μ from the latent dependent variables. We then have the system of equations, $\tilde{\mathbf{v}}_{\{ij\}} = X_{\{ij\}}\mu + \mathbf{e}_{\{ij\}}$, for an arbitrary pair $\{i, j\}$, where $\mathbf{e}_{ij} \sim N(0, \Sigma)$. We can write the full conditional as follows:

$$\text{(A3)} \quad p(\mu \mid \{\tilde{\mathbf{v}}_{\{ij\}}\}) = N(\hat{\mu}, \Omega_\mu),$$

where $\Omega_\mu^{-1} = \mathbf{C}^{-1} + \Sigma_{\{ij\}}X_{\{ij\}}'\Sigma^{-1}X_{\{ij\}}$ and $\hat{\mu} = \Omega_\mu[C^{-1}\eta + \Sigma_{\{ij\}}X_{\{ij\}}'\Sigma^{-1}\tilde{\mathbf{v}}_{\{ij\}}]$.

4. The full conditional for the heterogeneity parameter θ_i is a multivariate normal. We again begin by creating adjusted utilities and rate parameters, such as by subtracting all terms that do not involve θ_i. Let be the vector of adjusted utilities for the three relationships. Then we have the system. Again, we can use standard Bayesian theory for the multivariate normal to obtain the resulting full conditional:

$$\text{(A4)} \quad p(\theta_i \mid \{\tilde{\mathbf{v}}_{\{ij\}\theta}\}) = N(\hat{\theta}\%_i, \Omega_\theta),$$

where $\Omega_\theta^{-1} = \Sigma_\theta^{-1} + (N-1)\Sigma^{-1}$ and $\hat{\theta}_i = \Sigma^{-1}\Omega_\theta\Sigma_{j\neq i}\tilde{\mathbf{v}}_{\{ij\}\theta}$.

5. The full conditional for $\mathbf{z}_i$ that contains all the latent space vectors associated with an individual i is multivariate normal. Creating adjusted utilities such as $\tilde{u}_{ij1z} = u_{ij1} - \mathbf{x}_{ij1}'\mu_1 - \alpha_{i1} - \beta_{j1}$, we can form the vector of adjusted utilities and latent rate parameters $\tilde{\mathbf{v}}_{\{ij\}z} = \mathbf{Z}_j\mathbf{z}_i + \mathbf{e}_{\{ij\}}$, where $\mathbf{Z}_j$ is an appropriately constructed matrix from the latent space vector of actor j. This is a seemingly unrelated regression system. Given the prior $\mathbf{z}_i \sim N(0, \Sigma_z)$, where Σ_z is constructed from the different Σ_{zr} matrices, we can write the full conditional as $N(\hat{\mathbf{z}}_i\ \Omega_{z_i})$, where $\Omega_{z_i}^{-1} = (\Sigma_z)^{-1} + \Sigma_{j\neq i}\mathbf{Z}_j'\Sigma^{-1}\mathbf{Z}_j$ and $\hat{\mathbf{z}}_i = \Omega_{z_i}[\Sigma_{j\neq i}\mathbf{Z}_j'(\Sigma)^{-1}\tilde{\mathbf{v}}_{\{ij\}z}]$. The model depends on the inner product of the latent space vectors, which is invariant to rotations and reflections of the vectors. Thus visual representations of these vectors require that they be rotated to a common orientation, This can be done by using a Procustrean transformation as outlined in Hoff (2005).

6. The variance–covariance matrix of the errors for the weighted relationship, Σ, has a special structure as described in the "Modeling Framework" section. Given this special structure, we follow the separation strategy of Barnard, McCulloch, and Meng (2000) in setting the prior in terms of the standard deviations and correlations in Σ. The covariance matrix Σ can be decomposed into a correlation matrix, $\mathbf{R}$, and a vector, $\mathbf{s}$, of standard deviations—that is, $\Sigma = \text{diag}(\mathbf{s}) \times \mathbf{R} \times \text{diag}(\mathbf{s})$, where $\mathbf{s}$ contains the square roots of the standard deviations. Let ω contain the logarithms of the elements in $\mathbf{s}$. We assume a multivariate normal distribution $N(\mathbf{0}, \mathbf{I})$ for the nonredundant elements of $\mathbf{R}$, such that it is constrained to the subspace of the p-dimensional cube $[-1, 1]^p$, where p is the number of equations that yields a positive definite correlation matrix. Finally, we assume a univariate standard normal prior for the single log-standard deviation in ω.
   - The full conditional distribution for the free element in the vector of log-standard deviations ω of errors can only be written up to a normalizing constant (recall that the terms associated with the binary utilities in ω are fixed to 0 for identification purposes). Given our assumption of a normal prior for the single free element, we use a Metropolis–Hastings step to simulate the standard deviation in ω. A univariate normal proposal density can be used to generate candidates for this procedure. If is the current value of kth component of ω, a candidate value is generated using a random walk chain $\omega_k^c = \omega_k^{(t-1)} + N(0, \tau)$, where τ is a tuning constant that controls the acceptance rate.
   - Many different approaches can be used to sample the correlation matrix $\mathbf{R}$. Here, we use a multivariate Metropolis step to sample a vector of nonredundant correlations in $\mathbf{R}$. We used adaptive MCMC (Atchade 2006) for obtaining the tuning constant so as to ensure rapid mixing.

7. The full conditional distribution associated with the set of latent utilities and latent rate parameters in $\mathbf{u}_{ij}$ is again unknown. We sample the utilities and log-rate parameters using univariate conditional draws. Sampling the utilities is straightforward, because these are truncated univariate conditional normal draws. The log-rate parameters log λ_{ij} and log λ_{ji} are sampled such that these are univariate normal draws if the corresponding observation involves a zero count, and for an observation in which a positive count is observed, we use a univariate Metropolis step that combines the likelihood for a truncated Poisson distribution with a conditional normal prior.

APPENDIX B: MODEL SPECIFICATIONS FOR APPLICATION 2

Covariates for the Friendship Relationship

CRegion: CRegion is equal to 1 if both artists in a pair are from the same region; 0 otherwise.

CGenre: CGenre is equal to 1 if both artists in a pair are from the same genre; 0 otherwise.

BothPopular: BothPopular is equal to 1 if both artists in a pair are viewed (online) by more than the population median; 0 otherwise.

BothNotPopular: BothNotPopular is equal to 1 if both artists in a pair are viewed by fewer than the population median; 0 otherwise.

BothBigSongs: BothBigSongs is equal to 1 if both artists in a pair post more songs than the population median; 0 otherwise.

BothSmallSongs: BothSmallSongs is equal to 1 if both artists in a pair post fewer songs than the population median; 0 otherwise.

BothBand: BothBand is equal to 1 if both artists in a pair post are from a band; 0 otherwise.

BothNoBand: BothNoBand is equal to 1 if both artists in a pair post are not from a band; 0 otherwise.

BothBigAudience: BothBigAudience is equal to 1 if both artists in a pair had large concerts with more than 700 spectators; 0 otherwise.

BothSmallAudience: BothSmallAudience is equal to 1 if both artists in a pair had small concerts with more than 700 spectators; 0 otherwise.

BothLongActive: BothLongActive is equal to 1 if both artists in a pair had being active for more than six years; 0 otherwise.

BothShortActive: BothShortActive is equal to 1 if both artists in a pair had being active for less than six years;0 otherwise.

Covariates for the Communication Relationship

CRegion: CRegion is equal to 1 if both artists in a pair are from the same region; 0 otherwise.

CGenre: CGenre is equal to 1 if both artists in a pair are from the same genre; 0 otherwise.

SPviews: SPviews represents the number of sender paged views.

SSongs: SSongs represents the number of songs posted on the sender web page.

SBand: SBand is equal to 1 if the sender belongs to a band; 0 otherwise.

SAudience: SAudience is equal to 1 if the sender performed in front of an audience larger than 700 people; 0 otherwise.

SYActive: SYActive is equal to 1 if the sender was active for more than six years; 0 otherwise.

RPviews: RPviews represents the number of receiver paged views.

RSongs: RSongs represents the number of songs posted on the receiver web page.

RBand: RBand is equal to 1 if the receiver belongs to a band; 0 otherwise.

RAudience: RAudience is equal to 1 if the receiver performed in front of an audience larger than 700 people; 0 otherwise.

RYActive: RYActive is equal to 1 if the receiver was active for more than six years; 0 otherwise.

Covariates for the Music Download Relationship

CRegion: CRegion is equal to 1 if both artists in a pair are from the same region; 0 otherwise.

CGenre: CGenre is equal to 1 if both artists in a pair are from the same genre; 0 otherwise.

PPviews: PPviews represents the number of provider paged views.

PSongs: PSongs represents the number of songs posted on the provider web page.

PBand: PBand is equal to 1 if the provider belongs to a band; 0 otherwise.

PAudience: PAudience is equal to 1 if the provider performed in front of an audience larger than 700 people; 0 otherwise.

PYActive: PYActive is equal to 1 if the provider was active for more than six years; 0 otherwise.

DPviews: DPviews represents the number of downloader paged views.

DSongs: DSongs represents the number of songs posted on the downloader web page.

DBand: DBand is equal to 1 if the downloader belongs to a band; 0 otherwise.

DAudience: DAudience is equal to 1 if the downloader performed in front of an audience larger than 700 people; 0 otherwise.

DYActive: DYActive is equal to 1 if the downloader was active for more than six years; 0 otherwise.

REFERENCES

Atchade, Yves F. (2006), "An Adaptive Version for the Metropolis Adjusted Langevin Algorithm with a Truncated Drift," *Methodology and Computing in Applied Probability*, 8 (June), 235–54.

Barnard, John, Robert McCulloch, and Xiao-Li Meng (2000), "Modeling Covariance Matrices in Terms of Standard Deviations and Correlations, with Application to Shrinkage," *Statistica Sinica*, 10, 1281–1311.

Dekker, David, David Krackhardt, and Tom A.B. Snijders (2007), "Sensitivity of MRQAP Tests to Collinearity and Autocorrelation Conditions," *Psychometrika*, 72 (December), 563–81.

DeSarbo, W.S., Y. Kim, and D. Fong (1998), "A Bayesian Multidimensional Scaling Procedure for the Spatial Analysis of Revealed Choice Data," *Journal of Econometrics*, 89 (1/2), 79–108.

Faust, Katherine (2007), "Very Local Structure in Social Networks," *Sociological Methodology*, 37 (December), 209–256.

Fienberg, Stephen E., Michael M. Meyer, and Stanley S. Wasserman (1985), "Statistical Analysis of Multiple Sociometric Relations," *Journal of the American Statistical Association*, 80 (March), 51–67.

Frank, Ove and David Strauss (1986), "Markov Graphs," *Journal of the American Statistical Association*, 81 (September), 832–42.

Gelman, Andrew, Xiao-li Meng, and Hal Stern (1996), "Posterior Predictive Assessment of Model Fitness via Realized Discrepancies," *Statistica Sinica*, 6 (October), 733–807.

Granovetter, Mark (1985), "Economic Action and Social Structure: The Problem of Embeddedness," *American Journal of Sociology*, 91 (November), 481–510.

Handcock, Mark S., Adrian E. Raftery, and Jeremy Tantrum (2007), "Model-Based Clustering for Social Networks," *Journal of the Royal Statistical Society*, 170 (March), 301–354.

Hoff, Peter D. (2005), "Bilinear Mixed-Effects Models for Dyadic Data," *Journal of the American Statistical Association*, 100 (March), 286–95.

——, Adrian E. Raftery, and Mark S. Handcock (2002), "Latent Space Approaches to Social Network Analysis," *Journal of the American Statistical Association*, 97 (December), 1090–1098.

Holland, Paul W. and Samuel Leinhardt (1981), "An Exponential Family of Probability Distributions for Directed Graphs," *Journal of the American Statistical Association*, 76 (March), 33–50.

Iacobucci, Dawn (1989), "Modeling Multivariate Sequential Dyadic Interactions," *Social Networks*, 11 (December), 315–62.

—— and Nigel Hopkins (1992), "Modeling Dyadic Interactions and Networks in Marketing," *Journal of Marketing Research*, 26 (February), 5–17.

—— and Stanley Wasserman (1987), "Dyadic Social Interactions," *Psychological Bulletin*, 102 (September), 293–306.

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.

Leskovec, Jure, Kevin J. Lang, Anirban Dasgupta, and Michael W. Mahoney (2008), "Statistical Properties of Community Structure in Large Social and Information Networks," in *WWW 08: Social Networks and Web 2.0 -Discovery and Evolution of Communities*. New York: Association for Computing Machinery, 695–704.

Li, H. and E. Loken (2002), "A Unified Theory of Statistical Analysis and Inference for Variance Component Models for Dyadic Data," *Statistica Sinica*, 12, 519–35.

McPherson, M., L. Smith-Lovin, and J. Cook (2001), "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology*, 27, 415–44.

Nair, Harikesh S., Puneet Manchanda, and Tulikaa Bhatia (2010), "Asymmetric Social Interactions in Physician Prescription Behavior: The Role of Opinion Leaders," *Journal of Marketing Research*, 47 (October), 883–95.

Narayan, V. and S. Yang (2007), "Bayesian Analysis of Dyadic Survival Data with Endogenous Covariates from an Online Community," working paper, Johnson Graduate School of Management, Cornell University.

Pattison, Philippa E. and Stanley Wasserman (1999), "Logit Models and Logistic Regressions for Social Networks: II. Multivariate Relations," *British Journal of Mathematical and Statistical Psychology*, 52 (November), 169–93.

Rivera, Mark T., Sara B. Soderstrom, and Brian Uzzi (2010), "Dynamics of Dyads in Social Networks: Assortative, Relational, and Proximity Mechanisms," *Annual Review of Sociology*, 36 (August), 91–115.

Robins, Garry L., Philippa E. Pattison, Y. Kalish, and D. Lusher (2007), "An Introduction to Exponential Random Graph (p*) Models for Social Networks," *Social Networks*, 29 (2), 173–91.

———, Tom A.B. Snijders, Peng Wang, Mark S. Handcock, and Philippa E. Pattison (2007), "Recent Developments in Exponential Random Graph (p*) Models for Social Networks," *Social Networks*, 29 (May), 192–215.

Snijders, Tom A.B. (2005), "Models for Longitudinal Network Data," in *Models and Methods in Social Network Analysis*, Peter J. Carrington, John Scott, and Stanley Wasserman, eds. New York: Cambridge University Press.

———, Philippa E. Pattison, Garry L. Robins, and Mark S. Handcock (2006), "New Specifications for Exponential Random Graph Models," *Sociological Methodology*, 36 (December), 99–153.

Stephen, Andrew T. and Olivier Toubia (2010), "Deriving Value from Social Commerce Networks," *Journal of Marketing Research*, 47 (April), 215–28.

Sweeting, Trevor (2007), "Discussion on the Paper by Handcock, Raftery and Tantrum," *Journal of the Royal Statistical Society, A*, 170 (March), 327–28.

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

Trusov, Michael, R.E., Bucklin, and K. Pauwels (2009), "Estimating the Dynamic Effects of Online Word-of-Mouth on Member Growth of a Social Network Site," *Journal of Marketing*, 73 (September), 90–102.

Tuli, Kapil R., Sundar G. Bharadwaj, and Ajay K. Kohli (2010), "Ties That Bind: The Impact of Multiple Types of Ties with a Customer on Sales Growth and Sales Volatility," *Journal of Marketing Research*, 47 (February), 36–50.

Van den Bulte, Christophe and Rudy K. Moenaert (1998), "The Effects of R&D Team Co-location on Communication Patterns Among R&D, Marketing and Manufacturing," *Management Science*, 44 (November), Part 2 of 2, 1–18.

——— and Stefan Wuyts (2007), "Social Networks and Marketing," working paper, Marketing Science Institute.

Warner, R., D.A. Kenny, and M. Stoto (1979), "A New Round Robin Analysis of Variance for Social Interaction Data," *Journal of Personality and Social Psychology*, 37 (10), 1742–57.

Wasserman, Stanley and Dawn Iacobucci (1988), "Sequential Social Network Data," *Psychometrika*, 53 (June), 261–82.

——— and Philippa Pattison (1996), "Logit Models and Logistic Regressions for Social Networks: I. An Introduction to Markov Graphs and p*," *Psychometrika*, 61 (September), 401–425.

Watts, Duncan J., and Peter Sheridan Dodds (2007), "Influentials, Networks, and Public Opinion Formation," *Journal of Consumer Research*, 34 (May), 441–58.

Wedel, M. and W.S. DeSarbo (1996), "An Exponential Family Scaling Mixture Methodology," *Journal of Business and Economics Statistics*, 14 (4), 447–59.