# Dynamics of Network Structure and Content in Social Media

Vineet Kumar
Harvard Business School & i-lab, Heinz College, Carnegie Mellon Universityvkumar@hbs.edu,

Ramayya Krishnan
School of Information Systems and Management & i-lab, Heinz College, Carnegie Mellon University, rk2x@cmu.edu,

David Krackhardt
School of Public Policy and Management & i-lab, Heinz College, Carnegie Mellon University, krack@cmu.edu,

Organizations use social media to leverage knowledge contributions by individual employees, which also foster social interactions – activity in blogs, forums, wikis etc. is critical to ensuring a thriving online community. Prior studies have examined contributions to such media at the level of the individual, focusing on drivers of participation, whereas we investigate three different dimensions of dyadic interactions. Our setting is an online forum in an enterprise, where employees both exchange knowledge by query-response and interact socially.

Using a networks approach to query-response behavior, we characterize each interaction as a directed tie, and view the entire set of online forum interactions as a social network. We evaluate network constructs including Simmelian embeddedness and content of relationships (expressive or instrumental), to understand the mechanisms underlying online social interactions.

We find that content and embedded nature of the relationship strongly influence responses: Simmelian ties formed in an expressive setting have the highest positive impact on response propensity, i.e. both content and embeddedness are impactful and reinforce each other. Our results have implications for designing online social communities, specifically that practitioners ought to consider the benefits of purely social interactions through the forum that may serve to lubricate future instrumental interactions.

*Key words*: Online Communities, Social Networks, Simmelian Ties, MRQAP, Dyadic Relationships

## 1. Introduction

Knowledge management systems have proven the object of much investment by firms, with IDC estimating that the expenditure on these systems would be $ 4.8 billion in 2007 (Babcock 2004, Kankanhalli et al. 2005). The sources of knowledge would include employees of the organization, suppliers as well as clients, and the knowledge is coded and stored in multiple formats based on firm and industry practices. These varied forms include document repositories, databases as

well as directories that serve as a reference to experts in specific areas of interest. In fast-paced organizations, such systems may not fully support the informational needs of employees in a timely manner. For example, searching through the large number of resources available may be time consuming even in the presence of advanced search technology, because employees may not know what keywords to search. In addition, in several environments the validity of information over time may be questionable because of rapid technological change (Alavi and Leidner 2001). Even if the knowledge is available, users may also not know exactly what they want, and searching through a repository may be less efficient than an interactive system. As a consequence, static, hierarchical document-type forms of knowledge are increasingly giving way to dynamic collaborative platforms built by individuals associated with the organization.

These deficiencies of static non-interactive knowledge management systems has led firms to turn to organizational social media, where interactions and knowledge exchange are much more free-flowing and unconstrained. Such social media are increasingly proving to be effective, and the so-called "Enterprise 2.0" applications (McAfee 2006) like forums, blogs, wikis, messaging, social bookmarking etc. have been adopted by diverse organizations, including Fortune 100 firms like IBM and Oracle as well as small non-profits like the Little League. The emergence of these organizational micro-communities based on social media are driven by individual contributions, which naturally leads to the question of what factors motivate contributions to achieve a thriving community. Managers would naturally want to know whether responses are coming from the most qualified people, or from individuals known to the original query poster, and it is unclear which one is more desirable. As described in excellent detail by Wellman (2004), online media connect people via dynamic and flexible social networks, and enable participation and connection to be more directly linked.

As a primary component of their social media strategy, many organizations are increasingly implementing some version of an *employee forum* where employees may post queries detailing the information they require or a problem they currently face, and other employees respond to these questions. If the responding employee has the required expertise, this input can often be

more precise and more timely than alternative methods. Moreover, the searching expertise of the querying employee is less critical because of human interpretation. On the web, *Yahoo! Answers* demonstrates the utility of being able to tap into expertise available in specific community areas, ranging from pet food to high technology.[1]

In sum, the key practical differences between traditional knowledge management systems and current social media platforms include:

(i) The presence of an online community that develops in an organic manner, which enables rich multi-person interactions in real time. This also enables directed person-to-person response interaction, which could follow a query posted to the entire group.

(ii) The ability to proceed iteratively in querying information, which is especially useful when the querier is unsure of how to describe the problem or issue.

(iii) The ability to have purely social interactions online that may not be directly related to work, e.g. the weather, or sports.

We take an online social network approach to study interactions (both knowledge transfer and social) within an organization's query-response forum, choosing to focus on dyadic response behavior in addition to aggregate response behavior by individual users. We conceptualize a network tie to indicate a response message from a responding individual to the query poster[2]. Thus, we explore contributions to the forum by dissecting the interactions between individuals at a dyadic level.

Our focus in on understanding the dynamic determinants of online exchanges between individuals using social media, the factors that mediate interaction at the dyadic level, and the resulting implications for understanding design implications.

Our focus is on understanding how and to what degree individuals interact online, the factors that mediate interaction at the dyadic level, and the resulting implications for understanding social interactions and knowledge exchange in an organizational context. We investigate the effect

[1] The context we examine could be considered to be the organizational cousin of *Yahoo! Answers* where employees contribute to the *real-time knowledge* of the organization by answering work-related as well as social queries posted by their colleagues.

[2] We derive network ties exclusively from message traffic and do not observe offline social ties between individuals.

of three major factors on dyadic interactions: *network structure and embeddedness*, *content of interaction (or type of relationship)* and *homophily*. The literature has primarily focused on online community contributions aggregated to the level of the individual, rather than dyadic interchanges. Our objective is to take a more disaggregate approach focusing on analyzing drivers at the dyadic level, especially on how repeated interactions may form between specific members of the community. This question is of interest to academics who want a better understanding on and the extent to which social processes postulated for offline or face-to-face interactions carry over to online settings, where individuals interact with other identifiable people. To practitioners or managers who may be interested in encouraging more work-related versus social interactions, this work has strong implications on the desirability of "social" online interactions, and how they may benefit the firm.

In sum, our primary interest is to use the dyadic approach to answer the following questions:

(1) What kinds of individuals respond to what types of queries or queriers? Do traditional social forces like social closeness (homophily) translate to online communities?

(2) How does history matter? What is the effect of prior communication on future propensity to respond? In other words, do online participants exhibit social inertia?

(3) What is the effect of the type of tie on the likelihood of communication? How does the strength of tie between individuals affect response behavior differently depending on the characteristics of the query? Are relationships transferable across different types of ties? Does a tie established in a work-related (instrumental) content situation affect responses in a social (expressive or non-instrumental) query situation, and *vice versa*?

(4) What role does network structure play in dyadic response behavior? Do individuals in embedded (Simmelian) ties respond differently, compared with those in symmetric or asymmetric non-Simmelian relationships?

(5) From a practitioner's perspective, should managers encourage or discourage forum use for purely *social* topics not directly related to work (e.g. sports)?

To study the above questions and dissect the nature of social interactions in online social media, we build upon the ideas and methods from three disparate areas of study: the organizational

literature examining the content or type of dyadic ties, the structural literature on social network interactions and information retrieval from computer science.

Adapting methods from the structural networks literature, we examine the extent to which structural network theories of embedded relationship apply in online communities, more specifically the implications of Simmel's theory of micro-communities, where triadic embedding is known to influence dyadic relationships in organizational social media. From a methodological perspective, we bring into the IS literature recent enhancements of MRQAP to study dyadic interactions that have been demonstrated to be robust to multiple varieties of structural autocorrelation. We channel and adapt to our setting the ideas from the organizational behavior literature on the content of relationships, specifically the distinction made between expressive (affective or social or non-instrumental) versus instrumental (utilitarian) ties. To operationalize these ties in the setting of an online forum, we use methods from the Information Retrieval literature in Computer Science to construct a quantification of textual information to characterize the type (or content) of ties.

Depending on the designer's other objectives, there are several non-exclusive metrics along which the success of a firm-wide user forum can be judged: number of responses to queries, the quality of contributions, time to respond, average message traffic per day, appropriateness of queries to work-related performance can all be considered reasonable metrics in practice. Another goal might be to enable the transfer of knowledge between groups or divisions within the firm. Whatever the design objective may be, we first need to understand the nature and structure of contributions in order to successfully design an online community and analyze how users might respond to various design decisions.

We believe our results will aid in such design considerations discussed above. We find interesting aspects of response behavior that connect social network structure and embedding, the type or content of any prior ties, as well as the content of the query. Examining the responses at a query level, we find that users receive more responses when they have been active in responding to queries in the past, indicating support for reputational effects or reciporcity. Queries that have a higher information content receive more responses, and belonging to an organizational unit (vertical) with

more co-workers implies a higher likelihood of response. We also find that boundary spanners who are embedded in ties across verticals receive fewer responses than those who choose to focus their ties within their vertical. The query-level analysis helps us understand the response as a function of both the query poster's characteristics and past participation and the characteristics of the query itself.

At a dyadic response level, we show that homophily at the individual employee level (especially the organizational vertical and geographic distance) can positively impact response behavior. In addition, we find that users who have had prior communication in the past are strongly inclined to respond to one another, and the magnitude is especially strong when the past tie is of an expressive type as compared with an instrumental tie. We further find that the network structure established as a result of communications (and derived from communication patterns) significantly influences future propensity to respond. Specifically, expressive ties embedded in Simmelian micro-communities have a very strong impact on responses, suggesting that there is (future) value in establishing interactions that are primarily expressive in nature. Remarkably, an expressive tie has a positive impact even if a current interaction is of an instrumental type.

Our findings suggest that managers may want to create and support an online community environment that fosters social interactions, and looking upon such exchanges as a distraction may be a myopic approach that does not consider the positive long-run impact of such interactions. This unique focus on the dyadic response behavior helps us understand the nuances underlying communications depending on the content and prior network ties, and uncovers patterns that would remain hidden when using only user-level or query-level responses.

## 2. Theoretical Foundation

The study of online social interactions has a large number of studies drawing from several disparate disciplinary backgrounds, including organizational behavior, computer science, social psychology, human-computer interaction and information systems. Faced with the impossible task of providing a comprehensive review of each of these well-developed streams of literature, we instead opt to characterize our selective review the literature based on the substantive issues studied in this paper.

### Activity or Contribution Levels

Activity in a social media community drives individuals to participate both as contributors and consumers of information and knowledge, and researchers have found activity to be a major factor in successfully predicting whether communities thrive or wither away (Butler 2001, Kim 2000, Ludford et al. 2004). Specifically, in cases where users post a query or information request and do not receive any response, they are much more likely to abandon the community. Such factors are also especially important for individuals who have yet to reach a high degree of involvement with the community. Butler (2001) also asks the critical question of what makes communities sustainable and whether a small but committed set of users can make a community thrive or whether membership size is the most important factor. He finds that membership size of the community and its activity level are tightly interlinked in determining whether a community thrives: larger communities find it easier to attract members, but more difficult to retain them, whereas smaller communities face the opposite problem.

Social media have the property that certain actions by one individual (say, posting a query or a comment) induce further reactions by others in the community (responses). Thus, the activity of an individual can have a multiplier real-time effect on the community, especially if the issue is of the nature of a discussion. This property is distinct from traditional knowledge management systems. Even lurkers who may be considering whether to participate in the community are more likely to expect higher benefits when they observe other requests receiving attention, which increases the likelihood that the lurker's query receives a useful response.

Given the evidence that activity is critical to success, it becomes important to understand what drives activity, beginning at the level of the individual. Do individual users contribute more-or-less equally to communities or is there a wide divergence of participation, where most free-ride on the contributions of a few? Whittaker et al. (1998) evaluated community-level participation patterns in USENET newsgroups and found empirical evidence that participation is highly skewed and that a few individuals dominated discussion. In his examination of communities of practice, Wenger (1998) details the process of social learning by evaluating the different roles played by users

in online communities and how they influenced other users' decisions on how to participate. He also suggests several design criteria to ensure that communities support social learning effectively. Rashid et al. (2005) use predictions from social psychology and that contributions are helped when users are reminded of their value or the uniqueness of their contribution as well as when they are given specific goals. In a recent study by Kraut et al. (2010), individual users are conceptualized as forming a relationship with a community that can evolve over time, with the initial experience received by newcomers (who responds and how often, status of responders etc.) determining nature and degree of further participation. Also, the newcomers' nature of interaction with the group by signaling similarity or interest can influence how the group perceives them. A complementary study by Borgatti and Cross (2003) examines how individuals decide whether or not to seek information from specific others, and evaluates the factors that influence this decision. In contrast, our focus is on responses *after* individuals have decided to seek help from the online forum, and not from a specific person.

In addition to examining participation at an individual level, we provide a perspective that is complementary to the above studies by focusing on how individual users form ties with *specific* others in the community over time, and the effect of these ties on future interactions.

**Type (or Content) of Tie - *Instrumental versus Expressive or Non-instrumental***

The type or content of tie between individuals has been known to influence interactions, as well as the evolution of such relationships over time. A simple example would be to consider a work relationship with a colleague as compared with a social relationship with friends and family. To better characterize the content of the relationship or tie between individuals, we use the well-known distinction made in the social network literature between *expressive* and *instrumental* ties (Lincoln and Miller 1979, Fombrun 1982). Expressive ties are have a strong affective (or social) component, and it is possible for the affect to be positive or negative. Individuals form friendships in the workplace with colleagues that may not be in the same work-group, but with whom they may share social interests. Expressive ties can be thought of as interactions based on needs for

social belonging, identification with a group etc. and are more likely to be established with strong norms, interpersonal trust and social purpose. Instrumental ties are based on the idea of resources, either informational or physical, and individuals are conceptualized as forming the relationships that give them access to resources to achieve prestige, status or similar goals. These two different types of tie have served as the foundation of different paradigms of studying social activity, with the expressive tie approach developed by sociologists and the instrumental approach underlying the economic approach to examining social action. Coleman (1988) details these different approaches with the objective of providing an integrated view of rational action and social organization.

Most studies of communities whether online or offline do not characterize the content of conversations that occur between individuals who make up the community. This could be because observing and recording the content of conversations can be logistically challenging, but more importantly there is little research on how to characterize the content. Some studies in information systems have examined content-related metrics (Wasko and Faraj 2005, Jones et al. 2004) using measures like word counts and other metrics based on metadata, but do not focus on the textual content. Recent work by Arguello et al. (2006) and Burke et al. (2007) has identified the effect of different rhetorical strategies on response behavior by adopting a linguistic framework operationalized with automated tools from computer science text mining area. However, while such characterization provides a notion of complexity or familiarity, it does not serve our aim of differentiating ties as being of the instrumental or expressive variety.

We complement the above studies by developing measures to characterize the textual content of online interactions in an automated manner, which permits us to determine the type of ties established by the query-response interaction. In our analysis, we focus on interactions at the level of a thread, which is a query and the set of responses associated with the query, and use the textual content of the queries with the aid of techniques from information retrieval and text processing to establish the content of each interaction. In our setting, it is possible for a pair of individuals to have interactions with only an expressive tie, only an instrumental tie, both types of ties, or neither type of tie. Our object of interest is whether and how these different types of ties impact response

behavior at the dyadic level. We hypothesize that when individuals interact primarily interact in an instrumental setting, they are not likely to driven by social considerations or affect, and hence the effect of a tie is unlikely to influence future interaction behavior. On the other hand, ties that are primarily expressive in nature may be expected to have a stronger positive impact on future interactions.

**Homophily**

Succinctly stated as the aphorism 'birds of a feather flock together', the idea of homophily is simply that that people tend to form social ties with others who share similar characteristics with them. The interpretation of characteristics can be construed broadly to include both innate characteristics like gender, age or race and chosen characteristics like occupation or political affiliation (McPherson et al. 2001). Although homophily has been posited and commonly observed in offline settings, recent studies have also confirmed that individuals interacting in primarily online settings like e-mail also exhibit a propensity to interact with other "similar" to themselves (Yuan and Gay 2006, Aral and Alstyne 2007).

We are interested in examining homophily factors like age, tenure in firm, gender etc. as well as factors relevant to our setting like location and organizational vertical, to control for homophily factors.[3] The forum designed or practitioner may want to encourage diverse audiences to respond to postings, but may find that challenging if homophily is a strong factor. On the other hand, since the setting is online and equally accessible to all individuals associated with the organization independent of location or group, we might expect homophily to matter less in interactions.

**Network Structure and Embeddedness**

The final component of our theoretical framework incorporates the structure of ties within the social network formed by interactions in the online social medium, to characterize the embedding of individuals in micro-communities.

[3] In is common practice in information technology service firms to be organized according to client-facing units known as verticals, *e.g.* Insurance or Automotive

The basic nature of dyadic ties embedded in a community is based on Simmel (1950)'s notion of the importance of triads in changing the behavior of dyadic relationships. His approach countered the standard practice of considering interpersonal relations, or dyadic network ties, as the basic building blocks for social structure, arguing that the fundamental building block is the triad, not the dyad. The intuition is that triads provide contextual background that can change the meaning or importance of the dyads embedded within them, i.e. triadic features can impede, enhance or otherwise alter relations contained therein. In particular, when three or more actors all are joined together, they form a micro-community or "clique". Isolated dyads permit more individuality of each of the two actors, because the actors are on equal footing, whereas a dyad embedded in a clique reduces the individual's options because at any moment the individual can be outnumbered by others enforcing the norms of the group. Consider a bargaining situation: the power of the individual in an isolated dyad is enhanced by simply threatening to withdraw, leaving their partner with no alternatives, whereas in a clique such a threat of withdrawal has less impact since the remaining party still has the other clique members they can work with. In the case of conflict, when two conflicted partners are comparably connected to a third party, the mere presence of this third party can help to moderate the conflict as each conflicted party prefers not to appear unduly harsh in front of the neutral third other.

As micro-communities form and develop, they take on their own dynamics, identities, and norms of behavior. Not only the individuals but also the relations among these individuals become governed in part by the micro-community, with each dyad subject to the rules and obligations implicitly imposed by this overarching micro-community. Simmel argued that this notion of a micro-community is more critical than the sheer strength of a tie or frequency of interaction. [4]

Thus, what matters is whether the dyad is embedded in a group, not how big the group is, or how strong the relationship is. We have come to refer to such ties, those explicitly embedded in a

[4] To quote Simmel:
"Dyads thus have very specific features. This is shown not only by the fact that the addition of a third person completely changes them, but also, and even more so, by the common observation that the further expansion to four or more by no means correspondingly modifies the group any further." (Simmel, 1908/1950, p. 138)

micro-community of at least three individuals as "Simmelian Ties".

Several empirical investigations of Simmelian ties have allowed us to test and extend Simmel's theory. Researchers have shown that Simmelian ties reveal the strong underlying bedrock of structure that helps to clarify how embedded relationships operate. For example, Krackhardt (1999) found that roles inferred from Simmelian ties went a long way to explain why some participants in a unionization attempt quietly receded to the background while others were active agents (both for and against the union). A comparable role analysis based on all ties, including non-Simmelian ties, provided no such clarity, as if the non-Simmelian ties were simply adding noise to the structure. A very different setting involving scientists employed in corporations found that R&D scientists' patent productivity was strongly predicted by cross-cutting patterns of Simmelian ties, as opposed to normal ties, strong ties, or reciprocated ties (Tortoriello and Krackhardt 2010).

The strong evidence that Simmelian tie theory is useful in studying how networks influence structure and processes in organizations leads us to examine whether Simmel's theories and insights will apply to interactions in online community structures with the same predictive power. An interesting issue is whether norms can develop in online settings. If it could, then users are not just engaging in words, but understanding and responding to affective components underlying the interaction. These arguments are related to media richness theory, which classifies media into different degrees of "richness" based on the types of cues that can be communicated with the medium – face-to-face is the richest medium and e-mail is considered a leaner medium (Daft and Lengel 1986). Researchers applying this to online media have found distinctions in how individuals respond based on the medium, supporting the theoretical predictions (Dennis and Kinney 1998). The key question then is the following: can the online forum provide a medium rich enough to support the social factors necessary for the maintenance of norms, which allow Simmelian ties to influence interactions?

A critical question for our purposes here is whether Simmelian ties when formed in an online forum setting have any bearing on further activities between the individuals embedded in those ties. To our knowledge, this question has not been examined within the IS literature on networks. One

reason for that may be the relative anonymity of online forums, where users may find it difficult to associate individuals with their online identities. The setting we examine involves a forum where employees use their real names, not assumed identities, to post both query and response messages to the forum. Hence, the issue of Simmelian ties, where identity in a group or community is key, is more likely to be relevant in our context.

In his seminal article on the strength of weak ties, Granovetter (1973) posited that tie strength can be characterized by the following elements: the amount of time spent interacting, the emotional intensity of the interaction, the extent of mutual confiding and the degree of reciprocal services performed. Therefore, a related issue we examine is how Simmelian ties differ from dyadic ties, especially compared to strong and reciprocated, but non-Simmelian dyadic interactions. If these two types of ties have different effects, then a natural question is to examine the interaction between the structural and content aspect of ties. In this study, we include non-Simmelian ties but dichotomize them to indicate presence or absence of ties, but do not specifically evaluate the effect of tie strength in our online community.

## 3. Data and Operationalization of Constructs

We begin by describing the underlying interaction data in the online community that serves as the basis for deriving social interactions, which in turn permits the construction of the overall social network, as well as the textual content that helps determine the types of ties between individual users over time.

### 3.1. Data and Measurements

We use archival data from the enterprise employee forum of a top-5 global technology service provider firm. The firm's enterprise systems collect data at the message level as well as the user level. Our data covered the period from the beginning of the forum's launch in August 2006 to August 2007, implying that the left-truncation problem is minimized in our setting. Each message is classified as either a query or a response and the firm captures metadata like employee identity of the message poster, the date and time it was posted and a unique identifier indicating the

thread the message belonged to. A query and its corresponding responses are organized in the form of a thread. We also obtain the textual content of each message that was posted, including both the subject and the body of the message. A separate database collected from the human resources department captured data on individual employees including the unique identity number of the employee, the client-facing vertical (similar to division or department) of the employee, the physical location characterized by city and country, the tenure of the employee in the firm as well as demographic variables like age and gender. The data is summarized in Table 1.

**Table 1 Employee Forum Data**

| **Message-level Data** | **User-level Data** |
|---|---|
| Identity (Employee ID of the message poster) | Employee ID |
| Type of message (Query or Response) | Age |
| Timestamp | Tenure in the firm |
| Thread ID of the thread containing the message | Gender |
| Message Content (Subject and text of message) | Location |
| | Vertical (or Group) |

**Table 2 Summary Statistics**

| **Characteristic** | **Value** |
|---|---|
| Total number of users participating | 2974 |
| Total number of queries | 20090 |
| Total number of responses | 59038 |
| Average responses per query | 2.9 |
| Average messages per day | 162 |
| Average time to first response in minutes | 58 |
| Number of users only posting queries | 343 |
| Number of users only posting responses | 1377 |
| Number of users posting queries and responses | 1004 |

We divide the message traffic data in the online forum into two periods to evaluate how the forum evolved, and to characterize the effect of prior ties on participation behavior. The first period is August 2006 – March 2007 inclusive, whereas the second period covers April 2007 – August 2007.

This split between periods was chosen to ensure a roughly equal volume of message traffic (queries and responses) across the two periods.[5]

We present summary statistics of messages posted in the forum covering both periods in Table 2, and there are a few points worth expanding upon. First, the number of users participating in the forum is close to 3,000, and which is comparable to several Internet forums. The message traffic per day is much more than most focused Internet forums, but several general purpose forums can have thousands of messages per day. There is a significant segment of users who only post queries or only post responses as well as a large subset of users who actively participate along both the query and response dimensions. The average time to first response indicates that queries on average receive responses within an hour of posting, and is probably adequate for all but the most time-sensitive users, as long as the response is satisfactory.

### 3.2. Operationalization of Theoretical Constructs

In this section, we describe how we formulate each of the factors that we previously theorized about. For response behavior and ties, we obtain the network from querying and response postings made by individual users. Note that this online community network structure is induced *exclusively* by the messages posted in the forum, and is different from other offline social networks that may exist between the employees. We characterize the sequence of queries and responses as a social network between the users who participate in threads.

**A network model of queries and responses** The directed response graph in Figure 1 captures the dyadic nature of responses in each time period.[6] When a user $i$ responds to user $j$, a directed link is formed from $i$ to $j$. The tie strength of the directed link $i \rightarrow j$ is simply the number of times that $i$ has responded to $j$ in a given time period. So, an arrow represents at least one response from a user to a query poster, with the arrow head pointing to the query poster. When this information is aggregated over all threads, we obtain a directed graph of response behavior.

[5] Splitting the data into identical-length time periods results in similar qualitative effects.

[6] Non-numbered ties are assumed to have a value of 1

In Figure 1, user 2 has posted eight responses to queries posted by user 1 and one response to a query by user 4 in time period 1. Also, user 2 has received two responses from user 5 and one response by user 4 to her queries. From this directed response graph, we derive several structural constructs that we list in Table 3. The query-response network can be represented by an adjacency matrix $\mathbf{R}$ (which we call response matrix) which captures the response behavior at a dyadic level, *i.e.* $\mathbf{R}_{i,j}$ indicates the number of responses by user $i$ to user $j$. Note that this value is expected to be less than the number of questions user $j$ has asked, provided users don't respond multiple times to the same query. The diagonal elements are represented by 'X', which are structural zeros.

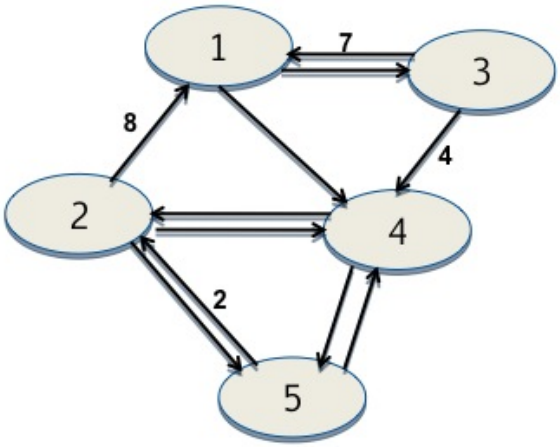


**Figure 1** **Directed Graph of Responses in Period** 1

**Table 3** **Structural Properties of Response Network**

$$\mathbf{R} = \begin{pmatrix} X & 0 & 1 & 1 & 0 \\ 8 & X & 0 & 1 & 1 \\ 7 & 0 & X & 4 & 0 \\ 0 & 1 & 0 & X & 1 \\ 0 & 2 & 0 & 1 & X \end{pmatrix}, \mathbf{T_{NZ}} = \begin{pmatrix} X & 0 & 1 & 1 & 0 \\ 1 & X & 0 & 0 & 0 \\ 1 & 0 & X & 1 & 0 \\ 0 & 0 & 0 & X & 0 \\ 0 & 0 & 0 & 0 & X \end{pmatrix}, \mathbf{T_Z} = \begin{pmatrix} X & 0 & 0 & 0 & 0 \\ 0 & X & 0 & 1 & 1 \\ 0 & 0 & X & 0 & 0 \\ 0 & 1 & 0 & X & 1 \\ 0 & 1 & 0 & 1 & X \end{pmatrix}$$

$\mathbf{R}$ is adjacency matrix, $\mathbf{T_{NZ}}$ indicates non-Simmelian ties, $\mathbf{T_Z}$ indicates Simmelian ties

**Simmelian Ties** In the setting of the query-response network, how is a Simmelian tie to be defined? We require each user in a Simmelian tie to be both a query poster and a responder in a period, to ensure that ties are bi-directional. A Simmelian tie is established only when a user is

part of triad or larger group where a group is defined by the graph theoretic notion of clique – all users are connected to all other others by a bi-directional relationship. Referring to Figure 1, we find that three users are embedded in a Simmelian relationship. These are users $2, 4$ and $5$, so each pair in this triad, *i.e.* $(2,4)$, $(2,5)$ and $(4,5)$ is said to have a Simmelian tie. However, users 1 and 3 have a symmetric tie between them that is not embedded within a larger micro-community, so the relationship between 3 and 5 would be characterized as non-Simmelian.

Mathematically, we can represent the presence of a Simmelian tie between two users as:

$$\mathbf{T_Z}(i,j) = 1 \iff \mathbf{R}(i,j) > 0 \wedge \left[\mathbf{R^T}(i,j) > 0\right] \tag{1}$$

$$\wedge \left(\exists k \notin \{i,j\} \;\; \texttt{s.t.} \;\; \mathbf{R}(i,k) > 0 \wedge \mathbf{R^T}(i,k) > 0 \wedge \mathbf{R}(j,k) > 0 \wedge \mathbf{R^T}(j,k) > 0\right) \tag{2}$$

The firm term, in square brackets, requires there to be a direct bi-directional interaction between $i$ and $j$, and the second term, in parentheses, requires the presence of a third individual $k$ who has a bi-directional tie with both $i$ and $j$. The presence of a Simmelian tie between users can also be derived from the response matrix $\mathbf{R}$ as follows:[7]

$$\mathbf{T_Z} = \mathbf{S} \cdot \mathbf{S^2}, \text{ where } \mathbf{S} = \text{sign}\left(\mathbf{R}\right) \cdot \text{sign}\left(\mathbf{R^T}\right) \tag{3}$$

where the operator $\cdot$ indicates element-by-element product of matrices. This evaluation of Simmelian ties dichotomizes the response matrix, so we do not consider information regarding the strength of ties. Recall that Simmel's theoretical argument posits that the triadic structure is what matters, not the tie strength in a dyadic or triadic context.

A non-Simmelian response tie is simply a response tie in the absence of a Simmelian tie between the users. We use the example from Figure 1 where user 3 has a tie with user 4 that is not embedded in a Simmelian tie. The asymmetric non-Simmelian tie matrix is defined as:

$$\mathbf{T_{NZ}} = \text{sign}(\mathbf{R}) - \mathbf{T_Z} \tag{4}$$

Note that since $\text{sign}\left(\mathbf{R}\right) = \mathbf{T_Z} + \mathbf{T_{NZ}}$, these two types of ties are mutually exclusive and exhaustive.

[7] Prior to performing this multiplication, all diagonal elements of $\mathbf{R}$ must be replaced by zero.

**Characterizing Instrumental and Expressive Ties** One of the objectives of this study is to explicitly examine the differences between online interactions of an expressive nature from those of an instrumental nature. The idea of distinguishing between the two types of ties is to get at the underlying social versus utilitarian motivation for providing responses. Since instrumental queries are more likely to be of interest to several users, those who make a contribution to a high-instrumentality query are more likely to receive recognition from their peers in the community than other users who contribute in a low-instrumentality query setting. When an individual $i$ responds to a query by $j$ that is very general, and of interest to many others, then we consider that to be a tie of an instrumental type $i \stackrel{I}{\rightarrow} j$. On the other hand, if the query deals with a narrow subject or topic that is unlikely to be of general interest to the community, then we characterize the tie as being expressive, or $i \stackrel{E}{\rightarrow} j$. In pursuit of operationalizing this distinction, we use the textual content of a query message to derive a measure of the type of tie. We formalize the measure of instrumentality, called *subject popularity* based on the procedure described below.

The procedure to provide the subject popularity for each query involves two steps. The first step is to use a reference corpus of documents, to create a measure of popularity for each document, and the second step is to create a matching between each query and the set of reference documents that match the query most closely. We detail the steps as follows:

(i) We use Wikipedia as a broad general-purpose database of articles and subjects that are of interest to the population. [8] Within Wikipedia, some articles are much more popular than others, e.g. an article on Cancer is more likely to be of general interest than an article on DNA mutations. In a technical setting, an article on operating systems may be more popular than the article on microkernels. It remains to construct a measure the subject popularity of each article.

We use the fact that more general-purpose articles will have many more links to them from

[8] We have used the entire set of forum postings as a corpus (as an alternative to Wikipedia) and this yields similar results.

other articles within Wikipedia. Links from other articles serve as a vote on the subject popularity of the article in question. Thus, the subject popularity of a focal article in Wikipedia is characterized as a function of the number of other articles which link to the focal article.

(ii) Having developed the measure of subject popularity for Wikipedia articles that form our reference corpus, we next collect the words and phrases in each query posted in the online community forum. The text of the query is used to find articles in Wikipedia that are most similar to the query. We use the open-source INDRI search engine that examines the relationship between the words and phrases in a query with the words in each of the Wikipedia articles to determine which articles provide the best set of matches for the query. As the output of the process, INDRI generates a likelihood of match between each document and the query. This process is illustrated in Figure 2.[9]

For our purposes the search and matching process above provides the likelihood that a given article matches the query under consideration. This is used to generate a ordered list of Wikipedia articles $A_1, A_2, \ldots, A_N$ and their associated log-likelihood scores $L(A_1), L(A_2), \ldots, L(A_N)$.

To evaluate the subject popularity (SP), which is our instrumental construct we note that the higher the number of in-degree (or in-links) to an article, the higher is its prestige. It is well known that in-degree is a first-order approximation to eigenvector centrality[10]. Therefore, we take the articles returned by INDRI and evaluate how many other articles link to the returned articles. Thus we calculate the subject popularity of query $k$ denoted as $\mathbf{SP_k}$ to be:

$$\mathbf{SP_k} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{INDEGREE}(A_i(k)) \tag{5}$$

[9] First, we obtain a snapshot of the entire set of articles in Wikipedia, provided by the INEX project (Denoyer and Gallinari 2006). We index the articles using the INDRI search engine to evaluate all the words and terms in the set of articles. (See Strohman et al. (2004) and `http://ciir.cs.umass.edu/~metzler/indriretmodel.html` for details of the INDRI engine.) This captures the frequency of each word in the entire corpus of Wikipedia articles to assess which terms are commonly used and which ones are relatively infrequent. Next, we pass the full text of each query posted in the forum to INDRI, including both subject and the message. INDRI searches through the index to retrieve a set of documents (Wikipedia articles), which most closely match the words and terms present in the query. INDRI uses a language modeling approach (Ponte and Croft 1998), which determines the probability that a query's textual terms are capable of being generated by any of the indexed documents. The language model is combined with an inference network, which essentially combines information from multiple sources (Turtle 1991).

[10] Google's original PageRank™ algorithm to calculate the quality of web pages was based on ideas derived from eigenvector centrality

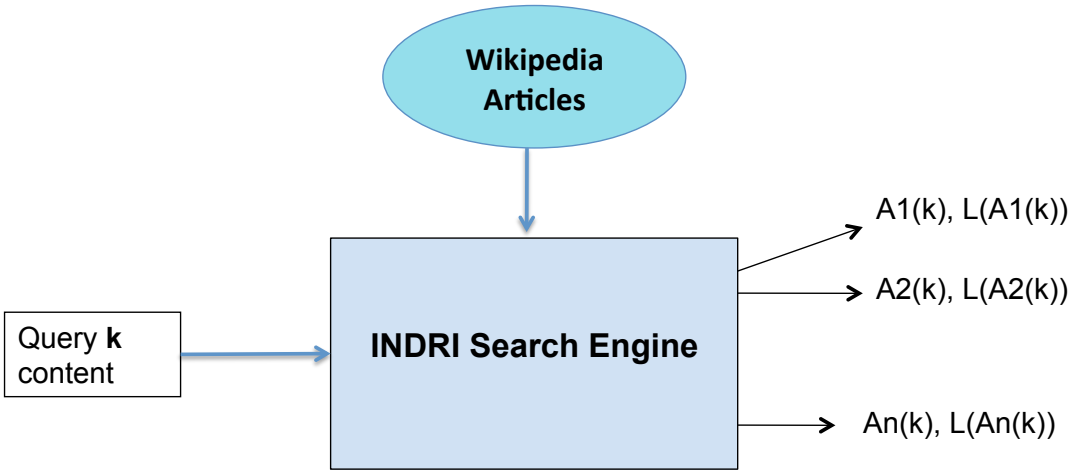


**Figure 2 INDRI Retrieval Process**

Finally, we must categorize queries as instrumental or expressive. Towards this end, we evaluate the median subject popularity of all queries $k = 1, 2, \ldots, K$ as $\tilde{\mathbf{SP}} = median(\mathbf{SP_1}, \mathbf{SP_2}, \ldots, \mathbf{SP_K}$. The type of tie $t_k$ established when responding to query $k$ is determined as:

$$t_k = \begin{cases} E & \text{if } \mathbf{SP_K} < \tilde{\mathbf{SP}} \\ I & \text{if } \mathbf{SP_K} \geq \tilde{\mathbf{SP}} \end{cases} \tag{6}$$

In each period, we aggregate the number of dyadic interactions across all threads that are posted in that period. Consider the following example: in period 1, user $i$ has posted 3 responses to user $j$ that involved expressive queries, and 4 responses to instrumental queries. Then we note that $i$ has both expressive *and* instrumental interactions with $j$ in period 1. Thus, each dyadic pair of individuals can have only an instrumental tie or interaction, only an expressive interaction, both types of interactions or neither type of interaction in each time period.

To derive the Simmelian relationships established in a particular tie type, say expressive, we begin by deriving the network considering *only* expressive ties. If $i$ and $j$ have a Simmelian relationship as defined in (1) in this expressive tie network, then $i$ and $j$ have an expressive Simmelian relationship.

In addition to the subject popularity metric to represent instrumental versus expressive ties, we also derive a measure called *Information Content* (IC) to capture how much informational detail is present in each query.

$$\mathbf{IC_k} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{L}(A_i(k)) \tag{7}$$

Recall that the likelihood of a match with an article indicates strong similarity of query terms with article terms, the higher the likelihood of match of query terms with the articles in Wikipedia, the higher is the information content. We expect queries with a higher information content to receive more responses, simply because more information is provided and readers are likely to be less uncertain about the subject under discussion. On the other hand, if queries have low IC, it could also indicate the matter is more open to interpretation and discussion and could receive more responses.

## 4. Empirical Methodology

Since our objective is to examine both behavior at a thread (i.e. a query and corresponding responses) level as well as evaluate determinants of dyadic interaction, we employ two separate empirical procedures. In the thread-level evaluation, we examine characteristics of a thread using a count model whereas for the dyadic effects, we use a robust version of the MRQAP approach to evaluate dyadic response behavior in the presence of network structural autocorrelation. For both approaches, we use the data divided into two periods, with approximately equal amounts of message traffic, as detailed in §3.1. The list of variables used in the thread-level and QAP regressions is summarized in Table 4.

### 4.1. Thread-level Regression

Our dependent variable of interest is the number of responses to a query (thread length), which is a count variable. Therefore, the appropriate methodology would be to use Poisson regression or a

similar count regression model. Since we do not want to constrain our data to a model where the mean and variance are the same, and the number of responses is at least one in most cases, we choose to use a Negative Binomial regression model instead of a Poisson model or a zero-inflated model (Cameron and Trivedi 1998). In our analysis, both the dependent and explanatory variables are at the thread-level, and we incorporate the characteristics of the query poster.

To determine the determinants of collaboration across organizational units (verticals), we evaluate the effect on the number of internal and external responses that are received by a query posted in the second period (April 2007-August 2007). Here, internal implies that the query poster and the responder belong to the same vertical and external refers to them being in different verticals. For thread $k$, we denote the user who posted the query as $p(k)$. Our model specification is:

$$\mathbf{N_{k,t=2}} = \mathbf{f^{thread}}\Big(\mathbf{NQUERY^{p(k)}_{t=1}}, \mathbf{NRESPONSE^{p(k)}_{t=1}}, \mathbf{SIMMELTIE^{p(k)}_{t=1}}, \mathbf{EISIMMEL^{p(k)}_{t=1}}, \mathbf{VSIZE^{p(k)}}, \mathbf{LOWSPHIGHIC_k}, \mathbf{HIGHSPLOWIC_k}, \mathbf{HIGHSPHIGHIC_k}\Big)$$

where the dependent variable can be $\mathbf{N_{k,t=2}} = \mathbf{N^{I}_{k,t=2}}$ or $\mathbf{N_{k,t=2}} = \mathbf{N^{E}_{k,t=2}}$ or $\mathbf{N_{k,t=2}} = \mathbf{N^{T}_{k,t=2}}$, i.e. the number of internal, external or total number of responses to thread $k$.

The function $\mathbf{f^{thread}}$ relating the dependent and independent variables results in the standard Negative Binomial count regression (Cameron and Trivedi 1998). Our specification remains the same for all three dependent variables, which include internal, external or total responses. Among the independent variables, we evaluate the effect of characteristics of the user posting the query: $\mathbf{NQUERY^{p(k)}_{t=1}}$ is the number of queries posted by the current query poster in period 1, $\mathbf{NRESPONSE^{p(k)}_{t=1}}$ is the number of responses posted by the query poster in period 1, $\mathbf{SIMMELTIE^{p(k)}_{t=1}}$ is a dummy variable denoting whether the query poster was part of a Simmelian group in period 1, $\mathbf{EISIMMEL^{p(k)}_{t=1}}$ denotes the E-I index[11] of Simmelian ties for the query poster in period 1. $\mathbf{VSIZE^{p(k)}}$ is the number of employees who are part of the query poster $p(k)$'s vertical.

[11] The E-I index for an individual depends on the number of internal ties (I) and the external ties (E) and is defined as: $EI = \frac{E-I}{E+I}$. The range for the index is $[-1, +1]$ and individuals with only internal ties have an EI of $-1$ whereas those with only external ties have an EI of $+1$. See Krackhardt and Stern (1988) for details.

In addition, we characterize the textual content of the query itself by using three dummy variables: $\mathbf{LOWSPHIGHIC_k = 1}$ if the query is classified as a low subject popularity and high information content, $\mathbf{HIGHSPLOWIC_k = 1}$ if the query has high SP and low IC and $\mathbf{HIGHSPHIGHIC_k = 1}$ if the query has high SP and IC. Note that the baseline case is that of low subject popularity and low information content.

### 4.2. Dyadic MRQAP Regression

A major focus of our study is the evaluation of dyadic effects in the context of the online social communities. The above thread-level analysis does not capture who responds to whom, i.e. whether certain users have a propensity to respond to queries by certain other users. Rather than examine the overall number of responses, the dyadic model seeks to drill down to interactions between a pair of users, to determine the effects of prior ties as well as similarities in demographic and organizational constructs. The general model specification is:

$$\begin{aligned}\mathbf{NRESPONSE_{i,j}^{t=2}} =& \beta_0 + \beta_1 \mathbf{AGEDIFF_{i,j}} + \beta_2 \mathbf{TENUREDIFF_{i,j}} \\ &+ \beta_3 \mathbf{LOCATIONDIFF_{i,j}} + \beta_4 \mathbf{VERTICALDIFF_{i,j}} + \beta_5 \mathbf{NQUERY_j^{t=2}} \\ &+ \beta_6 \mathbf{NRESPONSE_{i,j}^{t=1}} + \epsilon_{ij}^{t=2}\end{aligned}$$

This is the first specification where the dependent variable is the number of responses $i \rightarrow j$, i.e. by user $i$ to queries posted by user $j$ in the second period. The independent variables are a combination of dyadic time-invariant characteristics and time-varying relationship variables. There are four homophily-related characteristics: $\mathbf{AGEDIFF_{i,j}}$ is the absolute difference in age between users $i$ and $j$, $\mathbf{TENUREDIFF_{i,j}}$ is the absolute difference in tenure between users $i$ and $j$ in months, $\mathbf{LOCATIONDIFF_{i,j}}$ is a dummy variable equal to 1 if both users are in the same city, and $\mathbf{VERTICALDIFF_{i,j}}$ is a dummy variable equal to 1 if both users are part of the same organizational unit (vertical). We also expect prior structural (network) relationships to influence current period response behavior. $\mathbf{NRESPONSE_{i,j}^{t=1}}$ is simply the lagged dependent variable.

This basic MRQAP regression should inform us regarding the relative importance of prior network ties and homophily factors. However, we are interested in a more in-depth examination of the

structural network effects. Recall that to classify queries as instrumental or expressive, we calculate the subject popularity (SP) of the query, and evaluate whether it lies below the median SP for all queries or above it. The former are expressive whereas the latter are classified as instrumental queries. Similarly, we determine whether users $i$ and $j$ had a Simmelian tie between them in period one (represented as $i \Longleftrightarrow j$), implying whether that they were part of a micro-community involving at least one other person. We then decompose the response structure along both dimensions, i.e. the type of prior tie depends on both tie content (expressive or low SP versus instrumental or high SP) and the structural relationship (Simmelian, Non-simmelian or None). This decomposition of ties is captured by the following variables: $\mathbf{INSTRUMENTAL\_SIMMEL^{t=1}_{i,j}}$ indicates the presence of a Simmelian relationship that was established between user $i$ and user $j$ constructed from only instrumental ties, (i.e. when considering only responses to instrumental queries, $i \Longleftrightarrow j$). Similarly, $\mathbf{INSTRUMENTAL\_NONSIMMEL^{t=1}_{i,j}}$ indicates whether there was a non-Simmelian tie ($i \rightarrow j$) when considering only prior instrumental ties.

When the prior tie content is expressive (low SP), we have the corresponding variables $\mathbf{EXPRESSIVE\_SIMMEL^{t=1}_{i,j}}$ and $\mathbf{EXPRESSIVE\_NONSIMMEL^{t=1}_{i,j}}$ respectively. Note that a pair of users $(i, j)$ may have either a Simmelian or a non-Simmelian tie or no tie for a tie of specified content: these alternatives are mutually exclusive and exhaustive. However, the expressive and instrumental settings are not exclusive, so that a pair of users may have positive values for both Simmelian instrumental and Simmelian expressive relationship variables.

We evaluate the dependent variable separately to examine the effect of *content of current query*, *i.e.* is the current query more appropriately characterized as instrumental or expressive? We split the second period data to separate out response behavior to instrumental and expressive queries. So, instead of $\mathbf{NRESPONSE^{t=2}_{i,j}}$, we have $\mathbf{INSTRUMENTALRESPONSE^{t=2}_{i,j}}$ and $\mathbf{EXPRESSIVERESPONSE^{t=2}_{i,j}}$.

So, the full specification has two dependent variables regressed against the same set of explanatory variables:

$$(\mathbf{EXPRESSIVERESPONSE^{t=2}_{i,j}}, \mathbf{INSTRUMENTALRESPONSE^{t=2}_{i,j}}) =$$

$$\beta_0 + \beta_1 \mathbf{AGEDIFF_{i,j}} + \beta_2 \mathbf{TENUREDIFF_{i,j}} + \beta_3 \mathbf{LOCATIONDIFF_{i,j}}$$
$$+ \beta_4 \mathbf{VERTICALDIFF_{i,j}} + \beta_5 \mathbf{NQUERY_j^{t=2}}$$
$$+ \beta_6 \mathbf{EXPRESSIVE\_SIMMEL_{i,j}^{t=1}} + \beta_7 \mathbf{EXPRESSIVE\_NONSIMMEL_{i,j}^{t=1}}$$
$$+ \beta_8 \mathbf{INSTRUMENTAL\_SIMMEL_{i,j}^{t=1}} + \beta_9 \mathbf{INSTRUMENTAL\_NONSIMMEL_{i,j}^{t=1}} + \epsilon_{ij}^{t=2}$$

The main methodological issue is the type of regression to use: it is well known that OLS or related methods like GLS and NLS yield biased estimates of the standard errors, which in turn may spuriously show significance (Krackhardt 1988). The primary reason for this problem is that network data are characterized by structural autocorrelation between observations that belong to the same row or to the same column. An example of this can be seen clearly when considering an individual $i$ who has a strong propensity to form ties (or in our context, respond to any posted query). This propensity would imply that elements in a row $i$ of the adjacency matrix denoting ties formed by the user are likely to be highly correlated, because $i$ would be more likely to interact with *everyone*. The case is similar when a given user is likely to have several ties from other users, which may happen in our setting when several others have a high propensity to respond to a specific user's queries. These forms of structural autocorrelation imply that assuming observations to be independent and using OLS etc. can result in highly erroneous conclusions.

The Quadratic Assignment Procedure (QAP) has been shown to be an effective method to account for this structural autocorrelation (Krackhardt 1987). QAP regression has been used to study dyadic properties in several contexts, ranging from friendship networks to trade between countries to differences in perceptions of justice in organizations (Umphress et al. 2003). In contrast to other network statistical methods like Exponential Random Graph Models (ERGM), MRQAP permits us to take advantage of *valued* dyadic relationships whereas the former was designed for data that is dichotomous or can be reasonably dichotomized. An additional reason is that we are more interested in examining dyadic response behavior as our dependent variable, and not network properties like the emergence of triadic or other structures, for which ERGM would be an appropriate methodology.

A basic QAP methodology like the $y$-permutation test works by initially obtaining an OLS estimate for the coefficients. Instead of using the OLS standard error or significance level, we would systematically permute both the rows and columns of the dependent variable. For example, if we have a dependent variable $y$ to be a $k \times k$ matrix, one possible permutation is $1 \to j, j \to k, k \to 1$. This permutation gives us:

$$\begin{pmatrix} y_{11} & \cdots & y_{1j} & \cdots & y_{1k} \\ \vdots & \ddots & \vdots & & \\ y_{j1} & \cdots & y_{jj} & \cdots & y_{jk} \\ \vdots & & & & \vdots \\ y_{k1} & \cdots & y_{kj} & \cdots & y_{kk} \end{pmatrix} \longrightarrow \begin{pmatrix} y_{kk} & \cdots & y_{k1} & \cdots & y_{kj} \\ \vdots & \ddots & \vdots & & \\ y_{1k} & \cdots & y_{11} & \cdots & y_{1j} \\ \vdots & & & & \vdots \\ y_{jk} & \cdots & y_{j1} & \cdots & y_{jj} \end{pmatrix}$$

The idea is that permutations simply relabel the nodes in the network, and doing that ensures that the graph structure and consequently properties like density etc. do not change. However, permuted elements (as shown above) are not likely to have any relationship. So, we permute S=100 or 1000 times to sample from the possible $k!$ permutations. For each of the S permutations, we again run an OLS regression and obtain permuted coefficients. For each independent variable, we calculate the fraction of permutations where the unpermuted OLS coefficient is larger (or smaller) than the permuted coefficient. This fraction represents a non-parametric significance level for the coefficient.

Whereas the above procedure provides a basic intution into how QAP-based methods work, in our empirical analysis we use a recently developed MRQAP that uses a pivotal test statistic and is robust in the presence of skewness and multi-collinearity etc. known as *Double Semi-Partialling* (DSP) suggested by Dekker et al. (2007). The DSP method involves partialing out the effects of each independent variable on the other independent variables.[12]

### 4.3. Weakened Simmelian Ties

The query-response network we consider in by nature directional, with ties directed from the responder to the query poster. In the previous sections, we have assumed the requirement for a Simmelian tie to be bidirectional, i.e. a necessary condition is that both individuals involved in a

[12] The implementation of DSP in R is available by request from the first author.

Simmelian tie (say A and B) must have posted queries and must have responded to the queries of each other. Of course, the other requirement is the existence of another individual C with whom both A and B have bidirectional ties. These requirements are fairly strong and it is worthwhile exploring how a weakened criterion for ties to qualify as Simmelian would might alter the results. We consider a weakening of the definition of a Simmelian tie to include those ties that qualify as Simmelian after ignoring the directionality of the tie, i.e. we convert all unidirectional and bidirectional ties to be bidirectional and then examine whether two individuals are involved in a Simmelian tie. For example, if A has responded to B's queries but B has not responded to A's query we characterize the tie to be bidirectional under this weaker condition. However, if A and B do not have any ties or have a non-Simmelian tie, the weakened criterion does not change that tie status. We test a regression model that captures the effects of both a strong definition and a weakened definition of Simmelian ties to evaluate whether considering directionality is important in this context. For brevity, we leave out the homophily variables in our description and results although they are included in the actual regression estimation.

$$
\begin{aligned}
&\mathbf{NRESPONSE}_{\mathbf{i,j}}^{\mathbf{t=2}} \\
&= \beta_1 \mathbf{EXPRESSIVE\_NONSIMMEL}_{\mathbf{i,j}}^{\mathbf{t=1}} + \beta_2 \mathbf{EXPRESSIVE\_WEAKSIMMEL}_{\mathbf{i,j}}^{\mathbf{t=1}} \\
&+ \beta_3 \mathbf{EXPRESSIVE\_STRONGSIMMEL}_{\mathbf{i,j}}^{\mathbf{t=1}} + \beta_4 \mathbf{INSTRUMENTAL\_NONSIMMEL}_{\mathbf{i,j}}^{\mathbf{t=1}} \\
&+ \beta_5 \mathbf{INSTRUMENTAL\_WEAKSIMMEL}_{\mathbf{i,j}}^{\mathbf{t=1}} \\
&+ \beta_6 \mathbf{INSTRUMENTAL\_STRONGSIMMEL}_{\mathbf{i,j}}^{\mathbf{t=1}} + \epsilon_{\mathbf{i,j}}^{\mathbf{t=2}}
\end{aligned}
$$

The explanatory variables in this regression are all dummy variables which indicate the tie types, i.e. whether the first period tie between two individuals $i$ and $j$ was a strong Simmelian (considering only bidirectional ties) or a weak Simmelian (converting all ties to be bidirectional) or a non-Simmelian tie We do this separately for expressive and for instrumental ties. Thus, the above regression captures both the base effect of strong Simmelian ties and the residual effect of the weakened definition of Simmelian ties.

## 5. Results and Discussion

We have examined participation behavior at the thread-level which characterizes how many responses a query receives as influenced by the prior participation of the individual who posted the query as well as the characteristics of the query. We have also evaluated the dyadic response behavior that focuses on how an individual $i$ responds to another individual $j$ depending on the query as well as prior tie characteristics between these individuals.

### 5.1. Thread-level Regression

For the thread-level regression, we specified a negative binomial count model for the number of internal, external and total responses. We demonstrate the results for number of internal and external responses in Figure 3 below.

First, we find that there is evidence for a reputational effect: the more active users who have responded to queries posted by others in the past receive more responses for their queries. Second, the presence of a Simmelian tie for the query poster has a strong positive effect on the number of responses received, suggesting significant returns to establishing Simmelian ties. Third, the information content of the query is significant and positive, implying that high IC queries receive more responses. This result intuitively means that queries that contain more detailed information and may be more easily comprehended receive more responses. Fourth, the coefficient of the E-I index is negative and significant indicating that users who maintain more external Simmelian ties that span organizational boundaries receive ferwer responses to their queries. Thus, the community does not seem to reward boundary spanners, and those who maintain internal Simmelian ties receive a higher response rate. Finally, the size of the vertical positively influences the number of responses received, perhaps because users within a vertical are more inclined to respond to their peers resulting from homophily with respect to the organizational unit.

Overall, the thread-level regression informs us that it is important for users in identifiable online communities to take into account the returns from reputational effects when determining whether to contribute, given that the community seems to reward contributors. The subject popularity does

not strongly influence the number of responses at the thread-level indicating that dyadic factors differ significantly from aggregate factors, and that a deeper level of analysis can disentagle the effects of instrumental versus expressive ties.

| | Internal Responses | | | External Responses | | |
|---|---|---|---|---|---|---|
| **Independent Variable** | **Estimate** | **Std. Err.** | **Sig.** | **Estimate** | **Std. Err.** | **Sig.** |
| Number of queries in Period 1 | 0.0025 | 1.29E-03 | . | -0.0037 | 8.65E-04 | *** |
| Number of responses in Period 1 | 0.0037 | 1.30E-03 | ** | 0.0034 | 8.36E-04 | *** |
| Eigenvector centrality of poster in Period 1 | -2.5620 | 2.31E+00 | | 3.6480 | 1.52E+00 | * |
| Presence of Simmelian in Period 1 | 0.3662 | 5.39E-02 | *** | 0.1781 | 3.56E-02 | *** |
| Simmelian EI Index in Period 1 | -0.7267 | 9.19E-02 | *** | -0.0951 | 5.54E-02 | . |
| Dummy for Low SP, Hgh IC | 0.0397 | 3.95E-02 | | 0.1699 | 2.42E-02 | *** |
| Dummy for High SP, Low IC | -0.0967 | 4.14E-02 | * | 0.0513 | 2.51E-02 | * |
| Dummy for High SP, Hgh IC | 0.0875 | 3.78E-02 | * | 0.2265 | 2.33E-02 | *** |
| Size of Vertical of Query Poster | 0.0001 | 2.40E-06 | *** | 0.0000 | 1.46E-06 | *** |

**Figure 3 Thread Regression Results**

### 5.2. Dyadic MRQAP

We estimate the dyadic MRQAP as a linear regression model, and use QAP with DSP for significance tests for all models as detailed in §4.2.[13]

We have estimated multiple model specifications in evaluating the effect of content of tie, i.e. instrumental or expressive, and have considered the effect of the type of prior tie along with the current query setting.

First, we demonstrate the results for a basic MRQAP model with the only time-varying explanatory variable being the number of responses in period 1 is given in Figure 4 (all factors are significant at 95% level). We confirm that prior responses (lagged dependent variable) is the strongest predictor for current number of responses. However, all the homophily variables are also confirmed to

[13] For the MRQAP regression, the adjacency matrix of responses with all users was very sparse with a density of $5.5 \times 10^{-3}$. To counter this sparseness and address practical estimation issues caused by computational complexity and computing resource availability, we only include users who satisfied one of the following three criteria over a 1-year period: (a) posted at least 4 queries (b) posted at least 9 responses or (c) posted at least 8 messages (queries or responses). We set these criteria to obtain a reasonably dense subnetwork that also contained occasional posters. Our main results are not qualitatively sensitive to different cutoff values. This left us with $N = 729$ users and a density of $3.7 \times 10^{-2}$, which is still quite sparse but captures most of the factors driving response behavior. Note that factors that affect the sparseness of the matrix are the number of queries as well as the average length of threads.

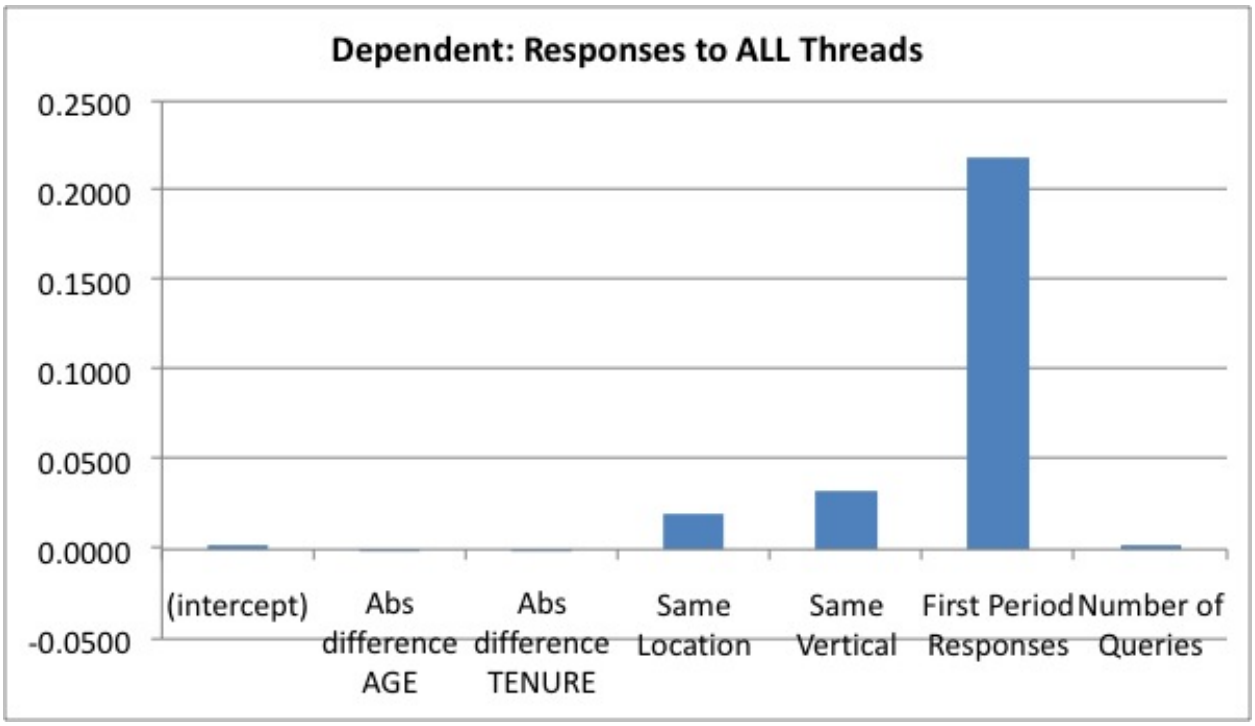


**Figure 4 Basic MRQAP Regression Results**

be significant at $p = 0.01$ (1% level) but have lower magnitude than the coefficient of prior tie.

Having established this, we delve deeper into the content of tie in for dyadic relationships as well as the current query type. The full results of these regressions are in the appendix in 6, but the essential insights are captured by Figure 5 below. The figure demonstrates two structural types of ties, Simmelian and non-Simmelian as well as the content type of the ties. We have also evaluated the effect on responses to a query, which can also be affected by the content (subject popularity) of the current query.

There are several noteworthy points here. First, the presence of a prior tie positively influences response probabilities, i.e. both Simmelian and non-Simmelian ties are significant and greater than zero in all cases. Second, non-Simmelian ties are independent of the content of tie established in prior periods as well as content type of the current query. Third, content type of the current query matters less than the content type of the tie between the pair of individuals. This is surprising since it demonstrates that individuals responses are less sensitive to the type of query than to the type of relationship they maintain with the query poster. This has strong implications for managers who

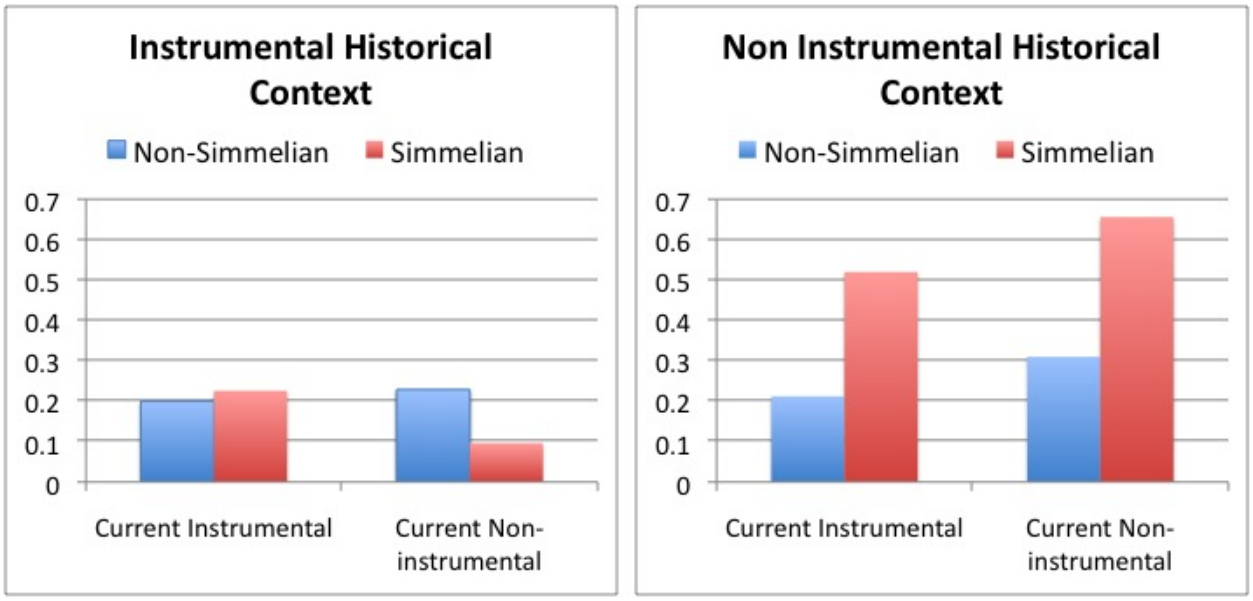


**Figure 5 MRQAP with Detail on Structural Factors and Content of Tie**

may expect that in a well-functioning forum, anyone who posts a query in a clear manner will receive responses. Fourth, if the individuals share an instrumental tie, there is little distinction whether the tie is a Simmelian or non-Simmelian one, i.e. having a Simmelian instrumental relationship brings no further advantage in terms of receiving responses. Last, and perhaps most important, if the individuals have an expressive type of tie, the difference between Simmelian and non-Simmelian ties is stark as shown by the panel on the right.

What exactly does this mean and why? Recall that users were more likely motivated by non-social utilitarian reasons to respond when encountering an instrumental query (high SP). High SP queries being associated with higher status or more prestige, and social or affective aspects are not of primary importance in response behavior in instrumental settings. In contrast, for an expressive (low SP) the non-social reasons offer less power and motivations for responding. The following intuition is consistent with these results: the instrumental types of ties formed have little staying power over time because of the lack of an affective component, and individuals who find themseleves in such ties do not feel bound by norms to respond to peers who interact with them in the future. For the expressive type of tie, the mere presence of such ties provides a higher response likelihood, but is very strongly reinforced when the expressive tie by being embedded in

a Simmelian relationship that is completely established in an expressive setting. Thus, individual users contribute more out a desire to maintain such a social tie, and in a future period such ties have a far stronger impact on the response behavior than either instrumental Simmelian ties or expressive non-Simmelian ties. For a practitioner considering restricting or advocating the use of social media for purely work-related or instrumental purposes, our results offer some caution. The expressive interactions, especially ones that take place in a group or micro-community setting, develop strong bonds and can serve as a social lubricant even for future instrumental interactions.

### 5.3. Weak Simmelian Ties

Our interest here is to examine whether a weakened definition of Simmelian tie captures much of the same effect on response behavior as a strong Simmelian tie that that we have evaluated previously. Referring to Figure 7, we find that most the strongest factor is the strong Simmelian established in a non-instrumental content setting. The weakened definition of Simmelian tie does capture some part of the effect and is in the correct positive direction, but its magnitude is much smaller than the strong Simmelian. Thus, we can infer that users who participate in a strong Simmelian expressive tie setting receive the most future responses.

## 6. Design Implications for Organizational Social Media

What are the implications of our research to designing or managing enterprise social media? We believe that our findings point to the importance of enabling users to establish links that are purely social in content, indeed even encouraging them to do so. Users ought to be encouraged to participate initially by receiving information about how many responses their queries have received as well as how timely the responses were.

In addition, they could be pointed to specific others who have responded to them to make the relationship more salient. This not only results in higher response propensity overall but also with respect to specific other users, thus establishing a relationship mediated by social media. Expressive or social relationships are more likely to result in cooperative response behavior, and these relationships are even more powerful when embedded in groups, i.e. Simmelian expressive relationships. This result suggests that while encouraging bilateral expressive exchanges, interactions

made at a group level are likely to prove even more useful. This is especially important in cases where users seek specialized knowledge that may not be known or be relevant to most others in the organization. In such cases where responses are in short supply, the relationships of an expressive or social nature are much more likely to prove responsive.

A somewhat negative finding from an organizational perspective is that boundary-spanners who establish relationships across organizational boundaries do not necessarily receive more responses. In fact, we find that users who focus their embedded relationships within silos are more likely to be rewarded with more responses. This implication may point to the need to have other mechanisms for recognizing users who work across boundaries since social media usage may not be sufficient in this matter.

## 7. Conclusion

To understand the dynamics of contributions in organizational social media, we have investigated how individual, dyadic as well as group-level factors determine the evolution of an online query-response community established in a multi-national technology services firm. Our primary focus and distinguishing feature in this study is the explicit evaluation of contributions made at a dyadic level, rather than the individual-level analyses that are common in the literature.

Specifically, we contribute to the literature in the following ways: we have dissected the dyadic interactions by distinguishing ties along two dimensions: the first is the content of ties, i.e expressive or social versus instrumental or utilitarian) and the second is the structural characteristics of the ties (Simmelian, non-Simmelian or none). We have contributed to translating Simmel's theory in online social media, and by demonstrating that the type of interactions in which Simmelian ties are established affects current and future cooperative behavior between individuals users. To the best of our knowledge, the content type of interactions and how its interaction with structural features like Simmelian embeddings have not been investigated in either online or offline settings.

We find that multiple factors including those motivated by homophily as well as structural factors are significant predictors of dyadic response behavior. We find that 'history matters' both

at a dyadic level as well in the setting of a larger micro-community in the following sense: when individual users form expressive ties that have a more social or affective aspect, it strongly results in more responses, whereas instrumental ties demonstrate the effect to a lesser degree and are less influential on future cooperative response behavior. In addition, the effect of expressive ties is reinforced when such ties bind a micro-community together with shared norms, than when established in a purely dyadic or user-to-user setting.

From a methodological viewpoint, we develop metrics for instrumental (motivated by external recognition) and expressive (motivated by social cooperation) ties by using techniques from the field of information retrieval to capture textual content in queries. We also demonstrate that the information content of a query is important in obtaining a larger number of responses but don't affect dyadic response behavior.

For future research, we plan to examine the time for establishment of Simmelian and non-Simmelian ties, and evaluate which factors are conducive to the establishment and renewal of these different types of ties. Other avenues that may prove useful is the evaluation of how individual users decide whether to join an online community, the interaction between online and offline social ties and between multiple online media like blogs, wikis and forums.

## Acknowledgments

We acknowledge the helpful comments contributed by participants at the International Symposium on Information Systems, Indian School of Business, Hyderabad and the Statistical Challenges in Electronic Commerce Research conference.

## References

Alavi, M., D. E Leidner. 2001. Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly* 107136.

Aral, S., M. Van Alstyne. 2007. Network structure & information advantage.

Arguello, Jaime, Brian S. Butler, Elisabeth Joyce, Robert Kraut, Kimberly S. Ling, Carolyn Ros, Xiaoqing Wang. 2006. Talk to me: foundations for successful individual-group interactions in online communities. ACM, New York, NY, USA, 959968. doi:http://doi.acm.org/10.1145/1124772.1124916.

Babcock, P. 2004. Shedding light on knowledge management. *HR MAGAZINE* **49**(5) 4651.

Borgatti, S. P, R. Cross. 2003. A relational view of information seeking and learning in social networks. *Management science* **49**(4) 432445.

Burke, M., E. Joyce, T. Kim, V. Anand, R. Kraut. 2007. Introductions and requests: Rhetorical strategies that elicit response in online communities.

Butler, Brian S. 2001. Membership size, communication activity, and sustainability: A Resource-Based model of online social structures. *Information Systems Research* **12**(4) 346362. doi:10.1287/isre.12.4.346.9703. URL `http://isr.journal.informs.org/cgi/content/abstract/12/4/346`.

Cameron, A. C., P. K. Trivedi. 1998. *Regression Analysis of Count Data*. Cambridge University Press.

Coleman, J. S. 1988. Social capital in the creation of human capital. *American Journal of Sociology* **94**(S1) 95.

Daft, R. L, R. H Lengel. 1986. Organizational information requirements, media richness and structural design. *Management science* **32**(5) 554571.

Dekker, D., D. Krackhardt, T. A. B Snijders. 2007. Sensitivity of MRQAP tests to collinearity and autocorrelation conditions. *Psychometrika* **72** 563581.

Dennis, A. R, S. T Kinney. 1998. Testing media richness theory in the new media: The effects of cues, feedback, and task equivocality. *Information Systems Research* **9** 256274.

Denoyer, Ludovic, Patrick Gallinari. 2006. The wikipedia XML corpus. *SIGIR Forum* .

Fombrun, C. J. 1982. Strategies for network research in organizations. *Academy of Management Review* **7**(2) 280291.

Granovetter, Mark S. 1973. The strength of weak ties. *American Journal of Sociology* **78**(6) 1360. doi: 10.1086/225469. URL `http://www.journals.uchicago.edu/doi/abs/10.1086/225469`.

Jones, Q., G. Ravid, S. Rafaeli. 2004. Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration. *Information Systems Research* **15**(2) 194210.

Kankanhalli, A., B. C.Y Tan, K. Wei. 2005. Contributing knowledge to electronic knowledge repositories: An empirical investigation. *MIS Quarterly* **29** 113143.

Kim, A. J. 2000. *Community building on the web: Secret strategies for successful online communities*. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA.

Krackhardt, D. 1987. QAP partialling as a test of spuriousness. *Social Networks* **9**(2) 171186.

Krackhardt, D. 1988. Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Social Networks* **10**(4) 359381.

Krackhardt, D. 1999. The ties that torture: Simmelian tie analysis in organizations. *Research in the Sociology of Organizations* **16** 183210.

Krackhardt, D., R. N. Stern. 1988. Informal networks and organizational crises: An experimental simulation. *Social Psychology Quarterly* **51**(2) 123140.

Kraut, R., X. Wang, B. Butler, E. Joyce, M. Burke. 2010. Beyond information: Developing the relationship between the individual and the group in online communities. *Working Paper, Carnegie Mellon University* .

Lincoln, James R, Jon Miller. 1979. Work and friendship ties in organizations: A comparative analysis of relation networks. *Administrative Science Quarterly* **24**(2) 181199. URL `http://www.jstor.org/stable/2392493`.

Ludford, P. J, D. Cosley, D. Frankowski, L. Terveen. 2004. Proceedings of the SIGCHI conference on human factors in computing systems. 631638.

McAfee, A. P. 2006. Enterprise 2.0: The dawn of emergent collaboration. *MIT Sloan Management Review* **47**(3) 21.

McPherson, M., L. Smith-Lovin, J. M Cook. 2001. Birds of a feather: Homophily in social networks. *Annual Review Of Sociology* **27** 415444.

Ponte, J. M., W. B. Croft. 1998. A language modeling approach to information retrieval. ACM New York, NY, USA, 275281.

Rashid, A. M., K. Ling, G. Beenen, P. Ludford, X. Wang, K. Chang, X. Li, D. Cosley, D. Frankowski, L. Terveen. 2005. Using social psychology to motivate contributions to online communities. *Journal of Computer-Mediated Communication* **10**(4).

Simmel, G. 1950. *The Sociology of Georg Simmel, ed. KH Wolff* . New York: Free Press.

Strohman, T., D. Metzler, H. Turtle, W. B. Croft. 2004. INDRI: a language model-based search engine for complex queries.

Tortoriello, M., D. Krackhardt. 2010. Activating cross-boundary knowledge: the role of simmelian ties in the generation of innovations. *The Academy of Management Journal (AMJ)* **53**(1) 167181.

Turtle, H.R. 1991. Inference networks for document retrieval. Ph.D. thesis, MIT. URL `citeseer.ist.psu.edu/turtle91inference.html`.

Umphress, E. E, D. J Brass, L. Scholten. 2003. The role of instrumental and expressive social ties in employees' perceptions of organizational justice. *Organization science* **14**(6) 738753.

Wasko, Molly, Samer Faraj. 2005. Why should i share? examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly* **29**(1) 3557.

Wellman, B. 2004. Connecting community: On-and offline. *Contexts* **3**(4) 513.

Wenger, Etienne. 1998. *Communities of practice: Learning, meaning, and identity*. Cambridge: Cambridge University Press.

Whittaker, S., L. Terveen, W. Hill, L. Cherny. 1998. Proceedings of the 1998 ACM conference on computer supported cooperative work. ACM Press New York, NY, USA, 257264.

Yuan, Y. C, G. Gay. 2006. Homophily of network ties and bonding and bridging social capital in computer-mediated distributed teams. *Journal of Computer-Mediated Communication* **11**(4) 10621084.

**Table 4 Variable List for Regressions**

| | |
|---|---|
| | *Thread-level Explanatory variables* |
| $SIMMELTIE_{t=1}^{p(k)}$ | Number of Simmelian ties for poster of query $k$ in $t=1$ |
| $EISIMMEL_{t=1}^{p(k)}$ | E-I index of Simmelian ties for poster of query $k$ in $t=1$ |
| $VSIZE^{p(k)}$ | Number of employees in vertical that poster of query $k$ belongs to |
| $LOWSPHIGHIC_k$ | Whether query $k$ is below median Subject Popularity, above median Information Content |
| $HIGHSPLOWIC_k$ | Whether query $k$ is above median Subject Popularity, below median Information Content |
| $HIGHSPHIGHIC_k$ | Whether query $k$ is above median Subject Popularity and Information Content |
| | *Dyadic Explanatory Variables* |
| $AGEDIFF_{i,j}$ | Difference in age of $i,j$ in months |
| $TENUREDIFF_{i,j}$ | Difference in tenure at company of $i,j$ in months |
| $LOCATIONDIFF_{i,j}$ | Indicator equal to 1 if $i,j$ work at same location |
| $VERTICALDIFF_{i,j}$ | Indicator equal to 1 if $i,j$ belong to the same vertical |
| $NQUERY_j^{t=2}$ | Number of queries posted by $j$ in period 2 |
| $\mathbf{EXPRESSIVE\_SIMMEL_{i,j}^{t=1}}$ | $=1$ if $i,j$ had a Simmelian relationship only with expressive ties in $t=1$ |
| $\mathbf{EXPRESSIVE\_NONSIMMEL_{i,j}^{t=1}}$ | $=1$ if $i,j$ had a non-embedded relationship only with expressive ties in $t=1$ |
| $\mathbf{INSTRUMENTAL\_SIMMEL_{i,j}^{t=1}}$ | $=1$ if $i,j$ had a Simmelian relationship formed only with instrumental ties in $t=1$ |
| $\mathbf{INSTRUMENTAL\_NONSIMMEL_{i,j}^{t=1}}$ | $=1$ if $i,j$ had a non-embedded relationship only with instrumental ties in $t=1$ |
| $\mathbf{EXPRESSIVE\_NONSIMMEL_{i,j}^{t=1}}$ | $=1$ if $i,j$ had an expressive tie in $t=1$ is not embedded |
| $\mathbf{EXPRESSIVE\_WEAKSIMMEL_{i,j}^{t=1}}$ | $=1$ if $i,j$ had an expressive tie in $t=1$ is embedded when direction is ignored |
| $\mathbf{EXPRESSIVE\_STRONGSIMMEL_{i,j}^{t=1}}$ | $=1$ if $i,j$ had an expressive tie in $t=1$ is embedded even when direction is considered |
| $\mathbf{INSTRUMENTAL\_NONSIMMEL_{i,j}^{t=1}}$ | $=1$ if $i,j$ had an instrumental tie in $t=1$ is not embedded |
| $\mathbf{INSTRUMENTAL\_WEAKSIMMEL_{i,j}^{t=1}}$ | $=1$ if $i,j$ had an instrumental tie in $t=1$ is embedded when direction is ignored |
| $\mathbf{INSTRUMENTAL\_STRONGSIMMEL_{i,j}^{t=1}}$ | $=1$ if $i,j$ had an instrumental tie in $t=1$ is embedded when direction is considered |

Variable List for Regressions

MRQAP Regression Results for Full Specification

**Dependent: Responses to ALL Threads**

| | Estimate | Pr(<=b) | Pr(>=b) | Pr(>=\|b\|) |
|---|---|---|---|---|
| (intercept) | 0.0014 | 0.66 | 0.34 | 0.64 |
| Abs difference AGE | -0.0017 | 0.01 | 0.99 | 0.01 |
| Abs difference TENURE | 0.0000 | 0.02 | 0.98 | 0.05 |
| Same Location | 0.0195 | 1.00 | 0.00 | 0.00 |
| Same Vertical | 0.0331 | 1.00 | 0.00 | 0.00 |
| Non-Simmelian First Period Non-Instrumental Responses | 0.5217 | 1.00 | 0.00 | 0.00 |
| Simmelian First Period Non-Instrumental Responses | 1.1742 | 1.00 | 0.00 | 0.00 |
| Non-Simmelian First Period Instrumental Responses | 0.4301 | 1.00 | 0.00 | 0.00 |
| Simmelian First Period Instrumental Responses | 0.3526 | 0.99 | 0.01 | 0.01 |
| Number of Queries | 0.0028 | 1.00 | 0.00 | 0.00 |

**Dependent: Responses to Instrumental Threads**

| | Estimate | Pr(<=b) | Pr(>=b) | Pr(>=\|b\|) |
|---|---|---|---|---|
| (intercept) | 0.0018 | 0.82 | 0.18 | 0.25 |
| Abs difference AGE | -0.0009 | 0.01 | 0.99 | 0.01 |
| Abs difference TENURE | 0.0000 | 0.00 | 1.00 | 0.00 |
| Same Location | 0.0093 | 1.00 | 0.00 | 0.00 |
| Same Vertical | 0.0166 | 1.00 | 0.00 | 0.00 |
| Non-Simmelian First Period Non-Instrumental Responses | 0.2128 | 1.00 | 0.00 | 0.00 |
| Simmelian First Period Non-Instrumental Responses | 0.5183 | 1.00 | 0.00 | 0.00 |
| Non-Simmelian First Period Instrumental Responses | 0.2014 | 1.00 | 0.00 | 0.00 |
| Simmelian First Period Instrumental Responses | 0.2562 | 1.00 | 0.00 | 0.00 |
| Number of Queries | 0.0030 | 1.00 | 0.00 | 0.00 |

**Dependent: Responses to Non-Instrumental Threads**

| | Estimate | Pr(<=b) | Pr(>=b) | Pr(>=\|b\|) |
|---|---|---|---|---|
| (intercept) | -0.0005 | 0.42 | 0.58 | 0.73 |
| Abs difference AGE | -0.0008 | 0.01 | 0.99 | 0.03 |
| Abs difference TENURE | 0.0000 | 0.13 | 0.87 | 0.24 |
| Same Location | 0.0101 | 1.00 | 0.00 | 0.00 |
| Same Vertical | 0.0165 | 1.00 | 0.00 | 0.00 |
| Non-Simmelian First Period Non-Instrumental Responses | 0.3091 | 1.00 | 0.00 | 0.00 |
| Simmelian First Period Non-Instrumental Responses | 0.6561 | 1.00 | 0.00 | 0.00 |
| Non-Simmelian First Period Instrumental Responses | 0.2287 | 1.00 | 0.00 | 0.00 |
| Simmelian First Period Instrumental Responses | 0.0963 | 0.99 | 0.01 | 0.01 |
| Number of Queries | 0.0027 | 1.00 | 0.00 | 0.00 |

**Figure 6 MRQAP Results - Current Query Content (dependent variable) and Prior Tie Content – Instrumental versus Expressive or Non-Instrumental**

**Dependent: Number of responses in Period Two**

| | Estimate | Pr(<=b) | Pr(>=b) | Pr(>=\|b\|) |
|---|---|---|---|---|
| Non-Simmelian in Non-Instrumental Context | 8.701E-01 | 1.00 | 0.00 | 0.00 |
| Weak Simmelian in Non-Instrumental Context | 2.958E-01 | 1.00 | 0.00 | 0.00 |
| Strong Simmelian in Non-Instrumental Context | 1.006E+00 | 1.00 | 0.00 | 0.00 |
| Non-Simmelian in Instrumental Context | 2.426E-01 | 1.00 | 0.00 | 0.00 |
| Weak Simmelian in Instrumental Context | 3.179E-01 | 1.00 | 0.00 | 0.00 |
| Strong Simmelian in Instrumental Context | 6.081E-01 | 0.98 | 0.02 | 0.02 |

**Figure 7 MRQAP with Weakened Simmelian**