PREFERENTIAL ATTACHMENT AND MUTUALITY IN ELECTRONIC KNOWLEDGE NETWORKS

Steven L. Johnson and Samer Faraj
Robert H. Smith School of Business
University of Maryland
College Park, MD U.S.A.
steven.l.johnson@gmail.com **sfaraj@rhsmith.umd.edu**

Abstract

The rapid adoption of Internet technology has accelerated the establishment of platforms for virtual interaction that overcome the inherent time and space limitations of face-to-face communication. The objective of this study is to investigate the individual and network level mechanisms that characterize interactions on these electronic knowledge networks (EKNs). Toward that goal, we develop a simulation model of a thread-based asynchronous EKN and provide results based on 330 runs of the model (simulating a total of 3,643,942 messages generated by 38,860 authors). This study contributes to our understanding of electronic knowledge networks by demonstrating the importance of structural characteristics in influencing participant behaviors. We focus specifically on the role of preferential attachment (the tendency to associate with the most popular participants) and mutuality (the tendency to maintain symmetry in relationships with others) in network formation. By using a simulation method and taking into account the nature of interpersonal ties, the study extends previous mathematical models of network formation to the specific setting of online knowledge exchange between individuals.

Keywords: Electronic knowledge networks, knowledge management, mutuality, preferential attachment, reciprocity

Introduction

The rapid adoption of Internet technology has accelerated the establishment of platforms for virtual interaction that overcome the inherent time and space limitations of face-to-face communication. We propose the term *electronic knowledge networks* (EKNs) to refer to collectivities that rely on electronic communication to exchange knowledge over time and therefore sustain a social network. These networks are sustained through a range of technology including e-mail (Finholt and Sproull 1990; Wu et al. Tyler 2004), USENET newsgroups (Butler 2001; Jones et al. 2004), and organizational discussion groups (Constant, Sproull, and Kiesler 1996; Ravid and Rafaeli 2004; Wasko and Faraj 2005). EKNs are important new ways of organizing that have not been well studied so far (Butler 2001; Fulk and DeSanctis 1995). As with any new phenomena, a number of researchers have offered different definitions each based on a specific emphasis. For example, EKNs focused on practice have been called electronic networks of practice (Wasko, Faraj, and Teigland 2004), those focused on a task have been called virtual groups (Ahuja et al. 2003), and those focused on shared interests and social support have been called virtual communities (Rheingold 1993). We use the electronic knowledge network definition to emphasize the communality of these collectives in terms of electronic mediation, knowledge exchange orientation, and social network structure.

A small, but growing, body of research has explored some aspects of EKNs. Studies emphasizing large groups dynamics have provided insight into group evolution and size limitation (Butler 2001; Jones et al. 2004). Other studies have elaborated the motivations and social mechanisms that lead individuals to voluntarily offer help to unknown others (Constant, Kiesler, and Sproull 1994; Wasko and Faraj 2000, 2005). Most recently, studies inspired by the new science of networks (Adamic et al. 2003; Ravid and Rafaeli 2004) have suggested that EKNs may exhibit scale-free properties (an uneven distribution of ties) (Barabasi and Albert 1999) and small-world properties (partially linked clusters) (Watts and Strogatz 1998). Whether EKNs share consistent and identifiable patterns of interactions in line with the scale-free and small-world phenomena identified in a variety of physical and genetic systems is a tantalizing question that could have important theoretical and practical implications for the study and management of EKNs.

The objective of this study is to investigate the individual and network level mechanisms that characterize interactions on EKNs. Toward that goal, we develop a simulation model of a thread-based asynchronous EKN and provide results based on 330 runs of our model (simulating a total of 3,643,942 messages generated by 38,860 authors). By applying variance reduction techniques (Law and Kelton 2000) we isolated the relative impact of three link-generation mechanisms on three network-level structural characteristics measures. We also report on data collected from six technology-oriented EKNs (72,368 messages over approximately 2 years). From these six groups, we derive actual measures of EKN structure. Thus, the simulation generated measures are then compared to the existing characteristics of six EKNs to evaluate model fit against actual data. This study contributes to our understanding of electronic knowledge networks by demonstrating the importance of structural characteristics in influencing participant behaviors. By using a simulation method and taking into account the nature of interpersonal ties, the study extends previous mathematical models of network formation to the specific setting of online knowledge exchange between individuals.

Theory

Individual motivations to participate in social communication networks are reflected in multiple theories (see Monge and Contractor 2003) and multiple network formation mechanisms (see Newman 2003). Table 1 provides a summary description of three of the network formation mechanisms particularly relevant to EKNs. We focus on these three mechanisms as a parsimonious set supported by previous theoretical and empirical social network studies.

Preferential Attachment in Scale-free Networks

In recent years, research falling under the heading of the "new science of networks" has shown that many physical, biological, informational, and social systems follow network structures that are highly similar and that can be characterized by known statistical properties (for a review, see Newman 2003). These networks are generally characterized by a scale-free distribution of links where a small number of nodes have an extremely high number of linkages to other nodes. For example, the distribution of in-degree, a measure of relative popularity, of websites (Adamic and Huberman 2000), e-mail recipients (Wu et al. 2004), and message board participants (Adamic et al. 2003; Ravid and Rafaeli 2004) have all been identified as examples of scale-free networks.

Significant research interest has been focused on understanding network topology and modeling the processes that lead to such organizing principles (e.g., Barabasi and Bonabeau 2003; Watts 2003). A clearer understanding has emerged regarding the statistical properties of large scale-free networks and how clustering within the network can lead to a small-world phenomenon (every node is reachable from any other node by crossing relatively few links; for a discussion, see Watts 2004). Less well understood are the mechanisms that lead to the development of the network in the first place. In this paper, we focus on the interaction mechanisms that lead to the development of a large-scale knowledge network. Because of our interest in EKNs, where the network links social actors as opposed to physical, biological, or economic entities, we specifically focus on mechanisms that characterize interactions between social actors.

Table 1. Network Formation Mechanisms

Formation Mechanism	Description
Preferential Attachment	The tendency of participants to link to the most popular participants (Barabasi and Albert 1999).
Mutuality	The tendency of pairs of participants (dyads) to maintain symmetry in their relationships (see Wasserman and Faust 1994).
Randomness	The tendency to link to other participants serendipitously (Kilduff and Tsai 2003) or based on nonstructural reasons (e.g., message content). Random links are a common mechanism in many network formation models (see Newman 2003).

Barabasi and Albert (1999) have proposed *preferential attachment* as a the primary mechanism for the formation of scale-free networks. Through preferential attachment, the preferences of new entrants, who they choose to link to in the network, is influenced by the preferences expressed by existing participants. The two key conditions for preferential attachment are (1) an open system with new entrants and (2) the new entrants being aware of and basing their preferences of those of existing participants.

There exists no agreed-upon measure to directly assess the presence or absence of preferential attachment in a network. Because each network may be different, the mechanism may be specific to the setting. Nonetheless, through a combination of mathematical proofs and simulations, a large body of research has described how various functional forms of preferential attachment do lead to the formation of networks with power–law degree coefficients in the range observed through empirical measurements (see Newman 2003). One objective of this study is to assess the applicability of preferential attachment to the setting of EKN formation. That is, to what extent are the actions of new participants influenced by the popularity conferred by existing participants?

The Impact of Social Exchange

In a social network, an important influence on participant behavior is the history of personal interactions between a specific pair of participants. Social network analysis has a long history of studying dyadic relationships within networks (Katz and Powell 1955; Moreno and Jennings 1938). *Mutuality* refers to extent that the relationship between two participants is symmetric. In theory-oriented literature in particular, the terms *reciprocity* and *mutuality* are often used synonymously (e.g., Monge and Contractor 2003). In descriptions of empirical measures, however, reciprocity is a more common term. The most frequently used empirical measures of reciprocity are based on unweighted ties and differentiate only between null and asymmetric (unreciprocated) dyads versus mutual (reciprocated) dyads (Wasserman and Faust 1994). In this paper, we have adopted the use of the term mutuality to emphasize our consideration of both weighted and unweighted measures of dyadic relationships. Both conceptually and empirically, we are concerned with mutuality as it encompasses the symmetry (the numerical balance) of weighted ties.

Multiple theories support the importance of mutuality in electronic knowledge networks. Social exchange theory (e.g., Blau 1964) suggests that as people share knowledge resources with one another, they creates obligations for future exchange. Providing help with an expectation of receiving future aid in return binds the giver and receiver together in a reciprocal gift exchange (Fulk et al. 1996; Kollock 1999). Multiple studies have identified that mutuality is indeed one of several user participation motivations (Constant, Sproull, and Kiesler 1996; Lakhani and von Hippel 2003).

An alternative view of network formation is that participants make their decisions entirely based on personal preferences, not on the preferences of anonymous others. Indeed, although potential members of an electronic knowledge network may lurk for a substantial period of time before joining in conversation and can read archived messages, it is an open question as to what extent that would reinforce existing group preferences or aid them in forming their own personal preferences. Indeed, participants may choose to respond to messages solely on the basis of originating message content or for variety of other private motivations. This deeply subjective aspect of human action, referred to as serendipitous by Kilduff and Tsai (2003), cannot be accounted for by structure-based mechanisms. We refer to this subjective link formation factor as randomness whereby participants form links without regard to relative popularity or past dyadic history.

In summary, theory suggests that preferential attachment, randomness, and mutuality all influence participant preferences during electronic knowledge network formation. Together these three mechanisms provide a parsimonious set of mechanisms supported by theoretical and empirical evidence. Furthermore, compared to alternative mechanisms—such as those based on triadic relationships (e.g., transitivity or cyclicality)—this set is both cognitively less complex for participants to manage as well as computationally less complex for researchers to model!

Next, we develop a model of EKN formation that explores the relative influence of these mechanisms toward specific outcomes. Due to the emergent nature of the EKN formation model and limited prior empirical tests, we do not propose specific hypothesis to test. Instead, we use the model in a simulation experiment to analyze the relationships between formation mechanisms and measures of key network outcomes.

Modeling Electronic Knowledge Network Dynamics

A great deal of research in network theory is analytical. However, analytical models are generally appropriate for a small number of nodes, which make the problem tractable, or when the rules of model construction can be specified in a deterministic way (see Watts 1999). For large-scale networks where the interactions are primarily random, such as EKNs, a more fruitful approach is to use a rigorous process of numerical simulation. The use of simulation avoids the need to make simplifying mathematical assumptions for the sake of generating a tractable solution and enables the direct measurement of multiple network measures. The simulation approach also enables the modeling of more complex network behaviors such as taking into account tie strength and directionality.

In this section, we present three related representations of a threaded electronic knowledge network. First, in the agent-based model, we describe the behaviors, rules, and tendencies that govern a threaded discussion group. Next, we propose an analytical model, including specific functional forms and distributions for a threaded EKN. Finally, we describe our implementation of the analytical model in an experimental simulation design. Because of the limited amount of existing research in this area, we used empirical observations from reference groups to derive some of the analytical model and simulation parameters. As such, we took particular care when designing and analyzing the simulation runs and results to ensure that manipulated parameters, not the fixed input parameters themselves, were responsible for the observed effects.

Agent-Based Model

To explore the role of participant structural properties in the formation of EKNs, we focus on a specific, common platform for EKN formation: threaded discussion groups. We take an agent perspective where each node represents a participant on the network. The communication activities reflected by the stream of messages in a threaded discussion group reflect a distinct network configuration. Once they enter the system (corresponding to the step of joining the EKN), each participant can initiate a new thread, contribute to an existing thread, or leave the system. Table 2 summarizes the relationships between threaded discussion participant behavior and the association network representation.

Consistent with previous studies (Ravid and Rafaeli 2004; Wasko and Faraj 2000, 2005) we model EKN links as being formed between a single poster and a single recipient. For each message that is posted to a thread (including the message that starts a thread), the next message sent by a different author to that thread is considered the “from” message for the immediately preceding message in that same thread (the “to” message). Unlike some “help desk” EKNs that develop the convention that the thread starter asks a question to which all other thread respondents respond (or else they are considered impolitely “off-topic”), this study models the more general threaded discussion group behavior where a thread continues in a free-flowing discussion and each participant response is, in general, typically focused on the immediately preceding message.

This modeling approach is consistent with structural network theories in that (1) activity such as unanswered messages and messages sent in response to oneself are not reflected in the network representation, (2) activity is cumulative and directed resulting in a network representation that is weighted and directed, and (3) the number of responses a participant receives (their in-degree) is constrained by the number of messages they send (their out-degree).

Table 3 summarizes participant behaviors, system rules and assumptions, and participant tendencies that characterize a typical thread-based discussion forum.

Table 2. Network Formation from Threaded Discussion Participant Behavior

Threaded Discussion Participant Behavior	Network Representation
Enter System	Eligible to Initiate Ties
Start New Thread	Eligible to Receive Ties
Add to Thread	Establishes Tie Rij
Exit System	No Further Ties Initiated

Table 3. EKN Formation: Threaded Discussion Group
Participant Behaviors Participants are limited to these actions: • Enter system • Post a message that starts a new thread • Post a message that adds to existing thread • Exit system
System Rules and Assumptions • All participants share one class of behaviors and tendencies • Participants may arrive at any time • Any participant in the system can start a new thread • Any participant in the system can add to an existing thread • Participants may depart the system at any time • After a participant exits the system they may not reenter
Participant Tendencies Participants are *inclined* to • Start new threads • Add to a thread immediately after a message is posted by an author who has a relatively large number of previous replies compared to all other authors (Preferential Attachment) • Add to a thread immediately after a message is posted by an author they have never replied to before but who has replied to them before (Mutuality) • Add to a thread immediately after a message is posted by an author they have replied to less often than that author has replied to them (Mutuality) • Add to a thread without regard to structural characteristics (Random) Participants show a *disinclination* to • Add to a thread immediately after a message is posted by an author they have replied to more often than that author has replied to them (Mutuality)

Analytical Model

In this section, we provide an analytical model with mathematical specifications for the agent-based model described above. We performed a preliminary analysis of 9,060 messages over 2 years from a technology-oriented EKN to develop initial functional forms for system entry, system exit, and thread creation. Preferential attachment, the core mechanism to explain the emergence of scale-free networks, was modeled according to the Barabasi and Albert (1999) model. Because previous empirical measures of mutuality represented the construct as simply the presence or absence of a tie (see Wasserman and Faust 1994), we developed our own representation of mutuality as a weighted and directed form of reciprocity between two nodes.

In order to develop robust characterizations of EKN networks, we gathered full interaction data from six technology-oriented EKNs. Table 4 provides summary statistics for these six reference groups. As part of a larger study of electronic knowledge networks, we gathered a full set of summary site statistics for approximately 600 threaded discussion groups that use a common technology platform. Because the intent of this study is to develop a normative model of successful electronic network formation, we eliminated outlier cases of extremely large groups as well as the very smallest and newest groups. From the remaining set of groups, we randomly selected a convenience sample of six groups that retain a full set of online message history and member profiles.

For example, during the first 30 months of existence 1,356 members registered at the CNEW site with the cumulative membership growth pattern closely fitting (adj. $R^2 = 0.91$) a linear model. The cumulative distribution function for node departure at the CNEW site is $0.06919 + 0.0607 \text{ LN } (M)$ where M is the number of months of membership (adj. $R^2 = 0.97$). Although nearly 70 percent of participants remain active for less than 1 month, the probability of departure diminishes with each successive month of participation. Similarly, we model participant arrival with a fixed probability in each period and node departure with an equivalent departure CDF (scaled for the simulation duration).

Table 4. Representative Electronic Knowledge Networks							
Site*	**Description**	**Inception**	**Active Authors**	**Messages**	**kIN**	**CC1**	**MI**
PRES	ASP and web programming	May 20, 2003	360	3481	1.305	0.259	0.577
CNEW	Programming for beginners and experts	May 6, 2002	420	9060	1.540	0.425	0.651
DEEG	Support forum for a consumer product	October 1, 2001	596	14872	1.716	0.489	0.516
DBPW	Support forum for a PC software package	April 1, 2002	892	13103	1.111	0.240	0.607
WINX	Windows OS	July 29, 2001	1298	16602	1.268	0.330	0.533
PFLW	Support forum for a PC software package	March 2, 2002	1515	15250	1.072	0.380	0.610

*URLS: PRES: **www.programmersresource.com**; CNEW: **codenewbie.com**; DEEG: **www.deegruenig.com/forum**; WIN: **forums.winxpcentral.com**; DBPW: **forum.dbpoweramp.com**; PFLW: **www.pictureflow.com/forum**.

The likelihood of a participant posting a message that starts a new thread is a simple probability function. For our simulation model, we use the value of 0.18 observed for CNEW. The likelihood of a participant posting a message that adds to a thread is more involved. The probability of a participant *i* adding to a thread immediately after a message is posted by an author *j* is specified as

$$P(r_{ij}) = \beta_{PA} * \frac{j.indegree}{\sum_{j=1..N} j.indegree} + \begin{cases} 0 & if\ (r.ij + r.ji) = 0 \\ \beta_{MUTL} * \frac{(r.ji - r.ij)}{(r.ij + r.ji)} & if\ (r.ij + r.ji) > 0 \end{cases} + \beta_{RAND}$$

where β_{PA}, β_{MUTL}, and β_{RAND} are values that determine the magnitude of the preferential attachment, mutuality and randomness effects (respectively).

As a rigorous test of the model's robustness that is consistent with a holistic multilevel multi-theoretical view of the network, we use three distinct outcomes measures.[1] First, a common measure of scale-free networks is the power-law degree distribution coefficient (Barabasi and Albert 1999), often referred to as *k* (Newman 2003). As we are specifically measuring the distribution of in-degree, the total number of messages sent to an author, we refer to this measure as *kIN*. Second, another frequently measured network property is clustering coefficient (Wasserman and Faust 1994). Of the two conceptually related (although empirically distinct) measurements of clustering coefficient, we use the measure *CC1* (Newman 2003). Also called the ratio of transitive triples, it can be thought of most generally as "the probability that two people with a common acquaintance know one another" (Newman 2004). As our third measure, we use the mutuality index (Katz and Powell 1955), which we refer to as *MI*, a measure of tendency toward dyadic reciprocity within a network. In summary, the combination of a weighted network level measure (*kIN*), an unweighted, undirected triadic measure (*CC1*), and an unweighted, directed measure (*MI*) provides a robust multilevel method for analyzing network outcomes.

[1]There are numerous alternative network-level measurements we could also consider. We selected these three as outcomes of practical and theoretical relevance to EKNs that also encompass multiple measurement characteristics. These measures all frequently appear in studies of large-scale networks.

Simulation Parameters

With a generative probabilistic model including three primary mechanisms and three network outcome variables with not only directed and undirected relationships but also weighted relationships, the model is not amenable to a direct mathematical solution. Therefore, we developed a simulation following the agent-based model of an EKN formed in threaded discussion, one consistent with the functional forms described earlier.

As part of data collection in a larger study, we gathered (via a web agent) the entire message history from inception through November 2004 for the online bulletin board **codenewbie.com**. The tagline of this electronic knowledge network is “A Programming Environment for Everyone.” The primary resource of the site, outside of a small number of articles and code snippets, are active forums covering a wide range of topics related to computer programming. As noted in Table 5, this archival data served as the basis for several input values in a simulation design.

We used a variation of the 2^k simulation experimental design (Law and Kelton 2000) whereby we ran our simulation multiple times (10 per) at various combinations of four levels of the preferential attachment factor and three levels each of the mutuality and randomness factors. This generates a total of 36 (4 × 3 × 3) different scenarios and 360 (36 × 10) different simulation runs.

Table 6 shows the actual values used for each factor as well as the fixed probability used for starting new threads.

Our initial testing of the simulation demonstrated that the outcomes measures were highly sensitive to differences in the number of messages generated during a scenario run. Because the intent of a simulation method is to compare alternative configurations in comparative conditions, we used the variance-reduction technique of common random numbers (Law and Kelton 2000) for the system exit probabilities. Thus, by holding the experimental condition of node departure constant, it eliminated a major source of undesirable variance within groups of multiple runs of the simulation with the same values for the four levels. These initial tests were also used to select the range of the simulation experiment values with the objective of generating networks with a similar number of messages as those in the reference groups.

Table 5. Simulation Experimental Design

Constructs	Source
# of Iterations	180
Enter System	Based on observed values
Exit System	
Start New Thread	
Preferential Attachment	Tested at 3 or 4 levels for each mechanism (36 total combinations)
Mutuality	
Random Ties	
Power-Law Distribution Coefficient	Outcome measures
Clustering Coefficient	
Mutuality Index	

Table 6. Simulation Experiment Variables

Behavior	Mechanism	Simulation Values
Start New Thread	Fixed Probability	0.18
Add to Existing Thread	Preferential Attachment	0, 0.5, 1, 2
	Mutuality	0, 0.5, 1
	Randomness	0.05, 0.1, 0.2

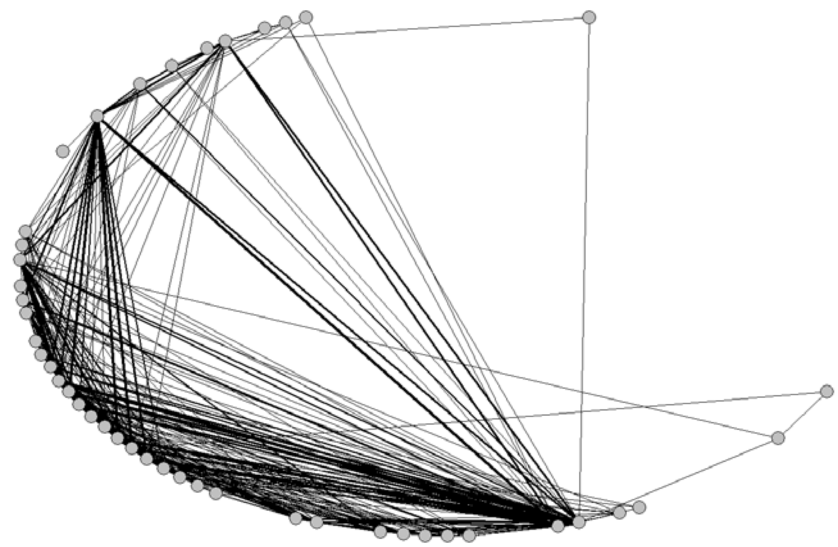

Figure 1. Network Representation of a Simulation Run

Each simulation run consisted of a set of nodes entering according to the probability functions described above. Each run of the simulation had 180 iterations of new nodes enter the system, existing nodes exiting the system, nodes starting threads, and nodes adding messages. The average simulation run resulted in 108 authors posting a total of 9,790 messages.

At the conclusion of each simulation run, the message history for that run was converted into a network representation in order to calculate the outcomes measures. An example of a visual representation of a simulation network is included as Figure 1. Circles represent message authors and the lines represent the relative number of messages sent or received by the authors. The diagram provides a visual indication of both a skewed popularity distribution (thick lines are concentrated among a relatively small number of nodes) as well as clustering (pairs of lines both heading to popular nodes suggest the appearance of transitive triples).

Results

Control Variables

The first analysis we perform is to assess the extent to which our outcomes measures reflect differences between groups. In an experimental simulation design, if the manipulated variables perfectly manipulate the outcome measures, then the adjusted R^2 should approach 1.0 (Law and Kelton 2000). Two reasons why an ANOVA might not approach complete explanation of variance are (1) that the outcome measures are influenced by unaccounted for values such difference in size or (2) that higher-order effects, not included in the model, are responsible for differences (Law and Kelton 2000). Of our three outcome measures, only the mutuality index is specifically intended as a statistical measure of tendency that includes an explicit adjustment for network size.

Table 7. Adjusted R^2 Values

Dependent Variable	Base Model: Controls Var. Only	Full Model: Base + 3 Mechanisms
kIN	.703	.921
CC1	.577	.753
MI	.099	.944

Table 8. Comparison of Simulation and Reference Group Values						
	Simulation Results (n=330)			**Reference Groups (n = 6)**		
Variable	**Min.**	**Avg.**	**Max.**	**Min.**	**Avg.**	**Max.**
kIN	0.875	1.589	1.955	1.072	1.335	1.716
CC1	0.245	0.429	0.520	0.240	0.354	0.489
MI	0.153	0.725	0.927	0.516	0.582	0.651

To demonstrate that the three mechanisms themselves account for variation in the outcome measures, rather than other sources of unintended variances, we can compare two ANOVA models: one with the control variables of the number of authors and the number of messages to one with those two variables plus the three mechanisms. As shown in Table 7, the control-variable only model has adjusted R^2 values of kIN = .702, CC1 = .577; and MI = .099 compared to .921, .753, and .944 for the model including the main effects of the three mechanisms. This provides strong evidence that the three mechanisms account for significant variance in the simulation outcomes above and beyond the control variables.

Finally, as a further robustness check of the model, we compared the range of simulation results against our reference groups. As shown in Table 8 the range of simulation output values match up well with the range of reference group values. The average simulation result falls within the range of reference group values for two of the three variables (kIN and CC1). Also, the simulation result range generally covers the reference group range—the only exception is that the minimum reference group value for CC1 of 0.240 is less than the minimum simulation result of 0.245. Together, the model R^2 values along with the range of output variables provide reasonable evidence that the manipulated variations in the formation mechanisms are responsible for significant variance in the outcome measures and that the simulation model successfully represents the phenomenon of interest.

Primary Analysis

The results of ANOVA analysis for each of the three dependent variables are presented in Table 9. The table includes a relative indicator of effect size, partial Eta-squared (η_p^2) (Cohen 1973), "the proportion of total variation attributable to the factor, partialling out (excluding) other factors from the total nonerror variation" (Pierce et al. 2004, p. 918). Figures 2, 3, and 4 provide graphical depictions of the relative variance explained by the main effects for the three outcome variables.

As shown in the Table 9 and Figure 2, only the mutuality and preferential attachment mechanisms are statistically significant ($p < .001$), the randomness mechanism is not. The mutuality mechanism has the largest relative effect on the power-law distribution coefficient with preferential attachment having a relatively small effect. This is a surprising result. It suggests that the presence of preferential attachment by itself is neither a necessary nor sufficient condition for the existence of a power-law in-degree distribution in an electronic knowledge network. Within the constraints of a threaded discussion board, mutuality has a much larger relative contribution to the formation of scale-free networks.

Table 9. Results of ANOVA Analysis (n = 360)						
	kIN		***CC(1)***		***MI***	
Variable	**SS**	η_p^2	**SS**	η_p^2	**SS**	η_p^2
Intercept	0.447	0.011^{***}	0.018	0.067^{***}	0.048	0.050^{***}
Authors (Control)	0.262	0.135^{***}	0.001	0.003	0.024	0.025^{**}
Messages (Control)	0.039	0.023^{**}	0.021	0.078^{**}	0.214	0.189^{***}
Preferential Attachment	0.447	0.210^{***}	0.077	0.239^{***}	0.187	0.170^{***}
Mutuality	2.011	0.545^{***}	0.022	0.083^{***}	5.427	0.855^{***}
Randomness	0.023	0.014	0.019	0.073^{***}	0.305	0.249^{***}
Error	1.680		0.245		0.917	
Corrected Total	21.811		1.018		16.791	
Model Adj. R^2	0.921		0.753		0.944	

* $p < .05$, ** $p < .01$, *** $p < .001$

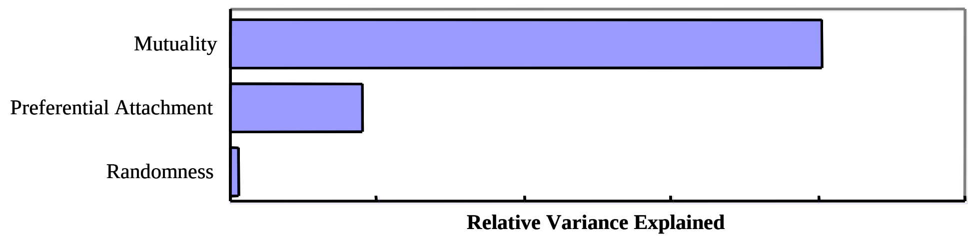

Figure 2. Relative Effects on Power–Law Distribution Coefficient (*kIN*)

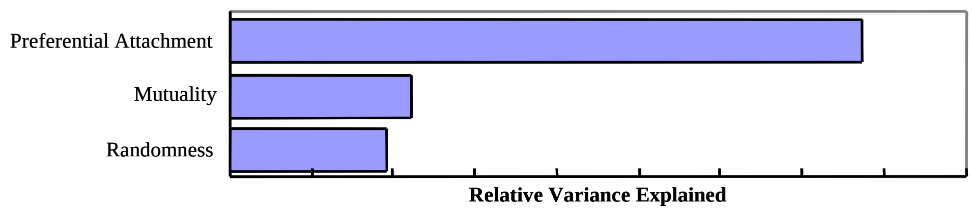

Figure 3. Relative Effects on Clustering Coefficient (*CCI*)

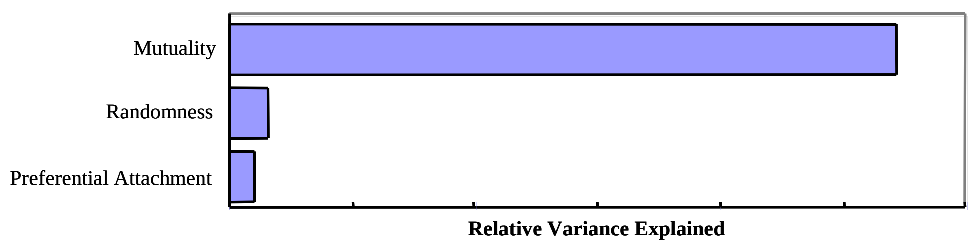

Figure 4. Relative Effects on Mutuality Index (*MI*)

An intuitive explanation for this finding is that mutuality serves as a critical mechanism to motivate the most popular participants to create messages that they would otherwise not create. It is these additional messages, in turn, that provide others with the opportunity to fulfill their motivation to respond to the most popular others. This, in turn, reinforces the perception of popularity, thereby feeding the cycle of the "richer getting richer." Following this reasoning, mutuality is critical for the appearance of a power–law distribution in threaded discussion boards, as compared to other network models, specifically due to the constraint that each message can have at most a single response.

As shown in the Table 9 and Figure 3, all three mechanisms are significant ($p < .001$) with preferential attachment having the largest effect on the clustering coefficient. The overall model fit (adjusted R^2 of 0.753) with statistically significant mechanisms is a surprising result in that the clustering coefficient is a triadic measure and none of the mechanisms in the model are specifically designed to influence triadic relationships. An intuitive explanation is that preferential attachment increases the probability that any two participants with a link will also have a link to the most popular participants. Thus, the concentration of popularity would, in turn, increase CC1, the percentage of transitive triples.

As shown in the Table 9 and Figure 4, all three mechanisms are statistically significant ($p < .001$) with the mutuality mechanism having by far the largest relative effect on the mutuality index. This is an expected result as the mutuality index is an unweighted measure of mutuality. Randomness and preferential attachment have statistically significant effects on the mutuality index yet their effect sizes are of little practical significance.

Additional Analysis

In order to test our results in a multivariate sense, we performed a MANOVA analysis to investigate the relation between our model and the three measures of network structure. We found that both control variables (number of authors and number of messages), and all four main effects were significant ($p < .0001$) when tested with Pillai's trace, Wilk's lambda, Hotelling's trace, or Roy's largest root. This provides further statistical evidence of a strong relationship between the link formation mechanisms and the outcome variables.

Because several of the simulation functional forms and parameters are based on a specific electronic knowledge network, a rival explanation is that the constraints of the threaded discussion system itself, rather than any of the theorized structural link mechanisms, could be responsible for the observed outcomes. To assess this possibility, we compared the mean values of 50 simulation runs with only randomly generated ties against the observed values for the CNEW electronic knowledge network (*kIN*, 1.540; *CC1*, 0.425; *MI*, 0.651). A *t*-test found a statistically significant difference ($p < .001$) for all three outcome variables. This provides further support for the expression of nonrandom link formation mechanisms such as preferential attachment, reciprocity, and commensurateness in actual EKNs.

Discussion

Research Implications

This study used a simulation to explore the nature of interaction on electronic knowledge networks. Little was known about such network-level interaction dynamics. Our contention was that any network model of EKNs may need to go beyond the mechanism of preferential attachment and take into account the nature of the ties linking actors. We found that both mutuality and preferential ttachment have a strong effect on network structure. The effect of each factor varied based on the specific measure of network structure. Mutuality had the strongest effect on scale-free network formation, while preferential attachment had the strongest effect on network clustering measured by the percentage of transitive triples. Finally, mutuality has the strongest effect on the network mutuality index. Taken together, these results provide a more complex representation of how complex networks in general, and of EKNs in particular, develop.

So far, few studies of EKN dynamics and knowledge exchange have targeted the network level. In focusing on structural relationships between participants, our work complements previous empirical research that relied on surveys to measure EKN participant motivations and social processes (e.g., Faraj and Wasko 2003; Wasko and Faraj 2000, 2005). By studying the phenomenon at the network rather than individual level, we are able to contribute a deeper understanding to the reasons why, at an aggregate level, individuals may participate and contribute knowledge in online communities. Our finding regarding the importance of mutuality confirms previous individual level findings as to the importance of social capital and trust in virtual settings (Jarvenpaa and Leidner 1999; McKnight et al. 2002; Stewart 2003). Further, the modeling approach allows us to investigate the impact of changes in parameters on network structure.

Our model also contributes to the growing body of research focused on complex or large-scale networks. Our findings regarding the effect of mutuality on scale-free network formation suggest an alternative mechanism to Barabasi and Albert's (1999) widely studied preferential attachment mechanism. In the quest to develop a general model for scale-free network formation, researchers in the new science of networks have focused on a simplified representation of actor behavior. Such a representation may be appropriate to a variety of physical or biological settings, but for networks emerging from human interaction, a more careful examination of the network context is required. Our findings thus provide support for a more nuanced view of online dynamics and the interplay of resources, motivations, and social ties that drive individual participation in online groups (e.g., Butler 2001; Wasko and Faraj 2005).

A contribution to network research is in the modeling of mutuality as an index of the reciprocity balance in a relationship. Previous work has used a simple representation of mutuality as whether a tie had or had not been reciprocated. While this representation may be beneficial in terms of analytical tractability, the lack of a standard term for weighted mutuality and the dearth of empirical studies measuring weighted mutuality speaks to the poverty of attention to this form of mutuality. Our findings of the significant effect of relationship imbalance in scale-free network formation speak to the importance of studying this aspect of mutuality further.

Practical Implications

Sponsors and participants of electronic knowledge networks face a myriad of challenges when they seek to intentionally motivate sharing of information in electronic forums. A challenge is the lack of clearly identifiable metrics for monitoring the state of an EKN. Further, there are complex indirect relationships between changes in individual level behaviors and network level measures. A practical implication of this study is that interventions designed to increase participation motivation in an EKN, such as fostering a welcoming environment through intentional introduction and acknowledgement activities (e.g., norms of mutuality), could also increase the likelihood of a power–law distribution and, therefore, be interpreted as a failure even when successful. Alternatively stated, fostering a norm of mutuality may increase both the concentration of participation by a small number of highly popular participants while also increasing the number of different people who participate, albeit with far less (relative) frequency.

Conclusion

This study is not without limitations. The agent-based simulation approach faces external validity concerns similar to those of a laboratory study. A second limitation relates to the subjectivity of the modeling effort. However, our use of empirical data from six different EKNs does provide partial relief against the threat of model relevance. Finally, although the experimental design is based on standard simulation techniques, further replication and additional sensitivity testing is required to confirm that the results are not idiosyncratic.

Future research can extend this study in several different directions. By testing a range of simulation parameters, the model can be used to further our understanding of actual electronic knowledge networks. By comparing observable input parameters and observable output parameters for an EKN against simulations with similar parameters, we can impute the relative weight of mechanisms such as preferential attachment that are not directly measurable. Further, this model can be extended beyond threaded discussion groups to a wider range of electronic knowledge networks. Finally, the assumption in this model of a single class of agents can be extended to multiple agent classes with differentiated tendencies (e.g., classes based on agent tenure, experience, or nonnetwork-related attributes).

In closing, this study demonstrates the value of a network level approach to studying electronic knowledge networks. The simulation model presented in this paper contributes by identifying, for the first time, the relative effect of network formation mechanisms on three core measures of network structure. The study results provide support for the importance of taking the strength and symmetry of social ties into account in generating network structure. We believe that elaborating the model and explaining the role of mutuality is a contribution to both the new science of networks as well as to online knowledge exchange.

References

Adamic, L. A., Buyukkokten, O., and Adar, E. "A Social Network Caught in the Web," *First Monday* (8:6), 2003.

Adamic, L. A., and Huberman, B. A. "Power–Law Distribution of the World Wide Web," *Science* (287), 2000, 2115a.

Ahuja, M. K., Galletta, D. F., and Carley, K. M. "Individual Centrality and Performance in Virtual R&D Groups: An Empirical Study," *Management Science* (49:1), 2003, pp. 21-38.

Barabasi, A.-L., and Albert, R. "Emergence of Scaling in Random Networks," *Science* (286), 1999, pp. 509-512.

Barabasi, A.-L., and Bonabeau, E. "Scale-Free Networks," *Scientific American* (288:5), May 2003, pp. 60-69.

Blau, P. M. *Exchange and Power in Social Life*, Wiley, New York, 1964.

Butler, B. S. "Membership Size, Communication Activity, and Sustainability: A Resource-Based Model of Online Social Structures," *Information Systems Research* (12:4), 2001, pp. 346-362.

Cohen, J. "Eta Squared and Partial Eta Squared in Fixed Factor ANOVA Designs," *Educational and Psychological Measurement* (33:1), 1973, pp. 107-112.

Constant, D., Kiesler, S., and Sproull, L. "What's Mine Is Ours, or Is It? A Study of Attitudes about Information Sharing." *Information Systems Research* (5:4), 1994, pp. 400-421.

Constant, D., Sproull, L., and Kiesler, S. "The Kindness of Strangers: The Usefulness of Electronic Weak Ties for Technical Advice," *Organization Science* (7:2), 1996, pp. 119-135.

Faraj, S., and Wasko, M. M. "Knowledge Contribution in Electronic Networks of Practice," Working Paper, 2003.

Finholt, T., and Sproull, L. "Electronic Groups at Work," *Organization Science* (1:1), 1990, pp. 41-64.

Fulk, J., and DeSanctis, G. "Electronic Communication and Changing Organizational Forms," *Organization Science* (6:4), 1995, pp. 337-349.

Fulk, J., Flanigan, A. J., Kalman, M. E., Monge, P. R., and Ryan, T. "Connective and Communal Public Goods in Interactive Communication Systems," *Communication Theory* (6:1), 1996, pp. 60-87.

Jarvenpaa, S. L., and Leidner, D. E. "Communication and Trust in Global Virtual Teams," *Organization Science* (10:6), 1999, pp. 791-815.

Jones, Q., Ravid, G., and Rafaeli, S. "Information Overload and the Message Dynamics of Online Interaction Spaces: A Theoretical Model and Empirical Exploration," *Information Systems Research* (15:2), 2004, pp. 194-210.

Katz, L., and Powell, J. H. "Measurement of the Tendency Toward Reciprocation of Choice," *Sociometry* (18), 1955, pp. 659-665.

Kilduff, M., and Tsai, W. *Social Networks and Organizations*, Sage Publications, Ltd., London, 2003.

Kollock, P. "The Economies of Online Cooperation: Gifts, and Public Goods in Cyberspace," in *Communities in Cyberspace*, M. A. Smith and P. Kollock (Eds.), Routledge, London, 1999. pp. 220-239.

Lakhani, K., and von Hippel, E. "How Open Source Software Works: 'Free' User-to-User Assistance," *Research Policy* (32:6), 2003, pp. 923-943.

Law, A. M., and Kelton, W. D. *Simulation Modeling and Analysis*, McGraw-Hill, New York, 2000.

McKnight, D. H., Choudhury, V., and Kacmar, C. "Developing and Validating Trust Measures for E-Commerce: An Integrative Typology," *Information Systems Research* (13:3), 2002, pp. 334-359.

Monge, P. R., and Contractor, N. S. *Theories of Communication Networks*, Oxford University Press, New York, 2003.

Moreno, J. L., and Jennings, H. H. "Statistics of Social Configurations," *Sociometry* (1), 1938, pp. 342-334.

Newman, M. E. J. Personal Correspondence, 2004.

Newman, M. E. J. "The Structure and Function of Complex Networks," *SIAM Review* (45:2), 2003, pp. 167-256.

Pierce, C. A., Block, R. A., and Aguinis, H. "Cautionary Note on Reporting Eta-Squared Values from Multifactor ANOVA Designs," *Educational and Psychological Measurement* (64:6), 2004, pp. 916-924.

Ravid, G., and Rafaeli, S. "Asynchronous Discussion Groups as Small World and Scale Free Networks," *First Monday* (9:9), 2004.

Rheingold, H. *The Virtual Community*, Addison-Wesley, Reading, MA, 1993.

Stewart, K. J. "Trust Transfer on the World Wide Web," *Organization Science* (14:1), 2003, pp. 5-17.

Wasko, M. M., and Faraj, S. "'It Is What One Does': Why People Participate and Help Others in Electronic Communities of Practice," *Journal of Strategic Information Systems* (9), 2000, pp. 155-173.

Wasko, M. M., and Faraj, S. "Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice," *MIS Quarterly* (29:1), 2005, pp. 35-47.

Wasko, M., Faraj, S., and Teigland, R. E. "Collective Action and Knowledge Contribution in Electronic Networks of Practice," *Journal of AIS* (5:11-12, 2004, pp. 493-513.

Wasserman, S., and Faust, K. *Social Network Analysis: Methods and Applications*, Cambridge University Press, Cambridge, UK, 1994.

Watts, D. J. *Six Degrees: The Science of the Connected Age*, W. W. Norton and Company, New York, 2003.

Watts, D. J. "The 'New' Science of Networks," *Annual Review of Sociology* (30), 2004, pp. 243-270.

Watts, D. J. 1999. Small worlds: the dynamics of networks between order and randomness. Princeton, NJ: Princeton University Press.

Watts, D. J., and Strogatz, S. H. "Collective Dynamics of Small-World Networks," *Nature* (393), 1998, pp. 440-442.

Wu, F., Huberman, B. A., Adamic, L. A., and Tyler, J. R. "Information Flow in Social Groups," *Physica A* (337), 2004, pp. 327-335.