12-31-2004

Membership Dynamics and Network Stability in the Open-Source Community: The Ising Perspective

Wonseok Oh
McGill University

Sangyong Jeon
McGill University

MEMBERSHIP DYNAMICS AND NETWORK STABILITY IN THE OPEN-SOURCE COMMUNITY: THE ISING PERSPECTIVE

Wonseok Oh
Faculty of Management
McGill University
Montreal, QC Canada
wonseok.oh@mcgill.ca

Sangyong Jeon
Physics Department
McGill University
Montreal, QC Canada
jeon@hep.physics.mcgill.ca

Abstract

In this paper, we address the following two questions: (1)How does a participant's membership decision affect the others (neighbors) with whom he has collaborated over an extended period of time in an open source software (OSS) network? (2) To what extent do network characteristics (i.e, size and connectivity) mediate the impact of external factors on the OSS participants' dynamic membership decisions and hence the stability of the network? From the Ising perspective, we present fresh theoretical insight into the dynamic and reciprocal membership relations between OSS participants. We also performed simulations based on empirical data that were collected from two actual OSS communities. Some of the key findings include that (1) membership herding is highly present when the external force is weak, but decreases significantly when the force increases, (2) the propensity for membership herding is most likely to be seen in a large network with a random connectivity, and (3) for large networks, at low external force a random connectivity will perform better than a scale-free counterpart in terms of the network strength. However, as the temperature (external force) increases, the reverse phenomenon is observed. In addition, the scale-free connectivity appears to be less volatile than with the random connectivity in response to the increase in the temperature. We conclude with several implications that may be of significance to OSS stakeholders.

Keywords: Open source, membership herding, Ising theory

Introduction

In recent years, we have witnessed an immense proliferation of open source software (OSS)[1] development and the radical shift of paradigm over which traditional software engineering practice had duly prevailed. At the fore of this burgeoning movement are the participants—project leaders, developers, users, and the like—who provide voluntary "community service" (i.e., software development and improvement, bug fixes, peer reviews, etc.) in return for nothing but intrinsic rewards (Hippel and Krogh 2003). These participants, who are globally scattered, form virtual communities in which dynamic interactions take place in the pursuit of common goals (Jones et al. 2004; Von Krogh et al. 2003).

Interestingly, however, since these participants provide service under no formal contract, they can freely leave the community, just as they can freely join. Recently, the stability of an OSS community has been severely threatened by the presence of various external forces, such as commercial vendor intervention (O'Mahony 2003), lucrative financial incentives and career offers by established firms (Lerner and Tirole 2002), and the availability of other OSS projects, which can provoke a participant to change allegiance. These dangers always exists in any OSS community (Butler 2001; Markus et al. 2000).

---

[1]OSS is defined as "software where the source code is freely distributed with the right to modify the code, and on the condition that redistribution is not restricted" (**www.opensource.org**).

In light of this, questions arise: (1) How does a participant's membership decision affect the others (neighbors) with whom he has collaborated over an extended period of time in an OSS network? (2) To what extent do network characteristics (i.e, size and connectivity) mediate the impact of these external factors on the OSS participants' dynamic membership decisions and hence the stability of the network? This paper addresses these two questions. Whereas the first question is addressed theoretically from the Ising (1925) perspective, the answer to the second is provided by performing simulations with the empirical network data that were collected from two actual OSS communities, Linux and Hypermail.

We consider the Ising perspective a suitable mechanism for exploring membership dynamics formulated by OSS participants. The Ising theory is generally concerned with basic patterns in dynamic interactions among physical objects or economic agents. A computer simulation is carefully designed to validate our theoretical propositions. Based on the results obtained from the simulation, we explore the implications for OSS communities with respect to maintaining network stability. Our main contribution in this study is the provision of a theoretical vantage point from which one can observe and analyze the dynamic interactions between OSS participants and understand the impact of network characteristics on the longevity of OSS communities.

Theoretical Background

A resource-based perspective of social structure posits that social structure must continuously provide members with benefits in order to ensure its sustainability. Butler (2001, p. 347) argues that "to be sustainable, social structures must maintain access to a pool of resources and support the social process that converts those resources into valued benefits for the participants." The benefits necessary to sustaining the social structure include opportunities for affiliation or championship (McClelland 1985) and opportunities to influence people (Winter 1973), both of which denote the importance of social interactions with their affiliated peers. The changes in their affiliations or ties (i.e., discontinued membership by an active member) are therefore likely to influence the level of benefits perceived by the members and hence their membership status.

Before illustrating the theoretical underpinnings of our study, it is worth discussing our motives in applying the basic principles of physics to the domain of human interaction. Researchers have argued that there is an operational similarity between physical and social systems (Gunn et al. 1968). In fact, some of the social phenomenon (e.g., organizational survivability, stock market dynamics, employee motivation) in fact have been explained by theories in natural science, such as biological evolution and electromagnetism. In social network theory, for example, the main idea underlying Granovetter's (1973) weak tie theory was inspired by the classic chemistry principle that shows how weakly connected hydrogen bonds can hold huge water molecules together. Many areas in social science, sociology in particular, deal with dynamic social interactions and use qualitative approaches to observe and explicate the complexity inherent to the behavior of social entities. Our goal here is to identify and understand, from a quantitative perspective, the basic pattern of interactions between network constituents via the Ising theory, which is concerned with magnetic interactions.

The primary essence of the Ising model is rather simple and straightforward, yet its conceptual underpinning has been applied to many different contexts, one of which is the exploration of herding behavior between market participants (Banerjee 1992). A crucial premise in the application of this theory is that human agents (investors, consumers, and sellers alike) possess an intrinsic desire for consensus (Banerjee 1992; Bikhchandani et al. 1998). The micro-structure field extensively employs the principles of the Ising and herding models in order to illustrate the way in which stock prices fluctuate given various market idiosyncrasies. Herd behavior theory denotes that "everyone will be doing what everyone else is doing" (Banerjee 1992, p. 798), positing that price movements are determined by the liquid interactions between market participants rather than the arrival of new information (Graham 1999; Scharfstein and Stein 1990). Herding literature suggests that investors tend to engage in nonrational herd behavior driven by an "animal spirit" that has them acting as "imitative lemmings" (Shleifer and Summers 1990).

In his seminal article, "A Simple Model of Herd Behavior," Banerjee (1992) examines a sequential decision-making process in which a decision maker imitates a decision made by previous decision makers. He warns that herd behavior can create inefficiencies for society, inflicting negative externalities on other market agents. Similarly, Bikhchandani et al. (1998) assert that behavior convergence can lead to a socially inefficient equilibrium in which a small shock can result in large shifts in mass behavior. Very recently, several information systems researchers have applied the notion of herding behavior to address various IT-related issues, particularly the phenomenon of IT adoption herding (Au and Kauffman 2003; Li 2004). These authors assert that IT managers tend to deliberately imitate other firms' investment behaviors when making their own IT investment and adoption decisions. These managers tend to follow the herd, even when the benefits of joining the herd are suboptimal compared to those made possible by standing alone. In summary, although herding behavior has been viewed as economically and socially undesirable, it appears that the phenomenon exists and persists (Banerjee 1992; Bikhchandani et al. 1998).

Interaction Dynamics in the OSS Network

A Conceptual Model

In order to lay a foundation for our propositions, we briefly exemplify in an intuitive fashion the building blocks that constitute the Ising model. Suppose we have a certain number of participants for an OSS project, and each participant is assigned a numeric state either +1 or –1. If he decides to participate actively, his state is numeritized as +1. On the other hand, if he decides to be idle or eventually leave the project, his state is assigned as –1. Each participant has n neighbors with whom he cooperatively contributes in the community (for simplicity, in our base model we begin with a regular network in which $n = 4$). Each participant is assumed to regularly poll his neighbors to keep track of how many of them actively participate in the project. He then calculates the energy (U) level, which reflects the magnitude of the dynamic interaction between OSS network members.

Figure 1 depicts graphically a spatial setting in which a participant (m_0) and his neighbors (m_1, m_2, m_3, m_4) are located in a two-dimensional network space. According to the Ising theory, $U > 0$ (energy level is greater than zero) indicates that the majority of his neighbors maintain the state opposite to his. In contrast, $U < 0$implies that both he and the majority of his neighbors are on the same side ($U = 0$ implies that his neighbors are evenly divided). He then tosses a coin to decide whether to stay or leave.

$$U = -m_0\,(m_1 + m_2 + m_3 + m_4) \text{ or } U = -(\textit{chosen square})\text{x}(\textit{sum of its neighbors})$$

U = *energy*, m_0: *state of the particpant*, m_1: *state of his neighbor* 1m m_2: *state of his neighbor* 2, m_3: *state of his neighbor* 3, m_4: *state of his neighbor* 4

The specific probabilities are calculated as follows:

$$P(\textit{flip}) = 1/(1 + \exp(\Delta U/T) \tag{1}$$

$$P(\textit{stay}) = 1 - P(\textit{flip})$$

where the energy difference, ΔU, is the difference between the current energy and the projected energy of changing the side.[2] T represents temperature (external factors).

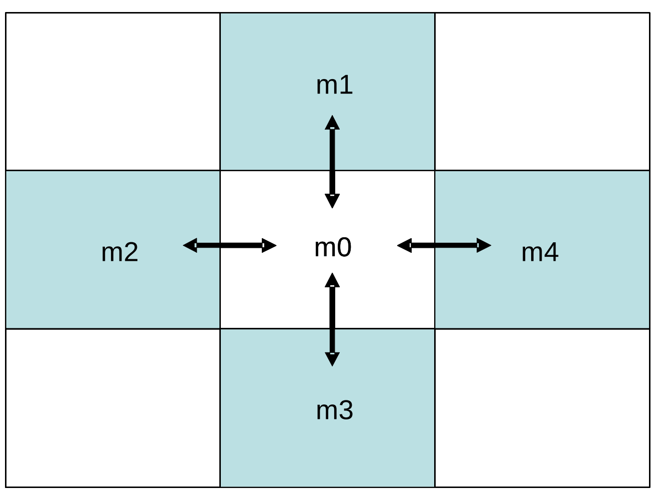


Figure 1. A Spatial Setting for a Network Participant and His Neighbors
(The Ising model assumes that each square can be either occupied or empty, but that only nearest neighbors interact.)

[2]From the expression, when the sign of S_0 is flipped, its energy is also flipped. Hence, the difference in energy is given by $\Delta U = U(\textit{final}) - U(\textit{initial}) = (-U) - (U) = -2U$.

In the probability formula specified above, ΔU and T measure the cost of changing states and temperature (external forces), respectively. If the cost associated with switching position is high, the participant is unlikely to make a move. In contrast, if the switching cost is low, he is more likely to change his state.

Illustration of Herding on a Regular Lattice

Drawing from the herding theory, we argue that an OSS participant's stay or leave decision is greatly influenced by the behavior of his neighbors. We believe that this is a legitimate assumption. Although individual participants in OSS communities have no formal obligations to contribute, a community spirit instilled as a result of long-term cooperation between members may in turn instill a sense of responsibility, urging them to make contributions to the community on a regular basis. However, when the majority of his neighbors, particularly those who are regulars, leave the community, he no longer feels such informal obligation or enjoyment. In the absence of those colleagues, he himself may not be as productive and valuable. Eventually he will follow those who have left. Therefore, we believe that these external forces influence the manner in which dynamic reciprocity occurs between network participants, resulting in the reconfiguration of the network.

The dynamics of this model may be illustrated further through the following example. Suppose *A* is an OSS participant in the network where each member has a fixed number of neighbors. Of paramount importance to *A*'s "life" is being in agreement with his neighbors. If there is no external force, sooner or later all individuals in the OSS network agree with each other and maintain the same membership state, either all +1 (actively participatory) or all –1 (idle or leave the community). However, changes in the temperature (representing exogenous uncertainties) may influence such a herding phenomenon. Simply put, high temperature (a strong force) implies a high chance of breaking up the network, while low temperature (a weak force) indicates a low chance of such dissolution. At low temperatures, external uncertainties play a limited role and the herding instinct would still govern *A*'s attitude, a result of which is that the propensity to herd still prevails. As the temperature rises (e.g., lucrative financial incentives by an established firm), however, the energy that *A* devoted to satisfy his herding instinct diminishes. As this occurs uniformly throughout this OSS community, holes start to appear in the active network as the number of individuals who venture out of the herd grows.

If the size and the number of the holes are small enough, the herd is still maintained. At a certain temperature, however, the outside world becomes much more engrossing for *A* than being part of the herd. Therefore, an attitudinal change is likely to occur and *A* will no longer regard the herd as being of paramount importance to his life. If the critical number of individuals behaves the same way as *A* does, then the *holes in the social fabric* become too large and the network simply tears apart. At temperatures higher than this critical level, the life in this world is no longer controlled by the herding instinct. What is more threatening, however, is that the breakup of the network (herd) can occur abruptly and in a short period of time.

An Extension of the Ising Model: External Forces and Membership

In the original Ising model (Figure 1), all individuals in the network are equally important, since everyone has the same number of contacts. Therefore, whatever happens locally also occurs globally. Real world networks, however, may not be as simple as a regular lattice; to reflect the real-world situation, we need to reconstruct the setting assumed in the base Ising model from regular lattices, converting it to a more complex network where the number of neighbors per node is distributed according to an arbitrary probability distribution $p(k)$. Intuitively, the differing number of connections should influence the individuals in the following way: suppose *A* has only a single contact. At low temperatures, the membership state of *A* is completely determined by his sole neighbor. As temperature grows, however, *A* becomes more easily influenced by the external force because he has only one neighbor to be concerned about. Accordingly, *A* can detach from the herd very early on, and no one will particularly care.

Now suppose *A* has a large number of neighbors and becomes a leader in the community. If the majority of his neighbors have a small number of neighbors, then *A*'s state will have a strong influence on his neighbors' membership decisions. This, in turn, is likely to reinforce *A*'s position, since more of his neighbors now share the same state as *A*. This will trigger more neighbors to align with *A*, which further strengthens *A*'s position and so forth. Hence, once his state is assigned, there isn't much chance that the state of *A* will change. This indeed illustrates the case for networks that have a power law distribution (Barabasi and Albert 1999).If the cluster around *A* is a large enough chunk of the whole network, then very quickly the state of the whole network will follow *A*'s position. On the other hand, if most of *A*'s neighbors have a large number of neighbors, then the influence of *A* on his neighbors is relatively weak, since the neighbors do not depend heavily on *A* for social contacts. Therefore, the system, as a whole, will be more or less indifferent to the state of the leaders. Furthermore, since there is very little reinforcement of his position, the leader state can be easily reversed.

Showing exactly how and when the phase transition occurs in a general Ising model is extremely difficult. Thus far, exact solutions for finite systems have been found only for simple one-dimensional and two-dimensional regular lattice cases, and there is an indication that this problem is as complex as the famous Traveling Salesman problem (Goldman et al. 1999). One way to approach such a complex problem is to use statistical methods through the use of Monte Carlo simulation techniques in order to accumulate a large number of numerical data sets (as we have done in this study). An alternative is to analyze the general Ising model by using the so-called mean field approximation (Leone et al. 2002). Although this method does not yield exact results, it is easy to understand, yet provides deeper insights into many of the essential features of the model.

The standard mean field approximation can be explained as follows. We begin by considering a single *spin* (a single OSS participant), without any neighbors, placed in an external magnetic field of strength B which can be regarded as an external bias. In this case, the energy of the spin is solely determined by the external field: if the spin is in parallel with the magnetic field (if the participant agrees with the bias), it gets a negative energy $-B$, whereas if the spin is anti-parallel with the magnetic field (if the participant disagrees with the bias) it gets a positive energy, $+B$. At a finite temperature, the probability to align with or anti-align with the magnetic field is exactly the same as in equation (1). The average magnetization of the spin (i.e., the average participation by an OSS participant) is then given by

$$<m> = tanh(B/T) = \frac{\sinh(B/T)}{\cosh(B/T)} = \frac{e^{(B/T)} - e^{-(B/T)}}{e^{(B/T)} + e^{-(B/T)}} \tag{2}$$

where B is the strength of the magnetic field and T is the temperature.

The direction of the average magnetization is the same as the external magnetic field. This simple problem is then completely solved.

We attempt to use the above solution of a simple system for more complicated Ising models. Now consider an individual spin in an Ising model. The mean field approach starts by assuming that each spin is in an effective external magnetic field B_{eff} that is caused by all of its neighbors (Figure 2). The average magnetization of each spin is then given by

$$<m> = tanh(B_{eff}/T) \tag{3}$$

The effective external magnetic field itself is given by the number of neighbors times the average magnetization of the neighbors. In the original Ising model, there is no distinction between individual members of the network. Therefore, the average magnetization of any neighbor is also given by $<m>$. Equation (3) then turns into a self-consistency condition for $<m>$,

$$<m> = tanh(n <m>/T) \tag{4}$$

where n is the number of the nearest neighbors.

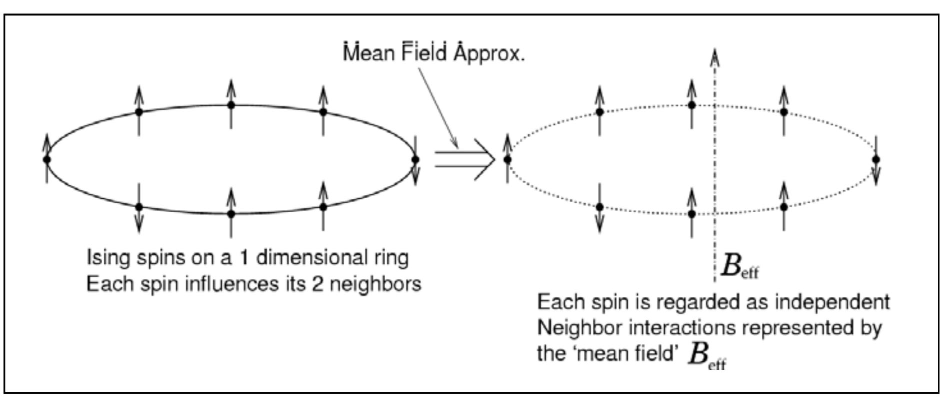


Figure 2. Illustration of Mean Field Approximation

This equation has no analytic solution. However, it is easy to calculate extreme cases. Suppose the temperature is very low; the magnitude of $x = n\langle m\rangle/T$ is, then, very large. The value of *tanh(x)* approaches +1 when x is a large positive number, while it approaches –1 when x is a large negative number. Therefore, at low temperatures $\langle m\rangle$ can be either +1 or –1. In other words, all spins have the same m value and, as a result, membership herding is likely to occur in this scenario. Conversely, if the temperature is large compared to $|n\langle m\rangle|$, then $x = n\langle m\rangle/T$ is small. In this case, we can use the approximation of $\tanh(x) \approx x$ and equation (4) becomes

$$\langle m\rangle = n \langle m\rangle/T \tag{5}$$

For arbitrary values of T, the only consistent solution is $\langle m\rangle = 0$. In this case, neighbors no longer influence each other and herding behavior is not present. The division between these two behaviors occurs at the temperature when $\langle m\rangle$ just becomes 0. This point can be obtained by differentiating equation (5) with respect to $\langle m\rangle$ and then setting $\langle m\rangle = 0$. Doing so yields the value of the critical temperature as

$$T_c = n,$$

where the subscript c stands for critical point.

This would indicate, for instance, that the Ising model on a regular 2-D lattice has a critical temperature of 4, when in reality the exact critical temperature is known in this case to be $T_c = 2.27$. The results obtained from the mean field approximation in this particular case may not be very accurate. Nevertheless, it does capture the heart of the matter in a very simple way. The break-up of a herd can occur suddenly with a small change of temperature across T_c.

If the number of neighbors is not fixed, then the above mean field approximation has to be modified (see Appendix A for the arbitrary number of neighbors). However, the essence of the arguments presented above remains the same with only the substitution of $n \rightarrow \langle n^2\rangle/\langle n\rangle$ where $\langle n\rangle$ is the average number of connections per node and $\langle n^2\rangle$ is the average of the square of the number of connections per node (Leone et al. 2002).

Network Size and Connectivity on Membership Dynamics

To reflect even more precisely on the characteristics of OSS networks, we study how the two dimensions of network properties (i.e., network size and connectivity) mediate the influence of external forces on the stability of a network.

Network Size: Butler (2001) has demonstrated that in contrast to the critical mass model (Markus 1990), membership size has both positive and negative impacts on the sustainability of online communities. An increase in network size attracts more new members to join the community, while it simultaneously influences more members to leave the community (Butler 2001). We explore this dual impact of network size from the vantage of membership dynamics.

Network Connectivity: In studying the structure of the World Wide Web, Barabasi and Albert (1999) discovered a new type of network connectivity, which they named *scale-free*, in which the connection of nodes is unevenly distributed according to a power-law distribution. In this network structure, a few nodes (called hubs) are dominantly well-connected to other nodes, and to far greater extent than regular nodes. Prior to the discovery of this *scale-free* connectivity, random connectivity (Erdos and Renyi, 1960), in which large networks form randomly with the distribution of neighbors given by a Poisson distribution, has prevailed as a legitimate model in explaining network connectivity. Structurally, due to the presence of the small number of hubs, scale-free networks show a greater level of hierarchical structure (not in the sense of the hierarchies seen in traditional organizations) than do randomly distributed networks. Barabasi and Albert argue that scale-free networks naturally emerge because new network nodes tend to attach preferentially to existing nodes that are already well-connected.

In summary, we test in this study the following scenario via the Ising model: We first begin with an all-active initial state and then evolve the initial state according to the model procedure to later points in the time frame. This test is carried out in the context of varying sizes of networks (in terms of total number of participants as well the magnitude of external forces) and in different network connectivity environments (i.e., random versus scale-free networks). Great attention is paid to the analysis of how these network properties collectively mediate the interactions of network participants in response to external forces.

Simulation Design

Empirical Data

To reflect real-world OSS networks, we collected actual network data from the two OSS communities (Linux and Hypermail) and used the data to design simulation models. These two communities were chosen because they use the UNIX mailbox format necessary for extracting the relevant data. The Linux and Hypermail OSS networks represent a large and a small open-system community, respectively. In order to identify the network characteristics, we initially downloaded nearly 100,000 archived (between 1997 and 2003) LINUX Kernel and Hypermail newsgroup messages posted in a UNIX mailbox format. We then proceeded to extract the relevant data written in RFC-1036 standards (see **http://www.faqs.org/rfcs/rfc1036.html** for actual message formats). A program was written using Java Mail APIs, which parse through the headers of individual messages and later save the extracted information in a MySQL database server for further network analysis. Specific information was obtained regarding the characteristics of these two OSS communities, including the size, the number of average connections per participant, and the hierarchy of each community. We identified the network connectivity of the Linux community as being of a scale-free type, but it was not possible to make an identification for Hypermail due to its small size. Utilizing UCINet software (Borgatti et al. 2002), we obtained the size of ego network for each participant active in the Linux community and then plotted its distribution.

Figure 3 shows the histogram of the number of connections that each node has in the Linux network. The shape of this distribution determines whether the given network is *scale-free* (straight line in a log-plot, or the distribution has the form $p(k) = Ck^{-\gamma}$ where k is the number of connections and C is the normalization constant). The x-size of the histogram is 10, so the first slot is the number of individuals with 0 to 9 connections. The second slot is the number of individuals with 10 to 19 connections, and so on. A log-transformation was performed. So the plot should look like $Log(p(k)) = -\gamma\, Log(k) + Log(C)$. There are two points worth noting in this connectivity distribution for Linux: (1) the linearity shows that the Linux developer network is indeed a *scale-free* network and (2) the exponent is smaller than 2. This means that the distribution has no single giant node that takes over almost all connections.

Simulation Algorithm Procedure and Measurement

Based on the empirical data, we have created simulation-generated, open-source communities that vary in size (both small and large) and connectivity (both random and scale-free). SimHyper and SimLinux were created with 54 and 3,475 participants, respectively. These network sizes were determined from the seven year annual average of actual number of participants in each community. Each community is then further broken down according to connectivity.

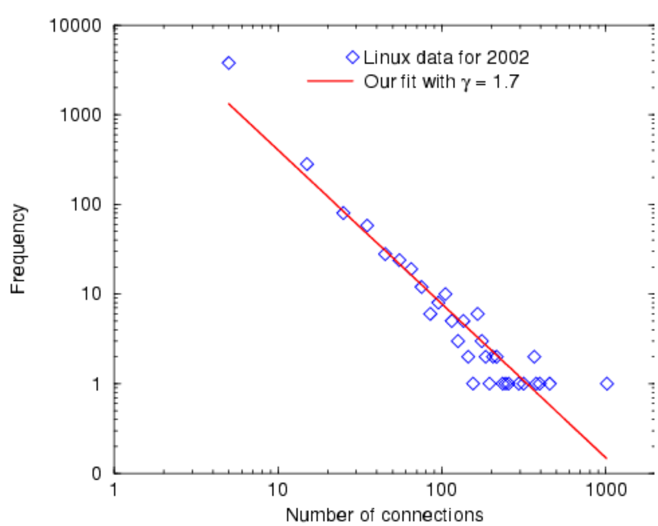


Figure 3. Scale-Free Network Connectivity of Linux

The primary focus of this study is (1) to understand the extent of herding behavior by OSS developers in response to the external forces and (2) to predict the stability (success) of the network based on member participation intensity. The criterion for success of a project is the "global robustness" of its network integration. A global network will be formed if all members are actively involved in the community. The first aforementioned issue is addressed by examining the degree of dispersion (D, measured by standard deviation) across the members with respect to their membership status, while the second is answered by looking at the average state value (m, with the average value of + 1 indicating the strongest network integrity). For a given network configuration with N nodes, the average (m) is given by

$$<m> = (1/N)(m_1 + m_2 + \dots m_N),$$

where m_1 is the state of the i^{th} node.

The dispersion is given by

$$D = (1/N)[(<m> - m_1)^2 + (<m> - m_2)^2 + \dots + (<m> - m_N)^2]$$

We measure these two numbers at 100 different times during the late period of the evolution of the system and take the average.

The simulation program[3] we develop starts by constructing a network structure according to the specified connectivity. After a structure is constructed, we assign uniform values of +1 to all nodes. In addition, we permanently fix the state of the most highly connected node as +1 throughout the simulation. Time-evolution sequences designed to mimic the evolution of a system (neighbors) are then repeatedly applied. Specifically, at each time step, a participant is randomly chosen. We then calculate the energy associated with that participant according to the expression,

$$U = - m_0 (m_1 + m_2 + m_3 + \dots m_{k0}),$$

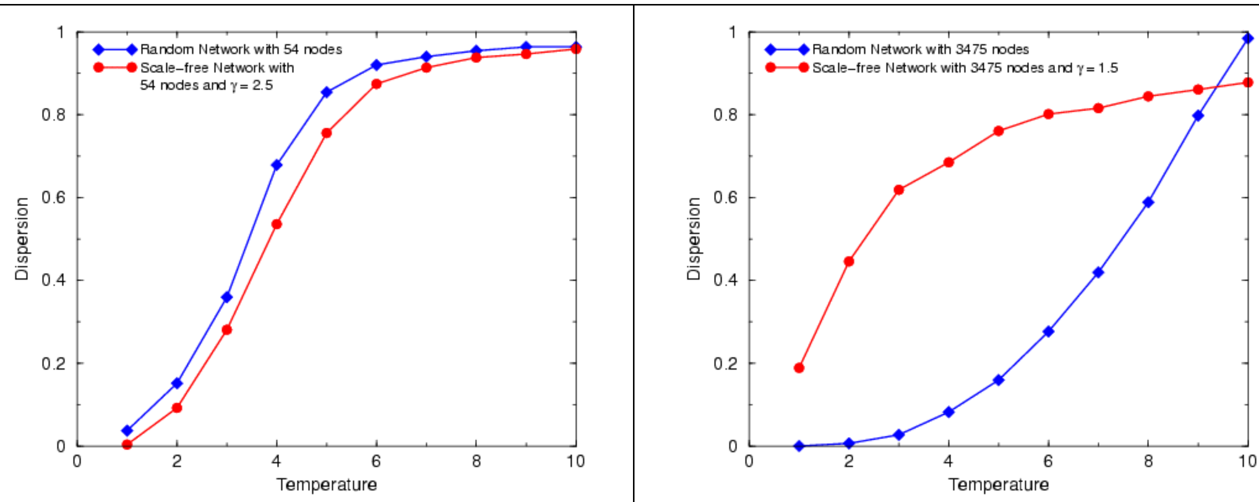
where $k0$ is the number of neighbors for m_0.

As described earlier, the decision to stay or change sides is then probabilistically made according to its energy level. Note that the higher the energy costs, the less likely one is to change states. In addition, the weaker the external forces, the less likely one is to change states. After a "sufficient" amount of time elapses (typically when all the members are checked 250 times), the program records the average state of the desired observables at periodic time intervals. After 100 data points are obtained in this way, they are averaged and recorded. We do this for each temperature starting afresh from the same initial state.

Results of the Simulation

The design of our analysis deals with one independent variable (temperature), two mediating variables (network size and connectivity) and two dependent variables (herding and network success). Figure 4 shows the *dispersion* of the average state value per participant as a function of the external forces (temperature) for the two OSS communities (SimHyper and SimLinux). Note that the x-axis indicates the strength of the external forces or temperature, while the y-axis denotes the *dispersion* of the average state value that represents the degree of herding: the smaller the dispersion, the greater the propensity of the participants to show the herding behavior in their membership decision. The two graphs, Figures 4(a) and 4(b), show that at low temperatures in which external forces are weakly present, the members, regardless of network size, tend to show a high degree of herding in their membership status as indicated by the low degree of dispersion. However, as the temperature (external force) increases, the magnitude of dispersion increases, which indicates a reduced propensity to herd. The small network (SimHyper) appears to be more volatile as the temperature increases, showing a much steeper increase of dispersion compared to the large network (SimLinux). This is because small networks tend to have a small number of average connections as well. The volatility of a network at a given temperature depends very much on its connectivity as our previous mean field analysis shows. The higher the connectivity, the less volatile the network is, as the probability to change sides depends on the ratio of the energy (which depends on the number of neighbors) and the temperature.

[3]The codes used for the simulation program are available upon request from the authors.



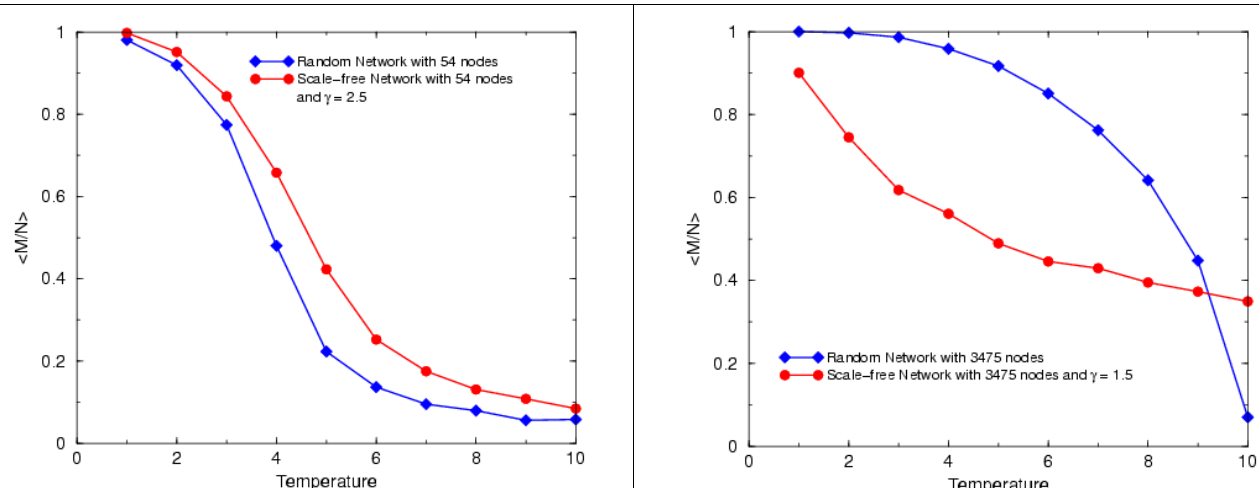
(a) SimHyper (N = 54)

(b) SimLinux (N = 3,475)

Figure 4. The Impact of External Forces on Membership Herding

In order to understand the network size effect, we compare the magnitude of the dispersion between the two graphs. For a given temperature (i.e., temperature = 5), SimHyper generally shows a much larger (smaller) dispersion (herding) (0.7 to 0.9) than the SimLinux (0.15 to 0.7). This result suggests that when the network connectivity is held constant, the herding phenomenon is more likely to be seen in a large OSS community. Regarding the network connectivity, in the small network (SimHyper) the difference in dispersion between the two network formations (random and scale-free) was deemed trivial, but that was not the case in the large network. Regardless of the network size and temperature, scale-free networks show a greater level of dispersion than random networks for a given temperature. For example, at temperature = 5, the magnitude of dispersion was 0.15 and 0.7 for random and scale-free networks, respectively. This result indicates that membership herding is less likely to occur in the network with scale-free networks in which nodes are unevenly distributed than in a randomly connected network.

Figures 5(a) and 5(b) show the average state values that literally determine the degree of network strength and eventually the success or failure of a community. As explained earlier, if the average state value of a participant moves toward +1, the project



(a) SimHyper (N = 54)

(b) SimLinux (N = 3,475)

Figure 5. The Impact of External Forces (Temperature) on the Average State Value

is pronounced a success, since at this state a global network emerges in which all participants are actively involved. Conversely, if the average state value approaches either –1 or 0, the project is considered a failure because people do not actively take part in the project or the state of the system is highly chaotic.

At the beginning of our simulation, all members of the community are active. After a sufficient period of time, the average state of a member changes according to temperature. In a small network, the average participation declines sharply as the temperature rises (Figure 5A). A radical phase transition occurs when the external force passes a certain critical value. Despite a relatively small temperature interval, the system completely changes from a well-organized state to a disorganized state, which may signal the failure of the project. Although the scale-free network is slightly superior to the random network in terms of network strength, the mediating role of network connectivity does not appear to be significant in the small network as indicated by the similar average state value and the pattern in which the average participation decreases in response to the temperature.

For the large network (Figure 5B), at low temperatures, random networks perform better than their scale-free counterparts in promoting average participation. This is expected because, at low temperatures, scale-free networks have a large chunk of nodes that are connected with only a few neighbors (a very low number of connections with others). These nodes can leave the herd relatively easily, causing the initial drop. In contrast, each node of a large random network has approximately the same number of connections, so nodes with a very small number of neighbors are rare. Therefore, at low temperatures, the number of participants who can easily leave the herd is relatively small. As the temperature rises, however, the average participation in random networks declines significantly. When the temperature passes a certain point (around 9), the position is indeed reversed and scale-free networks induce a higher level of average participation than do random networks. A more intriguing phenomenon is that large, scale-free networks are less volatile than large, random networks in response to an increase in temperature. This may be due to the fact that the leader in the network plays a significant buffer role in minimizing the impact of the external force, motivating the members to maintain the same state as his (note that the leader's state is always +1).

Discussion and Implications

Several important findings emerge from the simulation analysis and they are outlined in Table 1. Important implications can be drawn from these results, which pertain to membership herding propensity and network stability. Regarding the herding phenomenon, the most threatening is the possibility of herd exit. A snowball effect in which one member's retirement can lead his neighbors to leave can, in a cascading fashion, lead a neighbor's neighbors to leave, etc. This phenomenon can trigger a drastic phase transition, which results in network separation and disintegration. This may in fact be the very weak spot of the OSS community. As indicated by our results, even a well-established community with a large number of participants (such as Linux) may be vulnerable to the possibility of herd exit.

Table 1. Summary of Key Findings

	Key Findings
Membership Herding	• Membership herding is highly present when the external force (temperature) is weak (low). However, as the temperature increases, the herding propensity decreases significantly. • The herding phenomenon is more likely to be seen in a large network with random connectivity. • For small networks, membership herding is less likely to occur in scale-free structures than random structures. However, the reverse phenomenon was observed in large networks.
Average participation	• As the temperature increases, the average participation decreases significantly in the small network. However, such a sharp reduction was not observed in the large network. In fact, when the temperature passes a certain point (around 9), the position is indeed reversed and scale-free networks perform better than random networks. • For large networks, at low temperatures random networks perform better than their scale-free counterparts in promoting the participation. • For large networks, at low temperature, random networks perform better than scale-free counterparts in terms of network strength. However, as the temperature increases, the reverse phenomenon is observed. In addition, the scale-free connectivity is less volatile than the random connectivity in response to the increase in the temperature.

Recently, commercial companies have begun participating in OSS projects, taking control of some of the more successful projects (O'Mahony 2003). Although involvement of these companies may garner some benefits for the community, it may also threaten the fundamental reason why volunteers took part in the project in the first place (Lerner and Tirole 2002). Commercial vendors place their own people in the key technical and strategic decision-making positions in order to increase their control of the project (Long 2003). Due to differences in their orientation, motivation, and attitude, conflicts may arise over personal, technical, and strategic issues between the company, which is seeking to make a profit, and the volunteers, who gain value by giving value. When a commercial company takes control, the project is often redirected so that it is something other than what the volunteers originally wanted it to be. Commercial intervention may force the volunteers to leave the community because they may feel that they no longer have any stake in the project and that their ability to contribute is severely limited (Long 2003). Once some of the key volunteers have left the community, a snowball effect is expected to occur rapidly and can lead to abandonment of the project.

Active participation by the members is perhaps the most critical ingredient for the sustainability of any virtual community, including OSS (Jones et al. 2004). The results of our study reveal that the participation is significantly reduced in the presence of the high external force. Regardless of the network connectivity, small networks are found to be very fragile when met by an external force; even a small change in the force can dramatically break up the existing network, triggering the community to become very inactive and eventually disappear. Our results may provide some evidence for the argument related to the difficulties of establishing and maintaining critical mass in virtual communities (Butler 2001; Markus et al. 2000).

Reduced participation resulting from the external force was also observed in the large network. This result may explain, from the network perspective, the interesting phenomenon observed by Butler (2001), who discovered that membership loss increases as the size of a social community grows. Interestingly, network connectivity appears to matter in large networks: at low temperatures, random networks seem to structurally encourage more participation than do scale-free networks, but this state is reversed when the temperature increases significantly. Related to this, scale-free networks are less volatile than random networks in reaction to temperature increases. Leaders, who are present only in scale-free networks, seem to play buffer roles that, if in place, minimize the damage caused by external forces. Perhaps, this result may provide some explanation as to why Linux, which was found to form a "perfect" scale-free connectivity (see Figure 3), has been one of the most active and stable OSS communities. Virtual organizations do not operate in the same manner as traditional organizations (Ahuja and Carley 1999). Nevertheless, the structure under which the members interact influences their motivation to participate. A qualitative study is necessary to provide specific explanations as to why a hierarchical virtual structure is more effective than a loosely-coupled structure in motivating member participation.

Conclusion

Despite its paramount importance to the stakeholders, little progress has been made in studying why some OSS communities succeed and others fail. We address this concern in conjunction with the stability of dynamic relations between the community members who are the foundation components of the network. The main intention of this study was to provide fresh theoretical insight into the membership dynamics in the presence of external forces and the potential mediating role of network characteristics on preserving the stability of the OSS community.

The term *open* in open-source communities is generally perceived as having a positive connotation, and describes many of the benefits provided by these revolutionary practices. Many users as well as developers perceive open to mean unobstructed entrance, available, access to all, free from limitations, generous, etc. However, the term can also be interpreted as referring to some unconstructive characteristics, such as unobstructed exit, susceptible, vulnerable, fragile, lacking effective regulation, and so on. The unobstructed exit and lack of regulatory force inherent in the OSS community can result in a community's susceptibility and vulnerability to herded exits by its participants. Commercial vendor intervention, an alternative project becoming available, and licensing issues can result in some original core members ceasing to provide their loyal service for the community, which can prompt their coworkers to leave as well. Our main contribution in this study was to theoretically assess and test, through a simulation, such possibilities in light of the membership dynamics and the mediating effects of network characteristics. Further research should be carried out with particular emphasis placed on the empirical validation of our results. It is hoped that this paper will lay the groundwork for future and more comprehensive studies in this important field of research.

Acknowledgments

Sangyong Jeon is supported in part by the Natural Sciences and Engineering Research Council of Canada and by le Fonds Nature et Technologies of Quebec. Sangyong also thanks the RIKEN-BNL Center Brookhaven National Laboratory of Upton, New York, and the U.S. Department of Energy [DE-AC02-98CH10886] for providing facilities essential for the completion of this work.

References

Ahuja, M., and Carley, K. "Network Structure in Virtual Organizations," *Organization Science* (10:6), 1999, pp. 741-757.

Au. Y., and Kauffman, R. "What Do You Know? Rational Expectations and Information Technology Investment," *Journal of Management Information Systems* (20:2), 2003, pp. 49-76.

Banerjee, A. "A Simple Model of Herd Behavior," *Quarterly Journal of Economics* (7:3), 1992, pp. 797-818.

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

Bikhchandani, S., Hirshleifer, D., and Welch, I. "Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades," *Journal of Economic Perspectives* (2:3), 1998, pp. 151-170.

Borgatti, S., Everett, M., and Freeman, L. *UCINET for Windows: Software for Social Network Analysis*, Analytic Technologies, Harvard, MA, 2002.

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

Erdos, P., and Renyi, A. "On the Evolution of Random Graphs," *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* (5), 1960, pp. 17-61.

Goldman, D., Istrail, S., and Papadimitriou, C. "Algorithmic Aspects of Protein Structure Similarity," in *Proceedings of the 40th Annual Symposium on Foundations of Computer Science*, New York, 1999, pp., 512-522.

Gunn, B., Brenner, M., and Mjosund, A. "The Dynamic Synthesis Theory of Motivation," *Management Science* (14:10), 1968, pp. B601-B625

Graham, J. "Herding Among Investment Newsletters: Theory and Evidence," *Journal of Finance* (54), 1999, pp. 237-268

Granovetter, M. "The Strength of Weak Ties," *American Journal of Sociology* (78), 1973, pp. 1360-1380.

Hertel, G., Niedner, S., and Herrmann, S. "Motivation of Software Developers in Open Source Projects: An Internet-Based Survey of Contributors to the Linux Kernel," *Research Policy* (32), 2003, pp. 1159-1177.

Hippel, E., and Krogh, G. "Open Source Software and the Private-Collective Innovation Model: Issues for Organization Science," *Organization Science* (14:2), 2003, pp. 209-223.

Ising, E. "Beitrag zur theorie des ferromagnetismus," *Zeitschrift für Physik* (31), 1925, pp. 253-258.

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

Kogut, B., and Metiu, A. "Open-Source Software Development and Distributed Innovation," *Oxford Review of Economic Policy* (14:2), 2001, pp. 248-264.

Lerner, J., and Tirole, J. "Some Simple Economics of the Open Sourcem," *Journal of Industrial Economics* (2), 2002, pp. 197-234.

Li, X. "Informational Cascades in IT Adoption," *Communications of ACM* (47:4), 2004, pp. 93-97

Leone, M., Vazquez, A.,Vespignani, A., and Zecchina, R. "Ferromagnetic Ordering in Graphs with Arbitrary Degree Distribution," *European Physics Journal*, (28), 2002, pp. 191-197

Long, A. "How Firm Initiation and Control of Projects Affects Open Source Development," 2003 (available online at **http://opensource.mit.edu/papers/long.pdf**).

Markus, L. "Toward a 'Critical Mass' Theory of Interactive Media," in *Organizations and Communication Technology*, J. Fulk and C. Steinfield (Eds.), Sage Publications, Newbury Park, CA, 1990, pp. 194-218.

Markus, L, Manville, B., and Agres, C. "What Makes a Virtual Organization Work?," *Sloan Management Review* (42:1), 2000, pp. 13-26.

McClelland, D. C. "How Motives, Skills, and Values Determine What People Do," *American Psychology* (40), 1985, pp. 812-825

O'Mahony, S. "Guarding the Commons: How Community Managed Software Projects Protect Their Work, *Research Policy* (32), 2003, pp. 1179-1198

Scharfstein, D., and Stein, J. "Herd Behavior and Investment," *The American Economic Review* (80:3), 1990, pp. 465-479.

Shleifer, A., and Summers, L. "The Noise Trader Approach to Finance, *Journal of Economic Perspectives* (4), 1990, pp. 19-33.

von Krogh, G., Spaeth, S., and Lakhani, K. "Community, Joining, and Specialization in Open Source Software Innovation: A Case Study," *Research Policy* (32), 2003, pp. 1217-1241.

Winter, D. G. *The Power Motive*, Free Press, New York, 1973.

Appendix A. Average Magnetization: The Case of Arbitrary Neighbors

In this appendix, we explain the mean field approach to the Ising model on an arbitrary network structure, following Leone et al. (2002).

If $<m>_k$ denotes the average magnetization of a node with k neighbors, then the average magnetization averaged over the whole system is given by

$$<m> = \sum k <m>_k P(k) \tag{6}$$

where $P(k)$ is the probability distribution of the number of neighbors.

The mean field approximation is now constructed for each $<m>_k$ as

$$<m>_k = \tanh(k_{nn} <m> / T) \tag{7}$$

where k_{nn} is the average number of neighbors to which the neighbors of a chosen node are connected. In the absence of correlation, this is known to be independent of k and given by

$$k_{nn} = <k^2>/<k> \tag{8}$$

Inserting equation (7) into equation (6) then yields following mean field equation for the average magnetization

$$<m>_k = \tanh(k_{nn} <m>/T)$$

The low temperature and high temperature analysis of this equation is exactly the same as before with n replaced by $<k^2>/<k>$. At low temperatures, the self-consistent solution is again $<m> = +1$ or $<m> = -1$. The system is in highly organized herding state. At high temperatures, the approximation $\tanh(x) \approx x$ yields

$$<m> = (<k^2>/<k>)<m>/T$$

which gives $<m> = 0$ and herding behavior is absent.

The critical temperature where the transition occurs can be estimated by solving the slope equation at $<m> = 0$ as before. This yields $T_c = <k^2>/<k>$.

For the case of arbitrary node distributions, the exact formula for T_c is known for infinite size networks with an arbitrary $P(k)$ as long as $<k^2>$ is finite. The analysis for obtaining is quite involved. We only quote the result here:

$$1/T_c = -(½) \ln (1 - 2 <k>/<k^2>)$$

which reduces to the above mean field result when $<k>/<k^2>$ is small.