2010

A Empirically Validated Framework for Limiting Free-Riding in P2P Networks through the Use of Social Network Information

Fatmawati Zifa Zaukani
Seoul National University, fatzifa@gmail.com

Jörn Altmann
Seoul National University, jorn.altmann@acm.org

A EMPIRICALLY VALIDATED FRAMEWORK FOR LIMITING FREE-RIDING IN P2P NETWORKS THROUGH THE USE OF SOCIAL NETWORK INFORMATION

Fatmawati Zifa Zaukani, Technology Management, Economics and Policy Program, College of Engineering, Seoul National University, Seoul, South-Korea, fatzifa@gmail.com

Jörn Altmann, Technology Management Economics and Policy Program, Department of Industrial Engineering, College of Engineering, Seoul National University, Seoul, South-Korea, jorn.altmann@acm.org

Abstract

In order to overcome the problem of free-riding in current P2P system, we suggest applying social network theory. Based on our exploration of the overlapping research fields of social networks and peer-to-peer networks, we propose a new P2P framework within this paper. It specifies social network information that can be used in a P2P system to avoid performance inefficiencies caused by free-riding or by policies to overcome free-riding. To identify this specific social network information, we conduct a survey among a small group of students, who use Skype, a popular P2P system. We use descriptive analysis and multiple regression analysis to analyze the survey data. The results of the analyses provide an indication that the idea of using social network information in P2P systems is valid and that it is supported by P2P users. Based on our findings, we make recommendations for a successful implementation of social-network-information-based P2P systems that can overcome free-riding issues and, consequently, improve the performance of P2P systems.

Keywords: Peer-to-Peer Systems, User Behavior Analysis, Free-Riding, Social Network Information, Human and Social Factors in IT Adoption, IT Deployment, IT Use.

1 INTRODUCTION

Peer-to-peer (P2P) technology has attracted much attention because of its capability to share resources that are geographically distributed around the world. Because of its distributed architecture, P2P systems can scope with a large amount of users. However, this characteristic also brings the drawback that no means exist to check and enforce the behavior of P2P users, resulting in issues such as free-riding.

Existing P2P systems have emerged, particularly triggered by the need of resource sharing in a distributed environment in a simple and cost-efficient way. Current P2P systems are also designed to be anonymous, ensuring the identity of agents. Therefore, the agent can perform actions that remain hidden. P2P networks are very dynamic and, therefore, must be able to deal with “high turnover, asymmetry of interest, collusion, zero-cost identities, and traitors” (Feldman et al. 2004). In order to cope with this dynamic environment caused by peers entering and leaving the system, P2P systems have been designed in many different ways.

The convergence of both P2P systems and social networks, as recently claimed in the literature (Wang et al. 2006; Wang et al. 2006b), has promised better technology deployment of P2P systems in the future. However, the idea of implementing the combination of both networks also raises some questions and issues: What is the best strategy in applying social network theory in a P2P system? What are the social network properties that should be considered in a P2P system?

Despite the fact that many P2P systems exploit the advantages and features of P2P networks, many of these P2P systems are still facing performance-related issues. Technology enhancements alone did not provide a comprehensive solution to overcome the performance issues. The performance in P2P networks depends on two factors, namely technology and user behavior. This fact triggers the idea of analyzing user behavior by using social network theory and applying it into the design of P2P systems.

It is essential for our work to understand the connection between human social behaviors and technology in detail. Our interest is to understand the impact of using social network information in P2P networks. We use Skype, which is a representative P2P application, for our case study.

The objective of our work is to demonstrate that social network information could contribute to solving the performance issues in P2P networks caused by free-riding or by policies to overcome free-riding. In particular, this work is conducted to seek answers to the following questions:

- Can social network information improve the performance of a P2P system? The survey, which we created for understanding users’ willingness-to-share, indicated a positive answer to this question.
- Under what condition, when and how social network information can be applied? Although social behaviors are hard to predict and measure, we found indications (factors) on how social behaviors and technology will influence each other in a P2P system.
- How can the identified factors improve the design of current P2P systems? Based on the analysis results of a multiple regression analysis, we propose a new P2P framework that considers social network information in P2P systems. The analysis also showed that the framework can increase the performance of a P2P network.

The remainder of the paper is organized as follows: Section 2 introduces the research model and the literature on which our work is based. Section 3 describes the survey conducted, the results of the survey, the analysis of the result, and, finally, makes recommendations for future work on P2P systems with reduced free-riding. Section 4 concludes our work.

2 THE SNI-P2P MODEL

A peer in a P2P network refers to an actor in a social network. P2P networks are conceptualized as communities, which are created, if an entity has an interest in the same topic. The activity between two peers is considered a relationship (or tie) between two actors who share common interest, where any group of peers may contain a few peers that belong to another group at the same time (Wang et al. 2006). This is "similar to human social networks". Interactions in P2P networks resemble the "interaction in real world communities" (Wang et al. 2006b), where peers are communicating in a group that share common goods and interests. Despite of our understanding that a P2P network is a very large environment, the works of Wang et al. (2006 and 2006b) prove that social networks exist in P2P networks. Both works take the theory of 'small world' into account for their research. In addition to that, P2P networks are also considered as an "emerging social network for pooling network and information resources" (Li et al. 2004). Existing works also agree that social network theory can contribute to the performance improvement of P2P systems (Upadrashta et al. 2005; Fast et al. 2005; Lea et al. 2006).

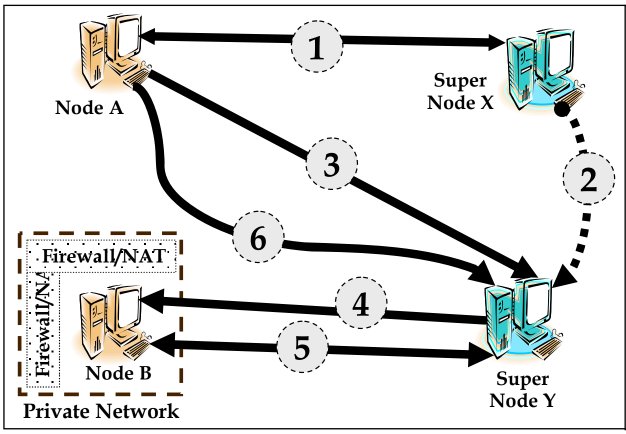
2.1 Case Study: Skype

Skype is a Voice-over-Internet Protocol (VoIP) application that is designed using the FastTrack protocol and implemented based on KaZaA. It is an overlay P2P network application that enables high quality telephone conferencing, instant messaging, file transfer, short message service and video conferencing. Skype is very popular mainly because of its ability to bypass network address translators (NAT) and firewalls (hereafter both will be acknowledged as NAT). Skype is considered proprietary software, since its source code and its documentation are not available publicly. It is assumed that Skype has a similar P2P protocol as KaZaA. A detailed analysis of the Skype architecture can be found in Baset and Schulzrinne (2006). The two-tier hierarchical architecture consists of super nodes and ordinary nodes:

- Super nodes (SN) are nodes that have high-performance resources (e.g., high bandwidth, large storage capacity, or high processing power). SNs are not behind a NAT and can be reached by any peer.
- Ordinary nodes (ON) are nodes with less resources and capabilities. Once they enter the network, the ONs connect to the SN appointed by Skype.

In many cases, NAT will hinder P2P communication. This usually makes peers behind a NAT unreachable for peers with globally valid IP addresses. A workaround for this situation is the relay nodes approach, which enables communication with peers behind a NAT (Suh et al. 2006). Given the situation, in which one of the nodes is located behind the NAT, Skype will use the super node relay approach to ensure the file transfer transaction. Using this approach, any communication of a peer that is located behind a NAT is relayed via one of the super nodes that the node is connected to. This requires super nodes to share their resources with other peers. Super nodes are automatically selected by Skype (which only considers the accessibility of the node and the bandwidth capacity to the Internet) and no Skype user can opt out of being a super node. Although this policy deals with the issue of free-riding, it might happen that a super node experiences performance degradation. In order to decrease the super nodes' burden, and therefore the risk of severe performance degradation, Skype limits the speed of a file transfer transaction to 5 Kb/s.

To illustrate this mechanism, the following scenario (Scenario I) is given. Node A is located in the public network while Node B is located in a private network (behind a NAT). Due to the condition that Node B is located behind the NAT; the file transfer activity has to be pursued through a super node. In detail, as Figure 1 illustrates, the following steps are executed in Scenario I: (1) Since Node A wants to send a file to Node B, Node A has to locate the super node of Node B through its super node, Super Node X. (2) Super Node X finds the super node of Node B, Super Node Y (Super Node X used designated searching algorithm (e.g., flooding message, random walk) to find Super Node Y). (3) Node A sends the file to Super Node Y directly. (4) Super Node Y forwards the file to Node B. (5) Node B receives file from Super Node Y and acknowledges it. (6) Super Node Y acknowledges it to Node A.



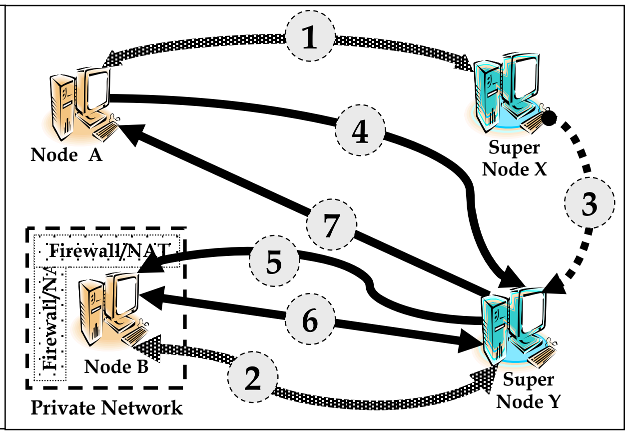


Figure 1. Scenario I: Skype Model *Figure 2. Scenario II: Extended Skype Model*

We conduct a simple test to illustrate the implications of Scenario I. For that, we compare two cases that involve a file transfer using Skype. In the first case (i), both PCs are located in the public Internet, and, in the second case (ii), one PC is located behind the NAT while the other PC is in the public Internet. The size of the file transferred is approximately 360MB. The file transfer time for both transactions is significantly different, 1,05 min in the first case and more than 40 minutes in the second case. For the second case, the slow transfer time is the result of the super node relay approach implementation and the limitation of transfer speed to 5Kb/s. This case eventually discourages file transfer activities in the Skype environment. People may only transfer files of small size, where this contributes to a limited use of the file transfer functionality of Skype. Therefore, if a user transfers a large file in the Skype network, the user would have to tolerate a very long file transfer time.

2.2 The Proposed SNI-P2P Model

In order to address this issue of very long transfer times for large files in the Skype network, our framework makes use of social network information for P2P systems. The use of social network information allows taking into consideration users' behaviors. The idea of our research is to gather social network information and apply the input to Scenario I. We believe, by using social network information, the lack of performance in current P2P systems can be reduced. Our model, the SNI-P2P model, differs from the current P2P implementations in the aspect of super node selections. We suggest that a P2P system allows a node to choose its super node from its contact list. Referring to the case of Scenario I, we propose that Node A and Node B can choose their own super nodes from their contact lists (containing their social networks), rather than depending on appointed super nodes of Skype. As depicted in Figure 2, the steps that are executed in this new scenario (Scenario II) are based on the steps in Scenario I: (1) Node A chooses a super node, which is also in its contact list, Super Node X. (2) Node B chooses a super node, which is also in its contact list, Super Node Y. (3) The remaining steps are identical to Scenario I.

Since we believe that "people will provide more resources for the benefit of their contacts", the restriction on the file transfer speed can be removed without burdening super nodes. For example, since Node B is one of Super Node Y's contacts or '*buddies*', Super Node Y is willing to give more resources and higher priority to the file transfer, i.e. Super Node Y allows a high transfer speed because it knows the user of his resources. Therefore, the file transfer can be executed faster, compared to current practice.

3 SURVEY, ANALYSIS, AND RESULTS

To seek answers to our questions as well as to validate our framework of applying social network information in P2P systems, we collect related data from Skype users. We use a survey methodology to obtain information about users' willingness-to-use our potential application. The feedback collected from the survey assists us in finding empirical evidence whether our idea will work in a real environment.

A survey data set is composed of observations, which consists of one or more pieces of variables. We received 60 observations, which is a good rate of response, considering the overall total of participants in the targeted group. Out of 60 observations, we omit 10 observations that belong to non-Skype users and perform the descriptive analysis and regression analysis on the remainder observations.

3.1 Descriptive Analysis

The analysis of the first set of survey questions shows that our subject group had acquired sufficient knowledge about Skype to answer the remaining questions in the survey. The questions address the usage frequency of Skype, the composition of the contact list, and the number of contacts. Out of five categories for the number of contacts in the contact list, we obtained data for three categories only: 'Less than 25', 'Between 26 and 50', and 'Between 51 and 75'. The analysis shows that 60% of the observations have a contact list size less than 25 people, 28% have a contact list size of 24 to 50, and only 12% have a contact list size of 51 to 75 contacts. We conclude that the high number of contacts in the contact list is an indication for the sophistication of our subjects in using Skype. This conclusion is supported by answers to the question on use frequency of Skype. It shows that all subjects use Skype frequently. The use percentage of 'PC-to-PC Call' and 'Instant Messaging' function is high.

In particular, it is to note that the usage percentage of 'File transfer' is only 56.2% on average, significantly lower than the usage for PC-to-PC calls. This indicates that Skype users do not mainly use Skype for file-sharing. It shows that the file sharing functionality of Skype is not fully utilized. Therefore, our objective to conduct this research is valid. Besides, on average, 31.30% of the contact list contains family members, 36.28% close friends, 3.74% loose contacts, 22.56% work acquaintances, 2.16% friend of a friend (FOAF), 3.68% people who sent invitations to make friends and finally 0.28% belongs to other groups. Therefore, in average, 67.58% of the contact list are people that are close to a Skype user.

For analyzing the willingness to share resources, we investigate the conditions that encourage people to do so. Figure 3 illustrates the obtained survey data.

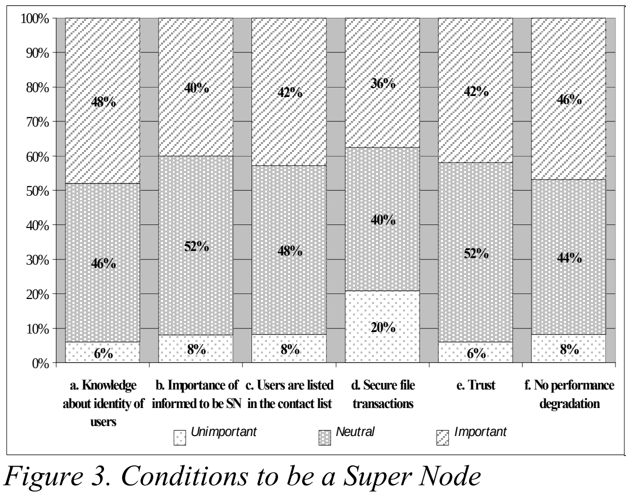


Figure 3. Conditions to be a Super Node

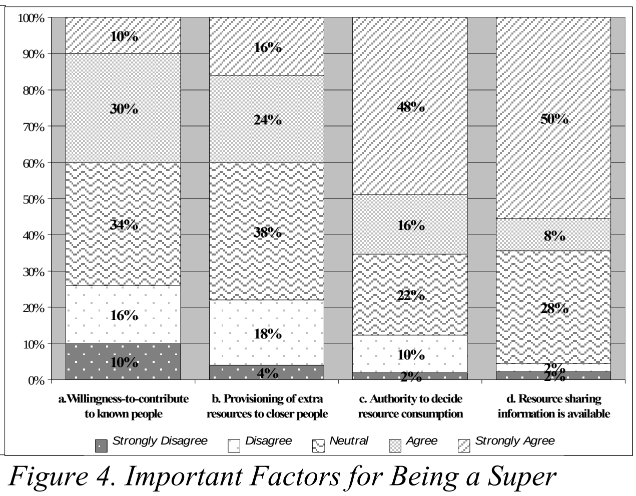


Figure 4. Important Factors for Being a Super Node

In particular, we checked the following conditions in our survey: (a) the importance of the knowledge about the identity of the user who consumes the resources, (b) the importance of knowledge about being selected as a super node, (c) the importance of letting only contacts of the contact list consume resources, (d) the importance of secured transactions, (e) the importance of trusting people that consume the resources, and (f) the importance that sharing does not result in a PC performance degradation. The results show that all conditions for sharing resources are important for our subjects.

The percentages of the choice '*Important*' are high for all questions. It shows that these conditions need to be considered.

Finally, we discuss those factors that should be considered for encouraging users to become a super node that is willing to share its resources with the contacts in their social network. We focus on four factors: (a) the willingness to share resources with people that a Skype user knows, (b) the willingness to provide extra resources to people that are closer, (c) the importance of having the authority about the resource consumption, and (d) the importance of information about the resource sharing status. Figure 4a shows the willingness of subjects to share resources with the people they know. At least half of the subjects are willing to share resources under this condition. Another interesting result is shown in Figure 4c, where 64% of the subjects want to decide about the resource allocation to other P2P users. This result and the results shown in Figure 4d, where at least 62% of the subjects would like to know about the resources that they provide, support the finding that P2P users are capable of differentiating between peers. Therefore, we can state that sharing resources with people of a social network is viable for P2P systems.

3.2 Multiple Regression Analysis (MRA)

The conditions, under which a super node is willing to share its resources, are determined now. In order to determine the appropriate variables, which describe the conditions, we seek the best-fitting multiple regression model (MRA). Fitting our research question, the dependent variable is chosen to be '*User's Willingness-to-Share Resources with People They Know*' (*sn_03*), as shown in Figure 4a.

All independent variables are chosen based on the literature reviews on social networks (Section 2). The selection is also based on our model on how social network information impacts users' behavior and the results of the descriptive analysis. Then, from all independent variables that have been identified for the survey, we have performed a series of regression experiments to include the best-fitting independent variables into our MRA model. The experiments identified six independent variables that fit best:

- **Knowledge about identity of the users, *cs_01*.** This variable measures the users' degree of importance in knowing those people, who are connected to their PCs and consume resources.
- **No performance degradation, *cs_06*.** This variable measures the level of importance of the fact that sharing will not affect the performance of PCs.
- **Provisioning of extra resources to close people in the contact list, *sn_06*.** This variable measures users' willingness to provide resource to those people who are closer to them.
- **Communication frequency, *iv_07*.** This variable measures the frequency of users' interactions with contacts of their contact list.
- **Contact list size, *buddy_list*.** This variable measures the size of a user's contact list.
- **File transfer using Skype, *ft_skype*.** This variable measures the frequency of file transfer activities.

Given the dependent variable y and the independent variables, we can estimate the parameters of the regression model as shown in the following equation:

$$sn\_03 = \beta_0 + \beta_1 cs\_01 + \beta_2 cs\_06 + \beta_3 iv\_07 + \beta_4 sn\_06 + \beta_5 buddy\_list + \beta_6 ft\_skype$$

The descriptive statistics for our MRA model is given in Table 1, providing the mean and the standard variation of the independent variables. Table 2 shows the correlation matrix. It indicates that there is little multi-collinearity between the variables. In particular, there is only strong correlation between *cs_01* and *sn_06* as well as between *buddy_list* and *ft_skype*. Although this correlation might lead to a situation, in which the coefficient estimates erratically change, a relative stability is achieved with the six variables.

Variable	Mean	Std. Deviation
cs_01	3.19	.895
cs_06	2.50	.648
sn_06	2.38	.637
iv_07	3.12	1.033
buddy_list	3.15	.834
ft_skype	1.69	1.011

Table 1. Descriptive Statistics

Variable	*sn_03*	*cs_01*	*cs_06*	*sn_06*	*iv_07*	*buddy_list*	*ft_skype*
sn_03	1	,485**	-,019	,386**	,217*	-,021	,223*
cs_01	,485**	1	,302*	,368**	,090	-,161	,058
cs_06	-,019	,302*	1	,170	-,107	-,234*	-,066
sn_06	,386**	,368**	,170	1	-,018	-,021	,208
iv_07	,217*	,090	-,107	-,018	1	,109	,019
buddy_list	-,021	-,161	-,234*	-,021	,109	1	,436**
ft_skype	,223*	,058	-,066	,208	,019	,436**	1

Table 2. Correlation Matrix

In order to interpret the goodness fit of the model, we calculate the *R family* values (i.e. the multiple correlation coefficient (R), the multiple coefficient of determination (R^2), and the adjusted R^2). The result shows that the model is fine. It has an R value of 0.837 and an R^2 value of 0.701. The adjusted R^2 is 0.607. The adjusted R^2 value shows that 60.7% of the variations in the dependent variable (*sn_03)* is explained by the independent variables. The standard error of the estimate is 0.562. We also analyze the variance. It assists us in evaluating the significance of the included variables. The value of the *F-Test* is 4.503 and statistically significant ($p<0.05$). This provides evidence of the existence of a linear relationship between the dependent variable 'willingness-to-contribute to known people' (*sn_03*) and the independent variables (*cs_01*, *cs_06*, *sn_06*, *iv_07*, *buddy_list*, and *ft_skype*). It also means that the combination of independent variables explains the dependent variable. The estimates of the coefficients β_i, which indicate the contribution of each independent variable to the model, are shown in Table 3. Therefore, the variables fit into the MRA model.

Model		Unstandard. Coeff.		Standard. Coeff.	t	Sig.
		β	Std. Err.	Beta		
1	(Constant)	.073	.776		.094	.926
	cs_01	.696	.243	.504	2.861	.010
	cs_06	-.636	.288	-.453	-2.209	.040
	iv_07	.347	.160	.323	2.162	.044
	sn_06	.607	.145	.700	4.182	.001
	buddy_list	-.451	.162	-.509	-2.780	.012
	ft_skype	.313	.134	.338	2.327	.031

Table 3. Coefficients table

3.3 Key Findings

The analysis of the survey responses enables us to answer all of our research questions positively. Our first research question about whether social network information can improve the performance of a P2P system could be answered based on the results of the descriptive analysis. The answers of the subjects of the survey indicated clearly that users are willing to share resources, if a social relationship exists between them. The amount of resources is even higher, if it is a close relationship. This information prevents any free-riding attempts, since such behavior would just not be considered between friends and family. The benefit of free-riding would be less than what is gained from a strong social relationship.

To answer the second research question (Under what condition, when and how social network information can be applied?), we analyzed the survey responses as well as performed a MRA. We identified those variables, which give an indication about the conditions under which users are willing to share their resources. The descriptive analysis and the MRA highlight the contributing factors for sharing resources in a P2P environment. The analysis results provide strong support for the selected variables.

To answer the third of our research questions (How can the factors identified improve the design of current P2P systems?), we propose a framework for developing P2P systems based on social network information. This framework lists the factors, which are essential for the development of efficient P2P systems, particularly at the level of super nodes. We state that a P2P system design that considers these factors (i.e. social network information) reduces free-riding and, therefore, improves performance.

The proposed framework will eventually change the topology formations of P2P systems (Altmann & Bedane 2009). Rather than formed randomly and anonymously, the P2P topology will be based on the topology of social networks. The new sharing environment is more conducive and based on trust. Consequently, transfer speeds that are higher than those of existing P2P systems can be achieved for peers of the same social network.

4 CONCLUSION

To find a solution to the performance issues in P2P systems, we used Skype as a study case and conducted a survey about the opinions of P2P system users. The analysis of the survey (descriptive analysis and multiple regression analysis) indicated that people are willing to share P2P resources within their social networks, i.e. with people, who are close to them.

We also proposed a new P2P framework on how to use social network information in solving the performance issue in P2P systems Our P2P framework can reduce the performance inefficiency of P2P systems caused by NATs and policies to avoid free-riding. For example, it allows removing the currently build-in speed limitation for file transfers, since users provide resources freely to their close contacts.

A P2P systems based on our framework enables users to pick their own super nodes, contrarily to the current implementation where super nodes are selected randomly for file transfers. A super node that is chosen by members of its own social network is willing to provide resources freely. It results in faster file transfers. Our idea also contributes to new P2P topology formation algorithms, which form the topology according to a node's social network rather than randomly as it is conducted currently.

References

Altmann, J. and Bedane, Z. (2009) P2P file sharing network topology formation algorithm based on social network information. IEEE NetSciCom2009 in conjunction with IEEE Infocom 2009.

Baset, S. A. and Schulzrinne, H. G. (2006) An analysis of the Skype peer-to-peer Internet telephony protocol. INFOCOM 2006.

Upadrashta, Y., Vassileva, J. and Grassmann, W. (2005) Social networks in peer-to-peer systems. HICSS 2005, Mini-track: The Peer to Peer Paradigm/Emerging Technologies.

Fast, A., Jensen, D. and Levine, B. N. (2005) Creating social networks to improve peer-to-peer networking. 11th ACM Intl. Conference on Knowledge Discovery in Data Mining, pp.568-573.

Feldman, M., Lai, K., Stoica, I. and Chang, J. (2004) Robust incentive techniques for peer-to-peer networks. 5th ACM Conference on Electronic Commerce, pp. 102-111.

Wang, F., Moreno, Y. and Sun, Y. (2006) The structure of peer-to-peer social networks. Phys. Rev. E73.

Wang, W., Zhao, L. and Yuan, R. (2006b) Improving cooperation in peer-to-peer systems using social networks. 20th International Parallel and Distributed Processing Symposium.

Lea, B. R., Yu, W. B. and Maguluru, N. (2006) Enhancing business networks using social network based virtual communities. Industrial Management & Data Systems, 106 (1), pp. 121-138.

Li, Y. M., Tan, Y. and De, P. (2004) Self-organized formation and evolution of peer-to-peer networks. International Conference on Information Systems 2004.

Suh, K., Figueiredo, D. R., Kurose, J. and Towsley, D. (2006) Characterizing and detecting relayed traffic: A case study using Skype. IEEE Infocom 2006.