Peer Production of Online Learning Resources: A Social Network Analysis

Beijie Xu and Mimi M. Recker
beijie.xu@aggiemail.usu.edu, mimi.recker@usu.edu
Department of Instructional Technology & Learning Sciences, Utah State University

Abstract. This paper describes methods for collecting user activity data in a peer production educational system, the Instructional Architect (IA), and then takes a social network perspective in analyzing these data. In particular, rather than focusing on content produced, it focuses on the relationship between users (teachers), and how they can be analyzed to identify important users and like-minded user groups. Our analyses and results provide an example for how to select the most important factors in analyzing the dynamics of an online peer production community using social network analysis metrics, such as in-degree, out-degree, betweenness, clique, and community. In this way, this paper contributes both to process and outcomes research.

1 Introduction

The increased pervasiveness of networked computing coupled with a vibrant participatory web culture has spawned new models of innovation and creation. These models, variously referred to as collective intelligence, crowd sourcing, wisdom of crowds, collective intelligence, or peer production, occasionally take the Internet by storm [1, 11]. As canonical examples, Wikipedia and YouTube need no introduction.

Similarly, in education, the scalable deployment of media-rich online resources supports peer production in ways that promise to radically transform teaching and learning [2, 5]. For example, a growing trend toward sharing online instructional resources has spawned the global *OpenCourseware* movement [19]. Likewise, online educational repositories such as the National Science Digital Library (NSDL.org) collect and curate online learning resources for a wide range of audiences and subject areas [12].

In our own work, we have developed a simple, web-based authoring tool, called the Instructional Architect (IA.usu.edu), which enables teachers to freely find, gather, and produce instructional activities for their students using online learning resources. Teachers can share these resulting activities, called *IA projects,* by making them publically available on the Web. These IA projects can then be *viewed*, *copied*, and *adapted* by other IA users, in ways that support innovative teacher peer production [17]. The latter corresponds a form of indirect collaboration between teachers.

Recent research, however suggests these peer production models may only succeed when they are aimed at focused tasks, coupled with incentives to harness the work of the best collaborators [11]. Without these enabling conditions, many online communities die an untimely death. In short, more is not simply better, and for educational peer production models to succeed, we need more nuanced understandings of how people participate in such environments to efficiently and effectively collaborate around learning resources.

The purpose of this exploratory paper is two-fold. First, we describe our methods for engineering the IA to collect and examine user activities as a means for better understanding peer production of online learning resources. However, instead of focusing on the content produced in the IA, we examine the users in the system and the kinds of networked relationships between them. In particular, we analyze the different social network structures within the IA formed around two user actions, *view* and *copy*, which will be explained shortly.

In the next sections of the paper, we first describe the educational peer production environment, the Instructional Architect. We then briefly review the literature around peer production and social network analysis. This is followed by descriptions of exploratory analyses on the *view* and *copy* social networks underlying the IA, and how they can be used to identify important users and like-minded user groups. In this way, this paper contributes both to process and outcomes research.

2 The Instructional Architect and its Social Networks

2.1 The Instructional Architect

The peer production context for this work is the Instructional Architect (the IA.usu.edu), deployed since 2001. The IA is a lightweight web-based authoring tool, which enables teachers to *quickly* and *easily* find and assemble NSDL resources into learning activities for their students [17].

To use the IA, a teacher must first register to create a free account. Then, once logged in, teachers can use the IA in several ways. The 'My Resources' area of the IA allows teachers to directly search for and collect Web resources, including interactive and Web 2.0 content, and add it to their list of saved resources. In the 'My Projects' area, teachers can create simple web pages in which they select a look and feel for their project, input selected online resources and provide accompanying text. Finally, teachers can 'publish' their resulting IA projects for only their students, or publically for the wider web world. A teacher can search for and view any IA public projects, and make a copy of the project he/she likes. These *view* and *copy* actions form the locus of the present research.

2.2 View Network and Copy Network

For each registered user, we determined the networks between users based on the following two pairs of relationships: 1) user A **viewed** user B's IA project, and user B's IA project was **viewed** by user A, 2) user A **copied** user B's IA project, and user B's IA project was **copied** by user A. Thus, the vertices in each network represent IA users, and a link represents the number of viewer/viewed or copier/copied actions between two users. These two networks (termed view network and copy network respectively) were represented as weighted, directed graphs. Note that the IA user interface does not directly display these relationships to users (instead displaying aggregate data), and thus can be characterized as hidden networks.

3 Related Work

3.1 Peer Production

Much research has examined patterns of participation in Internet-enabled collective intelligence communities. For example, Malone et al. [11] analyzed over 250 example communities, and synthesized these into two sets of related questions. The first set of questions examines whether a *crowd* or a *hierarchy* performs the task, and examines *motivations* and *incentives* underlying the task. In particular, who is performing the task and why are they doing it? The second set of questions examines goals in the system, for example *creation* or *decisions,* and addresses processes, for example *collecting* or *collaborating.* In particular, what is being accomplished and how is it being done?

This framework helps sharpen the focus on important elements of the particular peer production community under investigation, in our case the Instructional Architect. Couched in these terms, the IA relies on a *crowd* of teachers to produce instructional activities around online resources for use by their students, and potentially their peers. Second, the primary *motivation* of teachers is *creating* a collection of useful IA projects for their students. Some may be made from scratch; others may be copied or adapted from other publically available IA projects.

3.2 Social Network Analysis: Key Concepts

Social network analysis (SNA) is a well-established method for studying interactions among human organizations [8, 20]. SNA provides methods for studying the flow of information using mathematical graph theory. A SN consists of vertices (or nodes) and links (or edge) among them. Sometimes the links are directed to show particular relationships. A variety of SNA indicators can be used to identify patterns. For example, *in-degree, out-degree,* and *betweenness* are all measures of vertex importance. Link density measures can be computed with in-degree (the sum of the number of edges pointing to a vertex), and out-degree (the sum of the number of edges leaving a vertex). Another centrality metric, betweenness, measures the extent to which a vertex lies between others, giving a higher value for vertices that can bridge subgraphs together.

More recently, SNA has become widely applied and studied in the Web era, for example, see [13], and educational research. While the standard definition of "social network" implies a social relationship, emerging web tools (e.g., Facebook, Twitter) allows many more relationships to be formed, called online social networks. The networks don't necessarily uniquely consist of "friends", as traditionally defined.

Mislove et al. [14] presented a large-scale measurement study and analysis of the structure of four popular online social networks, including *Flickr, YouTube, LiveJournal,* and *Orkut*. They confirmed the power-law, small-world, and scale-free properties of online social networks. That is, the probability of a person connected with *k* people in a network is proportional to a power of *k* (power-law), most people can be reached by one another through a small number of hops (small-world), but the network itself is characterized by uneven distribution of connectedness with several hubs shaping the way the network operates (scale-free).

SNA has also been applied in educational research. In particular, patterns of social relationship revealed by SNA, coupled with results from other qualitative evaluation methods such as content analysis [21, 4], interviews [4], survey [9, 12], reports [12], and sociometry [9, 12], are frequently used in longitudinal study of the participatory aspects of computer-supported collaborative learning (CSCL). For example, Zhang et al. [21] used SNA to study students' collective cognitive responsibility in a 3-year knowledge-building community. They discovered dense note-reading networks throughout the three years, indicating students' awareness of collective responsibility.

However, it is important to note that in the era of unlimited access to information, human attention has become the critical and scarce resource [7, 16]. As such, simple measures of online social network connectivity may be exaggerating the nature and size of the underlying human social networks [7]. For example, Huberman et al. [7] analyzed the social network underlying *Twitter*. In Twitter, users can post short posts ("tweets") or follow other users' posts ("follower"). In addition, they defined a "friend" as someone to whom a user has directed at least two posts. They then examined relationships between number of tweets, followers, and friends. Their analyses found that the real driver of twitter activity is a much smaller, sparser network underlying the publically declared follower network. The sparse "friend" network, they argued, is the network that matters.

4 Purpose and Research Questions

The purpose of the present study is to take a social network perspective in analyzing a peer production system, the IA. Thus, rather than focusing on content, we focus on the relationship between users in the system. The following general research questions guide our study: What are the networks that seem matter in a group of teachers collecting and producing online resources? Can such approaches be used to help identify the IA's *important* producers? Can it help identify clusters of like-minded teachers? Do different activities have different assumptions about users' cognitive loads, associated costs, and expected payoffs?

5 Data Sources

Table 1 shows a summary for the present study, consisting of the view and copy actions occurring between September 2008 and February 2010. When calculating the average in-degree, we left out users who did not have an incoming edge; a similar procedure was used to calculate the average out-degree. As can been seen, the view network is generally bigger.

Table 1. Summary of the two networks

	# of in-degree		# of out-degree		# of viewers / copiers	# of viewees / copiees	Total # of users
	Average	Max	Average	Max			

View Network	5.12	83	3.63	36	988	700	1283
Copy Network	2.41	83	1.76	18	298	217	388

Figures 1A and 2B show scatter-plots of the number of users as a function of users' view/copy actions. Note that both graphs follow a Zipf (power-law) distribution, as is common in Internet usage data sets [14, 18]. Thus, small numbers of users account for a large amount of these actions.

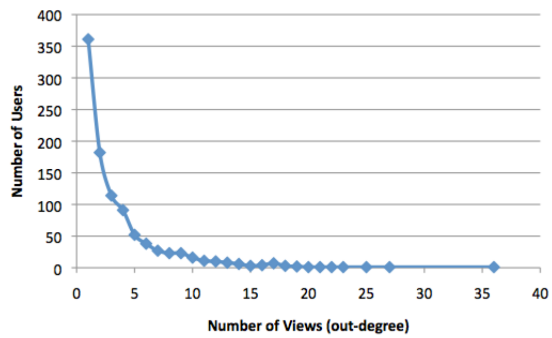

Figure 1A. Number of Users as a Function Of the Number of Peers a User Has Viewed (Out-degree in the View Network)

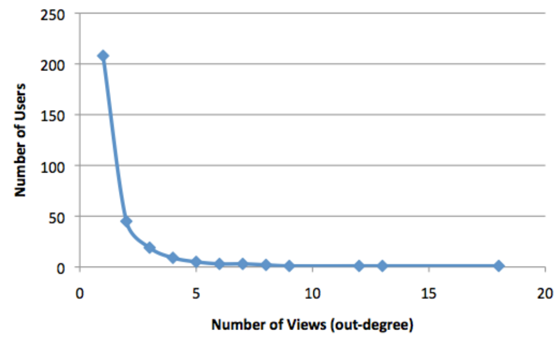

Figure 1B. Number of Users as a Function of the Number of Peers a User Has Copied (Out-degree in the Copy Network)

6 Data Analysis

6.1 Network Comparisons

The view and copy networks were represented within the freely-available SNA software Visone (Visone.info), which also computes key SNA measures for each network. Figures 2A and 2B show the viewer and copier networks, respectively. Each vertex represents a user, and its user ID. Users having higher betweenness scores are placed in the center. Note that the viewer network is much denser, as is also reflected in Table 1. From a user perspective, viewing represents an action with a much lower "cognitive" cost (a simple click) compared to a copy action (which represents a decision to use/adapt the content). Not surprisingly, this difference is reflected in the much sparser copier network compared to the viewer graph.

Betweenness is a measure of centrality based on shortest path [3], and higher scores indicate that a vertex lies on more shortest paths between other vertices. High degree does not necessarily correspond to high betweenness, and vice versa. For example, in the copy network, user 4629 has in-degree = 1 and out-degree = 2. However, this seemingly less active user ranks No. 6 in this copy network, because there is an outgoing edge from him to user 4635 (betweenness rank = 2) and an incoming edge from user 5068 (betweenness rank = 7) to him, and such connections dramatically promotes his degree of centrality (see Figure 2B). On the other hand, user 4517 with an out-degree of 23 is on the fringe of this network (betweenness score = 0), because he has zero in-degree, and is not able to serve as a bridge between vertices. Based on the above argument, we believe betweenness is a better measure of a user's centrality in a social network than the degree measure because it estimates how important one is in terms of helping connect the entire community together.

We arbitrarily set the top 5% (N = 64 and 19, respectively) of each network as the first tier of users, and then made comparisons between these two groups. The first tier is assumed to play an essential role in connecting a network, and we were curious about the degree of match between these two groups. Figure 3 is the distribution of those two groups after mapping from one network to the other. The majority (63.2%) of the first tier of the copy network remain the top users, and only 1 person falls on the fringe of the view network; however, the first tier users of the view network seem to lose their importance after this mapping, with only 20.3% still as top users, 45.3% on the fringe as peripheral players, and 28.1% not even in the copy network at all.

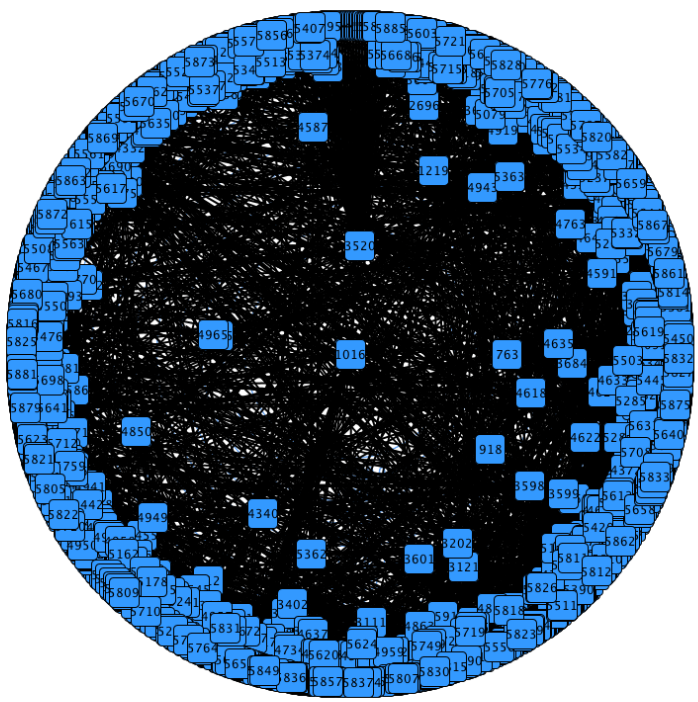

Figure 2A. A View Network with Users of High Betweenness Scores Placed at the Center

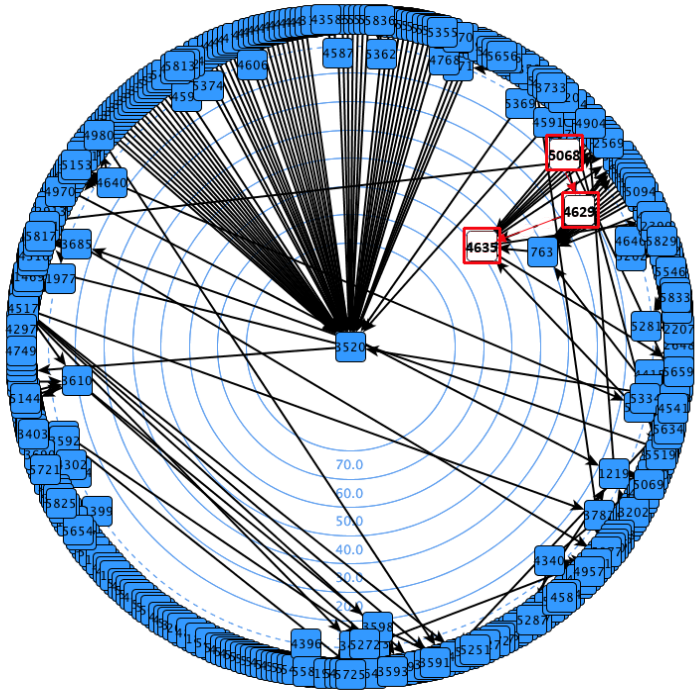

Figure 2B. A Copy Network with Users of High Betweenness Scores Placed at the Center

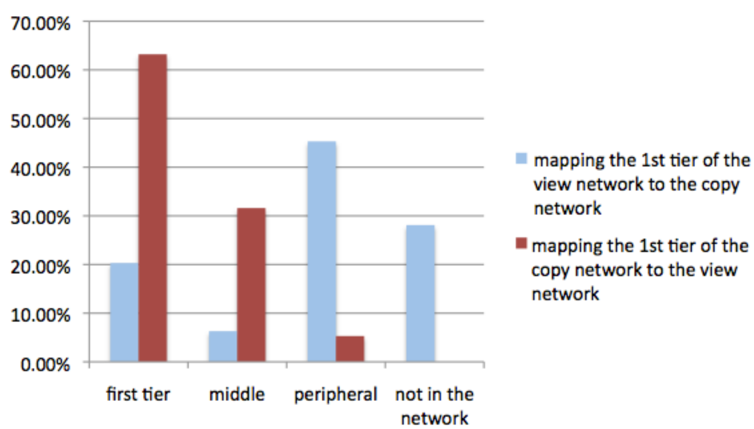

Figure 3. Mapping the First Tier of Users of One Network to the Other

The above analysis reveals that 1), the copy network is much smaller than the view network; 2) the top users of the copy network roughly represent the important users of the view network, but not vice versa. In sum, the copy network is more able to concisely represent the connection between users than the view network.

6.2 Relationship between Project Creation, Viewing, and Copying

Following [7]'s analysis, we studied the relationship between user production of IA projects, and viewing and copying actions. Figure 4A plots the mean number of IA projects created by users as a function of number of peers they have viewed. There was one case (out-degree = 36) far removed from others; and was thus considered an outlier and excluded from this analysis. As can be seen, the mean number of IA projects created initially increases as the number of viewer increases but then saturates except for a peak when out-degree = 25. This implies that users with a large number of views are not necessarily those who create a large number of IA projects.

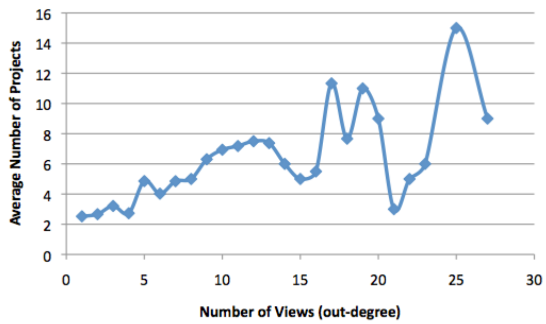

Figure 4A. Number of projects created as a function of a user's out-degree in the view network

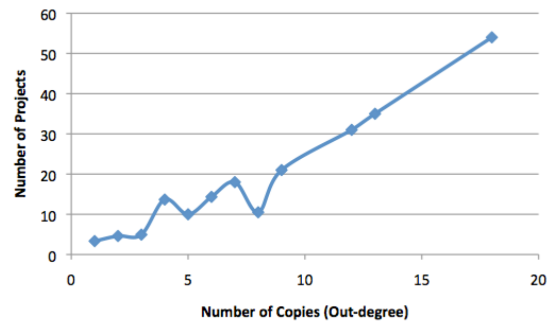

Figure 4B. Number of projects created as a function of a user's in-degree in the copy network

Conversely, Figure 4B plots the mean number of IA projects created by users as a function of the number of copy actions. The mean number of IA projects created does not saturate and exhibits an increasing trend. Thus, the number of copies is a more accurate signal than the number of views in estimating project creation magnitude. As above, the copy action appears to be a better metric for describing meaningful user's activity in a network, as opposed to the view action.

6.3 Cliques

Finally, we applied a *clique* analysis, for example, see [20], on the copy network – the more important network of the two. A *clique* is a subgraph in a network in which every two vertices are connected by an edge. When the number of vertices in such a subgraph is k, it is called a *k-clique*. A clique thus represents closely tied subset of the network. A *k-clique-community* is defined as the union of all k-cliques that can be reached from each other through a series of adjacent k-cliques. Two k-cliques are adjacent if they share k-1 vertices [6, 15].

CFinder is a software tool for network cluster detection, based on the Clique Percolation Method [6, 15]. We used CFinder to detect k-cliques inside the copy network, and eleven 3-cliques were found. These cliques suggest that some small subsets of users share common interests and that they could make use of each other's IA projects. We compared the users' declared subject areas (part of users' registration profile) within these 3-cliques, and found that users of the same clique indeed share similar interests. Eight cliques have all three users with the same declared subject area, and the other three 3-cliques have at least two users with the same subject area. The largest community in the copy network is a 6-clique-community formed by four adjacent 3-cliques (see Figure 5). Since this community represent a closely tied subset of the copy network, not surprisingly, all six users share the same subject area – language arts, and five of them check both math and science, and four of them check social studies. In sum, the clique analysis helped identify teachers with shared interests.

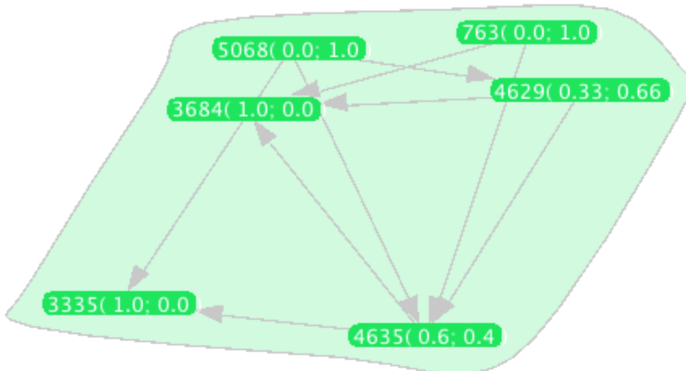

Figure 5. The Largest Community – a 6-clique-community – in the Copy Network. Each Vertex Shows User IDs (and Weights between in-degree and out-degree).

7 Conclusion

In this paper, we have presented a set of preliminary and exploratory analyses of a peer production system, the Instructional Architect, using social network analysis. We compared two implicit social networks – the view network and the copy network, and argued that the copy network is the more concise representation of drivers of peer production and knowledge dissemination within the IA community. From a cognitive point of view, this can be interpreted as the 'higher-cost/higher-benefit' of copy actions in this system.

While admittedly preliminary, our analyses provide an example for how to select the most important factors in analyzing the dynamics of an online peer production community. In addition we have presented several common social network analysis metrics, such as in-degree, out-degree, betweenness, clique, and community. They each serve different purposes. For example, in-degree and out-degree are used to measure one's connection with others; betweenness is used to measure a user's importance in terms of bridging users together; and clique and community metrics are used to identify like-minded groups.

In the future, we plan to extend the current study to examine other centrality measures such as closeness and hub, and their implications for the IA network. We also plan to use the current approach and results to triangulate with results from other educational data mining efforts on these data using latent class analysis. Applying social network analysis to understanding online educational systems is an ongoing research area, and we hope that our exploratory study contributes to its continued growth.

8 Acknowledgment

This material is based upon work supported by the National Science Foundation under Grants No. 840745 & 0840738, and Utah State University. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We thank members of the IA Research Group.

References

[1] Benkler, Y. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*, 2006. New Haven, CT: Yale University Press.

[2] Borgman, C. L., Abelson, H., Dirks, L., Johnson, R., Koedinger, K. R., Linn, M. C., ... Szalay, A. *Fostering Learning in the Networked World – The Cyberlearning Opportunity and Challenge: A 21st Century Agenda for the National Science Foundation* (Report of the NSF Task Force on Cyberlearning), 2008. Arlington VA: NSF.

[3] Brandes, U. A Faster Algorithm for Betweenness Centrality. *Journal of Mathematical Sociology,* 2001, 25(2), p. 163-177.

[4] Cho, H., Gay, G., Davidson, B., and Ingraffea, A. Social networks, communication styles, and learning performance in a CSCL community. *Computers & Education*, 2007, 49, p. 309-329.

[5] Computing Research Association. *Cyberinfrastructure for Education and Learning for the Future: A Vision and Research Agenda*, 2005.

[6] Derényi, I., Palla, G., and Vicsek, T. Clique Percolation in Random Networks. *Physical Review Letters,* 2005, 94(16).

[7] Huberman, B. A., Romero, D. M., & Wu, F. Social networks that matter: Twitter under the microscope. *First Monday* [Online], 2008, 14(1).

[8] Knoke, D. & Yang, S. *Social Network Analysis*, 2008. London: Sage.

[9] de Laat, M., Lally, V., Lipponen, L., and Simons, R. Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for Social Network Analysis*. International Journal of Computer-Supported Collaborative Learning,* 2007, 2(1), p. 1556-1607.

[10] Laghos, A., and Aphiris, P. (2007). Social network analysis of self-taught e-learning communities. *International Journal of Knowledge and Learning, 3*(4-5), 465-482.

[11] Malone, T., Laubacher, R., & Dellarocas, C. Harnessing Crowds: Mapping the Genome of Collective Intelligence. *MIT Sloan Research Paper No. 4732-09*, 2009.

[12] Martínez, A., Dimitriadis, Y., Rubia, B., Gómez, E., and de la Fuente, P. Combining qualitative evaluation and social network analysis for the study of classroom social interactions. *Computers & Education*, 2003, 41, p. 353-368.

[13] McArthur, D. J. & Zia, L. L. From NSDL 1.0 to NSDL 2.0: Towards a Comprehensive Cyberinfrastructure for Teaching and Learning. *Proceedings of Joint Conference on Digital Libraries*, 2008, p. 66-69. New York: ACM.

[14] Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., & Bhattacharjee, B. Measurement and analysis of online social networks. *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement,* 2007, p. 29-41.

[15] Palla, G., Derényi, I., and Vicsek, T. The Critical Point of k-clique Percolation in the Erdõs-Rényi Graph. *Journal of Statistical Physics,* 2007, 128(1/2), p. 219-227.

[16] Pirolli, P., Wollny, E., and Suh, B. So You Know You're Getting the Best Possible Information: A Tool That Increases Wikipedia Credibility. *Proceedings of the 27th International Conference on Human Factors in Computing Systems*, 2009, p.1505-1508. New York, NY: ACM.

[17] Recker, M. Perspectives on Teachers as Digital Library Users: Consumers, Contributors, and Designers. *D-Lib Magazine*, 2006, 12(9).

[18] Recker, M. and Pitkow, J. (1996). Predicting Document Access in Large, Multimedia Repositories, *ACM Transactions on Computer-Human Interaction*, 1996, 3(4), p. 352-375.

[19] Smith, M., & Casserly, C. The Promise of Open Educational Resources. *Change: The Magazine of Higher Learning*, 2006, 38(5), p. 8-17.

[20] Wasserman, S. & Faust, K. *Social Network Analysis Methods and Applications*, 1994. Cambridge: Cambridge University Press.

[21] Zhang, J., Scardamalia, M., Reeve, R., & Messina, R. Designs for collective cognitive responsibility in knowledge building communities. *Journal of the Learning Sciences,* 2009, 18(1), p. 7-44.