Visualizing and Analyzing Productive Structures and Patterns in Online Communities Using Multilevel Social Network Analysis

Hon Jie Teo, Aditya Johri, Raktim Mitra, Virginia Tech, USA
Email: hjteo@vt.edu, ajohri@vt.edu, raktim@vt.edu

Abstract: What does the social network structure of productive online communities look like? We present findings from a study that examined this question in the context of a popular Java programming help forum. Using techniques informed by social network analysis, we leveraged Gephi™ to visualize the social network of the online community and identified a core group of active help-givers central to the community. We then performed network motif analysis and found that a small selection of network motifs or patterns constitute the main building blocks of the entire network..

In this short paper we highlight two approaches for assessing learning interactions and participation in online communities. First, we demonstrate the viability of social network analysis (SNA) as an analytical tool in our study of an online community for programmers learning the Java language. SNA enabled us to identify the existence of a core group of users who enjoy high level of reciprocal ties and are the top participants within this open, voluntary community. Our investigation shows how SNA can be used to delineate different interaction patterns within a help forum resulting in a deeper understanding of learning and knowledge construction. Second, our study identifies productive structures and patterns of interactions that work within the context of a help-forum. Specifically, the sheer number of interactions between the core group of help-givers and a relatively large number of help-seekers necessitates the identification of the structural sub-communities that are formed around individuals of this core-group of help-givers.

Background

With the advent of the internet, learners of the Java programming language can now take their learning beyond the classroom by drawing on a variety of online learning platforms such as Usenet, MOOC and online discussion forums. Amongst the mentioned technologies, online discussion forums have flourished and attracted hundreds of thousands of learners who collaborate to increase their computer programming proficiencies and contribute to constructing knowledge for communal benefit (see Table 1 for an overview of participation statistics of the four main Java online communities). The large numbers of online discussions that take place in these forums indicate that they are potential sites of learning with significant reach.

Table 1: Participation information of four main Java programming online communities

Online Community	Membership	Post Count	Year of Establishment
Java Section of Oracle Forums* (forums.oracle.com) (*Site of study)	995,000	1,457,000	2001
Java Ranch (javaranch.com)	262,000	2,753,000	2000
Java Forums (java-forums.org)	40,000	306,000	2007
Java Programming Forums (javaprogrammingforums.com)	29,000	92,200	2008

Background

One useful methodology for examining interactions between learners in online communities is social network analysis (SNA). This multi-level approach, scholars argue, can help provide a better understanding of individual embeddedness in structural patterns (Frank, 1998; Powell et al., 2005). Social network analysis has found increased acceptance within education due to its ability to shed light on multiple levels of analysis and scholars have called for more network research in education (Hallworth, 1953; McFarland & Klopfer, 2010). A multi-level approach, for instance, helps to connect higher level of organizations – school districts – with micro level classroom data. Over the past few years social network analysis has gained popularity in the CSCL community (Jeong & Hmelo-Silver, 2010). In this context, social network analytics allows one to gain insights into the practices of a social group and can illuminate interactions within structural networks of learners (Haythornthwaite & de Laat, 2010; Suthers et al., 2012).This analytical method enables one to evaluate the nature of interactions between learners to understand the impact of learning activities so that informed instructional decisions can be made (Haythornthwaite, 2008). Overall, social network analysis is an innovative area of research with unique application within the context of our research as it is particularly pertinent for understanding learning analytics at multiple levels of analysis.

Method, Setting, and Analysis

We examined an open online help forum for Java programming language where an online text-based discussion forum facilitates voluntary and open asynchronous discussion oriented towards newcomers' explication of their learning needs. As there are limited studies conducted in out-of-school voluntary educational settings such as online discussion forums, it is helpful to highlight the distinctions in contrast with the formal classroom environment. First, the structure of the online community is informally construed and help is mainly provided by a core group of volunteers. Second, participation and collaboration is voluntary and not mandated by coursework. Third, the task structure in this study deviates from the common set-ups of common wrapper/starter roles and open-ended class discussions without pre-designated roles. In this educational setting, a help-seeker starts a discussion soliciting help from other community members who may emerge as voluntary help-givers to assist them with their learning needs. The data collection procedure involved parsing downloaded web data using Perl scripts and storing them as a relational database. The discussion forum comprised of 37,472 discussion topics created between 2001 and 2010. The activities over these forums were captured and a User information dataset contained details about all the registered users posting within the forum during the given time frame in addition to the replying user ID. We used this data to develop the dataset required for social network analysis (for details see: Mitra, 2011). The Gephi™ software (Bastian, Heymann & Jacomy, 2008) and Hu Yifan's Multilevel Layout force-directed algorithm (Hu, 2005) was leveraged for visualization purposes.

Community Structure

As demonstrated in Figure 1, the force algorithm has been applied to the entire network and the visualization in Figure 2 indicates the presence of a small core group of help-givers with high degrees of interactions with other community members. Given the positions of these core help-givers, it suggests that the community is highly reliant on these individual for their expertise and voluntary efforts.

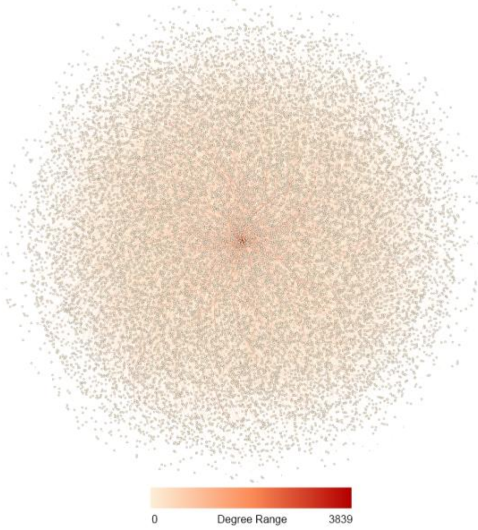


Figure 1. Network visualization of entire community

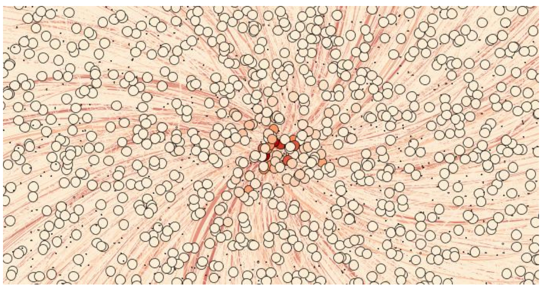


Figure 2. Close-up of visualization reveals core group of help-givers

Table 2 describes the main network properties such as the total number of mean in/out degree and network diameter. There are a total of 21,509 nodes and 125,944 edges for this network. The average path length for the network is 3.253 where the clustering coefficient, which indicates how embedded the nodes are in their neighborhood, is at a relatively low value of 0.335 and indicate that the network has relatively sparse connections. These figures suggest that the social groups in this community are not tightly connected to each other which may be a consequence of deriving the social network from standalone discussion threads.

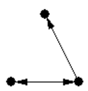
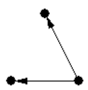
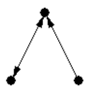
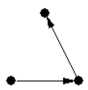
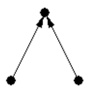
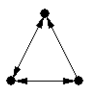
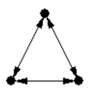
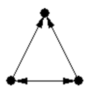
Table 2: Main Social Network Properties

Nodes	21509
Edges	125944
Mean Degree	5.981
Network Diameter	17
Average Path Length	3.253
Clustering Coefficient	0.335

Network Motifs

One formal – empirical and conceptual – avenue for understanding relationships in a community is "Network Morifs". Network motifs has been proposed by Milo et al. (2002) as recurrent patterns of local inter-connections that occur in complex networks at frequencies that are significantly higher (reflected by the Z-score) than those occurring in randomized networks with equivalent number of nodes, in degree and out degree. Motifs can be small subgraphs of typically 3 to 7 nodes and represent the basic building blocks of most networks (Milo et al., 2004; Mangan & Alon, 2003) to provide insights into the topology of complex networks (Juszczyszyn et al., 2008; Kastan et al., 2004). Referring to the motif network analysis (see Table 3), we found that the "branch with one mutual dyad" motif (M14) and the branch motif (M6) make up approximately 63.5% of all recurrent patterns in this community.

Table 3: Frequently Occurring Motifs from Network Motif Analysis

ID	Motif	Frequency [Original]	Frequency [Random network]	S.D.	Z-Score	p-value
14		43.2%	43.2%	3.284e-005	11.1	0
6		20.3%	20.2%	2.665e-005	49.7	0
164		6.0%	5.9%	4.479e-005	17.6	0
12		5.5%	5.4%	3.085e-005	32.5	0
36		0.6%	0.6%	1.583e-005	4.9	0
174		0.3%	0.3%	3.633e-005	2.2	0.007
238		0.2%	0.0%	4.840e-005	37.9	0
46		0.1%	0.1%	1.114e-005	39.6	0

The network motif M14 represents an interaction triad that suggests that a learner is engaged in a bidirectional interaction with one helper and a unidirectional interaction with another. Specifically, the latter two actors are not interacting with each other in this triadic interaction and in contrast to the fully reciprocal motif (M174), the motif is not complete. M6, as the second most frequently occurring motif, can be inferred as learning system with one actor having two unidirectional interactions. This finding is not surprising considering that each help discussion will usually involve helpers engaging in unidirectional interactions with the learner to assist the help-seeker with the learning task at hand. The dominance of the two motifs M14 and M6 suggests that a large number of interactions are not complete in this help-seeking online community. On the other hand, highly connected motifs with more than 2 edges such as M36, M174, M238 and M46 occurred less frequently and this finding suggests that interactions between help-givers and help-seekers are seldom reciprocal in the help discussions.

Conclusion

In this paper, we leveraged both social network analysis and motif analysis as a multilevel assessment approach to examine learning interactions in online communities. Through social network analysis, we found that a relatively large number of help-seekers are supported by a small core group of helpers and suggest that it is critical to consider the high help-seeker-to-helpers ratio in this setting. In addition, we found two network triad motifs (M14 and M16) make up more than half of the network triads and that highly connected motifs were very sparse. The next step to pursue in this research is to examine the quality of discussion in a sample of the discussion data and examine if the motifs play a role in determining the quality of discussion. We believe that our approaches and findings can inform assessment practices in online learning conducted on MOOC and web-based course management systems.

Acknowledgement(s)

This material is based upon work supported, in part, by the U.S. National Science Foundation under Grants EEC-0954034 & EEC-0935143. 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 funding agency.

References

Bastian M., Heymann S., & Jacomy M. (2009). Gephi: an open source software for exploring and manipulating networks. *Proceedings of International AAAI Conference on Weblogs and Social Media.*

Frank, K. A. (1998). Quantitative methods for studying social context in multilevels and through interpersonal relations. *Review of Research in Education,* 23, 171-216.

Hallworth, H. J. (1953). Sociometric relationships among grammar school boys and girls between the ages of eleven and sixteen years. *Sociometry*, 16(1), 39-70.

Haythornthwaite, C. & de Laat, M., (2010). Social networks and learning networks: using social network perspectives to understand social learning. *Proceedings of 7th International Conference on Networked Learning*, Aalborg, Denmark, (pp. 183-190).

Haythornthwaite, C. (2008). Learning relations and networks in web-based communities. *International Journal of Web Based Communities*, 42, 140-158

Hu, Y. F. (2005). Efficient and high quality force-directed graph drawing, *The Mathematica Journal,* 10, 37-71.

Jeong, H., & Hmelo-Silver, C. (2010). An overview of CSCL methodologies. *Paper presented at the 9th International Conference of the Learning Sciences.* Chicago, IL.

Juszczyszyn, K., Kazienko, P., & Musiał, K. (2008). Local topology of social network based on motif analysis. In *Knowledge-Based Intelligent Information and Engineering Systems*, 2008, 97–105.

Kashtan, N., S. Itzkovitz, S., Milo, R. & Alon, U. (2004). Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs. *Bioinformatics*, 20(11), 1746–1758.

Mangan, S. & Alon, U. (2003). Structure and function of the feed forward loop network motif. *Proceedings of National Academy of Sciences*, USA, 100 (21), 11980-11985.

McFarland, D., & Klopfer, E. (2010). Network search: A new way of seeing the education knowledge domain. The Teachers College Record, 112(10), 8-9.

Milo, R., Itzkovitz, S., Kashtan, N., Levitt, R., Shen-Orr, S., Ayzenshtat, I., Sheffer, M. & Alon, U. (2004) Superfamilies of evolved and designed networks. *Science*, 303(5663), 1538–42.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. and Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science*, 298, 824–827.

Mitra, R. (2011). Collaborative learning in Open Source Software (OSS) communities: The dynamics and challenges in networked learning environments. M.S. Thesis, Virginia Tech. http://scholar.lib.vt.edu/theses/available/etd-06272011-170238/

Powell, W.W., White, D.R., Koput, K.W., & Owen-Smith, J. (2005). Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences. *American Journal of Sociology*, 110(4), 1132-1205.

Suthers, D., Hoppe, H. U., Laat, M. & Simon Buckingham, S. (2012). Connecting levels and methods of analysis in networked learning communities. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, British Columbia, Canada (pp. 11-13).