

Using Data Analytics to Examine Expert/Novice Behavior in Informal Online Communities

Aditya Johri & Hon Jie Teo, *Virginia Tech*

Abstract—We present an overview of how data analytics can be used to examine informal learning in online communities and demonstrate the viability of the approach through examples from studies of online communities of engineering enthusiasts. First, we discuss the analysis of social network structure of online communities to discover the existence of a core group of community leaders. Next, we adopt network motif analysis to examine the micro-level interactions in triads and discuss how these network building blocks contribute to a better understanding of the quality of interaction in the community. Lastly, we demonstrate how a temporal analysis coupled with statistical tests can be useful for uncovering environmental factors that contribute to changes in newcomer participation. Our three approaches coalesce into one common purpose – to gain deeper insights into the interactional and participation dynamics of online communities.

Index Terms— Engineering Education, Informal Learning, Learning Analytics, Online Communities

I. INTRODUCTION

IN this paper we present an overview of how data analytics can be applied to examine informal learning in online communities. We use the term “informal learning” [1] to refer to a wide array of mechanisms for learning including “lifelong learning” [2], “non-formal education” [3], and “self-directed learning” [4]. Informal learning constitutes a multitude of activities that are “predominately unstructured, experiential, and non-institutional” [5] and are likely to take place outside the classroom (such as science museums [6]). Other characteristics associated with informal learning include unplanned and incidental cognition which occurs in daily routines or intentional cognition with a particular emphasis on collaborative knowledge building. In the domain of science education research is directed towards investigation of interest-driven learning of science in out-of-school settings museums or out-of-school groups, science camp and

enrichment programs [1, 6]. The focus of informal science learning has benefited from the deemphasizing of science information in favor of engaging participants and the development of interest in free-choice environments [6]. Overall, there have been numerous attempts to characterize learning in informal settings by using a continuum encompassing dissected forms of learning described as informal, non-formal and formal learning [3, 7]. For the purposes of our research within the context of engineering learning we define informal learning as: Social and cognitive behavior change relevant to engineering practice that occurs in engineering students as an outcome of their participation in activities outside of core curricular requirements (where teaching occurs at specific, institutionally dictated times, primarily in classrooms or in laboratories, under close supervision). Informal engineering learning is non-didactic and learner-centered and emphasizes participation in a community of practice.

II. ONLINE COMMUNITIES

Online communities that form around common interests have existed since the dawn of the Internet [9] but have proliferated as an increasing proportion of the population has gained access to the Internet. They are a modern incarnation of ‘list-serv’ discussion groups initiated by academic scholars in the early days of the Internet. Online communities allow participants to acquire information, skills, and other resources relevant to their work interests [10]. Scholars have studied online communities to understand motivation issues, issues of social cohesion and lack of social involvement, and have suggested ways to design online communities for sustained participation [11]. Online communities are increasingly regarded as important venues for promoting learning across the boundaries of time, space, and formal organization [12].

III. ONLINE COMMUNITIES OF ENGINEERING ENTHUSIASTS

One critical avenue for technology-driven informal engineering learning is online communities of engineering enthusiasts (OCEE). These communities have formed around a range of engineering topics and tools – such as microelectronics, rapid prototyping, open source software – and bring together engineering practitioners, hobbyists, and students, to engage in information sharing and product development (e.g. Do-It-Yourself or DIY projects). Although

Manuscript received March 30, 2013. This work was supported in part by the U.S. National Science Foundation under Grants EEC-0935143 & EEC-0954034.

Aditya Johri is Assistant Professor at Virginia Tech, Blacksburg, VA 24061 USA (phone: 540-231-0653; fax: 540-231-6555; e-mail: ajohri@vt.edu).

Hon Jie Teo is a doctoral candidate in Engineering Education at Virginia Tech, Blacksburg, VA 24061 USA. (e-mail: hjteo@vt.edu).

most of these communities do not start with the intent to educate others, it is one critical function that they serve. We believe it is important to examine OCEEs given their increased prominence in the field of engineering and because they attract Net Generation students. These communities have accumulated rich sources of materials and user data and can form the backbone to understand and implement technology-based informal engineering learning. Their examination can lead to an understanding of what works and how to accelerate innovations in cyber-enabled learning within engineering education. OCEEs are specifically appropriate for studying informal engineering learning for several reasons. First, they are informal in the manner in which they are structured and rather than being hierarchical based on age or tenure the hierarchy is informally established over time based primarily on expertise. Second, people participate in these communities due to their affinity [13] with the topic of the community and form “communities of interest” or, in many cases, communities of practice [14]. Finally, an important aspect of these communities, relevant to engineering education, is that people assemble here not only to discuss and share knowledge but to design knowledge-based products and services. Therefore, the ability to make and share expertise on making is critical for participation in these communities. As scholars who have applied constructivism to education have noted, building something for public exchange is often one of the strongest intrinsic motivations for learning [15]. We believe that these communities have a great potential for transforming engineering education particularly in this era where the engineering curriculum has become highly saturated with required courses and activities leaving little, if any, room to add new courses or activities.

IV. FINDINGS FROM STUDIES OF OCEES

Over the past couple of years, we have conducted several studies of informal learning in OCEEs with specific focus on computer and software engineering communities. The studies analyze learning on these communities using online digital data. The data is collected through parsing of forums on which discussion takes place. As an example, one of the biggest datasets we have is from a forum “New to Java” which is a community for those interested in learning the programming language Java. We have collected participation data for over 10 years consisting of over 200,000 discussion messages. Using this data we have examined several questions which we discuss below (also see [16]-[17]).

A. Community Structure Using Social Network Analysis

The overall functioning of OCEE is supported by collaborative activities – where learners and experts work together to answer questions and pursue common interests. The importance of social interaction and learner collaboration lead us to social network analysis, whereby the emphasis is placed on examining ties between social units. This approach is particularly useful due to the existence of a massive amount of discussions on the OCEE can be characterized as social networks. A macro-level social network analysis approach can

reveal the participation structure of the community of interest. This research approach is informed by advances in social network analysis and we leveraged Gephi for the visualization process [18]. As demonstrated in Figure 1, we apply the Hu YiFan [19] 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.

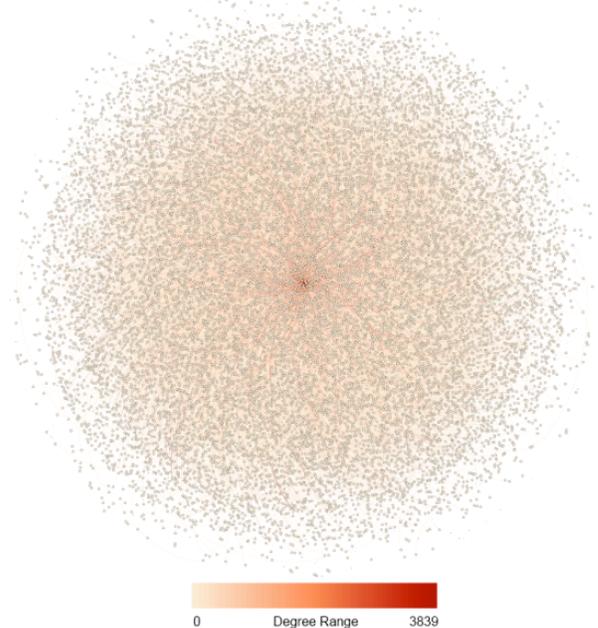


Figure 1: Network Visualization of Entire Community

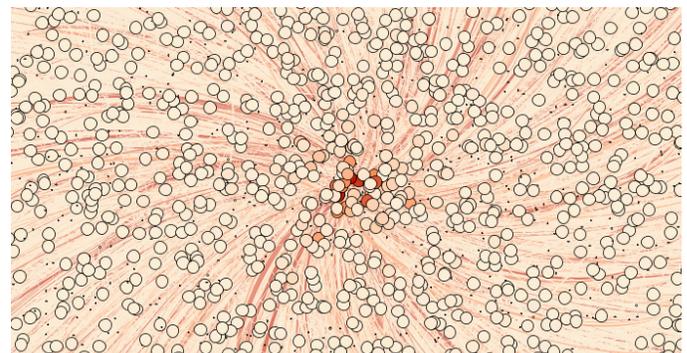


Figure 2: Close-up Visualization Showing Core Group of Help-givers

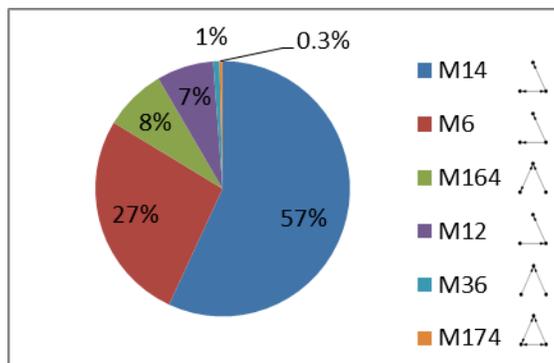
Table 1 describes the main network properties such as the total number of mean in/out degree and network diameter. There are a total of 21,509 nodes and 125,944 edges for this network. The average path length for the network is 3.253 where the clustering coefficient, which indicates how embedded the 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 1: Network Properties

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

B. Motifs – Patterns of Community Sub-structures

One formal – empirical and conceptual – avenue for understanding micro-level interactions in a community is “Network Motifs”. Network motifs have been proposed by Milo and colleagues [20] as recurrent patterns of local interconnections 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 (to provide insights into the topology of complex networks [21]). While its use in social network analysis has been limited, we adopt this approach to gain deeper insights to the micro-level interactions amongst individual grouped into triads. More specifically, we explored if certain types of motifs make up a significant portion of interactions in the OCEE.

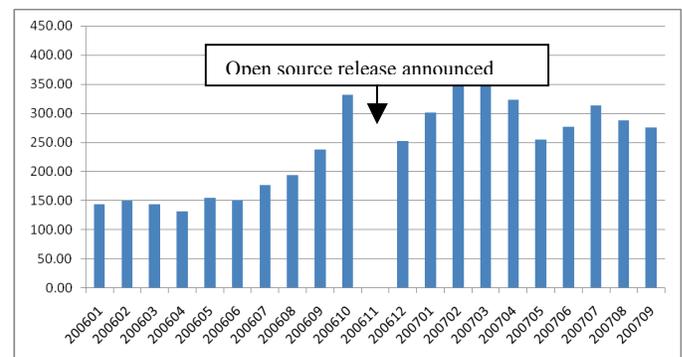
**Figure 3: Frequency of occurrence of motifs**

Referring to the motif network analysis (see Figure 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. 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.

C. Effects of External Events on Newcomer Participation

The attraction and retention of newcomers is a core concern to OCEEs as it reflects the much-needed community membership renewal. Researchers studying membership and persistence in organizations have argued that environmental factors rarely threaten well-designed organizations but can reveal how they react and adapt to them [22]. In one of our study, we drew from the concept of “environmental jolts” and conducted a temporal analysis of participation activities in a Java developer community triggered by two significant events: (1) open sourcing of Java by Sun (Nov. 19, 2006) and (2) acquisition of Sun (and consequently of Java) by Oracle (April 20, 2009). Both events represent a change either in the nature of product or in the ownership of the product and consequently the stewardship of the associated online community.

**Figure 4: Participation change due to environmental change**

We first captured a sizeable collection of archival data from Java developer community and parsed the data to capture the necessary user participation data with the goal of performing statistical tests on the four membership groups based on the amount of participation activity over time – bronze, silver, gold and platinum. The results of the quantitative analysis of forum participation before and after the announcement of open sourcing of Java by Sun reveal a significant difference in terms of both participation dimensions (see Figure 4): the average number of new forum users per day increased from 15.8 to 20.1 ($p < 0.01$, Mann-Whitney test), the average number of discussion threads per day increased from 29.4 to 42.8 ($p < 0.01$, Mann-Whitney test), and the average number of messages posted per day increased from 181.5 to 304.7 ($p < 0.01$, Mann-Whitney test). The number of discussion threads and the number of messages posted were highly correlated ($r = 0.96$ before the announcement; $r = 0.92$ after the announcement). Activity broken down by user status indicated a sharp increase in the number of Bronze users (Average of 776.75 before the announcement to an average of 1020.65 after the announcement ($t = -3.10$, $P < .01$). The other groups of users also showed a steady and significant increase in number although the jump in numbers was not large (see Figure 5 on the next page).

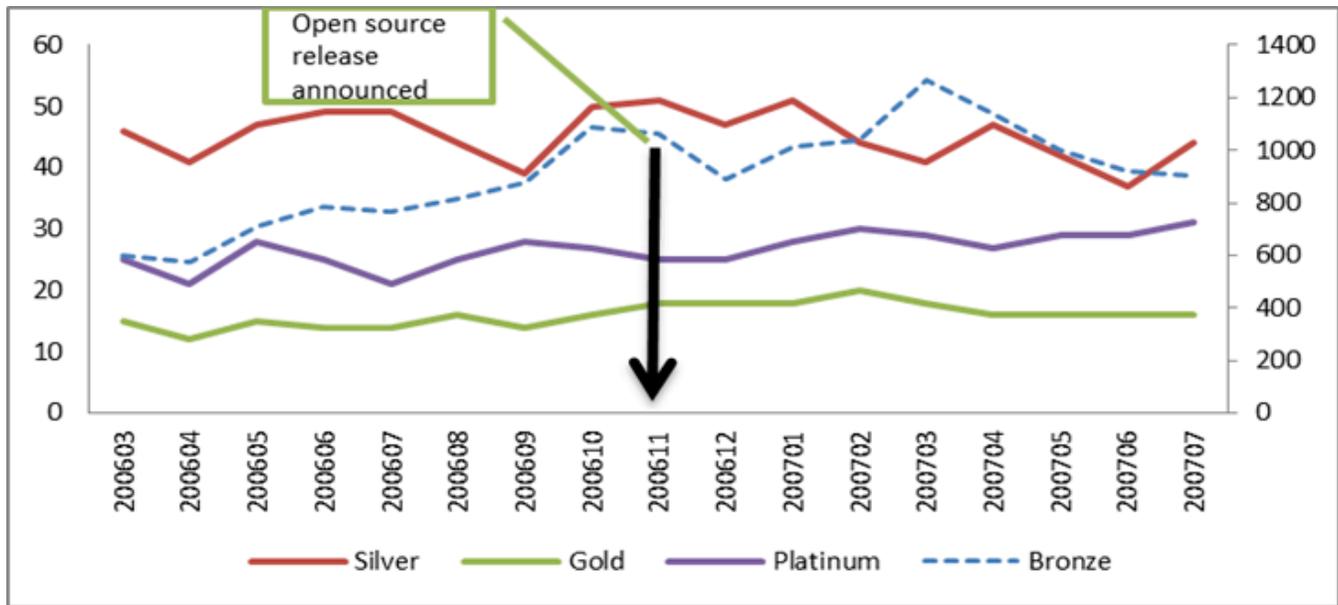


Figure 5: Differences in participation across four distinct membership groups

V. DISCUSSION AND CONCLUSION

In this paper, we describe three data analytics approaches: social network analysis, motif analysis and temporal analysis of participation. These analytics approaches can be extended to other platforms including formal online platforms complementing engineering coursework and MOOC. This is especially so considering that many engineering classes use course management software such as Blackboard that facilitates online discussion and learning. MOOC learning platforms on the other hand are supported with discussion forums to foster collaboration and knowledge sharing amongst learners and with instructors. Findings from our study show that by using data and analysis in novel ways we can uncover interactional and participation dimensions of learning communities that are not immediately representable and, hence interpretable, to an education practitioner.

REFERENCES

- [1] P. Bell, B. Lewenstein, A.W. Shouse and M.A. Feder, *Learning Science in Informal Environments: People, Places, and Pursuits*, The National Academy Press, 2009.
- [2] Accreditation Board for Engineering and Technology (ABET). *Criteria for Evaluating Engineering Programs*, 2011.
- [3] P. H. Coombs, P. H. and M. Ahmed, *Attacking Rural Poverty. How non-formal education can help*. Baltimore: John Hopkins University Press, 1974.
- [4] M. Knowles, *The Modern Practice of Adult Education. From pedagogy to andragogy*. Englewood Cliffs: Prentice Hall/Cambridge, 1980.
- [5] V. Marsick and M. Volpe. "The Nature of and Need for Informal Learning," in *Informal Learning on the Job, Advances in Developing Human Resources*, vol. 3, V. Marsick and M. Volpe, Eds. San Francisco: Berrett Koehler, 1999.
- [6] J. Falk and L. Dierking, *Learning From Museums*. Walnut creek: Altamira Press, 2000
- [7] Eraut, M. "Non-formal learning, implicit learning and tacit knowledge," in *The Necessity of Informal Learning*, F. Coffield, Ed. Bristol: Policy Press, 2000.
- [8] European Commission, *A memorandum on lifelong learning*. Brussels: SEC, 2000
- [9] M. Smith and P. Kollock, *Communities in Cyberspace*. London: Routledge, 1999.
- [10] L. Lakhani, and E. von Hippel. "How Open Source Software Works: "Free" User-to-User Assistance," *Research Policy*, vol. 32, no. 6, pp. 923–943, June 2003.
- [11] R. Kraut, M. Patterson, V. Lundmark, S. Kiesler, T. Mukopadhyay and W. Scherlis. "Internet paradox: A social technology that reduces social involvement and psychological wellbeing?" *American Psychologist*, vol. 53, no. 9, pp. 1017-1031, 1998.
- [12] M. Ahuja and K. Carley, "Network structure in virtual organizations," *Organization Science*, vol. 10, pp. 741-757, 1999.
- [13] J. Gee, "Semiotic Social Spaces and Affinity Spaces: From The Age of Mythology to Today's Schools," in *Beyond communities of practice: Language, power and social context*, D. Barton and K. Tusting, Eds, Cambridge: Cambridge University Press, 2005, pp. 214–232.
- [14] J. Lave and E. Wenger, *Situated Learning. Legitimate peripheral participation*, Cambridge: University of Cambridge Press, 1991.
- [15] S. Papert. *Mindstorms: Children, computers and powerful ideas*. New York, New York: Basic Books, 1980.
- [16] Teo, H. J., Johri, A. & Mitra, R. (2013). Visualizing and Analyzing Productive Structures and Patterns in Online Communities Using Multilevel Social Network Analysis. Proceedings of CACL 2013, Wisconsin, MI, USA.
- [17] Johri, A., Nov, O. & Mitra, R. (2011). "Cool" or "Monster"? Company Takeovers and Their Effect on Open Source Community Participation. Proceedings of iConference 2011, Feb. 8-11, Seattle, USA.
- [18] M. Bastian, S. Heymann and M. Jacomy, "Gephi: an open source software for exploring and manipulating networks." In *Proc. International AAAI Conference on Weblogs and Social Media*, San Jose, CA, 2009.
- [19] Y. F. Hu., "Efficient and high quality force-directed graph drawing," *The Mathematica Journal*, vol. 10, pp. 37-71, 2005.
- [20] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii and U. Alon. "Network motifs: simple building blocks of complex networks," *Science*, vol. 298, pp. 824–827, 2002.
- [21] K. Juszczyszyn, P. Kazienko, and K. Musia, "Local topology of social network based on motif analysis," *Knowledge-Based Intelligent Information and Engineering Systems*, pp. 97–105, 2008.
- [22] A. D. Meyer, V. Gaba, and K. A. Colwell, K. A. "Organizing far from equilibrium: Nonlinear change in organizational fields," *Organization Science*, vol. 16, no. 5, pp. 456–473, 2005.