# Social Motif Analytics: Network Building Blocks for Assessing Participation in an Online Engineering Community

**Mr. Hon Jie Teo, Virginia Tech**

# Social Motif Analytics: Network Building Blocks for Assessing Participation in an Online Engineering Community

## Abstract

The increased use of computers and greater accessibility of the internet have triggered numerous educational innovations such as online discussion forums, podcasts, Wikis, Open Educational Resources, MOOCs, to name a few. These advances have brought with them a wide range of instructional videos, written documents and discussion archives that provide opportunities for engineering learners to engage with one another beyond the engineering classroom. However, it remains a challenging task to assess the quality of participation on these learning platforms particularly due to the informal nature of engagement as a whole and the massive amount of learner-produced data. The purpose of this study is to assess the quality of participation and interactions by identifying building blocks of interactions in an online engineering community for electrical and electronics engineering learners (AllAboutCircuits.com). Data mining techniques are used to collect half a million messages from close to a hundred thousand pages of discussion and to extract learning traces to obtain network linkages between learners. The data is then examined by a micro-level social network analytical technique – network motif analysis – to identify and characterize building blocks of interactions in the online community of learners. Findings suggest that in-thread interaction occurred at a low rate and reciprocal interactions are equally likely to occur in discussion topics across varying levels of in-thread interactions. This research delineates a number of considerations why network motif analysis can be an apt approach for the assessment of the quality of in-thread interactions and the corresponding network reciprocity in thread-based online discussion forums. Overall, the study provides evidence that interaction dynamics between the learners based in their discussion roles is a critical consideration in evaluating the quality of interaction in online engineering communities.

## Introduction

With the exponential growth of information communication technologies, researchers and policymakers have coalesced around the use of technology to advance learning opportunities in and beyond the formal educational institutions[1]. Their projections and vision are essential as the current generation of students are highly proficient with using computers and have learning habits that are associated with the frequent use of the Internet[2]. Online communities represent an avenue for these digital native students to participate in activities that align with their academic needs and personal interests[3]. In this research, attention is cast on an online engineering community supported by text-based discussion forum software, which allow for open-ended and asynchronous communication without the constraint of time and geographic location[4].

In engineering education, researchers and practitioners have embraced Internet-powered applications and resources to drive innovation in engineering teaching and to enhance access to engineering learning resources[5]. While there is significant research work on online educational settings tied to formal classroom, there has been limited work on informal online communities which are supported by volunteers of varying levels of expertise. As a contribution to this pool of literature, this study looks at an informal learning platform that facilitates the discussion and

collaboration of both learners and experts who have academic interests in electrical circuits and electronics. In online engineering communities, the voluntary contributions from established members or experts represent a main resource for learning. This is particularly important considering an exponential increase in the availability of online resources, text, documents, videos and other forms of aids for learning found in learning platforms outside formal environments. Overall, this research addresses the heightened interest in assessing online learning in connection with formal institutions reveals a dire need to foster a deeper and more rigorous way at evaluating engineering learning experiences outside the formal educational ecosystem.

**Social Network Analysis as an Assessment Approach**

Social network analysis has emerged as a major research perspective within the educational research community as scholars undertake research endeavors stemming from the demands of "big data", examination of complex learning interactions and illuminating relational data in learning networks [7, 8]. The technique has found credibility as it allows researchers to make clear and illuminate interaction processes in their study of networks of learners, through an emphasis on relational properties and social structures [9, 10]. From this perspective, learning can be viewed as a relation that connects learners and as a network outcome of relations supported by interactions [9, 11]. Social network analytics allows one to gain insights into the practices and interests of a social group [12] and to examine interactions between individuals based on sharing common knowledge and practices [13].

Primary uses of social network analysis by educational researchers in learning analytics include clarifying and characterizing relationships between learners in online communities. For instance, Cambridge and Perez-Lopez [14] studied an online community of professional teachers and their interactions with content objects using bimodal social network analysis. They gain insights into highly influential individuals through egocentric usage maps and found that influenced users are persistently engaged with the communities through the advancement of content objects. Community detection techniques are widely used in research of learning analytics. Suthers and Chu [15] studied a professional network of educators with community detection algorithms to illuminate prominent actors and describe major communities within the social network. Their research found evidence that analytical algorithms can better the understanding of social structures of groups within large online communities. Zhuhadar and Yang [16] analyzed the HyperManyMedia user log files using community detection algorithms to support the design of a recommend system that is capable of proposing learning resources to the learners. SNA techniques can also complement another research methodology and an example of such an application is seen in discourse-centric learning analytics, where social network analysis complement discourse analysis to identify patterns of activities that correspond to knowledge construction and meaning-making[17].

Overall, social network analysis is an assessment approach that aids the evaluation of social relations and represents an innovative area of research with unique application within the context of this study. The reviewed research outlines potential examples of how practitioners and researchers can evaluate the nature of interactions between learners, to understand the impact of learning activities and make evidence-backed pedagogical decisions. However, a review of literature also suggests that social learning analytics has not been performed on online

engineering communities supported by off-the-shelf online discussion forums software. This research therefore serves to uncover the basic elements of social structures in an online engineering community.

**Research Setting and Data Collection**

This study examines an open online engineering community, All About Circuits, that is catered towards learners of electrical engineering topics such as circuits, electronics, microprocessors and programming (see Figure 1). This community is primarily supported by an online text-based discussion forum that facilitates voluntary and open asynchronous discussion. Community membership is open and persistent: an individual can register for a forum account with an email address and maintain an avatar within the community.

As there are limited studies conducted in out-of-school voluntary educational settings such as online engineering communities, it is pertinent 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 is voluntary and not mandated by coursework. Third, the task structure in this study deviates from the common discussion set-ups such having wrappers/starters and open-ended class discussions facilitated by an instructor. In contrast, discussion on the online community in this study is initiated by any participating member who creates a discussion thread with the title and first message as descriptors for the topic of discussion. As the discussion evolve, other community members may participate in the discussion to further the discussion and address the learning needs of the community member.



Figure 1. All About Circuits Online Community

Data collection is focused on the download and organization of ten years' worth of discussion web pages. As each of the discussion threads may have different discussion length, the section-level pages that have a list of the discussion topics are first downloaded to find out the number of messages in each discussion thread (see Figure 1). Based on this information, an automated downloader is then used to download all pages of discussion associated with each discussion

thread. After the targeted web pages have been downloaded, the social network data is extracted by means of Python software program based on BeautifulSoup library. Social network data is derived by means of extracting all instances indicative of a direct interaction between two users such as a reply and quote.

## Network Motif Analysis As An Assessment Approach

One network analytical technique for understanding micro-structures within networked systems is motif analysis. Network motifs has been proposed by Milo and colleagues [18] 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 are typically small sub-graphs of typically three to seven nodes and represent the basic building blocks of most networks [19] to provide insights into the topology of complex networks [18, 19].

Motif analysis originated from the field of bioinformatics where it has been used to analyze gene networks and the World Wide Web and had found its way into the research of social networks [20]. Findings from the use of network motif analysis are mixed. Researchers have studied a number of online communities such as Yahoo Answers!, music remix online communities and a Japanese video-sharing website [21,22,23]. The most frequently occurring motif was the feed-forward loop for Yahoo Answers! [21], branch pattern for a music remix community [22] and merge loop for the video-sharing website [23]. All of them differed and suggested a different social dynamic. Overall network motif analysis appears to be an exploratory measure for uncovering building blocks of interactions on online communities.

In this study, network motif analysis[24] is adopted as an educational assessment approach. This technique considers the corresponding discussion role of the user (node) in addition the directedness of the interactions in the motifs. The discussion role of the node refers to the discussion role of the user involved in a discussion and a user can be either an author or a participant (see Figure 2). An edge refers to a social link or interaction between two users whereas the directedness of the edge will indicate the direction of the interaction from the source node to target node.
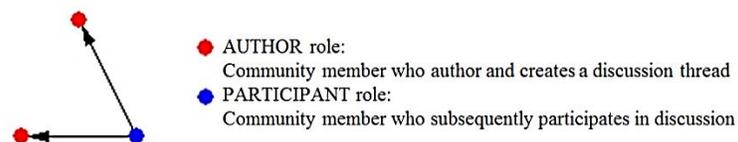


AUTHOR role:
Community member who author and creates a discussion thread
PARTICIPANT role:
Community member who subsequently participates in discussion

**Figure 2. Example of a Motif**

## Assessing In-Thread Discussion

One of the challenges of social network analysis is to assess social networks based on the discussion thread (or topic) that describes the interactions being made. This consideration is particularly essential as interactions on discussion thread may not necessarily take place within the discussion thread and can span across distinct discussion threads (See Figure 3). In this online engineering community, communication is facilitated through four discussion sections that comprise of a thread that contains messages contributed by participating users. In the

following sections, distinctions between A-type and B-type motifs are described and a discussion is carried out on how they can be leveraged to inform the assessment of interactions in thread-based online community.
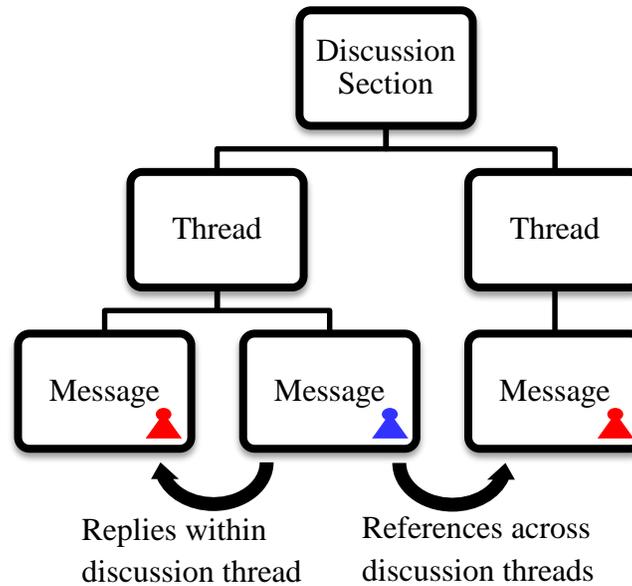


**Figure 3. Assessment Challenges of Interactions across Discussion Threads**

*A-Type Motifs*

Table 1 presents the most frequently occurring triad patterns between three users in the community. The top ten motifs make up approximately 72% of all motifs in the network. The top three motifs are A-type motifs that comprise of two users with the author role and one user with the participant role. A-type motifs make up close to 58% of all motifs in the network and the most common motif (A6) is characterized by two authors with direct interaction with each other and one participant with unidirectional edge with the two authors. The A6 motif indicates that a participant is engaged in two distinct discussion threads initiated by two distinct authors. As the interaction between the authors is non-existent, there is no cross-reference and direct connection between the two authored topics. Furthermore, the unidirectional edge between the authors and the participant suggest that there is no direct engagement between the authors and the participant in this triad network. According to these details, the occurrence of the A6 motif suggests that a relatively large number of the discussion topics have attracted contributions from community members but little participation from the members who initiated the topic with their problems of understanding. It also suggests that numerous discussion topics do not contain a cross-reference from the member who initiated the topic (author) although they are helped by the same community members (participant).

| ID | Motif | Frequency | ID | Motif | Frequency |
|----|-------|-----------|----|-------|-----------|
| A6 | | 15.8% | A164 | | 5.8% |
| A14 | | 11.3% | A36 | | 4.6% |
| A12 | | 10.1% | B12 | | 3.7% |
| B6 | | 8.0% | B164 | | 3.1% |
| B14 | | 6.8% | B168 | | 3.0% |

**Table 1. Top 10 Most Frequently Occurring Network Motif**

### B-Type Motifs

The B-type motifs (B6, B14, B12, B164, and B168) consist of two users with the participant role and one user with the author role. They make up approximately 35.1% of all triad network motifs in the network. Pertaining to interaction dynamics between the two participants, the interaction between the two participants may most likely take place in the same discussion topic or across two distinct discussion topics as the reply or quote mechanism facilitate interaction on the same thread. This is opposed to A-type motif where interactions take place across two distinct discussion topics due to the presence of two authors. In other words, an A-type motif is indicative of only one in-thread interaction. On the other hand, B-type motif represents a complete in-thread interaction and can be understood as better indicators of interaction and relations in the community. Overall, the tabulation of the frequencies of occurrences of network motifs in this study suggests that A-type motifs (58%) occurred approximately 50% more frequently than B-type motifs (37%). This implies that in-thread interactions occurred at a low rate and that users with author roles are less likely to make connections to other authors.

### Assessing Reciprocity

Social network reciprocity can be understood though the presence of a bidirectional edge between two nodes. The occurrence of a bidirectional edge in an A-type motif (such as A14) is a more relevant measure of reciprocity as compared to a bidirectional edge in a B-type motif (B14 and B168). To understand this point, let us consider the presence of a bidirectional edge in three motifs (A12, A14, and A6). The A12 and A14 motif are distinct from the A6 in one manner – the interaction between one of the author roles and the participant role is bi-directional and unidirectional respectively. These motifs suggest that the user with the author role may engage in different types of interaction with another user with the participant role. In the A12 motif, this

engagement is proactively sought by the topic author whereas in A14, this engagement is bidirectional between authors and participant.

Arguably, a bidirectional edge in an A-type motif is a stronger measure of reciprocity as compared to a bidirectional edge in a B-type motif. This can be illustrated in the motifs of A14 and B14. In the B14 motif, a bidirectional edge between the two participants suggests that a reciprocal social relationship exist between them. However, this reciprocal relationship is not between the participant and author as seen in A14.  Since discussion threads are focused on the needs of the community member who initiates the thread (author) rather than the community member who subsequently participates in the discussion (participant), a bidirectional edge between the author and participants is indicative of a more engaged community activity and usually found in A-type motifs.

Overall, the tabulation of the frequencies of occurrences of network motifs in this study suggests a relatively lower number of bidirectional and reciprocal relations in the community (A14, B14 and B168 have bidirectional edges). Out of these three motifs, only A14 represents a bidirectional edge between users with the author and participant roles. Overall, a quick takeaway from the examination of motifs with bi-directional ties suggests that there are fewer reciprocal in-thread ties between users with the author and participant roles in the community as compared unidirectional in-thread ties. This implies that numerous interactions in the online community took place in a unidirectional manner and without active engagement with the author and participants. Considering that there are the number of A-type motifs with bidirectional edges (10.1% of all motifs) is quite similar to the number of B-type motifs with bidirectional edges (9.9% of all motifs), reciprocal relationships are equally likely to occur in discussion with varying levels of in-thread interaction.

**Discussion and Conclusion**

This paper presented an exploratory application of network motif analysis in assessing participation in an online community hosted on a thread-based discussion platform. This approach offers a potential solution to the challenge of evaluating interactions on affordable and common thread-based discussion forum platforms. The study site has over sixty thousand discussion topics which mean that the research addresses the challenge of assessing social interactions based on the topics and threads in which interactions take place, rather than those that occur outside of topics. Motif analysis provides a potentially fruitful approach by tailoring the triadic network analysis based on the discussion roles of the community members engaged in interaction.

There are two merits of using role-based motif analysis. Firstly, the analysis of the occurrences of A-type and B-type motifs, through the tabulation of the frequencies of occurrences of network motifs in the community, allows for the examination of in-thread interactions and therefore the quality of interactions in the same discussion topic or thread. This is facilitated by a comparison of the number of B-type motifs to A-type motifs in the community network, whereby only the former is indicative of in-thread interaction. Secondly, network motif analysis provides a descriptive understanding of the number of bidirectional edges between two fundamental discussion roles in the forums. Since discussion threads are focused on the needs of the

community member who initiates the thread (author) rather than the community member who subsequently participates in the discussion (participant), a bidirectional edge between the author and participant is indicative of a more engaged community activity. In the context of this study, this analysis can be facilitated by comparing the number of A-type motifs to B-type motifs whereby bidirectional ties in A-type motifs are indicative of more engaged levels of interaction.

In tune with the call for more rigorous online assessment approaches in engineering education[5], network motif analysis offers a network perspective beyond educational research methodologies that focus on student self-perceptions and measurements. Network motif analysis is light-weight solution that is capable of evaluating large amounts of educational data in educational platforms that cater to the need of many students (such as MOOCs) and rely on the participation of masses (Q&A sites and online communities). The proposed approach will resolve some assessment challenges in examining student participation across different bodies of social groups and online engineering spaces. In online environments that host thousands of engineering learners, network motif analysis will offer descriptive accounts of recurring interaction patterns between novices and experts, as well as consistent forms of interactions between groups of engineering learners that is indicative of sustained participation in online learning.

Future research plans include the examination of the types of learner discourse and artifacts that support the formation of recurring network motifs in online communities. In addition, further research will consider the occurrence of C-type and D-type motifs, which consist of all author and all participant roles respectively. These were not discussed in this research as they make up less than 3% of the all network motifs. It may be useful to examine C-type and D-type motifs to understand the implications of these motifs, as well as the contrast against the learning processes that underlie the occurrence of A-Type and B-type motifs.

## References

[1] Atkins D. E. 2010.*Transforming American education: learning powered by technology, National Education Technology Plan.* U.S. Department of Education, 2010.

[2] Oblinger, D., and Oblinger, J. 2005. Is it age or IT: First steps towards understanding the net generation. In D. O. A. J. Oblinger Ed., *Educating the Net Generation* pp. 2.1-2.20.

[3] Bruckman, A. 2006. Learning In Online Communities. In R. K. Sawyer Ed., The Cambridge handbook of the learning sciences pp. 461–472. NY: Cambridge University Press.

[4] van de Sande, C. and Leinhardt, G. 2007. Help! Active student learning and error remediation in an online calculus e-help community. Electronic Journal of e-Learning. 5(3), 227-238,

[5] Bourne, J., Harris, D., and Mayadas, F. 2005. Online engineering education: Learning anywhere, anytime. *Journal of Engineering Education*, 94(1), 131–146.

[6] van de Sande, C. 2010. Free, open, online, mathematics help forums: the good, the bad, and the ugly. *Paper presented at the Proceedings of the 9th International Conference of the Learning Sciences - Volume 1*, Chicago, Illinois.

[7] Wellman, B. 2012. Networked individualism: how the personalized Internet, ubiquitous connectivity, and the turn to social networks can affect learning analytics. In *Proceedings of the LAK '12,* (Vancouver, British Columbia, Canada, April 29- May 02, 2012).

[8] Siemens, G. 2012. Learning Analytics: Envisioning a Research Discipline and a Domain of Practice. In *Proceedings of the LAK '11* (Vancouver, Canada, April 29-May 02, 2012).

[9] Haythornthwaite, C. 2011. Learning networks, crowds and communities. In *Proceedings LAK '11,* (Banff, Alberta, Canada, February 27 – March 01, 2011).

[10] Suthers, D., Hoppe, H. U., Laat, M. and Simon Buckingham, S. (2012). Connecting levels and methods of analysis in networked learning communities. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, (Vancouver, British Columbia, Canada, April 29 – May 02, 2012). LAK '12. ACM, New York, NY, 11-13.

[11] Ferguson, R. and Buckingham Shum, S. 2012. Social learning analytics: five approaches. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, (Vancouver, Canada, April 29 – May 02, 2012). LAK '12.

[12] Haythornthwaite, C. and de Laat, M., 2010. Social networks and learning networks: using social network perspectives to understand social learning. In *Proceedings of 7th International Conference on Networked Learning*, Aalborg, Denmark.

[13] Haythornthwaite, C. 2006. Learning and knowledge exchanges in interdisciplinary collaborations. *Journal of the American Society for Information Science and Technology*, 57, 8, 1079-1092.

[14] Cambridge, D. and Perez-Lopez, K. 2012. First steps towards a social learning analytics for online communities of practice for educators. In *Proceedings of the LAK '12* (Vancouver, British Columbia, Canada, April 29 – May 02, 2012).

[15] Suthers, D., and Chu, K. 2012. Multi-mediated community structure in a socio-technical network. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, (Vancouver, British Columbia, Canada, April 29 – May 02, 2012). LAK '12. ACM, New York, NY, 43-53.

[16] Zhuhadar, L., and Yang, R. 2012. Cyberlearners and learning resources. In *Proceedings of LAK'12*, (Vancouver, British Columbia, Canada, April 29 – May 02, 2012).

[17] De Liddo, A., Shum, S. B., Quinto, I., Bachler, M. and Cannavacciuolo. 2011. Discourse-Centric Learning Analytics. In *Proceedings of the LAK 2011* (Banff, Alberta, Canada, February 27 – March 01, 2011).

[18] 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.

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

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

[21] Adamic, L. A., Zhang, J.; Bakshy, E.; and Ackerman, M. S. 2008. Knowledge sharing and yahoo answers: Everyone knows something. In *Proceedings of WWW 2008*.

[22] Cheliotis, G., and Yew, J. 2009. An analysis of the social structure of remix culture. In *Proceedings of the 4th international conference on Communities and technologies*, 165–174. University Park, PA, USA.

[23] Hamasaki, M., Takeda, H., Hope, T. and Nishimura, T. 2009. Network analysis of an emergent massively collaborative creation community. In *Proceedings of the Third International ICWSM Conference,* 222-225, San Jose, CA.

[24] Wernicke, S. and Rasche, F. 2006. FANMOD: a tool for fast network motif detection. *Bioinformatics,* 22(9), 1152-1153.