

First Steps Towards a Social Learning Analytics for Online Communities of Practice for Educators

Darren Cambridge

American Institutes for Research
1000 Thomas Jefferson St NW
Washington, DC, 20007, USA
+1-202-403-6924

dcambridge@air.org

Kathleen Perez-Lopez

American Institutes for Research
1000 Thomas Jefferson St NW
Washington, DC, 20007, USA
+1-202-403-5651

kperez-lopez@air.org

ABSTRACT

Learning analytics has the potential to provide actionable insights for managers of online communities of practice. Because the purposes of such communities and the patterns of activity that might further them are diverse, a wider range of methods may be needed than in formal educational settings. This paper describes the proposed learning analytics approach of the U.S. Department of Education's Connected Online Communities of Practice project, and presents preliminary applications of social network analysis to the National Science Teachers Association Learning Center as an illustration.

Categories and Subject Descriptors

J.1 [Administrative Data Processing] *Education*; K.3.1 [Computer Uses in Education] *Collaborative learning, Computer-assisted instruction (CAI), Computer-managed instruction (CMI), Distance learning*

General Terms

Measurement, Human Factors

Keywords

Learning Analytics, Online Communities of Practice, Professional Development, Social Network Analysis, Education, Paradata

1. INTRODUCTION

This paper provides an overview of research and development work being conducted by the American Institutes for Research through the Connected Online Communities of Practice (COCP) project, under contract from the United States Department of Education's Office of Educational Technology. In accordance with the National Educational Technology Plan, COCP intends to help educators at the school, district, and state level across the U.S. become more connected with resources and with each other in order to enhance their professional effectiveness [13].

Online communities of practice (OCoPs) are a key means for helping educators connect. The project's initial environment scan yielded considerable evidence that online communities of practice are becoming increasingly prevalent and effective means for professional learning in education [9]. The research synthesized in our initial report has shown that OCoPs can help educators access, share, and create knowledge and develop professional identity in ways that go beyond what is possible through face-to-face engagement alone.

This research also shows that effective leadership and moderation is key to the success of OCoPs in supporting these activities [1, 6, 7]. In small OCoPs, a community manager may be able to read all of the member-contributed content and discussion and come to

know most or all of the participants. Once an OCoP reaches a certain scale, however, this coverage becomes impossible. Division of labor is one approach to being responsive to the emerging dynamics of community activity and relationships, but it is also helpful for the manager to have a systematically generated, holistic picture of what is going on. Learning analytics may help provide that picture, drawing on the considerable volume of "data exhaust" generated by online community activity. Beyond basic Web analytics, these data are largely an untapped resource for practitioners.

COCP is eager to explore ways in which pioneering work in the field can be applied to OCoPs for educators. However, most learning analytics work of which we are aware has so far focused on structured learning experiences for students, such as semester-long online courses or discussions on a blog. Learning within OCoPs likely differs from learning in these contexts in several ways. First, the learning experience is not time bounded. Second, participation is usually voluntary, with individual participants coming to the community with different needs that correspond with a range of styles of engagement [10]. Third, the primary motivation for engaging with OCoPs is often solving problems of practice rather than learning for its own sake. Finally, experience has shown that the value OCoPs offer is multi-dimensional, that causal relationships between different types of value are challenging to establish, and that understanding of the success of the shared enterprise may change over time [15]. All these differences suggest that it may be more difficult to determine which outcomes of collective activity are valuable and what patterns of activity are more significant in OCoPs than in formal educational settings.

Learning analytics for OCoPs may, therefore, need to be more exploratory than for formal educational experiences. Learning analytics should help community managers see a range of potentially notable patterns rather than simply tracking progress towards pre-defined indicators of success. Precisely because the range of potentially valuable patterns of content and activity are so diverse, learning analytics is likely to be powerful in enabling community managers and moderators to invest their expert interpretive attention more efficiently.

2. PLANNED APPROACH

Making visible the wide range of patterns of activity that are potentially actionable likely requires multiple methods. Initially, we plan to utilize learning analytics in two of the five categories proposed by Buckingham Shum and Ferguson [3], social learning network analysis and social learning context analysis. Social network analysis is our chosen method in the first category, and we are employing pre-hypothesis narrative analysis in the second.

In future work, we hope to also explore content and discourse analysis.

2.1 Social Network Analysis

Our work with social network analysis (SNA) is furthest along, and an example of its application is presented in the next section. SNA has been used in previous research on OCoPs in education, but primarily to analyze discrete discussions or “friend” networks [2, 8]. In contrast, we are using SNA to explore as wide a range of relationships between community participants and content objects as is possible for a given community, beginning from whole-network maps such as those presented in the next section and only narrowing the scope of the analysis when potentially significant patterns or individuals have been identified.

By representing OCoP platform usage data as a unimodal network of relationships between individual community members—the approach common in learning analytics applications of SNA, such as Social Networks Adapting Pedagogical Practice (SNAPP) [5]—we are seeking to identify community members who are particularly significant to the health of the community, either because they are highly influential, as signified by metrics such as eigenvector centrality, or because they frequently connect subgroups, as signaled by high betweenness centrality. These individuals may merit additional support or recognition from the community managers.

In addition, we are visualizing the data as bimodal social network diagrams that explore the relationships between people and content objects, as illustrated in the next section. This alternative representation allows us to see patterns in subgroup activity that might not otherwise be detectable. For selected individuals identified as significant in either representation, we will create egocentric maps of their usage. If permission can be obtained, we will also create egocentric maps of their usage of Facebook and Twitter to obtain a fuller picture of how they connect with other professionals online. Patterns in this usage may suggest strategies individuals can use to increase the impact of their participation, as well as criteria for identifying potentially effective participants to recruit into the OCoP.

2.2 Pre-Hypothesis Narrative Analysis

Analysis of usage data can yield a rich representation of the dynamic of online activity, but understanding the impact of that activity on offline professional practice can be furthered through the collection of self-report data. To collect and analyze self-report data, we are employing Snowden’s [12] narrative analysis approach to identify patterns in the context of OCoP members’ experiences. In this method, narrative fragments—brief stories—are collected online using the CognitiveEdge SenseMaker Collector software, which also asks the respondent a series of closed survey questions about the context of the narrative, called filters. Community members recording and classifying narratives are given unique identifiers, making it possible to add their centrality metrics from SNA.

Snowden calls this approach “pre-hypothesis narrative” research because it is designed to help analysts identify emergent patterns—“weak signals” in the terminology of complex systems theory—that might be missed if the impact of their cognitive biases—in this case, the beliefs they hold about what aspects of community context are significant enough to attend to—are not minimized. Techniques for minimizing bias include using prompting questions that encourage both strongly positive and negative stories, developing filter questions that mask the desired

outcomes, and soliciting stories from a large, diverse group of respondents.

Most significantly, analysts begin analysis by using the SenseMaker Explorer’s powerful visualization capabilities to look for patterns in the quantitative filter question data (including, in our case, SNA metrics) prior to analyzing the content of the narratives themselves. This abstraction helps to reduce the influence of the analysts’ interpretive predilections. Once a pattern is identified, researchers can drill down into the content of the narratives associated with it, which can become the subject of content analysis. This method of selecting samples of stories to analyze has the added benefit of making narrative analysis more time efficient as a formative evaluation strategy.

2.3 Content and Discourse Analysis

Eventually, we also hope to systematically analyze the content of members’ contributions to the communities, such as discussion posts and blog entries, as well as the narratives collected from them. We are particularly interested in extracting significant semantic concepts using automated tools such as Open Calais. Such automated analysis has the potential to help community managers discover emerging topics of interest to the community that could be incorporated into its editorial calendar.

We are also evaluating tools for discourse analysis. These tools have the potential to help community managers understand what styles of discourse are most often associated with which community outcomes, enabling them to encourage the style most likely to help the community achieve its purpose.

3. CURRENT WORK

At present, we have begun SNA work in three OCoPs. Here, we present some very preliminary findings on the National Science Teachers Association Learning Center (NSTA LC) to illustrate our direction in working with these communities. The NSTA LC, launched in April 2008 through the efforts of Dr. Al Byers and his colleagues, aims to support science teachers in increasing their knowledge of science and of pedagogy [4]. It provides a rich source of (mostly free) learning materials and experiences for science teachers. The NSTA LC also hosts an online community through its “Community Forums.” NSTA members can initiate topics within any of a number of forums, or post to existing topics. They can share and comment upon collections of material composed of their own or NSTA resources.

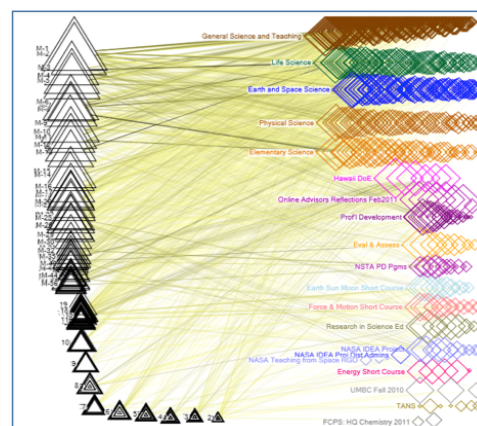


Figure 1: One year of posts made by 492 NSTA members to 557 topics within 21 forums.

We received NSTA LC forum posts for the full year, from 9/24/2010 to 9/28/2011 24 September 2010 to 28 September 2011. As a preliminary analysis of this forum activity, we used NodeXL [11], an open-source template for Microsoft® Excel®, to create bimodal network diagrams of the 6,978 posts made by 492 members to 557 topics within 21 forums during that time. Figure 1 is a low-resolution depiction of the patterns of these posts as edges between member nodes (black triangles along the left) and topic nodes (diamonds colored according to their forums along the right). Member nodes are sized according to the number of different topics to which they posted; topic nodes are sized according to the number of posts made to them. Topic nodes are grouped by forum; within their forum group they are placed left to right by number of posts made to them. The member nodes and the forum groups are positioned top to bottom by their total number of posts. The 59 large nodes of heavy-posting members—those with 20 or more posts, with a maximum of 582—take up almost two-thirds of the vertical space, while the other 433 are squeezed and stacked along the lower one-third and across the bottom of the figure. Edge color (black to yellow) and opacity are logarithmically proportional to the number of times a member posted to the topic. A frequency analysis indicates greatly skewed distributions within the post data: a few forums received most of the posts; a few members initiated most of the topics and made most of the posts; and a few topics received half of the posts. The figure captures this, with dense dark edges in the upper portion of the network, between the relatively few members and topics.

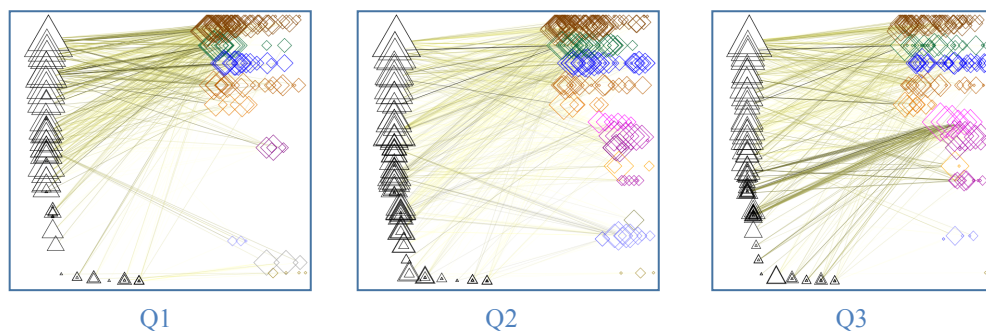
However, the distribution of edges is not smooth from the upper portions of the network to the lower, and there appear to be a large number of edges mid-figure. There might be some interesting activity by members in those regions, but with this static view, it is difficult to see what that could be.

To tease out this information, we separated the data into 5 contiguous periods, Q1–Q5, each containing one-fifth of the

posts, and created network diagrams for each period. These diagrams are shown in Figure 2. Nodes are placed exactly as in Figure 1 (i.e., according to total annual number of posts), but their sizes are relative to the number of posts made during the quintile. Likewise, edge color and opacity are relative to the data in the quintile.

During the initial period, Q1, the activity is mostly by a very few active members, and there is very little activity in the lower part of the figure. During Q2 there are many new members, but their posting activity is fairly light. In Q3 something interesting develops: very heavy posting to the forum named Hawaii DoE (pink, mid-diagram), mostly by moderately active posters, but also from a number of new members, along the bottom of the diagram. The Hawaii DoE posts disappear in Q4, but the mid-active members keep somewhat active during this period. By Q5, they are posting quite heavily, and now to the more “standard” forums of Life Science, Earth and Space Science, and Physical Science. We need to examine these data further, but it is possible that this time series of network views has highlighted something that could prove useful to community managers. It might show how time-bounded activity targeted at some subgroup could be leveraged into more sustained and general engagement.

The next steps in our analysis will include using the topic initiator information to create initiator-topic and member-initiator networks; transforming the bimodal data to create unimodal, member-member diagrams; obtaining additional member information, such as which online seminars they are attending, who is a member of a “cohort” (a district-wide NSTA LC professional development plan), and the points and badges they have attained for their activity in the Learning Center. We will also examine the social network analysis metrics produced by NodeXL, such as the centrality measures discussed in the previous section.



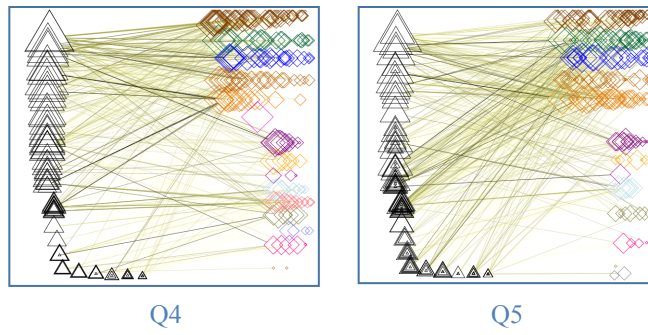


Figure 2. Network diagrams for five quintiles. Nodes are placed exactly as in Figure 1 (i.e., according to total annual number of posts), but their sizes are relative to the number of posts made during the quintile period. Likewise, edge color and opacity are relative to the data in the quintile.