

Simple and Computational Heuristics for Forum Management in the NSTA Learning Center: A Role for Learning Analytics in Online Communities of Practice Supporting Teacher Learning

Kathleen Pérez-López
American Institutes for Research
kperez-lopez@air.org

Darren Cambridge
American Institutes for Research
dcambridge@air.org

Al Byers
National Science Teachers Assoc.
abyers@nsta.org

Abstract

Prior research has demonstrated that the quality of moderation and management in online communities of practice is key to their successful support of learning. However, as communities grow in size and complexity, it becomes increasingly difficult for unaided experts to fully understand and take action in response to the activity of participants within them. Learning analytics has the potential to provide the support that community of practice leaders need to improve their performance. The National Science Teachers Association and U.S. Department of Education's Connected Educators project are exploring three approaches for managing forums to make them accessible and to synthesize the knowledge they generate: archiving, summarizing, and reorganizing. This paper describes manual heuristics for the first two of these, as well as the use of social network analysis to help develop algorithms to automate the third, community forum reorganization.

1. Introduction

Prior research has clearly demonstrated that the quality of moderation and management in online communities of practice is key to their successful support of learning [2,11,13]. However, as communities grow in size and complexity, it becomes increasingly difficult for unaided experts to fully understand and take action in response to the activity of participants within them. Learning analytics has the potential to provide the support that community of practice leaders need to improve their performance.

One way of defining learning analytics is as tools and techniques for analyzing data to support decisions made in the course of supporting or engaging in learning. As others have pointed out, this

contrasts with approaches to using data to support learning that attempt to replicate or replace human expert judgment with algorithms [32]. The National Science Teachers Association (NSTA) and the U.S. Department of Education-funded Connected Educators project are working together to develop a social learning analytics approach that combines the strengths of community moderators' expertise with analytical support from computational tools.

Cultivating online communities of practice can be seen as a subset of networked learning, one means to be used in conjunction with a range of others to build professional social learning capacity [27,18,20]. In contrast to recent trends in informal social learning theory, and social relations more generally, the communities of practice approach retains an emphasis on the defined collective as well as the connected individual. Healthy communities of practice are committed to supporting the learning of both newcomers and well-established community members. They enable newcomers to engage in "peripheral participation" alongside more experienced members working together to share and solve problems in their professional practice [5,23]. Part of what enables this support for the learning of members with varying levels of expertise is that the community develops a shared purpose and common language and that members identify with them.

This emphasis on both the individual and collective is important because it reflects both the concrete needs of the teachers of science it serves and the health of the profession. Beyond their preservice training, science teachers in the United States need to learn continuously and broadly throughout their careers. Many teachers, especially those in grades K-6, have little formal training in the science subjects they are charged to teach [7,34], while many at the 6-8 level have general science education degrees and may be required to teach "out of field" [17]. For example, at the middle-school level a teacher who specializes in biology may be asked to take on earth

science. The Next Generation Science Standards that will soon be implemented in most states have a topical organization that often similarly crosses traditional boundaries between subject areas [26]. In addition to facing these concrete challenges, science teachers must negotiate a changing body of “pedagogical content knowledge” that reflects the rapid pace of scientific discovery and its growing social importance [1,14,24,30,31].

This body of knowledge is not easily transferred or translated for classroom application and requires discourse among teachers to internalize and facilitate its application. Teachers themselves desire and need to participate in its creation through their practice. Effective professional learning that keeps up with the pace of new knowledge creation and the changing demands of practice depends on teachers co-creating knowledge in community, collectively establishing priorities, and defining what constitutes excellence.

Learning in community also enables teachers to collectively strengthen and chart the future of their profession. In online communities, teachers of science can come together to define excellent practice and to determine what is needed from the various stakeholders—such as schools, districts, states, the federal government, science education programs, and the scientific community—to support and recognize that practice. The development of a shared repertoire of knowledge, tools, and techniques and a shared identity as members of the profession in online communities of practice has the potential to support the health of the profession.

NSTA developed its Learning Center to address both of these needs. The Learning Center, begun in April 2008 with a current participation of over 100,000 members, provides a rich source of learning material and experiences for K12 science teachers. It also hosts an online community through its community forums, where NSTA members can initiate and post to topics within any of a number of forums.

Many teachers first come to the Learning Center to address an immediate challenge. For example, how do I teach students the difference between weather and climate tomorrow morning? By providing quick and easy access to information and to the guidance of experts and more experienced teachers, the Learning Center can address these immediate needs. However, such engagements should also help teachers establish a network of relationships and a coherent sense that the larger community understands and is co-creating the body of knowledge about teaching science [36]. This emergent body of knowledge will support their sustained learning and enable them to contribute to the profession’s growth.

Effective forum management may contribute to fulfilling these goals. Forums ought to:

1. Be very accessible to newcomers, allowing rapid access to the information they need, as well encouraging them to contribute from their own experience;
2. Foster the understanding among members of the community of practice’s collected knowledge of a topic to date;
3. Connect members to each other, developing a sense of professional collective identity.

In very small communities, topics or threads within a topic are often short enough that newcomers can read through all the posts. They can then locate specific information, resources, or expertise, and also get a sense of the whole, of who’s who and where they might be able to contribute. In larger communities such as the Learning Center, in the absence of effective forum management, reading all these posts is clearly not practical, even for many very active and experienced community members.

We are exploring three techniques for managing forums to make them accessible, to synthesize the knowledge they generate, and to cultivate relationships: archiving, summarizing, and reorganizing. The first two of these techniques are currently manageable by community moderators and managers using very simple heuristics to guide expert judgment. In the next section we describe how Learning Center leaders are currently employing them.

Reorganization may be the most challenging of the three because it requires not just analyzing a particular topic but looking across the whole of the community’s activity. In Section 3 we describe one approach, which can be classified as a form of social learning network analysis within Ferguson and Buckingham Shum’s [12] social learning analytics taxonomy, that shows promise for guiding community leaders in reorganizing forums. Social network analysis has been used in a variety of ways to examine activity within online communities of practice for educators, ranging from close examination of specific discussion topics [4] to global characterizations of the relationships between people and content across very large communities [33]. Several projects applying social network analysis to education also share our focus on helping those who support learning (or learners themselves) use computational analysis to connect with others and to guide their practice. These projects may be in the context of student online learning [8] or within blended professional learning networks [3].

2. Existing NSTA forum management heuristics

The Learning Center has 22 trained online advisors who both provide live chat during peak usage times and moderate the integrated asynchronous NSTA community forums. These online advisors are science methods professors and experienced educators who are familiar with the digital resources within the Learning Center. The online advisor team has developed a set of heuristics and protocols to help ensure that the forums remain accessible and relevant to the large and growing base of active users and to curate the knowledge that members co-create within them.

A first challenge that these heuristics and protocols address is archiving topics and posts within the forums. As the number of posts increases—there are now over 10,000 posts by Learning Center members who are not online advisors—the number of topics within a forum and the number of posts within a specific topic make the knowledge it contains difficult to access and integrate. For example, as of June 2012, there were over 263 topics and over 2,700 posts for the General Science Teaching forum. Therefore, determining which topics to archive, and possibly to summarize, is imperative.

Literature suggests the flow of dialog within communities often develops consistent rhythms [35, 37]. In developing their protocols, the online advisors analyzed the flow in the Learning Center's five major forums—Chemistry; Earth and Space Science; General Science and Teaching; Life Science; and Physical Science—examining:

- Total number of days threads existed
- Total number of posts to threads
- Total number of views of threads
- Average number of days of activity per thread
- Average number of posts per thread
- Total number of threads per forum

The analysis showed that the average duration of activity for a discussion topic within a particular forum is approximately 45 days. For many threads, activity (number of posts) slows down after this time, perhaps because the posting volume makes browsing unwieldy.

The analysis also demonstrated that the community has many more viewers (i.e., those who read but do not create posts) than post contributors, and that viewing follows a very different pattern from posting. Viewers, who may be newcomers to the topic or to the community as a whole, often read posts well after those actively engaged in the

conversation lose interest and move on to other topics. Well-designed forum management processes are likely to benefit these readers especially, possibly encouraging their more active participation.

One possible approach to archiving and reorganizing forums likely to be valuable to viewers (as opposed to contributors) would be to institute an explicit and/or implicit rating system for posts. The drawback of this solution is that, while it supports reading, it might actually discourage viewers from transitioning to contributors if they received poor ratings on their early contributions [22]. There is some evidence that rating systems encourage enculturation of new members into community norms, but negative feedback can also delegitimize peripheral participation [9]. Providing indicators of the value of contributions can be a motivator for some members, depending on their dispositions and experience, but multiple types of value that are difficult to measure through an automated or user contributed rating system need to be signaled [28].

In light of these reservations about rankings and on the basis of the analysis, the advisors developed an archival approach that involves both quantitative and qualitative analysis. On a regular basis, Learning Center staff and advisors make recommendations for collapsing and moving topics as they become unwieldy in the various forums. The first criterion is the length of time the topic has been active. After 45 days, it is auto-identified and assigned to one of the online advisors for review. The online advisor manually reviews the postings to the topic and the frequency of the most recent posts, determining if the discussion is active or waning. If the topic is active no action is taken, but if the conversation has lulled and there are more than 20 posts, the online advisor may generate a summary for it and recommend archiving the discussion. The online advisor also responds to any outstanding queries in the topic, suggesting digital learning resources, external URLs, or redirects to other discussion topics, content, or users within the forums. This helps build the relationships and social norms so crucial to a vibrant community.

By providing a way to archive and summarize, as well as a link to “reopen” a topic for additional discussion, the Learning Center is attempting to cull the knowledge of the community while still allowing it to grow as participants generate new queries and views.

3. Computational heuristics for forum reorganization

3.1 Sharpening forum focus while maintaining community connections

As just discussed, the Learning Center should maintain vibrant, accessible, and coherent forums comprised of member posts to topics within them. At the same time, it does not want to create isolated sub-sets of members who are only interested in a particular forum. Research being conducted through the Connected Educators project on a broader set of online communities for educators suggests that this is a common challenge. We have been analyzing Learning Center data to determine algorithms that might be used by forum advisors to help automate their tasks. Figure 1 portrays a network diagram of 6,792 Learning Center posts made by 307 members (triangles on the left), including online advisors (red triangles) to 556 topics (diamonds on the right) within 20 forums (rows of diamonds), from 9/24/2010 to 9/28/2011. Triangle size is proportional to number of topics the member posted to, and diamond size is proportional to the number of members posting to the topic. Forums are ordered from top to bottom by the total number of posts made to them; note that four of the five largest forums were analyzed by the online advisors. Posts are represented by edges, whose darkness and opacity represent the number of posts from a member to a topic. Forums labeled PrvF<n> are private forums, often focused on professional development courses. (All network diagrams here were produced using NodeXL [16].)

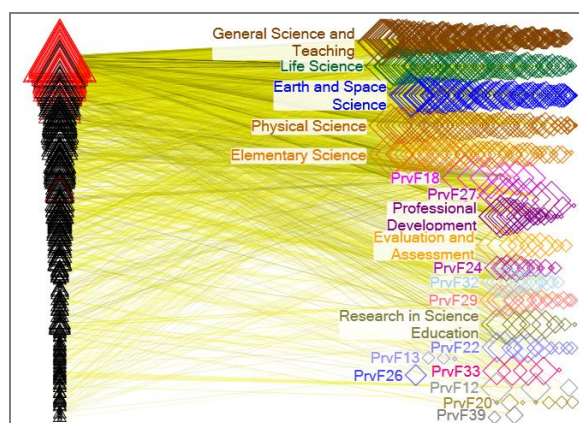


Figure 1. One year of Learning Center posts from members on left to topics, within forums, on right.

As the posting volume to forums increases, the content can become more diverse, making it difficult

for users to quickly find the information they need. This may require refocusing content by splitting off new forums, archiving inactive topics, or possibly resorting topics, to achieve forums that are comprehensive yet specific. That is, posts should be partitioned into tightly grouped forums that contain most of the information a member would be seeking on a subject, while excluding information that would be better placed in a different forum. However, since an overarching purpose of the Learning Center is to build and maintain a community, we want to constrain the forum sharpening so that it does not greatly reduce the mix of forums that members visit and interact within.

We begin by seeking two relationships, one between topics and one between Learning Center members, that would achieve a balance between two conflicting goals. The first goal is for the topic relationship to link pairs of topics that members

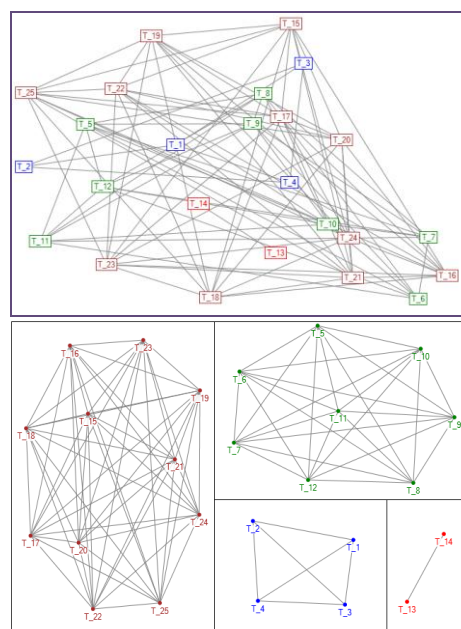


Figure 2. Hypothetical network of related topics: top, ungrouped; bottom, grouped. Given the relationships between these pairs of topics, there are no inter-group links, which is an ideal outcome.

would consider to have similar content, leaving any topics with unrelated content unlinked. Grouping on this relationship should then configure topics into well-structured forums. Once these new forums are created, we consider the social network of members created by defining a relationship that links two members if they post messages to topics within the same forums. That is, if members A and B post to at least one common forum, they would be linked in the

network. The second goal, then, is for the topic relationship to have been created in a way that this member relationship *not* segregate members into tight groups, but keep as many connections as possible between members of any subset.

The topic relationship determines a network graph; the nodes represent topics and an edge exists between any related topic. Figure 2, top, shows a network graph of 24 topics for a hypothetical topic relationship that was constructed to perfectly satisfy the requirement of producing tight topic groups. The bottom of the figure shows this network clustered into four groups, which would become the new forums. In this ideal clustering, there are no links between groups, only within them. (These groups were created using the Clauset-Newman-Moore (CNM) [6] algorithm in NodeXL [16].) Such an ideal relationship would allow members to go to a forum of interest and find all the topics related to that interest, and not have to weed through any unrelated material.

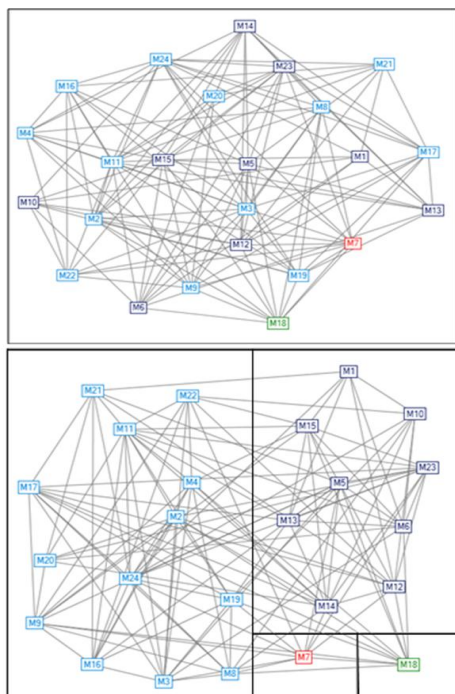


Figure 3. Hypothetical network of Learning Center members; top: ungrouped; bottom: grouped, with a desirable high number of inter-group edges.

The member relationship defines a network graph as well, and the constraint on the forum reorganization would be satisfied if this network looked something like the one depicted in Figure 2, created for illustrative purposes. The top image in this figure shows an ungrouped network of members

who are linked if they posted to the same forum; the bottom shows member groups derived from the CNM algorithm. We see that the grouping is actually quite weak and the network maintains a lot of interaction among the groups, a desirable outcome.

Summarizing our goals, we want to reorganize the topics into a new forum configuration that approximates the ideal network displayed in Figure 2, while maintaining cross-forum member interactions, represented by the network displayed in Figure 3.

3.2 Approach and implementation

In our initial attempt at creating restructured forums for the Learning Center that would satisfy the goals just described, we followed the steps outlined

1. Define a relationship between topics and create a network T of topics linked by it.
2. Choose a clustering algorithm for T and cluster the topics into a candidate forum configuration. Are the groups well formed, i.e., do they segregate topics into focused groups with few links between groups? If so, skip to Step 4; if not, continue to Step 3.
3. Are there other clustering algorithms to consider?
 - a. If yes, repeat Step 2.
 - b. If not, but there is another feasible choice for the topic relationship, continue at Step 1.
 - c. If there are no other feasible choices for the topic relationship, **the process has ended, unsuccessfully.**
4. Provisionally accept the derived topic clusters as the new candidate forums and check whether the members are overly segregated in the candidate forums:
 - a. From the bimodal network of members posting to *forums*, create a unimodal network M of members, where links are defined as members mutually posting to the same forum. Check that the density of M is not much lower than that of M_0 , the member network created from the original forum configuration.
 - b. Cluster the members of M and check that there are many links between clusters. If M is not overly segregated, continue with Step 5; otherwise, continue at Step 3.
5. Are the new forums a refinement of the previous version:
 - a. Ideas within each of the groups are (more) closely related
 - b. New structure is as fine as (or finer than) the original one
 If yes, accept it as the revised forum organization. **The process has ended, successfully.**

Figure 4. Steps to forum reorganization.

in Figure 4. For Step 1, using the Learning Center data of members posting to topics depicted in Figure 1, we created the relationship between topics by linking two topics if any member posted to both of them.

That is, we used the social network analysis technique of creating a unimodal topic network from a bimodal member-topic network. A bimodal network has two types of nodes, with edges only between nodes of one type and nodes of the other type, never between nodes of the same type. The network of Learning Center postings is bimodal, with member nodes and topic nodes. A unimodal network can be derived for each node type. In such networks, an edge exists between two nodes of one type if they are both linked to the same node—of the other type—in the bimodal network. We used the social network analysis program Pajek [10] to construct such a unimodal topic network, T , from the Learning Center posts.

Following Steps 2 and 3a, we clustered the topics in T by applying different algorithms available in Pajek [10] and NodeXL [16] to create F_1 , F_2 , F_3 , and F_4 candidate forum configurations (i.e., each configuration is an assignment of each topic to one and only one forum). The first attempt used the CNM algorithm; the second applied the Wakita-Tsurumi algorithm, an optimized version of CNM; and the third involved m-slices and k-cores (see [10] for definitions of these techniques used in social network analysis). None of these attempts worked very well, as the clustering algorithms could not separate the topics. Typically these algorithms assigned topics to only a few groups, with two or three very dominant groups containing most of the topics. Even between these large groups there was very little separation, as many topics were linked across them. Compared to the original forum arrangement, any of these new topic configurations would make it much harder for members to find the content they need.

We believe the clustering failed for two related reasons: we defined the topic relationship too inclusively—more topics were considered to be related than should be and posts may need to be filtered for content before using them to define a link between topics. Our definition for the relationship between topic nodes resulted in a topic network with density (the ratio of actual links to all possible links) 0.49: on average, nodes were connected to almost half of all the other nodes.

On our fourth iteration of Steps 2 and 3a, we achieved a more suitable candidate forum configuration by excluding posts made by the online advisors. These are red nodes in Figure 1, whose generally large sizes indicate posting to many topics,

which they were tasked to do. Given our definition of the topic relationship, their posts would artificially link many topics. Omitting these posts reduced the number of topics from 556 to 474 and drastically reduced the density of the topic network to 0.10. Grouping this reduced topic network using Wakita-Tsurumi clustering yielded F_4 , a much better candidate than the previous attempts, F_1 , F_2 , or F_3 . F_4 separated topics into 16 groups more finely than the others were able to do, and while the largest group contained 138 topics, or 29% of all topics, none of the groups totally dominated all the others as in the previous attempts.

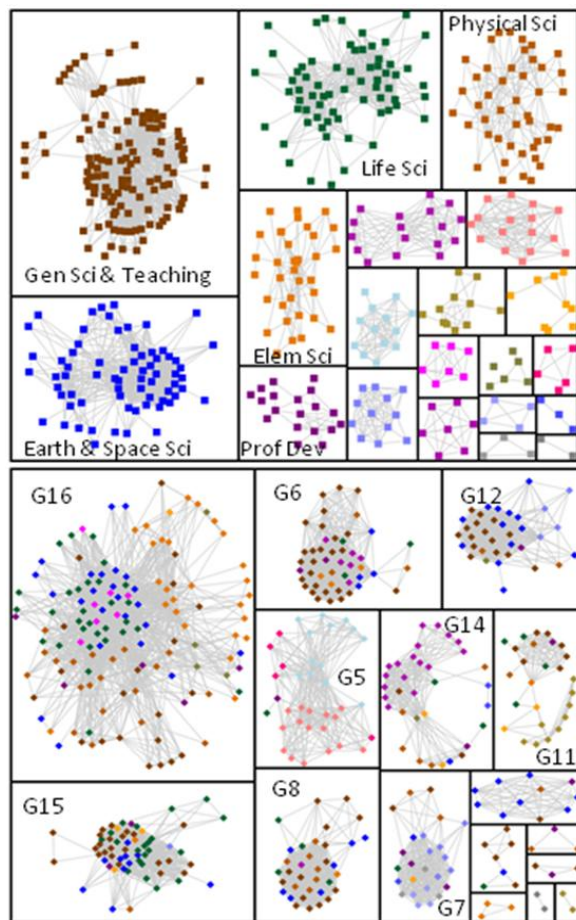


Figure 5 Top: topics in the 20 original Learning Center forums; bottom: the topic network partitioned into F_4 using the Wakita-Tsurumi algorithm, inter-forum edges not displayed.

Figure visually compares the 16 topic groups, G1 to G16, defined by F_4 (bottom) with those of F_0 , the original Learning Center forums (top). Both top and bottom images label the larger groups; color the topics as in the Learning Center forum diagram in

Figure 1; and omit the edges between groups in the configuration, as the inter-group edges are dense and obscure the group structures in both cases. All but the two smallest of the new topic groups in F_4 draw from a number of the original Learning Center forums, but each group comprises different mixtures of them. For instance, over half of G_6 's topics draw from the General Science and Teaching forum, while G_5 draws 80% of its topics from two private forums concerned with professional development.

One of the goals of reconfiguring forums is to refine them, particularly the largest ones, so that content is more focused and accessible. F_4 does not do that: it has four fewer groups than the original 20 Learning Center forums, and the size distribution of the groups—the number of topics within them—is very close to that of forums, as can be seen by comparing the blue and red lines in Figure 6, which shows the distribution of the number of topic nodes within the original forums and two new groupings. Each of F_4 and F_0 has one very large group containing 29% and 25% of the topics, respectively, and the three largest groups contain 53% and 55% of the topics, respectively. In other words, F_4 does not flatten the size distribution of the topic groups.

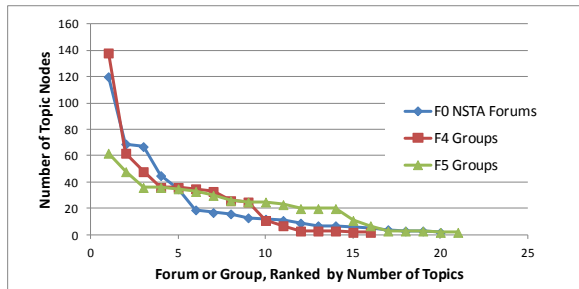


Figure 6. Distribution of topic nodes, F_0 , F_4 , and F_5 .

However, since F_4 's G_{16} draws from nine of the F_0 forums, with five of them accounting for 85% of its topics, a reasonable approach would be to split G_{16} according to these forums.

Figure 7 shows F_5 , a refinement of F_4 created from such a decomposition of G_{16} . From the figure and the green line in Figure 6 corresponding to F_5 , we see that the distribution of group sizes has flattened out.

Table 1 further compares these three configurations; the shaded numbers are the more desirable values in terms of the overarching goals. The top portion of the table contains basic information about the entire graph, in particular, that of the 110,215 possible edges, only 10,701, or about 10% of which, exist. This lower density—compared to the topic network containing online advisor posts—is an improvement because a highly dense network of topics is more difficult to cluster.

The lower portion of the table shows that F_4 is the closest approximation to the ideal clusters of Figure 1, since it has the highest concentration of edges within groups rather than between groups: 47% compared to 29% for F_0 and 38% for F_5 . However, part of this difference derives from the fact that the F_4 configuration is coarser than the other two, with only 16 groups, so its groups are the largest, averaging 29 topic nodes compared to 24 for F_0 and 22 for F_5 .

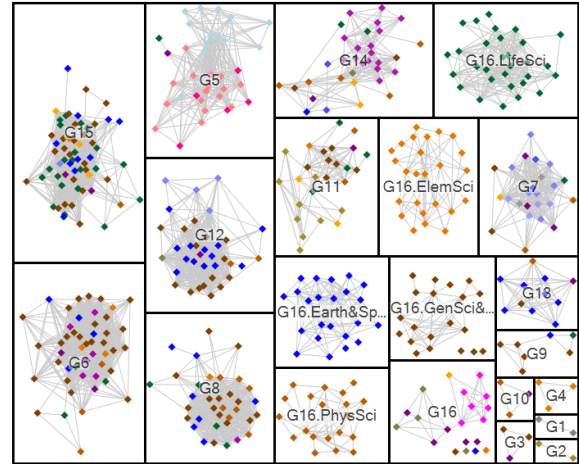


Figure 7. F_5 , created from splitting G_{16} , the largest group of F_4 , into its F_0 components.

With respect to most other considerations, including the most even size distribution, F_5 is superior to both F_4 and F_0 . F_5 groups are the densest, with an average weighted density of 0.51 across groups, while this value is 0.32 and 0.46 for F_0 and F_4 , respectively, where the weighting was by the number of nodes in each group or forum. This indicates that the topics in F_5 are more closely related to each other than for the other two configurations. Finally, with 21 groups compared to the original 20 forums, F_5 is no coarser than F_0 .

Table 1. Comparison of original Learning Center forums, F_0 , and configurations, F_4 and F_5 .

Entire Topic Graph			
Number of nodes	470		
Number of edges	10701		
Number of possible edges	110215		
Graph density	0.10		
Topic Forums or Groups	F_0	F_4	F_5
Number of groups	20	16	21
Percent of edges within groups	29%	47%	38%
Average number of topic nodes per group	24	29	22
Weighted average density over groups	0.32	0.46	0.51

Until we complete the analysis of the content of the posts to the topics, we will not know if this

redistribution of topics among the F_5 groups is more meaningful than the original forums and F_4 . In the meantime, however, proceeding to Step 4 of Figure 4, we provisionally accepted F_5 and checked the constraint that this grouping does not overly segregate Learning Center members.

We constructed the unimodal member network M_5 from the bimodal network of members posting to the groups in F_5 . That is, there would be an edge between Learning Center members if they mutually posted to any common group of F_5 . We then applied a clustering algorithm to the member network M_5 and checked whether it *did* have a lot of cross-group edges, i.e., whether its network diagram appeared similar to the one on the bottom in Figure 3.

We actually performed Step 4 for F_4 as well, constructing M_4 . Table 2 compares M_4 and M_5 to M_0 , the member network from the original Learning Center forums, restricted in the same way by omitting posts from the online advisors. (M_4 and M_5 have fewer nodes than M_0 because some members posted only to isolated topics that did not get assigned to any groups in F_4 or F_5 , while every topic is assigned to a forum in F_0 .)

Table 2. Comparison of member networks, M_0 , M_4 , and M_5 , derived from original Learning Center forums, F_0 , and topic reconfigurations F_4 and F_5 .

Entire Member Graph	M_0	M_4	M_5
Number of nodes	281	278	272
Number of edges	16965	17675	12585
Number of possible edges	39340	38503	36856
Graph density	0.43	0.46	0.34
Member Groups	M_0	M_4	M_5
Number of groups	11	7	8
Percent of edges within groups	33%	43%	44%
Average number of nodes per group	26	40	34
Weighted average density over groups	0.85	0.85	0.80

The density of M_4 is in fact higher than that of M_0 , 0.46 compared to 0.43, but the decomposition of G16 for M_5 reduces links between members, resulting in the less desirable lower density of 0.34 for M_5 . While the percent of all edges that are between groups declined from 67% in M_0 to 57% and 56% in M_4 and M_5 , this is likely due to the increased average size of groups, from 26 to 40 and 34 members. The average weighted density across groups (again, weighted by number of topics in the forums or groups) is essentially unchanged from M_0 to M_4 , at 0.85, but drops to 0.80 for M_5 .

In summary, F_5 is a more focused configuration of the Learning Center topics, with more evenly sized groups than the original forums, F_0 . While it does reduce the density of the member network by nine

percentage points, and decrease the percentage of edges between groups by 11 points, the new values do not seem so low that they would significantly discourage member interaction. The ultimate value of this revised configuration depends on whether it focuses the discussions.

4. Conclusions and future work

We are encouraged by this initial implementation of our approach to automating the reorganization of Learning Center forums using network analysis techniques. Judging from the improved results from our fourth attempt, which restricted the analysis to posts by ordinary members, we understand the need to carefully refine the input data. In our fifth attempt, incorporating the information contained in the original configurations resulted in more evenly sized and focused topic groups.

We hope that an automated version of these analysis procedures and visualizations will eventually become part of a community management dashboard that Learning Center staff can use monthly alongside their existing archiving and summarization procedures. Within this dashboard, managers and advisors would be able to drill down into posts within each cluster to examine their content and discover emerging issues and themes that might inform substantive programming choices, as well as increase the accessibility, coherence, and relationship-building capability of the forums. However, we will need to do some further refinement to the analysis techniques before investment in developing this functionality is justified.

Our subsequent work will involve (1) filtering posts by member attributes and post content; (2) applying the approach to post viewing as well as post authoring; and (3) considering temporal aspects of data.

The improved results achieved after refining the dataset of posts supports the first point. In addition, Rodríguez, et al., who take a very similar approach to ours by creating a one-mode topic network from a learner-topic network [29], get promising results by prefiltering more than we do. We have access to some member attributes on which to separate posts for more coherent analyses. Our current approach involves an implicit tension: we associate topics into forums based on the mutual postings by members, constrained to minimize the resulting segregation of the members posting to these reorganized forums. A better approach might be to associate topics based on their content, still preferring solutions that minimize member segregation. To do this, we will explore the

many tools available for automatic text analysis [15] to “denoise” the posts, separating out friendly banter and enabling linking posts by the relevant terms they contain.

In their analysis of activity in the Learning Center forums, the online advisors found a much greater volume of views than posts; viewing patterns were also very different from posting patterns in that viewers used material over longer periods. The Learning Center recognizes the importance of careful archival procedures; as post volume increases, automatic aids for this task resulting from our analyses will be very useful. The concerns of the current work also apply to the archived material: it will be important for archived topics to be well categorized, but not to segregate viewers. Viewed topics manually archived by online advisors will provide a comparison for our automated algorithm development.

The online advisors also found that there a strong temporal aspect to posts, and so we are analyzing the current data set plus additional posting data through June 2012 on a monthly basis. In particular, the Learning Center members are K12 educators and the rhythm of the school year, its vacation and testing schedules, may point to topic patterns and distinct analysis periods. Given that posts to a topic peak and then wane over an average 45-day lifespan, it might be best to assign posts to the time period when the peak occurred, as determined by burst detection [19,25]. Krause, et al. [21] combines data classification with intensity tracking, obtaining improved results for both; we will consider incorporating their approach.

Another important aspect of our future work will be to evaluate the extent to which the recommendations for reorganization provided by these techniques prove useful to the NSTA Learning Center managers and moderators in making authentic forum management decisions in support of their members’ individual and collective learning, and the degree to which members find value in the improvements they implement. Interview research underway using the value creation story framework for evaluating online communities of practice and informal learning networks [36] may provide some evidence to help make this determination.

5. Acknowledgments

Thanks to the reviewers of this paper, whose very thoughtful comments significantly improved it. Special appreciation goes to online advisors Susan German, Arlene Jurewicz Leighton, Patty Rourke,

and Maureen Stover who have worked together to identify the criteria and processes for archiving the NSTA Learning Center discussion forums; to Patty Rourke for compiling the forum data they used; and to Rachel Crossno, who helped us with careful document reviews. Finally, a special thanks to Flavio Mendez and Don Boonstra who manage the community forums and the online advisors at NSTA.

6. References

- [1] Abell, S. K., “Research on science teacher knowledge”, in S. K. Abell and N. G. Lederman (Eds.), *Handbook of research on science education*, Lawrence Erlbaum Associates Mahwah, NJ, 2007, pp. 1105-1149.
- [2] Babinski, L. M., B. D. Jones, and M. H. DeWert, “The roles of facilitators and peers in an online support community for first-year teachers”, *Journal of Educational Psychological Consultation*, Lawrence Erlbaum Associates Inc., Mahwah, NJ, 2001, pp. 151-169.
- [3] Bieke, S., and M. De Laat, “Network awareness tool - learning analytics in the workplace: Detecting and analyzing informal workplace learning. Multi-mediated community structure in a socio-technical network”, *Proceedings of LAK12*, May 2012, ACM, Vancouver, BC, 2012.
- [4] Bonsignore, E., D. Hansen, A. Galyardt, et al., “The power of social networking for professional development”, in T. Gray, T. and H. Silver-Pacuilla (Eds.), *Breakthrough Teaching and Learning: How Educational Assistive Technologies are Driving Innovation*, Springer, New York, NY, 2010.
- [5] Brown, J. S., A. Collins, and P. Duguid, “Situated cognition and the culture of learning”, *Educational Researcher*, SAGE Publications, Thousand Oaks, CA, Jan-Feb 1989, pp. 32-42.
- [6] Clauset, A., M. E. J. Newman, and C. Moore, “Finding community structure in very large networks”, *Physical Review E*, American Physical Society, College Park, MD, December 2004.
- [7] Council of Chief State School Officers, *State indicators of science and mathematics education: 2007*, Washington, DC, 2007.
- [8] Dawson, S., “‘Seeing’ the learning community: an exploration of the development of a resource for monitoring online student networking”, *British Journal of Educational Technology*, Wiley-Blackwell, Hoboken, NJ, September 2010, pp. 736-752.
- [9] Dellarocas, C., M. Fan, and C. Wood, “Self-interest, reciprocity, and participation in online reputation systems”, MIT Sloan Working Paper, MIT Center for Digital Business, Cambridge, MA, February 2004, Paper 205.
- [10] deNooy, W., A. Mrvar, and V. Batagelj, *Exploratory Social Network Analysis with Pajek*, Cambridge University Press, Cambridge, UK, 2005.
- [11] U. Farooq, P. Schank, A. Harris, J. Fusco, and M. Schlager, “Sustaining a community computing infrastructure for online teacher professional development: a case study of designing Tapped In”, *Computer Supported*

Cooperative Work, Kluwer Academic Publishers, Norwell, MA, 2007, pp. 397-429.

[12] Ferguson, R., and S. Buckingham Shum, "Social learning analytics: five approaches", LAK12 Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, ACM, New York, NY, May 2012, pp. 23-33.

[13] Gareis, C. R., and S. Nussbaum-Beach, (2007), "Electronically mentoring to develop accomplished professional teachers", *Journal of Personnel Evaluation in Education*, Springer, New York, NY, Dec 2007, pp. 227-246.

[14] Gess-Newsome, J., and N. G. Lederman, (Eds.), *Examining Pedagogical Content Knowledge: The Construct and its Implication for Science Education*, Kluwer Academic Publishers, Boston, MA, 1999.

[15] <http://dirt.projectbamboo.org/>.

[16] <http://nodexl.codeplex.com/>.

[17] Ingersoll, R., "The problem of underqualified teachers in American secondary schools", *Educational Researcher*, SAGE Publications, Thousand Oaks, CA, March 1999, pp. 26-37.

[18] Katz, S., and L. Earl, "Learning about networked learning communities", *School effectiveness and school improvement*, Routledge, London, UK, February 2010, pp. 27-51.

[19] Kleinberg, J. M., "Bursty and hierarchical structure in streams", 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, New York, NY, 2002, pp. 91-101.

[20] Kop, R., and A. Hill, "Connectivism: learning theory of the future or vestige of the past?", *International Review of Research in Open and Distance Learning*, electronic book, <http://www.irrodl.org/index.php/irrodl/article/view/1101/1920>, 2008, pp. 69-81.

[21] Krause, A., J. Leskovec, and C. Guestrin, "Data association for topic intensity tracking", *Proceeding ICML '06 Proceedings of the 23rd international conference on Machine learning*, ACM, New York, NY, 2006, pp. 497-504.

[22] Lampel, J., and A. Bhalla, "The role of status seeking in online communities: giving the gift of experience", *Journal of Computer-Mediated Communication*, Indiana University, Bloomington, IN, January 2007, pp. 434-455.

[23] Lave, J., and E. Wenger, *Situated Learning: Legitimate Peripheral Participation*, Cambridge University Press, New York, NY, 1991.

[24] Magnusson, S., J. Krajcik, and H. Borko, "Nature, sources, and development of pedagogical content knowledge for science teaching". In J. Gess-Newsome and N. G. Lederman (Eds.), *Examining pedagogical content knowledge: The construct and its implications for science education*, Kluwer Academic Publishers Boston, MA, 1999.

[25] Mane, K. K., and K. Borner, "Mapping topics and topic bursts in PNAS", *PNAS*, vol. 101, suppl. 1, National Academy of Sciences, online <http://www.pnas.org/>, April 6, 2004, pp. 5287-5290.

[26] National Research Council, *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and*

Core Ideas, The National Academies Press, Washington, DC, 2012.

[27] Rainie, L., and B. Wellman, *Networked: The new social operating system*, MIT Press, Cambridge, MA, 2012.

[28] Rashid, A. M., K. Ling, R. D. Rassone, P. Resnick, R. Kraut, and J. Riedl, "Motivating participation by displaying the value of contribution", *Proceedings of SIGCHI*, ACM, New York, NY, 2006, pp. 955-958.

[29] Rodríguez, D., M. Sicilia, M., S. Sanchez-Alonso, L. Lezcano, and E. García-Barriocanal, "Exploring affiliation network models as a collaborative filtering mechanism in e-learning", *Interactive Learning Environments*, Routledge, London, UK, May 24, 2011, pp. 317-331.

[30] Shulman, L. S., "Those who understand: Knowledge growth in teaching", *Educational Researcher*, SAGE Publications, Thousand Oaks, CA, Feb 1986, pp. 4-14.

[31] Shulman, L. S., "Knowledge and teaching: foundations of the new reform", *Harvard Educational Review*, 57(1), Harvard Education Publishing Group, Cambridge, MA, 1987, pp. 1-22.

[32] Siemens, G., "Learning analytics: envisioning a research discipline and a domain of practice", LAK12 Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, ACM, New York, NY, May 2012, pp. 4-8.

[33] Suthers, D., and K. Chu, "Multi-mediated community structure in a socio-technical network", LAK12 Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, ACM, New York, NY, May 2012, pp. 43-53.

[34] Weiss, I. R., E. R. Banilower, K. C. McMahon, and P. S. Smith, *Report of the 2000 National Survey of Science and Mathematics Education*, Horizon Research, Chapel Hill, NC, 2001.

[35] Wenger, E., *Communities of Practice: Learning, Meaning, and Identity*, Cambridge University Press, Cambridge, UK, 1998.

[36] Wenger, E., B. Tayner, and M. de Laat, *Promoting and Assessing Value Creation in Communities and Networks: A Conceptual Framework*, Ruud de Moor Centrum, Amsterdam, 2011.

[37] Wenger, E., R. McDermott, and W. M. Snyder, *Cultivating Communities of Practice*, Harvard Business Press, Boston, MA, 2002.