Digital environments for networked learning and professional networks may not comprise one “community”: identification of clusters of affiliated groups of participants that potentially constitute embedded communities is an empirical matter, and one of interest to managers of large learning and professional networks. Also, these socio-technical networks are typically multi-mediated, in that they offer multiple means of participation, each with their own interactional affordances. Different communities may be using the multiple media in different ways. We have developed an analytic framework for extracting events from log files and representing interaction and affiliations at different granularities as needed for analysis. In this paper we show how bimodal networks of actors and media artifacts can be constructed in which directed arcs relate actors to the artifacts they read, write or edit, and how the resulting graphs can be used to detect community structures that extend across different media. We illustrate these ideas with a study that characterizes community structure within the Tapped In network of educational professionals, and how the associations between members of this network are distributed across media (chat rooms, discussion forums and file sharing).

Categories and Subject Descriptors
K.4.3 [Computers and Society]: Organizational Impacts – computer supported collaborative work

Keywords
Socio-technical networks, community structure, distributed learning, networked learning, social network analysis.

1. INTRODUCTION

Learning in university settings, professional communities, and virtual organizations [3, 5, 12] is increasingly technologically embedded, with the rapid adoption of information and communication technologies in support of “online,” “distributed,” and “networked” learning and knowledge creation activities [2, 10, 37], and their blending with face-to-face venues [15]. A related trend is towards open learning communities. In corporate or other work settings, professional learning communities may cross team contexts rather than being isolated in work teams [38]. The sharing of resources in these networks benefits both the individual users within these networks and their collectives [18, 37], and the network and socio-technical infrastructures in of themselves constitute a form of socio-technical capital [16, 28]. We will refer to these various technologically embedded social networks as socio-technical networks [21]. A fundamental question in all of these settings is how learning and other enhancements of knowledge, skill and capital take place through the interplay between individual and collective agency. Such a question demands analyses that connect learning activity in specific times and places with the larger socio-technical network contexts in which they take place. A related analytic challenge is that the granularity at which events are recorded may not match analytic needs. Addressing these analytic challenges by connecting levels of analysis is one objective of our developing analytic approach [34].

Many digital environments for networked learning and professional networks are multi-mediated, in that they offer multiple means of participation, each with their own interactional and social affordances (e.g., asynchronous discussion forums, quasi-synchronous chats, and file sharing). Thus, there are different mediational means through which members may be affiliated. Licoppe and Smoreda [25] found that the choice of technologies by which people share personal news or keep in touch with each other both reflects and reaffirms the nature of the relationship between the interlocutors. The present line of research applies this idea at a collective rather than dyadic level: communities are embedded within and make use of technological media for interaction in ways that reflect and reaffirm their community nature. We address the question of community identification as an empirical matter, discovering clusters of affiliated groups of participants in socio-technical networks rather than assuming that the network constitutes one community or that that prior or external communities are replicated within the socio-technical system. We then examine how the discovered clusters correspond to participants’ organizational affiliations and projects. The fact that learning and knowledge creation activities in these networked environments are often distributed across multiple media and sites leads to a second analytic challenge. The networked learning environments we study offer mixtures of threaded discussion, synchronous chats, wikis, whiteboards, profiles, and resource sharing. Events in these media may be logged in different formats and databases, disassociating actions that for participants were part of a single unified activity. This disassociation is exacerbated when activity is distributed across multiple virtual sites or spread over time, and by the need to work at higher levels of description alluded to earlier in the first
analytic challenge. To address both of these analytic challenges, we developed an analytic hierarchy that support bridging between local analysis of sequential activity and global analysis of mediated affiliations and ties, and associated representations that abstracts from media-specific transcript representations [34]. This hierarchy includes an intermediate representation, the associogram, that supports structural network analysis while preserving information about mediation and direction of interaction.

The concerns discussed above are particularly salient in the professional learning network we are studying, SRI’s Tapped In professional network of educators [14, 30]. The network has consisted of both organizational “tenants” and individuals who come for their own enrichment. There are multiple forms of participation and mediational means by which participant associate with each other. This paper reports an analysis of community structure within the Tapped In network, and how this structure is distributed across media (chat rooms, discussion forums and file sharing). The paper also illustrates an application of our analytic framework. Affiliation networks of actors and media artifacts were constructed in which directed arcs relate actors to the artifacts they read, write or edit. Analysis of these networks exposes community clusters within the network and how they are distributed differently across media types.

The remainder of this paper begins with a description of Tapped In. We then summarize the analytic approach mentioned above. After describing how the Tapped In data was prepared for analysis in the associogram representation, we provide empirical results. We characterize the overall network studied, and then focus on the top six largest sub-networks detected by a “community detection” (modularity partitioning) algorithm. These sub-networks are described in terms of what we know about the most active participants in each network, showing that sub-networks derived on a purely structural basis correspond to clusters of institutional affiliations and purposeful collective activity in Tapped In. The paper ends with discussion of what this tells us about Tapped In, and what this tells us about analyzing distributed activity in socio-technical networks.

2. TAPPED IN

The study examines participant interaction in SRI International’s Tapped In (tappedin.org), an international (albeit mostly US) network of educators engaged in diverse forms of informal and formal professional development and peer support [14, 30]. Cumulatively, Tapped In has hosted the content and activities of more than 150,000 education professionals (over 20,000 per year in our study period) in thousands of user-created spaces that contain threaded discussions, shared files and URLs, text chats, an event calendar, and other tools to support collaborative work. More than 50 “tenant” organizations, including education agencies and institutions of higher education, have used Tapped In to meet the needs of students and faculty with online courses, workshops, seminars, mentoring programs, and other collaborative activities. Also, approximately 40-60 community-wide activities per month were explicitly designed by Tapped In volunteer members to help connect members. (Volunteers drive the majority of Tapped In activity.) Extensive data collection capabilities underlying the system captured the activity of all members and groups. SRI colleagues provided eight years of anonymized data to us. Out of this data, we selected a period of peak usage that occurred from September 2005 through May 2007 for analysis in this study.

Because Tapped In is populated with members of multiple tenant organizations as well as unaffiliated members, it is best seen as a network of education professionals rather than a single “community.” Members may move freely between most forms of participation. The question of what communities (or clusters) exist in this network is a matter for empirical investigation. We approach this question in terms of the artifact-mediated associations found between members.

3. ANALYTIC APPROACH

Our prior research is methodologically eclectic, taking insights from multiple traditions. At a fine granularity, rich descriptions and the unpacking of the sequential structure of situated activity leads to insights into the experiences of participants and the methods by which they accomplish their objectives. Yet microanalytic approaches do not capture emergent social structures that are constructed by yet influence local activity. At the other extreme, social network data and analytic techniques can uncover social regularities, but binary “ties” obscure the situated interactions that constitute and sustain these ties. Both levels of analysis are needed, and we also see the need for bridges between the two levels. There is a dialectic in socio-technical networks in which the local and the global influence and indeed constitute each other. These influences are mediated by cultural artifacts as well as by human actors [20]. Latour’s [23] version of Actor-Network Theory inspires us to trace out the artifact-mediated pathways of influence between actors, acknowledging that non-human entities or “actants” are influential in the network as well as human actors. We use the term “association” for these mediated pathways, preferring it to “relationship” or “tie” because the latter have socio-emotive connotations, while we include in our scope of consideration the diversity of other ways in which people interact with or influence each other [10, 18], or even potentially may do so in the future [31].

Motivated in this manner, we have developed an abstract transcript representation called the Entity-Event-Contingency (EEC) graph that provides a unified analytic artifact [33]; and an analytic hierarchy derived from the EEC that supports multiple levels of analysis [34]. To construct the EEC, log files are abstracted to collections of events (actors taking actions on media objects), supported by a domain model describing the participating entities (actors, actions and digital objects or “artifacts”). Other information is also recorded, such as time and virtual location. To support sequential analysis of interaction, directed graphs that record observed relationships between events called contingencies are constructed [33]. The present study does not use contingency graphs, but rather abstracts further to associograms, two-mode directed graphs that record how associations between actors are mediated by their creation and modification of and access to digital artifacts.
3.1 Associograms
Associograms are like affiliation networks, but they relate actors to mediating artifacts and can be directed: arcs are directed from actors to the artifacts they read, and from artifacts to the actors who wrote or edited the artifact (Figure 1). The arcs represent a form of dependency: they may be reversed to indicate direction of information flow. Existing social network analytic techniques for affiliation network analysis may be applied to associograms, or transitive closure can be used to convert associograms to sociograms to which other existing techniques can apply [36]. The results of network analysis can then be interpreted by reference to the other levels of analysis. Thus associograms bridge between interaction data and network analysis.

Associograms offer a different viewpoint on a network than sociograms. An associogram tie links an actor to an artifact; therefore, an artifact always mediates an association between two actors. Another way to visualize this is to expand a sociogram tie, identify the means by two actors have interacted, and then explicitly display the mediating factors. Because digital artifacts afford different types of interactions, associations between actors need not be the result of “direct” exchange or transmission of information. While the tie in a sociogram might represent a relationship (e.g. family member) or metric that defines the relationship between the actors (e.g. frequency of contact), an association defines the relationship in terms of the distribution of affiliations across a set of mediating artifacts. This enables study of how users connect with others within a socio-technical system, and whether the different media types affect the relationships and structures that are formed.

3.2 Data Preparation
We now transition to discussion of aspects of method specific to our analysis of Tapped In. Preparation of the data required many months of work, and was greatly facilitated by SRI colleagues. We parsed and filtered logs of user activity involving three different types of digital artifacts: files, asynchronous threaded discussion forums, and quasi-synchronous chat rooms. The relevant interactions consist of users accessing (reading and downloading) or contributing (posting and uploading) to one of these three artifact types. Private chats were excluded from our analysis, as we are focusing on observable public behavior. All activity in the K-12 (student) campus was excluded, as our research (and human subjects permission) focuses on the professional community. Guest accounts were also filtered, as different rotating individuals use these. File access was filtered to include only intentional access such as clicking on a link to a file such as a Word document, Powerpoint presentation, etc. (excluding downloads incidental to entering a room). Data from several different EEC and Tapped In database tables, including old system backups that provided missing data, were consolidated to help with later analyses, and the arcs (directed edges) were weighted according to the number of times the given actor-artifact affiliation was observed, as discussed below.

We constructed an associogram in which vertices represented actors, files, discussions, and chat rooms. A file vertex represents a single file, a discussion vertex represents an entire threaded discussion, and a chat vertex represents a chat room. The direction of the arcs (directed edges) is a form of state dependency: the source of an arc is dependent on the target of the arc in that information or content has potentially moved from the latter to the former. A file vertex points to an actor if the actor created (uploaded) that file; an actor vertex points to a file vertex if the actor has downloaded the file. A discussion vertex points to an actor vertex if the actor has posted a message in the corresponding discussion forum; an actor vertex points to a discussion vertex if the actor has accessed messages in the discussion forum (having loaded the discussion page). A chat room points to an actor if the actor posted a chat contribution in that chat room while someone else was present; an actor vertex points to a chat vertex if the actor was present in the room when another actor posted a chat contribution. Finally, each of these arcs is weighted according to the number of time the events just described were seen. For example, if an actor's arc to a chat room has a weight of 375 then the actor was present for 375 contributions made by other actors.

3.3 Data Analysis
We used the Gephi [6] software package to analyze the associogram for our data. Gephi is an open-source tool that provides a suite of network metrics and interactive visualizations. At this writing, Gephi is still in the beta stage of development (version 0.8), and it is not as mature as Pajek or UCINet in terms of available sociometrics. However, unlike these older software tools (which use matrix representations of graphs), Gephi uses an adjacency list representation of graphs, so is able to handle very large graphs. Gephi provides several of the more recent layout algorithms for large graphs and a recent community detection algorithm with good properties (discussed below). Once our graph was partitioned into candidate communities, we used Gephi to

![Figure 1. Constructing an associogram from events. For example, the event of P2 writing message m1 is represented as a dependency of m1 on P2, and the event of P1 reading m1 is represented as a dependency of P1 on m1. The cycle between P1 and P2 indicates an interactional relationship, while P3 is a consumer of artifacts created by P1 and P2.](image)
visualize the partitioning in various ways. We also used Gephi’s “Data Laboratory” to inspect members of the partitions. The suite of metrics available in Gephi was run to generate basic descriptive statistics. We also exported spreadsheets of various partitions to compute sums of unweighted and weighted in-degree and out-degree in order to interpret the distribution of communities across artifact types. When needed, we accessed our original database to look up additional information on the entities involved or inspected the live Tapped In environment.

3.3.1 Data Visualization
We applied Gephi’s implementation of the OpenOrd [26] layout algorithm to visualize the network. OpenOrd is a force-directed algorithm that scales well to support very large networks. OpenOrd is based on Fruchterman-Reingold, an O(n^2) force-directed layout algorithm, but resolves two of the latter’s defects: Fruchterman-Reingold is too slow for large graphs, and may obscure global structure. OpenOrd uses a multilevel approach, where a sequence of successively coarser graphs is constructed by clustering vertices according to edge weights and distances and representing each cluster with a single vertex in the coarser graph. Layout is computed with the coarsest graph obtained and then the constituent vertices are placed in the location of their cluster to initiate the next finer iteration of layout. Additionally, edge cutting is used to prevent long edges from unduly pulling together vertices that are best displayed in distinct global structures. Several phases are used in which parameters are varied to resolve tradeoffs between initial clustering, expansion, tightening of clusters, and fine-tuning. The OpenOrd algorithm is very fast and exposes structure in large graphs.

3.3.2 Community Detection
In the network analysis literature, “community” refers to clusters of mutually associated vertices under graph-theoretic definitions rather than to the sociological concept, and “community detection” refers to finding sub-graphs that constitute such structural clusters. However, a good graph-theoretic definition should capture the intuition that individuals in a sociological community are more closely associated with each other than they are with individuals outside of their community. In this paper, we understand “community” as empirically associated actants for whom it is also possible to identify some focus of their shared activity. This sense of “community” is much looser than the traditional gemeinschaft [35], and does not make claims about participant’s own identities [11]. We believe that a looser definition is appropriate for the networked age [10, 37]. We use graph theoretic terms (e.g., “partition”) when discussing algorithmic results that are candidates for interpretation as a community, and usually reserve “community” for when we are entering into the realm of such interpretation by trying to identify what a cluster has in common, except when referring to the larger endeavor of “community detection”.

A variety of graph-theoretic definitions of communities are available, and there are multiple algorithms for each. Algorithms based on the modularity metric are widely used. The modularity metric compares the density of weighted links inside partitions to weighted links crossing between partitions, ranging from 1 (high modularity) to -1 (no modularity). A partitioning of a graph into highly modular partitions (a.k.a. “modularity classes” in Gephi) defines nonoverlapping communities. Finding the best possible partition under a modularity metric is computationally hard (impractical to compute on large networks) [9]. Blondel et al. [7] offer a fast approximation algorithm that gives good results. On each pass, each vertex is initially placed into its own separate partition, and then the algorithm examines the neighbors of each vertex to see whether moving the vertex to a neighbor’s partition can increase modularity. This process iterates over vertices until no vertices move between partitions. Then each partition so constructed is collapsed into a single vertex, with the edges to other partitions merged, summing their weights. The algorithm then repeats on the collapsed vertices until there is no further change in partition membership. At this point, a local maximum has been reached that is not guaranteed to be the most optimal partition under the modularity metric, but has been shown to be a good approximation.

4. EMPIRICAL RESULTS
In this section we summarize metrics for the overall socio-technical network studied: all activity within Tapped In surrounding chats, discussions and files (filtered as discussed previously) for the period from 9/2005 to 5/2007. We also summarize the metrics for media-specific networks. Then we present the modular partitions found by the Blondel et al. algorithm and interpret the several largest partitions as communities. We do so by examining who the top actants (actors and artifacts) are, what their affiliations and/or intended purpose seems to be, and how their activity distributes across media. These results are used to illustrate the utility of this kind of analysis.

4.1 Overall Network Metrics
In the combined network, there are 40,490 vertices (a.k.a entities in the EEC, or actants in the world). These comprise 19,842 Actors (49.00%), 12,037 Discussions (29.73%), 5,862 Files (14.48%), and 2,749 Chat Rooms (6.79%). The combined network has 229,072 edges. The presence of an edge represents the existence of a person-artifact association. Each edge may represent one or more events involving that person and artifact. Edges are weighted according to the number of events. The sum of weights on all edges gives us the number of events encompassed by the total analysis. The sum of weighted degree in the entire network is 20,431,944, which constitutes the number of events (as we have defined them) analyzed.

The weighted out-degree gives the number of “write” events for the given artifact (the arc indicates that the state of the artifact depends on the target actor), so the sum of weighted out-degree across all artifacts of a given type gives the number of ‘write’ events for that artifact type. Similarly, the weighted in-degree gives the number of “read” events for an artifact (the arc indicates that the actor has accessed the artifact), so the sum gives the total number of such events for an artifact type. The sums shown in Table 1 indicate how activity distributes across artifact types, with the caveat that the units are different activities. There are over twice as many events in the chat rooms as in discussions, and few file events, which reflects typical frequencies with which one

<table>
<thead>
<tr>
<th>Table 1. Weighted degrees (events)</th>
<th>Weighted Degree</th>
<th>In-degree</th>
<th>Weighted Out-degree</th>
<th>Row Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chat Rooms</td>
<td>12,220,792</td>
<td>2,512,887</td>
<td>14,733,679</td>
<td></td>
</tr>
<tr>
<td>Discussions</td>
<td>5,592,946</td>
<td>45,085</td>
<td>5,638,031</td>
<td></td>
</tr>
<tr>
<td>Files</td>
<td>54,372</td>
<td>5,862</td>
<td>60,234</td>
<td></td>
</tr>
<tr>
<td>Artifacts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td>17,868,110</td>
<td>2,563,834</td>
<td>20,431,944</td>
<td></td>
</tr>
<tr>
<td>Actors</td>
<td>2,563,834</td>
<td>17,868,110</td>
<td>20,431,944</td>
<td></td>
</tr>
</tbody>
</table>
might interact with each of these artifact types. These results illustrate how the directed, weighted, and multimodal properties of associograms provide information about mediated activity not available in simple sociograms.

Some further graph metrics help characterize this network in comparison to other networks. The density is less than .001 (does not display any significant digits in Gephi), so this is a sparse graph. The network diameter is 17, and the average path length is 4.398, smaller than the “6 degrees of separation” found in other networks [e.g., 24]. This result is especially remarkable given that the associogram is a bimodal graph in which an artifact must be present between every person connected. If the graph were collapsed to direct actor-actor ties the average path length would be about half, so this network is more closely knit than typical networks, although others have found similar results [4].

The close connectivity is partly due to associations via a single artifact, the Tapped In Reception (R1), which most users pass through after logging in. This room has a normalized betweenness centrality [8] of 0.665, and has by far the highest degree (2,511,057 weighted or 18,810 unweighted) of any actant. If this mediating artifact is deleted from the graph, average path length goes up to 6.02, or about 3 if intermediate artifacts are removed. We return to the relevance of R1 shortly.

4.2 Modular Partitions

The Blondel et al. algorithm [7] constructs 171 partitions with an modularity score of 0.817. The combined network is visualized in Figure 2. Color represents the modular partitions generated by the algorithm. Figure 3 shows the six largest partitions by number of vertices. Henceforth we refer to these six partitions as partitions 1-6. The overall visualization in Figure 2 shows a strong central cluster, as do several of visualizations of the six largest partitions in Figure 3. The fact that these central clusters are co-located in the OpenOrd layout (which places more strongly related vertices near each other) indicates that there is overlap between the cores of these potential communities.

The visualization also juxtaposes vertices for several important actants on top of each other. In order to separate the important actants and see the relationships between them, Figure 4 shows only those actants of unweighted degree greater than 282, with vertex size scaled by weighted degree, and using a radial layout with a non-overlapping filter. Several of the major actants are labeled anonymously, and will be discussed below. A, B, C, and D are actors, D1 is a discussion forum, and R# indicate chat rooms. The colors indicate partitions, as in Figures 2-3. (If color is not visible, of the numbered vertices, R1, R3, R4, and R5 are in partition 1; A, B, C, D, and R6 are in partition 2; R8 is in partition 3; D1 in partition 4; and R2 in partition 6.)

Figure 4 shows that there are strong relationships between partition 1 and the other networks, particularly partition 2. We were concerned that the large degree of R1, the Tapped In Reception chat room around which partition 1 forms, might skew the community analysis. Its degree and weighted degrees are a factor of 10 larger than the next largest actant. This may be due to the fact that R1 is the default room into which anyone without a specific tenant affiliation is placed when logging in, and an association will form if anyone chats while they are there. The “help desk” is located in R1, and help desk volunteers often greet others who enter this room. R1 plays an important role in the functioning of Tapped In, but from the standpoint of community analysis R1 presents a dilemma. On the one hand, associations via
Blondel et al. community detection algorithm (and many other community detection algorithms): it does not allow vertices to lie in multiple communities. Thus, any actant placed into the modularity class surrounding R1 will not be available for membership in other modularity class, potentially obscuring their intentional participation elsewhere.

In order to assess the risk of superficial associations via R1 obscuring more purposeful communities, we conducted an analysis with R1 (and all of its associations) deleted. As expected, the large partition centered on R1 (partition 1) disappeared and partition 2 became the largest partition, absorbing many of the important actants formerly in partition 1. The interpretation of this new collapsed partition was very clear, and the other largest partitions still produced similar interpretations as communities. However, we decided to leave R1 in for this final analysis, for two reasons: First, many actors were orphaned by the removal of R1: 2178 isolates appear. Second, we felt that we should grapple with interpreting the partitions resulting from the unedited data, as we may learn something about both the Tapped In network and what a community detection algorithm on an associograms can show us. As it turns out, the partition (#1) forming around R1 has a useful interpretation that can be distinguished from partition 2.

### 4.3 Example Interpretations of Partitions

In this section we discuss our interpretation of the highest modularity partitions discovered to illustrate the method. Each of the networks is interpreted for what kind of human network it represents by examining what is known about the actants (actors and artifacts) that are ranked highest by degree within the associated partition, indicating their importance to the activity of that partition.

#### 4.3.1 Partition 1

The largest modular partition has 8452 vertices (20.87% of the total graph) 29698 edges (12.96% of the total edges). There are 6953 actors, 673 chat rooms, 495 discussions, and 331 files in this partition. We illustrate how this partition is interpreted by examining top ranked actants and the distribution of activity across media.

#### 4.3.1.1 Top actants by degree

Recall that unweighted degree is a measure of how many other actants a given actant has been associated with, and weighted degree is a measure of how many contact events there have been, i.e., a measure of level of activity.

The top 20 actants by unweighted degree in this partition are all rooms. Out-degree indicates the number of persons chatting in each room, and in-degree indicates the number of persons hearing someone chat in each room. Total degree counts ingoing and outgoing associations, so will count someone who both chatted and heard a chat twice. The top ranked actants are:

- **R1**, the Tapped In Reception: 7173 out, 11,637 in, 18,810 total.
- **R4**, the public room for Tapped In’s After School Online (ASO) events: 898 out, 986 in, 1884 total.
- **R10**, the Floor Lobby for the Tapped In Groups floor: 489 out, 601 in, 1090 total. All groups are on the third floor; this is the lobby through which you enter that floor.
- **R5**, the personal office of Actor E, a faculty member focusing on teachers’ use of technology, and the active owner of a Tapped In group on teacher education: 313 out, 319 in, 632 total.

![Figure 4. Radial layout of high degree vertices (degree > 282; Gephi does not currently filter by weighted degree). Vertex size is weighted degree. Color is modularity class.](image)

**R9**, “ComfyConf”, a public conference room available for use by Tapped In members: 285 out, 310 in, 595 total.

**R12**, the personal office of Actor B, an important volunteer to be described in the next section: 289 out, 296 in, 585 total.

The top 20 weighted degree vertices in this community are also all rooms. Weighted out-degree indicates the number of chat events in each room; weighted in-degree indicates the number of events of someone being present when someone else chatted (loosely, “hearing” events); and the total can be taken as a measure of total level of chat interaction in the rooms:

- **R1**, the Tapped In Reception: 549,572 out (chatting events) 1,961,485 in (“hearing” events), 2,511,057 total. This room hosts 12.29% of all chat events in the network.
- **R4**, the ASO Public Room: 51618 out, 253,611 in, 305,229 total.
- **R5**, the personal office of Actor E, 19448 out, 213474 in, 232922 total.
- **R11**, a group room owned by Actor B: 10872 out, 172479 in, 183351 total.
- **R9**, the “ComfyConf”: 16785 out, 106596 in, 123381 total.

Most of the rooms on both top-20 lists are public, reception or group rooms. Some these are owned by groups with diverse purposes (Art, Blogging, Math, Portfolios, Robotics, Writing), making it difficult to identify a specific purpose or activity for this partition. But the very fact of this diversity and consideration of other information leads to a clearer interpretation of this partition.

Many of the highest ranked rooms (R1, R4, R9 and R10) are owned by Tapped In, and explicitly function either to welcome newcomers and route them to their destinations or as venues for public events open to all. The personal offices involved mostly belong to people who keep open office hours to help others. Also,
consider the structural fact that when R1 is deleted, 2178 actor vertices (5.3% of all actants and 10.98% of all actors) become isolated, so were only linked to R1. Of these, 1519 are in partition 1, meaning that 17.97% of actors in partition 1 are there solely because of their association via R1. Furthermore, the distribution of degrees is highly skewed: the average is 9.31 (4.62 for actors), the median 3, and the mode 2. Thus this partition consists of many actors who have weak affiliations within Tapped In, but are bound together by their mutual association with the major entry points and centers of activity for new members or those unaffiliated with tenant organizations: the reception, rooms in which public events take place, and offices of volunteers, as well as some public group rooms for popular topics. We interpret this partition to represent not a separate community with its own purpose or activity, but rather a large network in which more specific communities represented by other partitions are embedded. It may also reflect a phase of participation in which new members are becoming familiar with Tapped In, after which they may or may not deepen their participation in specific groups. (Temporal analysis is a topic for further research.)

4.3.1.2 Media distribution
The degrees for each artifact type give us a general overview of the distribution of activity in this partition. Unweighted degree summarizes the number of other actants a given actant has some affiliation with. Weighted degree is a better overall measure of level of activity bearing on or emanating from an actant, keeping in mind that events involving chats, discussions and files each are of a different nature.

Examining the degrees in Table 2, we can see that the bulk of events involving these 6953 actors are overwhelmingly chat-based. This is consistent with the fact that the top actants are chat rooms, and with the interpretation that this partition gathers together activity related to newcomers and chat sessions popular with participants not strongly active in tenant organizations.

<table>
<thead>
<tr>
<th>Table 2. Media Associations in Partition 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>(6953 Actors)</td>
</tr>
<tr>
<td>Unweighted</td>
</tr>
<tr>
<td>Weighted</td>
</tr>
<tr>
<td>In</td>
</tr>
<tr>
<td>Chat Rooms (673)</td>
</tr>
<tr>
<td>Discussions (495)</td>
</tr>
<tr>
<td>Files (331)</td>
</tr>
</tbody>
</table>

4.3.2 Partition 2
The next largest sub-network (modularity class) has 5826 vertices (14.39% of the actants) and 20459 edges (8.93% of events), with 2485 actors, 782 chat rooms, 1828 discussions, and 731 files. Henceforth we will report the top five actants by weighted degree to indicate where the activity lies, and then add those that are in the top five by unweighted degree to ensure we also include actants with high connectivity. For brevity we now report only total degree in parentheses. The top five weighted degree actants in this partition are as follows:

**Actor B** (weighted 370,604; unweighted 2,892), a volunteer who was given Facilitator status and paid for some (but not all) of her activities. Account B is the second most active account in the system. The real-world actor was given a second account B’ (121,259; 238), so that she could facilitate two events at once. Account B’ is a member of a tenant organization for a partnership of multiple school districts during the time our data was gathered. Taken together, the real world actor B/B’ is as active as Actor A.

**R13** (441,670; 229), the personal office of a faculty member in the school of education at a community college. R13 does not appear in Figure 4 due to the lower unweighted degree.

**Actor C** (336,002; 305), a middle school technology support staff and a Tapped In help desk volunteer.

**R6** (221,078; 369), an educational technology group room owned by a university education faculty member.

The top five actants by unweighted degree include Actors A and B, and the following:

**R14** (48,116; 2,355), the personal office of Actor A. This office has nickname “Online Support” and is described as offering “sustained online support”.

**R15** (31,847; 625), a resource room for a primary school center, owned by **Actor G**, a primary school teacher and Tapped In help desk volunteer who is involved in many groups.

**Actor D** (166,568; 583), a middle school technology teacher who became a Tapped In help desk volunteer during the period of this study.

Most of the actors on this list are help desk volunteers. This suggests that partition 2 has some overlap in function with partition 1: helping support the broader Tapped In community. But scanning ranked lists of chat rooms and discussions, a striking difference emerges: while partition 1 has a mixture of personal offices and group rooms, the top ranked chats and discussions in partition 2 are overwhelmingly group rooms. The topics are diverse, including assessment, climate change, librarianship, online teaching and learning (two groups), music, special education, teachers in training, the WWW (two groups), and a system-wide Tapped In festival. We found that all of these chat rooms have one thing in common: they are the site of regular (repeating) and public After School Online (ASO) events announced in the calendar. Furthermore, actors A and B are known for facilitating or supporting the facilitators of many ASO events (i.e., were present in those rooms when the chatting took place). These facts suggest that partition 2 is the broad public community associated with After School Online, arguably the largest and most significant activity in Tapped In. We acknowledge that some relevant rooms such as R4, the After School Online room, R9, the “ComfyConf” room, and R10, the floor lobby for Tapped In groups, are in partition 1. However, all of these actants ended up in the same partition in the other analysis conducted without R1.

Examining the distribution of activity across media in Table 3, most of the events are chat-based, which makes sense given that ASO events are chat based. However, although there is less than half the number of actors as in partition 1, both discussion activity and file sharing are double those in partition 1. These facts are consistent with the interpretation that partition 1 focuses more on brief chat interactions such as when persons enter R1 and are given help, while partition 2 includes topic focused public groups, some of which include asynchronous discussion as well as
scheduled chats. Help desk volunteers are involved in both of these activities: they are represented by the rooms where help is given in partition 1, and by themselves as actants participating in various group rooms in partition 2.

Table 3. Media Associations in Partition 2

<table>
<thead>
<tr>
<th>(2485 Actors)</th>
<th>Unweighted</th>
<th>Weighted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chat Rooms (673)</td>
<td>4,862</td>
<td>1,198,117</td>
</tr>
<tr>
<td>Discussions (495)</td>
<td>7,228</td>
<td>3,654</td>
</tr>
<tr>
<td>Files (331)</td>
<td>2,841</td>
<td>731</td>
</tr>
</tbody>
</table>

The remaining partitions will be described much more briefly due to limitations of space, but sufficiently to illustrate how modularity partitioning on associograms identifies clearly interpretable communities once the densely connected core has been taken care of.

4.3.3 Partition 3

The next largest modularity class has 2565 actants (6.33% of total), including 851 actors, 103 chat rooms, 1286 discussion forums, and 325 files. Interpretation of this sub-network is much more straightforward. Sorted by degree, two rooms followed by 40 actors have the highest degree, and all 42 of these actants have tenant affiliation with a public school district of a city in the midwestern US. The top ranked actors by degree are all teachers at various levels in this public school system. When sorted by weighted degree, there is a mixture of rooms, discussions and actors, the majority of which are again associated with this public school system or with a related teacher education center.

These results suggest that there was formal involvement of this school system during 2005-2007. We checked this interpretation out with our SRI colleagues. According to Judi Fusco (an SRI researcher on the Tapped In project; personal communication October 2011), Tapped In began working with this school system in 2002 to help them support new teachers online with a professional development center (NPD), both located in the same state as Actor H. Sorting by unweighted degree adds three more actors, all teachers in the same Midwestern region. The media usage distribution for these 857 actors is weighted on the chats (104,666 in, 35,111 out). There are low figures for the 2313 discussions—most have only one posting and a few reads. This appears to be a series of events run by Actor H.

4.3.6 Partition 6

Finally, the sixth largest modularity class has 1037 vertices (2.56%) and 6858 edges (2.99%), with 729 actors, 153 chat rooms, 71 discussions, and 220 files. The affiliation with a university on the west coast (we’ll call it WU) is clear. The top five actants by weighted degree include four personal offices of WU faculty members in education, and one group room owned by a researcher at WU and SRI. Sorting by unweighted degree we add one more WU faculty member office, and two WU public rooms. Almost all of the 100 top ranked actors by weighted degree have WU as their tenant affiliation. The interaction is clearly chat based (sum of weighted degree 4,519,127 compared to only 27,789 for discussions). Heavy use is made of personal offices: 41 of top 50 chats are in personal offices. Asynchronous discussions play a lesser role, but are affiliated with a nearby public school system, suggesting involvement of this university with local public schools. Clearly, this cluster is associated with WU teacher education, possibly in collaboration with local public schools. Judi Fusco confirms that WU was one of the major Tapped In tenants during this time period. They have Masters and EdD programs, and apparently had advanced students doing professional development with teachers in the schools.

4.4 Small Communities

The visualization of Figure 2 shows that there are also many small peripheral networks embedded in the larger network. We inspected some of the small clusters shown in Figure 5, and found a diversity of actors and topics. (1) One centers around a group room for educational technology students at a liberal arts college with multiple campuses and online programs. This room is owned by an elementary school teacher who is also teaching online for this institution. (2) A cluster of actors is connected by several discussions, all owned by a high school language arts teacher in the Midwest. (3) Another room owned by an education faculty member at a west coast state college is surrounded by many actors. Their user IDs are distributed in a manner suggesting clustered creation over a number of years. Sampling the actors we have three actors, all being secondary school teachers located in the SPD state. Of the top 50 ranked discussions, 35 of them are affiliated with the SPD tenant. Of the remaining 15, the Tapped In Reception hosts 12. This is an online professional development effort involving persons in multiple southern states but with the greatest activity focused in one state. It is possibly a collaboration between the state PD center and the national PD center. Media figures show that this is primarily a discussion-based effort, so it was conducted asynchronously.

4.3.5 Partition 5

This modularity class has 1251 actants (3.09%), including 112 actors, 35 chat rooms, 1006 discussion forums and 98 files, and 4219 edges (1.84%). Top ranked actants by weighted degree include a room for “final task collaboration” and a room that appears to be for a course on using the internet in K-12 schools. Both are owned by Actor H, a language arts high school teacher in the Midwest, who is the third highest weighted actant. The next two actants are the office of a language arts teacher and the account of a middle school language teacher, both in the same state as Actor H. Sorting by unweighted degree adds three more actors, all teachers in the same Midwestern region. The media usage distribution for these 857 actors is weighted on the chats (104,666 in, 35,111 out). There are low figures for the 2313 discussions—most have only one posting and a few reads. This appears to be a series of events run by Actor H.

4.3.4 Partition 4

The fourth largest modularity class has 1630 actants (4.03%), including 857 actors, 26 chat rooms, 605 discussion forums and 142 files. The top actants involve a state professional development center (anonymously abbreviated SPD), and a national professional development center (NPD), both located in the southern US. Top actants by weighted degree include a discussion on 21st century learning in the SPD group room (D1 in Figure 4); a chat room owned by a teacher educator; an NPD actor with facilitator status who is a university faculty member, a middle school teacher in the same state as SPD, and another discussion in a SPD group room. Adding top five unweighted degree actants, we
find a consistent pattern, with accounts being created around the same time in 2004, another cluster in 2005, and another in 2006. All of the sampled actors are educators in nearby school districts, ranging from elementary through high school and including math, science, technology and social studies. It appears that the state college actor owning this room is running a recurring course or program. These small clusters attest to how Tapped In enables participants to create hundreds of small communities embedded in and synergizing with the larger Tapped In network.

5. DISCUSSION

These communities were detected based purely on structural characteristics. Although our database contains other information such as institutional affiliation, job role, geographic location, and demographics for actors and descriptions for discussions and rooms, none of this information was used in constructing the partition. We only used such descriptive information associated with top-ranked actants in interpreting the partitions after they have been constructed. The fact that we can find a clear interpretation based on the characteristics of actants in each partition attests to the power of the structural method for finding partitions that have some external validity.

A caveat: one should not conclude that all actants in these partitions are engaged in the activity identified by our interpretations. Modularity optimization—maximizing intra-group links while minimizing inter-group links—is appealing as a graph-theoretic definition having some correspondence to our understanding of community, but does not capture traditional aspects of community having to do with shared geography, identity or purpose. Each partition discussed above has up to a few hundred highly connected actants, followed by a long tail of other actants that have some weaker or peripheral association with the network partition. The institutional affiliations, job roles, etc. of the core actants ranked highly by degree give us an idea of the activity that resulted in closer affiliations among these actants, but many other actants who did not necessarily participate in the identified activity but are weakly associated with the network will be placed in the corresponding partition by the modularity optimization algorithm if they do not have stronger associations elsewhere. The same point applies to participation of artifacts. Thus, for example, we will sometimes see very large numbers of discussions involved in a partition, but this does not mean that all of these discussions were devoted to the identified activity.

Yet, this caveat does not diminish the power of the analytic technique. On the contrary, it shows that we can find both the core purpose of a partition and the extent to which it involves others not directly identifying with this core purpose. Also, such expanded structures illustrate the synergistic power of embedding task-specific activities within a larger “transcendent community” [19, 32], a point on which Tapped In was clearly successful.

A limitation of the analysis is the use of a non-overlapping community detection algorithm. Clearly, actors and artifacts may play a role in multiple communities, but many community detection algorithms force each vertex into one community. Recently, various algorithms for overlapping community detection have been proposed, including clusters of k-cliques [27], vertex splitting [17], and approaches that find communities of edges [1, 13]. Ongoing work is evaluating these algorithms with respect to associograms. K-cliques do not apply to bipartite graphs because there are no triads or higher order cliques in a bipartite graph. Edge communities look most promising from both empirical and theoretical standpoints: the results reported in [1]

Figure 5. Closeup of small peripheral sub-networks, many of which constitute topic-focused communities.

are strong, and the approach is consistent with our theoretical position that relationships between entities are primary rather than entities in isolation (see for example [33, 34]). Once we find an overlapping community detection algorithm that is suitable for associograms, we will redo this work to study how actants (including TI-Reception and highly active volunteer facilitators) may bridge between communities.

6. Conclusions

These results demonstrate the vibrancy of Tapped In, in which multiple tenant individuals and organizations support their own instrumental purposes and also lead to the emergence of a larger encompassing socio-technical network that is not in the same sense a “community” but that constitutes the synergistic value of the network—what we have termed a “transcendent community” [19]. As can be see in Figure 2 and its decomposition into sub-networks in Figure 3 and our interpretations above, tenant organizations drive significant activity in Tapped In, but they are entangled with participants from other organizations and the larger public sphere of the volunteer-based After School Online series.

For the purposes of this paper, these results also demonstrate our method. Social network analysis and its use of the classic sociogram generally take a high level view of user relationships, but do not elucidate how those relationships are formed or mediated. Various micro-analytic methods provide detailed information for different forms of interactions, but do not expose structures of the larger network. This study illustrates a middle level of analysis: associations between participants that take place through digital media. Associograms capture valuable information, abstracting enough for aggregate structure analysis across multiple media while preserving some information on the quantity and directionality of interaction. A “community detection” algorithm applied to the Tapped In associogram found actual communities without applying any knowledge of participant affiliation or other demographics. This further illustrates the importance of considering activity across all available mediational means for interaction, not just limiting research to (for example) chats or discussion alone.

7. ACKNOWLEDGMENTS

This work was supported by NSF Award 0943147. The views expressed herein do not necessarily represent the views of NSF. The authors thank SRI and Patti Schank for providing us with
access to the anonymized data, Judi Fusco for her help with our interpretations, and Nathan Dwyer and Devan Rosen for their prior collaborations on this work.

8. REFERENCES


