# Analyzing the Flow of Knowledge in Computer Mediated Teams

**Joshua Introne**
Center for Collective Intelligence
MIT
Cambridge, MA
jintrone@mit.edu

**Marcus Drescher**
School of Management
Technische Universität München
München, Germany
drescher@tum.de

## ABSTRACT
In this article, we present an analysis of communication transcripts from computer-mediated teams that illustrates how different kinds of decision support impact collaborative knowledge construction. Our analysis introduces an algorithmic technique called Topic Evolution Analysis (TEvA), which tracks clusters of words in conversation, and illustrates how these clusters change and merge over time. This analysis is combined with measurements of group dynamics to distinguish between teams using different kinds of decision support.

Our analysis offers evidence that some kinds of decision support improve the apparent rationality of a team, but at the cost of collaborative knowledge construction. This result is not apparent when simply measuring team decision performance. We use this finding to motivate the utility and importance of the approach when assessing the impact of technology on collaborative knowledge processing.

### Author Keywords
Collaborative knowledge construction; decision support; computer-mediated communication; online conversation; topic tracking; knowledge flow

### ACM Classification Keywords
H.5.3. Group and Organization Interfaces: Computer-supported cooperative work, Evaluation/methodology, Theory and models; I.2.7 Natural Language Processing: Text analysis

## INTRODUCTION
One desirable outcome of collaboration is the birth of knowledge that is somehow more than the sum of its parts. Yet collaborative knowledge construction is subject to the vagaries of interpersonal dynamics and hampered by socio-psychological biases, and can thus be an elusive end product for groups. A challenge for designers is how to build technology that can reliably improve a community's ability to exchange and fluidly recombine its available information to develop new insights.


*CSCW '13,* February 23–27, 2013, San Antonio, Texas, USA.


Robust analytical techniques are necessary to determine how technology impacts collaborative knowledge construction and knowledge processing. In an ideal case, we would be able to witness the historical evolution of knowledge as it flows across a socio-technical network, and describe the impact of technology in these terms. For example, is a community better or less able to combine different pools of knowledge over time? What is the average lifespan of a new idea? How homogenous is the knowledge held by a group? Being able to observe the evolution and flow of knowledge in collaboration would allow us to better understand the broader impact of technology on the knowledge processing properties of a group.

Approaches have been developed recently to track the dynamics of topics in document corpora [2,37]. These approaches seek to label each document in a corpus with one of more topic descriptors, but topics can appear, disappear, or change over time. Using these techniques, it is possible to examine how topics can change, merge, and split, generating a rich picture of how knowledge evolves throughout the corpus [7].

This is precisely the sort of analysis we would like to perform with online communities to discover how knowledge flows across an online conversation. Unfortunately, online conversation may be too noisy and dynamic to apply document-centric topic tracking. Unlike a document, which offers a snapshot of a relatively stable distribution of topics, a conversation is an unfolding process in which the definition of a topic is continually re-negotiated. The lack of a relatively well-structured, intentionally designed document limits the utility of the machine-learning approaches that drive modern topic-tracking algorithms.

Consequently, we have sought to develop an approach that makes no assumptions about the existence of documents or their structure. Instead, we take as our unit of analysis a sequence of replies, seek to understand how clusters of words in these reply sequences change, merge, and split over time. We call this approach Topic Evolution Analysis (TEvA).

In this article, we present the algorithm, and use it to analyze a dataset collected in a prior study on decision support technology. By comparing TEvA with other indices of group dynamics we are able to identify important differences in collaborative knowledge processing of groups and better characterize the differential impact of collaborative technologies. Specifically, we show that one type of decision support appears to dampen collaborative knowledge construction, even though there is no clear impact on performance in the decision task.

The goals of our analysis are twofold. First, we demonstrate that a normative approach to collaborative decision support, intended to help collaborators engage in more rational knowledge processing, can have the unintended effect of disrupting a process that allows people to combine their individual knowledge synergistically. This bounded result contains cautions about evaluating the design of collaborative systems and the difficulty of generalizing team performance from targeted measures.

By way of this example we also hope to introduce a new approach to the assessment of online groups via the analysis of collaborative knowledge processing. By examining the movement and evolution of word co-occurrence clusters in online conversation, we can speak with more precision about the knowledge processing capabilities and characteristics of online groups. This in turn can support new ways of distinguishing between online communities and assessing the differential impacts of technology upon them, and ultimately improve our ability to design technology that supports more effective collaborative knowledge processing.

## LITERATURE REVIEW

There is a growing interest in being able to observe the dynamics of information in technologically mediated communities. Approaches based on interaction analysis offer tremendous resolution on the dynamics of knowledge processing in small groups (e.g. [14,26,35]) but these techniques involve manual coding and cannot scale to the volumes of unbounded and informal collaboration now available in online settings.

To scale up to such settings, some researchers have made opportunistic use of structured content that appears in online interactions to observe information dynamics. For instance, social tagging is an affordance of some social networking platforms that users use to identify relevant information. It also provides researchers with clear markers that can be readily followed across a network. Social tagging has been used to illustrate the miscibility of different ideas and the diffusion patterns of information [6,28].

Web users have also often adopted conventional behaviors of copying and pasting snippets of text to create “memes” with enough regularity that researchers can track them across multiple networks based on their lexical similarity [20]. Unlike social tags, lexical memes have internal structure that can be analyzed, allowing researchers to observe how these memes are transformed across networks–mutating, splitting, and even merging [30].

Moving beyond markers like memes and tags, information retrieval and topic modeling techniques have been employed to trace the movement of words or groups of words across networks (e.g. [10][27]). These approaches are a more general approach to analyzing the flow of information in social networks, but as they have been applied offer only a limited view of the kinds of transformations information undergoes over time.

In the last several years, topic-modeling techniques have been applied to elucidate detailed topic dynamics in document corpora [7]. These techniques leverage statistical regularities in word frequencies that result from the structure of the documents and corpora in which they are found. A document is a designed artifact; a self-contained package of knowledge (e.g. an scientific paper, or a news story) that is intended to convey a fixed set of ideas that do not typically change over the course of the document. These topics are relatively stable across documents in a corpus at a given point in time because they describe external phenomena (e.g. news events), and may be reinforced by other communication channels (e.g. academic discourse).

Unfortunately, topic-modeling techniques are not well-suited for analyzing multi-party dialog. One difficulty is that dialog is often so dynamic that it lacks the statistical regularities that enable topic detection in document corpora. A more serious issue is that the subjective meaning of any given passage may only be clear when the hierarchical context of the overall dialog is also considered [12], so topic models based upon the local, atemporal distribution of words may not be sufficient.

Thus, we have developed an approach from the ground up that may be better suited to tracking the flow of information through online conversation. In our approach we model topics as evolving “communities” of words over sequences of replies using a well-known community identification technique developed in the social networking literature. Our approach allows us to observe both how topics evolve within threads, as well as connect across them.

## INSTANCES OF KNOWLEDGE CONSTRUCTION

Our initial inspiration for TEvA was drawn from a subjective reading of transcripts of teams attempting to solve a murder mystery. In the following, we offer examples of effective and ineffective knowledge construction in different teams.

Both transcripts are taken from teams of five collaborators trying to solve a murder mystery using a threaded-chat-like tool [17]. Each post is broadcast to the team, and there is a 10–20 second lag between when a post is made and when it

| Time (mm:ss) | Author | Content |
|---|---|---|
| 02:41 | P1 | *[talking about Billy]* he was there, picked up thecrowbar and then ran. kinda guilty sounding |
| | | ***--- posts skipped---*** |
| 03:17 | P2 | *[talking about Billy]* the fact that it was found in the bushes makes it seem like he was framed. why would he leave the murder weapon at the scene of the crime? it looks like someone planted it there with the intention of it being found |
| | | ***--- posts skipped---*** |
| 16:54 | P3 | I think Eddie is guilty but I'm still looking for a sufficient motive, probably with his daughter. |
| 17:44 | P2 | I agree |
| 17:54 | P4 | I agree with this, but I want some theory on how he got billy's prints on the crowbar |
| 19:05 | P3 | Good point. I think it's possible that he removed it from the scene of the crime without thinking for some reason |
| 19:07 | P1 | the crowbar was leaning on the garage door so billy had to move it to the side to get the lawnmower out, seeing how billy cam every saturday to mow lawn, eddie could have planned that |
| 19:37 | P2 | BINGO |
| 19:54 | P3 | I didn't know that |
| 20:30 | P5 | AHA |

**Figure 1: Collaborators (P1-P5) constructing a story to solve a murder mystery. Reported time is from the beginning of the conversation.**

appears on all of the participant's screens. Conversations are roughly 30 minutes long and all teams were offered a small prize (movie tickets) if they unanimously came to the correct conclusion about the culprit. Several of the participants in each team have critical pieces of information that the others do not have. Participants were told that they might have non-identical information, but they were not aware how information is distributed or which pieces were critical.

Figure 1 is a portion of conversation extracted from a team that correctly identified the culprit, Eddie. In the first post (at 02:41) P1 reports that Billy picked up the crowbar (the murder weapon), and then ran. This evidence explains how Billy's fingerprints got on the crowbar, but none of the team members seem to notice. In a subsequent post (at 03:17) P2 observes that it was strange that the murder weapon, with Billy's fingerprints, was just lying around near the scene of the crime, and suggests that perhaps he was framed. Once again, no one else picks up on this idea.

The third post in the figure occurs roughly seventeen minutes into the conversation. By this point, several of the collaborators are suspicious of Eddie, but they have no good explanation for his motive or the presence of Billy's fingerprints on the crowbar. The conversation proceeds until the post at 19:07, when P1 repeats (but also reframes) the information she had previously presented in the first post. This time, the other team members pick up on the importance of the information—it becomes clear that Eddie framed Billy by placing the crowbar in his way, forcing him to pick it up.

| Time (mm:ss) | Author | Content |
|---|---|---|
| 11:07 | P1 | I think Eddie killed him heat of the moment over something with his daughter, and then tried to cover it up because he regretted it. |
| 12:01 | P2 | how could he have put billy's finger prints on his crowbar? |
| 12:36 | P1 | Billy might have touched the crowbar when finding the body. |
| 12:52 | P1 | Or Eddie could have paid Billy off to keep quiet. |
| 17:25 | P3 | That's not evidence. |
| 18:48 | P4 | Yes. Billy admitted to touching the crowbar. That doesn't mean he used it to kill Blake. |
| 19:56 | P5 | Did he? I didn't have that. |
| 21:51 | P2 | why would he touch it? |

**Figure 2: A missed opportunity for knowledge construction**

The bits of evidence held by different collaborators are synthesized at this point to become a complete story that explains the murder. The collaborators are aware that this has happened, as indicated by P2's enthusiastic "BINGO" and P5's "AHA." The dynamics of the conversation change in several ways at this point as well. All collaborators are present and engaged, peripheral conversations (in other active threads) die down, and posting speed increases. Conversation at this point is inclusive, focused and excited.

A second example (Figure 2) presents a section of dialog from another group that also successfully solved the same mystery. The passage begins much as the passage in Figure 1; collaborators suspect Eddie is the culprit, but lack sufficient evidence. Yet, when P4 provides the relevant piece of missing evidence (18:48), the others meet it with skepticism, and even P4 himself seems unable or unwilling to integrate it into the emerging story. Ultimately, this group chooses Eddie as the most likely suspect, but never agree upon a consistent story about how the murder occurred.

In both of the preceding examples, individuals bring different pieces of knowledge to the same passage of conversation. But in the first case, something special happens—the participants notice how the pieces of information fit together and synthesize them to create a story that makes sense. The latter example, on the other hand, can best be described as a case of missed opportunity. All of the necessary information is present, but collaborators fail to weave it together.

In the remaining sections, we describe a quantitative approach to identifying such patterns, and then apply this approach to the entire dataset. This analysis will show that the degree to which knowledge convergence correlates with posting dynamics differs significantly between groups using different kinds of decision support. With these findings in hand, we offer an explanation for the differences, and reflect upon the approach and its potential in other online contexts.

## APPROACH

We seek to identify moments of a conversation where (1) collaborators bring multiple pieces of information together, and (2) the dynamics of the conversation (as observed in the posting patterns of users) change. In the following we first describe the Topic Evolution Analysis (TEvA) process, and then describe a composite measure of posting dynamics.

### Topic Evolution Analysis (TEvA)

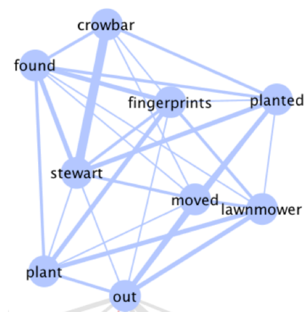
To identify topics in conversation we use patterns of co-occurrence frequencies between words in the text, and model how these frequencies change over time. There are some similarities between our approach and topic tracking in that a topic can ultimately be resolved to a set of co-occurring words found throughout the corpus, and each sample of the corpus is labeled with a set of topics. Unlike many approaches to topic-tracking, which model topics as probability distributions over words in a dictionary, and seek to identify the underlying probability distributions that generate words in documents, we use word adjacency in reply sequences as a strong hint about what a conversational topic is, and then connect these topics across different reply sequences.

Within TEvA, a topic at any point in time is represented as a network of connections between words. The approach is based upon an algorithm for modeling the dynamics of communities in social networks [23]. The procedure works by identifying communities of words—referred to here as topics—in networks extracted at a series of overlapping time intervals, and then describing how these topics change, merge, and split as the network evolves.

The overall procedure is as follows:

1. Split the conversation into overlapping windows.
2. Transform data in each window into a single network for each time slice.
3. Extract topics in each window.
4. Connect topics across windows.

Before any analysis is performed, the entire dataset is preprocessed to remove stop-words and replace alternative spellings with identical tokens (e.g. mower → lawnmower).




| | |
|---|---|
| P1 | His fingerprints were found on Stewart's crowbar. |
| P2 | Stewart may have planted them somehow |
| P3 | I don't think you can plant fingerprints |
| P4 | Yes, but he moved it to get it out of the way of the mower. |

**Figure 3: Extracting the network from a transcript. Only the highlighted words (non stop-words) are processed, and stronger relationships are indicated by link width. The "Stewart-crowbar" relationship is the strongest in this frame.**

Transcripts are then divided into overlapping windows. Allowing windows to overlap helps the algorithm to establish continuity between emerging topics. We expect that different datasets will require different windowing parameters depending on the amount of traffic and nature of the forum. We chose to use ten-minute windows slid over the data in one-minute increments, such that each window overlapped the previous by nine minutes. The size of the window was based on our impressions about the typical lifetime of a topic in our dataset, so that topics not mentioned for ten minutes will be "forgotten" by the topic extraction (step 3) step. The amount of time by which each window was advanced was chosen to be the approximately equal to the average time between posts in any given sequence of replies. Thus, each subsequent window on average replaces one post per thread. More general techniques for choosing a windowing method and parameters are a subject for future work.

Once the data has been split into windows, text within each is transformed into a network using a technique called Wordij [8] (see Figure 3 for an example). In Wordij, strong ties connect adjacent words, weaker ties link words that are adjacent to a common word but are not themselves adjacent, and so forth. This approach may be applied up to some maximum degree of indirection. Within a given time window, the approach is applied separately to each thread (a sequence of replies). The availability of reply sequences is important, because adjacent posts tend to be about the same topic, and adjacency relationships are what drive topic extraction. The resulting networks are merged, adding edge-weights where edges overlap, to reflect the possibility that multiple threads may be about the same topic.

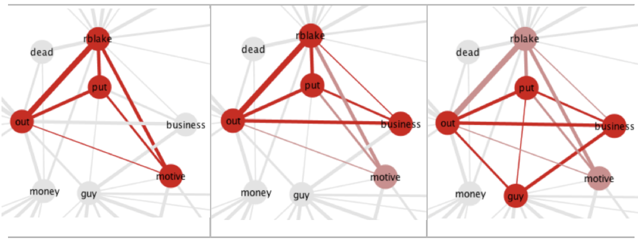



**Figure 4: Rolling a *k=4* clique template across a network to identify a community. The final community consists of three cliques.**

To identify topics in each network, we apply the Clique Percolation Method (CPM; [24]). CPM works by identifying a clique (a fully connected set of nodes) of some pre-determined size, and then replacing a node in this clique to obtain a new, overlapping clique of the same size. This process continues with each new clique, until no new replacement nodes can be found (see Figure 4 for an example). This is euphamistically described as "rolling" a clique template across a network. When a clique can be rolled no further, the boundary of a topic has been found. A complete set of topics can be discovered by exhaustively applying this process to the word co-occurrence network.

In the final step, we seek to connect topics across time-windows. Because some topics may disappear and than reappear, it is necessary at each point in time to consider each previously seen topic as a potential progenitor of a currently observed topic. We make a simplifying assumption that only most recent version of each historical topic is a viable candidate as progenitor of a current topic. Thus, if topic **A** existed from time *t-n-m* (*n>0;m>1*) to time *t-n*, we would only check to see if a connection can be drawn from **A** at *t-n* to a novel topic **B** at time t. This assumption reflects our intuition that knowledge tends to accumulate for small groups that are engaged in a time bounded problem-solving task. For larger, longer-lived groups with more dynamic membership, this may not be a realistic assumption.

A simple approach to connecting topics in subsequent time steps would be to identify those pairs that have the highest relative overlap, where the relative overlap between A and B is defined as:

$$S(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

However, as described by Palla et al [23], this approach will lead to spurious matches in instances where community growth is rapid and many communities overlap. Thus, following the procedure described in [23], the connections to communities at time *t* are established as follows:

1. CPM is used to extract topics from a network formed by merging networks at times *t-1* and *t* along with the most recent version of all topics that have previously been identified. We denote this set of topics as $\mathbf{C}^{merged}$ .
2. We seek to connect the most recent version of topics

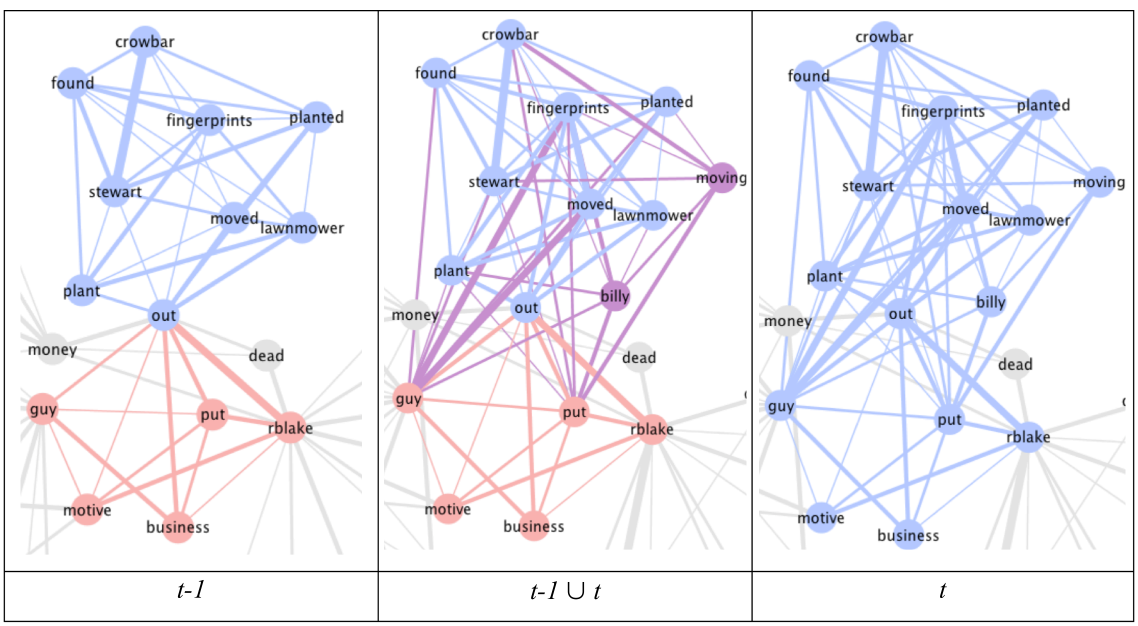



**Figure 5: Establishing the connection between topics in adjacent time windows. The network at time *t* adds numerous links, causing the two topics at time *t-1* to become joined together. Note that because there is a greater overlap between topic A at *t-1* and topic identified at *t,* the new community adopts topic A's label.**

that have already been identified, $\mathbf{C}^{u<t}$, to the set of topics found in the current time window, $\mathbf{C}^{t}$ as follows. For each community $c_j \in \mathbf{C}^{merged}$:

a. Identify each community $c_i \in \mathbf{C}^{u<t}$ and $c_k \in \mathbf{C}^{t}$ whose links are completely contained in $c_j$.

b. Compute $S(c_i, c_k)$ for all pairs ( $i,k$ ).

c. Order pairs by descending score, and draw connections between the pairs of communities until all topics $c_i$ and $c_k$ have been accounted for.

3. Assign labels to the communities at time $t$ according to the connections. Where a single community $c_i \in \mathbf{C}^{u<t}$ is determined to be the progenitor of $c_k \in \mathbf{C}^{t}$, $c_k$ is simply assigned the label of the progenitor, indicating that the community has evolved. In cases where multiple communities in $\mathbf{C}^{u<t}$ are connected to a single community in $\mathbf{C}^{t}$, the destination community is labeled with the $c_i \in \mathbf{C}^{u<t}$ that accounts for the largest proportion of members.

The output of the algorithm is a set of topics that exist for some length of time during the conversation, and a set of links that indicate how different topics are connected. Topics can appear or disappear, evolve, merge, and split (Figure 6).

A final step in the analysis attempts to label the actual chat traces with the discovered topics. This is a simple matter of using the networks originally developed for each window and each thread, and finding the best match (in terms of relative overlap) with the extracted topics. At most one topic for each distinct thread in a time window is chosen.

### Integrated Collaborative Intensity

As described with our knowledge construction example above (Figure 1), collaborative insights among the murder-mystery solving teams analyzed in this article often seemed to occur during portions of conversation where:

1. Posting activity among a team of collaborators is focused in one thread,
2. Posting speed increased, and
3. Most of the users were directly involved in the conversation.

Together, (1) and (3) indicate that a majority of the collaborators are contributing and attending to the conversation. A number of recent studies suggest that such

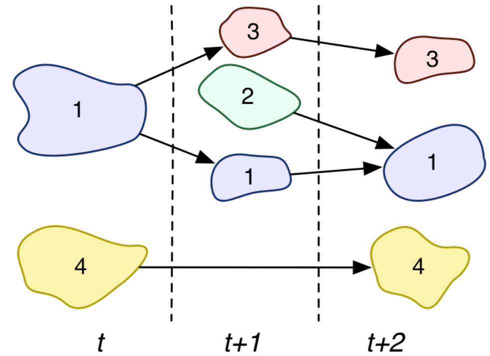



**Figure 6: Illustrating the types of relationships between topics at different time steps. Topics can spawn new children (1➔3), be consumed (2➔1), evolve (3), or disappear and then reappear (4).**

group dynamics are indeed important variables for group performance. For instance, Woolley, et al. [36] found that groups that have more balanced interaction patterns (i.e. lower variance in conversational turns among team members) are more collectively intelligent.

We developed a measurement, which we refer to as *integrated collaborative intensity* (ICI), which combines the preceding factors. The observation that activity is focused on one thread is quantified as $\theta_w$, which is the inverse of the number of active threads (a thread is considered to be active in a time window $w$ if there is a post in that thread in that time window) normalized by the maximum number of threads in any time window.

The speed of posting, $\alpha_{tw}$, is determined by the number of posts in topic $t$ at time window $w$, normalized by the maximum number of posts in any window in the conversation.

Finally, we adopt a measurement of *communication integration* ($\mu_{tw}$) for every topic in every time window to describe how balanced and inclusive the conversation is [1]. The value is at a minimum where no team member talks to another, and at a maximum where everyone speaks directly to everyone else. In online forums where a sequence of replies may be established, communication integration is derived from the graph of who replies to whom.

More precisely, let $l_{ij}$ be the smallest number of links between team members $i$ and $j$ in the reply graph for the section of conversation that is under consideration. The longest possible chain between any two members in a team of N members is N-1; if no chain exists between two members, we set the path length to N. Thus, following [1], communication integration is defined as:

$$\mu_{tw} = 1 - \frac{1}{N^2} \sum_{i \in \{1,2 \ldots N\}} \sum_{j \neq i} \frac{l_{ij}}{N-1}$$

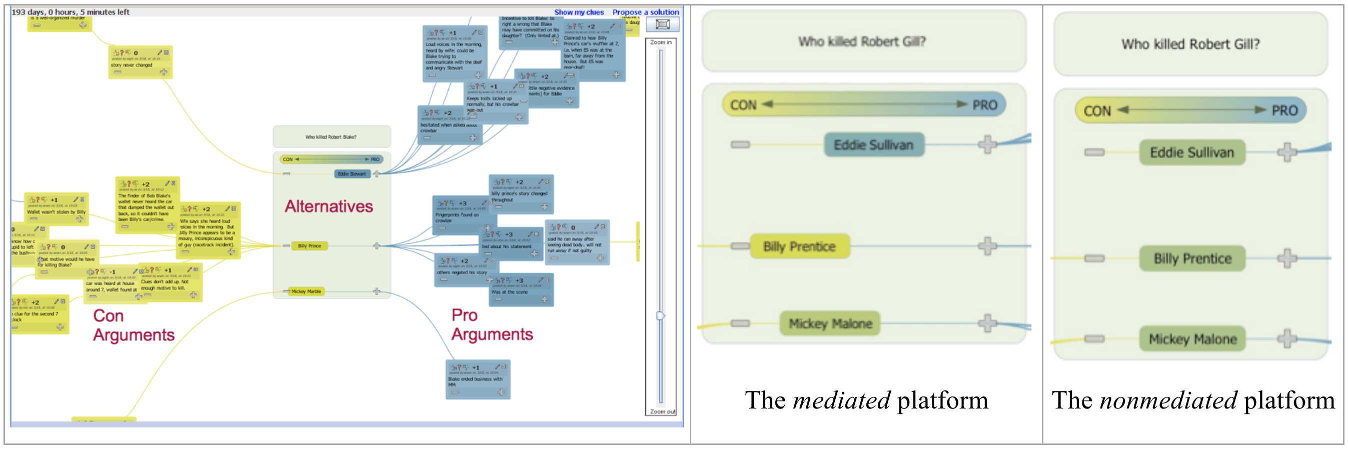



**Figure 7: The platform interface. Both platforms used the same a graph-based interface. top level implicating arguments drifted to the right. In the mediated platform, cells representing the three suspects moved to the left or right depending upon the system's assessment of the suspect's guilt. In the non-mediated platform, these cells were anchored in the middle.**

We adopt communication integration instead of simply counting the number of distinct users in a time window because we are interested not only in the number of unique sources of information, but also how that information might flow between people to create new collaborative insights. Our hypothesis is that such insight is most likely when online collaborators see different pieces information physically juxtaposed, but it is possible that any single individual might also carry information between two others. Thus, communication integration offers a rough approximation of the likelihood each team member will have been exposed to each other team member's information in a section of the conversation.

Each of the constituent measurements is normalized to the interval [0-1], so we express ICI for topic $t$ in window $w$ simply as:

$$ICI_{tw} = \theta_w \alpha_{tw} \mu_{tw}$$

## DOMAIN

We used TEvA and ICI to analyze a dataset introduced at the beginning of this article. The dataset is derived from a controlled study that examined two versions of a decision platform [17]. The purpose of the study was to investigate whether a particular form of decision support could help groups avoid a pitfall of group decision making called the *common knowledge phenomenon*[31]. This phenomenon describes groups' tendencies to focus upon knowledge held by a majority of participants, and discount information held by a minority.

Both versions of the decision support platform were similar to a threaded chat forum with some specialized features (Figure 7). Each thread of discussion was anchored at a decision option (one of the suspects), and users were required to express whether a top-level post was for (implicating) or against (exonerating) the decision option it was attached to. Respondents to any existing post had to indicate whether their reply agreed or disagreed with that post. Finally, participants could vote for or against anyone else's posts.

The two platforms were identical with the following exceptions. In the *non-mediated* version the platform, teams were provided with the structure described above and a simple tool for negotiating consensus. In the *mediated* version, the system provided continuous feedback about which decision option was winning (based on analysis of the structure of the conversation and user votes). Feedback was provided via a set of sliders that moved according to the system's assessment of guilt for each suspect (Figure 7; see the detail for the two platforms). In addition, the team's final decision was constrained to match the winning option. If a team did not agree with the system, they could continue to deliberate to change the system's assessment.

The study was performed in a lab, and no contact was allowed between users outside of the system. Teams consisted of five people, and their task was to identify the correct suspect in a murder mystery. Team members were provided with a set of affidavits from characters in the mystery. The information contained in these materials was the same across all team members, but three of the team were provided with some additional information no one else had. Participants were told that some team members might have information others did not have, but no one knew which information was unshared. Teams were required to form a consensus by the end of a time-limited decision-making period, or else forfeit an opportunity to win a small prize for making the correct decision.

Twenty teams participated in the study, evenly split between the two platforms. Together, they generated roughly 1200 posts, evenly distributed across the conditions. As reported in [17], the mediated platform was found to improve the correlation of group decisions with the evidence mentioned in their conversations, and hence appeared to help participants avoid the common knowledge bias, but there was no appreciable difference in decision performance.

Subjectively however, teams in either condition appeared to solve the mystery quite differently. Teams with the non-mediated platform tended to pool their information in a single thread and attempted to create a shared story in periods of intense collaborative activity. Individuals in teams using the mediated platform appeared to operate more independently, and seemed less interested in generating a shared story to explain the evidence. We sought to use TEvA and ICI to expand upon these observations.

## ANALYSIS

We applied the TEvA and ICI analyses to the collected dataset. Settings for the CPM algorithm and word co-occurrence network creation were chosen following guidance in [24] and [8], and parameters for the windowing strategy were chosen as described above. In all cases, a time windows were ten minutes long and were slid across the data in one-minute increments. Thus, a thirty-minute thread would contain at twenty windows, and each window after the first would overlap the previous window by nine minutes.

We analyzed our results in three steps. To provide initial validation of our procedure, we manually coded the dataset to extract conversational topics, and then compared the performance of TEvA and two established topic-modeling algorithms against the coded data. We then visualized TEvA's results to gain some insight into the data. Finally we performed a quantitative analysis to help make our more subjective impressions more concrete. The following sections describe each of these steps.

### Validation

The validation of topic modeling approaches is somewhat complicated. Most topic modeling approaches adopt the common machine learning technique of training the probabilistic model with some subset of a corpus, and then testing the ability of the model to predict held-out documents (documents that were withheld from the training set) [33]. There have also been some efforts to determine if topic models produce psychologically meaningful topics [18]. This work has examined both the coherence of topic definitions and their fitness as labels that summarize the contents of a document. These approaches are not as useful in a discourse context where topic definitions change rapidly over time and the documents are very small. Thus, we validated the topics detected by TEvA and their assignment to posts against manually coded data. We also compared TEvA's topic assignments to two established topic-modeling approaches.

To code the data, two coders first worked together to establish a consistent methodology. Two sessions of data were randomly drawn from the corpus and used in this step. Topic assignment proceeded as follows. Within a sequence of replies (a thread), each post was considered to belong to the same evolving topic as long as local coherence was maintained and the thread did not deviate in a significant way from the inferred high-level topic for that thread. Opening and procedural posts (e.g. "Hi guys!" or "We need an answer soon!") were assigned the same topic as the temporally closest content post within a thread. If a post was encountered that for the most part replicated the contents of a thread that began earlier, it was labeled with the earlier thread's topic ID.

To establish inter-rater reliability, the coders processed two additional sessions independently. Because the number of topics and their meanings could vary, we could not apply approaches to inter-rater reliability that use a fixed classification scheme. Therefore, we approached the problem as one of assessing clustering similarity, where documents are clustered by topic ID.

We report two accepted clustering similarity metrics. The adjusted Rand index [16] compares how different clustering procedures assign pairs of items (posts, in our case) to the same or different clusters (topics). The adjusted Rand index ranges is equivalent to Cohen's Kappa [5,34], and is appealing because it is easy to interpret. However, it has the drawback of assuming a null hypothesis in which the two clusterings are drawn randomly from all possible partitions with a fixed number of clusters and fixed number of elements per cluster[22]. This is not a realistic assumption in our domain.

Because of this drawback, we also report Variation of Information (*VI*), which is a metric that addresses problems with the adjusted Rand index but is somewhat harder to interpret. *VI* is an information theoretic criterion that measures the distance between two clustering procedures based on the information gain and loss by switching from one to another method [22]. *VI* ranges from 0 to $log(n)$ with $n$ equal the number of items to be clustered. Smaller values of *VI* indicate smaller distances between the results of two algorithms.

The two coders achieved a mean adjusted Rand index of .93 and *VI* of .05. Coders coded one session consisting of 88 posts exactly the same, and disagreed about the topic assignment for five posts in the 64 posts of the other session. This result satisfied us that the coding methodology was sufficiently robust, and the two coders split the remaining sessions to code the entire corpus.

Using the manually coded data as ground-truth, we compared the performance of TEvA and two probabilistic topic-modeling approaches: latent Dirichlet allocation (LDA) [4] and correlated topic models (CTM) [3]. We used the sliding window approach used with TEvA to prepare documents for LDA and CTM, and followed the procedure suggested by [13] to retrieve topics. Because it is necessary to specify the number of topics in advance for both LDA and CTM, we estimated models ranging from 5 to 40 topics (the maximum number of topics identified by the coders in any given session) and used the model yielding the best results for each session in our analysis.

Topics were assigned on a per window, rather than per post, basis. To compare the results of topic extraction to the manually coded data, it was therefore necessary to assign each post to a topic. To do this we labeled each post with the topic label that was assigned to the greatest number of time windows that the post appeared in. In the case of ties, we used the topic that reached its highest frequency first.

| | $TEvA_{mean}$ | $LDA_{mean}$ | t | p |
|---|---|---|---|---|
| **Adj. Rand** | 0.85 | 0.67 | 6.36 | <.001 |
| **VI** | 0.10 | 0.26 | -9.27 | <.001 |
| | $TEvA_{mean}$ | $CTM_{mean}$ | t | p |
| **Adj. Rand** | 0.85 | 0.68 | 6.72 | <.001 |
| **VI** | 0.10 | 0.27 | -10.17 | <.001 |

**Table 1: Performance differences between TEvA and other topic extraction approaches (TEvA vs. LDA, TEvA vs. CTM) . In both cases, a two-tailed paired t-test was used. Sample size *m*=20 sessions.**

Table 1 reports mean values of the adjusted Rand index and *VI* for TEvA, LDA and CTM compared against the manual coding. The results show that TEvA performs well, and is significantly more similar to the manual topic assignments than either LDA or CTM. We conclude from this initial validation step that TEvA is an effective method for extracting psychologically meaningful topics from our corpus.

### Visualization

Figure 8 and Figure 9 present visualizations of TEvA's output for groups in the mediated and non-mediated conditions respectively. Time flows from left to right in the graphs, and each vertical line marks a minute of conversation. Each node is a segment of conversation (possibly containing multiple posts) about a topic detected by TEvA, and a node's size is proportional to ICI. Each color represents a different topic. The layout algorithm attempts to keep successive, connected nodes representing the same topic at the same vertical position in the graph until they merge with other topics. In some cases, topics that are active at different times appear at the same vertical position to economize on vertical space. Links indicate how topics evolve and become merged together. Because we are primarily concerned with the convergence of topics, instances where topics split are not shown and are not considered here.

Upon visualizing the data, we discovered an important difference between the mediated and non-mediated conditions. In non-mediated conditions, periods of maximum ICI tended to occur at points in the conversation where collaborators brought together the maximum number of topics. The portion of the transcript provided in Figure 8 traces the evolution of the conversation to such a point. Topic 3 occurs early on in the discussion, and concerns Billy's lying, his fingerprints on the crowbar, and the lawnmower. Topic 8 is a separate discussion about the wife not hearing Billy's loud muffler. Topic 3 merges with another topic (not shown) to become Topic 13, which is about Eddie potentially trying to frame Billy by claiming he heard his muffler. Finally, Topic 17 brings together information about the crowbar, the muffler, and Eddie's possible intent to frame Billy to arrive at the final shared story that Eddie must have framed Billy.

By the end of the session, Topic 17 has collected information from a total of five previous topics. It also reaches the maximum ICI for any point in the conversation at the second to last time window. This corresponds to the behavior observed in the subjective analysis of non-mediated teams, and is the moment where the group appears to form the story that determines the group's final decision.

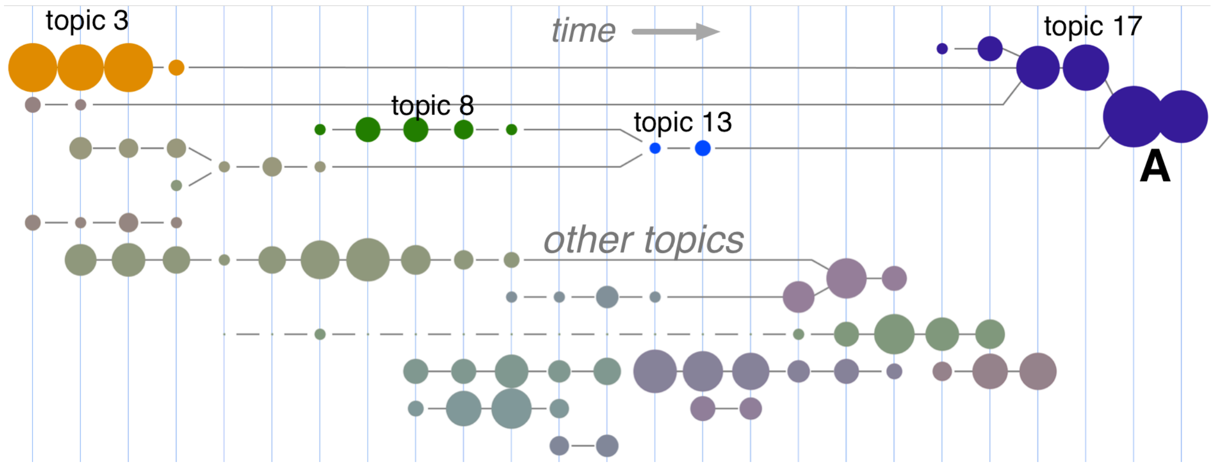
In the mediated groups, periods of high ICI and topic convergence were observed, but they did not occur together as frequently. This is illustrated by Figure 9, which highlights the evolution of the conversation up to the points of highest ICI and maximum topic convergence for a mediated group. Topic 10 begins with a discussion of Billy's motive and theft of the victim's wallet, and this flows into Topic 9, which continues the discussion about the missing wallet. Topic 9 becomes quite heated as participants discuss the importance of the wallet, leading to the highest point of ICI for the conversation.

Topic 10 also flows into Topic 12, which focuses on Marble and his alibi. Topic 12 is informed by Topics 0 and 11, which cover the victim's wife's report of voices outside and Marble's motive. Topic 1 develops in parallel, and is initially concerned with Billy's tire tracks at the house on Saturday morning. Finally, Topic 16, which is about the victim's wife hearing a loud noise outside, merges along with the previously mentioned topics into Topic 1. At the point of convergence, information about Billy's car is brought in to the discussion about Marble to support the argument that Marble did not visit the victim that morning. Five topics flow into Topic 1 at this point, which is the maximum topic convergence for the conversation, but ICI remains relatively low.

Based on our observations about the difference in correlation between topic convergence and ICI across the conditions, we performed a quantitative analysis of both topic convergence and ICI across all sessions, and report the results below.

### Quantitative Analysis

In the preceding analysis, we refer to *topic convergence* as the number of topics that flow into another topic at any given point in the conversation. More concretely, we can define topic convergence $V$ as:



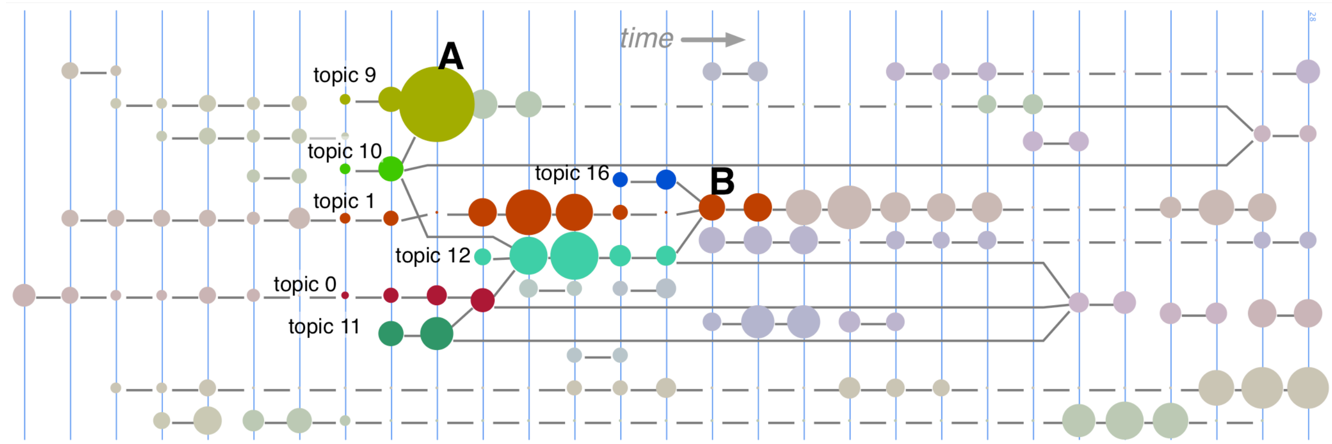

| Topic 3 (Billy) | |
|---|---|
| P1 | billy was VERY withholding of information |
| P2 | Has lied to police and was caught twice. |
| P1 | his fingerprints were on the crowboar, but he denied using it |
| P3 | he's just a stupid scared kid |
| P1 | and it's not his crowbar |
| P1 | then why had he handled the supposed murder weapon? |
| P4 | his fingerprints were on eddie's crowbar that was found in the bushes |
| P1 | He handled it to get to the mower - claims he never used it, though. |

| Topic 8 (Billy) | |
|---|---|
| P5 | does his car have a loud muffler? |
| P3 | mrs blake didn't say anything about a loud muffler |
| P4 | if his car was there also there would have been two cars, they only heard one noise |

| Topic 13 (Eddie) | |
|---|---|
| P1 | eddie said he heard the loud muffler, and said it was billie's. was he trying to frame him? |

| Topic 17 (Eddie) | |
|---|---|
| P5 | very much disregarded the question did he find the crowbar |
| P1 | how would billie's fingerprints get on eddie's crowbar? |
| P2 | Billy said he moved the crowbar to get to the mower. |
| P1 | and would that really require throwing the murderweapo- crowbar into bushes, to be hidden? |
| P3 | no, but eddie might of used the opportunity to frame him |
| P2 | He never said he threw it into the bushes, just that he moved it - Eddit could move it |
| P2 | Eddie also wants to pin this on Billy from the beginning. (muffler) |
| ***(A) Maximum Topic Convergence & Maximum ICI*** | |
| P1 | what's eddie's motivation to frame billy? |
| P2 | It's somebody that is not Eddie. |

**Figure 8: Non-mediated condition – Topic evolution graph with a portion of the conversation it describes. Time flows from left to right in the graph, and each vertical line represents a minute of time. Nodes represent conversation about a particular topic in a time window, and links on the graph illustrate how topics become merged together in time. Node colors distinguish between different topics, and node size reflects ICI. Passages of chat corresponding to the labeled nodes are shown beneath the graph. Point of maximum intensity and maximum convergence is indicated.**




| Topic 10 (Billy) | |
|---|---|
| P4 | Motive? |
| P3 | Always needed extra money. Remember Blake's wallet was stolen, and all the cash inside was taken. |
| P2 | That's a good point, but remember he also saw the body. Maybe he just took the wallet an ran, without committing the murder. |

| Topic 9 (Billy) | |
|---|---|
| P5 | Cash was missing. He's the only one who seems to need money |
| P4 | They give him cash when he needs it. No need to steal. |
| ***(A) Maximum ICI*** | |
| P2 | Stealing the wallet does not mean he committed the murder. |
| P5 | Bob told his wife he was keeping an eye on billy re: money |

| Topic 0 (Marble) | |
|---|---|
| P4 | Motive: Buisiness would be ruined by Blake |
| P5 | Wife heard voices arguing outside |
| P2 | Okay, but she also was alseep and is probably in shock. |

| Topic 12 (Marble) | |
|---|---|
| P1 | called victim Saturday morning |
| P2 | That doesn't suggest he killed the guy. He needed to talk about the golf game. They didn't want to both be there are the same time. |
| P4 | Mrs. Blake heard a car on the driveway at 6:40, could have come and gone in that time. |
| P1 | he left his home at 6:30-6:40 and it takes 15 minutes to get to golf, and he was there |

| Topic 11 (Marble) | |
|---|---|
| P1 | got to his golf game on time |
| P4 | The voices were heard at 6 am, plenty of time to get to the golf game |

| Topic 16 (Billy) | |
|---|---|
| P5 | A very loud noise at 6:40 according to Wife; billy's car is very loud |

| Topic 1 (Billy + Marble) | |
|---|---|
| P5 | Tire Tracks at the house on the day |
| P2 | No, not on the day. He was there the day before. That was confirmed. |
| P3 | The tracks were from Saturday morning. Billy admitted that he went Saturday to Blake's house and found his dead body. |
| | ... |
| ***(B) Maximum Topic Convergence*** | |
| P3 | Eddie identified the car as Billy's because he could hear its loud muffler. Billy admitted to being at Blake's Saturday morning |
| P3 | He was at the golf course by 6:50 |
| P4 | That still fits in with the time frame of him killing Blake |

**Figure 9: Mediated Condition - Visualization of topic evolution with portions of the conversation it describes. Topic 1 spans multiple threads about Billy and Marble. Note that the points of maximum ICI and topic convergence occur at different points in the conversation.**

$$V(t_{iw}) = \frac{|t_j \in Tree(t_{iw})|}{|\mathbf{T}| - 1}$$

For topic $t_i$ in window $w$, $\boldsymbol{V}$ is the number of unique topics in the tree formed by incoming edges and rooted at $t_{iw}$, divided by the total number of topics identified in a session minus one.

We sought to assess the relationship between the type of decision support and the two dependent variables, ICI and $\boldsymbol{V}$. A histogram analysis revealed that both ICI and convergence were exponentially distributed, so a standard analysis of variance (which presumes normally distributed data) was inappropriate. Therefore, we used a variety of tests that are recommended for non-normally distributed data.

| | Mediated (*n=989*) | Non-mediated (*n=921*) | Significance |
|---|---|---|---|
| **ICI (mean)** | .041 | .054 | $p<.01$ |
| ***V* (mean)** | .046 | .060 | n.s. |
| **Correlation (Spearman's $\rho$)** | .27 | .39 | $p<.001$ |

**Table 2. Descriptive statistics for ICI and Convergence (V); *p* values for the individual variables were determined using the Mann-Whitney's *U*. Fisher's transformation was applied to determine the significance of the difference in ρ.**

Table 2 summarizes statistics for ICI and $\boldsymbol{V}$. Both measurements were higher on average in the non-mediated condition. We used the Mann-Whitney test, which is a non-parametric alternative to Student's $t$ and is appropriate for non-normal data [29], to evaluate the significance of this difference. The result indicates the difference in average ICI is highly significant, but that the difference in $\boldsymbol{V}$ is not.

To assess the relationship between $\boldsymbol{V}$ and ICI, we used Spearman's ρ, and compared the difference using Fisher's $z$ transformation ([29]; p. 1071). Spearman's ρ is defined as Pearson's product-moment correlation between ranked variables, and can be interpreted similarly to $r$. The results indicate that correlation between ICI and $\boldsymbol{V}$ is significantly higher in groups that used the non-mediated platform.

To further elaborate these findings, we restricted our analysis to just points of maximum ICI for all groups and ranked the topic under discussion at that point according to $\boldsymbol{V}$. Histograms reporting the percentile rank of the convergence value for the topic under discussion at points of maximum ICI are shown in Figure 10. Seven of the ten groups in the non-mediated condition experienced a maximum value of $\boldsymbol{V}$ at the same time they experienced maximum ICI. In the mediated groups, periods of maximal ICI were more highly correlated with topics that had relatively low $\boldsymbol{V}$ values.

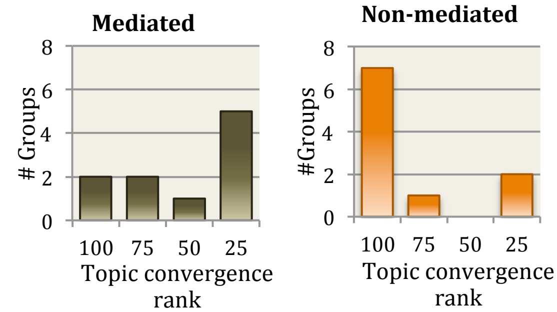



**Figure 10: Topic convergence (V) at points of maximum ICI. Y-axis indicates the number of groups, x-axis the percentile rank of the topic according to V.**

These results confirm our initial impressions upon reading the transcripts. Groups in the non-mediated condition in experienced significantly more excited, focused collaboration as measured by ICI. Groups in either condition experienced a similar degree of convergence, but the degree of convergence correlated more strongly with the intensity of collaboration (ICI) in the non-mediated groups. It was also more likely that the most intense period of collaboration in the non-mediated groups would occur at the same time the group pooled a large amount of their information together.

In the following discussion we offer an explanation of these results, and then reflect upon our analysis methodology.

## DISCUSSION

### Decision Support and Knowledge Construction

Across the non-mediated groups, the passages of high topic convergence and high ICI were strongly correlated. These appeared to be passages where individuals were attempting to combine their information and develop a story that they could all agree to.

There is other evidence that individuals create narratives to solve certain kinds of problems. Pennington and Hastie demonstrated in a series of empirical studies that jurors maintain a mental model of evidence in the form of a narrative, and that these models are predictive both of their memory for evidence and ultimate verdict [25]. Based on these studies, the authors suggested that a group of jurors ultimately decide on a verdict by selecting a story that best satisfies a handful of criteria (e.g. completeness, consistency, plausibility) [15].

Our findings also suggest that story formation can be a natural part of decision making for groups if they are allowed to manage their own decision process. However, this does not appear to be a simple process of selecting the best story. Rather, individuals actively seek to construct their shared stories during highly collaborative phases of discussion.

Results from our previous work indicated that the non-mediated groups appeared to be less rational than the mediated groups because their final decision did not conform to the evidence mentioned in conversation. The analysis presented here offers an explanation; that the shared story is more important than the individual pieces of evidence. Groups may discount evidence that doesn't fit within an emerging consensual story, or infer evidence that completes it.

This is a useful perspective upon a large body of research on the common knowledge phenomenon and information pooling in general (e.g. [9,21]), which has focused on how group information pooling deviates from a rational norm. Studies in this thread of research typically rely upon domain problems in which individual pieces of information should be considered independently, and consistently report on groups' inability to make the "correct" decision under assumptions of independence. The findings have been robust enough (and troubling enough) to make their way into the popular literature [32].

The mediated platform used in this study was designed to correct the common knowledge problem. It intervened in a team's evidence aggregation procedure, and did indeed help groups to appear more rational. Yet our extended analysis shows that it also disrupted a natural collaborative process of story formation. Individuals may have still sought to create stories to explain the evidence, but the dynamics associated with teams weaving their information together to form a shared narrative were absent.

There are a variety of possible reasons for this. The presence of a referee (the mediating platform) may have reduced the scope of negotiation and hence the need for a group to form a common understanding. The system may have also had the appearance of an intransigent partner, and so participants in mediated teams would often simply accede to the opinion reflected by the system as the clock ran out. Furthermore, continuous feedback about which option was winning may have created a sense of competition that dissuaded collaboration.

More subtly, we noted during our analysis that effective narrative creation tended to occur in long, meandering threads where many pieces of information were juxtaposed in close temporal proximity. These threads did not conform well to the argumentative structure enforced by the platform. On the other hand, the combination of features in the mediated platform seemed to encourage posting behavior that did conform to this argumentative structure—threads were shorter and more focused. We might tentatively conclude that adherence to a purely argumentative formalism as a means for both communicating about and evaluating individual pieces of information makes it harder for groups to draw connections across the different pieces of information.

An approach to remedying this problem would be to provide representational support for both narrative and argument structures. This might allow collaborators to remain within the formalism yet still benefit from computational support. It has been noted elsewhere that individuals have tremendous difficulty in bringing evidence to bear upon their working theories [19], and so developing hybrid narrative-argumentative formalisms may ultimately be of great value.

### Theory Driven Design and Technology Impact

In the study described here, simply assessing performance with respect to the common knowledge problem would not have uncovered the impact of the technology on collaborative knowledge construction. This reaffirms a more general lesson that is well understood by ethnographers, activity theorists and practitioners of distributed cognition: when evaluating technology that is designed to address a specific aspect of collaboration, it is important to consider how that technology affects all aspects.

How effectively any group of people can assimilate information, generate ideas, or make decisions depends upon a large number of interacting processes. By virtue of these interactions, technology designed to mediate any specific process or property can fundamentally alter the abilities of the group as a whole. We advocate the development of analytical approaches that can use digital trace data to provide a broad view of such system-wide processes, and the application of these when assessing the impact of collaborative and social technologies.

Researchers have begun to develop new methodological frameworks around systemic analyses of this kind (e.g. group informatics [11]). TEvA is a tool in this family, intended to serve as an imaging device for a socio-technical organism. It is our hope that others will begin to use TEvA and similar techniques not only for descriptive purposes, but also to critically evaluate the impact of technology on the health and intelligence of socio-technical systems.

## FUTURE WORK

Our near term goal is to apply TEvA to corpora extracted from online communities that are orders of magnitude larger than the dataset analyzed here. There are challenges in doing this. One difficulty is the algorithmic complexity of the CPM algorithm, which slows exponentially if the underlying network is too densely connected. However, we can take advantage of the fact that networks of words change incrementally to reduce total computation time. If the algorithmic complexity of the CPM algorithm ultimately proves to be an insurmountable hurdle, other community detection algorithms might be used.

Another challenge in applying the algorithm to larger and more varied datasets is the variety of purposes online conversation serves. The transcripts we analyzed here contained highly focused, task-oriented discussion that was used to exchange a relatively small number of underlying ideas. In the wild, message-forums are used for a variety of additional purposes including planning, socialization, and emotional support, and they are informed by an external context that is much richer than the constrained domain of a single decision making task. We are not certain how TEvA will process such a varied substrate, but anticipate that the attempt will lead to new insights.

## CONCLUSION

We have presented an analysis of computer-mediated teams using a novel approach called TEvA. This analysis allowed us to observe the historical evolution of knowledge as it flowed across a team, leading us to perceive the impact of technology in a new way. We concluded that the particular form of decision support under study disrupted a natural collaborative process, even while it improved decision making from a theoretically rational point of view.

We used the example of this study to reaffirm a more general lesson about the design of collaborative technology. The performance of a socio-technical system is governed by a complex set of interacting processes, and it is possible for a relatively small technological intervention to have much broader impacts. To understand these impacts, it is necessary to be able to observe the system as a whole.

The growing availability of digital trace data has enabled a new era of tools that allows us to analyze the global dynamics of collaborative systems in a new way. We hope that TEvA will be a useful contribution to this growing palette of techniques, and that the study presented here helps demonstrate the value and importance such analyses.

## ACKNOWLEDGEMENTS

We would like to thank the reviewers for their careful review of this paper and excellent suggestions. We would especially like to thank Peter Gloor and Rob Laubacher for their many good ideas and continued encouragement throughout this project. Finally we would like to thank Thomas Malone for generously supporting this work at the MIT Center for Collective Intelligence.

## REFERENCES

1. Van Alstyne, M. and Brynjolfsson, E. Global Village or Cyber-Balkans? Modeling and Measuring the Integration of Electronic Communities. *Management Science 51*, 6 (2005), 851 –868.
2. Blei, D.M. and Lafferty, J.D. Dynamic topic models. *Proceedings of the 23rd international conference on Machine learning*, ACM (2006), 113–120.
3. Blei, D.M. and Lafferty, J.D. A Correlated Topic Model of Science. *The Annals of Applied Statistics 1*, 1 (2007), 17–35.
4. Blei, D.M., Ng, A.Y., and Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res. 3*, (2003), 993–1022.
5. Cohen, J. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin 70*, 4 (1968), 213–220.
6. Conover, M.D., Ratkiewicz, J., Francisco, M., Gonc, B., Flammini, A., and Menczer, F. Political Polarization on Twitter. *Networks 133*, 26 (2011), 89–96.
7. Cui, W., Liu, S., Tan, L., et al. TextFlow: Towards Better Understanding of Evolving Topics in Text. *Visualization and Computer Graphics, IEEE Transactions on 17*, 12 (2011), 2412–2421.
8. Danowski, J. WORDij: A word-pair approach to information retrieval. *NIST special publication*, 500207 (1993), 131–136.
9. Dennis, A.R. Information Exchange and Use in Group Decision Making: You Can Lead a Group to Information, but You Can't Make It Think. *MIS Quarterly 20*, 4 (1996), 433–457.
10. Gloor, P.A., Krauss, J., Nann, S., Fischbach, K., and Schoder, D. Web Science 2.0: Identifying Trends through Semantic Social Network Analysis. *Computational Science and Engineering, IEEE International Conference on*, IEEE Computer Society (2009), 215–222.
11. Goggins, S., Mascaro, C., and Valetto, G. Group Informatices: A Methodological Approach and Ontology for Understanding Socio-Technical Groups. *Journal of the American Society for Information Science and Technology*, (2012).
12. Grosz, B.J. and Sidner, C.L. Attention, intentions, and the structure of discourse. *Comput. Linguist. 12*, 3 (1986), 175–204.
13. Grün, B. and Hornik, K. topicmodels: An R Package for Fitting Topic Models. *Journal of Statistical Software 40*, 13 (2011), 1–30.
14. Gunawardena, C.N., Lowe, C.A., and Anderson, T. Analysis of global on line debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing. *Journal of Educational Computing Research 17*, 4 (1997), 397–431.
15. Hastie, R. and Pennington, N. Explanation-Based Decision Making. In *Judgment and Decision Making: An Interdisciplinary Reader*. Cambridge University Press, 2000, 212–228.
16. Hubert, L. and Arabie, P. Comparing partitions. *Journal of Classification 2*, 1 (1985), 193–218.
17. Introne, J.E. Supporting group decisions by mediating deliberation to improve information pooling. *Proceedings of the ACM 2009 international conference on Supporting group work*, (2009), 189–198.
18. Jonathan Chang, Jordan Boyd-graber, Sean Gerrish, Chong Wang, and Blei, D.M. *Reading Tea Leaves: How Humans Interpret Topic Models*. CiteSeerX, 2010.
19. Kuhn, D. *The skills of argument*. Cambridge Univ Pr, 1991.
20. Leskovec, J., Backstrom, L., and Kleinberg, J. Meme-tracking and the dynamics of the news cycle.

*Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM (2009), 497–506.
21. Lightle, J.P., Kagel, J.H., and Arkes, H.R. Information Exchange in Group Decision Making: The Hidden Profile Problem Reconsidered. *Manage. Sci. 55*, 4 (2009), 568–581.
22. Meilă, M. Comparing clusterings—an information based distance. *Journal of Multivariate Analysis 98*, 5 (2007), 873–895.
23. Palla, G., Barabasi, A.-L., and Vicsek, T. Quantifying social group evolution. *Nature 446*, 7136 (2007), 664–667.
24. Palla, G., Derenyi, I., Farkas, I., and Vicsek, T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature 435*, 7043 (2005), 814–818.
25. Pennington, N. and Hastie, R. Reasoning in explanation-based decision making. *Cognition 49*, 1–2 (1993), 123–163.
26. Poole, M.S. Decision development in small groups, III: A multiple sequence model of group decision development 1. *Communication Monographs 50*, 4 (1983), 321–341.
27. Ramage, D., Dumais, S., and Liebling, D. Characterizing microblogs with topic models. *International AAAI Conference on Weblogs and Social Media*, (2010).
28. Ratkiewicz, J., Conover, M., Meiss, M., et al. Truthy: mapping the spread of astroturf in microblog streams. *Proceedings of the 20th international conference companion on World wide web*, ACM (2011), 249–252.
29. Sheskin, D. *Handbook of Parametric and Nonparametric Statistical Procedures*. CRC Press, 2004.
30. Simmons, M.P., Adamic, L.A., and Adar, E. Memes Online: Extracted, Subtracted, Injected, and Recollected. *ICWSM 2011*, (2011).
31. Stasser, G. and Titus, W. Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology 48*, 6 (1985), 1467.
32. Sunstein, C.R. *Infotopia: how many minds produce knowledge*. Oxford University Press US, 2006.
33. Wallach, H.M., Murray, I., Salakhutdinov, R., and Mimno, D. Evaluation methods for topic models. *Proceedings of the 26th Annual International Conference on Machine Learning*, ACM (2009), 1105–1112.
34. Warrens, M. On the Equivalence of Cohen's Kappa and the Hubert-Arabie Adjusted Rand Index. *Journal of Classification 25*, 2 (2008), 177–183.
35. Wise, A.F. and Chiu, M.M. Analyzing temporal patterns of knowledge construction in a role-based online discussion. *International Journal of Computer-Supported Collaborative Learning 6*, (2011), 445–470.
36. Woolley, A.W., Chabris, C.F., Pentland, A., Hashmi, N., and Malone, T.W. Evidence for a collective intelligence factor in the performance of human groups. *Science 330*, 6004 (2010), 686.
37. Zhang, J., Song, Y., Zhang, C., and Liu, S. Evolutionary hierarchical Dirichlet processes for multiple correlated time-varying corpora. *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, (2010), 1079–1088.

.