CS224W
Group #2

Toward Collaboration Sensing: Applying Network Analysis Techniques to Collaborative Eye-tracking Data

Bertrand Schneider
Stanford University
schneibe@stanford.edu

Sami Abu-El-Haija
Stanford University
haija@stanford.edu

Jim Reesman
Stanford University
jreesman@cs.stanford.edu

ABSTRACT

In this paper we describe preliminary applications of network analysis techniques to eye-tracking data. In a previous study, the first author conducted a collaborative learning experiment in which subjects had access (or not) to a gaze-awareness tool: their task was to learn from neuroscience diagrams in a remote collaboration. In the treatment group, they could see the gaze of their partner represented dynamically on the screen. In the control group, they could not. Dyads (group of two students) in the treatment group achieved a higher quality of collaboration and a higher learning gain. In this paper, we describe how network analysis techniques can further illuminate these results, and contribute to the development of 'collaboration sensing'. More specifically, we describe two contributions: first, one can use networks to visualize and explore eye-tracking data. Second, network metrics can be computed to interpret the properties of the graph and make basic prediction regarding dyads' quality of collaboration. We conclude with comments on implementing this approach for formal learning environments.

Categories and Subject Descriptors

K.3.1 [**Computer Uses in Education**]: Collaborative Learning

Keywords

Network Analysis, Joint Attention, Eye-tracking, CSCL.

1. INTRODUCTION

Nowadays, massive datasets are becoming available for a wide range of applications. Education is no exception to this phenomenon: cheap sensors can now detect every movement and utterance of a student. On the web, Massive Open Online Courses (MOOCs) collect every click of users taking classes online. This information can provide crucial insights into how learning processes unfold *in situ* or in a remote situation. However, researchers often lack the tools to make sense of those giant datasets; our contribution is to propose additional ways to explore massive log files, such as eye-tracking data.

Our work can be summarized by 2 main tasks. First, process large log files containing eye-tracking data, and representing their information as graphs. Then, we run graph analysis algorithms to detect patterns in the log files, which correspond to patterns in how subjects jointly gazed at the displayed diagram. Our final goal is to correlate the characteristics of those graphs with the subjects' quality of collaboration during the task.

In other words, our motivation is to find measurable differences between the two test conditions. We aim to model the learning strategies that subjects in the "gaze" condition adopt, and investigate how those strategies differ from what's adopted at the "no gaze" condition. Those patterns can then be used to understand and predict the quality of collaboration in small collaborative learning groups.

2. RELATED LITERATURE

Our work lies in the intersection between basic network analysis and studies of the effects of gaze awareness on collaborative learning. While there is literature in both of these areas, there appears to be none squarely in the intersection of those two domains; as such, we believe the proposed work is novel and relevant to generating insights. We discuss the literature from related areas in order to justify our proposed work. In this section we look briefly at eye-tracking studies on collaborative learning, basic network analysis techniques, and at examples employing simple network analysis of eye tracking data.

First of all, we need to define the main phenomenon of interest for this study. Joint attention is defined as "the tendency for social partners to focus on a common reference and to monitor one another's attention to an outside entity, such as an object, person, or event. [...] The fact that two individuals are simultaneously focused on the same aspect of the environment at the same time does not constitute joint attention. To qualify as joint attention, the social partners need to demonstrate awareness that they are attending to something in common" [20]. Joint attention is fundamental to any kind of social coordination: young infants communicate their emotions by being in a state of synchrony with their caregivers, which in turn helps them achieve visual coordination when learning to speak [19]. Parents use deictic gestures (i.e. pointing at an event or object of interest to establish joint visual attention) to signal important features of the environment to their children [2]. Professors and mentors teach by highlighting subtle nuances between students' and experts' conceptual understanding of a domain [16]. Groups of students rely on the coordination between its members to reach the solution of a problem [1], which in turn impacts their level of abstract thinking [18].

Previous studies in CSCL (Computer-Supported Collaborative Learning) have used eye-trackers to study joint attention in collaborative learning situations. For instance, Richardson & Dale [15] found that the number of times gazes are aligned between individual speaker–listener pairs is correlated with the listeners' accuracy on comprehension questions. Jermann, Nuessli, Mullins and Dillenbourg [9] used synchronized eye-trackers to assess how programmers collaboratively worked on a segment of code; they contrasted a 'good' and a 'bad' dyad, and their results suggest that a productive collaboration is associated with high joint visual recurrence. In another study, Liu [12] used machine-learning techniques to analyze users' gaze patterns, and was able to predict the level of expertise of each subject as fast as one minute into the collaboration (96% of accuracy). Finally, Cherubini, Nuessli and Dillenbourg [5] designed an algorithm which detected misunderstanding in a remote collaboration by using the distance between the gaze of the emitter and the receiver. They found that if there is more dispersion, the likelihood of misunderstandings is increased. In all those studies, however, no data-mining techniques were used to uncover more complicated patterns. We

thus propose to build large networks based on eye-tracking data. Our work deals mainly with basic graph property determination, since it is an exploratory attempt at building networks on top of gaze movements. This includes but is not limited to network size, degree distribution, clustering coefficient, and so forth [7]. By analyzing the attributes of the networks, we lay the foundation for future research, which can control for various network properties to determine their effect on study outcomes.

By understanding subjects' gaze patterns via network analysis techniques, we hope to shed new light on collaborative learning processes. In the next section, we describe our dataset and our attempt at modeling it in terms of a series of networks.

3. THE CURRENT STUDY

The first author previously conducted an experiment where dyads of students remotely worked on a set of contrasting cases [17]. The students worked in pairs, each in a different room, both looking at the same diagram on their computer screen. Dyads were able to communicate through voice. Their goal was to use the displayed diagram to learn how the human brain processes visual information (Fig. 2). In the "gaze" condition, members of the dyads saw the gaze of their partner on the screen; in the control "no gaze" group, they did not see their partner's gaze. This intervention helped students achieve a higher quality of collaboration and a higher learning gain compared to the control group.

To measure their learning gain, the subjects were asked to participate in a test by asking subjects to participate in a test, which measures how much subjects learned from the diagram. The "gaze" condition dyads outperformed those in the "no-gaze" condition for the total learning gain: $F(1,40) = 7.81$, $p < 0.01$.

After each experiment, each participant was categorized as the "leader" or the "follower" of the dyad. The leader is the one initiating conversations, and leading the problem-solving processes during the experiment. Followers learned significantly more when they could see the gaze of the leader on the screen.

In addition, dyads from the "gaze" condition exhibited higher collaboration quality. In particular, "gaze" dyads were better at sustaining mutual understanding, pooling information, reaching consensus, and managing time.

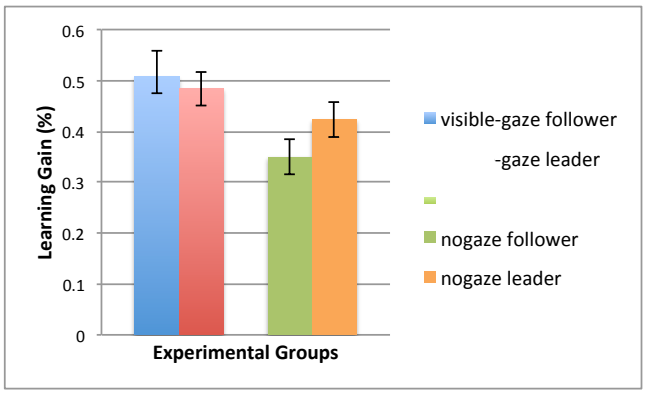

Figure 1: results of the learning test in the previous study. We found that students using the gaze-awareness tool outperformed the students who did not have access to it on the learning test. "Followers" in particular benefited from this intervention.

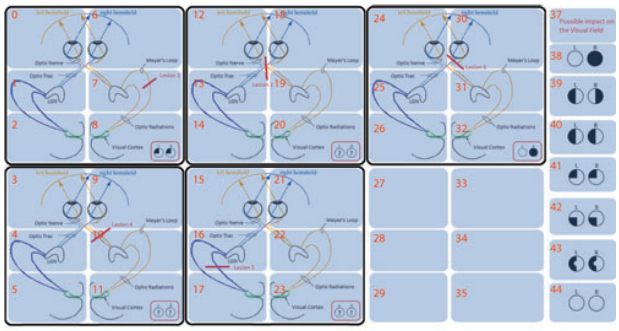

Figure 2: To create the nodes, we choose to divide the screen in 44 different areas based on its visual configuration.

In the next section, we explain how we constructed graphs from the eye-tracking data and how we analyzed them. Then we isolate the attributes that differ between the "gaze" condition and the "no-gaze" condition to gain further insights into our data.

4. CONSTRUCTING GRAPHS WITH EYE-TRACKING DATA

4.1 Goals

Our goal is twofold: first, we want to provide an alternative way to explore eye-tracking data. This approach involves data visualization techniques, such as force-directed graphs. We believe that uses of visualization techniques for representing massive datasets can provide interesting insights to researchers. Previous work has tried to develop visualizations to represent dyads' moments of joint attention [9] (see figure 3); we want to propose an alternative and more intuitive way of visualizing this particular kind of data. Second, we want to compute network measures based on those graphs; our goal is to examine whether some metrics are statistically different across our two experimental groups. Those metrics can provide interesting proxies for estimating dyads' quality of collaboration.

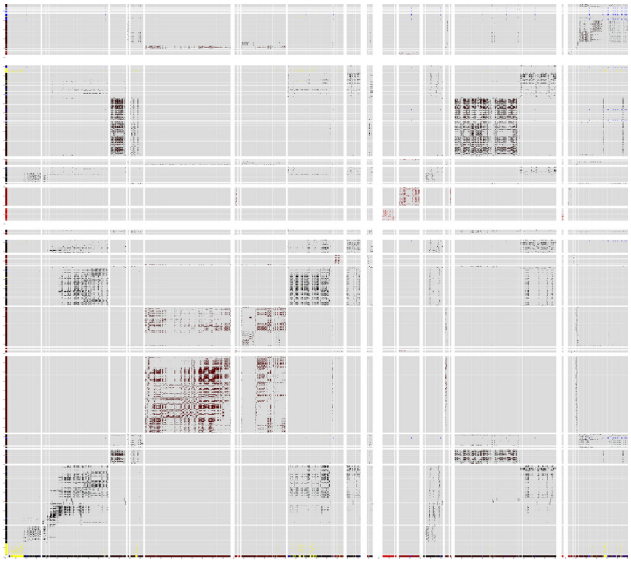

Figure 3: A cross-recurrence gaze plot [9] is the standard way of representing social eye-tracking data in the scientific literature. A dark line on the diagonal means that two collaborators looked at the same area on the screen.

4.2 Movements

To construct graphs from gaze data, we divided the screen into 44 different areas based on the configuration of the diagrams (Fig. 2). Students had to analyze 5 contrasting cases; the answer to the top left and top right cases were given. Possible answers were given on the right. Thus, we have 30 areas that cover the diagrams of the human brain and 8 areas that cover the answer keys.

Edges are created between nodes when we observe saccades between the corresponding screen regions. The weight of an edge is proportional to the number of saccades between the corresponding screen end-points.

In this section we describe graphs created with individuals as the unit of analysis: each network is built by using the eye-tracking data of one subject (Fig. 3). The label on each node corresponds to a screen region as defined in Fig 2. The size of a node shows the number of fixations on this area. Node colors correspond to screen section. Blue nodes correspond to a diagram region (major/left side of the screen). Orange nodes correspond to answer keys (right column of the screen). An edge represents saccades between two regions. The width of an edge shows the number of times a subject compared those two regions. Those graphs were implemented with a force-directed layout and can be directly manipulated on a web page[1].

This approach already shows interesting results: we can observe that subject 1 (on the left) spent a lot of time understanding the diagram on the top right corner of the screen; however (s)he mostly neglected the answers on the right. Subject 2 (on the right), had a completely different strategy: (s)he intensively compared answers and diagrams. Thus, with this visualization one can quickly identify patterns and build hypotheses to investigate.

One limitation of this approach is known as the "hair ball" problem in data visualization: since the graph is quite dense, every node is connected to a lot of other nodes and thus makes interpretations difficult. This problem is inherent to eye-tracking dataset: since an edge is a saccade, each node is going to be connected to at least two other nodes. Moreover, due to the limited amount of potential nodes, our graphs are bound to be highly connected and highly clustered. We then tried to use standard data visualization techniques to facilitate the interpretation of our graphs. One attempt at solving this problem involved creating "edge-bundling graphs", where nodes are arranged on a ring and edges are bundled together to show strong connectivity between vertices. This approach was unsuccessful at isolating key patterns, unfortunately[2]. Another attempt tried to apply the *Prism* technique [10], also without success[3]. Graphs looked similar in both conditions and did not show interesting patterns.

Another limitation of our graphs is the fact that this kind of visualization totally ignores the collaborative aspect of the study [17]. In previous results, dyadic synchronization was found to be a critical factor for a positive learning experience. This is why in the next section, we describe how we included the social aspect of our eye-tracking data into our visualizations. We sought to create smaller and more informative graphs by focusing on dyads instead of individuals. Those graphs provide a different window into our dataset.

[1] stanford.edu/~schneibe/cgi-bin/d3/examples/force/force.php

[2] stanford.edu/~jreesman/bundlegallery.html

[3] stanford.edu/~jreesman/conn-inds.html

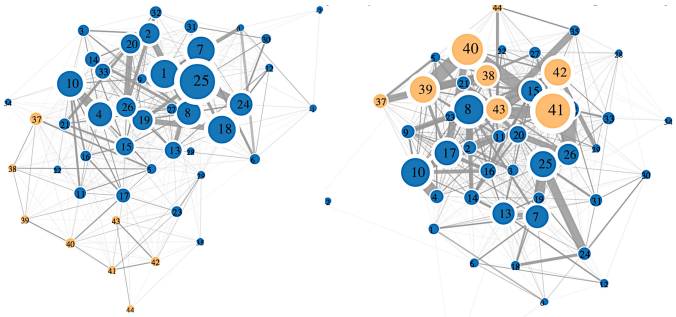

Figure 4: Two graphs based on individuals' data. Blue means brain diagram, orange means answer key on the right of the screen. Both graphs suffer from the "hair ball" problem since they contain a lot of edges.

4.3 At the Dyad Level (Joint Attention)

Our next attempt involved building one graph for each dyad. Here, we want to capture the moments in which dyad members were jointly looking at the same area on the screen. The nodes correspond to the screen areas, as previously defined. At the dyad level, however, a node will only appear in the dyad graph if both dyad members gazed at the corresponding screen area within a 2 second window. Small graphs with few nodes are characteristic of poor collaboration, and large graphs with highly connected nodes show a potential flow of communication across members of the dyad. Figure 4 provides an example of this kind of contrast.

The color scheme of the nodes is the same as that used in the graphs of individual subjects. However, the node size in the dyad graphs is proportional to the number of times dyad members looked at the respective screen area within a 2 second window. The choice of 2 seconds is based on the work done by Richardson & Dale [15], where they find that it takes a *follower* about 2 seconds to look at the screen area that the *leader* is discussing. Edges are defined as previously (i.e., number of saccades between two areas of the screen).

Again, from a data visualization perspective, this approach conveys key patterns in collaborative learning situations. The top graph in Fig. 4 shows a dyad in the "no-gaze" condition; one can immediately see that students did not share a common attentional focus very often. Nodes are small and poorly connected. The graph on the bottom represents a dyad in the "visible-gaze" condition and is a strong contrast to the previous example: here students are looking at common things much more frequently and those moments of joint attention provide opportunities to compare diagrams or answers. Nodes are bigger and better connected.

Based on this new dataset, we computed various network metrics. We found that in the visible-gaze condition, there was significantly more nodes ($F(1,30) = 8.57$, $p = 0.06$), their average size was bigger ($F(1,30) = 22.15$, $p < 0.001$), there was more edges ($F(1,30) = 5.63$, $p = 0.024$), and more reciprocated edges ($F(1,30) = 7.31$, $p = 0.011$). Those results indicate that we can potentially separate our two experimental conditions solely based on network characteristics.

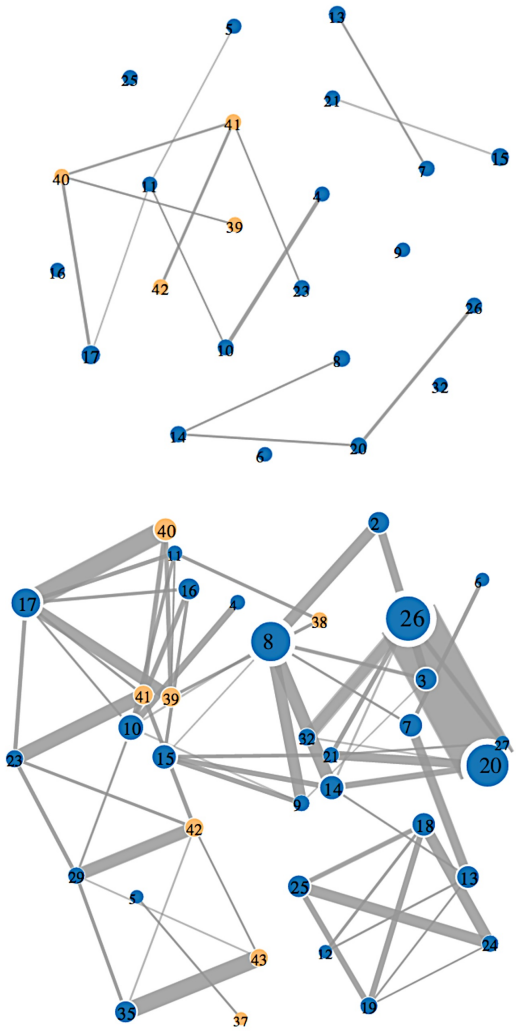

Figure 5: Graph based on dyads' data. The size of each node reflects the number of moments of joint attention members of the group shared on one area of the screen. Graph on the top is from a dyad in the "no-gaze" condition; graph on the bottom is from a dyad in the "visible-gaze" condition.

More interestingly, several measures were significantly correlated with the groups' quality of collaboration (as defined by the rating scheme described in the methods' section [13]; see table 1):

The average size of a node was correlated with the overall quality of collaboration ($r(32) = .64$, $p = 0.039$), and all the sub-dimensions of the collaborative rating scheme. The number of nodes (and edges) in the graph was correlated with the sub-dimensions *Reaching Consensus* ("Decisions for alternatives on the way to a final solution (i.e., parts of the diagnosis) stand at the end of a critical discussion in which partners have collected and evaluated"): $r(32) = .57$, $p < 0.001$, *Information pooling* ("Partners try to gather as many solution-relevant pieces of information as possible"): $r(32) = .52$, $p = 0.002$, and *Time management* (Partners monitor the remaining time throughout their cooperation to finish the current subtask or topic with enough time to complete the remaining subtasks): $r(32) = 0.63$, $p < 0.001$.

Betweenness centrality was the only measure (with the number of nodes in the graph) to be correlated with the sub-dimension "*Sustaining Mutual Understanding*" ("Speakers make their contributions understandable for their collaboration partner, e.g., by avoiding or explaining technical terms from their domain of expertise"): $r(32) = .41$, $p = 0.037$. The largest node in the graph was more sensitive to subjects' orientation toward the task (Each participant actively engages in finding a good solution to the problem): $r(32) = 0.54$, $p < 0.001$. Other measures were correlated only with one sub-dimension, which makes them ideal candidates for making precise predictions regarding the quality of collaboration of a dyad. For instance, the average size of the various strongly connected component in the graphs was correlated only with the sub-dimension *Reaching consensus* ($r(32) = 0.41$, $p < 0.05$). Similarly, the betweenness centrality of the graph was negatively correlated with the sub-dimension *Information pooling* ($r(32) = -0.39$, $p < 0.05$).

As a side node, we also correlated our set of 30 graph metrics[4] with the learning outcomes of the activity (i.e. the results of the post-test filled by the students). The only significant result that we found was that the total number of moments of joint attention was significantly correlated with students' learning gain ($r = 0.39$, $p < 0.05$). This suggests that the kind of graph described above (where nodes are built using dyads' shared attention on an area of the screen) is useful for describing patterns of collaboration, but not so much for predicting learning outcomes. This is why we will focus our attention on understanding and predicting the quality of collaboration of a dyad for the remaning of this paper. More specifically, in the following section we will try to predict collaboration scores using machine learning algorithms.

4.4 Attempt at Interpreting the Correlations Found Between Graphs' Features and Quality of Collaboration

In the section, we provide a first attempt at interpreting the correlations found in table 1. We believe that those graph metrics reflect different collaborative processes. First of all, the average node size seems to be the strongest predictor for our desired outcome (i.e. overall quality of collaboration). This makes sense on a theoretical level: the size of the nodes conveys the number of moments of joint attention of a dyad. From the scientific literature in developmental psychology [4], psychoanalysis [19], the learning sciences [1], and educational cognitive psychology [18], it is a well-established fact that joint attention plays a crucial role in any kind of social interaction. What is interesting is that the raw count of moments of joint attention is strongly associated with an overall high quality of collaboration; additionally, it is also correlated with all its sub-dimensions (table 1). This suggest that merely counting the number of time subjects share the same attentional focus provides a good approximation for the quality of their collaboration.

More specifically, the number of nodes and edges in the graph are associated with the sub-dimensions *information pooling* and *reaching consensus*. Again, it makes sense that the more nodes subjects explore and compare, the better they will be at gathering information and reaching similar conclusions

[4] We gathered a set of 30 measures using built-in functions from the python package networkX (http://networkx.lanl.gov/).

	Average Node Size	Count of Nodes	Count of Edges	Betweenness Centrality (undirected)	Largest Node	Largest SCC	Betweenness Centrality (Directed)	Average SCC size
Dialogue management	**.359***							
Reciprocal Interaction	**.408***							
Division of Work	**.378***							
Sustaining Mutual Understanding	**.411***			**.370***				
Information Pooling	**.531****	**.523****	**.441***				**-.395***	
Reaching Consensus	**.571****	**.652****	**.650****	**.412***		**.393***		**.416***
Time Management	**.633****	**.400***	**.410***	**.363***	**.465****	**.386***		
Task Orientation	**.397***			**.434***	**.539****			
Quality of Collaboration (Total)	**.643****	**.450****	**.415***	**.425***	**.489****			

Table 1: Pearson's correlations of the graphs' features and the 8 dimensions of the collaboration rating scheme (*<0.05, **< 0.01)

The betwennness centrality (i.e. the number of shortest paths from all vertices to all others that pass through a node; in other words, the node's centrality in a network) is more difficult to explain. Mostly because in the directed version of the graph, it is negatively correlated with the sub-dimension *information pooling*; in the undirected version of the graph, it is correlated with the overall quality of collaboration and five of its sub-dimensions. Since the correlations go in two different directions, we currently do not have any relevant explanation regarding this result.

Measures related to the Strongly Connected Components (size of the largest SCC, average size of the SCCs) were associated with a better ability to reach a consensus. We believe that SCCs are likely to represent the clusters on the screen where subjects were working together to solve a sub-problem. For instance, they can be identical regions across different brain diagrams (e.g. compare how the lateral geniculate nucleus is affected in different situations). A small SCC may mean that a dyad shared a moment of joint attention on a sub-region of the screen, but did not connect this node to other components of the graph. Conversely, a large SCC may mean that the dyad worked together on an area of the screen, and then jointly moved to another area on the screen to compare cases or find information to explain the sub-problem. On a higher level, the average size of the graph's SCCs is likely to represent the level of synchronization inside groups.

Finally, the size of the largest node was correlated with the subjects' orientation toward the task; a really large node means that the dyad spent a lot of time focusing intensively on one area of the screen. One can imagine that involving so much attention and effort on one thing reflects subjects' engagement toward the problem at hand.

The following section describes a model that deals specifically with the following question: to which extent did subjects worked together? We tried to measure the extent to which dyads would transition to another area on the screen in a synchronized way.

4.5 Using Machine Learning to Predict Patterns of Collaboration

Using our current dataset, our next goal was to classify dyads into two groups: 1) dyads with a high quality of collaboration, 2) dyads with a lack of collaboration. We divided our dataset in two equal groups using a median split on the overall collaborative score and assigned a dummy variable for each subject (0 = poor collaboration, 1 = good collaboration). Our set of features included the 30 graph's characteristics previously mentioned as well as various demographic data (gender, age, GPA,). Finally, the dataset was completed with a last dummy variable representing the experimental group of the dyad (i.e. "visible-gaze" or "no-gaze" condition). We used three different machine learning algorithms to predict the desired outcome (Naïve Bayes, logistic regression and Support Vector Machine) using a "leave-one-out" cross validation. Since we obtained our best results with SVM (Support Vector Machine [6]), we will only report our prediction accuracy using this technique.

Our results are summarized in table 2. We were able to predict the quality of collaboration using SVM with a multi-layer perceptron (mlp) kernel (93.75% classification accuracy) and applying a forward search feature selection. The algorithm used the following 4 features to make its classification (number in parenthesis indicate the classification accuracy when each feature is added to the model): load centrality (68.75%), size of the largest edge in the graph (84.38%), average degree coefficient (84.38%), nodes' centrality (93.75%).

	Dialogue Management	Reciprocal Interaction	Division of Work	Sustaining Mutual Understanding	Information Pooling	Reaching Consensus	Time Management	Task Orientation	Quality of Collaboration (Total)
Accuracy	96.88%	87.50%	93.75%	100.00%	87.50%	93.75%	90.62%	90.62%	93.75%
Kernel	polynomial	polynomial	polynomial	quadratic	polynomial	mlp	quadratic	polynomial	mlp
Features	7	11	4	6	5	5	20	3	4
Top 3 features	number of nodes with out-degree > 5, nodes' load centrality, nodes' closeness centrality	average node size, square clustering, number of nodes with in-degree < 5	size of the largest node, average degree coefficient, experimental condition	betweenness centrality, average node size, triangle clustering	Average node size, number of nodes, nodes with in-degree smaller than 5	Number of nodes, closeness centrality, triangle clustering	average node size, size of largest node, nodes' centrality	betweenness centrality, closeness centrality, size of largest node	Load centrality, nodes with an out-degree larger than5, circuit average

Table 2: classification accuracy using Support Vector Machine on the graphs' characteristics. Our goal was to predict dyads' level of collaboration (highly productive versus ineffective) for each sub-dimensions of the rating scheme.

It should be noted that those predictions were made *using solely graphs' measures*. When using additional information (e.g. demographic data, dummy variable representing the experimental condition of each subject), we reached a classification accuracy of 100% for the overall quality of collaboration.

When looking at the rating scheme's sub-dimensions, the performances of the learning algorithm were similar. We found a 96.88% classification accuracy for *Dialogue Management* (7 features, polynomial kernel), 87.50% accuracy for *Reciprocal Interaction* (11 features, polynomial kernel), 93.75% accuracy for *Division of Work* (4 features, polynomial kernel), 100% accuracy for *Sustaining mutual understanding* (6 features, quadratic kernel), 90.62% accuracy for *Information Pooling* (3 features, polynomial kernel), 84.38% accuracy for *Reaching Consensus* (2 features, polynomial kernel), 90.62% accuracy for *Time Management* (20 features, quadratic kernel), and 90.62% accuracy for *Task Orientation* (3 features, polynomial kernel). Averaging over all those results, we can show that for this particular task and dataset, our classification accuracy is around 92.71%.

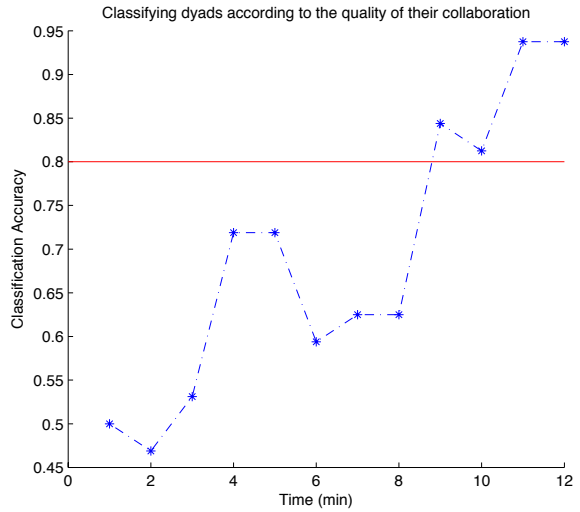

Figure 6: predicting the overall quality of collaboration during the learning activity

4.6 Evolution over Time of our Predictions

Considering the results shown in the previous section, it may not be necessary to wait until the end of the activity to make relevant predictions. In figure 7, we show the evolution of our predictions during the activity using the best learning algorithm for the overall quality of collaboration (SVM with mlp kernel using the 4 specific features described in section 4.6). We see that one minute before the end of the activity, our algorithm already converged to the best classification accuracy (93.75%). Additionally, we reach a classification accuracy higher than 80% three minutes before the end of the activity. This shows that ~10 minutes is the minimum amount of time required by our algorithm to make acceptable predictions.

On a practical level, those results have implications beyond this particular learning activity. With more training data and additional features, one can imagine teachers assessing students' collaboration in real time. This would allow for a more rigorous and informal evaluation of students' abilities to collaboratively solve problems. This is an increasingly relevant question as recent educational reforms start focusing on *21th century skills* [14], such as communication, collaboration, innovation, creativity, critical thinking and problem solving. Using state of the art machine learning techniques may help educators assess students' abilities and thus diagnose the areas where improvement is needed.

5. CONCLUSION

Our preliminary results show the relevance of using network analysis techniques for eye-tracking data. In particular, we found this approach fruitful when applied to social eye-tracking data (i.e., a collaborative task where the gaze of two or more participants are recorded at the same time).

More specifically, we found that different aspects of collaborative learning were associated with different network metrics. The average size of a graph's nodes appeared to be a good proxy for the overall quality of collaboration of the group; the number of nodes and edges in the graph can be used to estimate to what extent dyads try to reach a consensus and pool information to find a good solution to the problem at hand. The size of the largest node in the graph seems to be associated with subjects' orientation

toward the task. Finally, measures related to SCCs (size of the largest SCC, average size of the SCCs) were associated with dyads' efforts to reach a consensus. Of course, more work is needed to replicate those results. But overall, we found that network analysis techniques can be used advantageously to further our understanding of group collaboration processes.

More specifically, we found that applying machine learning algorithms produced interesting results. We were able to classify dyads' quality of collaboration with an accuracy of 92.71% in average (across the various sub-dimensions of the collaboration rating scheme we used). We develop the implications of those results for classroom instruction in the discussion section.

Our work has limitations. First, we studied only one particular kind of collaboration (i.e., remote collaboration). It is likely that *in situ* interactions are different from online collaborative work. In other words, it is unclear how those results can generalize to other collaborative situations. Secondly, we computed those metrics with only 32 students (16 dyads); with more subjects we would probably find more statistically significant patterns. In summary, those results need to be replicated and extended to other collaborative situations.

6. FUTURE WORK

One promising extension of our work is to provide a case study where our graph visualizations help researchers gain insights into their dataset. We believe that network visualizations should complement existing plots and graphs for initial data exploration.

Another direction for future work it to include voice data to the machine learning algorithm. A moment of joint attention can be accidental or coordinated (e.g. via verbal instructions). Differentiating between those two categories would certainly allow our predictions to be more accurate early on during the interaction. Processing the voice characteristics (for instance variation in pitch) would also help us refine our features: certain patterns are known to reflect a high arousal, which can signal dyads reaching an insight.

Finally, the indicators described in table 1 (network metrics correlated with a positive quality of collaboration) should be analyzed in depth to provide further insights into the graph structure. For some of them, it is yet not clear why they are associated with a positive collaboration. A more fined-grained analysis of those indicators would probably provide additional information concerning our dataset.

7. DISCUSSION

This work provides two significant contributions. First, we developed new visualizations to explore social eye-tracking data. We believe that researchers can take advantage of this approach to discover new patterns in existing datasets. Second, we showed that simple network metrics can serve as acceptable proxies for evaluating the quality of group collaboration. As eye-trackers become cheaper and widely available, one can develop automatic measures for assessing people's collaboration. Such instrumentation would enable researchers to spend less time coding videos and more time exploring patterns in their data. In formal learning environments, such measures could be computed in real time; teachers could employ such metrics of 'collaboration sensing' to target specific interventions while students are at work on a task. In informal networked learning, collaboration sensor metrics could trigger hints or provide other scaffolds for guiding collaborators to more productive coordination of their attention and action. We also envision the extension of such network analyses as these for eye-tracking during collaboration to other interaction data related to interpersonal coordination and learning, such as gestures and bodily orientation. This pioneering work could be quickly implemented in classroom as the hardware becomes widely available.

Those results may also have implications beyond the classroom, for instance in any situation resulting in a social construction (e.g. diplomatic compromises, business meetings, group projects, negotiations). As previously mentioned, interpreting and using subtle social signs as predictors may help us define the essential characteristics of a good collaboration in a more nuanced way; and consequently, suggest ways to improve and teach collaborative skills

8. REFERENCES

[1] Barron, B. When Smart Groups Fail. *Journal of the Learning Sciences 12*, 3 (2003), 307–359.

[2] Bates, E., Thal, D., Whitesell, K., Fenson, L., and Oakes, L. Integrating language and gesture in infancy. *Developmental Psychology 25*, 6 (1989), 1004–1019.

[3] Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of research in education*, 61–100.

[4] Brooks, R. and Meltzoff, A.N. Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: a longitudinal, growth curve modeling study. *Journal of Child Language 35*, 01 (2008), 207–220.

[5] Cherubini, M., Nüssli, M.-A., & Dillenbourg, P. (2008). Deixis and gaze in collaborative work at a distance (over a shared map): a computational model to detect misunderstandings. *Proceedings of the 2008 symposium on Eye tracking research & applications, ETRA '08* (pp. 173–180). New York, NY, USA: ACM.

[6] Cortes, C. & Vapnik, V N. (1995). Support-Vector Networks, *Machine Learning, 20* (3), 273-297.

[7] Erdős, P., & Rényi, A. (1960). On the Evolution of Random Graphs.

[8] Freeman, L. (1977). A set of measures of centrality based upon betweenness. *Sociometry* **40**: 35–41.

[9] Jermann, P., Mullins, D., Nuessli, M.-A., Dillenbourg, P. (2001). Collaborative Gaze Footprints: Correlates of Interaction Quality. *In Proceedings of CSCL 2011*. Hong Kong.

[10] Kairam, S., MacLean, D., Savva, M., and Heer, J. GraphPrism: compact visualization of network structure. *Proceedings of the International Working Conference on Advanced Visual Interfaces*, ACM (2012), 498–505.

[11] Kleinberg, J. (2000). The small-world phenomenon: an algorithm perspective. *Proceedings of the thirty-second annual ACM symposium on Theory of computing, STOC '00* (pp. 163–170). New York, NY, USA: ACM.

[12] Liu, Y., Hsueh, P.-Y., Lai, J., Sangin, M., Nussli, M.-A., & Dillenbourg, P. (2009). Who is the expert? Analyzing gaze data to predict expertise level in collaborative applications. IEEE International Conference on Multimedia and Expo, 2009. ICME 2009 (pp. 898 –901). *Presented at the IEEE International Conference on Multimedia and Expo,* 2009. ICME 2009.

[13] Meier, A. et al. 2007. A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning*. 2, 1 (Feb. 2007), 63–86.

[14] Pellegrino, J.W. and Hilton, M.L. 2012. Education for life and work: developing transferable knowledge and skills in the 21st century. *National Research Council, Washington, DC*. (2012).

[15] Richardson, D. C., & Dale, R. (2005). Looking To Understand: The Coupling Between Speakers' and Listeners' Eye Movements and Its Relationship to Discourse Comprehension. *Cognitive Science, 29*(6), 1045–1060.

[16] Roth, W.-M. Gestures: Their Role in Teaching and Learning. *Review of Educational Research 71*, 3 (2001), 365–392.

[17] Schneider, B., & Pea, R. (submitted). Using Eye-Tracking Technology to Support Visual Coordination in Collaborative Problem-Solving Groups. *CSCL 2013.* Madison, USA.

[18] Schwartz, D.L. The Emergence of Abstract Representations in Dyad Problem Solving. *Journal of the Learning Sciences 4*, 3 (1995), 321–354.

[19] Stern, D.N. *The First Relationship: Infant and Mother*. Harvard University Press, 2002.

[20] Tomasello, M. Joint attention as social cognition. In C. Moore and P.J. Dunham, eds., *Joint attention: Its origins and role in development*. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, England, 1995, 103–130.