AUTHOR'S PROOF

Intern. J. Comput.-Support. Collab. Learn.
DOI 10.1007/s11412-014-9202-y

Toward collaboration sensing

Bertrand Schneider · Roy Pea

Received: 10 March 2014 / Accepted: 3 October 2014
© International Society of the Learning Sciences, Inc. 2014

Abstract We describe preliminary applications of network analysis techniques to eye-tracking data collected during collaborative learning activities. This paper makes three contributions: first, we visualize collaborative eye-tracking data as networks, where the nodes of the graph represent fixations and edges represent saccades. We found that those representations can serve as starting points for formulating research questions and hypotheses about collaborative processes. Second, network metrics can be computed to interpret the properties of the graph and find proxies for the quality of students' collaboration. We found that different characteristics of our graphs correlated with different aspects of students' collaboration (for instance, the extent to which students reached consensus was associated with the average size of the *strongly connected components* of the graphs). Third, we used those characteristics to predict the quality of students' collaboration by feeding those features into a machine-learning algorithm. We found that among the eight dimensions of collaboration that we considered, we were able to roughly predict (using a median-split) students' quality of collaboration with an accuracy between ~85 and 100 %. We conclude by discussing implications for developing "collaboration-sensing" tools, and comment on implementing this approach for formal learning environments.

Keywords Collaborative learning · Dual eye-tracking · Network analysis

Introduction Q1

Nowadays massive datasets are becoming available for a wide range of applications, with education no exception: Cheap sensors can now detect every student movement and utterance. Massive Open Online Courses (MOOCs) over the web 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 large datasets; our contribution is to propose additional ways to explore massive log files and describe how collaboration unfolds based on gaze patterns. Eye-tracking data is of

B. Schneider (✉) · R. Pea
Stanford University, Stanford, CA, USA
e-mail: schneibe@stanford.edu

R. Pea
e-mail: roypea@stanford.edu

AUTHOR'S PROOF
JrnlID 11412_ArtID 9202_Proof# 1 - 07/10/2014

particular interest for us, because the technology is becoming ever cheaper and ubiquitous. Several eye-tracking devices are now affordable to the general public, not just to researchers, and there have been multiple interesting attempts at using regular webcams (such as the ones integrated in laptops) to perform basic eye-tracking tasks. Even though the data generated by those low-cost devices is still far from being perfect, there is a trend suggesting that their price is steadily decreasing and their accuracy improving. On the long run, we believe that every single device found in the market will be equipped with some kind of eye-tracking technology.

Given that eye-tracking will become ubiquitous over the next decade, our work pursues three primary aims. First, we want to be able to process large log files containing eye-tracking data and visually represent this information to facilitate the generation of research questions and hypotheses explaining collaborative patterns (cf. the **data visualization** section below). Eye-tracking dataset are generally massive, because eye movements are captured 30–60 times per second. As an example, the dataset we present below contains almost a million data points. There is no way that this amount of data can be interpreted without some kind of data reduction, and data visualization techniques are ideal candidates for this task. Secondly, we want to run graph analysis algorithms to detect patterns in the log files, which correspond to patterns in how subjects jointly gazed at the displayed diagram (cf. the **proxies for rating collaboration** section below). Our final goal is to investigate the relationships between the characteristics of those graphs with the subjects' quality of collaboration during their task (cf. the **prediction of dyads' quality of collaboration** section). Those three contributions are significant, because they contribute to several important areas of research in the CSCL community (e.g., visualizing, analyzing and predicting levels of collaboration in small groups of students).

In the next section, we start by reviewing the literature on using dual eye-tracking setups in collaborative settings. We then introduce the study we conducted to collect our data and describe the measures that we used to rate students' collaboration. Next, we go through each of the contributions mentioned above (data visualization, proxies for rating collaboration, and prediction of dyads' quality of collaboration). We conclude by discussing the implications of each contribution for using multiple eye-trackers in learning environments.

Related literature

Our work lies in the intersection between traditional social network analysis and dual eye-tracking studies in collaborative learning settings. 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 and inspiring future research. We discuss the literature from related areas to justify our proposed work. More specifically, we 1) define visual attention for individuals and small groups of students; 2) review studies that have used dual eye-tracking setup to study social interaction; and 3) look at existing visualizations for representing collaborative eye-tracking data.

In the context of this paper, we are interested in visual attention both for individuals and dyads (groups of two students). For individuals, visual attention is defined as "the behavioral and cognitive process of selectively concentrating on one aspect of the environment while ignoring other things" (Anderson 2004). Visual attention is of particular interest in learning scenarios, because it provides researchers with precise information regarding which resources students processed and which ones they neglected. For dyads, a particularly interesting type of visual attention is when participants synchronize their gaze with their partner (i.e., achieve joint attention). 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" (Tomasello 1995). 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 (Stern 2002). 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 (Bates et al. 1989). Professors and mentors teach by highlighting subtle nuances between students' and experts' conceptual understanding of a domain (Roth 2001). Groups of students rely on the coordination between their members to reach the solution of a problem (Barron 2003), which in turn impacts their level of abstract thinking (Schwartz 1995).

Since collaboration is the main focus of this paper, we concentrate our attention on previous studies in CSCL (Computer-Supported Collaborative Learning) that have used eye-trackers to study joint attention. A foundational work is Richardson and Dale (2005), who found that the number of times gazes are aligned between individual speaker–listener pairs is correlated with the listeners' accuracy on comprehension questions. In another study, Jermann et al. (2001) 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 more joint visual attention. In another study, Liu et al. (2009) used machine-learning techniques to analyze users' gaze patterns, and were able to predict the level of expertise of each subject as rapidly as one minute into the collaboration (with 96 % accuracy). Finally, Cherubini et al. (2008) designed an algorithm that detected misunderstanding in a remote collaboration by using the distance between the gaze of the emitter and the receiver. They found that with more gaze dispersion, the likelihood of misunderstandings is increased. In summary, there are multiple studies showing that computing a measure of joint attention is an interesting proxy for evaluating the quality of social interaction.

Additionally, some prior work has tried to visualize collaborative eye-tracking datasets. The preferred way of looking at how joint attention unfolds over time is by creating *cross-recurrence graphs* (Fig. 1). However, interpreting those graphs is not necessarily obvious for readers unaccustomed to this type of data visualization. To provide a clear and concise description of cross-recurrence graphs, we will quote the excellent explanation from Jermann et al. (2011): "[in a cross-recurrence graph,] the horizontal axis represents time for the first collaborator and the vertical axis represents time for the second collaborator. Each pixel of the plot corresponds to 200 milliseconds time slice (the duration of short gaze fixations are around 100 ms). For a pixel to be colored, the distance between the fixations of the two collaborators has to be lower than a given threshold (70 pixels in our case)." In summary, a dark line on the diagonal represents two collaborators continuously looking at the same area of interest at the time (Fig. 1, right side), while a white or light gray diagonal means no or little joint attention (Fig. 1, left side). Interestingly, those graphs also show when joint attention is preceded and followed by a temporal lag: Dark pixels below the diagonal means that the first collaborator looked at a screen area *after* the second collaborator looked at it, and vice-versa for the pixels above the diagonal (i.e., the second collaborator looked at a screen area *after* the first collaborator).

To our knowledge, however, no prior work has tried to build complex abstractions on top of collaborative eye-tracking data. Prior studies have mostly dealt with raw data and tried to visualize it as cross-recurrence graphs or use it as features for machine learning algorithms. We

UNCORRECTED PROOF

Fig. 1 A cross-recurrence gaze plot (Jermann et al. 2011) 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 screen area. The *left graph* represents a poor collaboration, and the *right graph* represents a "good" collaboration

UNCORRECTED PROOF

thus propose to build large networks where nodes are visual fixations and edges are eye movements between those fixations. Our work deals mainly with basic graph property determination, since it is an exploratory attempt at building networks on top of gaze movements. This emphasis includes but is not limited to network size, degree distribution, clustering coefficient, and so forth (Erdos and Rényi 1960). 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.

The current dataset

We previously conducted an experiment where dyads of students (N=42) remotely worked on a set of contrasting cases (Anonymous for blind review 2013). 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 an audio channel over the network. Their goal was to use the displayed diagram to learn how the human brain processes visual information (Fig. 2). Two Tobii X1 eye-trackers running at 30 Hz captured their gaze during the study. In the "gaze" condition, members of the dyads saw the gaze of their partner on the screen, shown as a light blue dot, and they had the opportunity to disable this overlay by pressing a keystroke (interestingly, none of the students chose to deactivate the gaze awareness tool); in the control "no gaze" group, they did not see the gaze of their partner on the screen. Dyads collaboratively worked on this task for 12 min; they then read a textbook chapter for another 12 min. This text provided them with explanations and diagrams about visual processing in the human brain. The structure of the activity followed a PFL (Preparing for Future Learning; Schwartz et al. 2011) type of learning task (i.e., contrasting cases followed by a standard instruction). Students finally took a post-test and received a debriefing about the study goal. We found that our

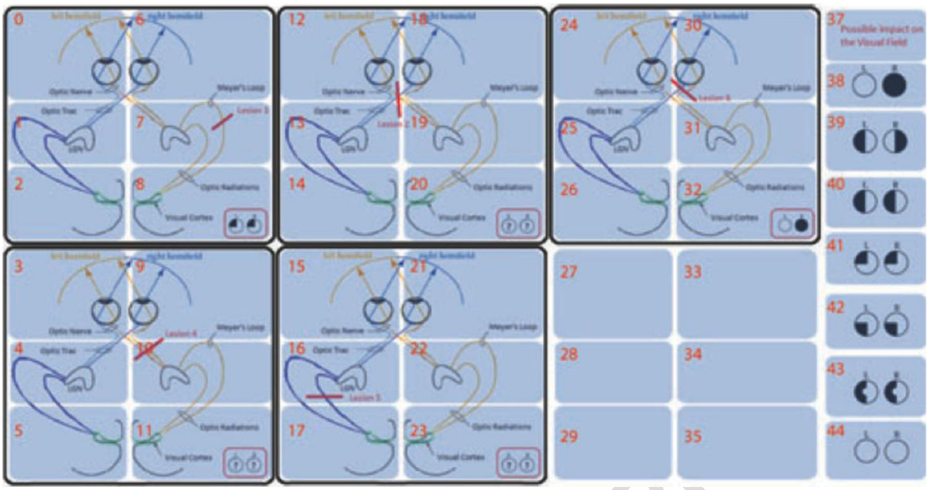

Fig. 2 To create the nodes, we choose to divide the screen into 44 different areas based on the visual configuration of the contrasting cases

UNCORRECTED PROOF

intervention—being able to see the gaze of their partner in real time on the screen with the gaze awareness tool— helped students achieve a significantly higher quality of collaboration and a significantly higher learning gain compared to the control group. Additionally, the two eye-trackers running captured students' eye movements during the study and stored these data as logs; because of technical issues, we only have the complete eye-tracking data for 16 pairs (N=32).

We measured learning gains by using a pre-test and a post-test capturing students understanding of the terminology used, the concepts taught, and their ability to transfer their new knowledge to different situations. We measured collaboration by using Meier et al. (2007) rating scheme. Since our measures of collaboration are central to the analyses conducted below, we describe them in more detail in this section. Meier, Spada and Rummel's rating scheme distinguishes nine dimensions of a social collaboration (see Table 1). At the end of the learning activity, one researcher rates all the dyads using those nine categories and gives each group a score between −3 and +3. In our case, a second judge double-coded 20 % of the video data. Inter-reliability index using Krippendorff's alpha was 0.81 (a value higher than 0.8 is considered as a reliable agreement between judges; Hayes and Krippendorff 2007). Among those nine dimensions, we only considered eight of them because the category "Technical interaction" was not applicable to our experiment: students did not need any technical skill to complete the activity.

Goals

As mentioned above, we have three goals for this paper. The first is to provide an alternative approach for exploring eye-tracking data, involving data visualization techniques, such as force-directed graphs (Fruchterman and Reingold 1991). We conjecture that uses of visualization techniques for representing massive datasets can provide interesting insights to researchers. Previous work has sought to develop visualizations for representing dyads' moments of joint attention (e.g. Fig. 1; Jermann et al. 2011); we want to propose an alternative

AUTHOR'S PROOF JrnlID 11412_ArtID 9202_Proof# 1 - 07/10/2014

t1.1 **Table 1** Rating scheme used to assess students' quality of collaboration (from Meier et al. 2007)

	Dimension	Definition
t1.2	Dimension	Definition
t1.3	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 or by paraphrasing longer passages of text from their materials"
t1.4	Dialogue management	"A smooth "flow" of communication is maintained in which little time is lost due to overlaps in speech or confusion about whose turn it is to talk. Turn-taking is often facilitated by means of questions or explicit handovers."
t1.5	Information pooling	"Partners try to gather as many solution-relevant pieces of information as possible. New information is introduced in an elaborated way, for example by relating it to facts that have already been established, or by pointing out its relevance for the solution."
t1.6	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. arguments for and against the available options."
t1.7	Task division	"The task is divided into subtasks. Partners proceed with their task systematically, taking on one step toward the solution after the other with a clear goal or question guiding each work phase. […] Partners define and take on individual subtasks that match their expertise and their resources. The work is divided equally so none of the collaborators has to waste time waiting for his or her partner to finish a subtask"
t1.8	Time management	"Partners monitor the remaining time throughout their cooperation and make sure to finish the current subtask or topic with enough time to complete the remaining subtasks."
t1.9	Technical coordination	"Partners master the basic technical skills that allow them to use the technical tools to their advantage (for example, they know how to switch between applications, or how to "copy and paste")."
t1.10	Reciprocal interaction	"Partners treat each other with respect and encourage one another to contribute their opinions and perspectives. Critical remarks are constructive and factual, never personal"
t1.11	Individual task orientation	"Each participant actively engages in finding a good solution to the problem, thus bringing his or her knowledge and skills to bear. He or she focuses attention on the task and on task relevant information, avoids distractions"

UNCORRECTED PROOF

and perhaps more intuitive way of visualizing this particular kind of data, e.g., by building networks that represent students' shared visual attention. Our second goal is to compute network measures based on those graphs, so as to examine whether some metrics are significantly different across our two experimental groups. Those metrics (defined on the last page of this paper) can provide interesting proxies for estimating dyads' quality of collaboration. Finally, we try to automatically predict students' quality of collaboration by feeding network features into machine learning algorithms.

In the next section, we provide the rationale for using network analysis techniques as an alternative visualization for exploring eye-tracking data.

Data visualization—Constructing graphs with eye-tracking data

Rationale for using networks to represent collaborative eye-tracking data

The main advantage of using cross-recurrence graphs is being able to analyze the temporal evolution of joint attention in a collaborative group. One can easily determine if a dyad started with a low visual synchronization and progressively became more coordinated; or if a group started with a good synchronization, and then lost their visual coordination because of a

AUTHOR'S PROOF

disagreement or a conflict. The main disadvantage of using a cross-recurrence graph is the inability to analyze *where* dyads jointly gazed during the interaction. There is no way to recover this kind of information from this graph, which limits the hypotheses that researchers can generate when conducting more in-depth analyses. In summary, cross-recurrence graphs display highly granular *temporal* information, but poor *spatial* representation of a group's visual coordination.

Our goal is provide researchers with a complementary representation of a dyad's synchronization: We would like to produce visualizations that show highly granular *spatial* information. Since our goal is not to replace cross-recurrence graphs but to augment them with additional visualizations, we will not focus on including any kind of temporal information in our graphs. We also want to go beyond merely counting the number of times that dyads jointly gazed at the same area of interest (AOIs) on the screen; we want to show how those areas are connected, for instance if students went back and forth between particular diagrams. This is especially important when looking at our learning activity, where students had to analyze contrasting cases: the only way that students can understand the material taught is by *comparing* features of the diagrams shown. We found that networks lent themselves well for this purpose: Fixations are easily represented by nodes, and comparisons between areas of interest (i.e., gaze movements) can be represented by edges in a network. Finally, networks have been intensely studied for the past decades. We can stand on the shoulders of giants by reusing previously defined network metrics, such as network size, density, centrality of nodes, number and properties of sub-graphs, and so on. This allows us to leverage knowledge from other fields of research when analyzing eye-tracking networks for studying collaborative learning.

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

Using fixations as nodes and saccades as edges in a network To construct graphs from gaze data, we divided the screen into 44 different areas based on the configuration of the diagrams learners were shown during the study (Fig. 2). Students had to analyze five contrasting cases; the answer to the top left and top right cases were given. Possible answers were given on the right. Students had to predict the answer of the three remaining cases. We segmented the screen into squares, which provides us with 30 areas that cover the diagrams of the human brain and 8 areas that cover the answer keys.

In our approach, edges are created between nodes when we observe eye movements between the corresponding areas of interest. The weight of an edge is proportional to the number of visual transitions between the corresponding screen end-points.

In this section, we describe graphs created with individuals as the units of analysis: Each network is built by using the eye-tracking data of one subject. 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.

Yet even this basic approach already reveals interesting patterns: 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),

UNCORRECTED PROOF

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 collaborative learning patterns.

One limitation of this data visualization approach is known as the “hair ball” problem (Fig. 3): 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 quantity 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 these graphs. One of our attempts at solving this problem involved creating “edge-bundling graphs” (Selassie et al. 2011), 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. Graphs looked similar in both conditions and did not exhibit any interesting pattern.

Even though this kind of visualization already provides some interesting ways to represent eye-tracking data (Fig. 4), it is unfortunately too dense to provide us with any relevant visual patterns or network metrics that cannot be obtained with simpler methods. One way to reduce the size of those networks is to include the collaborative aspect of the study, by filtering out nodes based on students’ visual synchronization. In previous results (Schneider and Pea 2013), we found the amount of joint attention to be a critical factor for a student’s learning experience. This is why in the next section we describe how we incorporated 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.

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, and edges are defined as previously (i.e., number of saccades between two areas of the screen for an individual).

UNCORRECTED PROOF

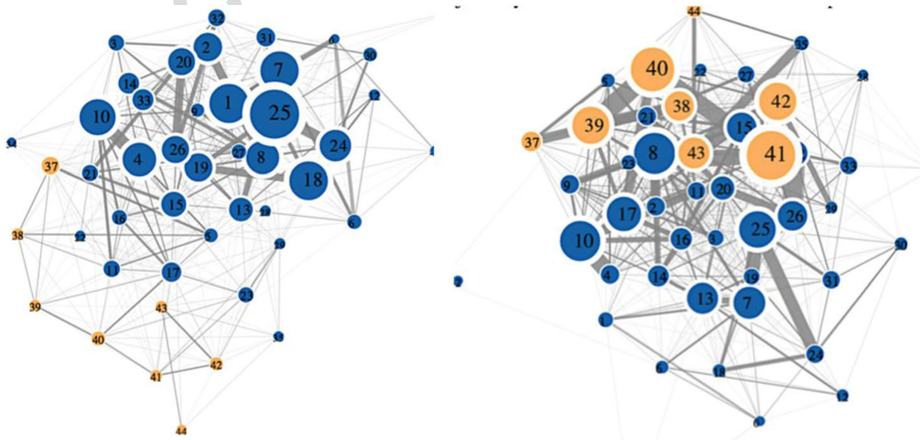

Fig. 3 Two graphs based on individuals’ data. *Blue* means ‘brain diagram’, *Orange* means ‘answer key’ on the right of the screen. Both graphs suffer from the “hair ball” problem since they contain many edges (i.e., each node is connected to every other node in the graph). *Note: in black and white prints, orange will appear as light gray and blue as dark gray*

Fig. 4 The complete set of networks for individuals. One can notice that some networks sometimes have one big node (i.e., one diagram was thoroughly analyzed by a student) and large edges (i.e., two diagrams were intensely compared). Most of them are highly connected (i.e., there are a large number of edges)

Thus, those graphs contain information at both the individual and group level, which is why we create a network for each participant. 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-s 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 5 provides an example of this kind of contrast.

The color scheme of the nodes is the same as used above for 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-s window. The choice of 2 s is based on the work done by Richardson and Dale (2005), where they find that it takes a

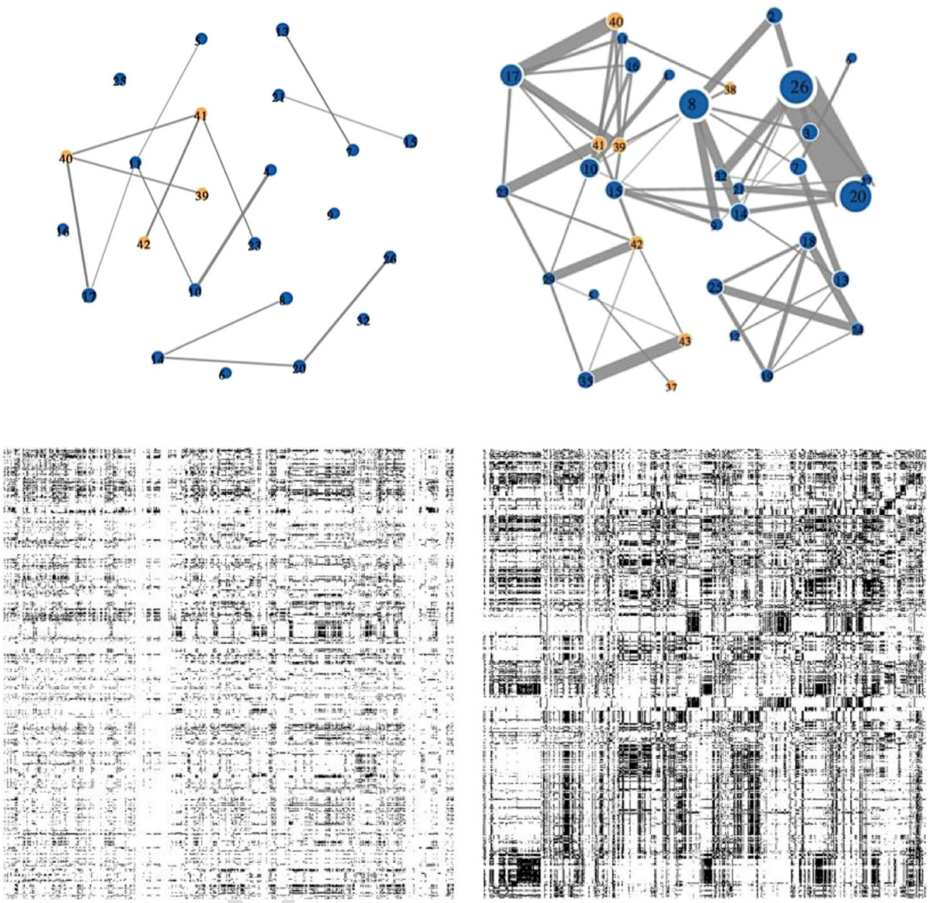

Fig. 5 Graphs based on dyads' data (*top*). The size of each node reflects the number of moments of joint attention members of the group shared on one area of the screen. The graph on the *top left* is from a dyad in the "no-gaze" condition; that on the *top right* from a dyad in the "visible-gaze" condition. Cross-recurrence graphs (*bottom*) are shown for the same two groups as comparison; one pixel represents one second of the collaborative task

follower about 2 s to look at the screen area that the leader is mentioning. Edges are defined as previously (i.e., number of saccades between two areas of the screen for an individual).

Again, from a data visualization perspective, this approach conveys key patterns in collaborative learning situations. The top left graph in Fig. 5 shows a dyad in the "no-gaze" condition; one can immediately see that these students rarely shared a common attentional focus; nodes are small and poorly connected. The graph on the top right represents a dyad in the "visible-gaze" condition and is a strong contrast to the previous example: here students are looking at common items much more frequently and those moments of joint attention provide opportunities to compare diagrams. Nodes are bigger and better connected.

Based on this new dataset, we computed basic network metrics. The variables below satisfied the parametric assumptions of the analysis of variance that we used (i.e., homogeneity of variance and normality). We found that in the visible-gaze condition, there were significantly more nodes (F(1.30)=8.57, p=0.06), with bigger average size (F(1.30)=22.15, p<0.001), 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*.

The main goal of a visualization, however, is to generate insights or hypotheses about a particular dataset. We believe that cross-recurrence graphs and networks allow researchers to generate alternative interpretations of their data. For instance, by looking at the network in Fig. 5 (top right) one can generate the following hypotheses: the strategy of this group seemed to be to compare particular diagram regions (in blue) with answer keys (in orange): for instance, there are several strong connections between area 35 and 43, 29 and 42, 23 and 41, and so on. Additionally, the participants spent a lot of time comparing diagram two and three (as shown by node 26 and 20). When looking at the cross-recurrence graph (Fig. 5, bottom right), one can see that see that there are "clusters" of joint attention along the diagonal (as represented by dark squares). One can hypothesize that participants go through cycles of collaboration: they first jointly analyze an area of the screen (dark section of the diagonal), then explore the other diagrams on their own (light section of the diagonal), and then share their observations with their partner (diagonal becoming dark again). These observations can be used to guide qualitative data analysis when watching the videos of the experiment and for isolating cycles of collaboration.

In summary, the contribution of this section is that we have shown how visualizing dual eye-tracking datasets as networks provides us with information not available on cross-recurrence graphs. Networks encode *where* dyads jointly looked at the same area on the screen, while cross-recurrence graphs describe *when* dyads share a joint attentional focus. The hypotheses that we generated from the visualizations in Fig. 6 show that both graphs can be used in a complementary way to construct hypotheses about collaboration patterns. Another contribution is illustrating how networks are useful when visualizing collaborative eye-tracking data, but of limited use when applied to individuals.

In the next section, we discuss how we computed more complex metrics from those network and how we relate them to the dyads' quality of collaboration. The extensive literature on network analysis (i.e., Erdos and Rényi 1960) provides us with numerous measures that describe relevant networks properties (see Appendix 1 for some examples).

UNCORRECTED PROOF

Proxies for rating collaboration

Furthermore, several measures were significantly correlated with the groups' quality of collaboration (discussed above): the average size of a node was correlated with the overall quality of collaboration (r (32)=0.62, p=0.039), as well as all the sub-dimensions of the collaboration quality rating scheme. The number of nodes (and edges) in the graph was correlated with the sub-dimensions:

(1) **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 arguments for and against the available options"): r (32)=0.71, p<0.001.

(2) **Information Pooling:** ("Partners try to gather as many solution-relevant pieces of information as possible"): r (32)=0.56, p=0.002.

(3) **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)=0.36, p<0.05.

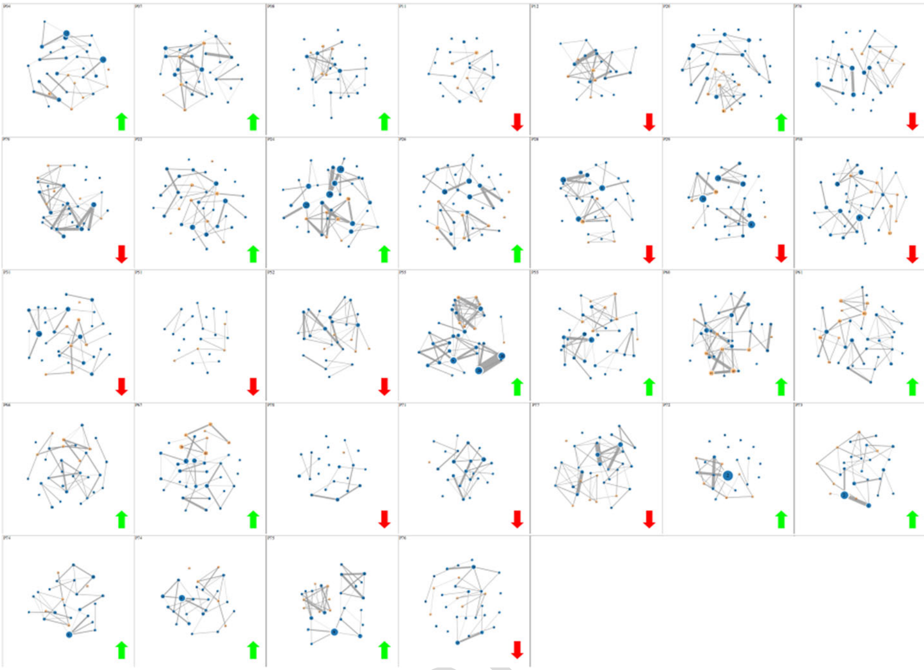

Fig. 6 The complete set of graphs built on the dyads' data. *Upward arrows* mean that the quality of collaboration was above the median split, and *downward arrows* mean below

UNCORRECTED

Betweenness centrality is a measure of a node's centrality in a network. It is equal to the number of shortest paths from all vertices to all others that pass through that node. In our case, the average betweenness centrality of all the nodes of the graph was the only measure 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)=0.42, 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)=0.52, p<0.001, **Reciprocal Interaction** (""): r (32)=0.59, p<0.001 and **Division of Work** (""): r (32)=45, 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 a dyad's collaboration. For instance, in graph theory one can define subgraphs in a particular network; e.g., a subgraph is *strongly connected* if every node is reachable from every other node. Thus, a strongly connected component (SCC) of a directed graph forms a partition into subgraphs that are themselves strongly connected. In our graphs, we found that the average size of the strongly connected component was correlated only with the sub-dimension **Reaching Consensus** (r (32)=0.39, p<0.05). Similarly, the betweenness centrality of the graph was negatively correlated with the sub-dimension **Information Pooling** (r (32)=−0.35, p<0.05).

We note that we also correlated our set of 30 graph metrics with the learning outcomes of the activity (i.e., results of the post-test students completed). The only significant result ascertained was that the total number of moments of joint attention was significantly correlated with students' learning gain (r=0.39, p<0.05). This finding suggests that the kind of graph described above (where nodes are built using dyads' shared attention on an area of the screen),

AUTHOR'S PROOF

while useful for describing collaboration patterns, may not be as useful for predicting learning outcomes. This is why we will now focus our analytic attention on understanding and predicting the quality of a dyad's collaboration.

On Interpreting the correlations found between features of graphs and collaboration quality

In this section we offer an attempt at interpreting the correlations found in Table 2. More specifically, we hypothesize that those graph metrics reflect different collaborative processes. For instance, the average node size appears to be the strongest predictor for our desired outcome (i.e., overall quality of collaboration). This finding makes sense on a theoretical level: the size of the nodes conveys the number of moments of dyadic joint attention. From the scientific literature in developmental psychology (Brooks and Meltzoff 2008), psychoanalysis (Stern 2002), the learning sciences (Barron 2003), and educational cognitive psychology (Schwartz 1995), it is a well-established fact that joint attention plays a crucial role in any kind of social interaction. What we find intriguing 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 2). This suggests that merely counting the number of times 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 Collaboration Quality 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.

It is more difficult to account for the finding for betweenness centrality (defined as 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). This is principally challenging to explain 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 compelling account for this result.

Measures related to the Strongly Connected Components (SCCs; sub-graphs where there is a path from each vertex in the graph to every other vertex) exposed interesting patterns. Both the size of the largest SCC, as well as the average size of the SCCs, was positively associated with greater success in reaching a consensus. We expect that SCCs are likely to represent the clusters on the screen where subjects were working closely 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 for 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 together intensively on one area of the screen. One can imagine that devoting so much attention and effort to one place reflects subjects' engagement toward the problem at hand.

The complete correlation matrix can be found at the end of this paper (Appendix 2). It should be noted that we followed Rothman's advice (1990) to not adjust our results for multiple comparison, since we are conducting exploratory data analysis (as opposed to

UNCORRECTED PROOF

AUTHOR'S PROOF

t2.1 **Table 2** Spearman's correlations of the graphs' features and the eight dimensions of the collaboration quality rating scheme (*<0.05, **<0.01)

		Joint attention (percent of time)	Avg. node size	Count of nodes	Count of edges	Betweenness centrality (undirected)	Largest node	Largest SCC	Betweenness centrality (Directed)	Avg. SCC size
t2.3	Dialogue management									
t2.4	Reciprocal interaction						0.587**			
t2.5	Division of work		0.356*				0.452**			
t2.6	Sustaining mutual understanding		0.418*			0.418*				
t2.7	Information pooling	0.627*	0.589**	0.561**	0.480*				−0.351*	
t2.8	Reaching consensus	0.679**	0.625**	0.707**	0.653**	0.441*	0.475**	0.356*		0.386*
t2.9	Time management	0.690**	0.648**	0.363*	0.410*	0.374*	0.537**	0.372*		
t2.10	Task orientation		0.413*	0.364*		0.394*	0.582**			
t2.11	Quality of collaboration	0.678**	0.617**	0.513**	0.388*	0.478**	0.518**	0.358*		

The first column (joint attention) serves as a baseline. See Appendix 2 for the complete correlation matrix

UNCORRECTED PROOF

hypothesis testing). As a consequence, there is a possibility that some of those results may be due to chance. For this reason, we stress that future work needs to replicate those results before our approach can be proved to detect multiple levels of collaborative work.

The contribution of this section is as follows: Our results suggest that network metrics are more powerful and more accurate than simply computing proportion of joint attention between two participants. Why? They are more powerful because the average node size of our graphs is correlated with most dimensions of our rating scheme (whereas the percentage of joint attention correlates only with some of them). One explanation is that the former measure takes into account the *dispersion* of joint attention on the screen, while the latter only considers whether or not two participants are gazing at the same area. It is likely that a good, dynamic collaboration is more likely to explore the problem space as much as possible rather than just jointly looking at a few screen regions. Our networks make this distinction possible. Our networks are also more accurate, because the network metrics shown in Table 2 are more sensitive to the various facets of a good collaboration: for instance, betweenness centrality and the average size of a SCC allows us to potentially discriminate between groups' tendency to pool information and/or reach consensus. Only using the proportion of joint attention, in contrast, does not allow us to discriminate between two dimensions because it correlates with both aspects of students' collaboration.

In the following section, we will seek to predict collaboration scores using machine-learning algorithms. Since our network metrics seem to be useful measures for predicting students' quality of collaboration, we hypothesize that feeding them into a supervised machine-learning algorithm should lead to accurate predictions. We acknowledge in advance that our dataset is rather small for this purpose and that our model is likely to over fit our training data. Nevertheless, we still believe that it is a reasonable first step in our overall research agenda to predict quality of collaboration in student dyads.

Prediction of dyads' quality 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 into 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 characteristics of graphs 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, Support Vector Machine) using a "leave-one-out" cross validation. Since we obtained our best results with SVM (Support Vector Machine; Cortes and Vapnik 1995), we will only report our prediction accuracy using this technique. In summary, our dataset had 32 rows (16 dyads) where members of a particular dyad had the same nodes but different edges. The output of our classification was a binary score reflecting our prediction for the subjects' quality of collaboration during the task. To minimize over fitting, we used a Leave-One-Out Cross Validation procedure (LOOCV) and repeatedly trained our model on N-1 rows (training data) and predicted the category of the remaining row (test data). The LOOCV procedure ensures that our model doesn't completely over fit the data and generalizes to new, unseen examples. Our results are summarized in Table 3.

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 four features to make its classification

UNCORRECTED PROOF

AUTHOR'S PROOF JrnlID 11412_ArtID 9202_Proof# 1 - 07/10/2014

t3.1 **Table 3** Predicting students' quality of collaboration based on network metrics using Support Vector Machine (SVM) with a Leave-One-Out Cross Validation procedure (LOOCV)

	Sub-dimensions	Accuracy	Kernel	# Features used	Top three features
t3.2					
t3.3	Dialogue management	96.88 %	Polynomial	7	Number of nodes with out-degree >5, nodes' load centrality, nodes' closeness centrality
t3.4	Reciprocal interaction	87.50 %	Polynomial	11	Average node size, square clustering, number of nodes with in-degree <5
t3.5	Division of work	93.75 %	Polynomial	4	Size of the largest node, average degree coefficient, experimental condition
t3.6	Sustaining mutual understanding	100 %	Quadratic	6	Betweenness centrality, average node size, triangle clustering
t3.7	Information pooling	90.62 %	Polynomial	3	Experimental condition, size of the largest node, number of nodes
t3.8	Reaching consensus	84.38 %	Polynomial	2	Experimental condition, average size of SCCs, average circuits' size
t3.9	Time Management	90.62 %	Quadratic	20	Average node size, size of largest node, nodes' centrality
t3.10	Task orientation	90.62 %	Polynomial	3	Betweenness centrality, closeness centrality, size of largest node
t3.11	Quality of collaboration (Total)	100 %	Polynomial	6	Experimental condition, betweenness centrality, average degree coefficient

The classification task was to predict whether dyads would be below or above the median split performed on the overall collaboration scores and subdimensions. The accuracy reported below is given for the test set

UNCORRECTED PROOF

(the proportion in parenthesis indicates 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 %), and nodes' centrality (93.75 %).

It should be noted that those predictions were made using *only measures from the graphs.* When using additional information—such as demographic data and a dummy variable representing the experimental condition of each subject—we reached a classification accuracy of 100 % for the overall quality of collaboration.

The performances of the learning algorithm were similar when considered for the rating scheme's sub-dimensions. 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 those results, we show that for this particular task and dataset, our classification accuracy is around 92.71 %.

Those results are impressive, but they need to be hedged with healthy skepticism. The small size of our dataset suggests that our model is probably over fitting our data, even though we used a LOOCV procedure. Secondly, we used a large number of features for a simple prediction task (i.e., binary classification). It is likely that SVM is, to some extent, cherry-picking the best features for separating productive versus unproductive collaborative groups.

Thus, those results should be replicated with a larger sample size to be convincing that the accuracy scores reported are indeed generalizable to the larger population of students. A last limitation is that the samples of our data are not strictly independent: we created a network for each individual, even though students completed this task in dyads. This decision was motivated by 1) the fact that our dataset is already small (N=32), 2) the networks were vastly different between individuals of the same group (edges were taken from individual student, and most of our measures were about *how* nodes were connected to each other), and 3) we wanted our algorithm to generalize to very similar and very dissimilar networks. But overall, even though we suffer from the limitations listed above, those results are encouraging and seem to suggest that network features have some predictive power regarding students' quality of collaboration.

Microgenesis of collaboration reflected in eye-movements and prediction accuracy

Considering the results described in the previous section, it may not be necessary to wait until the end of the dyad's collaborative learning activity to make relevant predictions about their collaboration quality. This makes especially relevant the important developmental concept of 'microgenesis', which we will explicate below for its applicability in this collaborative learning context. We then go on to show that the best learning algorithm for predicting the overall quality of a dyad's collaboration changes over the course of their activity together.

As the Stanford psychologist John Flavell has indicated (Flavell and Draguns 1957), his Clark University Professor Heinz Werner developed the concept of "microgenesis" (Werner 1926/1948, 1956) to unite the contents and methods of experimental and developmental psychology (also see Catan 1986) and to study the unfolding processes of perceptual, cognitive and social activities. As noted by Rosenthal (2004), "Microgenetic development concerns the psychogenetic dynamics of a process that can take from a few seconds (as in the case of perception and speech) up to several hours or even weeks (as in the case of reading, problem solving or skill acquisition)." This vital concept of microgenesis and its associated microgenetic method is integral to the developmental studies of Werner and the Soviet socio-historical school as represented in the works of Vygotsky (1978) and Luria (1928/1978), as well as more recent socio-cultural process-oriented studies by Scribner (1984, 1985) on cognitive development in social context, in her case, for adults in the workplace.

In our present case of dyads collaboratively learning about a neuroscience phenomenon employing diagrams and traces of one another's gaze behaviors as displayed in our new hybrid representation (in which they can see both the neuroscience diagrams and, superimposed, their partner's gaze patterns investigating those diagrams in real time), it is of substantial scientific interest to investigate the microgenesis of their collaborative processes when mediated by these representational resources.

What are the temporal dynamics of dyadic gaze behavior in collaborative learning conditions when one can perceive the gaze of the other (or not, for the no-gaze condition) and track its shadowing, leading, or diverging nature as turns in the gaze interactivity of the dyad unfold? The screen to which each dyad member is attending has both the learning-relevant information depicted in the diagrams, and the unfolding movements of gaze patterns overlaid on those diagrams as they are explored by the partner. Consider that as well, each participant can both see and come to anticipate how his or her own gaze patterns are serving as a stimulus for the partner's next gaze behavior, which also provides each of them with feedback on the consequences of their provision of a meaningful signal to the other as to where one is looking, which the other can conjecture to be useful for their joint task, and which they can elect to

UNCORRECTED PROOF

AUTHOR'S PROOF JrnlID 11412_ArtID 9202_Proof# 1 - 07/10/2014

follow, or choose to pursue their own next saccade. These couplings are providing occasions for learning as well, as to whether one is warranted in following the other in their gaze, or whether initiating one's own saccade is more effective in harvesting learning-task-relevant information in the diagrams displayed. So one other fruitful area of future inquiry concerns through what stages of identifiable activity dyads come to reveal one individual as predominantly a leader in the collective gaze behavior of the dyad, or as predominantly a follower of the other's lead.

In Fig. 7, we show the changing nature of our predictions during the activity using the best learning algorithm for the overall quality of collaboration (SVM with mlp kernel using the four 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 reached classification accuracy greater than 80 % three minutes before the activity ended. This result indicates that ~10 min is the minimum amount of time required by our algorithm to make acceptable predictions. Of particular interest will be further investigations that delve more deeply into the moment-by-moment microgenesis of the dyadic interchanges of gaze behaviors as they, over a short period of time, settle into a particular collaboration quality that comes to be defining of their session.

On a practical level, those results have implications already quite beyond this particular learning activity. With more training data and additional user interface features that make visible to the students' teacher these evolving collaboration patterns and their likely consequences if left unabated, one can imagine the teacher assessing students' evolving

UNCORRECTED PROOF

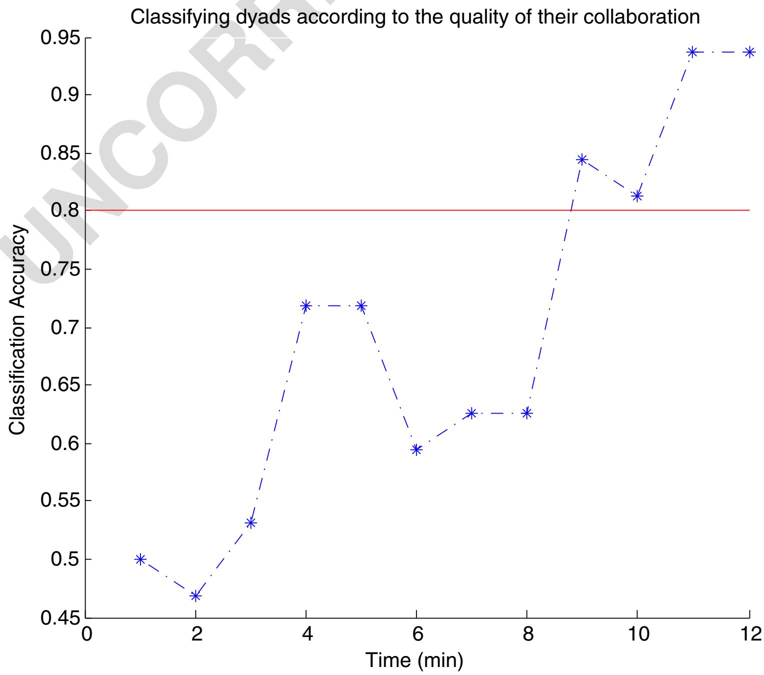

Fig. 7 Predicting the overall quality of collaboration during the learning activity

collaboration in real time. This would not only allow for a more informal evaluation of students' abilities to collaboratively solve problems, but potentially enable steering on the teacher's part of a more successful collaboration outcome. How to appraise development progress in collaboration and collaborative learning are increasingly relevant questions as recent educational reforms start focusing on what some call 21st century skills (Pellegrino and Hilton 2012), commonly considered to include collaboration, communication, innovation, creativity, critical thinking and problem solving. Using state-of-the-art machine learning techniques may enable educators to assess students' collaborative competencies and thus diagnose and scaffold (Pea 2004) the areas where improvement is needed.

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 behaviors of each member of a dyad are recorded simultaneously and made visible to the other member).

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 dyadic collaboration; 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 faced. The size of the largest node in the graph seems associated with subjects' orientation toward the task, division of work and tendency to maintain reciprocal interaction. Finally, measures related to SCCs (size of the largest SCC, average size of the SCCs) were associated with dyads' efforts to reach 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.

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 % on 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 worth noting. First, we studied only one particular kind of collaboration (i.e., remote collaboration). It is an open empirical question how well these results generalize to other collaborative situations, as it is likely that in situ interactions are different from online collaborative work because so many other streams of perceptual information are mutually available to participants in a co-located setting (Streeck et al. 2014). Another limitation is the type of task used in our study: we decided to ask participants to study a set of contrasting cases, where visual comparisons between diagrams are key to understanding the concepts taught. Thus, building networks based on collaborative eye-tracking data seems to be appropriate here, but it is not clear whether this approach would generalize to other types of tasks. It should also be noted that our approach was successful because we only considered static areas of interest; it is not clear how one would apply this method to dynamic AOIs. Additionally, we computed network metrics with only 32 students (16 dyads); a larger subject population may well yield more statistically significant patterns. Finally, as highlighted above, our dataset is relatively small and the machine-learning algorithm is likely to over fit our training data. In summary, those results need to be replicated and extended to other collaborative situations and larger datasets.

UNCORRECTED PROOF

AUTHOR'S PROOF JrnlID 11412_ArtID 9202_Proof# 1 - 07/10/2014

Future work

One promising extension of our work will be to provide a case study where our graph visualizations help other researchers gain further insights into their own datasets. We believe that network visualizations can advantageously complement existing plots and graphs for initial data exploration, and that various social settings could benefit from the visualization developed in this paper (e.g., parent-infant interactions, diplomatic negotiations, psychotherapeutic dialogues, brainstorming sessions, or sales activities).

Another direction for future work is to include voice data in 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 microgenesis of 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 (Pentland 2010), which can signal dyads that they may be reaching an insight.

Finally, the indicators described in Table 2 (network metrics correlated with a positive quality of collaboration) should be analyzed in greater depth to provide further insights into the graph structure. For some of the indicators, it is yet not clear why they are associated with a positive collaboration. A more fine-grained analysis of those indicators would probably provide additional information concerning our dataset.

UNCORRECTED PROOF

Discussion

This work provides three 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 might serve as acceptable proxies for evaluating the quality of group collaboration. Third, we fed network measures into machine learning algorithm, which seems to suggest that those features can predict multiple dimensions of a productive collaboration. As eye-trackers become cheaper and widely available, one can develop automatic measures for assessing the dynamics of people's collaborations. Such instrumentation would enable researchers to spend less time coding videos and more time designing studies and exploring patterns in their data, thus providing augmentation tools that enable humans and computers to each play to their strengths in the human-machine systems for studying collaboration. In this regard, we pursue the vision of the co-evolution of human-computer intelligent systems envisioned by Licklider (1960) and Engelbart (1963). 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 emerging-edge work could be quickly implemented in classrooms as the hardware becomes widely available and privacy concerns are sufficiently addressed in human subjects protocols.

These 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, to suggest ways to improve and teach collaborative skills as well as to better understand 'collaboration' as a theoretical construct.

Appendix 1

t4.1 **Table 4** Social network analysis (SNA) glossary for eye-tracking graphs

	Construct	Definition	Mentioned on page #
t4.2			
t4.3	Directed graph	"A directed graph is a graph, or set of nodes connected by edges, where the edges have a direction associated with them." [w]	5,7,11,13
t4.4	Undirected graph	"An undirected graph is a representation of a set of objects where some pairs of objects are connected by links, and where links do not have a direction associated with them" [w]	13
t4.5	Weight of an Edge	In an eye-tracking graph, the weight of an edge corresponds to the number of gaze movements between two predefined areas on a screen.	6
t4.6	Node of a dyadic graph	The node of a dyadic graph represents a moment of joint attention between the two members of the dyad. The size of the node shows how many times students looked at the same area on the screen at the same time.	9,10
t4.7	Highly connected nodes	A highly connected node is connected to most of the other nodes in the graph, which creates a large number of edges.	8,9,10
t4.8	Reciprocated edges	In a directed graph, an edge from node A to node B is reciprocated if the graph also has an edge from node B to node A.	10
t4.9	Betweenness centrality	"Betweenness centrality of a node v is the sum of the fraction of all-pairs shortest paths that pass through v" [n]	11,13,15
t4.10	Largest node of the graph	The largest node of a dyadic eye-tracking graph is the area of the screen where students jointly spent most of their attention.	11,13,14,18
t4.11	Strongly connected component (SCC)	"A graph is said to be strongly connected if every vertex is reachable from every other vertex. The strongly connected components of an arbitrary directed graph form a partition into subgraphs that are themselves strongly connected." [w]	11,13,14

General definitions (not related to eye-tracking data) are from networkx.github.io [n] or Wikipedia.com [w] (retrieved on 03/08/2014)

UNCORRECTED PROOF

AUTHOR'S PROOF

Appendix 2

t5.1 **Table 5** Spearman correlation matrix for network metrics and quality of collaboration (*<0.05, **<0.01)

		Sust. mutual understanding	Dialogue Mgmt	Info Pooling	Reaching Consensus	Division of work	Time Mgmt	Reciprocal interaction	Task orientation	Quality of collaboration
t5.3	Square clustering	−0.18	0.015	−0.27	−0.18	−0.006	−0.168	−0.105	−0.252	−0.264
t5.4	SCC count	−0.109	−0.017	−0.121	−0.287	−0.224	−0.243	−0.239	−0.09	−0.211
t5.5	Out degree centrality	−0.017	−0.048	−0.233	−0.184	0.09	−0.002	0.078	−0.072	−0.083
t5.6	Centrality directed	−0.017	−0.048	−0.233	−0.184	0.09	−0.002	0.078	−0.072	−0.083
t5.7	Circuits count	0.239	0.068	0.373*	0.499**	0.244	0.39*	0.18	0.221	0.397*
t5.8	In degree smaller 10	0.254	0.127	0.561**	0.707**	0.256	0.363*	0.173	0.364*	0.513**
t5.9	Largest edge	−0.118	0.032	−0.194	−0.062	0.115	0.02	0.222	0.18	−0.029
t5.10	Load centrality	0.201	−0.011	0.167	0.286	0.238	0.279	0.266	0.179	0.304
t5.11	Betweeness centrality directed	−0.109	−0.073	−0.351*	−0.318	−0.005	−0.132	−0.006	−0.16	−0.212
t5.12	In degree centrality	−0.017	−0.048	−0.233	−0.184	0.09	−0.002	0.078	−0.072	−0.083
t5.13	Betweeness centrality undirected	0.418*	0.054	0.336	0.441*	0.333	0.342	0.165	0.394*	0.478**
t5.14	Triangles clustering	0.079	0.012	0.075	0.191	0.169	0.216	0.085	0.22	0.203
t5.15	Average edge size	−0.081	0.068	−0.231	−0.479**	−0.092	−0.271	−0.096	−0.159	−0.246
t5.16	Average degree coeff undirected	−0.187	0.032	−0.241	0.078	0.305	−0.183	0.242	−0.016	0.026
t5.17	Undirected edges	0.209	0.098	0.48**	0.653**	0.28	0.374*	0.189	0.284	0.469**
t5.18	Largest circuit	0.208	0.08	0.305	0.43*	0.34	0.337	0.282	0.176	0.388*
t5.19	Reciprocated edges	0.225	0.099	0.493**	0.667**	0.284	0.387*	0.222	0.3	0.478**
t5.20	Largest node	0.187	0.315	0.239	0.475**	0.452**	0.537**	0.587**	0.582**	0.518**
t5.21	Directed edges	0.22	0.108	0.475**	0.637**	0.274	0.392*	0.209	0.276	0.469**
t5.22	Largest SCC	0.178	0.112	0.273	0.356*	0.227	0.372*	0.293	0.184	0.358*
t5.23	Centrality undirected	−0.109	−0.073	−0.351*	−0.318	−0.005	−0.132	−0.006	−0.16	−0.212
t5.24	SCC average size	0.115	0.019	0.19	0.386*	0.248	0.279	0.277	0.117	0.274
t5.25	Circuits average	0.202	0.058	0.268	0.359*	0.236	0.36*	0.187	0.152	0.33

UNCORRECTED PROOF

AUTHOR'S PROOF
JrnlID 11412_ArtID 9202_Proof# 1 - 07/10/2014

t5.26 **Table 5** (continued)

		Sust. mutual understanding	Dialogue Mgmt	Info Pooling	Reaching Consensus	Division of work	Time Mgmt	Reciprocal interaction	Task orientation	Quality of collaboration
t5.27	Closeness centrality undirected	0.274	0.146	0.18	0.264	0.26	0.309	0.166	0.15	0.314
t5.28	Average node size	0.418*	0.338	0.589**	0.625**	0.356*	0.648**	0.328	0.413*	0.617**
t5.29	Closeness centrality directed	0.153	0.054	0.383*	0.4*	0.064	0.195	0.077	0.248	0.279
t5.30	Nodes	0.191	0.084	0.145	0.249	0.219	0.27	0.187	0.168	0.271
t5.31	Average clustering	0.051	0.005	0.103	0.336	0.181	0.181	0.159	0.116	0.18

UNCORRECTED PROOF

References

Anderson, J. R. (2004). Cognitive psychology and its implications (6th ed.). Worth Publishers.

Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences, 12*(3), 307–359.

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

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: a simple proposal with multiple implications. *Review of Research in Education*, 61–100. Q6

Brooks, R., & Meltzoff, A. N. (2008). 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*(1), 207–220.

Catan, L. (1986). The dynamic display of process: Historical development and contemporary uses of the microgenetic method. *Human Development, 29*(5), 252–263.

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. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications, ETRA'08* (pp. 173–180). New York: ACM.

Cortes, C., & Vapnik, V. N. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297.

Engelbart, D. (1963). Conceptual framework for the augmentation of man's intellect. In P. Howerton & D. Weeks (Eds.), *Vistas in information handling* (pp. 1–29). Spartan Books.

Erdos, P., & Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci, 5*, 17–61.

Flavell, J. H., & Draguns, J. (1957). A microgenetic approach to perception and thought. *Psychological Bulletin, 54*(3), 197.

Freeman, L. (1977). A set of measures of centrality based upon betweenness. *Sociometry, 40*, 35–41. Q7

Fruchterman, T. M., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience, 21*(11), 1129–1164.

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures, 1*, 77–89.

Jermann, P., Mullins, D., Nuessli, M.-A., & Dillenbourg, P. (2011). Collaborative gaze footprints: Correlates of interaction quality. In *Proceedings of CSCL 2011*. Hong Kong.

Kleinberg, J. (2000). The small-world phenomenon: an algorithm perspective. In *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing, STOC'00* (pp. 163–170). New York: ACM. Q8

Licklider, J. C. R. (1960). Man-computer symbiosis. *IRE Transactions on Human Factors in Electronics, 1*, 4–11.

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. In *IEEE International Conference on Multimedia and Expo, 2009. ICME 2009* (pp. 898 –901).

Luria, A. (1978). The development of writing in the child. In M. Cole (Ed.), *The selected writings of A. R. Luria*. New York: M. Sharpe. Originally published, 1928.

Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning, 2*(1), 63–86.

Pea, R. D. (2004). The social and technological dimensions of "scaffolding" and related theoretical concepts for learning, education and human activity. *The Journal of the Learning Sciences, 13*(3), 423–451.

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

Pentland, A. S. (2010). *Honest signals*. Cambridge: MIT Press.

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.

Rosenthal, V. (2004). Microgenesis, immediate experience and visual processes in reading. In *Seeing, thinking and knowing* (pp. 221–243). Netherlands: Springer.

Roth, W.-M. (2001). Gestures: Their role in teaching and learning. *Review of Educational Research, 71*(3), 365–392.

Rothman, K. J. (1990). No adjustments are needed for multiple comparisons. *Epidemiology, 1*(1), 43–46.

Schneider, B., & Pea, R. (2013). Real-time mutual gaze perception enhances collaborative learning and collaboration quality. *International Journal of Computer-Supported Collaborative Learning, 8*(4), 375–397.

Schwartz, D. L. (1995). The emergence of abstract representations in dyad problem solving. *The Journal of the Learning Sciences, 4*(3), 321–354.

Scribner, S. (1984). Organizing knowledge at work. *Quarterly Newsletter of the Laboratory of Comparative Human Cognition, 6*(1&2), 5.

Scribner, S. (1985). Vygotsky's uses of history. In J. Wertsch (Ed.), *Culture, communication, and cognition: Vygotskian perspectives* (pp. 119–145). New York: Cambridge University Press.

UNCORRECTED PROOF

Selassie, D., Heller, B., & Heer, J. (2011). Divided edge bundling for directional network data. *IEEE Transactions on Visualization and Computer Graphics, 17*(12), 2354–2363.

Stern, D. N. (2002). *The first relationship: Infant and mother*. Cambridge: Harvard University Press.

Streeck, J., Goodwin, C., & LeBaron, C. D. (Eds.). (2014). *Embodied Interaction: Language and body in the material world*. Cambridge: Cambridge University Press.

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103–130). Hillsdale: Lawrence Erlbaum Associates.

Vygotsky, L. S. (1978). *Mind in society: The development of higher mental process*. Cambridge: Harvard University Press.

Werner, H. (1940). *The comparative psychology of mental development* (2nd ed.). Chicago: Follett, 1948. Originally published, 1926.

Werner, H. (1956). Microgenesis and aphasia. *Journal of Abnormal Social Psychology, 52*, 347–353.

UNCORRECTED PROOF