Intelligent Social Network Diagnoses in Teams

We collect and analyze dynamic social network interactions of 18 student project teams in social sciences and engineering fields from spring 2020 to spring 2021. We build a "DR GAS" machine-learning pipeline for (i) speaker Diarization and speech Recognition, (ii) Giving/Asking/other information classification, and (iii) Sentiment analysis. From results produced by the pipeline, coupled with email and longitudinal survey data, we provide real-time diagnoses of communication problems in project team networks for the future improvement of human decisions and team performance through network interventions. The machine learning analysis of giving-information decisions and sentiments can help detect early on, via periods of absence or sentimental abnormalities in communication, members who eventually have low engagement and receive low peer ratings from surveys, while email exchanges or mid-project surveys cannot achieve this goal. Our analysis also suggests potential differences in individuals' engagement in teams by gender, culture, and race. Accordingly, we will develop team and individual level tactics of interventions for improvements in future project teams.

CCS Concepts: • **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

Additional Key Words and Phrases: datasets, neural networks, gaze detection, text tagging

ACM Reference Format:
. 2018. Intelligent Social Network Diagnoses in Teams. In . ACM, New York, NY, USA, 16 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

As complex, large-scale projects become common, it is increasingly important for workers to develop knowledge and skills and make high-quality decisions to support highly interdependent work contributions in complex social and task structures. A trans-disciplinary team of engineering, computer science, education, social networks, organizational psychology, and economics experts is developing a model to address: (i) preparing individuals for the future of work in complex social systems [2, 4]; (ii) social network interventions to bolster multiteam coordination [16]; (iii) machine learning to automate sociogram development and diagnose network problems to augment human cognition for multiteam coordination [11]; and (iv) economic and social implications [5, 7, 8].

This paper focuses on the development of machine-learning diagnosis of action and communication problems through student team members' decisions to speak during meetings. We build what we call a **DR GAS** pipeline: (i) speaker **d**iarization and speech **r**ecognition, (ii) **g**iving/**a**sking/other information classification, and (iii) **s**entiment analysis. Over the past year and a half (from the spring semester of 2020 to the spring semester of 2021), we collected dynamic social network interactions of 18 student project teams in classes in several social science and engineering fields in a large public university in Big Ten. We collected their emails, meetings, Groupme messages, and eight individual-level surveys for each team, in addition to their objective performance for the projects. We use the pipeline to detect problems in team communication.

Our results show that problems and disconnects can be detected by individuals' speaking contents (from GAO classification) and tones (from sentiment analysis) as early as the first meeting (out of 6 to 7), suggesting the role of team interventions. In addition, demographic characteristics, especially for historically disadvantaged groups (women, immigrants, and underrepresented minorities), may be associated with low level of engagement in the team, suggesting the role of individualized intervention. Our next step will be to intervene during the project by solving these real-time machine-diagnosed problems to improve team performance.

This paper makes three contributions: (i) we obtain complete network interaction data of many student project teams in several different fields ranging from social sciences to engineering, notably, during the COVID period, for future intelligent diagnosis and intervention; (ii) we develop a machine-learning pipeline to enable decision and sentiment analysis of meetings; and (iii) we provide frequency, network, and sentiment analysis of the results from the pipeline in the hope of early detection of communication and decision problems in teams, and student project teams in particular [1].

2 MACHINE LEARNING METHODS

In this section, we describe the machine learning technologies we used for diagnoses of communication problems in project team networks. Specifically, we build a pipeline, which conducts **speaker diarization and speech recognition**, **giving/asking/other information classification** and **sentiment analysis** in turn on raw meeting audio or video records and produces structured information for the following analysis. Next, we introduce the corresponding components in detail.

2.1 Speaker Diarization and Speech Recognition

The speak diarization module aims to split audio signals based on speaker identity. It first partitions the audio into small segments, each of which ideally contains a variable-length utterance from a unique speaker. Then, by comparing the audio signals in each segment, the module determines the number of speakers in a conversation, and judges which speaker each segment belongs to.

In practice, we apply a deep learning based speaker diarization model released in Kaldi speech recognition toolkit [14]. The model first encodes utterances to fixed-dimensional embeddings which are called "x-vectors" [15], and then performs agglomerative hierarchical clustering on x-vectors to produce an initial diarization output. Finally, a variational Bayes Hidden Markov model is applied over x-vectors to improve the diarization results. In our implementation, we use the pre-trained model from Kaldi[10] and run it in ONNXRuntime[12].

The speech recognition module transcribes speakers' utterances in audio into text. After meeting audio is split by the speaker diarization module into segments, each of them is passed through the speech recognition module to get the utterance in text format.

Similar to the speaker diarization module, we directly adopt a chain transcription model released in the Kaldi toolkit as our speech recognition module. The transcription is completed in three steps: (1) acoustic feature extraction, where an acoustic model is applied to convert raw audio signals into acoustic features; (2) word selection, which chooses candidate words according to the acoustic features from the system's dictionary; and (3) sentence-level matching, where a language model is applied to determine the final words based on their contexts. In our implementation, we use the pre-trained ASpIRE chain model [13] from Kaldi.

2.2 Giving/Asking/Other Information Classification

The giving/asking/other information classification module is a text classifier that classifies utterances from speakers into three categories: giving information (G), asking information (A), and others (O). This module aims to determine whether an utterance is giving information to others, asking for information from others, or neither, to help us understand the information exchange in communications among team members. Subsequently we can focus on pertinent information exchanges in teams.

We train a deep text classifier for this goal. Specifically, we build a 1-layer recurrent neural network (RNN) with gated recurrent units (GRU) as the text classification model. The size of word embeddings is set as 50 and the hidden size of the RNN is 128. The model is trained on 932 training instances manually labeled by human coders with a batch size of 64 for 20 epochs.

2.3 Sentiment Analysis

The sentiment analysis module predicts the sentiment polarity of the utterances from speakers. It judges whether the sentiment expressed in an utterance is positive, negative or neutral, and measures the strength of the sentiment. Based on the sentiment analysis results, we can trace the emotional changes of team members in their communications.

We adopt the public sentiment analysis tool Vader [9] as our sentiment analysis module. Given an utterance, the tool outputs a normalized sentiment intensity score ranging from -1 to 1, where -1 and 1 indicate extremely negative and positive sentiments, respectively, and 0 indicates neutral sentiment. In our experiments, we use the implementation of the Vader model in the NLTK toolkit [3].

3 RESULTS

3.1 Summaries

Giving Information Frequencies and Queue Networks. Based on the machine-learning pipeline, we obtain transcriptions of meetings, which are speaker-identified and labeled as giving/asking/other information. We first conduct frequency analysis to measure how long each member is giving information during the conversation, to measure their real contribution to the team (Figure 1). We then provide conversation queue network analysis (Figures 2 and 3), where sociograms are constructed for each team based on the duration of giving information (size of the node) and the frequency one speaks before/after another speaker (an undirected edge aggregates the interaction of two speakers whenever they give information next to each other, and directed edges distinguish between when person 1 is speaking before and after person 2). These results reveal that teams interact differently in meetings. Teams have (i) a naturally arising "leader," who communicates with everyone on the team (e.g., Teams B, C, and D), (ii) a pair of active leaders, who communicate much more frequently with each other (Teams K and N), (iii) evenly contributing members (e.g., Team T), and (iv) evenly contributing members and some less involved participants (e.g., Teams R and S).

Sentiment. We assign a sentiment score to each sentence. Based on these scores, we can investigate each team's sentiment by member and by video. Figure 4 provides an average sentiment score for each team member, weighted by the duration of giving information. It provides some information, but the more detailed video-by-video sentiment analysis will be more helpful.

Emails and Progress Loops. For completeness, we also include the email frequencies (#emails sent, Figure 5) and email exchange networks for each team (Figure 6); #emails received and undirected networks are omitted but can be constructed. From emails, meetings, and milestones extracted from the meetings, we can construct progress loops, which are widely used in construction management and civil engineering [6]. Figure 7 provides two examples (the two teams we will elaborate on in the subsequent subsection). However, due to the low frequency of email exchanges—especially after the first team meeting starts—the detection of communication problems through meetings is much more important.

3.2 Qualitative Analysis

We found instances in which women, international students, and underrepresented minorities participated in meetings at lower frequencies and were rated lower by other group members, even though they had normal amount of email exchanges and regularly participated in our research surveys.

International Students and Underrepresented Minorities. Student teams A and B are from the same social science class and work on the same project. They each had six members in Spring Semester 2020, which was disrupted by COVID. From speaker diarization (Figures 8 and 9), we see the lack of speaking/giving-info time by A1 and A4 and B5 and B6, who are the only international students in the two teams. In comparison, email communication does not reveal such a stark contrast: There is no significant difference in frequency of emails sent and received (Figures 5 and 6). A lack of communication eventually led team B not to finish the project after the COVID disruption in March 2020. Team C of 8 members from the same class but a different section had 3 international students and 1 underrepresented minority. However, one international student (C3) and the URM (C5) dropped out of the team without showing up to any of the meetings. The other two international students, C7 and C8, also had relatively low participation. The cultural differences—coupled with a lower communication skill—may have contributed to the stark differences in contribution in meetings.

Female Students. Teams R and S each had a female student receiving relatively low ratings from their team members (Figure 10). Team R consists of 1 female student and 5 male students in an engineering class, and team S consists of 5 female students in a social science class (different subject from teams A–C). From the aggregate speaker diarization data, we cannot tell R1 and S1 are the students receiving significantly low ratings. Meeting-by-meeting breakdown provides a more detailed and complicated story (Figure 11). R1, the only female in the team, dominated the first two meetings, but chose not to show up to meetings 3–7. S1 showed up and spoke a minimum amount in the first meeting, skipped the next four meetings, and showed up to the last meeting and all three mock presentations. R1 and S1 shared the feature that they were both "disconnected" from the team for an extended period, but they differed in that R1 never reconnected and S1 did. The gender composition of the team may be one factor that affected the outcome.

Early Detection of Problems. A closer look at the directed conversation network suggests that the disconnect from the team may be detected early on, even in the first meetings. For example, the extremely uneven contribution by members in team B in the two meetings might have led to the team deciding to quit when COVID hits. Figure 12 shows the conversation exchange of teams R and S, with the node size indicating the speaking duration and the thickness of the edges the number of exchanges between the two members. R1 dominated the conversation, but the exchange was mostly with R2. S1 spoke the least amount, but the exchange was evenly distributed with other members of the team. These delicate patterns are not easily seen from total speaking duration, showing the role of detailed analysis. Furthermore, sentiment analysis of the speakers' tone (Figure 13) indicates that R1 was significantly more positive than other members of the team, but S1 was arguably the least positive. Though being the most cheerful, R1 arguably did not "fit in" the team. Overall, paying closer attention to certain demographic groups in our intervention may result in better individual and team performance.

4 FUTURE WORK

We demonstrate the promise of the "DR GAS" machine-learning pipeline to generate speaker diarization for frequency and network analysis, generate G/A/O information classification to eradicate useless speeches and conduct sentiment analysis on the speech for early detection of communication problems in teams. We combine these machine-generated results with emails and surveys to construct progress loops for teams and to detect issues in team coordination and communication.

Potential additional automations to the existing pipeline include but are not limited to (i) speaker recognition (which is hand-performed after speaker diarization), and (ii) identification of team-set milestones (e.g., "complete this writeup by Wednesday"). Furthermore, we can try to recruit more underrepresented minorities to be able to have more powerful analyses. These analyses of machine-generated data all serve the ultimate goal of early intervention to problems in teams. Interventions, which we are implementing this semester, include (i) team meetings to set milestones at the beginning of the project, and (ii) individual-level nudges to improve individual—and ultimately team—performance; results suggest the role of nudges based on individual and team demographic characteristics. Furthermore, the same exercises can be applied to the three authentic project teams tracked by our team, and will eventually be applied to more project teams outside.

ACKNOWLEDGMENTS

XX

REFERENCES

[1] Richard Arum and Josipa Roksa. 2011. *Academically adrift: Limited learning on college campuses.* University of Chicago Press.
[2] Hasan Gokberk Bayhan, Sinem Mollaoglu, Hanzhe Zhang, and Kenneth Frank. 2021. Project Team Collaborations During Time of Disruptions: Transaction Costs, Knowledge Flows, and Social Network Theory Perspective. In *Construction Research Congress.*
[3] Steven Bird. 2006. NLTK: the natural language toolkit. In *Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions*. 69–72.
[4] Rob Cross, Peter Gray, Shirley Cunningham, Mark Showers, and Robert J Thomas. 2010. How to Make Employee Networks Really Work. *MIT Sloan management review* (2010), 83–90.
[5] David J Deming. 2017. The growing importance of social skills in the labor market. *The Quarterly Journal of Economics* 132, 4 (2017), 1593–1640.
[6] Angelo Joseph Garcia, Sinem Mollaoglu-Korkmaz, and Vernon D. Miller. 2014. Progress Loops in Interorganizational Project Teams: An IPD Case. In *Construction Research Congress 2014*. 2011–2020.
[7] Luis Garicano. 2000. Hierarchies and the Organization of Knowledge in Production. *Journal of political economy* 108, 5 (2000), 874–904.
[8] Luis Garicano and Esteban Rossi-Hansberg. 2015. Knowledge-based hierarchies: Using organizations to understand the economy. *economics* 7, 1 (2015), 1–30.
[9] Clayton Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 8.
[10] Kaldi Project. 2021. Kaldi Project. https://github.com/kaldi-asr/kaldi
[11] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111–3119.
[12] ONNXRuntime. 2021. Optimize and Accelerate Machine Learning Inferencing and Training. https://github.com/Microsoft/onnxruntime
[13] Daniel Povey. 2021. ASpIRE Chain Model. http://kaldi-asr.org/models/m1
[14] Daniel Povey, Arnab Ghoshal, Gilles Boulianne, Lukas Burget, Ondrej Glembek, Nagendra Goel, Mirko Hannemann, Petr Motlicek, Yanmin Qian, Petr Schwarz, et al. 2011. The Kaldi speech recognition toolkit. In *IEEE 2011 workshop on automatic speech recognition and understanding*. IEEE Signal Processing Society.
[15] David Snyder, Daniel Garcia-Romero, Gregory Sell, Daniel Povey, and Sanjeev Khudanpur. 2018. X-vectors: Robust dnn embeddings for speaker recognition. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 5329–5333.
[16] Thomas W. Valente. 2012. Network Interventions. *Science* 337, 6090 (2012), 49–53.

A APPENDIX

In the appendix, we present the figures we used in the main text.

List of Figures

Fig. 1. Giving-information duration by speaker

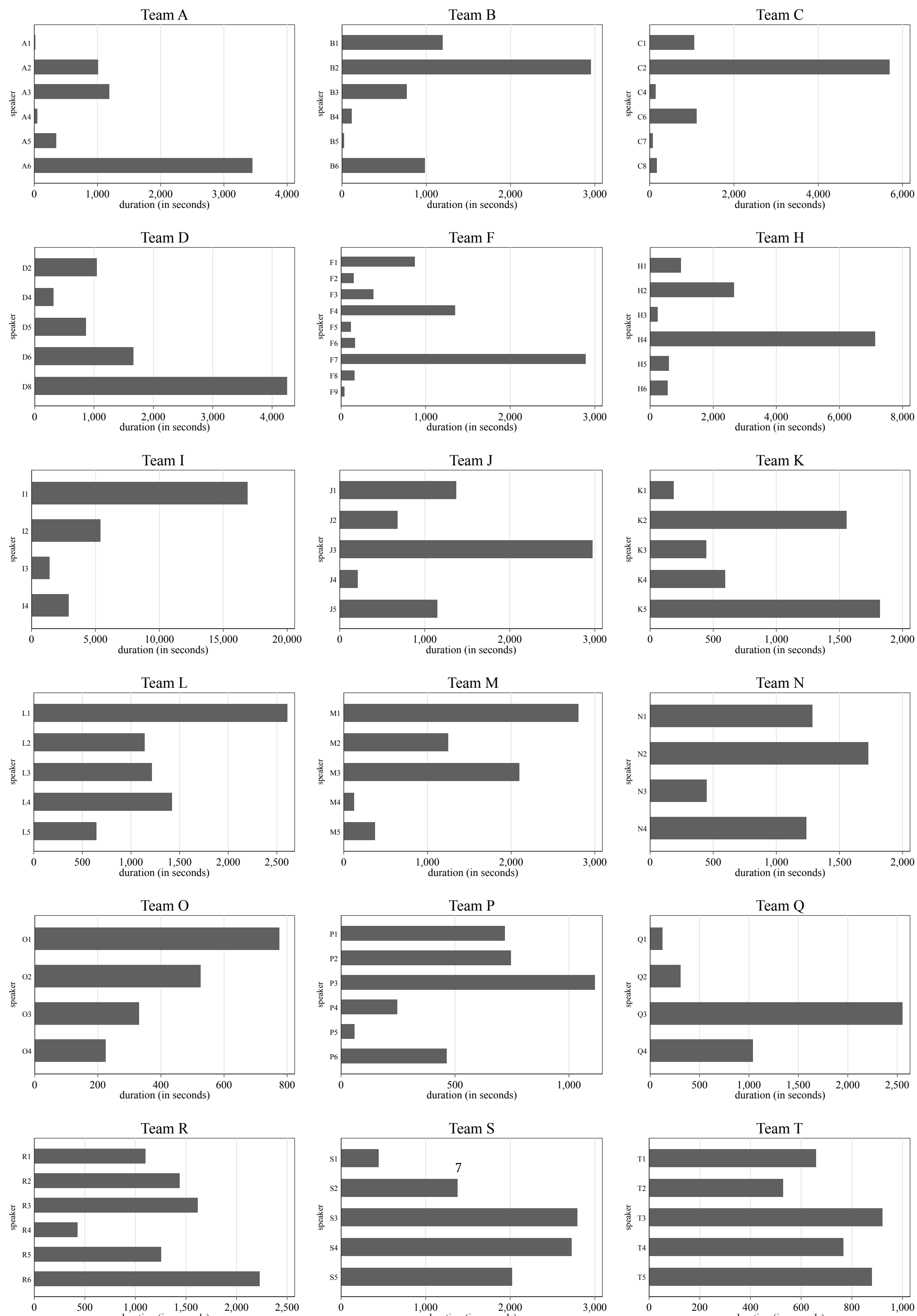


Fig. 2. Undirected giving-information queue network

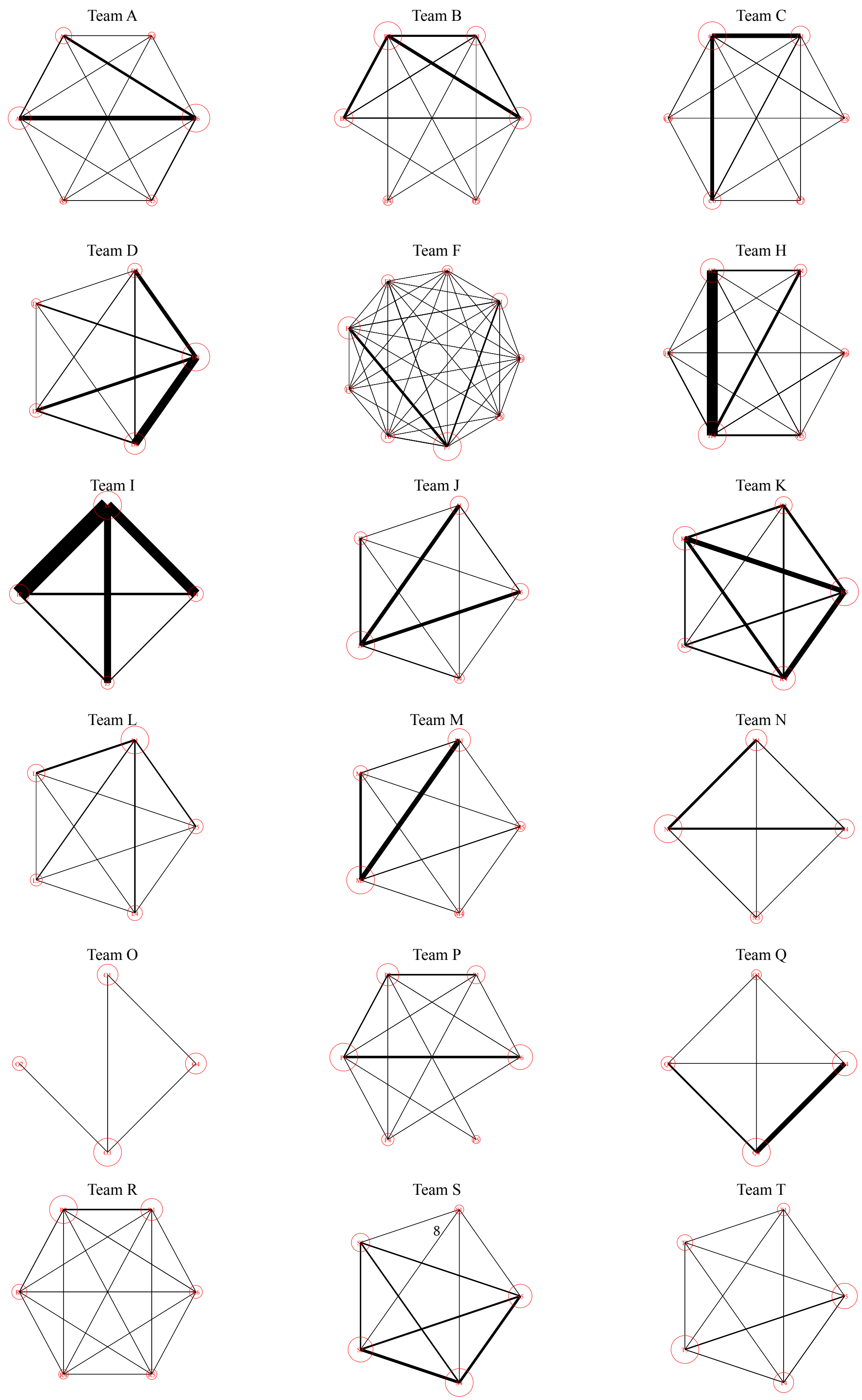


Fig. 3. Directed giving-information queue network

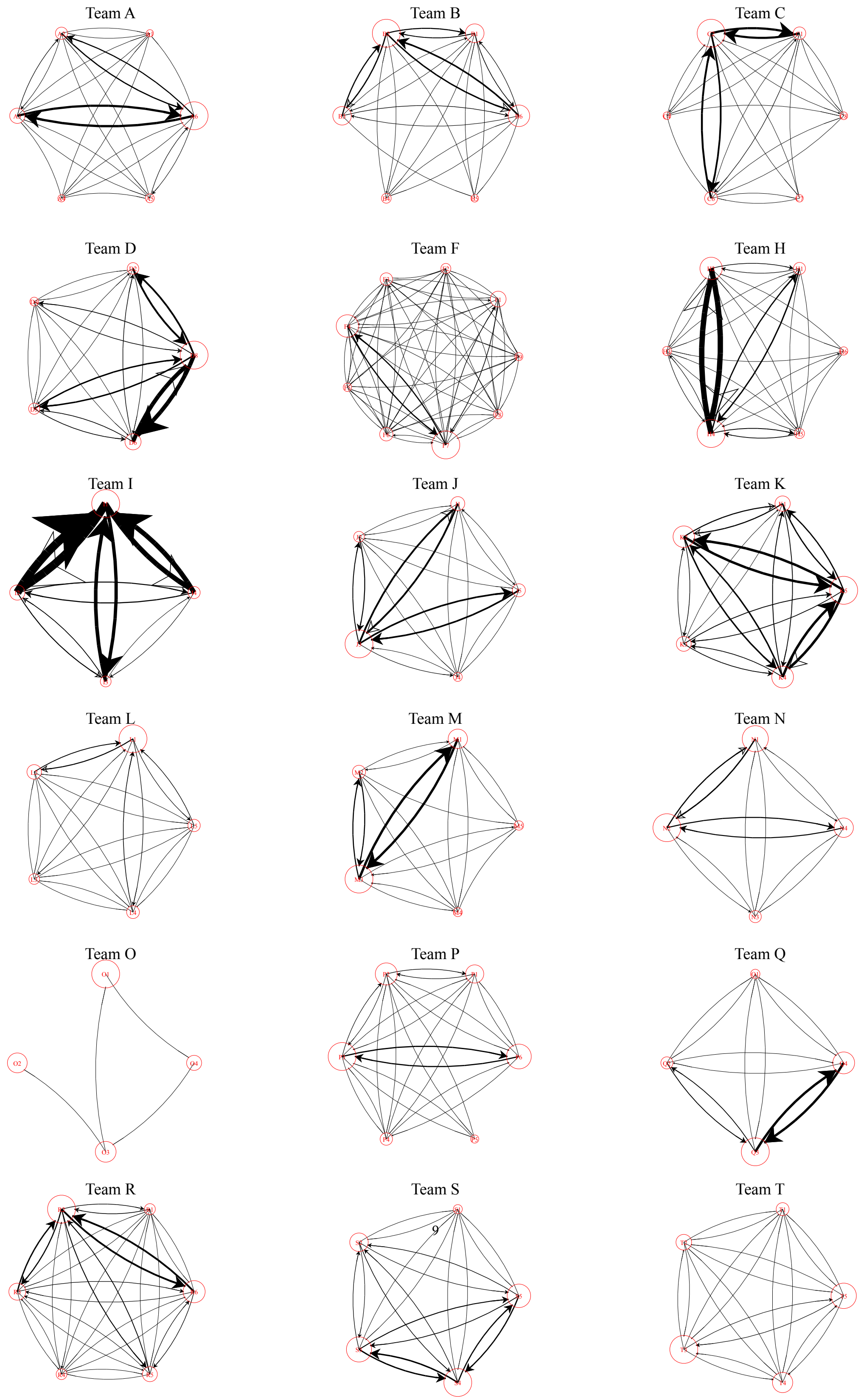


Fig. 4. Giving-information-duration-weighted nltk by speaker

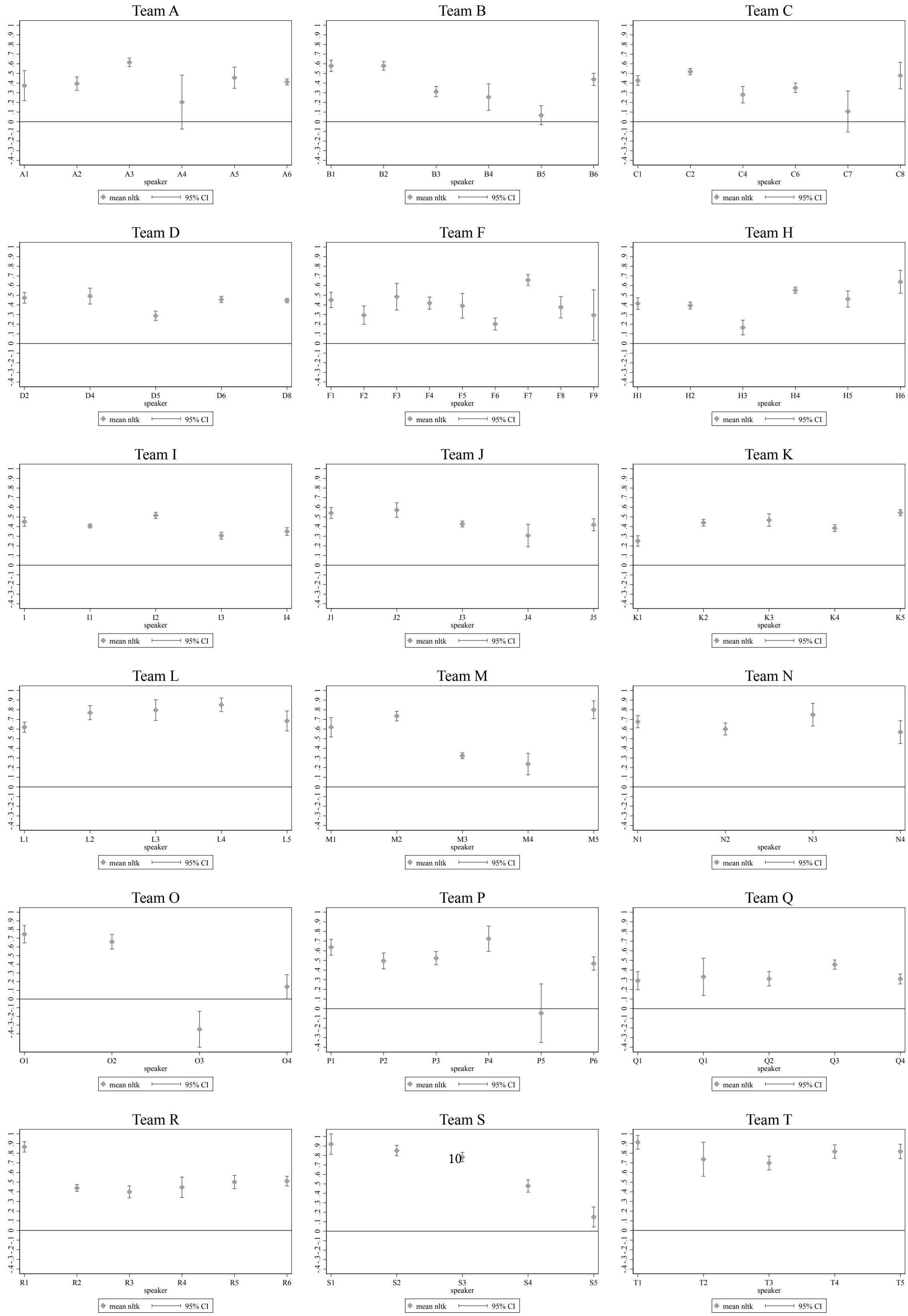


Fig. 5. Number of emails sent by member

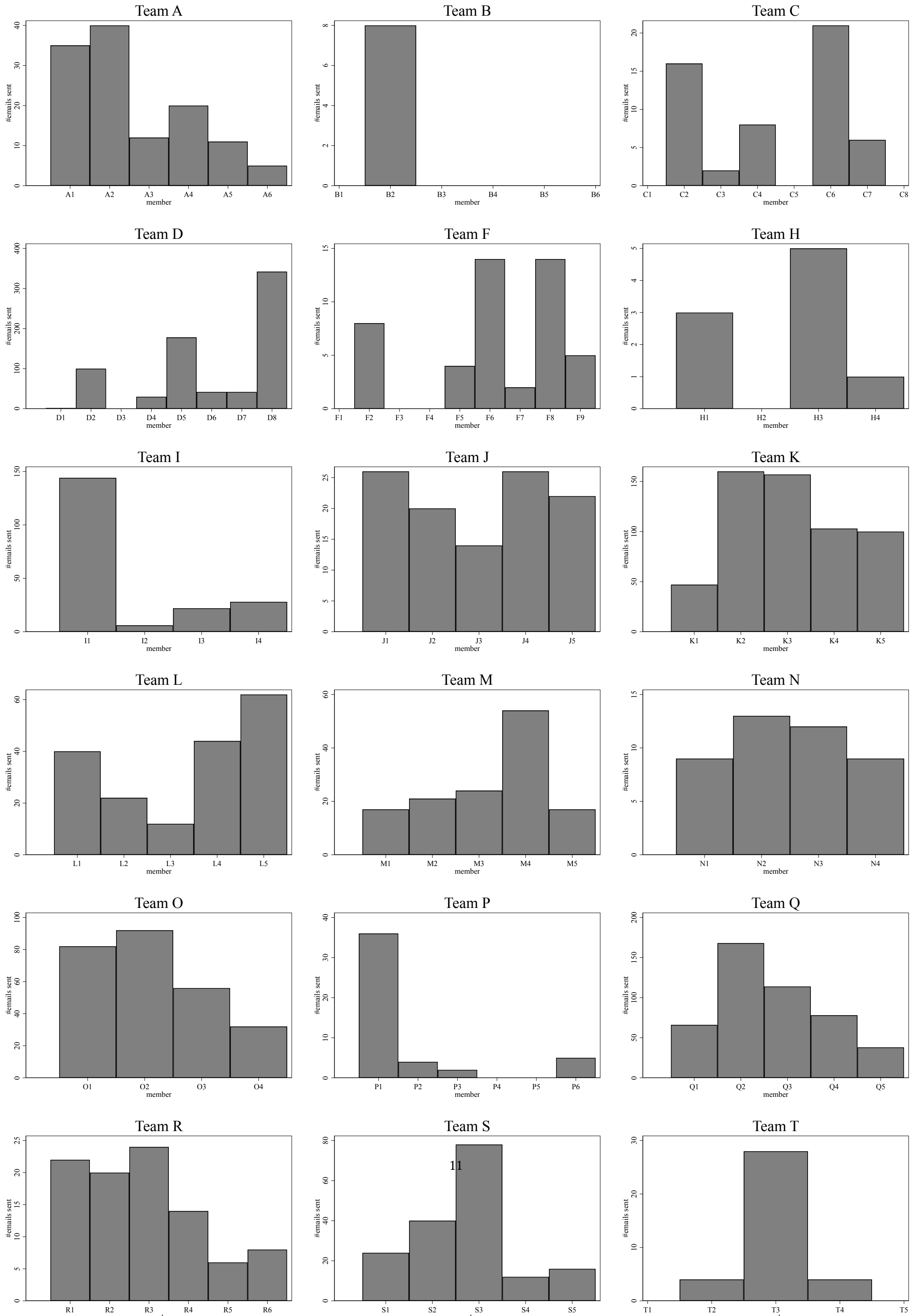


Fig. 6. Emails exchange networks (undirected)

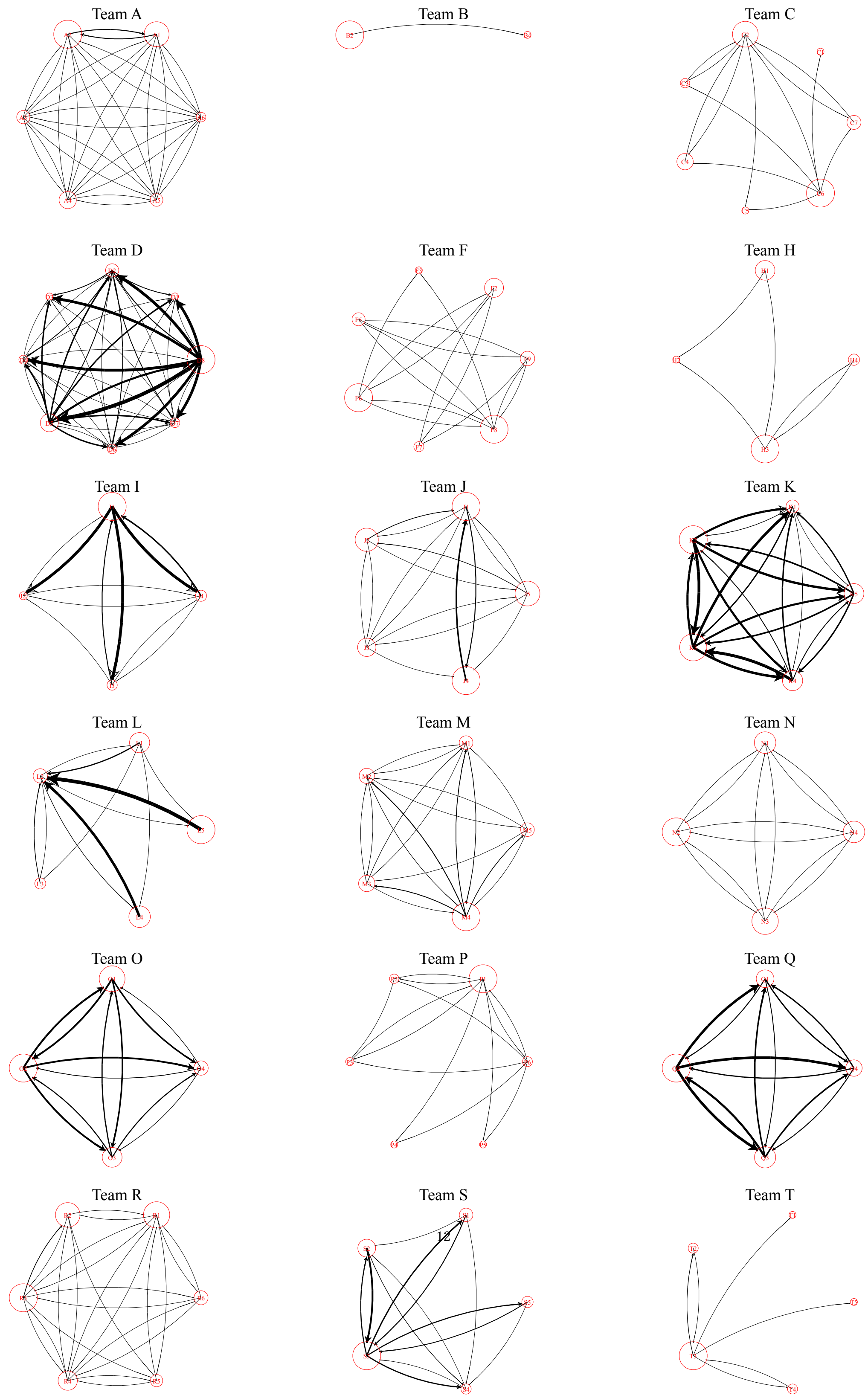


Fig. 7. Examples of progress loops: Teams R and S

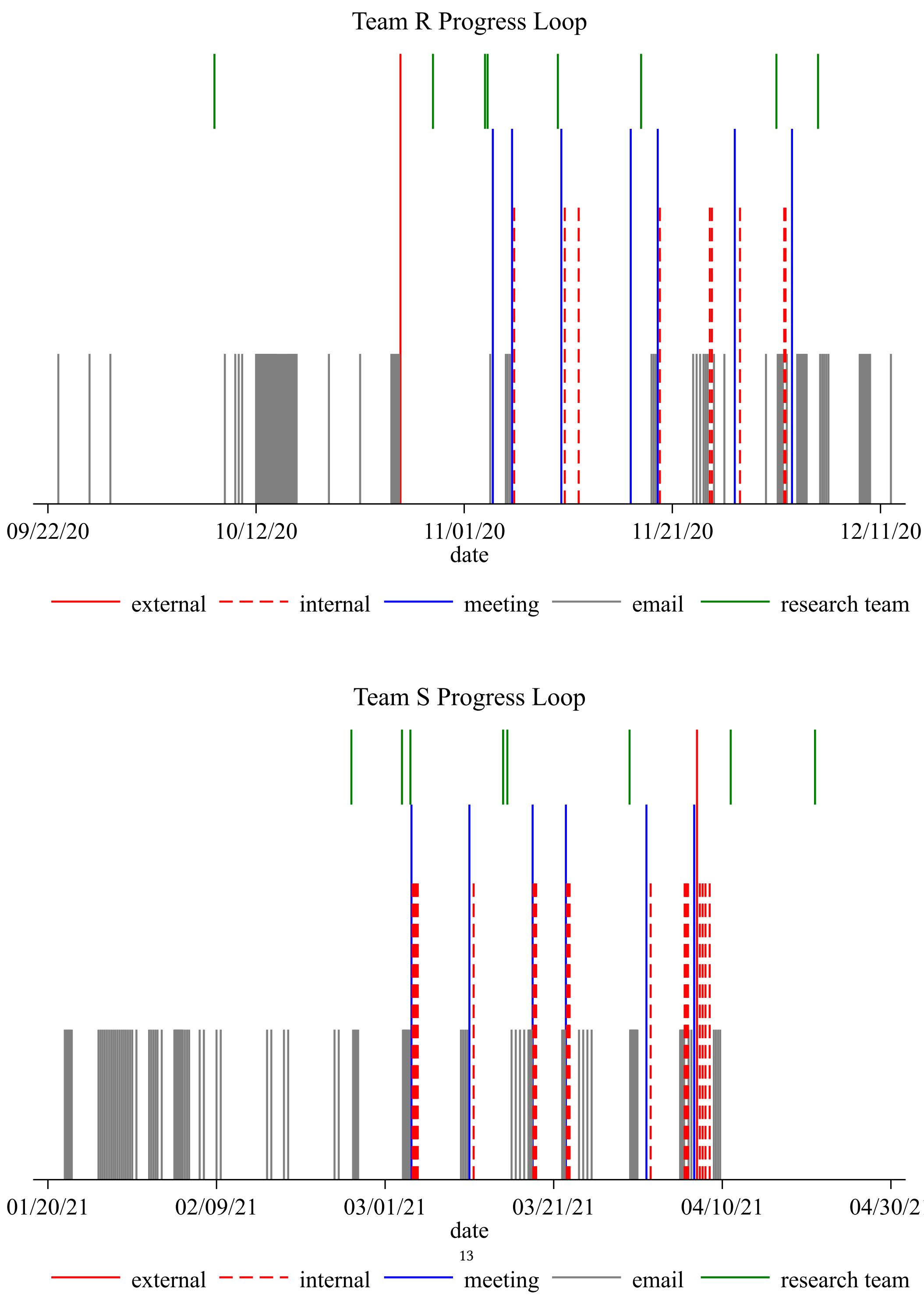
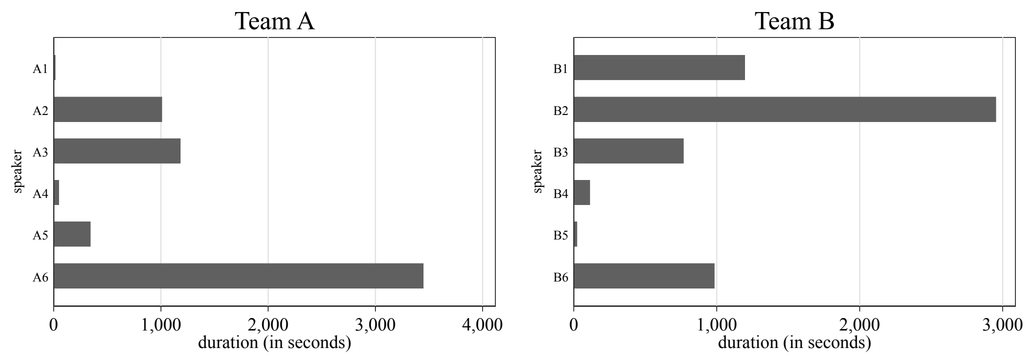
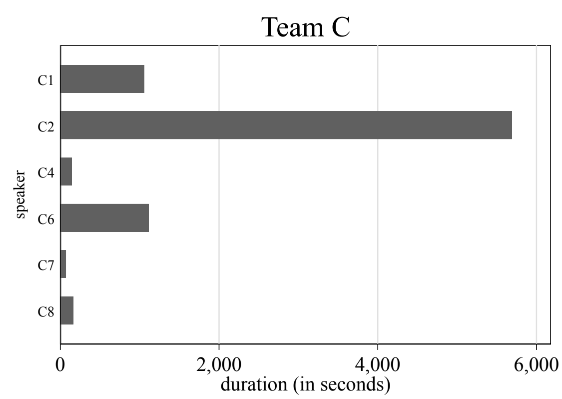


Fig. 8. Short giving-information duration by international students



Note. Members A1 and A4 and B4 and B5 are international students.

Fig. 9. Short giving-information duration and low participation by international students and underrepresented minorities



Note. Members C3, C7, and C8 are international students. Member C5 is an underrepresented minority. C3 and C5 dropped out of the team before the first meeting.

Fig. 10. R1 and S1 received low ratings from team members

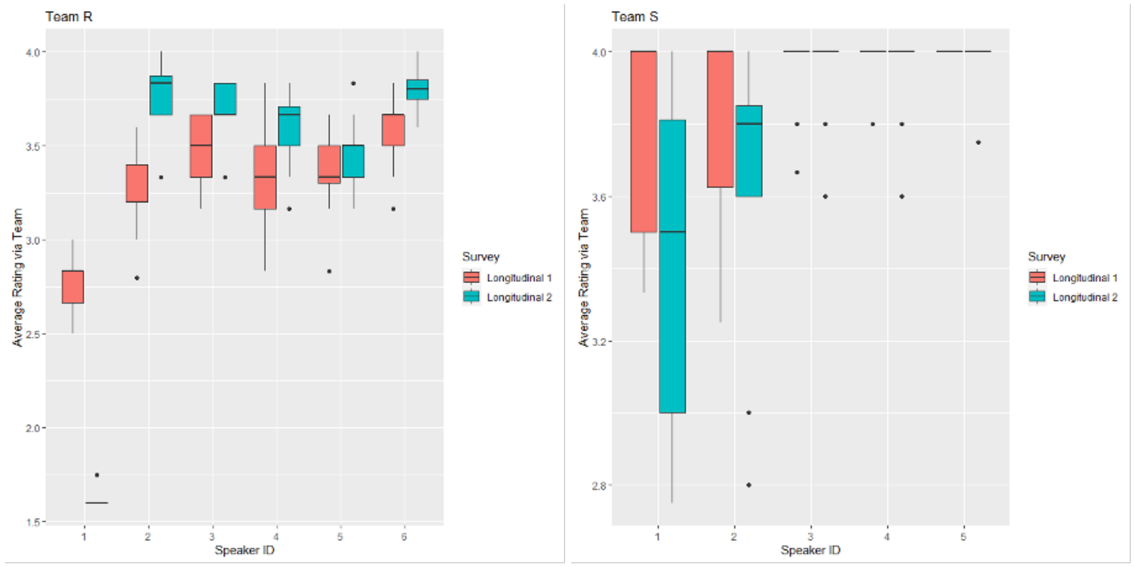


Fig. 11. Giving-information duration by video

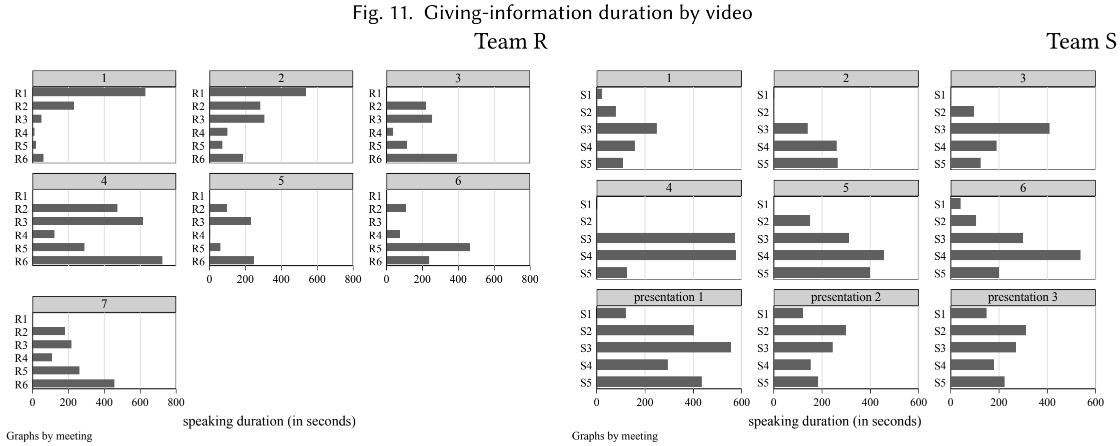


Fig. 12. Conversation network, first meeting

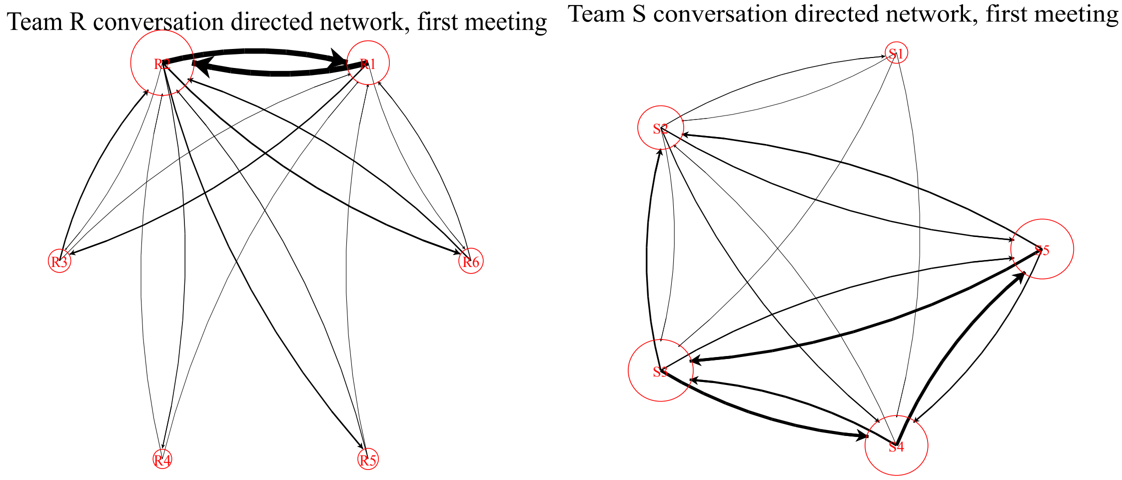


Fig. 13. Average giving-information duration-weighted sentiment score