GEDII GENDER DIVERSITY IMPACT – Improving research and innovation through gender diversity

WWW.GEDII.EU

Jörg Müller

Project Deliverable

D2.4 Comparative Case Studies Report

1st March 2018 – **FINAL 1.0**

Project Consortium

R&I

IN3
Internet
Interdisciplinary
Institute

Universitat Oberta de Catalunya, Spain

Project Coordination

Dr. Jörg Müller
Dr. Milagros Sáinz
Dr. Rosa Borge
Dr. Julià Minguillón
Dr. Julio Meneses
Dr. Sergi Fàbregues

Hochschule Furtwangen, Germany

Prof. Dr. Ulrike Busolt
Sandra Klatt
Wiebke Kronsbein

Örebro University, Sweden

Dr. Ulf Sandström
Dr. Anne-Charlott Callerstig
Prof. Dr. Liisa Husu

Verband der Elektrotechnik Elektronik Informationstechnik e.V., Germany

Dr. Walter Börmann
Nicole Bienge
Andreas Landwehr

Cranfield University, UK

Dr. Anne Laure Humbert
Dr. Elisabeth Anna Guenther

Horizon 2020
European Union Funding
for Research & Innovation

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 665851.

Table of Contents

Acknowledgments

This research would have been impossible without the collaboration and the predisposition of the 8 research groups as well as their research centers. In particular we wish to express our gratitude to the team leaders and their team members for their generosity to participate in this study and share their views on their team and working environment. Without all their help, this research would not have been possible.

Data collection was carried out by Elisabeth Günther and Jörg Müller, including sociometric badges and semi-structured interviews. Preparation of sociometric data as well as exploratory analysis was carried out by Jörg Müller. The development of the interview code book and the codification of the transcribed interviews was carried out by Lidia Arroyo, María José Romano, Elisabeth Günther and Jörg Müller. Interviews were codified by María José Romano, Lidia Arroyo and Efrem Melian. Individual case study reports were written by Elisabeth Günther, María José Romano and Jörg Müller. The initial steps of REM modeling have been discussed with Anne Laure Humbert and Elisabeth Günther. The design and execution of the controlled experiments were lead by Julio Meneses. This comparative report has been written by Jörg Müller, based upon the collective work of all the above mentioned colleagues.

Executive Summary

- Existing research has documented gender differences in non-verbal behavior during interactions and conversational styles. Status and gender matter for the interaction system. Sensor technology and computational approaches are rapidly evolving, potentially being able to detect and measure these behavioral differences.
- The present research has used Sociometric Badges by the US based consultancy company Humanyze[1] to assess the potential for measuring gender differences during interactions and conversational dynamics. Sociometric Badges incorporate different types of sensors: Bluetooth (proximity), infrared (for face-to-face interaction), microphone (for speech profiles and turn-taking) as well as accelerometer (for body activity).
- Concurrent with other research, basic reliability issues regarding the sensor data were detected. Several controlled experiments were carried out in order to assess the reliability of microphone and interaction based measurements. Whereas speech duration is measured with a more or less consistent error, turn-taking measurements are neither precise nor consistent. The audio derived measures of Sociometric badges can only be used with great caution; turn-taking counts do not reflect actual conversational shifts.
- 8 case studies were carried out. Field work involved the wearing of sociometric badges during one week in each team. A total of 35 semi-structured interviews were carried out as well as a short questionnaire distributed to obtain basic socio-demographic data of participants. A comparative analysis across the case studies demonstrates commonalities and differences regarding leadership and collaboration patterns. The sociometric profiles and derived measures provide basic information regarding the working style as well as the level of explicit steering carried out by the team leader.
- The sociometric data did not produce any consistent gender pattern across the teams. No statistically significant differences could be found on the individual level regarding speaking duration, listening duration, or the overall amount of interaction detects between women and men. However, differences in terms of mean speaking and listening duration as well as the participation of men and women in interactions do exist among members within teams. Gender differences regarding several face-to-face aggregated network statistics (closeness and betweenness centrality) tend to be significant for gender, although the effect is very small.
- A promising approach for analyzing time-based data is used. Relational Event Modeling allows to analyze well ordered event sequences. It allows distinguish the effects of covariates (such as gender) from endogenous dynamics such as the relative frequency with which certain individuals in the team interact. Considering the exploratory character of the REM models used in this report, gender as well as other

1 https://www.humanyze.com/

effects can be detected across the 8 research teams. However, results are still preliminary and should be used with caution.

- Using sociometric badges in research is far from a straight forward process. Rather, the often black-box “nature” of the technology complicates not only data collection but also the interpretation of the resulting data. The usage of this data in social science research is still in its early stages. There is a clear need for shared field work protocols as well as open standards for data processing and analysis to guarantee the reproducability of results.

Introduction

The GEDII project has carried out 8 case studies with research and development teams in order to produce empirically grounded knowledge about the contexts and logic of “gender diversity” and its impact on research performance. Instead of addressing a specific set of hypothesis regarding gender diversity in R&D teams, however, the principal concern of work package 2 is a methodological one, namely to develop methods for studying gendered team dynamics using “sociometric badges.”

Sociometric badges and other sensor based devices are becoming increasingly widespread including the social sciences – and certainly will do so in the immediate future. Proximity measurements through RFID, Bluetooth or infrared enabled devices have received considerable attention, both from organizational researchers as well as the computer science community (Atzmueller, Ernst, Krebs, Scholz, & Stumme, 2014; Cattuto et al., 2010; Do, Kalimeri, Lepri, Pianesi, & Gatica-Perez, 2013; Hung, Englebienne, & Kools, 2013; Lee, Seo, Choe, & Kim, 2012; A. Pentland, 2007). The possibilities of video- or audio recordings, although being part of social sciences research methods for quite some time, are expanded as smaller sensors, cameras or microphones uproot previous limitations of size and storage capacities. New devices, including smart-phones, deliver high resolution data (down to the microsecond) on proximity, connectivity, or behavioral (body movement) measures that open up new possibilities for unobtrusive, real-time monitoring – the implications and possibilities of which are only beginning to be spelled out by social sciences researchers (George, Osinga, Lavie, & Scott, 2016; Tonidandel, King, & Cortina, 2016).

The potential of sociometric devices are promising for the field of gender research in order to probe and expand on previous insights how gender differences are expressed in group interactions.

Sociometric badges could offer insights into, what Alex Pentland has called “Honest Signals” (Pentland, 2008), i.e. non-conscious dimensions of human behavior. Especially the influence of stereotypes and bias appear at first sight as an interesting field for testing sociometric technology, as certain behavior such as turn-taking in conversations is both regulated by highly fine-tuned, automated conversational rules while also having been shown to exhibit gendered patterns.

As already described in the D1.1 Conceptual Framework, “expectation states theory” explains how gender stereotypes bias competency expectations in social interactions, conditioning “the likelihood that people speak up in a situation, whether others attend to them if they do, how their suggestions are evaluated, whether they become influential, and the extent to which others and they, themselves, are willing to infer high ability in them based on their performance.” (Ridgeway, 1992, 2007, pp. 324–25). There are important non-verbal behavioral cues that structure interaction patterns in small groups: more powerful individuals in a group tend to speak more and tend to interrupt others more often than lower status members. Speaking time especially has been identified as a strong indicator of dominance (Hall, Coats, & Smith LeBeau, 2005; Mast, 2002). This finding has been corroborated from gender studies perspective, where high status men dominate group conversations compared to lower status women (Mast, 2001, 2002). Turn-taking, together with visual dominance, has been identified as one of the key nonverbal features for expressing dominance in social interaction (Cappella, 1985; Ellyson, Dovidio, & Brown, 1992; Mast & Cousin, 2013; Palmer, 1989; Ridgeway, 1992).

Although the computer science- and artificial intelligence community is actively exploring "social signal processing" as a new discipline (Vinciarelli, Pantic, & Bourlard, 2009), the attempts to address the potential of sensor based data from a gender perspective are rather scarce to non-existent. To our knowledge, currently there exists just two studies that explicitly combine gender and sociometric technologies. On the one hand this is work done by Onnela et al., (2014) who analyzed gender, talkativeness and interaction style in small groups, showing that women are more "talkative" in small group settings. On the other hand, Stehlé et al., (2013) used RFID beacons to monitor spatial behavior within a school to investigate gender homophily among children. Some publications targeting gender aspects use sociometric badges although it is not the main focus of their study such as for example Woolley, Chabris, Pentland, Hashmi, & Malone, 2010. Members of the original development team of the badges at MIT and later on Humanyze have published short online pieces at Bloomberg and HBR specifically addressing gender issues in teams (Turban, Freeman, & Waber, 2017; Waber, 2014). According to the authors, gender differences are not apparent in the sociometric data. However, the lack of details regarding methodology and data makes it hard to explore these results in more depth.

Given the increasing importance of big data for social sciences – and sensors being one source of big data – it is important to start more detailed worked on the potentials for gender research within this area. As the following report will show, sensor based data poses several challenges, including the reliability of the measurements and their subsequent interpretation and "construct" validity. Despite providing supposedly "objective" measures, a central finding of this report is certainly that sensor based interpretations and insights regarding any topic under consideration are a far cry from being straight forward and "easy". Rather on the contrary: to the degree that measurement involves complex technological instruments, the interpretation of the produced data becomes more challenging, error prone and necessarily cautionary. It seems that the ease and volume with with data is generated nowadays stands in stark contrast to the difficulties of their interpretation and fertility for generating new theory.

Overview of the report

The present report gives a summary overview of the work carried out during the WP2 Case Studies within the GEDII project. It concentrates on the 8 case studies carried out with research teams during the period of June 2016 to March 2017. In parallel to the actual work carried out with researchers, the scientific literature published during this period started to raise concerns regarding the basic reliability of the sociometric badges involved in our case studies (Chaffin et al., 2015). This prompted us to carry out additional controlled experiments to assess the reliability of the audio (derived) measures. The report incorporates these different strands of work and provides overall new estimations on the overall reliability of sociometric badges.

Hence, the report is structured as follows: in the first section, the reliability of the sociometric data is described. There is mounting evidence that sensor data is much less “objective” and reliable than its name would suggest (Chaffin et al., 2015; Chen & Miller, 2017; Kayhan et al., 2018; Yu et al., 2016). The GEDII project contributes to this work by providing new insights especially regarding turn-taking, speech duration and badge orientation (angles).

The second section explores in a comparative manner the data across the 8 case studies, using each of the four data dimension – proximity, infrared, sound, accelerometer - available. A unique contribution of this research is the availability of sociometric data from real-world research teams. Thus, whereas the numeric values are hard to interpret “on their own”, a comparative view already allows to qualify and contextualize the data in important ways. Reporting basic indicators such as for example the (mean) counts of face-to-face contacts across teams, facilitates the construction of benchmarks for similar studies in the future.

Section three then establishes a “data dialog” in order to contextualize further the sociometric findings in relation to the interview materials and the short questionnaire. The main features of each team are used to construct a basic matrix of commonalities and differences across the cases. In a second step, the sociometric profiles are consulted in order to detect possible correspondences between the descriptive accounts and the sociometric data. This third section proves especially challenging from a methodological point of view since it applies a mixed-method design combining quantitative “big-data” approaches with qualitative ones.

The fourth chapter introduces the Relational Event Model (Butts, 2008) for analyzing time based data. Sociometric data enable dynamic approaches to the analysis of interaction in teams. Using the Relational Event Model, the current report engages in team research beyond the aggregation of static snapshots in favor of taking genuine time-based concepts such as duration, cyclicity, and timing of events into account.

Apart from a concluding section, several Annexes are provided that summarize the sociometric data profiles for each team as well as different visualizations.

Methodology

The overall methodology adopted by WP2 is based on “case studies”. A case study entails the detailed and intensive analysis of single cases. It aims at an in-depth understanding of the phenomena in its real-life context (Yin, 1994). We aim at deeper insights how individuals interact on the team level and how gendered hierarchies and bias conditions team interaction and the sharing of information. We need to explore how teams and team members see the environment in which they work, including their team collaborations, the wider organizational context and how this influences “performance” from their perspective and in relation to established indicators.

Case Study Design and Fieldwork

The GEDII project has carried out 8 case studies[2]. Table 1 provides an overview of the main features of each case study. 5 out of 8 teams work in Spanish Universities or Research Performing Organizations while 3 were located in the UK. Recruitment of research teams was a laborious and difficult process. Originally, WP2 foresaw to carry out the case studies in Germany where Partner VDE has strong ties to the industry in order to facilitate field access. However, after considerable efforts to secure unsuccessfully the participation of German research teams, we decided to expand our search to other countries, mainly UK and Spain due to the GEDII partner contacts. Difficulties during the recruitment process mainly explain the skewed sampling of the research teams: where we originally foresaw to have at least one research team in the public sector and one in the private sector for the two thematic choices, namely Transport research on the one hand and Biomedical Engineering on the other (within the same country), most research teams pertain to the public Biomedical Engineering sector in Spain.

Personal contacts were key in the end to secure the participation of research teams and establish a sufficient level of trust with team leaders to implement the actual study with sociometric badges. Hence, concessions had to be made in terms of the ideal sampling of research teams across sectors and thematic focus, in order to secure the participation of a sufficient number of teams within the tight framework of the project. However, although the sample of research teams is not ideal, it covers the public and private sector as well as different scientific disciplines (Biomedical field as well as Energy / Transport research). Most importantly, however, as the following results will demonstrate, the case studies are sufficiently diverse from each other as to provide a rich matrix of commonalities and differences for interpreting the data.

2 The original outline of the work to be carried out counted on a total of 4 case studies within the GEDII project. This was one point criticized by the reviewers of our proposal. We managed to double the original number to 8.

Case Study	Country	Field / Organization		Team size	Data collected
1	ES	Biomedical Eng / University A		8	5 Interviews
2	ES	Biomedical Eng / Research Inst. B		10	3 Interviews
3	ES	Biomedical Eng / University A		8	3 Interviews
4	ES	Biomedical Eng / Research B		9	3 Interviews
5	ES	Biomedical Eng / Research B		11	5 Interviews
6	UK	Energy Eng / University C		10	6 Interviews
7	UK	Transport Eng / Private D (two locations!)		16	6 Interviews
8	UK	Transport Eng / Private D		8	4 Interviews
			Total	**N=80**	**35 Interviews**

Table 1: Overview of case studies carried out during the GEDII project

Field Access

As mentioned, field access was initially organized through industry contacts by Partner VDE in Germany. Calls for participation through their industry networks produces some interest and requests for further information; however, after initial negotiations spanning even month involving small and quite large multinational companies, no team could be secured to participate during 2016.

In a second step, the search was then escalated to the UK, Spain and Sweden through personal networks of Consortium members. Contacts to management of research institutes could finally secure the participation of 6 teams, 5 in Spain and one in the UK for the autumn of 2016. Through further personal contacts, two more case studies then could be carried out during the Spring 2017 in the UK as well.

Field access to the teams involved a presentation regarding the GEDII project, the topic of (gender) diversity in research teams and the requirements of the case studies. The presentation also entailed an explanation regarding the ethical guidelines and the distribution of the consent agreements for both the badges and the interviews. In all cases, a contact email was left with the team in order to voice any concerns in private between individual participants and researchers. This concerns specifically to make a decision for (not) wearing the badges in the presence of the rest of the group. Research participants were informed that dummy badges were available for them in case they do not want to wear badges but also avoid being “outed” as rejecting participation.

It is interesting to note that while data protection and privacy issues of sociometric badges were our (the research team) primary concerns when introducing our study to potential participants, this was rather of minor importance for the actual team members. The primary interest and question that was present throughout all presentations touched upon the possibility to “fake” the data, i.e. to which degree the mere presence of badges could influence the interactions among team members. Prior studies have demonstrated however

– and this is a finding corroborated by our own participants – that the wearing of badges is relatively quickly forgotten. The attention and energy necessary to actively influence the measurements is quite high, first and foremost since some variables are targeting unconscious bodily reactions that are hard to control at all – such as excitement during interactions or the mirroring of non-verbal gestures.

Field Logistics and Data Collection

In order to ensure comparability across teams, each case study has been carried out according to the same protocol.

- Five days of data collection with sociometric badges during the working hours of the team. On the start of the field work, a short introduction was given to all team members on how to wear the badge, how to turn it on/off and where to pick it up in the morning and drop it off in the afternoon or whenever people were leaving work. Usually, one central place was agreed where people could pick up or drop-off the badge. Badges were delivered in the morning by a GEDII researcher and picked up in the evening in order to download the data and re-synch the badges internal clock (by connecting them to a computer). Team members were also instructed to note down any exceptional occurrences when the badges were turned off or people absent from work. A sheet with the participant names and "their" badge number was also deposited with the team; team members had to make sure they wear the same badge all days.
- At least three semi-structured interviews with team members from each team, including the team leader, one senior and a junior member were conducted. Interviews lasted in general between 50 minutes to 1 hour + 15 minutes. In addition, at least two interviews were carried out with management staff of the participating organization, such as human resources or scientific managers. Interviews were carried out before, during and after the fieldwork with the sociometric badges, depending on the availability of the participants. A total of 35 interviews were carried out (see Table 1). Three interview guidelines had been developed during the preparatory phase of WP2: one for team members, one for the team leader and one for management staff.
- A short online questionnaire was administered to each team member during or shortly after the actual fieldwork with the sociometric badges with the intention to collect socio-demographic variables of each member as well as additional information on certain team characteristics that would allow to contrast and interpret the results of the sociometric profile. The questionnaire gathered information on the following variables: Year of birth, Team Role, Gender, Highest Qualification, Team tenure, Gender Stereotype, Big Five Personality Traits. In addition it included three round-robin items by which team members rated each other regarding psychological safety, social relations and advice seeking. The questionnaire was filled out by 73 out of 80 participants (response rate: 91%).
- In addition, information on the distribution of working places was collected in order to understand the co-location of the team members among each other in terms of a

shared workbench, office space, etc. The actual co-location of team member would provide important information for interpreting the Bluetooth/proximity signals which depend upon physical proximity.

Two pilot studies were carried out with research teams in order to test the overall field logistics (dropping off and picking up of badges), the viability of downloading data for teams of 10 people or more and performing first analysis. The pilot studies lasted one week each; no supporting information such as interviews or questionnaire was collected.

Each GEDII researcher kept a field log where any occurrences – such as problems with the badges, additional comments from participants, etc. – during the data collection were recorded.

Sociometric Setup and Data Export Settings

The sociometric badges were configured using the firmware version 3.1.2669 with the default settings recommended by the user guide (Sociometric Solutions, 2014). The accelerometer data resolution and audio data resolution were adjusted the record at 0.1 second time intervals instead of the default 0.5 second setting. As the guidebooks suggests, 0.5 second resolution is ideal of longitudinal data collection while not being fine-grained enough to allow for accurate turn-taking or body movement mirroring analysis, which are the required metrics for analyzing structured meetings. In order to capture structured meetings, the more fine-grained resolution of audio- and accelerometer data setting is required. However, since team members go into and come out of structured meetings during their working day, the switching of the firmware configuration while in the field is not feasible. Therefore, the firmware was configured to work with the 0.1 resolution by default. This does not diminish the collection of longitudinal data but enables fundamentally the analysis of any structured meetings as they occur during the working days of teams. Overall, the 0.1 resolution produces simply more data which can be dealt with during the posterior analysis but does not condition the actual collection of information per se. The firmware settings are provided in ANNEX IV - Sociometric Badges Firmware Setting on page 126.

The data of the badges was downloaded into the Sociometric Datalab (version 3.1.2468) at the end of each day for 4 out of 5 days. Each case study also had a two-day period where badges were left with the team for two consecutive days – since badges have the capacity to run in an uninterrupted fashion for up to 40hours and store up to 4GB of data, sufficient for two days. On average, one day of data collection per badge produced between 200 to 350 MB of data, depending on the activity of the badge wearer. Across all teams, approximately 2000 hours were recorded.

Overall, the default export settings from the Sociometric Datalab were used, adjusting where necessary the resolution of the audio and body mirroring values. Depending on the necessities, different time resolutions were used: body mirroring values for example were exported at the 1 second aggregation level while audio profiles for the entire team used a 60 second aggregation level.

The export process is quite slow. Depending on the selected resolution as well as the size of the team, body or audio mirroring values take a long time to export basically because mirroring values are calculated in a pairwise fashion between all badges. Exporting the Excel files for interaction, body movement or audio can take several days, since it usually has to be broken down into individual days and variables. Not all possible matrices have been exported for all teams but only those that were necessary for the analysis carried out so far in the GEDII project.

Data Pre-processing and Analysis

Data cleaning and management was carried out exclusively in the R language for both the short case study questionnaire as well as the sociometric data itself. The scripts used for data cleaning will be published on the Github and the Zenodo repository account of the GEDII project. Please check the GEDII website for updates.

The cleaning and preparation of sociometric data involved basically the conversion of Excel exported files into "tidy" data format for easier manipulation and analysis with the "tidyverse" software packages.[3]

The Sociometric Solutions Datalab – which is used for downloading the recorded data from the badges – exports "raw" measurements as Excel files. What "raw" data here means is a little tricky, since for certain variables, the numerical values are not the values recorded by sensors but actually aggregated values. Thus, to give an example, under a firmware configuration of 0.1 seconds, audio volume and fundamental frequency (pitch) is stored every 0.1 seconds. However, when exporting the audio activity sheet, the export parameters allow to specify the aggregation level of the resulting data starting from 1 second upwards. This means that the exported values which have been recorded at a resolution of 0.1 seconds have already been aggregated to at least 1 second (or more) when exported to Excel.

The Datalab also offers the possibility to export certain derived measures such as network statistics on interaction data or turn-taking statistics. However, these standard indicators offer data on a highly aggregated level. In order to work with the "original" data sheets, custom functions had to be written in R in order to clean, process, analyze and visualize the data. All code and functions will be made available on the Github repository.

R packages used included: *circlize*[4] for producing Chord Diagrams; *sna*[5], *igraph*[6] and *ggraph*[7] packages for social network analysis and visualizations; *relevent*[8] package for Relational Event Modeling; *sjPlot*[9] for generating regression tables.

3 See https://www.tidyverse.org/ and (Wickham & Grolemund, 2017)

4 See https://cran.r-project.org/web/packages/circlize/

5 https://cran.r-project.org/web/packages/sna/index.html

6 https://cran.r-project.org/web/packages/igraph/index.html

7 https://cran.r-project.org/web/packages/ggraph/index.html

8 https://cran.r-project.org/web/packages/relevent/index.html

9 https://cran.r-project.org/web/packages/sjPlot/index.html

Open Access Policy

The sociometric data files will be made public in the near future on the GEDII project website (and Zenodo repository) as stipulated by the data management report and Open Access Policy of the project. This involves the interaction data, the speech profile as well as body mirroring values. The data will also include basic socio-demographic variables of team members. Please check the GEDII project website for updates.

Validity of Sociometric Badges Data

The following section contributes to the assessment of the validity of measurements done with sociometric badges. When consulting the provided documentation by Humanize but also early publications that involve sociometric badges, there is little concern regarding the veracity of the generated data (Sociometric Solutions, 2014). The basic modi operandi is to explore possible correlations between sociometric measurements and other variables such as "trust", "operating glitches", "business pitches", etc. without questioning if the incoming data reliably represents the underlying phenomena. This suggests that sensor data is largely equated with being "objective" data simply by the fact that this data is produced without the direct intervention of humans at the time of measurement. However, for some dimensions of sociometric data this assumption does unfortunately not hold up to closer scrutiny.

The first academic publication to seriously draw attention to the validity of the sociometric measurements is Chaffin et al., (2015). The authors highlight the large variability of interaction and audio data and the need to define thresholds for Bluetooth signal strength that influence the characteristics of the observed network. They also highlight the particularly problematic nature of audio data and its derived measures such as speaking duration. Yu et al., (2016) have validated sociometric badges in a simulated emergency department setting, finding that body movement data distinguished reliably between stationary or non-stationary activities; face-to-face detects were under-reported by badges under ideal conditions and badges underestimate speech duration consistently. Most recently, Kayhan et al., (2018) expand this line of critical investigation discovering crucial issues in relation to the synchronization of the badges internal clock: without a precise synchronization of timestamp between badges, however, the ability of identifying events across badges is severely diminished. As a consequence of the lack of precision in detecting speech timing and duration, the badges overestimate turn-taking counts (ibid.). Another publication by Chen & Miller (2017) explores in greater detail the accuracy for total speaking length and the precision of speaking time, confirming largely the variability of measurements depending on length of measurement and contextual factors such as ambient noise. Tripathi & Burleson (2012) look at the correlation between body movement measurements and creativity self-ratings.

As can be seen by the dates of these publications, efforts to scrutinize the reliability of sociometric measurements are fairly recent and coincide basically with the implementation of the GEDII case studies. Alerted by the Chaffin et al. paper in 2015, the GEDII team started itself to carry out a carefully designed experiment assessing in more detail the accuracy of the turn-taking and speech duration measurement capabilities of the badges. Due to the close association of turn-taking for speaker dominance in general and specifically in the context of gendered group dynamics, the reliability of the audio recordings are of crucial importance for the project. Unfortunately, as the present report will show and the forthcoming results (Müller & Meneses, forthcoming) indicate, turn-taking analysis is too imprecise as to be useful for any research purposes.

Our own results stemming from two controlled experiments contribute to the existing evidence in important ways:

- For audio data: By carefully designing the conditions that might influence speech detection and turn-taking analysis, we can quantify the error with respect to different experimental conditions: sex, distance of mouth to badge and distance between speakers.
- For Bluetooth and Infrared: we provide new evidence how distance and angle of positioning of badges influence RSSI signal strength and infrared detection rate.

Both of these experiments will be described in the following sections.

It is probably worth mentioning two consequences of these reliability issues of sociometric badges up front: First, the results cast a shadow on the quality of studies that take sociometric data at “face-value”. The surprising fact indeed is, that impressive results are reported in terms how certain measurements correlate with “creativity” (Tripathi & Burleson, 2012) or creativity and personality (Gloor et al., 2011), or the outcome of speed dating events, elevator pitches, the outcome of salary negotiations, or trading contact information on conferences (all reported in Pentland, 2008) without mentioning these basic reliability issues. Since the results in Pentland (2008) have not necessarily been carried out with the current version of the badges, nor with the standard features of the Sociometric Datalab, it is hard to assess and/or potentially reproduce these results. However, it might be well worth to revisit the results in light of the reported findings by our own study and the critical literature already cited and try to reproduce the results.

Second, as Kayhan et al., (2018) point out, some of the measurement errors can be considered systematic and could be addressed through more powerful analytic techniques such as Machine Learning. This drastically expands the possibilities but also work and knowledge expertise necessary to analyze the data in a more reliable fashion. These possibilities are outlined as future work towards the end of this report. It also reinforces the need for better guidance and protocols to collect more reliable and comparable data from the beginning.

Interaction Data

Sociometric badges produced two types of “interaction” data: on the one hand Bluetooth sensors detect other Bluetooth enable devices within a radius of 1 to several meters. Bluetooth detection can include other devices such as cellphones if configured in the firmware settings. Sociometric badges do not provide any information an absolute positioning (GPS). A mutual detect is stored as the numeric value of the Radio Signal Strength Indicator (RSSI) of the Bluetooth signal. RSSI values usually range between -40 to -90, where smaller values indicating a weaker signal. The strength of the signal can vary due to distance between the devices, due to obstacles (clothing, walls, other bodies between device), the angle in which badges are positioned to each other or simply due to the inherent variability of the measurement itself. The RSSI therefore can give an indication – in the best of cases – of how close or distant badge wearers were to each other at a certain period in time; however, given that the angle between badges as well as obstacles can also influence the signal strength, it only provides a very crude approximation on how distant speakers are positioned to each other. On the other hand, the sociometric badge incorporate an infrared sensor which is much more restrictive in detecting proximity: for a face-to-face detect to take place, badges need to be within a range of 1-1.5 meters from each other and within a 30° cone, according to manufacturer specification. An Infrared sensor detect is binary, i.e. it does not indicate signal strength but simply a timestamp of when it occurred.

Several controlled experiments are described in the following paragraphs with the aim to better estimate the precision and reliability of the interaction data collected during the actual GEDII case studies.

Controlled Experiment 1

Experiment 1 was specifically designed to assess the accuracy of microphone derived measures, particularly speech duration and turn-taking. As a by-product of this controlled turn-taking experiment, proximity data is available, demonstrating the variability of measurements under very "stable" conditions: a) two badges per speaker, speakers facing each other sitting at a table at a b) distance of 0.7 meters and at a c) distance of 1.4 meters.[10]

Proximity - Measurements between two badges worn by same person

First, since each person was wearing two badges, the Bluetooth signal strength between the badges worn by the same person can be visualized (while all other Bluetooth detects being eliminated). Badges in this case are not facing each other, but since Bluetooth works independent of the orientation of badges, this should not influence the data.

Distance	Mean	sd	Min.	1st Qu.	Median	3rd Qu.	Max.	N
10 cm	-60.85	4.99	-81.00	-64.00	-61.00	-57.00	-47.00	580

Tabla 2: Summary of RSSI values of badges worn by same person

As one can see from the minimum and maximum values, although the distance between badges is fixed, the range of values is relatively wide, spanning a relatively strong signal at -47 to the weakest signal at -81, covering the full range of possible values. Even though badges are very close to each other, one can find RSSI signals at the very lower end of possible values.

Proximity - Measurements between badges of face-to-face speakers

When comparing the RSSI values between speakers, a similar observation emerges: the measurements span a wide range of values while a difference in mean values indicates distance between speakers. As the following table demonstrates, the mean value for 0.7 meters is -57.82 while the mean RSSI value at a distance of 1.4 meters is -70.13. Although there is a clear difference that allows to distinguish between the two conditions, again, the range of values produced is relatively wide when comparing maximum and minimum values and the standard deviation.

10 For detailed description of the experimental setting see page 28. As long as not otherwise reported, the data is based on the experiment conducted the 6th of October 2017.

Distance	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	sd	N
0.7 m	-76	-61	-57	-57.82	-53	-47	7.83	512
1.4 m	-90	-78	-69	-70.13	-63	-53	7.45	516

Table 3: Summary of RSSI values during the controlled experiment

In both cases, RSSI values fall into the range of the mean +/- 7.8 and 7.5 units respectively. The relatively strong overlap can be seen in Illustration 1 which displays the RSSI signal strength for 0.7m and 1.4m of distance between speakers. The further away the badges are from each other the more they “fail” to produce strong signals, although maximum values discriminate only weakly between the two conditions, with -47 (0.7m) and -53 (1.4m).

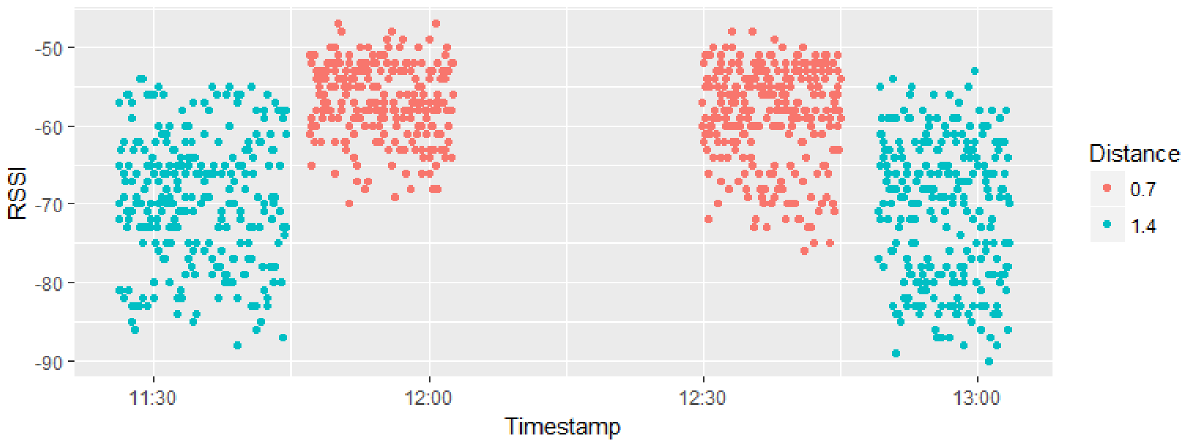


Illustration 1: RSSI signal strength under 0.7 and 1.4 distance of badge wearers

Comparing the measurements between badges worn by the same person (see previous section) with the more “realistic” measurements of badges worn by different persons produces, however, a counter-intuitive result. Surprisingly the RSSI mean value (-60.85) of badges worn by the same person is weaker than the mean value (-57.82) of badges worn by people sitting across from each other, although they are physically closer (0.1m vs. 0.7m). Illustration 2 demonstrates this finding: despite the clear difference in mean RSSI values for the 0.7m vs 1.4m condition (B), the mean RSSI value for the 0.1m condition (A) is not smaller but larger. This suggests that the orientation of badges does indeed matter even for Bluetooth signals.

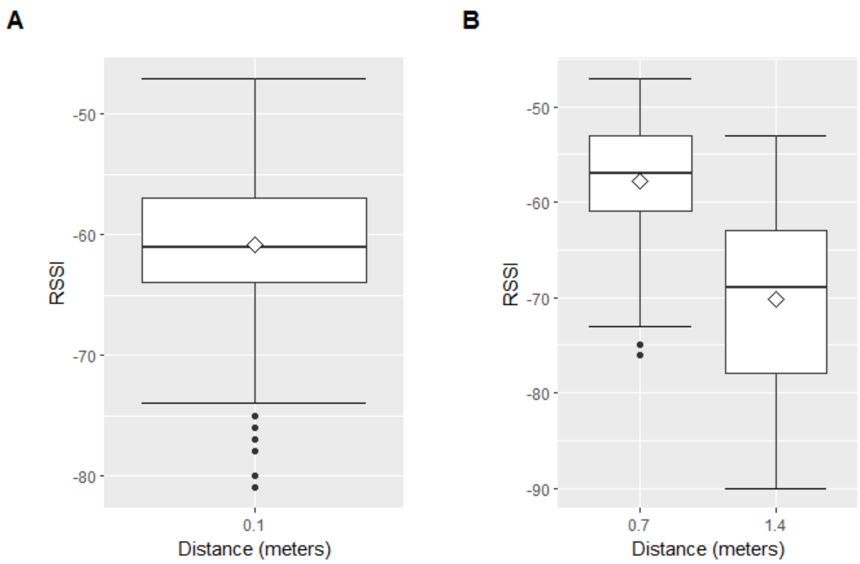


Illustration 2: Boxplots RSSI values for 0.1 side-by-side, vs. RSSI values at 0.7 and 1.4 meters distance for badges facing each other

Infrared / Face-to-face

Regarding face-to-face (or Infrared detects), the number of detects should be relatively constant given that the manufacturer specifications situate the ideal distance from 1 to 1.5 meters. As the following table shows, the number of detects is conditioned by the distance between badges:

Date of Experiment	Distance	Count
2017-10-06	0.7m	9424
2017-10-06	1.4m	4036

Tabla 4: Number of infrared detects by distance

The closer the badges are to each other, the more frequently are face-to-face interactions detected. The detects roughly double as the distanced is halved from 1.4 to 0.7 meters!

Controlled Experiment 2

A second controlled experiment was carried out in order to assess the a) mean RSSI value as proxy for physical distance and b) test how the angle between badges affect face-to-face detection count.

Badges were placed in fixed positions on small wooden pedestals 16 cm above the floor as show in Illustration 3. Overall there were 6 positions in 0.5 meters increments ranging from 0.5 to 3 meters. Four axis were defined, positioning badges in various angles to each other:

- P1 – P2: 0°
- P1 – P3: 10°

- P1 – P4: 45°
- P1 – P5: 90°

The different angles correspond to the manufacturer specification which defines face-to-face interaction as happening within a cone of +/- 15°. This means that P1-P2 as well as P1-P3 are within the given angle for optimal face-to-face detects while P1-P4 and P1-P5 are not. While the badge at P1 faces with its sensor "upward" towards P2 the relative rotation of each badge changes between positions. P2, P4 and P5 all have their infrared sensor pointing towards P1 (the front of each badge is marked with the "v"). In other words, while P1-P2 point their infrared sensors towards each other, P5 points towards P1, while P1 always faces towards P2.

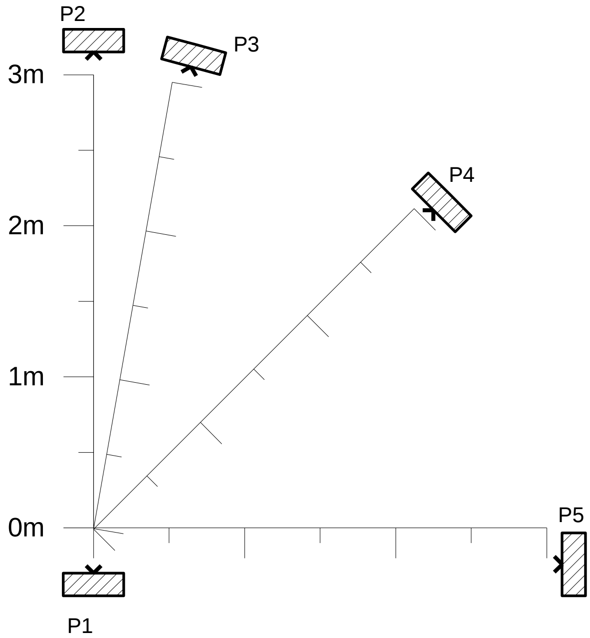


Illustration 3: Birds-view perspective: badges experimental setup

Badges were occupying each of the positions 1 to 5 in a clock wise fashion, such that each badge would "visit" each positions for each of the 6 different distances. Badges were at rest during 10 minutes in each position and then move towards the next position/distance. This produces an overall of 30 sessions (6 distances on 5 positions) each lasting for 10 minutes. Badges were not turned off during the rotation/moving; however, the start of each observational period was marked by a loud beep. This allowed to identify the exact timestamp of the start of the observational period and filter out interaction detects happening during the moving of the badges.

Data preparation involved to match the badge number with the observational position (P1, P2, P3, P4, P5) and add the angle of each dyad. Distances and angles between position P2, P3, P4, P5 were initially not included in the analysis.

Bluetooth / Proximity

Table 5 shows the mean RSSI values as the distance between badges increases across all badges (on the left) and by varying angle (on the right). Although there is a clear difference in mean RSSI value from 0.5 to 3 meters, the 0.5 increments do not discriminate very well. The mean values for 1 meter and 2 meters are almost identical and the difference between 0.5 to 3 meters is also quite weak. It is also noteworthy that the mean figures differ quite substantially from the first experiment (see Table 3) where 0.7 meters produced a mean RSSI value of -57 and the 1.4 meters distance a mean value of approximately -70! Thus, although there is a relative difference between the physical distance and the RSSI values, where larger distances produce weaker signals, this does not correspond to a fixed, physical distance. Other factors seem to influence the overall RSSI mean apart from the actual physical distance.

Distance	Mean RSSI	SD	Count	Angle	Mean RSSI	SD	Count
0.5	-48.06	6.05	1748	0	-44.50	3.76	221
				10	-44.42	4.64	212
				45	-47.81	4.04	216
				90	-55.24	4.01	225
1.0	-58.57	8.58	1688	0	-53.25	5.15	212
				10	-53.56	7.15	216
				45	-58.76	6.05	199
				90	-68.58	4.87	217
1.5	-55.44	7.93	1660	0	-50.20	4.10	224
				10	-51.30	5.29	208
				45	-55.24	4.17	198
				90	-65.83	6.31	200
2.0	-58.01	8.04	1788	0	-52.69	4.61	232
				10	-54.48	5.37	217
				45	-57.45	4.03	221
				90	-67.51	7.79	224
2.5	-59.32	7.21	1716	0	-55.71	5.29	220
				10	-54.72	4.40	204
				45	-59.05	4.46	209
				90	-67.28	6.30	225
3.0	-61.48	6.88	1676	0	-57.22	4.39	216
				10	-56.98	3.49	198
				45	-61.91	4.47	210
				90	-69.53	5.76	214

Tabla 5: Mean RSSI values by distance and angle

As can easily be seen, one other factor influencing the strength of the RSSI signal is the orientation of the badges towards each other. When badges face each other directly, the RSSI signal is stronger as when badges are perpendicular to each other. Although – according to manufacturer specification, the orientation of badges should not play a significant role, in the current experiment it clearly influences the mean RSSI value. In fact, taking the angle in which badges are placed as the sole variable irrespective of distance, shows a considerable difference of 12.78 units which is just a little bit smaller than the difference of mean RSSI values by distance (61.48 – 48.06 = 13.42)!

Angle	Mean RSSI	SD	Count
0	-52.23	6.15	1325
10	-52.53	6.57	1255
45	-56.66	6.40	1253
90	-65.61	7.69	1305

Tabla 6: Mean RSSI by angle

This implies that both the angle in which badges are placed as well as the distance influence to a considerable amount the actual mean RSSI value available.

Infrared / Face-to-face

This second controlled experiment also addressed specifically the precision with which Infrared detects depend on the distance and orientation of the badges towards each other. The manufacturer specifies that badges have to be within a cone of 30 degrees and within a distance of 1 – 1.5 meters in order to detected.

As table 7 shows, badges have Infrared detects well beyond the 1.5 meters distance reaching 2 and even 3 meters, if they face each other directly. However, the detection count is indeed dependent upon the angle in which badges are placed. At 0° and 10° there is virtually no difference, which is correct since both badges are within the manufacturer specified cone of 30°. Even at 45° and distances and 1.5 meters distance, face-to-face detects are happening. What the badges detect reliably are larger angles: two badges being placed perpendicular to each other have no face-to-face detect at any distance, even at 0.5 meters.

Distance	Count	Angle	Count
0.5	11556	0	4591
		10	4367
		45	2598
		90	-
1.0	10797	0	4188
		10	4405
		45	2204
		90	-
1.5	8008	0	3986
		10	3660
		45	362
		90	-
2.0	5465	0	3387
		10	2078
		45	-
		90	-
2.5	3502	0	2246
		10	1256
		45	-
		90	-
3.0	393	0	387
		10	6
		45	-
		90	-

Tabla 7: Face-to-face counts by distance and angle

Overall, looking at Bluetooth and infrared detects simultaneously indicates that face-to-face detection counts is both a function of distance and angle. Similar, RSSI values vary relatively little with regard to distance, especially when considering a distance from 1 meters – 3 meters, ignoring mean values below < 1 meter. The angle appears to be at least as decisive: the mean RSSI value at 1m/90° is weaker (-68.57) than the mean value at 3m/0° (-57.22)!

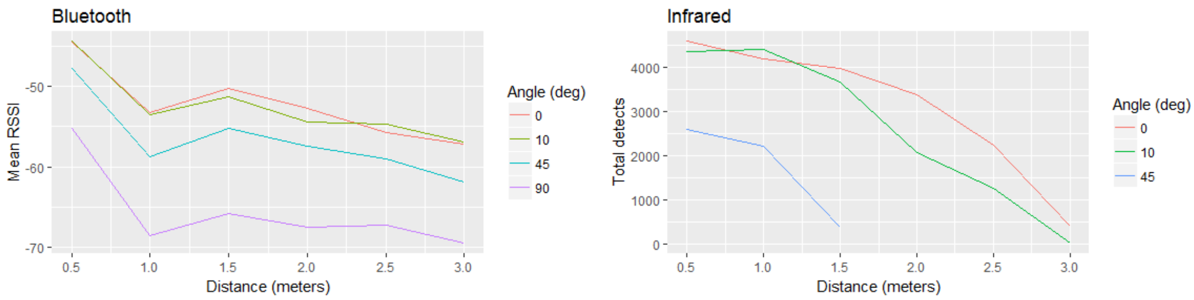


Illustration 4: Bluetooth and Infrared detects by distance and angle

Infrared - Contribution Index

This second experiment also provides the opportunity to check on one of the infrared derived measures, namely the "contribution index". The contribution index is calculated based upon the face-to-face detects, where a difference can be drawn between in-coming and out-going detects (which badge detects the other as being within its infrared cone). The contribution index is defined as the (number of send messages - number of received messages) / total messages. It will be -1 for badges that only receive (are only looked at) and +1 for badges that only send (look at others but never are looked at). Theoretically, the CI is dependent upon the angle at which badges are positioned to each other. Badges facing each other with their sensors directly should detect each other to equal amounts. Incrementing the angle at which badges are placed should increment the count at which face-to-face detects occur.

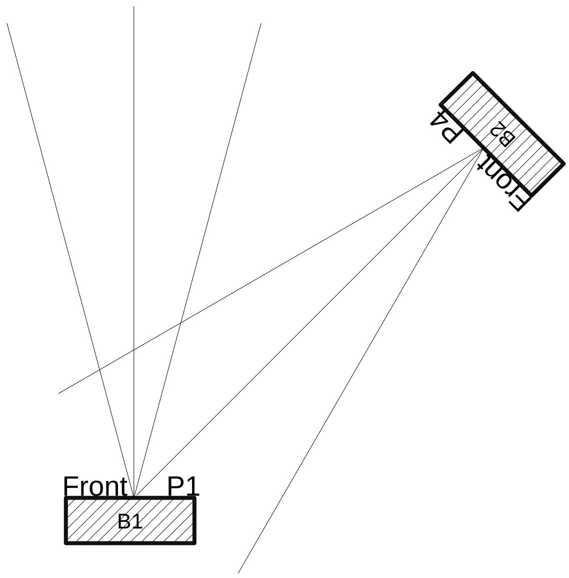


Illustration 5: Contribution Index Schema

Positions	Angle	Count	Contribution Index
1-2	0	8917	-0.05
2-1	0	9868	0.05
1-3	10	8653	0.09
3-1	10	7119	-0.09
1-4	45	4681	0.81
4-1	45	483	-0.81

Tabla 8: Direction of Infrared detects

The most pronounced difference occurs at the 45° angle where there is a clear difference between the “detecting” and the “being detected” badge. However, considering the given detects in relation to Illustration 5 which demonstrates the placement of badges in relation to each other, the surprising fact is that for P1-P4 most detects are attributed to P1 and not P4, although P1 is inside the “visibility” cone of P4 while P4 is not inside the “visibility” cone of P1!

Noteworthy is also the fact that there are no detects for position P1 – P5, i.e. when infrared sensors of badges are placed perpendicular to each other. Although P1 is placed inside the “detection” zone of P5 – because it faces towards P1 – there are no face-to-face detects taking place.

Audio Data

Sociometric badges have a build in microphone that records audio signals. Audio data is sampled at a frequency of 8 kHz. There are two microphones integrated, one in the back and one at the front. By default, badges do not record actual audio files (they do not record what people say) but simply the numerical values of the volume and pitch. Building upon these raw values, the Sociometric Datalab exports several data files, including speech profiles (indicating if badge wearer was talking), speech participation, and turn-taking analysis.

The quality of the sociometric audio data is of paramount importance for the GEDII project. As mentioned the scientific literature does suggest marked gender differences in speaking dominance, leader emergence and turn-taking (see introduction). According to the manufacturer specification and various publications using the audio features, the collection of speech data seems rather unproblematic and reliable. However, Chaffin et al. (2015) raises serious concerns regarding the reliability of sociometric audio data. Thus, a specifically designed experiment was carried out to assess turn-taking counts under strictly controlled conditions.

Controlled Experiment 1

Method

The experiment was designed to test the accuracy of the audio derived measurements of the sociometric badges. The microphone samples audio data at 8 KHz. Based on the raw data, the Sociometric Datalab derives the turn-taking as well as speech duration measures. While data is recorded by the badge, it is during the exporting of turn-taking and speech (profiles) that the internal algorithm produces the corresponding measures.

The principal difficulty for each badge is not only to detect when a person speaks but also to correctly assign the speaker to “its” badge. Badges do not “know” around which neck they hang and have to distinguish between “its” speaker, ambient noise and other speakers standing close by. In order to test specifically the badges capability to determining “own” vs. “other” speakers, the following set of conditions were established:

(a) distance of badge to mouth. The closer the badge to the mouth of the speaker, the “easier” it will be for the badge to detect when a person speaks or is silent due to the stronger vs. weaker audio signal. Each person wore two badges, one closer to the mouth and one further away.

(b) Same- and mixed sex. Voice pitch is a powerful indicator for differentiating speakers. The voice of women has a higher pitch than the voice of men. As a consequence, it should be “easier” for badges to detect the correct speaker in mixed sex settings when the difference of pitch is wider than in same sex settings.

(c) Distance between speakers. The further away speakers are from each other, the “easier” it will be for badges to differentiate between “own” vs. “other” speaker due to the strength of the audio signal. Speakers further away will be less loud and thus fail to activate the speech detection of the “other” badge. Speakers were thus placed 0.7m and 1.4m from each other.

Three carefully scripted text fragments were created to simulate a dialog between two persons A) without interruptions, B) unsuccessful interruptions and C) successful interruptions based upon the specification of the Sociometric User Guide. Each script roughly lasts for about 2 minutes. The scripts were read out by two persons, each following its assigned part. The whole experiment was carried out in one large session, in a quite room.

Each of the three scripts was read out by three times for each of the conditions (except the distance to the mouth), yielding a total of 3 (repetitions) x 3 (texts) x 2 (sex) x 2 (distance speakers) = 36 sessions. The following table gives an overview of the conditions:

Condition	
Text	No Interruptions
	Unsuccessful Interruptions
	Successful Interruptions
Sex	Same Sex
	Mixed Sex
Distance (Badge Mouth)	Close
	Distant
Distance (Speakers)	0.7 meters
	1.4 meters

Tabla 9: Conditions of turn-taking experiment

The export of the speech duration and turn-taking was carried out with the Sociometric Datalab settings for “Structured meeting” which export audio signals at 1 second resolution. Since each start of the reading of the script was marked with a loud beep, the precise timestamped sessions for each script/condition was determined. The sessions were then imported into Sociometric Datalab in order to export the speech duration and turn-taking for the exact duration of the actual reading of the script.

In addition, the overall experiment was video recorded and time-coded using Atlas.TI indicating the speaking time of each person as well as the overlap duration. The actual speaking time was used in the subsequent analysis and compared to the speaking duration detected by the sociometric badges.

Through the video coding as well as the design of three different types of texts, the target values for turn-taking as well as speech duration were given and used in the subsequent analysis.

Speaking Duration – Results

The Sociometric Datalab exports a speech profile containing several measurements.

- Speaking time for each badge: only the person wearing the particular badge is peaking, while all other persons are silent
- Speaking overlap: the total duration the badge wearer was speaking while other badges were also speaking
- Total speaking time: the total duration a given badge wearer was speaking (while others are silent) + the total duration s/he was overlapping with others.

In addition, the Datalab exports further metrics such as the total duration of “Listening” as well as “Silent” time for each badge. However, these metrics were not considered in this experiment.

The following table provides the mean values across all texts for each of the 6 conditions for speaking duration measured (badge) versus video coded. As it appears, the measurements are quite consistent across the different conditions at first sight; the badges consistently under-estimate the real speaking duration ranging from a couple of seconds up to 7 seconds.

Dist. to mouth	Sex	Dist. to speaker	Speak (badge)	Speak (video)	Speak total (badge)	Speak total (video)	Overlap (badge)	Overlap (video)
Close	Mixed sex	0.7	25.46	32.85	38.22	45.29	12.76	12.44
Close	Mixed sex	1.4	29.52	33.40	36.94	45.63	7.42	12.23
Close	Same sex	0.7	29.06	33.44	36.17	45.27	7.11	11.84
Close	Same sex	1.4	32.32	35.91	39.17	47.76	6.86	11.85
Distant	Mixed sex	0.7	24.71	32.85	33.53	45.29	8.82	12.44
Distant	Mixed sex	1.4	26.44	33.40	38.27	45.63	11.83	12.23
Distant	Same sex	0.7	27.65	33.44	36.88	45.27	9.23	11.84
Distant	Same sex	1.4	31.51	35.91	38.94	47.76	7.43	11.85

Tabla 10: Mean speech duration per badge over all texts - badge vs. video coded duration

When aggregating the speaking time across all sessions, the overall mean approximates the video-coded speaking time. As the following boxplots demonstrate, when examining individual sessions, outliers do exist, suggesting the speech recognition becomes more precise the longer it is carried out.

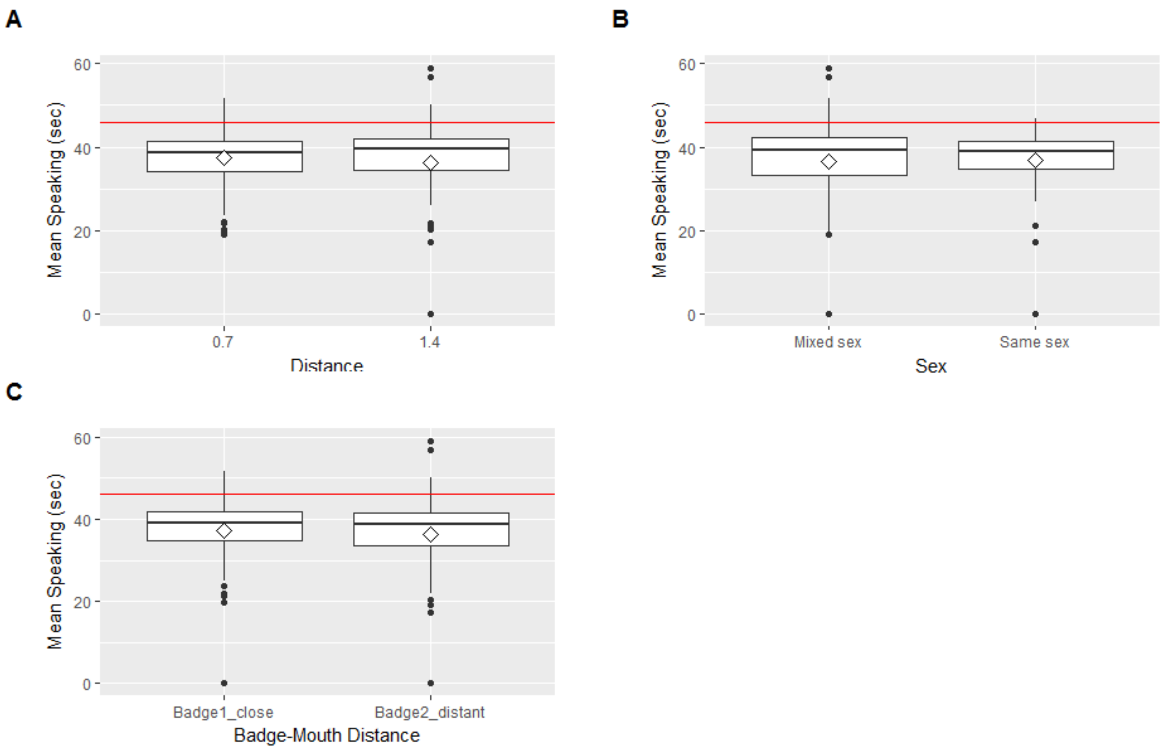


Illustration 6: Boxplots of mean speaking time for each of 6 experimental conditions. Red lines indicates the video coded duration

Turn-Taking

Examining the turn-taking capabilities of the sociometric badges provides a rather different view. As the following table demonstrate, the algorithm over-estimates considerable the number of real turns. Given the simplest script where each speaker has precisely 5 turns, one after the other, the Sociometric Datalab detects on average 12.25 turns for the mixed sex and 0.7 meter distance condition. Surprisingly, the count on turns is better for same-sex setting (8.87) and best (as expected) when speakers are more distant (1.4m) to each other (5.0).

However, problematic is the fact that the algorithm detects consistently approximately 5 “successful interruptions” when there are actually none and even unsuccessful interruptions when there is absolutely no overlap between speakers.

Text	Sex	Dist. Speaker	Turns (badge)	Turns (real)	Success (badge)	Success (real)	Unsuccess. (badge)	Unsuccess. (real)
No Inter.	Mixed	0.7	12.25	**5**	4.25	**0**	8.00	**0**
No Inter.	Mixed	1.4	6.17	**5**	5.08	**0**	1.08	**0**
No Inter.	Same	0.7	8.67	**5**	4.92	**0**	3.75	**0**
No Inter.	Same	1.4	5.00	**5**	5.00	**0**	0.00	**0**
Success Int.	Mixed	0.7	10.25	**1**	4.17	**3**	6.08	**0**
Success Int.	Mixed	1.4	6.92	**1**	4.33	**3**	2.58	**0**
Success Int.	Same	0.7	8.25	**1**	5.08	**3**	3.17	**0**
Success Int.	Same	1.4	6.58	**1**	5.08	**3**	1.50	**0**
Unsucc. Int.	Mixed	0.7	15.58	**2**	4.67	**0**	10.92	**2**
Unsucc. Int.	Mixed	1.4	11.58	**2**	4.00	**0**	7.58	**2**
Unsucc. Int.	Same	0.7	11.58	**2**	4.17	**0**	7.42	**2**
Unsucc. Int.	Same	1.4	9.83	**2**	4.92	**0**	4.92	**2**

Tabla 11: Mean (three scripts) turn-taking badge measure vs. real counts

The precise error rate for each of the conditions is subject to forthcoming publication (Müller & Meneses), however, it can be seen from the following graph that some conditions (specifically the distance to the speaker) has an effect on the precision of the turn-taking measures. The further away the speakers are from each other, the higher the correct number of turns (blue points on red line) whereas the shorter distance (red points) are almost never correct.

Illustration 7: Distance and sex conditioning turn-taking measurement

The error rate of turn-taking counts also emphasizes the need to have a better grip on the actual distance during interaction. The further away speakers are from each other clearly improves the error rate of the turn-count.

Body Movement & Mirroring

The sociometric badges are equipped with an accelerometer which measures the activity (body movement) of the badge wearer. Accelerometer data is sampled with a frequency of 20Hz and provides information about the activity level of the person wearing the badge. High activity levels (> 0.2) indicate that a person is moving (walking) around, while lower values (< 0.2) indicate a stationary activity (sitting). Similar to the interaction and audio data, the Sociometric Datalab exports several derived data files regarding the "consistency" and "mirroring" of body activity patterns between badge wearers.

One of the principal difficulties when examining these body-movement data and their derived "honest signals", is the fact that they constitute a new level of data collection that is hard to compare and match with other more conventional types of measurement and observation. One of the interesting aspects of "honest signals" is precisely that they provide access to a level of non-verbal communication that is hard, if not impossible to observe with traditional research methods. The fine-grained mirroring of body movements, or the variability of activity patterns in the voice is not accessible by the human senses directly and especially not over longer periods of time. However, this produces simultaneously the difficulty to interpret the available data. Correlating certain "mirroring" values with a desired outcome variable is one way as long as an outcome variable can be isolated. In the case of the case studies with teams, these variables do not exist, since interactions among team members are not bound to a certain unique, purpose.

During the GEDII project no dedicated experiment was carried out regarding the body movement measurements and derived values. Kayhan et al. (2018) has recently started to perform basic, mechanical checks regarding the movement of badges and resulting measurements. However, there are currently no controlled experiments that would probe further the reliability and the construct validity of accelerometer derived measurements. It should be noted that some of this work is currently carried out in the computer science and machine learning community. Work on "Social Signal Processing" aims to train algorithms to correctly identify "dominance" (Jayagopi, Hung, Yeo, & Gatica-Perez, 2009; Varni, Camurri, Coletta, & Volpe, 2009), "fights" (Fu, Leong, Ngai, & Chan, 2015), leader emergence (Sanchez-Cortes, Aran, Jayagopi, Schmid Mast, & Gatica-Perez, 2013; Wang et al., 2012) or different types of social actions (Sapru & Bourlard, 2015) in real-time interaction scenarios (Kennedy & Ellis, 2003).

Honest Signals?

Much of the attraction regarding sociometric badges resides in the possibility to track "honest signals", i.e. a basic layer of of semi-automatic, non-verbal communication that shapes human behavior (Gloor et al., 2011; A. Pentland, 2008). However, given the results of the measurement validation exercises carried out by others as well as by the present research casts a certain doubt upon the possibilities of sociometric badges. Nevertheless, the main dimensions of "honest signals" will be briefly rehearsed. There are four different honest signals, as described by Pentland (2008): emphasis, activity, influence and mimicry. These four ways of "signaling" draw mainly upon the audio features and accelerometer data. The following paragraphs summarize Pentland's description quite literally without necessarily critically engaging with the statements at this point.

Consistency / Emphasis

Emphasis is commonly considered a measure of the speaker's motivation and cognitive load. The consistency of variations provide cues of a speaker's mental focus and of her/his openness to influence from other people, respectively. Emphasis can be measured by variation in speech prosody — specifically, variation in pitch and volume. Prosody refers to non-linguistic speech features that are longer than one phonetic segment such as intonation or rhythm. These features guide listeners about the intent of the speaker. Lack of variation in volume and pitch can indicate mental focus and determination; lack of consistency in speech prosody can be sign of emotionality such as when a speaker goes from a whisper to shouting. Pentland and collaborators have measured emphasis by "extracting the speaking energy and the frequency of the fundamental format for each voiced segment, and then we calculated the standard deviation of the energy and frequency measures, each scaled by their respective means. We measured each speaker's emphasis by the sum of these scaled standard deviation."

To the best of our knowledge, there is no published research validating under controlled experimental settings the reliability of "consistency" measurements of sociometric badges.

Activity

Conversational dynamics can be characterized by different levels of activity. As a "honest signal" activity indicates interest and excitement. Activity is basically measured by conversational speaking time, using the speech profile of the sociometric badges. Activity can also be measured by the accelerometer readings which indicate if a person is stationary or walking for example.

Influence / Engagement

Influence indicates the amount of control and/or influence one person has over another. This can be measured by calculating overlapping speech segments. Influence is a signal of dominance. Moreover, its strength in a conversation can serve as an indicator of attention.

Pentland measured engagement by modeling each participant's individual turn-taking using a Hidden Markov Model and then calculating the conditional probabilities that connected these two HMMs to estimate the influence each participant had on the other's turn-taking dynamics. "By quantifying the conditional probability of person A's current state (speaking versus non-speaking) given person B's previous state, we obtain a measure of person B's influence over the turn-taking behavior."

The Sociometric Datalab does provide global figures regarding turn-taking. As demonstrated these figures are not reliable. An alternative consists of using the individual speech profiles which indicate when badge wearers are "speaking" or "silent". As these data is timestamped, a conditional probability of speakers influence over each other might be calculated. However, controlled experiments should be carried out to assess the reliability of these measurements and influence models.

Mimicry / Mirroring

Mimicry / mirroring is the un-reflected copying of one person (her gestures and prosody) by another one during a conversation. The mimicry is key for the smoothness of interaction and can signal emphatic understanding. Ideally it would provide a window on the "social sensitivity" of interaction partners. On the one hand, this copying is performed on the level of speech, i.e. through short interjections (e.g. "uh", "mhh", "ahh") or back-and-forth exchanges consisting of short words (e.g. "OK?", "wow!", "got!"). However, as Pentland remarks, since mimicry is the result of a complex behavior, it is difficult to computationally measure.

The Sociometric Datalab does not provide measurements regarding the mirroring of speech features. Turn-taking measurements go somewhat in this direction – however, they are not reliable as already mentioned. There are no "similarity" measures regarding speech profiles across badges.

On the other hand, sociometric badges produce "similarity" scores based upon the accelerometer measurements which give an indication of the mirroring of two body activity patterns. However, again, controlled experiments should be carried out in order to assess the reliability of these scores and derived models.

Descriptive & Comparative Insights (8 Case Studies)

The following section provides a comparative overview regarding the main sociometric data dimensions (interaction, audio, body movement) across the eight case studies. The comparative view is an important step in the interpretation of sociometric data since it provides a first benchmark regarding basic quantities. Do 5000 proximity detects per week and team member constitute a lot or few? What conclusions can we draw from comparing proximity vs. face-to-face counts across teams? Are the consistent differences regarding speaking time according to role?

Interaction

The most straightforward comparison between teams consists of the frequency of proximity detects (Bluetooth) and face-to-face detects (Infrared). As the following two Illustrations show, there exists considerable differences between across research groups as well as within the teams themselves regarding the proximity and face-to-face profile.

Illustration 8A shows the absolute proximity detect for each team, ranging from a minimum of 9998 (T1) to a maximum of 201.906 detects (T7)[11]. The absolute number of detects is of course not only an expression of how much team members are in proximity to each other but also dependent upon the size of the team: for each additional team member (and hence additional badge), the total number of detects will increase assuming that team members are not completely isolated from each other.

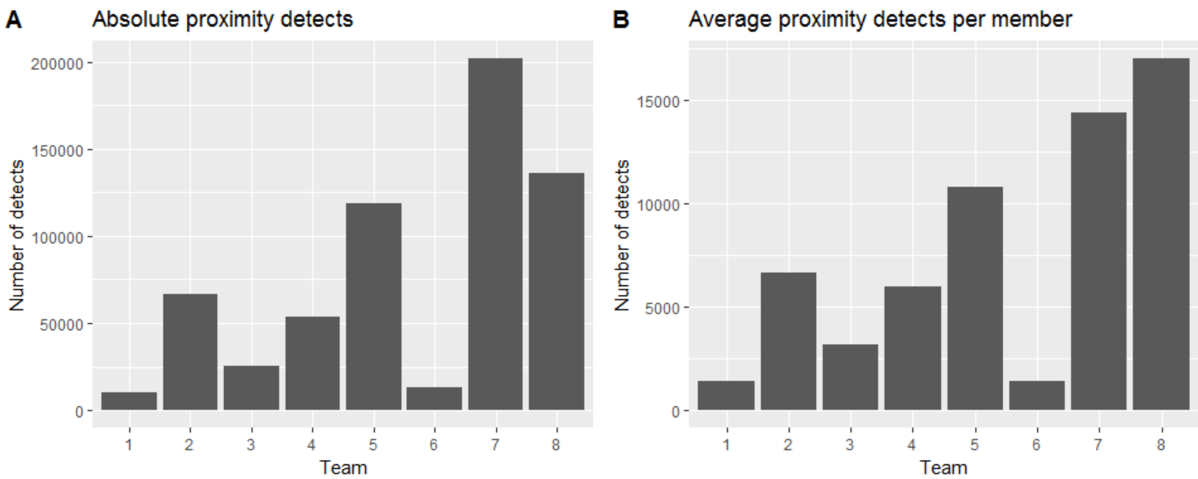


Illustration 8: Absolute and average proximity detects for each team

In order to correct for the effect of team size, Illustration 8B indicates the average proximity detects per team member for each group. Independent of team size, this graph gives a more adequate account of the frequency with which team members are in proximity to each other. For example, note the change between T7 and T8 between both graphs: taking into account

11 This covers a 5 day recording period for all teams except T7 which was monitored during two weeks. Hence is higher absolute detects.

the size of the team, T8 has most “interactions” compared to T7 which has the highest number of proximity detects in absolute terms.

With regard to face-to-face detects, some interesting observations can be made. Again, there is a difference in terms of absolute and mean detects per team member with visible effects for T3 and T5. Team 5 and Team 7 have equally high face-to-face interactions followed by T2 and T3.

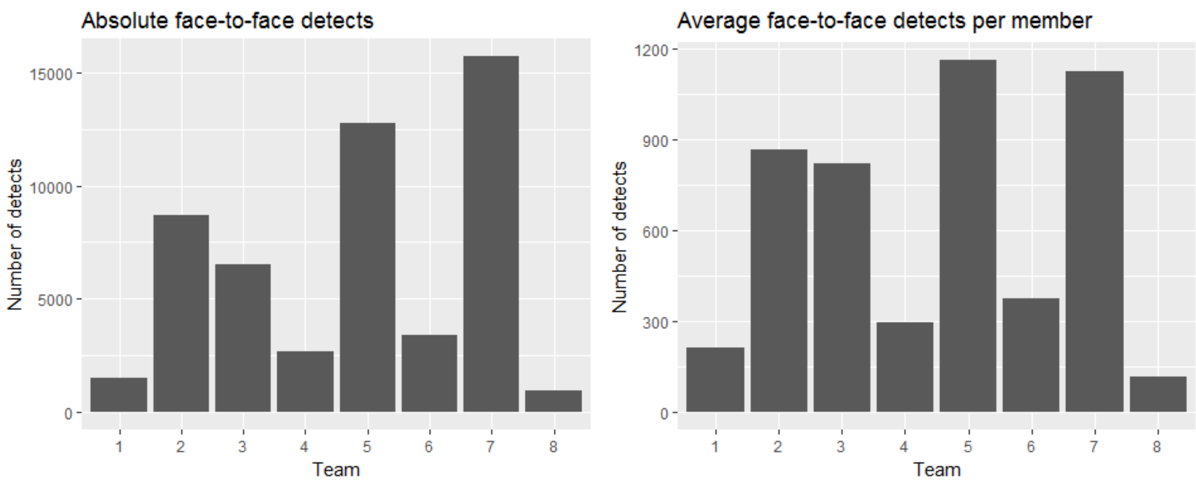


Illustration 9: Absolute and average face-to-face detects for each team

Comparing Proximity and Face-to-face

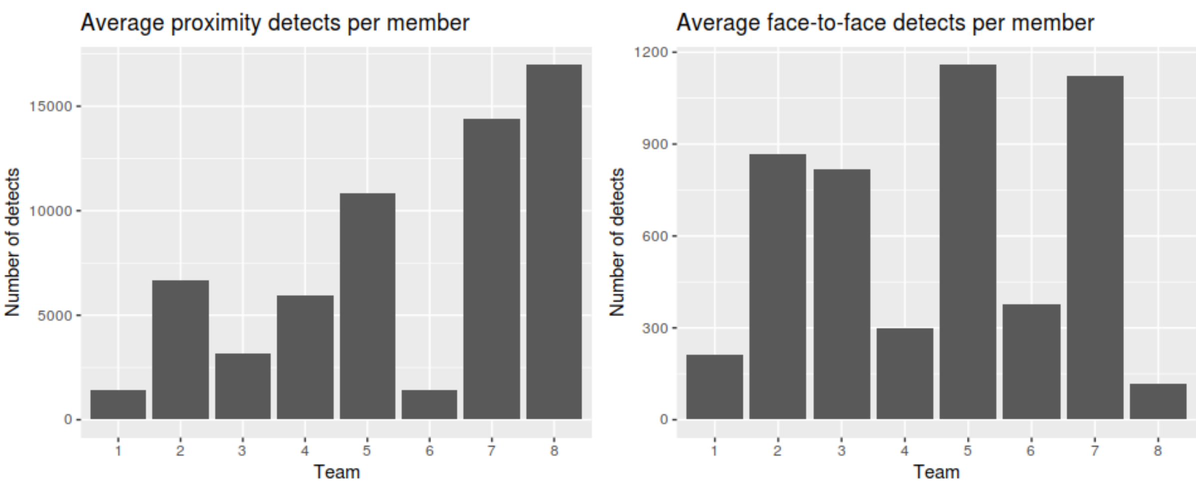


Illustration 10: Comparing average proximity with face-to-face counts

Examining frequency counts for each team on its own poses the questions on how to interpret the average (or absolute) counts. Are 53000 proximity detects a lot for one week of interaction or few? Comparing the average counts across teams provides some leverage in order to start answering these questions. Table 12 list the absolute and mean proximity as well as face-to-face detects. Mean proximity detects per team member is 15224 while the mean face-to-face detects per member is 1306 over a period of 5 days for all teams except

T7. A more intuitive measure is the mean detects per team member and day. Here, for example, members of T5 interact on average 464 times per day, while members of T3 do so 374 times and T8 47 times respectively over one working day.

Apart from giving average detects per team member and day, some interesting observations emerge when comparing proximity- and face-to-face detects (see Illustration 10). What can be easily noticed is the change between the average proximity count and average face-to-face count for each team: T8 for example has the highest proximity count but the lowest face-to-face count. T6 on the other hand has one of the lowest proximity counts but does comparatively well when examining the average face-to-face counts per member. Basically, the average proximity count indicates how "packed" people are at their workplace, i.e. how much space each team member has. The less space, the closer people are to each other providing more opportunities to the Bluetooth proximity signal to detect each other. If team members have separate offices, the badges will detect each other much less frequently, that is, only when people actually do meet. Thus, we can observe that T8 is sharing a very tight office space without meeting face-to-face. T5 on the other hand, equally does share the same lab/office space indicated by the high average proximity detects per team member. At the same time, these opportunities to interact are taken advantage of because members of this team have the highest face-to-face detects. For T3 and T6 finally, the opposite seems to hold: team members do not share a common office space, and when they do meet, i.e. when they are close to each other, they do so for explicit face-to-face meetings.

The teams could be analyzed according to the percentage of face-to-face / proximity detects (see Table 12) as indicated by the "Ratio" column. T6 and T3 have the highest proportion of face-to-face to proximity detects: 26%. This means that for every face-to-face detect we have approximately 4 proximity detects. T8 on the other hand has a proportion of 0.01, i.e. for every face-to-face detect there are approximately 100 proximity detects.

	Absolute[12] proximity	Absolute F2F	Mean[13] proximity	Mean F2F	Mean F2F x Day	Mean RSSI	Skew[14]	Kurtosis	Ratio
T1	9998	1493	2857	427	85	-76.57	0.64	3.00	0.15
T2	66570	8674	13314	1735	347	-78.11	0.83	3.08	0.13
T3	25336	6545	6334	1870	374	-75.44	0.65	2.81	0.26
T4	53571	2681	11905	596	119	-78.15	0.92	3.46	0.05
T5	119042	12772	21644	2322	464	-76.93	0.86	3.18	0.11
T6	12905	3381	2868	845	169	-70.52	0.03	2.15	0.26
T7	201906[15]	15719	28844	2418	242	-76.07	0.52	2.79	0.08

12 Absolute proximity and absolute face-to-face counts refer to pairwise detects

13 Average proximity and average face-to-face counts refer to mean detects per member (and not dyads). Each dyad detect therefore counts as 1 detect for each of the participating badges.

14 Skew and Kurtosis refer to the Histogram of RSSI values

15 Absolute proximity and face-to-face counts for T7 are based on data collection over two weeks instead of one.

	Absolute proximity	Absolut e F2F	Mean proximity	Mean F2F	Mean F2F x Day	Mean RSSI	Skew	Kurtosis	Ratio
T8	136107	948	34027	237	47	-77.52	0.64	3.21	0.01
		Mean	**15224**	**1306**					

Tabla 12: Proximity and face-to-face detects by team

Histogram of RSSI Frequency

When considering the proximity/face-to-face ratio in relation to the Histogram of frequency of RSSI values confirms the diagnosis of T6: it is the least skewed distribution where stronger signals are almost as frequent as weaker ones (this is also confirmed by the higher mean RSSI value, see the previous table). Again, this means that when people do meet and the possibility of Bluetooth detects are given, these signals are relatively strong, i.e. people are close to each other and interact directly (as indicated by the high face-to-face detects as well).

Relatively strongly skewed distributions (see the “Skew” values in Table 12) are T4 and T5 indicating ample opportunity of badges to pick up on Bluetooth signals, even when team members are quite distant to each other. For both teams there are ample opportunities to interact, which are taken up by T5 (ratio is 0.11) and to a lesser extend by T4 (ratio is 0.05).

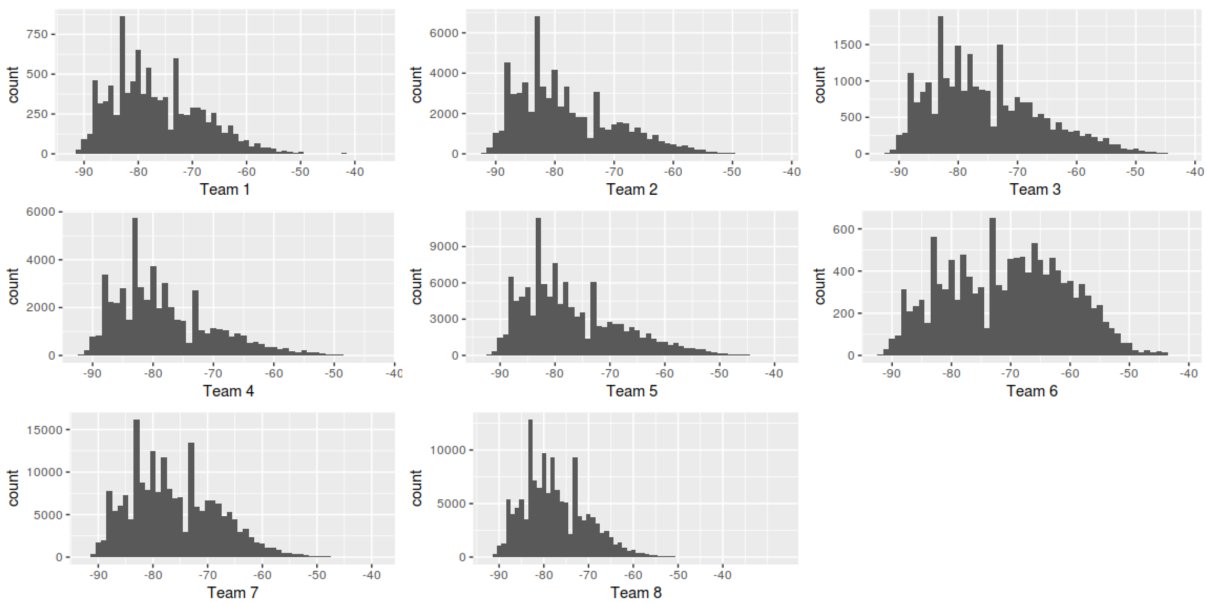


Illustration 11: Histogram of RSSI values for each team

Session Size – Shared Lab Index

A further comparison deduces session sizes for each team, i.e. how many badges are usually involved in “meetings”. This tells us something about the “meeting” culture of the team, i.e.

if it meets in large groups and rather infrequently or has ad-hoc, casual, interactions in small groups.

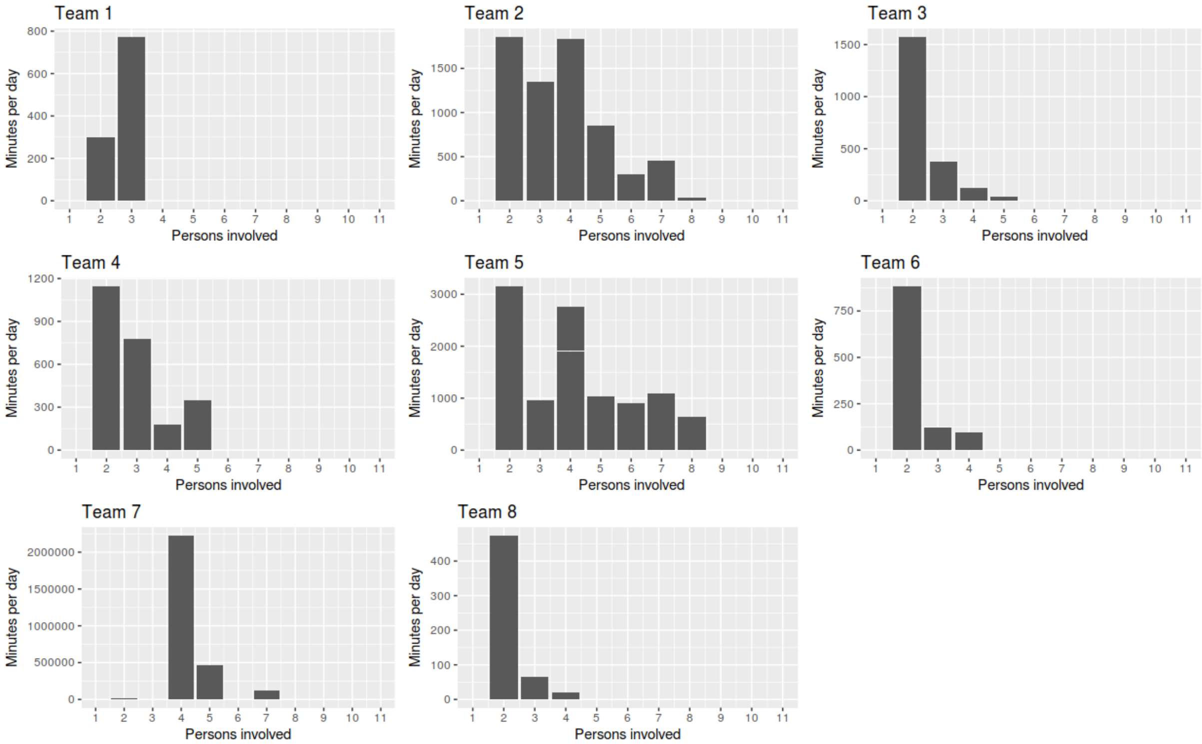


Illustration 12: Minutes per day of face-to-face detects by number of involved badges

As the above illustration shows, T1 for example interacts face-to-face more frequently in groups of three than in groups of 2: the accumulated duration of “meetings” with three badges is approximate 800 minutes per day while meeting with only two badges involved last for roughly 300 minutes per day. The inverse is true for T3 or T6 where the duration of two-tier meetings is much higher than for any other configuration: these groups usually interact on a one-by-one basis. T2 and T4 again have high two-person meetings while equally high amounts that involve 4 team members or even 5.

The following illustration provides information of the overall duration per session size for proximity detects. This is a much less restrictive count and the session size should increase – depending on the degree to which team members share the same office space.

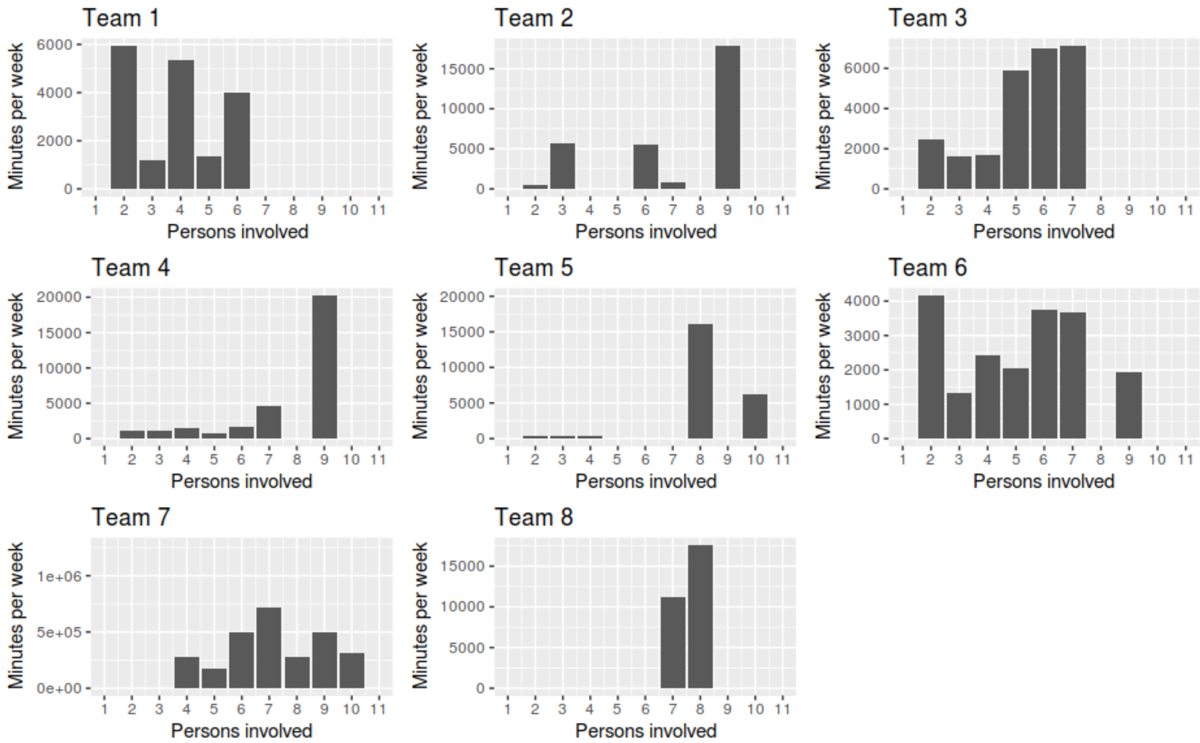


Illustration 13: Minutes per week of proximity detects by number of involved badges

In fact, Illustration 13 can be interpreted as a “Shared Lab” Index. What is the most frequent team size (how many team members are usually present for mutual detects)? Comparing the resulting figures to the actual team size gives a good indicator of shared office/lab space. This index is close to 1 when the size of the most frequent Bluetooth sessions is close to the actual size of the team: since everybody is working within the same space, all badges will continuously detect each other and produce a correspondingly large session size. The less team members are present simultaneously in the same session, the more likely it is that everybody has its own office space and the session size becomes much smaller than the actual team size. Table 13 shows the calculated Shared Lab Index which coincides with the observed situation. It is calculated by dividing the most frequently observed session size by the real team size. Interestingly, T7 was a big team with two office spaces in two locations. Office space was shared, but in relation to the overall team size, the index is smaller since most frequent sessions sizes correspond to the local, i.e. separate offices.

The Shared Lab Index is a crude measure developed within the context of the case studies and based upon the above analysis. Future, more sophisticated versions would be more sensitive to border-line situations. For example, instead of using the maximum session size value the mean or median could be used to account for not-so-clear cases like T6 or T1. In Team 6 for example, sessions involving 2 members accumulate the longest duration. However, this is closely followed by a duration for sessions involving 6 and 7 members. A small difference in duration would here quite drastically swap the interpretation from separated-offices to shared office space.

Team	Shared Lab Index	Observed Lab Space
1	0.25	Separate / Teaching
2	0.90	Shared
3	0.25	Separate / Teaching
4	1.00	Shared
5	0.72	Shared
6	0.20	Separate/Teaching
7	0.68	Shared
8	1.00	Shared

Table 13: Shared Lab Index

Session Duration / Detects and Face-to-face badge-diversity measures

The Histograms of the number of session detects provides further insights into the interaction patterns of teams. Sessions have been calculated by clustering all interactions into a single session if consecutive events are not separated by more than 120 seconds. Overall, the distribution of session duration resembles a power-law distribution, where shorter sessions are much more frequent than longer face-to-face meetings. The longer the session duration, the less frequent. Within this overall pattern, individual differences between teams can be observed.

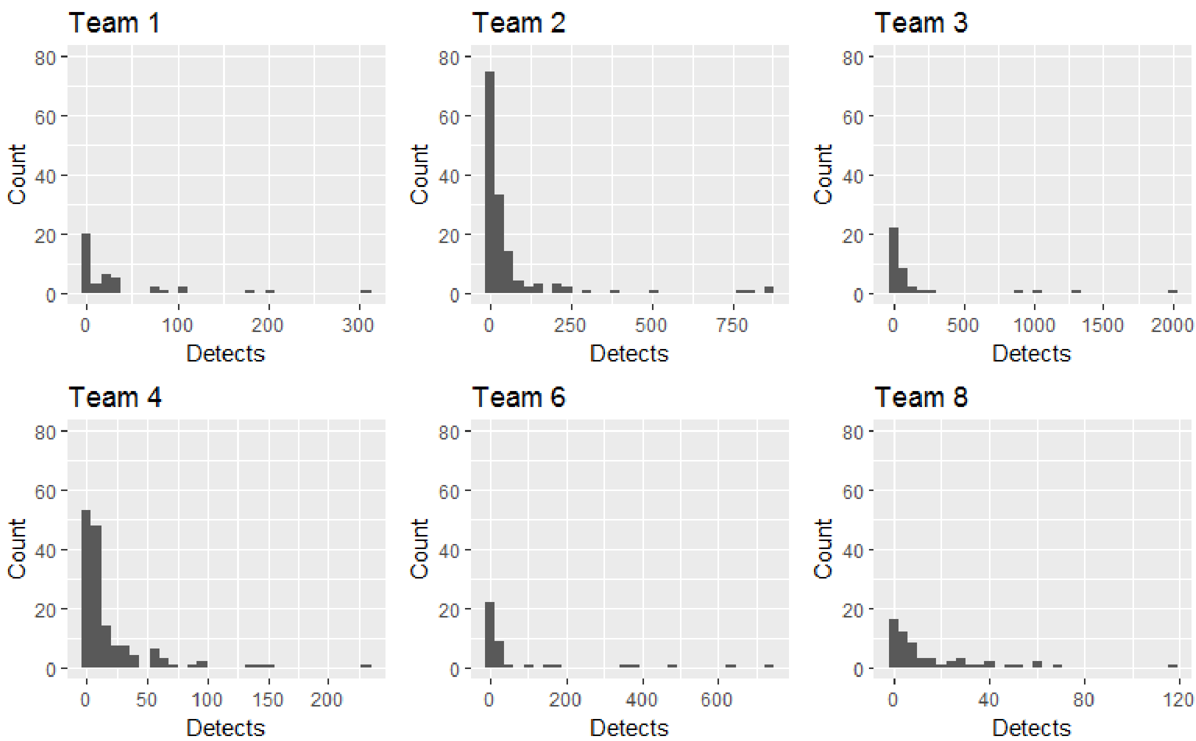


Illustration 14: Histogram of session detects for each T1, T2, T3, T4, T6, T8

First, what distinguishes T5 and T7 from the rest of team is clearly the number of detects (note the different y-scale!). Although T5 and T7 do not stick out in terms of the median, they do in terms of the number of average face-to-face detects (see also Table 12). Second, what also distinguishes the individual teams from each other are the presence of absence of longer sessions. Thus, T3, T5 and T7 do have relatively longer sessions, reaching around 40 minutes of face-to-face interaction. This is in stark contrast with the lower bound of sessions, such as for example for T8 which has a maximum of 4 minutes.

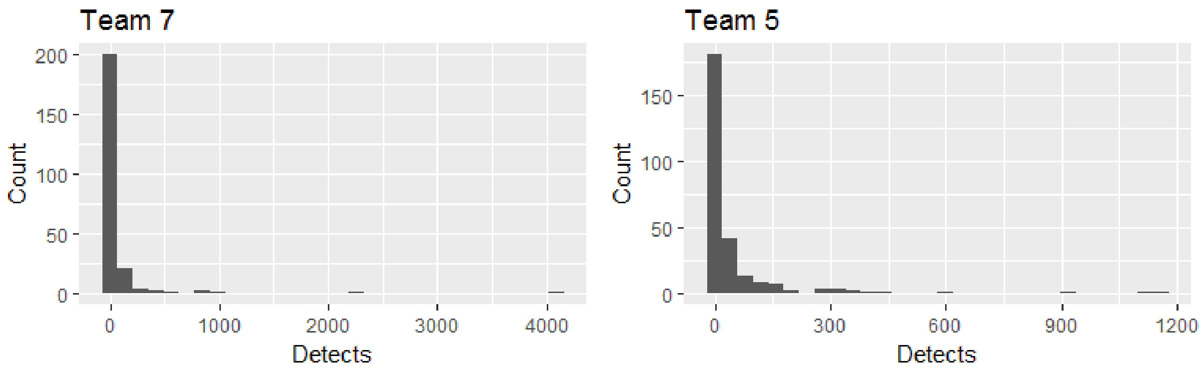


Illustration 15: Histogram of session detects for T5, T7

When comparing the session duration with the number of participants involved it emerges that T3 for example has most of its interactions in small groups (see Illustration 12). Although these may last up to 45 minutes, only a small portion of the group members is involved, as most interactions happens between 2-3 badges. The same holds true for T1, T6, T8 where the dominant session size is between 2-3 participants with shorter duration. This means that members of T1, T6 and T8 interact in small groups and during less time. On the other hand, face-to-face interactions in groups T2, T5 as well as T7 do involve more than 3 persons quite often reaching up to 7 and 8 members. T2 in this sense seems to have the strongest interaction pattern where group size ranges from 2-8 people, having the highest median value with just above 1 minute. There are frequent and “lasting” face-to-face interactions taking place between all members of the group. A similar finding holds for T5 and T7 with the slight difference that although its median values are inferior to T2, its maximum sessions are much higher (being close to 40 minutes). Again, in these teams, communication happens between many team members on an ongoing basis.

	Mean Duration	Median Duration	Max Duration	Mean Entropy	Mean Gini-C
T1	127.93 secs	42.11 secs	25.65 min	0.78	0.54
T2	163.52 secs	65.00 secs	23.88 min	1.40	0.33
T3	278.47 secs	40.50 secs	45.46 min	0.64	0.66
T4	79.78 secs	40.09 secs	10.19 min	1.16	0.40
T5	145.03 secs	42.77 secs	44.21 min	1.43	0.33
T6	138.47 secs	26.50 secs	13.78 min	0.53	0.71

	Mean Duration	Median Duration	Max Duration	Mean Entropy	Mean Gini-C
T7	135.40 secs	27.08 secs	38.75 min	0.72	0.62
T8	49.09 secs	25.13 secs	4.30 min	0.88	0.49

Table 14: Mean face-to-face session duration

The overall relation between persons involved and duration can also be expressed in terms of Shannon Entropy or the Gini-Coefficient of concentration. Shannon entropy[16] is highest for those teams where members communicate with many others in a very balanced fashion, i.e. both variety and balance are high. T2, T5 and T4 have all very high Entropy values, indicating that face-to-face interactions happens to equal amounts among all team members. The lowest values are to be found for T6 and T3 – which have correspondingly high Gini-Simpson concentration measures – who interactions with who is rather unbalanced and/or infrequent.

Network Analysis of Interaction Data

The interaction data collected by sociometric badges can be analyzed as social networks. However, any analysis collapsing the time dimension of the data into a static slice faces the difficulty to construct an over-determined network. Since research teams constitute relatively small groups (8-15 team members on average) that interact sometimes intensively, the static network measures discriminate weakly between the different roles and activity patterns within the team. When everybody interacts with everybody else over the duration of one week, network measures will be very similar.

What therefore gets visible first and foremost through these network statistics are global characteristics of the team, i.e. if they work closely together in the research lab or if their work spaces are rather separated. For example while Illustration 16 shows the interaction pattern of a research group in a teaching based university, Illustration 17 shows proximity and face-to-face interaction patterns for research lab based teams.

The network diagrams mirror here insights gained from the pure frequency counts of teaching vs. research lab based teams, the latter showing a much higher count of face-to-face but also proximity count.

ANNEX II – Selected Graphs / Charts from page 106 on wards provides an overview of these network graphs for each team. The next chapter on “Data Dialog” will draw extensively on these graphs and measures in order to explore possible correlations between the real world settings of the participating teams and their sociometric profiles.

16 The higher the “variety” and the “balance”, the higher the entropy.

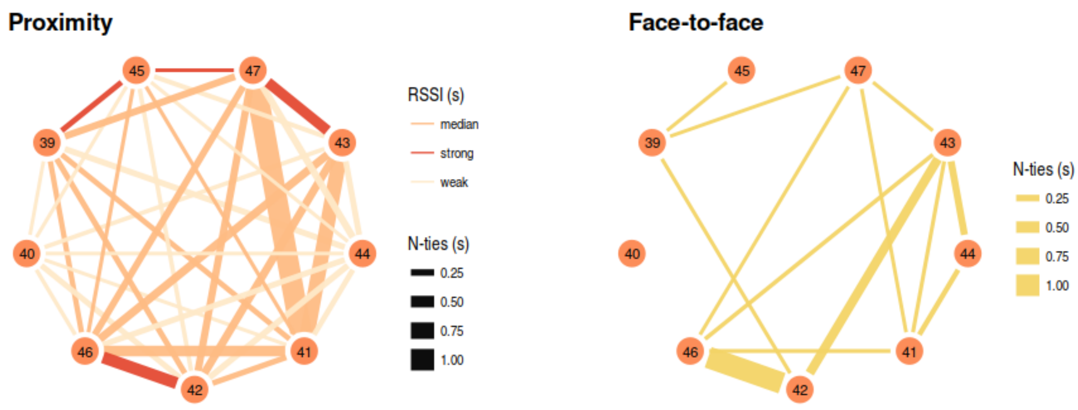


Illustration 16: Teaching based university research team

From a network analytic perspective the above face-to-face graph provides some inroads to calculate centrality measures that allow to characterize differences among nodes (and hence team members). However, this does apply in a very limited fashion to the next network in the below graph where most badges are in contact with each other.

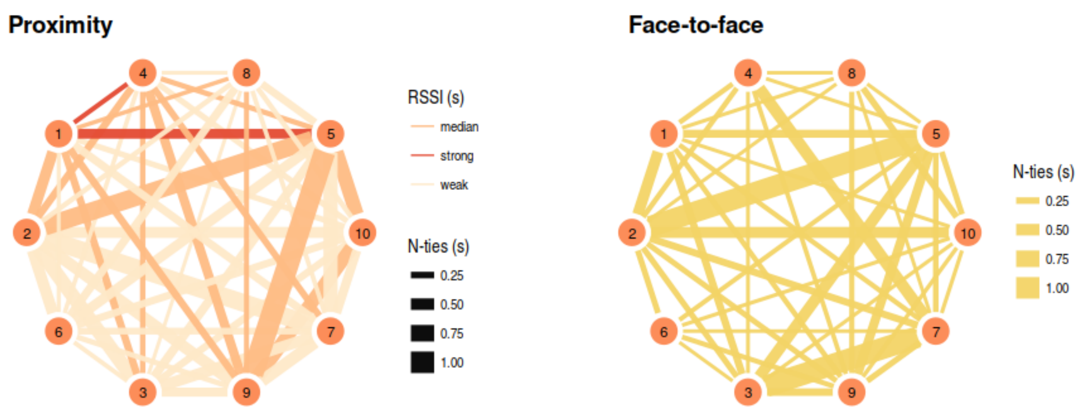


Illustration 17: Research based lab network graphs

A simple regression model has been fitted for “Closeness”, “Eigencentrality”, and “Betweenness” measures (Table 15). As the following table indicates, gender is significant for Closeness centrality and Betweenness centrality ($p < 0.05$). Closeness centrality slightly decreases for women while Betweenness centrality increases.

	Closeness			Eigencentrality			Betweenness		
	B	*CI*	*p*	*B*	*CI*	*p*	*B*	*CI*	*p*
(Intercept)	0.04	0.01–0.07	**.005**	0.64	0.30–0.97	**<.001**	1.47	-1.25–4.19	.285
Gender (Man)	-0.01	-0.02–-0.00	**.021**	0.00	-0.12–0.13	.969	1.08	0.05–2.11	**.039**
Age	0.00	-0.00–0.00	.998	0.00	-0.01–0.01	.541	-0.00	-0.07–0.06	.990
Tenure	0.00	-0.00–0.00	.232	-0.00	-0.00–0.00	.710	0.00	-0.02–0.02	.807
Role (Junior)	0.01	-0.00–0.02	.131	0.09	-0.07–0.25	.252	-1.17	-2.47–0.12	.075
Observations	76			76			76		
R^2 / adj. R^2	.118 / .068			.024 / -.031			.138 / .089		

Tabla 15: Regression results for face-to-face network centrality measures

Previous investigations indicate that a relationship between personality traits and network centrality measures (Alshamsi, Pianesi, Lepri, Pentland, & Rahwan, 2016; Lepri et al., 2012; Lepri, Staiano, Shmueli, Pianesi, & Pentland, 2016). Participants also answered the Big Five Personality questionnaire items (Rammstedt & John, 2007). However, as the following regression model indicates, none of the personality traits yields significant results regarding the three centrality measures.

	Closeness			Eigencentrality			Betweenness		
	B	*CI*	*p*	*B*	*CI*	*p*	*B*	*CI*	*p*
(Intercept)	0.08	0.03–0.13	**.001**	0.54	-0.00–1.09	.052	2.81	-1.80–7.43	.228
Extraversion	-0.00	-0.00–0.00	.385	-0.00	-0.01–0.01	.663	0.02	-0.05–0.08	.637
Agreeableness	-0.00	-0.01–0.01	.768	0.05	-0.04–0.13	.271	0.26	-0.42–0.95	.446
Conscientious.	-0.01	-0.02–0.00	.184	-0.04	-0.14–0.06	.458	-0.05	-0.89–0.79	.900
Openness	-0.00	-0.01–0.01	.968	0.04	-0.05–0.12	.431	-0.57	-1.32–0.18	.136
Emotional Stab.	-0.00	-0.01–0.01	.699	0.02	-0.07–0.11	.663	-0.01	-0.80–0.79	.983
Observations	76			76			76		
R^2 / adj. R^2	.056 / -.012			.032 / -.038			.075 / .009		

Interactions by Gender

Across the 8 teams there were 35 women and 45 men participating. As can be seen from the following table, the average detects for face-to-face interactions is slightly higher for women than for men considering a time span of 5 days. The inverse holds for proximity detects, where men have a slightly higher count.

Gender	Interaction	Total	Mean 5 Days	Daily Mean
Woman	Proximity	514071	14688	2938
Woman	Face-to-face	51885	1526	305
Man	Proximity	736799	17971	3594
Man	Face-to-face	52541	1347	269

Table 16: Total and mean detects by gender across all teams

A closer look regarding gender differences in interaction between groups is presented in the section on Relation Event Modeling – which provides a framework for testing statistically significant differences regarding time-based data. Depending on the team, there are differences of the mean detects for women and men. However, there is no consistent patterns across teams, that is, that men consistently would be more "visible" in interactions women or vice versa.

The following graph illustrates that there is no easily identifiable gender pattern. It shows the total pairwise detects for proximity and face-to-face interactions. There is also no clearly identifiable homophily pattern where women would interact more frequently among women only or men only among men, although the proximity data could suggest otherwise. It seems that men "hang around" other men more frequently due to the higher detects in Illustration 18. However, overall the proportion between men and women is also slightly skewed, having 35 women across all teams and 45 men. Thus it can be expected that there are more man-man interactions than women-women. This also suggests that there is an imbalance in the face-to-face interaction where we would also expect to have more men-to-men interactions than women only. To which degree these differences are significant has to be explored however in

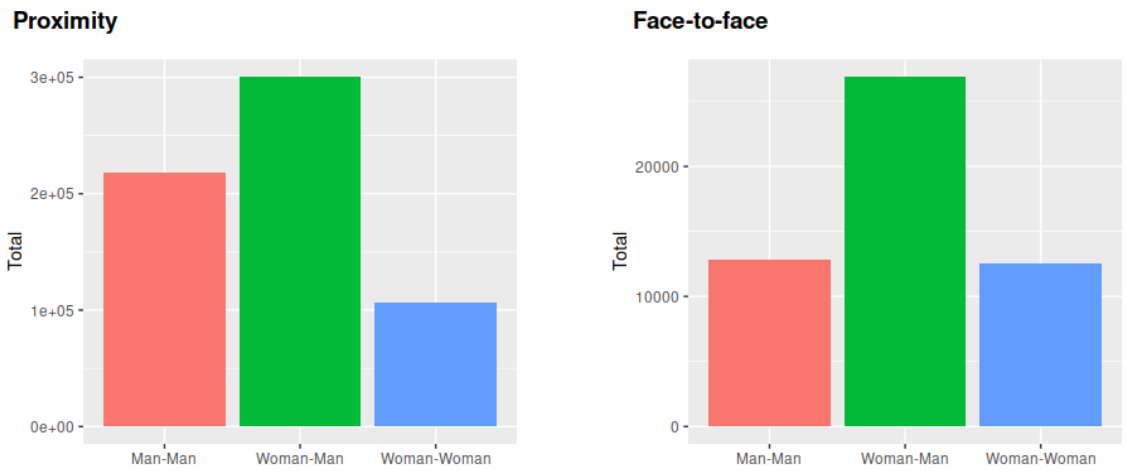


Illustration 18: Pairwise proximity- and face-to-face detects by gender across all teams

the context of the Relational Event Modeling, on page 78 on wards. As a basis, the following tables provide already an overview of the mean face-to-face interaction for each team according to gender. The right hand graph gives an impression of the relative participation of women and men during interactions. Thus, in Team 1 for example, women are over-represented in interactions whereas in Team 2 they are slightly under-represented. What more, the right chart for Team 2 suggests that most interactions happen among women whereas in Team 1, most interactions happen in mixed gender dyads.

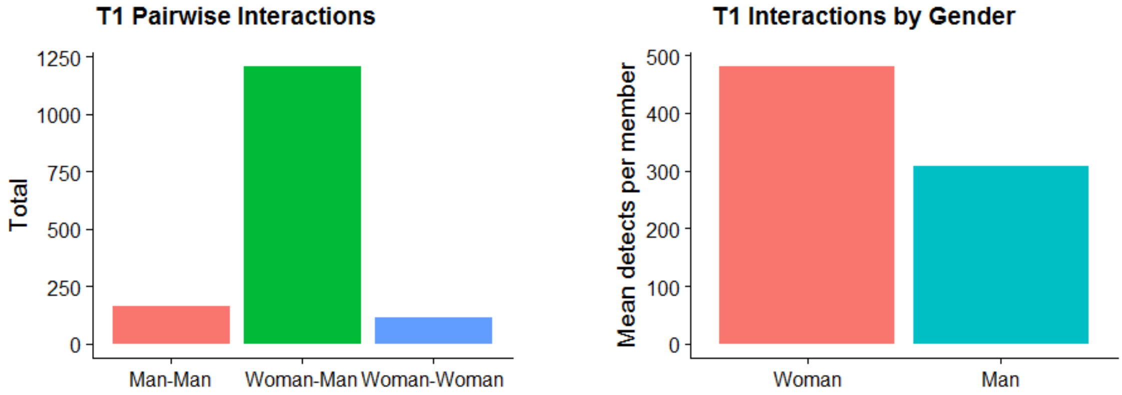


Illustration 19: Team 1 face-to-face interactions by gender

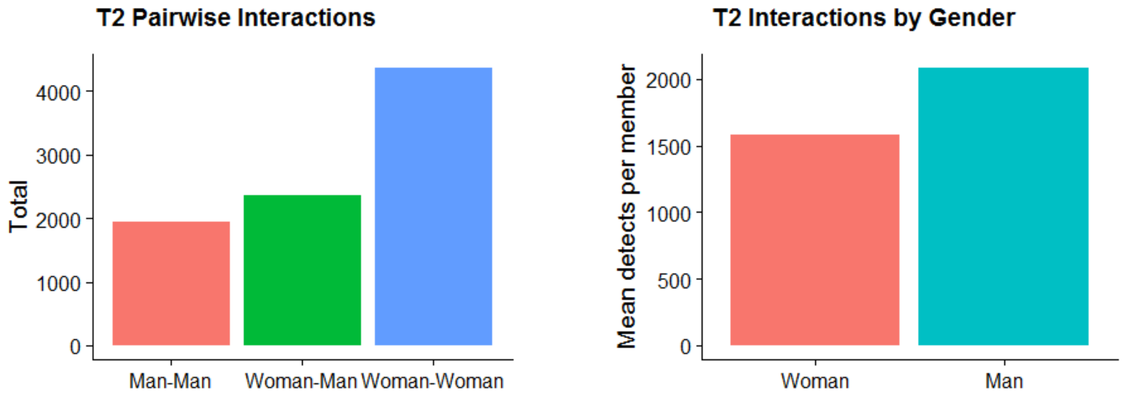


Illustration 20: Team 2 face-to-face interactions by gender

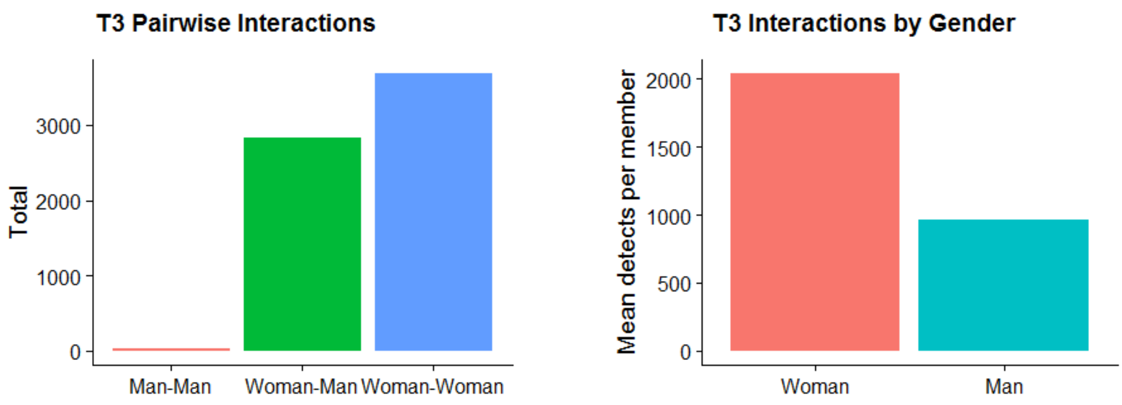


Illustration 21: Team 3 face-to-face interactions by gender

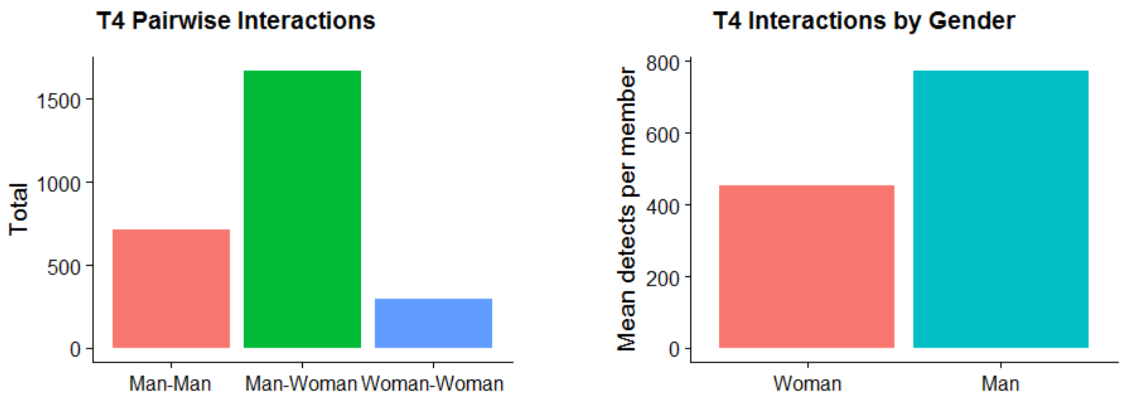


Illustration 22: Team 4 face-to-face interactions by gender

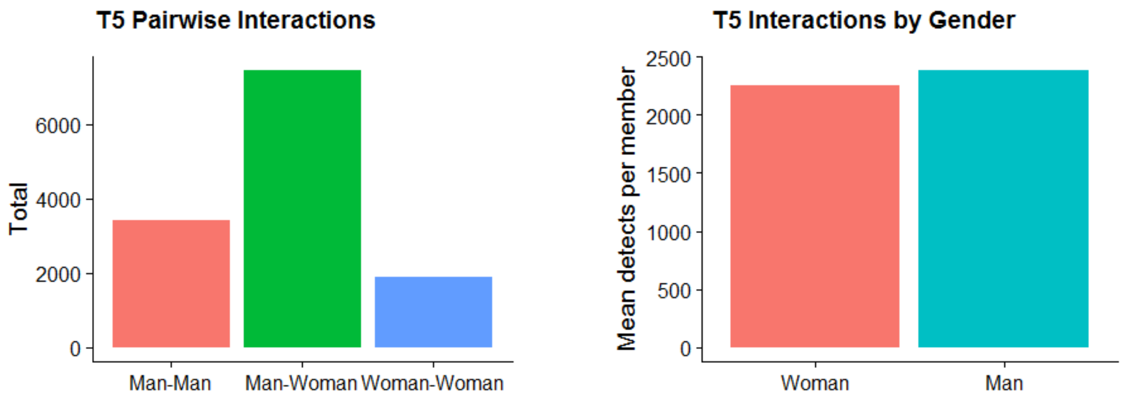


Illustration 23: Team 5 face-to-face interactions by gender

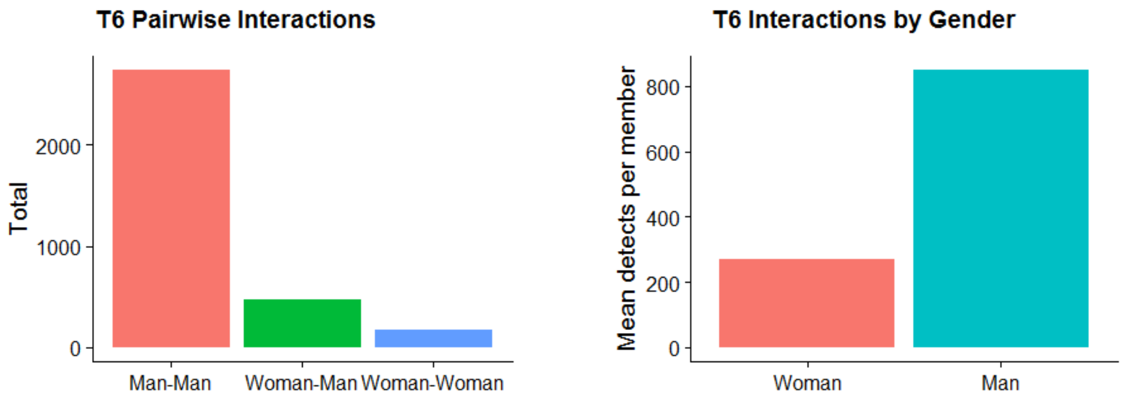


Illustration 24: Team 6 face-to-face interactions by gender

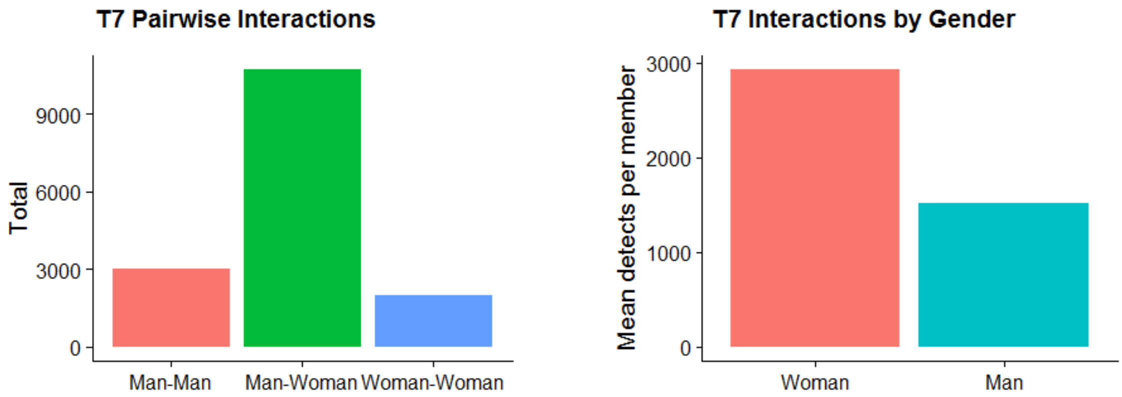


Illustration 25: Team 7 face-to-face interactions by gender

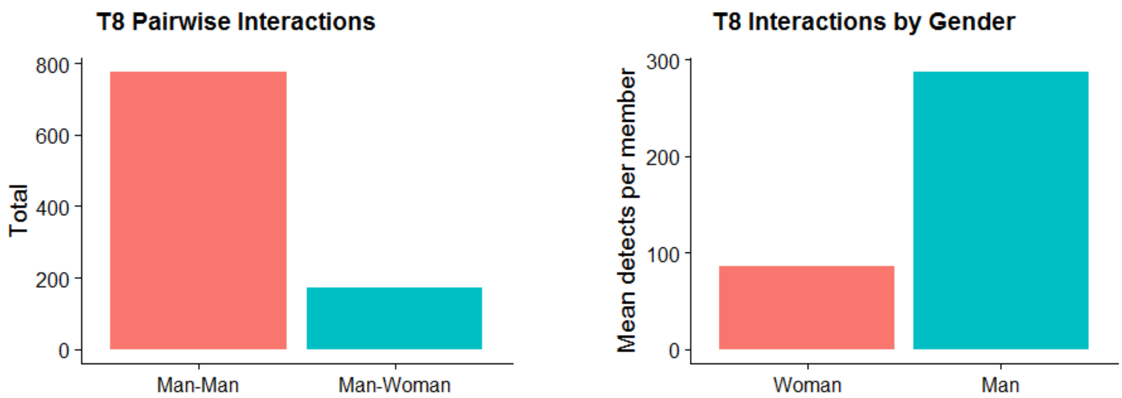


Illustration 26: Team 8 face-to-face interactions by gender.

Interactions by Team Role

The following graph demonstrates the mean number per team role for both proximity- as well as face-to-face detects for the six research/university based teams. The two remaining case studies from the private company are not included since they do not have PhD or MA positions.

Comparing face-to-face and proximity detects, the differential role of both Research/Lab Assistants as well as Administrative Assistants is noteworthy. Research/Lab Assistants are in proximity to the research staff as can be seen in Illustration 27B, having a mean count that is close to those of Postdocs, PhDs and MA students. However, when examining the face-to-face interaction profile, their frequency of detects is much lower: who is interacting in the lab are indeed PhD and Postdoc researchers as well as MA students while assistant roles are much less prominent.

The mean detects suggest that most face-to-face interaction in the lab is carried out by PhD and Postdocs. MA students also seem to play an important role. Senior researchers and team leaders are less involved in the actual experimental work of research teams.

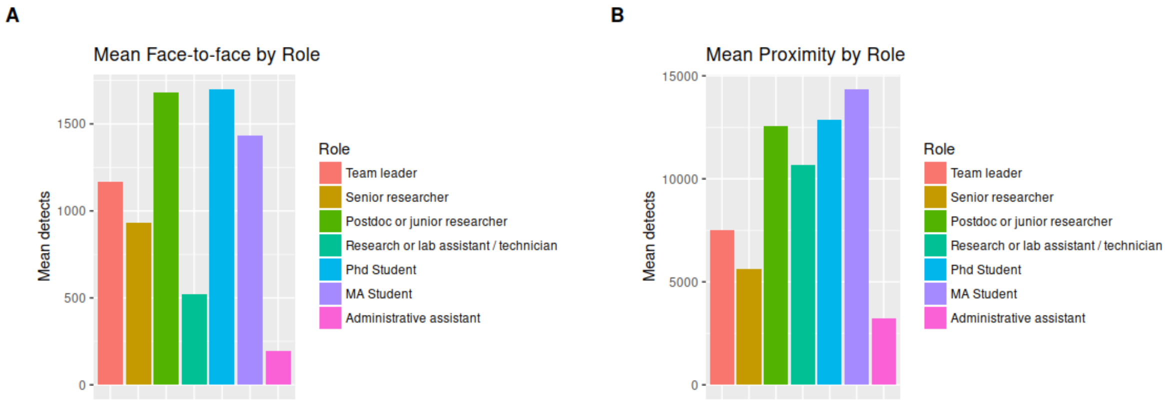


Illustration 27: Mean detects by role

The importance of the PhD, Postdocs and MA students for the actual work carried out is confirmed by the analysis of interaction dyads by role: most interactions do happen among PhD students, PhD-MA students, PhD and Postdoc/Junior researchers and less for example between Assistants and Team Leaders, or MA Students and Seniors.

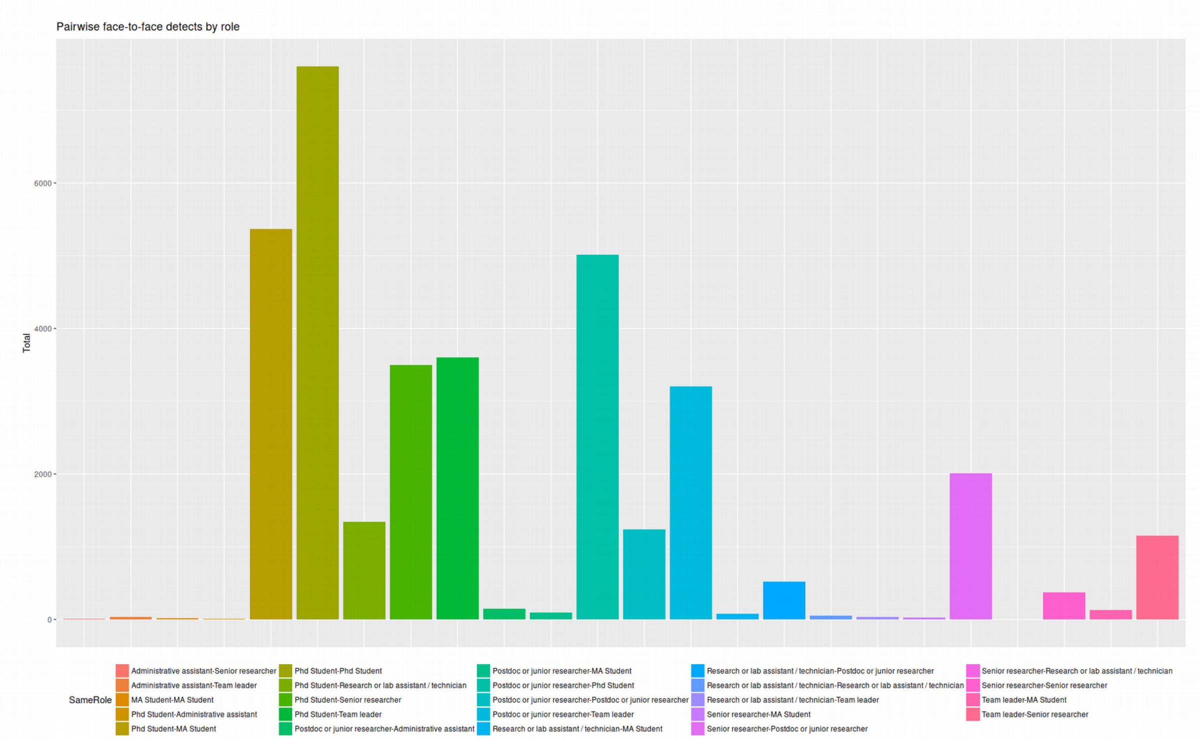


Illustration 28: Pairwise face-to-face detects by roles

If there are statistically significant differences regarding interaction patterns between these roles will be furthermore addressed in the section on Relational Event Modeling on page 78.

Regression Model for Interaction Data using Age, Gender, Role

Two simple regression models have been fitted using the interaction data across all teams. The first model predicts the total amount of proximity detects using the covariates of age, gender, role, gender stereotype and team tenure. Age is significant at the $p < .001$ level. A binary variable "team role" (recodifying Team Leader and Senior Researcher as "Senior" and Postdoc, PhDs, MAs as "Junior") is significant at $p < .05$. Since team roles do correlate with age, this is not surprising.

The second model uses face-to-face detects but does not produce any significant results. In both cases, gender is not predictive of the total amount of interaction detects, neither proximity nor face-to-face counts.

	Total Face-to-face Detects			Total Proximity Detects		
	B	*CI*	*p*	*B*	*CI*	*p*
(Intercept)	2578.56	457.12–4700.01	**.018**	46778.84	27892.19–65665.49	**<.001**
Age	-39.17	-79.29–0.95	.056	-668.47	-1026.28–-310.66	**<.001**
Gender (Man)	-122.25	-754.76–510.26	.701	4272.93	-1266.75–9812.61	.128
Role (Junior)	-658.26	-1459.51–142.99	.106	-7419.96	-14519.37–-320.54	**.041**
Gender Stereotype	379.31	-0.15–758.77	.050	-1008.51	-4345.56–2328.55	.549
Tenure	-10.09	-20.90–0.73	.067	-66.72	-163.26–29.82	.172
Observations	73			76		
R^2 / adj. R^2	.203 / .144			.259 / .206		

Table 17: Regression on total proximity and face-to-face detects using age, gender, and role as covariates

Audio

A comparative view of the mean speaking, listening, overlap and silent duration per team is given in Illustration 29. The figures reported refer to the average hours per team member over a five working day period. The average speaking duration does not show large differences; most team members speak between 2 to 3.5 hours during the entire week. Differences are more pronounced for the listening duration with members of T7, T8 and T5 being the most active listeners and members of T1, T3 and T6 having the lowest listening duration. This raises the question to which degree the listening profile does capture actual speaking within the team, since T1, T3 and T6 are university/teaching based teams where team members do not share office space or research labs. The problem is clearly suggested by T8 which has a very high listening duration, although it has one of the lowest face-to-face interaction detects. This means that the badges probably pick up on people talking even though team members are not interacting with each other directly.[17]

In order to explain how these four dimensions of the speech profile is calculated, let's use a simple example involving two persons A and B, each wearing a badge. Speaking time for person A indicates the duration this person A was speaking. If person B was present and silent during this time, its badge would record the same amount of time as “Listening”. If both persons A and B talk at the same time, this would be registered as “Overlapping”. If none of the persons is speaking, this would be registered as “Silence”. The average hours of “Silence” are therefore quite high because it indicates the duration that a badge was turned on – while nobody was speaking.

17 When interpreting the audio profiles in the context of other collected data sources, the office space suggests that badges “listen” to people talking in the environment without necessarily talking to each other.

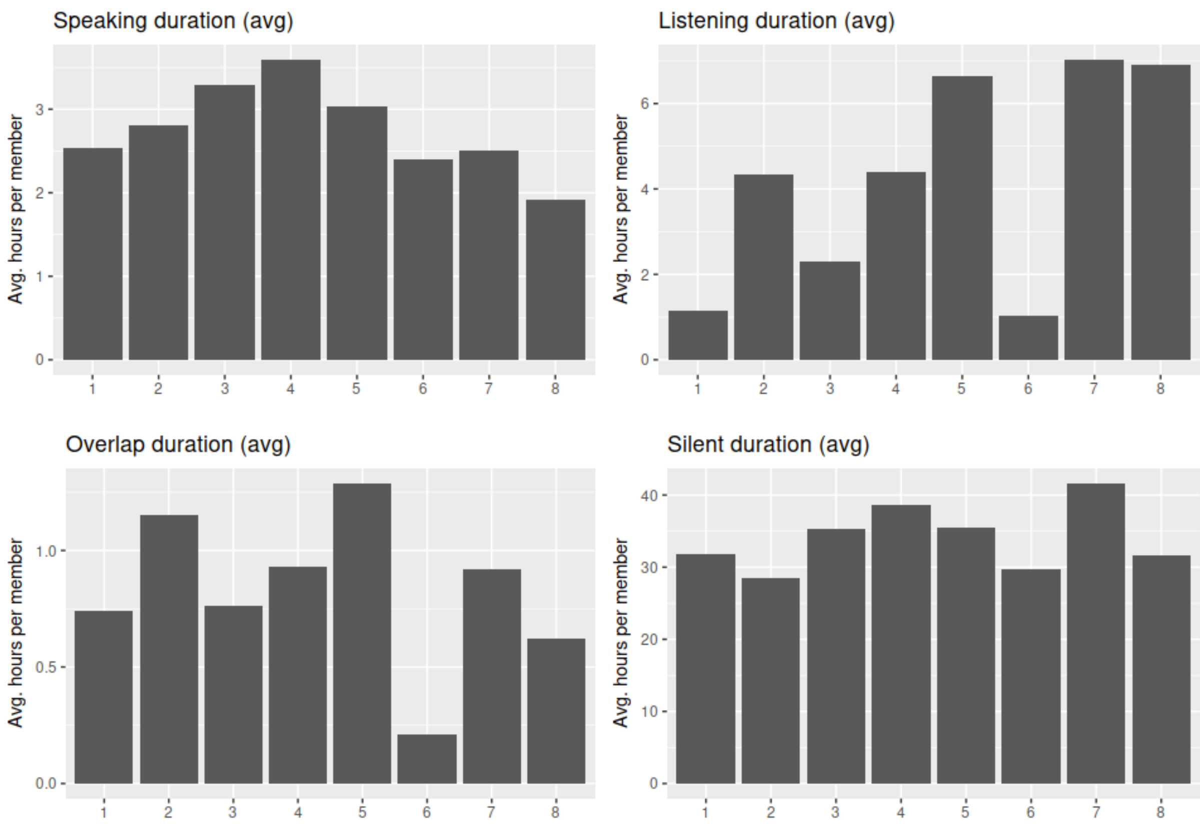


Illustration 29: Average speaking, listening, overlap and silent hours per team member over 5 days

Interestingly, the average speaking duration is highest for T4 which has at the same time a relatively low face-to-face interaction count. The same applies for T7 and T8, the latter having a very low face-to-face interaction count compared to T7. A possible interpretation is that team members were interacting substantially with non-badge wearers. To think in extremes: the maximum amount of speaking duration per team would consist of all team members speaking all the time. The only situation where this seems somehow feasible is, if team members are separated – otherwise they would be talking over each other. This interpretation fits the fact that T4 has indeed the third lowest face-to-face count, after T1 and T8. Thus, people don't interact a lot within the team but nevertheless have the highest speaking duration, suggesting that they talk on the phone or to other non-badge wearers, or that the badges pick up on ambient noise.

Team	Total Speaking	Total Listen	Mean Speaking	Mean Listen	SD Speaking	SD Listen
1	17.71	8.03	2.53	1.15	1.25	0.71
2	28.04	43.37	2.80	4.34	1.35	1.96
3	26.32	18.37	3.29	2.30	2.02	2.03
4	32.41	39.65	3.60	4.41	1.75	2.72

Team	Total Speaking	Total Listen	Mean Speaking	Mean Listen	SD Speaking	SD Listen
5	33.30	72.90	3.03	6.63	1.50	2.78
6	21.59	9.22	2.40	1.02	1.18	0.85
7	34.95	98.53	2.50	7.04	1.35	3.89
8	15.36	55.30	1.92	6.91	0.63	1.60

Table 18: Total and mean speaking and listening duration

Speech Profile by Gender

Examining speaking duration by gender does not display strong differences, neither for "speaking" nor for "listening" except for T1 and T2 where women do seem to speak more than men and T3 where women have higher listening duration than men.

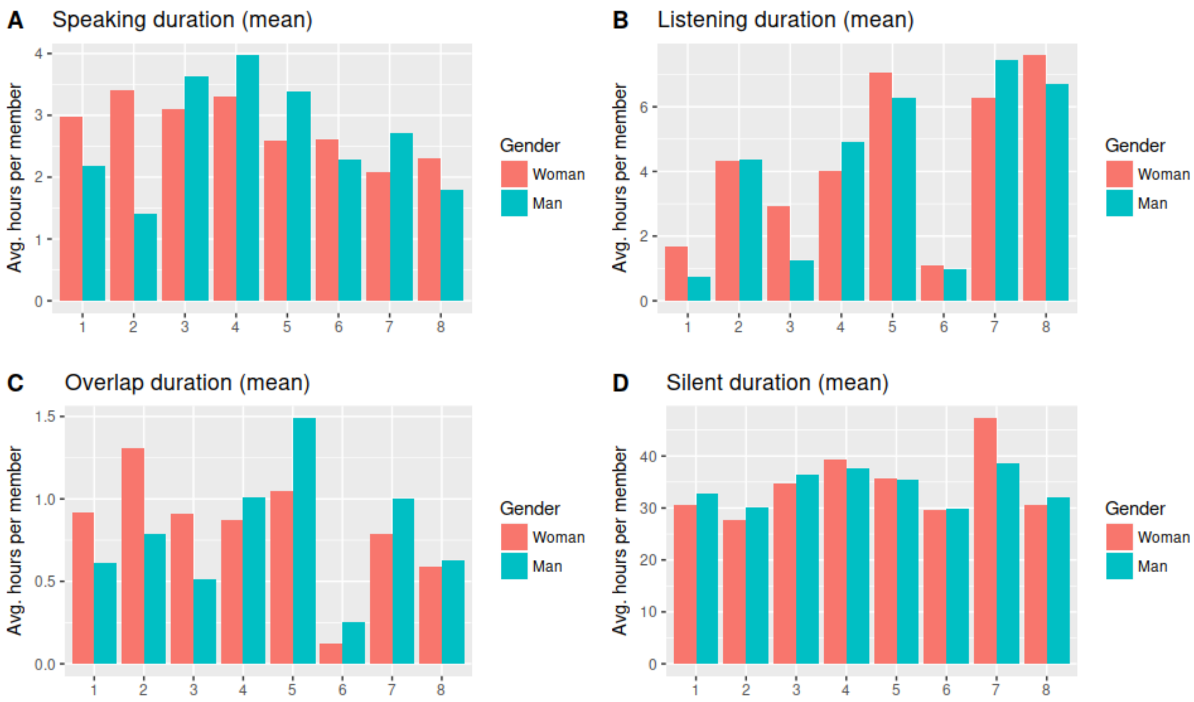


Illustration 30: Speech profile mean duration by gender

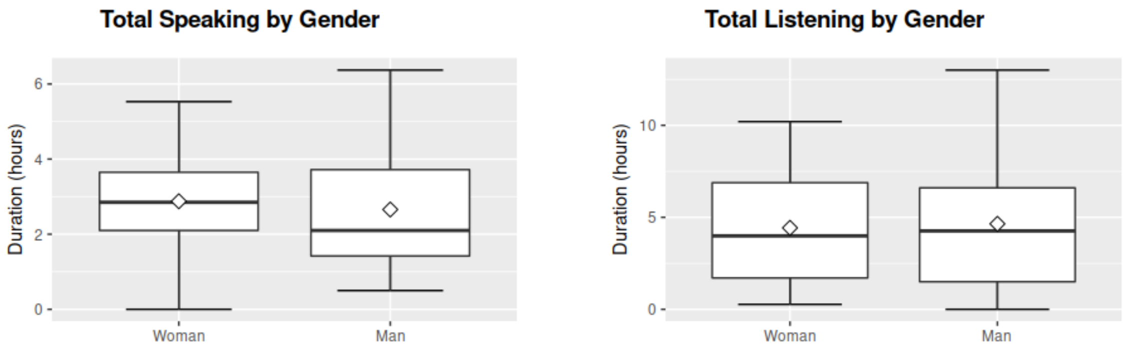


Illustration 31: Percentile and mean values for speaking/listening duration by gender

Speech Profile by Role

An interesting question concerns the speaking duration by junior vs. senior roles in the team[18]. The mean speaking duration for junior and senior roles is quite similar. Marked difference appear when looking at the average listening duration: senior positions seem to listen considerably less than the junior researchers. Not only is the distribution of listening time for senior positions less dispersed, there is also a clear difference for the median and mean values. However, significant differences between listening duration for senior/junior positions are achieved only if the private company teams are excluded, because the junior and seniority profiles are quite different (with senior researchers and PhD positions being absent in the private company).

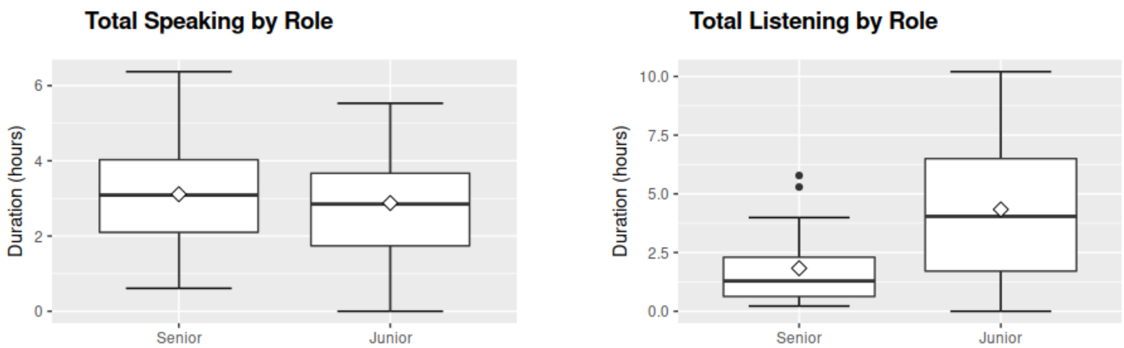


Illustration 32: Total Speaking and Listening Duration by Role for T1-T6 (public research)

Senior positions talk considerably more in T3 while the opposite is true for T4. Very clear is the lower listening duration for seniors across all teams (Illustration 33B), except T2 which is slightly more balanced.

18 Senior roles include “Team leader”, “Senior researcher”. Junior roles include “Postdoc/Junior researcher”, “PhD student”, “MA student” and assistant positions.

	listening		speaking	
	B	*CI*	*B*	*CI*
(Intercept)	9.39***	6.84–11.95	2.55***	1.31–3.78
Age	-0.13**	-0.21–-0.05	0.01	-0.03–0.05
Tenure	-0.01	-0.03–0.01	-0.00	-0.01–0.01
Observations	76		76	
R^2 / adj. R^2	.205 / .183		.002 / -.025	
Notes				** p<.05** p<.01*** p<.001*

Table 19: Regression for speaking / listening duration

If team members of the private companies are included in the sample, then it's not so much Team Role anymore that produces significant differences but Age. As mentioned, the team roles in the private company do not match the team roles in the research centers or university. However, team roles do correlate with age. Older team members have more senior team roles. As the regression model shows, the older the participants the less do they listen. On the one hand this is plausible considering team leaders for example which “should” at least talk as much as they listen to others; at least it would be strange to have a team leader that only listens but does not talk. On the other hand, more senior positions do participate less in daily lab work and hence will pick up less the speaking of other team members. PhDs and Junior team members interact more frequently which increments the chances that their listening duration increments, simply because they are close to other potential speakers.

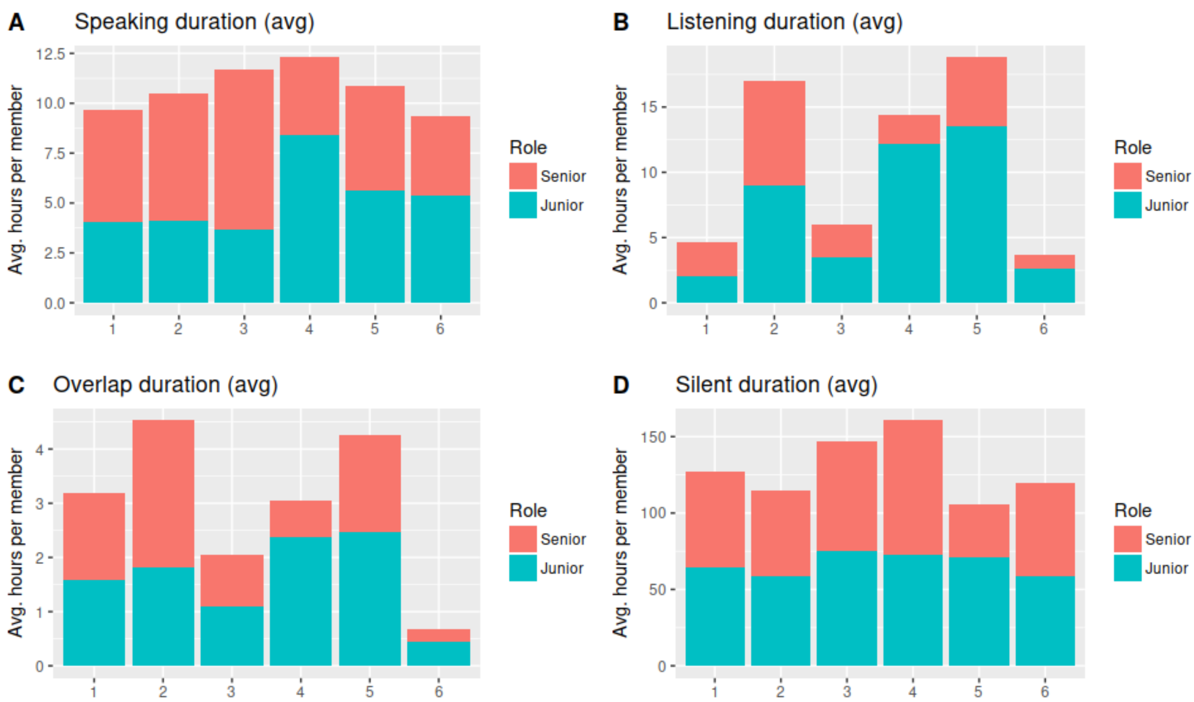


Illustration 33: Average speaking, listening, overlap and silent duration by role

Body Movement & Mirroring

As described, sociometric badges are equipped with an accelerometer, i.e. sensors to monitor body activity. Since each measurement is timestamped, "similarity" scores are calculated that indicate how "well" body movements between participants are synchronized and mirror each other. At the same time, the accelerometer readings of all badges also allows some insights regarding the movement pattern of the team members, especially how well people act in synchronicity indicating "centralized" or well orchestrated events.

Body Activity across Badges

The accelerometer readings of sociometric badges give insights into the overall activity pattern of the team. Activities that are "compulsory" for all team members such as team meetings but also social activities such as shared lunch display characteristic peaks of activity and immobility which coincide in time.

The following illustration shows the body activity measurements for 8 badges over an entire day. Activity measurements > 0.2 usually indicate walking or non-stationary activities. Apart from the irregular peaks, what clearly sticks out is the overlapping activity pattern at around 14:00 o'clock, where all team members seem to move uniformly during the same time period. For about 30 minutes, everybody was moving in a very synchronized way, reproducing even small irregular "dents" in the activity pattern across all badges.

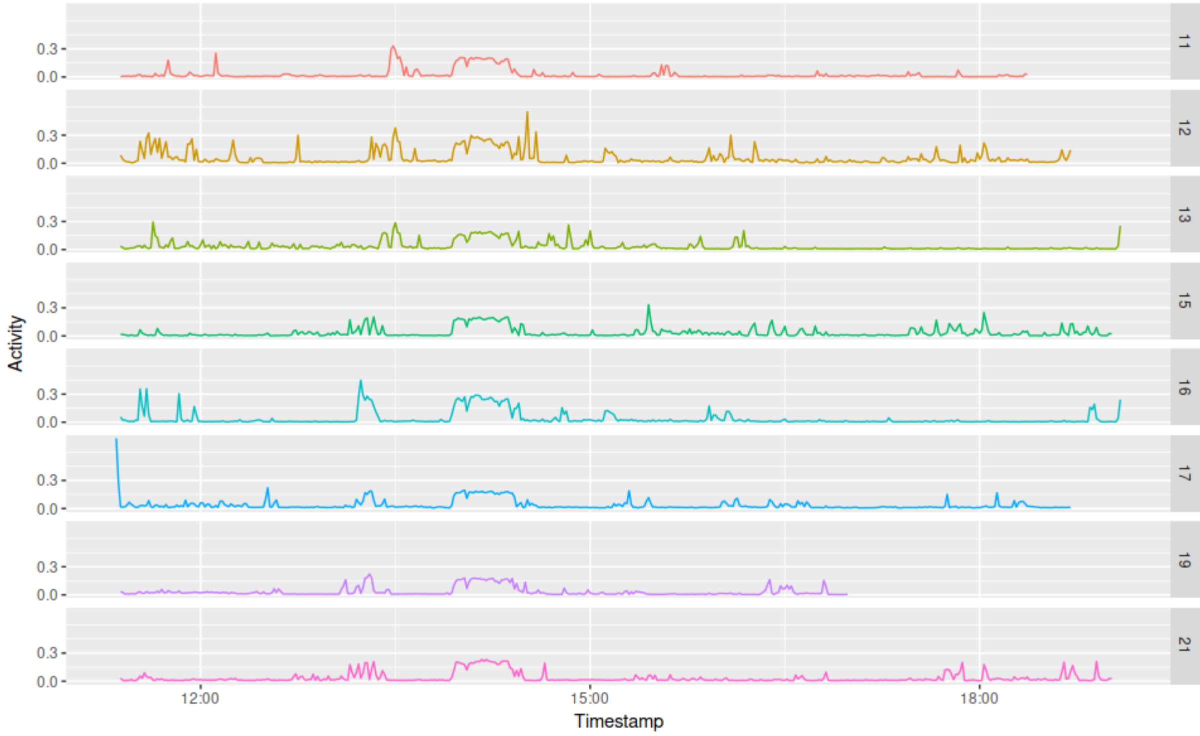


Illustration 34: Body activity pattern for one day and 8 badges

This synchronicity in body movement across all badges gives an indication about the number of “formalized” encounters within the research group. Opportunities where all team members act in concordance usually require a formal arrangement that tells us something about the level of centralized steering within the group.

A simple measure that indicates the degree to which the activity curves of badges coincide is the Pearsons' correlation coefficient. When calculating the correlation coefficient over the entire body activity curve, however, the peaks will likely be overshadowed by the larger periods of non-activity or lack of coincidence. After all, coincidence between many badges is rather unlikely. Thus, for our “synchronicity” measure, we slide a time window over the activity values and calculate the correlation for each time window separately. Illustration 35 shows the mean Pearson correlation coefficient for all badges of the previous graph (above). Clearly visible is the peak between cluster 10 and 15, where a Pearsons' correlation coefficient close to 1 indicates almost perfect matching of the activity patterns across badges for the given time cluster.

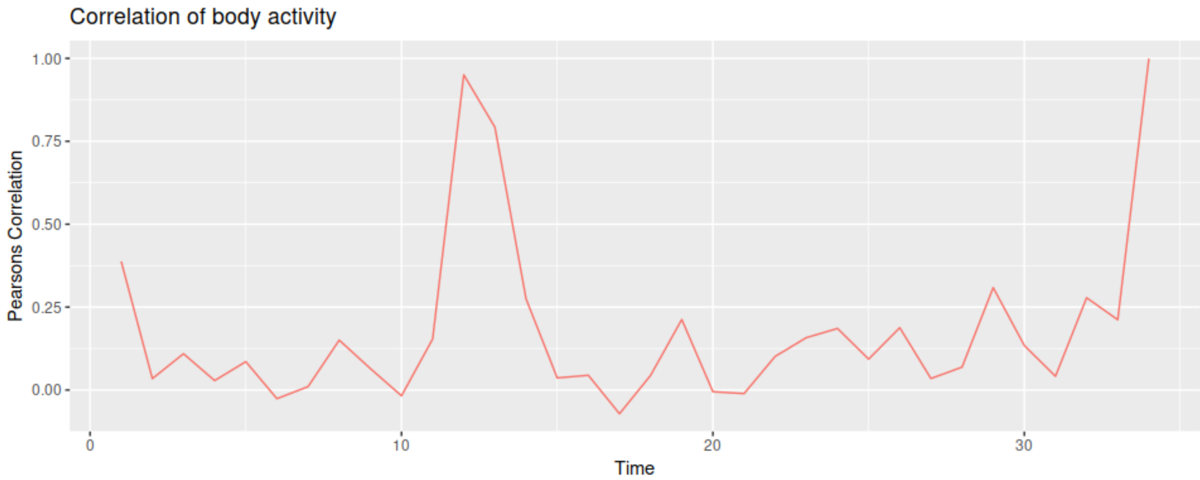


Illustration 35: Mean Pearson correlation coefficient over all badges combinations using a 800 second time window

Note, that the mean of the correlation coefficient is calculated for all combinations of available badges, i.e. it will be higher the more the activity measurements coincide between all members, not just two or three.

The correlation coefficient can be calculated for each day individually, producing a characteristic "synchronicity" curve for each day. Overall, by setting a certain threshold we can count the number of peaks, i.e. the number of highly centralized events for each team which gives an indication of the level of centralized events. Few "peaks" implies that the coordination of activities among team members is low; the team is rather individualized and there is little "formalized" cohesion. On the other hand, if peaks are frequent, then there is a strong formalized cohesion which brings team members together in a central place at specified times.

Our proposed correlation coefficient is calculated as soon as three badges are active at the same time. The current version of the synchronicity index does not take into account the size of the team, even though it is evidently harder to achieve high values for large teams than for smaller ones. This should be considered in a future version.

	Team 1	**Team 2**	**Team 3**	**Team 4**	**Team 5**	**Team 6**	**Team 7[19]**	**Team 8**
Nr. Peaks	4	3	2	1	7	1	5/2	2

Tabla 20: Number of highly correlated body activity patterns by teams

Team 5 has the highest synchronicity of body activity patterns vs. T4 and T6 which have the lowest. What can be seen from the correlation curve for T5 in Illustration 36 (top), is not only the high synchronicity values at certain moments during the day, but also how these coincide at exactly the same moment across two different days (green lines), indicated by the peaks

19 Body activity data spans two weeks. Scores have been calculated for each week independently.

between 11 and 12 on the x-axis.

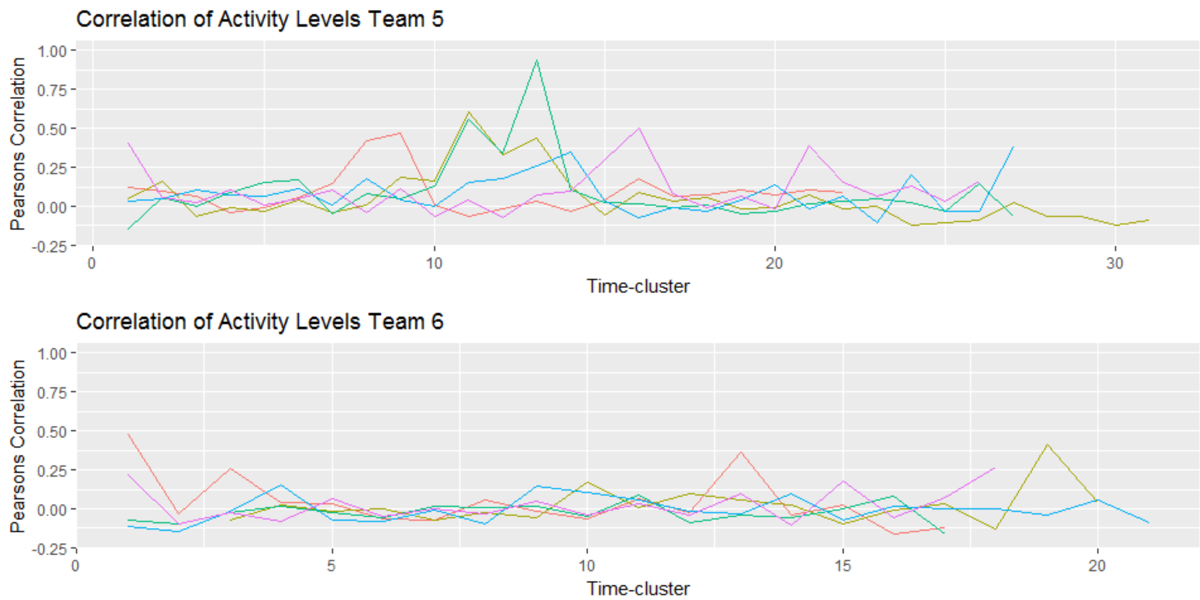


Illustration 36: High vs. low synchronicity in body activity patterns T5 (7 peaks) vs. T6 (1 peak). Different lines indicate different days.

Body Mirroring

The nonverbal mirroring data is based upon the face-to-face interaction only. Although mirroring values between all badges at all times could be calculated, it only makes sense to calculate these values when actual interactions take place. All face-to-face sessions with a duration longer than 60 seconds have been exported and provide the basis of the following analysis.

Mirroring values range from 0 to 1, where 0 means complete lack of mirroring data and 1 a perfect match. Mirroring values have been calculated on a per second basis for the whole duration of the interaction. However, in order to flesh out important difference between badge pairs, a threshold value isolates the relatively rare, higher mirroring values from the vast amount of low values. In the following illustration, only mirroring values > 0.5 are retained. Clearly visible is the high count for T5 in this case. However, when normalizing the higher body mirroring values with respect to the overall count of mirroring values the picture changes considerably: now T3 has the highest values and T5 resides with the normal range compared to other teams. This implies that the high count for T5 is rather a product of the frequent interactions than necessarily a stronger synchronicity between body activity.

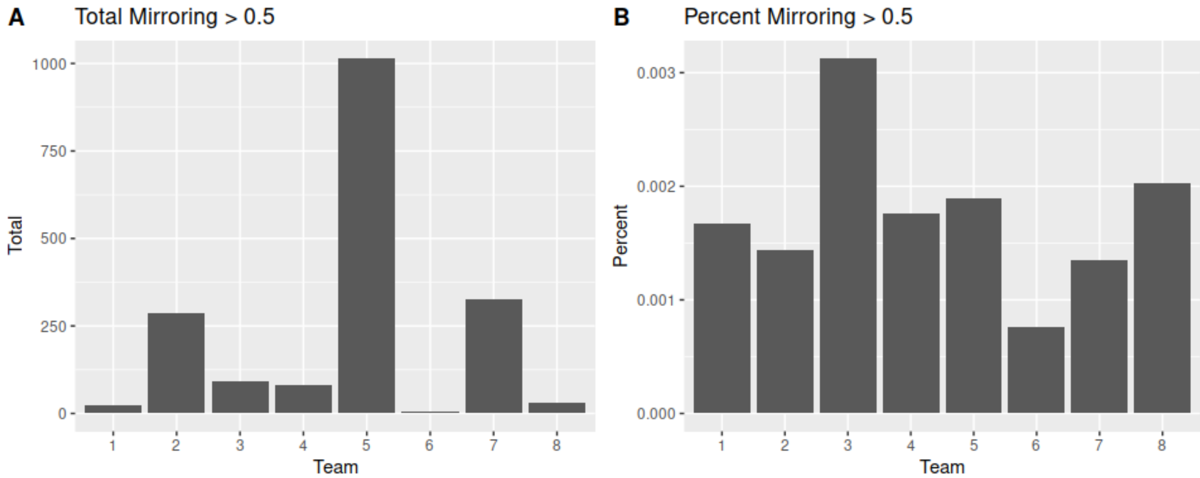


Illustration 37: Body mirroring values > 0.5 across all teams

Body Mirroring by Gender

Examining the body mirroring values by gender shows that there is some difference on the dyadic level, with women-men dyads having higher mirroring values. However, the differences are highly dependent upon the chosen threshold value and hence the number of observations available above the given threshold.

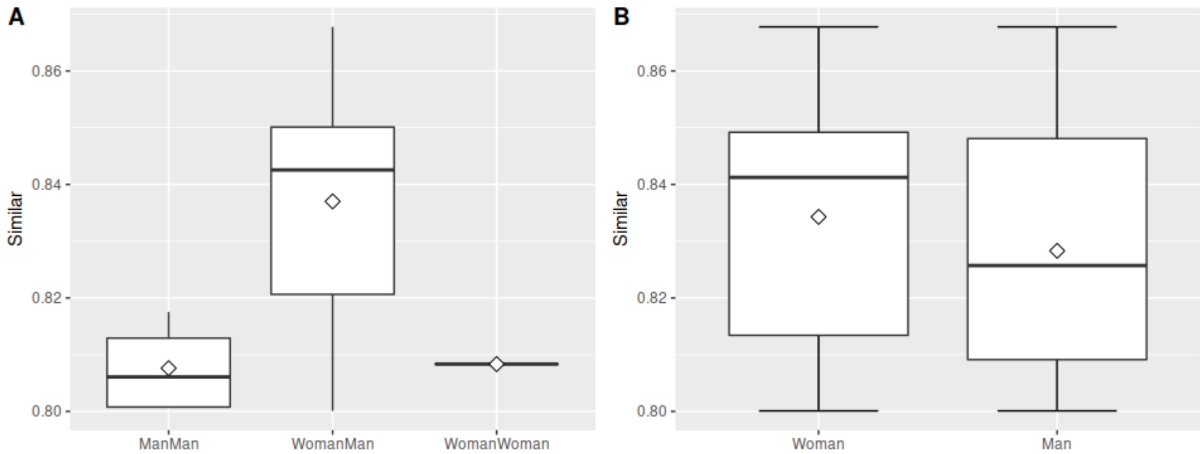


Illustration 38: Body mirroring > 0.8 by gender

Reducing the threshold of body mirroring values to 0.5 levels out the differences in values as can be seen from the following illustration:

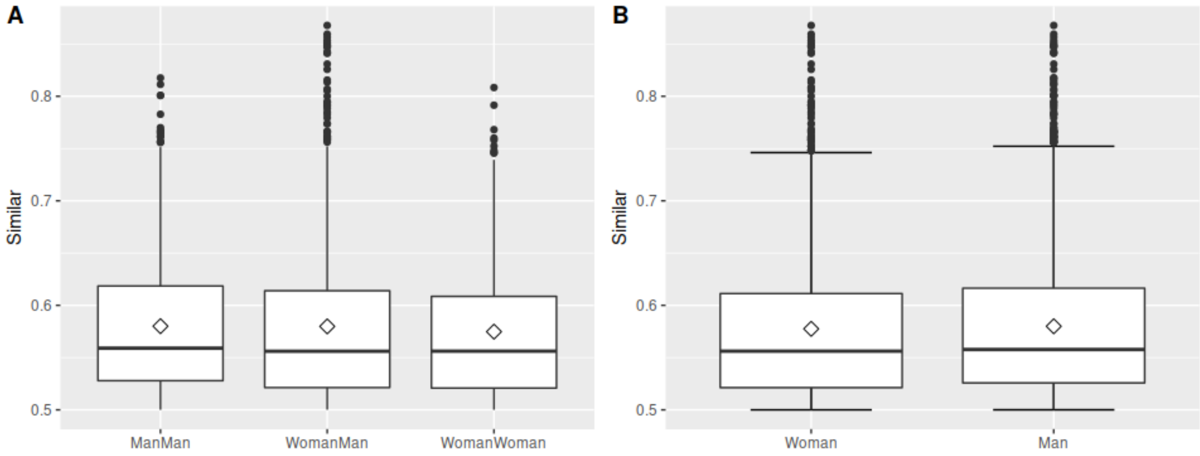


Illustration 39: Body mirroring > .5 by gender

There are no gender differences regarding body mirroring values.

Contextualizing Sociometric Data – a Data Dialog

The GEDII case studies offer a unique opportunity to explore sociometric data in a more qualitative manner. As the following paragraphs will show, the participating teams comprise a considerable variety along a number of key dimensions such as organizational setup, main goals, leadership style or wider organizational context. Thus, instead of examining commonalities and differences based on the sociometric data directly, the following section explores case study characteristics relying on the additional data sources collected, such as the semi-structured interviews and the short questionnaire. Based on the empirical research, important differences between research groups emerge – which, in a second step will be linked back to the sociometric data where possible. The descriptive accounts of the following paragraphs therefore also provide an opportunity to understand better the interaction pattern of each team before using more quantitative approaches in the next chapter.

The central contribution of this section consists of providing comparative insights how the four social concepts "leadership", "hierarchy", "collaboration", and "gender diversity" could become manifest in the sociometric data. Consulting the interview material for each case study, it is clear that important differences do exist for example concerning the "leadership style" between the teams. However, how "leadership style" or other social concepts of interest do become apparent in the sociometric interaction pattern and network statistics is far from clear.

Several dimensions of the sociometric interaction profile are examined for differences and commonalities:

Team collaboration

- refers to the "shared lab" index which indicates if a research group works predominantly in a shared space such as a lab or has separate offices.
- Assesses how "uniform" or "diverse" the aggregated face-to-face network statistics are. In some teams, few individuals dominate the interaction whereas in others, interaction patterns are more uniform and shared between all members.
- Analysis of role-based interaction: which roles interaction the most with each other (e.g. Leader – Senior, or PhD – Postdoc)?

Leadership

- Insights regarding the leadership style can be gained from the face-to-face network statistics of the leader vs all other team members.
- Furthermore the speech profile indicates the total and mean speaking and listening time of the leader (versus all other team members). More dominant leaders will have a relatively higher speaking time.
- The Shannon Entropy values of the team leader vs. all other team members should also be considered. A team leader should have Shannon Entropy diversity values in

the medium range, indicating interactions with a variety of team members but not the entire team on a constant basis.

- Last but not least the number of highly synchronized body activity events is interpreted in terms of the level of strong/centralized steering. The fact that team members move in a synchronized fashion has to be seen as the exception rather than the rule; the more group members move in an orchestrated way, the more a centralized steering of some sort can be assumed.

Role-based Hierarchy

Each team member has indicated her/his team role, ranging from "Team leader" to "Senior researcher", "Postdoc / Junior", "PhD student", "MA student", "Administrative Assistant" or "Research / Lab Assistant". Roles can be ordered in terms of a hierarchy with more senior roles being located at the top and more junior and assistant roles at the bottom. Team roles thus provide important information regarding the potential hierarchy within the group by analyzing interaction patterns or the speech profile according to these roles.

- Does the speech duration (of the most dominant roles) correspond to the role based hierarchy? In part, the leadership analysis already answers this question regarding the mean duration of the leader vs. all other team members.
- Does the face-to-face network statistics – which indicate for example influential team members coincide with the team roles of each member?
- Are there any patterns regarding the analysis of interaction patterns between roles? Hence, for research lab based groups, a "normal" pattern is a comparatively high collaboration between PhD and Postdocs within the lab since these team roles carry out most of the actual research work. The degree to which certain role dyads (e.g. PhD-PhD or PhD-MA) interacts provides insights into the hierarchy within the team.

Gender differences

- Gender differences can be analyzed according to the relative speaking and listening duration of women and men in the team.
- Gender differences are also apparent by the mean proximity and face-to-face detects of women and men within each team.

Altogether 12 dimensions are available to establish differences and commonalities across the 8 sociometric team profiles. As will become apparent, some of these map neatly unto the general insights that are provided by the interviews and the questionnaire data. On this qualitative level, the sociometric data characterizes reasonably well differences between the teams along the just mentioned dimensions of "team collaboration", "leadership" mainly. It needs to be seen, to which degree more precise statistical measures can be provided – especially when exploring the dynamic nature of the data with the Relational Event Modeling (see page 78)

During the course of WP2, eight individual case study reports have been produced that contain detailed information on the sociometric profile of each team. Each report also contains a summary of the insights derived from the other, non-sociometric data sources collected during each case study, including the semi-structured interviews and the short questionnaires. The individual case study reports are not public due to privacy issues. However, summary statistics as well as data visualizations are provided in ANNEX I – Comparative Case Study Table on page 103 and ANNEX II – Selected Graphs / Charts on page 106. The contents of the following paragraphs is based to a large degree upon these individual case study reports.

Individual Case Study Analysis

In the following section, each case study will be described very briefly, characterizing its main features in terms of leadership, team collaboration, hierarchy and gender diversity. In a second step the main features of the sociometric profile will be summarized in order to see if commonalities and differences between the cases map onto commonalities and differences regarding the sociometric data.

Team 1 – Main Features

T1 is a university/teaching based research team comprising 8 members (3 women, 5 men), with a mean age of 46 years and the mean team tenure is 5.14 years. The team has a long history and counts on senior members that form part of it for the past 7 years at least. The group comprises large age- and tenure differences between younger PhD students and senior figures. Research activities compete with the demands of teaching responsibilities which makes it hard to focus on a collaborative research process – as stated several times during the interviews. Leadership is shared between relatively young (and new) leader and the previous leader of the group (both men) which corresponds to two research lines developed within the group. Age and status differentials are large, comprising senior members with a long trajectory and much younger PhD students. The duality in research lines together with the teaching responsibilities make a shared, explicit vision of the group a challenge. Interactions of the leader largely happen on the dyadic level directly engaging with other seniors or PhD students directly. Decisions are taken by consensus and conditioned clearly through the close interaction among seniors. There is an explicit appreciation of diversity by senior team members in terms of recognizing the innovation need and potential of younger (PhD) students.

Sociometric Profile

The sociometric profile examines primarily face-to-face interaction frequencies, the aggregated network statistics (see Table 22 on page 122) as well as the speech profile summarized in Annex I.

Team collaboration. The shared lab index is 0.25, correctly indicating a teaching/university based team. The mean proximity detects per team member are 2857 (sd=1322) and the mean face-to-face detects are 427 (sd=290) situating it at the lower end of interaction frequency in comparison to the other teams. The mean face-to-face detect per member and day is 85. Examining face-to-face network statistics for T1 suggests three members as highly influential: the team leader, one senior position and one PhD position have very similar Eigencentrality scores and degree count. The three team members also share a similar high score for the Shannon Entropy measure. These 3 most active members are responsible for 60% of all interactions of the entire group of 8.

The face-to-face statistics for this team is rather diverse and conditioned by specific team members which influence the overall face-to-face interaction pattern.

Interaction happens predominantly across team roles, i.e. in a horizontal fashion involving senior researchers and PhDs as well as team leaders and PhDs directly, even though this is highly determined by the three dominant members who belong to a different role each.

Comparing the proximity and face-to-face profile, an inverted relationship is observed between PhD – senior roles: most face-to-face interactions happen between PhD and Seniors while there are relatively few proximity detects. This indicates intentional meetings in specific times and places. The inverse holds for Senior – Senior interactions which count on a high opportunity structure (many proximity detects) but only few direct face-to-face interactions.

Leadership. The network statistics suggest a certain ambiguity regarding the team leader – who has a similar position to PhD and another senior position. In fact, the similar scores correspond to the “dual” leadership where the other senior position corresponds to the previous team leader. Interaction happens across different roles directly, in a horizontal fashion.

The speech profile emphasizes the centrality of the leader since the highest mean listening and speaking duration pertains to the team leader. However, again, total speaking and listening duration for the three dominant team members is very similar (3.6, 3.6 and 3.4 hours).

The number of synchronized body activities is 4. In comparison to other teams, this is a medium range value, indicating a certain coherency where the team acts as a global unity.

Role-based Hierarchy. The face-to-face interaction is horizontal rather than pyramidal meaning that senior and junior roles interact directly with each other. There is no delegation from senior towards junior roles. Most influential positions (Eigencentrality) according to role in face-to-face network in decreasing order of importance: Senior, PhD+Leader, PhD+Senior, Senior. Leader has a strong presence in the speech profile, i.e. highest mean listening and speaking duration. Mean speaking and listening very similar among PhD, Senior and Leader.

Gender differences. Most interactions happen among mixed gender dyads. Although there are more men than women in the team, women (481) accumulate on average more face-to-face detects than men (385), i.e. women are over-represented in interactions. Women also have a higher mean speaking (2.98) and listening duration (1.67) than men (2.18/0.74).

Summary Insights for T1

Certain individuals dominate the sociometric profile, comprising a core-group of three team members involving a PhD, Senior and Leader position. The split between this highly influential group of three and the rest of the team which is less active indicates is visible in the diversity of sociometric profiles characteristic for teaching based groups. The status and age differentials within the team are not reflected in the sociometric profile where most interactions do happen across (senior vs. junior) roles. The strong interaction across hierarchies emphasizes the appreciation of diversity in the team. The sociometric profile indicates correctly a somewhat “split” leadership role, i.e. the equal importance of the formal leader and one other senior position. Leadership does not involve delegation. There are marked gender differences due to over-representation of women in interaction and speech profile.

Team 2 – Main Features

T2 is a pure research based team, comprising 10 members (7 women, 3 men), with a mean age of 34 years and mean tenure of 1.56 years. It is a quite young team, recently founded and with most members joining the team during the past 2 years. The leadership style could be described as “participatory” and “care oriented.” Without an explicit strategy, human relations within the team receive special attention by the leader, who is a woman, by caring for a “personal” space of each person and by being attentive to buffer the possible “frustrations” generated by lab- and experimental work. The team leader is one more team member, where the entire team operates in an open, trust-based climate. There are some centralized activities such as a bi-weekly “journal club”, but most interactions happens ad-hoc, on an informal basis. Decisions are taken by consensus. One of the main features of this team concerns the pro-active approach to interdisciplinary work; the team members have been recruited from different scientific disciplines like Biology, Chemistry, or Physics in order to examine research questions from different angles / knowledge fields. This requires active dialog among team members to bring different disciplinary backgrounds to bear on shared problems.

Sociometric Profile

The sociometric profile examines primarily face-to-face interaction frequencies, the aggregated network statistics (see Table 23 on 122) as well as the speech profile summarized in Annex I.

Team Collaboration. The shared lab index is 0.9 correctly indicating a research lab team. Mean proximity detects are 13314 (sd=6031) and mean face-to-face is 1735 (sd=1085) detects per member, which indicates abundant interaction in comparison to other teams. The mean face-to-face detects per member and day is 347. The interaction profile is also highly uniform, i.e. most face-to-face network measures are very similar: 7 out of 10 team members, including the team leader have Eigencentrality values between 0.9 and 1 and Degree scores between 8 and 9. Thus, there are no influential individuals which would condition the overall interaction profile. The 3 most prolific interaction members are responsible for 51% of all interactions within this group of 10. Interaction among PhD, Juniors and Seniors does happen in a horizontal fashion where PhD to Senior interactions are most frequent, followed by Senior to Postdocs and Postdocs to PhDs.

Team Leadership. The team leader blends in with most other team members regarding face-to-face network statistics. Shannon Entropy measures are highest for PhD, Postdocs and Leader, suggesting that communication of the leader with team members is relatively balanced. However, absolute and mean interaction counts signal a more nuanced picture with most interactions of the team leader being directed towards Seniors and Junior/Postdoc positions and only limited interaction with PhDs. This suggests delegation from leadership to Senior and Junior positions.

The number of synchronized body activity patterns is 3. There are few centralized activities and active steering.

The speech profile establishes highest mean speaking for the group leader but lowest listening duration.

Role-based Hierarchy. Interactions according to role suggests delegation from leader primarily to Seniors and then Junior/Postdoc positions. However, face-to-face network statistics does not seem to indicate an especially influential role to the team leader who shares with 7 others high Eigencentrality and Degree scores. Besides the team leader, the rest of team members do interact in a very horizontal fashion irrespective of hierarchical roles.

Gender differences. Although there are 7 women and 3 men on this team, the mean face-to-face detects for women is 1585 vs. 2085 detects for men. This means men are over-represented in face-to-face interactions within this team. Most interactions do happen among same-gender dyads. However, women have on average a higher speaking duration (3.4 hours) vs. men (1.4 hours) and a higher mean listening duration 4.43h (women) vs. 4.37h (men).

Summary Insights for T2

The network statistics of this team coincide strongly with the general description of a horizontally organized team where most team members interact with each other, especially considering Senior, Junior/Postdoc and PhD levels. The face-to-face statistics indicate shared, uniform team roles, including the team leader. This observation also holds according to the interaction profile where communication among team roles is quite balanced among Senior, Postdoc and PhD level. At the same time, leadership is somewhat hierarchical in that most interactions from the leader target senior and junior members and less PhDs. The leader also has a strong presence due to highest mean speaking duration. The 3 synchronized body activity events are in the middle range compared to other teams. Overall, this seems to be a high interaction team having most team members participate equally irrespective of hierarchical roles. The team leader, although “being one of the team”, has a certain presence as indicated by the speech profile and centralized activities.

Team 3 – Main Features

T3 again is a university/teaching based research team with 8 members (3 men and 5 women), with a mean age of 36 years and mean tenure of 3.19 years. One of the central characteristics of this research group concerns its high heterogeneity in terms of roles and academic trajectories mirroring diversity of responsibilities and experiences among staff of a university (and hence department) based environment. Leadership in this sense is conditioned by people's backgrounds, the wider organizational environment and overall teaching responsibilities. It is also marked by age differences that not necessarily correspond to years of experience and academic ranks. Hence, the possibilities to actively shape the group through a distinct leadership style are bound by external conditions. The team has a shared laboratory although most members have also their own office space. Research activities compete with a very demanding teaching agenda and are carried out in “individual” time slots where researchers work mostly on separate sub-tasks without direct interaction. The group therefore does not operate too much on the global level as a unity but addresses individual collaborations and work within the confinements of the overall university/departmental environment.

Sociometric Profile

The sociometric profile examines primarily face-to-face interaction frequencies, the aggregated network statistics (see Table 24 on page 123) as well as the speech profile summarized in Annex I.

Team collaboration. The shared lab index is 0.25 correctly identifying this team as university/teaching based group. The mean face-to-face interaction count is 1870 (sd=2292/) and the mean proximity count is 6334 (sd=5284) which is a comparatively high interaction count for a teaching based group. This is also confirmed by the mean face-to-face detects per day and team member, which is 374. The sociometric profile for T3 suggests an interaction profile that is dominated by three highly influential individuals (MA student, PhD student and team leader) in terms of frequency of face-to-face detects: these three team members are responsible for 93% of all interactions. The face-to-face network statistics of Eigencentrality and Degree scores is quite diverse, suggesting a sociometric profile that is highly dependent upon certain individuals. Little can be said regarding the interaction among the wider group of team members because there are very few detects in general.

Comparing proximity and face-to-face profiles shows important differences between the proximity profile vs. the face-to-face interactions. There are clearly more frequent detects among a wider variety within the team than observed face-to-face interactions. This underlines further the intentional nature of the interactions among the dominant pairs.

Team Leadership. The leadership role is not clearly identified by the sociometric data. Network statistics do not signal the leader as especially influential compared to other team roles. The Shannon Entropy score is in the medium range indicating an unbalanced interaction with few selected others by the team leader. Most interactions of the leader happen with a PhD student. There is no delegation nor cascading.

The number of synchronized body activity patterns is 2. The group does seldom “move” as a unity. There are little centralized activities.

The speech profile for the leader resides in the medium range. Highest as well as lowest speaking and listening duration pertain to other roles, not the leader.

Role-based Hierarchy. The interaction pattern does not produce insights regarding different behaviors of senior vs junior roles. Senior profiles are largely absent from the interaction with most face-to-face happening from Leader to PhD and from PhD to MA student directly.

Gender differences. There are 5 women in this team and 3 men. Mean detects of women is considerably higher (2039) than the mean face-to-face detects for men (1445). Women therefore are over-represented in interactions. However, mean speaking duration for women is 3.09 hours vs. 3.61 hours for men and mean listening duration is 2.92 hours (women) and 1.25 (men). Although women are disproportionately participating in interactions, men are more dominant considering speaking duration.

Summary Insights for T3

Interaction for T3 is dominated by three individuals. The interaction profile and network statistics mirrors correctly the selective and rather fragmented nature of the team. The speech profile as well as the network statistics for the team leader does not indicate a strong direction and unit, despite the high face-to-face detects. This sociometric data here coincides with the general observation regarding the rather fragmented research activities that are conditioned by the wider organizational context and teaching responsibilities. The sociometric profiles indicates highly influential individuals which furthermore underlines a rather less cohesive group in terms of research.

Team 4 – Main Features

T4 is a pure research based team, comprising 9 members (5 women and 4 men), with mean age being 35 years and mean team tenure being 4.19 years. The team has a relatively "longer" history, given that the mean tenure of senior members only is 7.29 years. During its history, leadership changed with the previous male leader leaving and the current team leader (woman) taking over. The legacy of its history and the trajectory of the group condition to a certain degree its present through the diversity of agendas, expectations and trajectories within the group. The necessity to secure funding for the team is the principal concern for the leader which also places constraints on her availability to the group. Interaction of the team leader happens predominantly with senior members and less with the entire group; leadership style has been described as not overtly demanding/pressing but leaving members a large degree of autonomy which on the one hand has been perceived as not sufficiently present/accompanying on the other. Most decisions are taken through consensus, although this vision is not necessarily shared by all members of the team.

T4 Sociometric Profile

The sociometric profile examines primarily face-to-face interaction frequencies, the aggregated network statistics (see Table 25 on page 123) as well as the speech profile summarized in Annex I.

Team Collaboration. The shared lab index is 1.0 and correctly identify this team as a research lab based group. The mean face-to-face interactions per member is 596 (sd=422) and the mean proximity detects per member is 11905 (sd=6646). Although this is a research lab team, the mean interaction detects is rather low in comparison to other teams. The mean face-to-face detects per member and day is 119. At the same time, the face-to-face network statistics suggests a diverse profile with team members participating to different degrees in the overall interaction. Examining the Chord Diagram, one can see that most interactions happen among PhD students and among PhD and MA students. What is furthermore characteristic is a lack of interaction between Seniors and PhD / MA students as well as an overall low participation of the team leader. The 3 most prolific team members (2 PhDs and 1 MA) are responsible for 57% of all interactions within this team of 9.

Opportunity structures for interaction match actual face-to-face interactions for most dyads, except for Research assistants – PhDs: they have a high proximity detect which do not translate into a similar amount of face concrete interactions.

Team Leadership. Most distinguishable feature of this team is the low participation profile of the team leader. Network statistics based on face-to-face interactions are low with lowest Eigencentrality and Degree scores of all team members. The leader has the second highest Gini-C Concentration coefficient. There is no delegation, nor cascading. The limited interactions of the Team Leader happen with senior positions and with PhD students, to similar degree.

The number of synchronized body activity patterns is 1, indicating that the group does seldom “move” as a unity. There are little centralized activities.

The group leader has a low speech profile. It has the lowest listening duration and the second lowest speaking duration.

Role-based Hierarchy. Most interactions happen between PhDs and PhD+MA students. Team Leader and Seniors are rather absent and thus little can be said about delegation. Most influential positions (Eigencentrality) according to role in face-to-face network in decreasing order of importance: Senior + PhD, MA + PhD, PhD showing influential positions across team roles.

Gender differences. There are 5 women in this team and 4 men. Mean face-to-face detects of men are considerably higher (774) than for women (453). Men are over-represented during interactions in this group. This also holds for the speech profile: mean speaking duration for men is 3.97 hours vs. 3.30 hours for women. Mean listening duration is 3.99 hours for women and 4.91 for men.

Summary Insights for T4

Among the most interesting features of this team concerns the diverse sociometric profile which is in contrast to the research lab based environment where it operates. Thus, the overall amount of interaction detects is quite low. The network statistics suggest a diversity of roles which overall matches quite well the impressions obtained throughout the interviews indicating equally diverse individual trajectories of team members. Part of the diversity of profiles is given by the less active sociometric profile of the team leader which runs across the face-to-face statistics, the overall low number of interaction detects as well as the speech profile. There are few synchronized body activities and the role based analysis suggests that there is a certain separation between senior levels and the junior roles which also has been confirmed by the interviews. Even though this team is research lab based, the interaction network statistics are diverse indicating unequal participation and roles within the group.

Team 5 – Main Features

T5 is a pure research based group, comprising 11 members at the time of the case studies (5 women and 6 men), with average team tenure being 0.92 years and average age 28 years. Clearly, this is a relatively young research group both in terms of age of its members as well as team membership. A professional management approach has been pursued by the team leader (man), carefully planning and fine-tuning the scientific- but also human relations within the group. Leadership in terms of work responsibilities is pyramidal, with the leader mainly interacting with senior members who collaborate with Juniors and PhDs. Decision making for some issues is nevertheless horizontally organized. There is a clear vision of the group as a research performing unit that is steered by the leader. The fact that the team is also quite new and the fact that it is a pure research setting favors the design of an explicit leadership agenda where objectives can be put into practice.

Sociometric Profile

The sociometric profile examines primarily face-to-face interaction frequencies, the aggregated network statistics (see Table 26 on page 124) as well as the speech profile summarized in Annex I.

Team Collaboration. The shared lab index for this team is 0.72 identifying it correctly as a research lab based group. Mean face-to-face detects per member is 2322 (sd=1432) and mean proximity detects per member is 21644 (sd=8638). Comparing with other groups, this is a highly interacting team where the mean face-to-face detects per member and day is 464. The sociometric profile for T5 establishes for 8 out of 11 team members identical network statistics with highest Eigencentrality values (1.0) and degree scores (10), including the team leader. Most team members therefore share uniform face-to-face network statistics making it impossible to single out any highly influential members. The 3 most prolific team members are responsible for 50% of interactions.

The collaboration profile is pyramidal and cascades from higher roles towards more junior positions where interaction increments in frequency for more junior roles: the team leader interacts predominantly with Junior/Postdoc positions which in turn interact with PhDs. Most interactions happen at the PhD and Junior level.

Comparing proximity and face-to-face profiles shows that the opportunity for interaction due to proximity detects with the leader are higher than the actual face-to-face interactions. This means that the team leader is available but real interaction happens predominantly with specific roles.

Team leadership. The face-to-face network statistics do not signal any special role for the team leader, other than being one more member of the team as he shares highest Eigencentrality, and Degree scores with 8 other members. Highest Shannon Entropy scores belong to senior- and postdoc positions with the team leader scoring in the middle/high range.

The number of synchronized body activity events is 7. This is the highest score across all teams. It indicates a strong unity and direction at the group level.

The speech profile indicates that the team leader has the highest absolute and mean speaking duration across all team members and a medium profile for listening duration.

Role-based Hierarchy. The interaction pattern according to roles is hierarchically ordered: the leader interactions with Postdocs/Juniors which in turn interaction with PhDs which among themselves collaborate massively. This is a pyramidal arrangement respecting hierarchy of roles. The frequency of interactions corresponds to this pyramidal arrangement: most interactions happen at the PhD levels, decreasing in frequency at more senior positions.

At the same time, the face-to-face network statistics indicate a uniform group profile where most team members form part of the overall communication pattern.

Gender differences. This team has 5 women and 6 men. Mean face-to-face detects are 2247 for women and 2384 for men. The participation between men and women in interactions is quite balanced. Men are more dominant speakers with mean speaking duration for women being 2.6 hours and 3.38 hours for men. Mean listening duration for women is 7.04 hours and 6.28 hours for men. Most interactions do happen among mixed-gender dyads.

Summary of T5

The interviews characterize this team as horizontal where decisions are made in a consensual manner. The face-to-face network statistics support this finding where 8 out of 11 members have identical centrality and degree scores. The uniform profile does not single out any especially influential team members. At the same time, work is distributed in a pyramidal fashion where frequency of interactions cascades from the leader- to senior- to junior positions, augmenting in intensity. The strong leadership is visible and consistent across the different sociometric dimensions, including the high mean speaking and listening duration for the leader, the high Eigencentrality and Degree scores, as well as the especially high centralized body activity pattern.

Team 6 – Main Features

Team 6 is a university based research team, comprising 10 members (7 men, 3 women), with average team tenure being 2.56 years and average age 38. The team has been reorganized in 2015 with a newly appointed leader (man). Leadership has been described as participatory and supporting autonomous decision making of its members. The leader has been described as high status while being very accessible at the same time, caring for team members and building trustful relations. The team integrates different roles and tasks that are not always connected to each other such as laboratory work, consultancy or teaching responsibilities. The work is also split between technicians, responsible for experimental setups, and academic roles who build computer models for certain processes. Technicians are shared with other teams. The overall impression is that team members are well informed regarding their roles and work rather autonomously on their individual tasks. There are no clear collaboration patterns emerging among team members given the split between teaching activities, research- but also consulting activities. Although there is weekly group meeting, these are not necessarily attended by the entire group.

Sociometric Profile

The sociometric profile examines primarily face-to-face interaction frequencies, the aggregated network statistics (see Table 27 on page 124) as well as the speech profile summarized in Annex I.

Team Collaboration. The shared lab index for this team is 0.2 correctly identifying it as a teaching/university based environment. Mean face-to-face detects per member are 845 (sd=1032) and mean proximity detects per member is 2868 (sd=1785). The mean face-to-face detects per day and team member is 169. The sociometric network statistics for team members is rather diverse, conditioned by 2 to 3 influential individuals. Most interaction does happen between three individuals (Leader and two Postdocs), with the rest of the team members being rather absent from the interaction profile. The three most prolific team members account for 87% of all team interaction in a group of 10.

The collaboration profile is pyramidal in that most interactions happen between the team leader and Postdoc positions. However, there is little interaction among the more junior roles. Again, the interaction profile is conditioned to a large degree by 2-3 highly influential individuals.

Comparing proximity and face-to-face profiles shows that the opportunity for interaction due to proximity detects is much higher and varied than the real face-to-face interactions. Indeed, one of the characteristic features of this group profile is the "intentional" character of the face-to-face meetings. This group has the highest mean RSSI values across all teams, meaning that when people come close to each other, this happens for explicit meetings where RSSI signals are very strong (due to the close range in which badges are). However, at the same time, the proximity profile also shows that PhD – Assistants as well as Postdocs – Assistants have relatively high proximity detects, without converting these into real face-to-face meetings.

Team leadership. The face-to-face network statistics does not signal a highly influential profile for the team leader. The number of synchronized body activity events is 2.

The leader has a low speaking profile; highest mean speaking and listening duration has been observed for the Assistant role.

Role-based Hierarchy. The interaction pattern according to roles respects to some degree the hierarchy based on roles: most interactions happen between Leader and Postdocs. However, senior positions are absent from the profile. Also, interaction among junior roles is quite scarce. Thus, the interaction profile is conditioned by three influential individuals which happen to be Leader, Postdoc and PhD. The face-to-face network statistics suggest the Postdoc, Senior and PhD as most influential.

The speech profile suggests an inverted hierarchy with the Assistant role having the most dominant position, followed by postdoc and senior team members.

Gender differences. This team has 7 men and 3 women. Mean face-to-face detects for women are 404 and for men 992. Most face-to-face interactions do happen among men. The mean speaking time for women is 2.62 and for men 2.28 hours. Mean listening time is 1.09 hours for women and 0.98 for men.

Summary of T6

Team 6 is a university based research team where members work on rather separate tasks while team members are seated in separate offices as well. This fragmented profile is visible in the sociometric data profile which is dominated by three individuals without providing a clear picture that would involve all team members. Interaction is rather scarce. At the same time, there is no clear leadership according to the face-to-face network statistics and speech profile when taking into account all team members, including assistant positions. Face-to-face meetings do occur at selected times and places, an observation that can be derived from the very high mean RSSI values as well as the discrepancy between the diverse proximity detects profile and the much more limited face-to-face detects.

Team 7

Team 7 could not be included in the comparative analysis due to time constraints in the preparation of the analysis.

Team 8

Team 8 could not be included in the comparative analysis due to time constraints in the preparation of the analysis.

Summary of Data Dialog across Case Studies

One of the central questions regarding the case studies concerns the interpretation of the sociometric data. The unique opportunity of the GEDII case studies lies in the combination of a comparative view across case studies with in-depth descriptions of each research group. This is a complex task, given the overall number of case studies in combination of the variety of data sources available as well as the relative novelty of the sociometric data used. When summarizing the findings across the case studies one should bear in mind this complexity involved and treat the interpretations with caution. Nevertheless, it seems that some general patterns between research groups can be observed that "match" reasonably well across the data sources.

The most consistent, basic and clear difference between research groups concerns their organizational environment with pure research groups on the one hand and mixed teaching/research based groups on the other. This difference seems to condition substantially the possibilities of research activities: naturally, research groups that operate within a university cannot dedicate 100% of their time to research activities. This affects to a considerable degree the interaction pattern of the research groups as observed with the sociometric profiles. In research labs, team members do work together on a more continuous basis conditioned by a shared space (the laboratory) as well as the shared time. Research groups, on the other hand, whose members have teaching responsibilities are much more restricted when and with whom they can interact; and, this limited possibilities to work together face-to-face is often exacerbated by individual office space in universities. This basic difference between research centers and universities is visible on the one hand in the “shared lab” index which compares simple the average group size of interactions with the team size. However, it is also visible through the uniform vs. diverse face-to-face network statistical profiles. Research groups that operate in dedicated research centers have a much more homogeneous profile conditioned by frequent and balanced interaction of most team members among each other. Research groups that are located in universities on the other hand produce much more diverse profiles; since the organizational environment prevents a uniform and full interaction among all team members due to the diversity of roles, office spaces and tasks, these differences become visible as distinguished sociometric profiles for certain individual team members.

Another relatively consistent observation concerns leadership style. The combination of the speech profile, face-to-face network statistics and synchronized body activities provides three dimensions to analyze commonalities and differences between the groups. There are marked differences regarding the centralized activities (synchronized body activity pattern) as well as the mean speaking duration for leaders. These differentiate quite reliably between different levels of “presence” of group leaders within their team. Thus, comparatively strong leadership could be observed in Team 5, Team 1 and Team 2 whereas leaders are rather absent from the sociometric profile of Team 6, Team 4 and Team 3. Exemplary in this sense is certainly the high number of synchronized activities for T5 vs. the few synchronized activities of T4 being both groups located in a research center/lab which also corresponds to high vs. low speaking duration for the leaders. Frequent synchronized activities suggest here a certain team cohesion and unity often from “above” that establishes the team on the group level. As such it does not necessarily say much about the actual collaboration pattern of the team: members can interact fluently on an individual, ad-hoc basis (see for example T2). However, providing a basis for the team to come together as a unity – no matter if this happens for leisure activities, shared lunch, coffee or seminars – says something about the leadership style of each team.

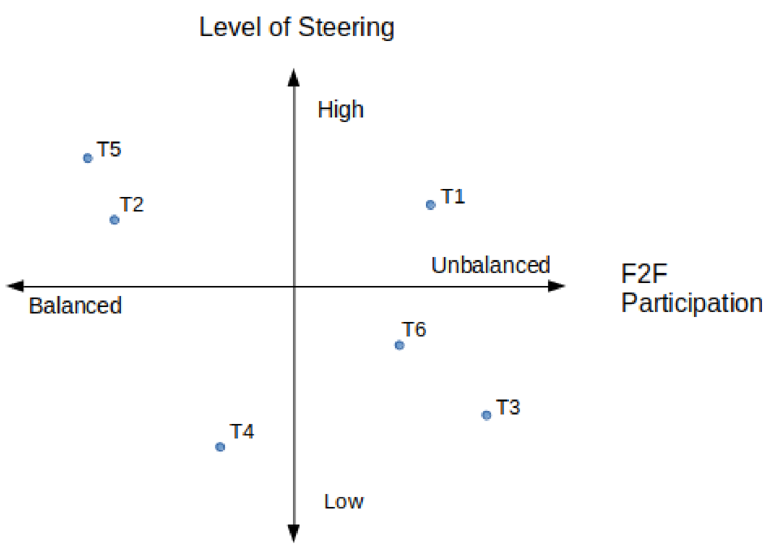


Illustration 40: Team typology

By combining insights regarding the relative “uniform” nature of sociometric profile (e.g. T2, T5) vs. more “diverse” and “individualized” profiles (e.g. T6, T3, T1) with insights regarding the centralization of activities and leadership style suggest further lines of comparison. Thus, Team 1 has an overall rather individualized profile that corresponds to a university/teaching based group. At the same time, it has a relatively strong leadership with 4 synchronized activities as well as high mean speaking and listening duration for the team leader. In comparison, T4 is a group working in a research center, with relatively few synchronized activities as well as low mean speaking and listening duration for the team leader. Altogether, both leadership as well as the overall interaction pattern opens a two-dimensional space of 4 quadrants to differentiate teams as shown in Illustration 40.

The team collaboration according to roles has produced a certain variety of profiles. On the one hand, those sociometric profiles where a limited number of individuals are highly influential do not allow to say much about the overall collaboration according to roles. The analysis in this respect is quite constrained to the specific roles of these highly visible members. Thus, T3 for example shows a strong interaction between the team leader and PhD. This is a horizontal, direct collaboration. However, since other team roles are not visible in the interaction profile, the observation does not say much about the overall role based collaboration. On the other hand, in research teams where the participation of members is more balanced, broader interaction patterns according to roles become apparent – that can be characterized either as pyramidal or more horizontally oriented. T5 for example shows a pyramidal arrangement where the frequency of interactions augments from senior to junior levels. It seems that task and responsibilities are delegated from the leader towards the Postdocs which in turn interact more with PhDs. On the other hand, Team 2 for example has a more horizontal arrangement: even though interaction between the team leader is limited again to the more senior roles, all other team members interact to a similar degree with each other. Overall, it seems that the role-based analysis of interaction can potentially provide interesting insights regarding the hierarchy and collaboration pattern among different roles within the team – as long as a balanced overall interaction profile is available (e.g. T2, T5).

Finally, the most difficult dimension to interpret commonalities and differences across the research teams concern gender issues. Neither the interviews nor the sociometric data did produce a clear pattern *on the level of the groups*. Certainly, the importance of gender for research careers, team leadership and research collaboration does emerge during individual interviews. Awareness of gender issues certainly are more prominent in interviews of women at more senior levels, often with children reflecting their own experiences. However, persistent gender differences that would characterize differences between groups beyond the simple mean face-to-face or proximity detects as well as speaking and listening duration were not apparent. A regression analysis on total proximity detects only identified “age” as a significant covariate but none of the other variables including gender. Comparing the speech profiles across all 80 participants did not produce statistically significant differences regarding gender[20]. Neither did the body mirroring values produce significant differences between women and men across teams. The analysis of gender differences has been furthermore limited due to the quality of the turn-taking data, which could not be used. However, as the following section will show, new analytic approaches such as the Relational Event Modeling might open up new perspectives for detecting gendered interaction patterns in time-based data.

20 However, “age” again was significant. See Table 19 on page 56 for differences regarding listening duration and age!

Relational Event Modeling

Most research on teams does not or only marginally consider the time dimension. Although there have been repeatedly calls to incorporate a more dynamic view in the study of small groups, these have not been entirely successful (Arrow, McGrath, & Berdahl, 2000; Cronin, Weingart, & Todorova, 2011; Humphrey & Aime, 2014; Roe, Gockel, & Meyer, 2012). The classical "Input-Process-Output" (I-P-O) model has informed much research on teams, focusing on how certain "Inputs" produce desired "Outputs" while black-boxing largely the dynamic "Process" in the middle (Ilgen, Hollenbeck, Johnson, & Jundt, 2005; Kozlowski, 2015). There is agreement that the I-P-O model "fails to capture the emerging consensus about teams as complex adaptive systems." (Ilgen et al, 2005, p.419).

Leenders, Contractor, & DeChurch (2016) summarize 4 challenges that hamper the incorporation of a dynamic, time-based view in the study of small groups. The first challenge assumes falsely a "homogeneity over time" where team events are aggregated into a single, static slice stipulating that input-output relations are time-invariant. This temporal "collapse" is unfortunate since it eliminates the analysis of path dependency, erroneously assuming that the order of events does not matter. A second challenge concerns the assumed homogeneity across team members and their interactions. Without being able to monitor interactions within the team over time, existing approaches assume that the behavior of individuals is constant as if they could not fluctuate in their intentions, opinions or goals over time. However, interaction dynamics across dyads do vary and affect in turn the overall interaction profile at the team level. A third challenge assumes that part of these previous problems can be resolved by aggregating repeated, "static" measurements in an attempt to approximate team dynamics. However, studying team processes as a limited series of snapshots fails to address genuine time-based issues regarding duration, feedback loops, or temporal scales. Research needs to address "continuous time", since multiple discrete time points are seldom reflective of "actual temporal processes". And finally, the fourth challenge concerns a lack of time-based theory and thinking that prevents researchers from formulating process-based hypothesis. Seldom does research inquire path dependencies, the "pace, trajectories, and cyclicality" of interactions in teams: "Many of our hypotheses of team processes (typically: ''teams higher on X are also higher on Y'') are static in formulation—they do not explicitly describe temporal relations between variables nor do they call for longitudinal data to test them." (ibid. p.96).

There are pragmatic reasons why a truly temporal approach to the study of team processes is relatively infrequent: collecting time based data is not only very resource intensive but also potentially very taxing on research participants. Recent technological developments such as Sociometric Badges, however, address some of these challenges. With the availability of sensor based technologies such as RFID Beacons, Bluetooth- or infrared sensors, real-time monitoring of interaction among people in groups has become feasible at a relatively low cost.[21]

21 The new technologies are of course not limited to research but actively deployed already in consumer and marketing research. See for example https://kontakt.io/ or https://www.mocaplatform.com/ or https://www.polestar.eu/

Concurrent to these technical developments new statistical approaches have become available to analyze social sequences (Cornwell, 2015) and temporal processes that go beyond the aggregation of static time slices. “Relational Event Models” (REM) in particular shift the unit of analysis from individuals to “interactions” that occur at a specific points in time between one or more members of a particular group. The term has originated in the social networks literature, specifically in the work of Carter T. Butts (Butts, 2008; Butts & Marcum, 2017; DuBois, Butts, McFarland, & Smyth, 2013; Marcum & Butts, 2015).[22] The innovative element of the REM is the fact that exchanges between senders/receivers are modeled as a sequence of time-stamped events where the unfolding of these events are both conditioned endogenously (prior events) and exogenously that is, by the wider environment. The most simple example of a relational event would just involve the sending of a message from Ego to Alter at a certain point in time. Endogenous effects take into account how the amount of past interactions between these two individuals condition future interactions. For example, “inertia” is a typical endogenous effect that assumes that the future rate of communication between two individuals depends upon their frequency of communication in the past; friends that are used to talk a lot will do so in the future. However, the unfolding of an event sequence is not only conditioned by path dependent internal events but also externalities such as for example the “absence” of a team member. Friends break up. The REM provides a mechanism to take into consideration both endogenous as well as exogenous events in order to predict the occurrence of future events. Thus, as Schecter & Contractor (2017, p. 224) maintain, a REM captures “in a single model the influences of individual, dyadic, triadic, and group-level characteristic on the dynamic unfolding of collaboration processes.” As a result, the previous drawbacks regarding the assumption of homogeneity, both among team members and over time, in the study of small groups can be adequately addressed.

Butts (2008) as well as Marcum & Butts, (2015) provide the basic modeling framework and R packages to work with the Relational Event Model. Although several papers exist that make use of REMs across a variety of topics, overall the available references are rather limited. Butts & Marcum (2017) provide a step-by-step introduction to REM with two different type of datasets; Vu, Pattison, & Robins (2015) study MOOCs, while Tranmer et al. (2015) demonstrate the general power of REMs for studing animal behavior sequences. Quintane et al. (2014) study Relational Events in an Open Source Software project, extending the original model to two-mode networks. And Pilny et al. (2016) and Schecter et al. (2017) provide two further examples of a REM analysis in the context of military communication settings. However, given the ten year time period since the original work on the REM has been published by Butts in 2008, these are very few studies.

Somewhat surprisingly, to the best of our knowledge, the REM has not be used with sociometric data directly. Pentland's et al. (2017) call for “action-centric research” contrasts all the more with the missing publications that would apply Relational Event Modeling with sociometric data. Most research analyses the network data using a static approach, aggregating the overall timestamped interactions into a-temporal snapshots of the entire network (Gloor et al., 2011, 2012; Onnela et al., 2014; Tripathi & Burleson, 2012; Wu, Waber,

22 Recent develops include also the Dynamic Network Actor Model (DyNAM) by Stadtfeld, Hollway, & Block (2017).

Aral, Brynjolfsson, & Pentland, 2008). Other research looks at temporal effects of (social) networks (Starnini et al., 2017; Ubaldi, Vezzani, Karsai, Perra, & Burioni, 2017) but do not necessarily use sociometric badges or focus on a specific social science research question. As mentioned, the emerging literature regarding sociometric sensors – including this report - are rather busy validating the different measurements available. Despite the scarcity of publications combining REMs and sociometric data, the following paragraphs present nevertheless a first approximation. Given the overall state of the art, this has to remain currently on a rather exploratory level. Too many questions could not be addressed within the timing and workload of the GEDII project. Nevertheless, we hope to provide fruitful first steps towards a more dynamic approach to studying group processes using sociometric face-to-face interaction networks and REM.

Methods and Data Preparation

According to Butts & Macrum (2017), the most important assumption of the REM is that events in time are well ordered. Sociometric data – both proximity as well as face-to-face interactions – are timestamped. They are thus well-ordered down to the microsecond-level. The way sociometric data is recorded also implies that actions are directed from A to B. Thus, on a purely formal level, sociometric interaction data fulfills the basic assumptions of the REM: a well ordered sequence of interaction events among dyads. However, some simplifying assumptions should nevertheless be mentioned.

First, although the sociometric badges record the direction of action, to a certain point this is arbitrary. If two persons face each other, which badge detects first the other (determining the direction of the interaction) depends on the angle in which badges are placed to each other but also the scanning frequency of each badge. Even though the badges might detect that A is the “target” of the interaction and B the “source”, these does not necessarily mirror who initiates and subsequently responds to an interaction. Other studies that work for example with radio communication do not have this problem since each “call” has an unequivocal direction from the caller to the receptor (see Butts & Macrum's 2017 World Trade Disaster example).

Second, even though the event series of sociometric badges is well ordered, the ordering is to a large degree arbitrary. During a typical interaction sequence, face-to-face detects will be registered one after the other, even though this does not necessarily imply a specific turn-taking sequence where previous events condition future events. Rather, the mutual detection seems to be conditioned by the technical capabilities and configuration of the badges than conversational dynamics and turn-taking.

Both of these issues – the somewhat artificial directionality and sequencing of events – can be accounted for by the REM, specifying in some instances the adequate model parameters. However, there are also several interesting modeling approaches offered by the REM for analyzing conversational shifts and turn-taking in particular which are not part of this initial analysis. A more thorough investigation is necessary to which degree the speech profiles, rather than the interaction sequences – would be suitable to analysis conversational turn-taking.

The REM model provides standard procedures for working with dyad based events that involve exchanges between two entities. This fits perfectly the required approach of the sociometric interaction data. However, the REM also provides for a more broader framework to model any sequences of events of any type. Model specification however comes more demanding for this second case – a research line to be explored in the future.

Finally, the REM distinguished between events that are simply ordered in time (ordinal) and event sequences that contain information on the precise timing of events (interval data). Since sociometric interaction data is timestamped, the “ordinal” parameter for the modeling was set to FALSE.

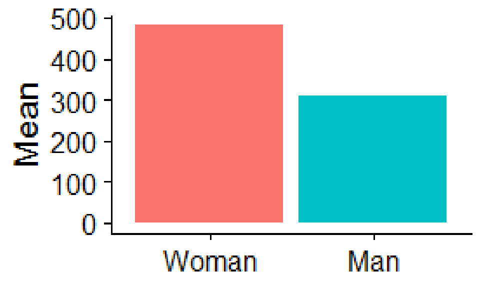
Overall, two REM models have been fitted to each of the 8 research teams using the face-to-face interaction data over the entire 5 days. External constraints such as for example night-time or absence of badges during certain days are ignored. The REM allows for event histories where only the order of event is known (“ordinal time”) and event sequences with exact timing (“exact”, “interval time”). Sociometric data provide “exact” timing.

A Simple Model with Gender

A first simple Relational Event Model is fitted to the face-to-face interaction data of each team using just one covariate namely “gender”. The reference category is set to “woman=TRUE”. As such, the model gives insights regarding the importance of gender as a variable for predicting interaction events. As can be seen from the following table, the effects of gender on this very basic level mirrors the relative participation of women and men in interactions. On an intuitive level, the REM picks up on the relative mean detects for women in relation to men: if women are over-represented, the effect is positive (T1, T3, T7 for example) and if women are underrepresented, the effects are negative (T2, T4, or T6). This can be easily seen by comparing the REM effects with the illustrations regarding the relative mean detects in the right column.

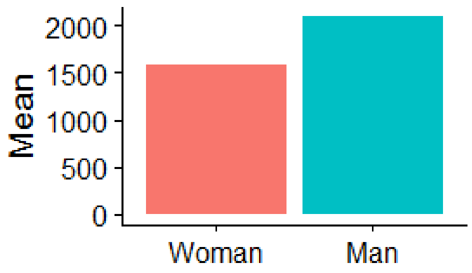
The more precise interpretation of the REM effects according to Butts & Marcum (2017) stipulates that effect estimates provide insights regarding the chances a certain event will happen given all other parameters specified in the model. The hazard (or relative rate) is calculated by taking the exponential function of the estimate: in Team 1, the estimate for “woman” is 0.28, which yields a hazard of $e^{0.28} = 1.37$. If a given event involves a women, then the next event is 1.37 times more likely to involve at least one women. Or, more precisely, events that involve a woman-man dyad have 1.37 times the hazard than events of man-man only. Team 2 on the other hand has a negative estimate of -0.34. Since gender is used as a binary variable in this case, changing the reference category to “man=TRUE” switches the sign of the estimate to 0.34 which yields a hazard of $e^{0.34}$=1.41. Thus, the propensity of events involving woman-man dyad is 1.41 times higher than events involving woman-woman dyads only for Team 2.

Team	Relational Event Model	Mean Interaction Detects by Gender

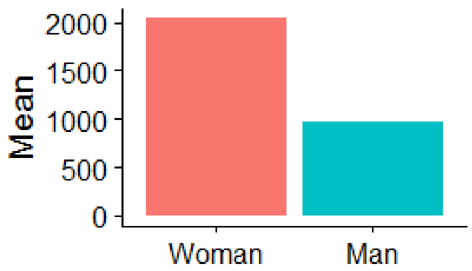
T1	<pre>Relational Event Model (Temporal Likelihood)
	
	                            Estimate Std.Err Z value Pr(>|z|)
	Intercept -4.741105 0.023558 -201.2539 < 2.2e-16 ***
	Woman 0.283008 0.040371 7.0102 2.38e-12 ***
	--- 	
	Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
	Null deviance: 30301.53 on 1482 degrees of freedom
	Residual deviance: 30252.53 on 1481 degrees of freedom
	        Chi-square: 48.99921 on 1 degrees of freedom, asymptotic p-value
	2.560618e-12
	AIC: 30256.53 AICC: 30256.54 BIC: 30267.13</pre>
		

<table>
<tr><td></td><td></td><td></td></tr>
<tr><td>T2</td><td>Relational Event Model (Temporal Likelihood)

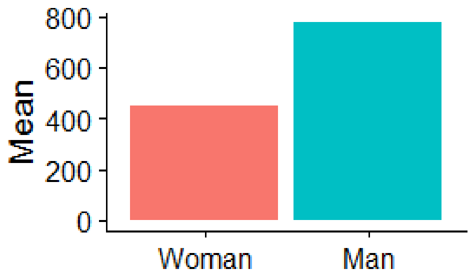
Estimate Std.Err Z value Pr(>|z|)
Intercept -4.027908 0.013966 -288.413 < 2.2e-16 ***
Woman -0.345761 0.019721 -17.533 < 2.2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null deviance: 120502.4 on 6331 degrees of freedom
Residual deviance: 120202.5 on 6330 degrees of freedom
Chi-square: 299.9104 on 1 degrees of freedom, asymptotic p-value 0
AIC: 120206.5 AICC: 120206.5 BIC: 120220</td><td>
</td></tr>
<tr><td>T3</td><td>Relational Event Model (Temporal Likelihood)

Estimate Std.Err Z value Pr(>|z|)
Intercept -4.385507 0.019940 -219.934 < 2.2e-16 ***
Woman 0.274243 0.024645 11.127 < 2.2e-16 ***

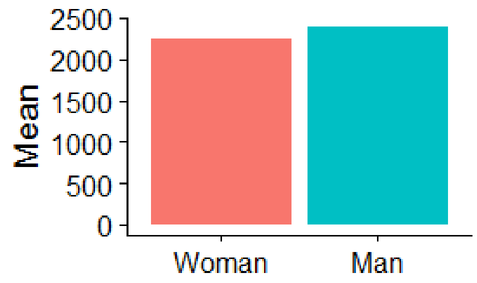
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null deviance: 98446.74 on 5255 degrees of freedom
Residual deviance: 98319.01 on 5254 degrees of freedom
Chi-square: 127.7336 on 1 degrees of freedom, asymptotic p-value 0
AIC: 98323.01 AICC: 98323.01 BIC: 98336.14</td><td>
</td></tr>
<tr><td>T4</td><td>Relational Event Model (Temporal Likelihood)

Estimate Std.Err Z value Pr(>|z|)
Intercept -4.301012 0.015779 -272.573 < 2.2e-16 ***
Woman -0.612523 0.029529 -20.743 < 2.2e-16 ***

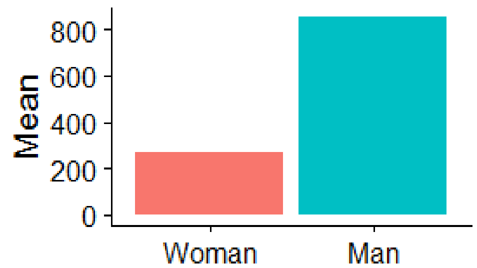
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null deviance: 54677.47 on 2680 degrees of freedom
Residual deviance: 54241.58 on 2679 degrees of freedom
Chi-square: 435.8863 on 1 degrees of freedom, asymptotic p-value 0
AIC: 54245.58 AICC: 54245.59 BIC: 54257.37</td><td>
</td></tr>
</table>

<table>
<tr><td>T5</td><td>Relational Event Model (Temporal Likelihood)

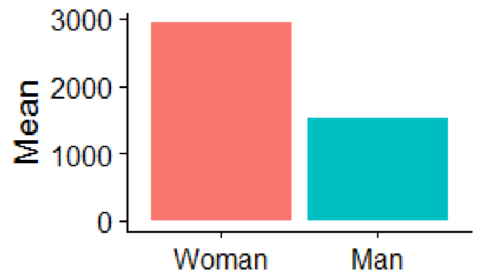
Estimate Std.Err Z value Pr(>|z|)
Intercept -4.0034283 0.0073301 -546.1593 < 2.2e-16 ***
Woman -0.0654767 0.0132827 -4.9295 8.245e-07 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null deviance: 231512.8 on 12769 degrees of freedom
Residual deviance: 231488.5 on 12768 degrees of freedom
Chi-square: 24.34831 on 1 degrees of freedom, asymptotic p-value 8.039727e-07
AIC: 231492.5 AICC: 231492.5 BIC: 231507.4</td><td></td></tr>
<tr><td>T6</td><td>Relational Event Model (Temporal Likelihood)

Estimate Std.Err Z value Pr(>|z|)
Intercept -4.315055 0.011253 -383.445 < 2.2e-16 ***
Woman -0.915502 0.043097 -21.243 < 2.2e-16 ***

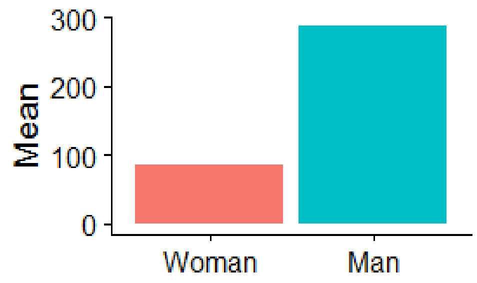
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null deviance: 51895.18 on 2603 degrees of freedom
Residual deviance: 51357.25 on 2602 degrees of freedom
Chi-square: 537.9326 on 1 degrees of freedom, asymptotic p-value 0
AIC: 51361.25 AICC: 51361.25 BIC: 51372.98</td><td></td></tr>
<tr><td>T7</td><td>Relational Event Model (Temporal Likelihood)

Estimate Std.Err Z value Pr(>|z|)
Intercept -4.744124 0.006813 -696.336 < 2.2e-16 ***
Woman 0.367410 0.011841 31.027 < 2.2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null deviance: 319885.6 on 15718 degrees of freedom
Residual deviance: 318933.2 on 15717 degrees of freedom
Chi-square: 952.4465 on 1 degrees of freedom, asymptotic p-value 0
AIC: 318937.2 AICC: 318937.2 BIC: 318952.5</td><td></td></tr>
</table>

<table>
<tr>
<td>T8</td>
<td>Relational Event Model (Temporal Likelihood)

Estimate Std.Err Z value Pr(>|z|)
Intercept -4.777758 0.017870 -267.367 < 2.2e-16 ***
Woman -1.336872 0.082398 -16.225 < 2.2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null deviance: 20805.88 on 947 degrees of freedom
Residual deviance: 20449.37 on 946 degrees of freedom
Chi-square: 356.51 on 1 degrees of freedom, asymptotic p-value 0
AIC: 20453.37 AICC: 20453.39 BIC: 20463.08</td>
<td>
</td>
</tr>
</table>

A More Complex Model: Recency, Inertia, Degree, Gender

The following tables presents a more complex Relational Event Model, now incorporating "Recency", "Inertia", and "Degree" effects in addition to the covariate of "gender". These and other endogenous effects are available with the REM model in order to account for genuine time-based patterns within the observed data.

"**NTDegSnd**" and "**NTDegRec**" estimates how the overall normalized degree of a given node affects its future sending rate or its future receiving rate respectively. Intuitively this effect captures how the "popularity" of a given node affects its sending/receiving, where "popularity" is defined as the overall number of interaction partners a given team member has within the network. The "NTDegSnd" effect indicates literally the degree to which a well connected node is more likely to form part of future interactions. "NTDegRec" on the other hand indicates how the normalized total degree of a badge affects its future receiving rate: to which degree is a "preferential attachment" responsible for future rates of interaction.

"Degree" effects cannot be modeled without taking into account the direction of interactions as "sending" or "receiving". However, the following tables only reports the estimates for "sending" even though several models incorporating both effects have been tested with the overall finding that the size of the effect is similar but inverted in direction.

"**RRecSnd**" and "**RsndSnd**" indicate both "recency" effects that estimate how the rate of a relational event occurring at a given time from A to B, is positively affected by the volume of prior instances of a relational event from B to A. "RRecSnd" signals how the recency of receipt of actions from B affects A's future rate of sending to B. "RSndSnd" on the other hand indicates how the recency of sending to B affects A future rate of sending to B. One can expect that receny effects with sociometric badges are quite strong, since face-to-face detects between badges happen with a certain frequency. Detects not only indicate that two badges are interacting with each other but are conditioned by the technical setup and configuration of the badges – which scan for each other every 25 seconds. Recency effects should be strong across all teams and not vary too much since they do happen independently of the actual sociometric interaction profile of each team. If they do vary, then this indicates their relative importance in relation to all other effect-types incorporated in the model.

"**FrSndSnd**" and "**FrRecSnd**" are two "inertia" effects, which capture routinization/habitation of interaction. Leenders et al. define inertia as "the rate of a relational event occuring at any given time from one team member, A, to another member, B, increases with the volumne of the prior instances of a relation event from A to B." (Leenders et al. 2016, p.99). "FrPSndSnd" thereby indicates how the fraction of A's past actions *directed* to B affects A's future rate of sending to B. And "FrRecSnd" indicates how the Fraction of A's past *receipt* of actions from B affects A's future rate of sending to B. Again, the direction of events cannot be ignored with "inertia" based events.

Given these time-based events, the following tables illustrate how Degree, Recency and Inertia affect the relative hazard of events in relation to the covariate gender (with “women=TRUE” again set as reference category). As can be seen, the time-based effects are much stronger than the covariate of gender in all cases. And, they considerably affect the sign and size of the estimates for gender.

Considering Team 1, the hazard that sociometric badges detect the exact same event as the previous one (recency) is multiplied by $e^{5.79} = 327$. This makes sense since the detection of events is conditioned by the technical configuration of badges which scan for other badges at determined time intervals. Thus, when team members are in a face-to-face situation, the badges will detect each other in a series of closely paced events. Even stronger is the hazard regarding the total degree of a node which produces a 437 ($e^{6.08}$) fold increase[23].

What is clearly visible is the overall affect of these time-based estimates on the covariate of gender. Except for Team 2 (and Team 1 which is 0), women have now across all research teams a decreasing propensity to be part of interaction events. All effects are negative, or, to state it the other way round: men are more likely to be part of interaction events taking into consideration the effects of Degree, Inertia and Recency. Thus, the model provides interesting insights regarding the relative importance of gender in time-based interaction dynamics, separating the temporal pattern of interaction from attribute variables such as gender.

T1

```
Relational Event Model (Temporal Likelihood)

     Estimate     Std.Err  Z value Pr(>|z|)
NTDegSnd   6.08978328  0.19409860  31.3747   <2e-16 ***
FrPSndSnd  1.57640941  0.07914538  19.9179   <2e-16 ***
RSndSnd    5.79622089  0.17244968  33.6111   <2e-16 ***
Intercept -7.59359034  0.09335416 -81.3417   <2e-16 ***
Woman      0.00013238  0.03782088   0.0035   0.9972
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null deviance: 30301.53 on 1482 degrees of freedom
Residual deviance: 24758.15 on 1478 degrees of freedom
   Chi-square: 5543.382 on 4 degrees of freedom, asymptotic p-value 0
AIC: 24768.15 AICC: 24768.19 BIC: 24794.65
```

T2

```
Relational Event Model (Temporal Likelihood)

    Estimate   Std.Err   Z value  Pr(>|z|)
NTDegSnd   6.500331  0.163363   39.7908 < 2.2e-16 ***
FrPSndSnd  0.431692  0.046635    9.2568 < 2.2e-16 ***
RSndSnd    6.373492  0.077102   82.6636 < 2.2e-16 ***
Intercept -6.959122  0.044038 -158.0253 < 2.2e-16 ***
Woman      0.190216  0.020078    9.4737 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null deviance: 120502.4 on 6331 degrees of freedom
```

23 The precise interpretation of this increase is not clear: does this 437 increase depend upon each increment of degree from 1 to the maxium degree possible in the given network?

```
Residual deviance: 94439.54 on 6327 degrees of freedom
   Chi-square: 26062.83 on 4 degrees of freedom, asymptotic p-value 0
AIC: 94449.54 AICC: 94449.55 BIC: 94483.31
```

T3

```
Relational Event Model (Temporal Likelihood)

           Estimate   Std.Err Z value  Pr(>|z|)
NTDegSnd   4.578619  0.092395  49.555 < 2.2e-16 ***
FrPSndSnd  4.170700  0.053234  78.346 < 2.2e-16 ***
RSndSnd    8.939699  0.194074  46.063 < 2.2e-16 ***
Intercept -8.261549  0.098029 -84.277 < 2.2e-16 ***
Woman     -1.610204  0.025764 -62.497 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null deviance: 98446.74 on 5255 degrees of freedom
Residual deviance: 69014.79 on 5251 degrees of freedom
        Chi-square: 29431.95 on 4 degrees of freedom, asymptotic p-
value 0
AIC: 69024.79 AICC: 69024.81 BIC: 69057.63
```

T4

```
Relational Event Model (Temporal Likelihood)

           Estimate   Std.Err   Z value  Pr(>|z|)
NTDegSnd   7.389114  0.210854   35.0438 < 2.2e-16 ***
FrPSndSnd  1.835531  0.072915   25.1735 < 2.2e-16 ***
RSndSnd    5.242027  0.100535   52.1412 < 2.2e-16 ***
Intercept -6.972734  0.055659 -125.2750 < 2.2e-16 ***
Woman     -0.262735  0.035962   -7.3058 2.756e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null deviance: 54677.47 on 2680 degrees of freedom
Residual deviance: 44497.44 on 2676 degrees of freedom
        Chi-square: 10180.03 on 4 degrees of freedom, asymptotic p-
value 0
AIC: 44507.44 AICC: 44507.46 BIC: 44536.91
```

T5

```
Relational Event Model (Temporal Likelihood)

           Estimate   Std.Err  Z value  Pr(>|z|)
NTDegSnd  11.002986  0.126391   87.055 < 2.2e-16 ***
FrPSndSnd  2.204847  0.032102   68.682 < 2.2e-16 ***
RSndSnd    6.369519  0.058002  109.815 < 2.2e-16 ***
Intercept -6.837261  0.028891 -236.659 < 2.2e-16 ***
Woman     -0.377968  0.012244  -30.869 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null deviance: 231512.8 on 12769 degrees of freedom
Residual deviance: 172298.5 on 12765 degrees of freedom
        Chi-square: 59214.37 on 4 degrees of freedom, asymptotic p-
value 0
AIC: 172308.5 AICC: 172308.5 BIC: 172345.7
```

T6

```
Relational Event Model (Temporal Likelihood)
```

	Estimate Std.Err Z value Pr(>|z|) NTDegSnd 8.199855 0.141940 57.770 < 2.2e-16 *** FrPSndSnd 2.423550 0.062414 38.830 < 2.2e-16 *** RSndSnd 8.492535 0.252750 33.600 < 2.2e-16 *** Intercept -9.256722 0.133424 -69.378 < 2.2e-16 *** Woman -0.545678 0.039688 -13.749 < 2.2e-16 *** --- Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 Null deviance: 51895.18 on 2603 degrees of freedom Residual deviance: 36861.52 on 2599 degrees of freedom Chi-square: 15033.66 on 4 degrees of freedom, asymptotic p-value 0 AIC: 36871.52 AICC: 36871.54 BIC: 36900.84
T7	Relational Event Model (Temporal Likelihood) Estimate Std.Err Z value Pr(>|z|) NTDegSnd 9.527234 0.079342 120.0786 < 2.2e-16 *** FrPSndSnd 3.566632 0.026443 134.8789 < 2.2e-16 *** RSndSnd 5.809057 0.049789 116.6730 < 2.2e-16 *** Intercept -7.653928 0.027177 -281.6280 < 2.2e-16 *** Woman -0.058731 0.014547 -4.0374 5.405e-05 *** --- Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 Null deviance: 319885.6 on 15718 degrees of freedom Residual deviance: 219766.7 on 15714 degrees of freedom Chi-square: 100118.9 on 4 degrees of freedom, asymptotic p-value 0 AIC: 219776.7 AICC: 219776.7 BIC: 219815.1
T8	Relational Event Model (Temporal Likelihood) Estimate Std.Err Z value Pr(>|z|) NTDegSnd 5.78776 0.38112 15.1864 < 2e-16 *** FrPSndSnd -0.24303 0.11722 -2.0733 0.03814 * RSndSnd 5.76342 0.19243 29.9502 < 2e-16 *** Intercept -7.31096 0.10324 -70.8172 < 2e-16 *** Woman -0.19470 0.10525 -1.8498 0.06434 . --- Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 Null deviance: 20805.88 on 947 degrees of freedom Residual deviance: 17253.86 on 943 degrees of freedom Chi-square: 3552.028 on 4 degrees of freedom, asymptotic p-value 0 AIC: 17263.86 AICC: 17263.92 BIC: 17288.12

Selecting Models

The comparative approach across several case studies is not the standard way to use the REM. Rather, the “normal” analytical process foresees to construct a complex model for each case that explains as much of the variance as possible. Model selection is based upon the reduction of the Bayesian Information Criterion (BIC) from simpler to more complex models. The following table demonstrates how the introduction of endogenous effects as well as covariates produces an increasingly better fitting model. Model fitting is applied to one particular case study, namely T1.

	Model 1			Model 2			Model 3		
	Estimate	*Std.Err*	*p*	*Estimate*	*Std.Err*	*p*	*Estimate*	*Std.Err*	*p*
(Intercept)	-4.8953	0.0280	**< .000**	-7.4262	0.0964	**< .000**	-8.2341	0.1018	**< .000**
Gender (Woman)	0.0649	0.0447	.146	-0.0548	0.0472	< .245	-0.2977	0.0554	**< .000**
PhD	0.5235	0.0445	**< .000**	0.6206	0.0486	**< .000**	0.2250	0.0557	**< .000**
RRecSnd				2.2933	0.1218	**< .000**	2.1520	0.1215	**< .000**
RSndSnd				5.0325	0.1843	**< .000**	5.4312	0.1909	**< .000**
NTDegSnd							4.7915	0.1729	**< .000**
NTDegRec							4.4464	0.1890	**< .000**
Null deviance	30301.53 on 1482 df			30301.53 on 1482 df			30301.53 on 1482 df		
Residual deviance	30252.53 on 1481			24693.22 on 1478 df			23958.31 on 1476 df		
AIC / AICC / BIC	AIC: 30256.53 AICC: 30256.54 BIC: 30267.13			AIC: 24703.22 AICC: 24703.26 BIC: 24729.73			AIC: 23972.31 AICC: 23972.39 BIC: 24009.42		

Tabla 21: REM Model Selection

Consulting Table 22 on page 122 and ANNEX II – Selected Graphs / Charts on page 106 provides insights on the basic characteristics of Team 1. Model one uses two covariates Gender and PhD since one of the most influential team members is a PhD candidate. It is plausible to assume that the PhD role could have a strong effect. This member is also a woman. As the model shows, the effects are quite small. Gender (with reference category “woman”) is not significant and the BIC is only marginally reduced. Model 2 ads recency effects since we can assume that exchanges between badges happen in a fast-paced manner given the technical scanning intervals. As can be seen, effect sizes is considerable and the second model is clearly preferred over the first according to the reduction in the BIC score. In a third model, “preferential attachment” effects are introduced, i.e. the propensity of actors to communicate with well connected others. Since T1 has a highly centralized interaction network where few individuals are responsible for most of the communication, this seems to be a plausible assumption. As can be seen, the normalized total degree for both sending as well as receiving interactions is strong, although the BIC does not indicate a much improved model fit. However, under model 3 Gender produces again a significant effect: considering all parameters of the given model, men are less likely to participate in interactions. The third model yields a deviance reduction from the null model at approx. 21%. The explorations could continue with further covariates to produce a best fitting model.

Summary and Concluding Remarks

This final section aims to synthesize the central findings of this report and hence the overall work carried out during work package 2 of the GEDII project. This is certainly not an easy task, given the complexity and wealth of data generated in the course of the case studies with 8 research teams. The exploratory incursions into the sociometric data proved extremely challenging and time consuming given the different types of non-standard data (interaction, audio and body activity). The fact that a shared research- and data reporting protocol does currently not exist for sociometric badges reduces the possibilities to draw upon similar projects by other researchers. Thus, the challenges for carrying out this research can also be taken as an overall indicator of its innovative character. New ground has been covered including basic consideration of field logistics using badges, data pre-processing and reliability as well as the qualitative and quantitative analysis of the produced data. The following points thus summarize the main contributions and insights.

Reliability of Sociometric Data

As described in great detail, we have put considerable effort in validating the data produced by sociometric badges, following similar efforts of other researchers (Chaffin et al., 2015; Kayhan et al., 2018). The measurement of Bluetooth based proximity and infrared based face-to-face interaction is highly variable and not necessarily to be identified with a “social interaction”. The RSSI signal strength of proximity detects is both indicative of the distance between badges as well as their angle towards each other, a finding that has so far not been reported in the literature. The microphone and its derived measures has to be used with great caution. Whereas the measurement of speaking time is not precise but relatively consistent in terms of the error rate, the subsequent calculation of turn-taking is utterly wrong. As demonstrated, turn-taking measurements do not match at all the experimental scripted situation, let alone measurements generated from an open field setting. This is certainly a mayor disappointment regarding the capabilities of the sociometric badges given the importance of turn-taking for gendered group dynamics. Finally, no special testing was performed using the accelerometers of the badges. Kayan et al. (2008) provide some insights in this regard. Contrary to Kayan et al. (2008), however, problems with the synchronicity of the badges internal clock did not occur during our work, neither in the experimental settings nor the synchronicity of the body activity patterns.

Comparative / Qualitative Insights across 8 Research Teams

To the best of our knowledge, the current research is the first of its kinds providing an in-depth view across 8 research teams using sociometric badges. This provides the opportunity to analyze commonalities and differences across the sociometric profiles of each team. As has been argued, research groups exhibit a reliable, global pattern regarding their collaboration pattern as well as more or less centralized steering through the leader. The interaction- as well as speech profiles establish interesting dimensions for cross-team comparison that seem to capture reasonably well differences among the team as derived from the interviews. However, there is no clear gender pattern across the participating teams. Although differences do exist in terms of the mean participation of women and men in interactions or in terms of speaking time, there is not a global, transversal trend across all groups where women (or men) would consistently be under- or over-represented. This holds for both, the interview data as well as the sociometric data. Of course, the case studies did not aim for producing statements that could be generalizable to a wider population of research groups. It might well be that a gendered pattern does emerge with a higher number of groups participating; the possibilities are certainly given by the fact that interaction as well as speech profiles capture differences between women and men.

Relational Event Model

Among the exciting new developments coming out of this work are the possibilities to analyze time-based data. As argued, this is an incipient field of research where sociometric data despite it inherently time-stamped “nature” is still absent. Thus, the present work opens up promising new ground not only to explore team dynamics in continuous time but also to do so with an explicit interest in gender aspects. Two basic models have been presented with already demonstrate some interesting findings, that is, relative changes of the importance of gender when considering endogenous effects of event sequences. However, a note of caution seems in place given the fact that expertise regarding REM is not widespread in the academic field. Thus the interpretation presented in the previous pages is temporary at best and rather a proof of concept that has to be developed in future analysis and work.

Building upon the time-based analysis of interaction sequences but also conversational turn-taking with REMs offers promising avenues of future research. One possibility should be briefly mentioned in closing: the “relevent” R package includes effects for modeling conversational turn-taking[24]. Assuming that the precision of data collection has been resolved, this is especially interesting for assessing the effects of gender status on turn-taking in the context of expectation states theory (Ridgeway, 1992; Ridgeway & Smith-Lovin, 1999). As Gibson (2003) argues, turn-taking is not only influenced by status but also conversational rules. That is, when people participate in conversations is not only determined by differences in status but by conversational norms, so-called “participation-shifts”. Expectation states theory has no formal way to differentiate between these two. “A consequence is that the effects hypothesized by expectation states researchers may sometimes be confounded, at least over short periods, and perhaps over longer ones as well. If, for instance, a high-status individual is expected to speak frequently, he or she may not be able to do so over some interval if the turn-taking rules prevent it – say, by fueling a sustained dyadic exchange involving two other individuals” (Gibson, 2003, p. 1339). A more thorough examination of the relative importance of status vs. conversational norms for speaker dominance is within reach and certainly an exciting opportunity given the Relational Event Modeling framework with its default incorporation of Gibson's p-shifts model.

Gender and “Honest” Signals

Part of the attraction of using sociometric badges in research is the potential “access” to

24 Altogether 13 conversational shifts can be modeled, including simple turn receiving or turn-interruptions according to the model provided by Gibson (2003).

“honest” signals, i.e. a level of non-verbal communication that although- or rather because it happens in a semi-automatic fashion heavily influences the process and outcomes of human interaction. As already mentioned during the introduction as well as D1.1. Conceptual Framework, previous gender research has documented consistently gender differences – often tied to power and status differentials – in non-verbal communication. Using sociometric badges appears as especially promising since it could provide access to these hard-to detect, implicit bias which conditions behavior during interactions.

So far, the evidence regarding gender differences of “honest” signals is scarce at best. As mentioned, members of the original Sociometric development team at MIT and later on Humanyze have published short online pieces at Bloomberg and HBR specifically addressing gender issues (Turban et al., 2017; Waber, 2014). The general tenor of this non-scientific articles is, that no gender differences could be detected in the sociometric data – suggesting that women and men do not behave differently and that factual differences in outcome (e.g slower careers) are in the “eye of the beholder”. In short: because similar behavior is interpreted differently according to the existing stereotypes of what constitutes adequate female or male behavior, inequalities do exist. Although there is nothing wrong with this insight – which has its parallels in other research on the importance of perceptions regarding gender differences and similarities[25] – the overall affirmation regarding the non-existence of gender differences in sociometric profiles should be treated with great caution. In fact, if some dimensions of the sociometric data is as unreliable as described in this report, the absence of any clear pattern should not be surprising. This concerns especially turn-taking patterns! The absence of any clear behavioral gender pattern could as well be the product of imprecise measurements rather than a reflection of reality.

There is certainly a need to conduct more tightly controlled experiments with sociometric badges in order to get a better grasp of the reliability of its measurements especially regarding body mirroring and activity. As described, on a very basic level of analysis no gender differences could be found for body mirroring values. This is in contrast to the literature on “social sensitivity” which suggests gender and sex differences in this regard (Derntl et al., 2010; Donges, Kersting, & Suslow, 2012; Grant & Berry, 2011; Koenig & Eagly, 2005; Schulte-Rüther, Markowitsch, Shah, Fink, & Piefke, 2008; Woolley et al., 2010). Exploring further the possibilities of sociometric badges for “social sensitivity” measurements would be a plausible next task.

Future Lines of Analysis

A final challenge concerns the limits of standard social sciences methods for analyzing the volume and type of data generated by sociometric badges. Standard statistical methods such as regression modeling are not suited for taking advantage of the richness of this type of data. In this context, Machine Learning approaches such as Support Vector Machines or Neural Networks could be explored in their utility to classify correctly types of (speech) behavior. Within the limits of the current project, this was not possible. However, existing first inroads in using machine learning for improving speech profiles including turn-taking are

25 A good example in this respect is an article published by Leslie et al. (2015) on the “expectations of brilliance” that explain the under-representation of women in certain STEM subject areas.

encouraging (Kayhan et al., 2018, p. 65). While these machine learning classifiers were trained on a data set produced by a controlled experimental setting, further work is necessary to assess how these models would improve the accuracy of speech- and turn-taking analysis based on other datasets collected in open field settings.

References

Alshamsi, A., Pianesi, F., Lepri, B., Pentland, A., & Rahwan, I. (2016). Network Diversity and Affect Dynamics: The Role of Personality Traits. *PLOS ONE*, *11*(4), e0152358. http://doi.org/10.1371/journal.pone.0152358

Arrow, H., McGrath, J. E., & Berdahl, J. L. (2000). *Small groups as complex systems : formation, coordination, development and adaptation*. Thousand Oaks Calif.: Sage Publications.

Atzmueller, M., Ernst, A., Krebs, F., Scholz, C., & Stumme, G. (2014). Formation and Temporal Evolution of Social Groups During Coffee Breaks. In *Postproceedings of the International Workshops MUSE & SenseML*. Berlin, Heidelberg: Springer Berlin / Heidelberg.

Butts, C. T. (2008). A Relational Event Framework for Social Action. *Sociological Methodology*, *38*(1), 155–200. http://doi.org/10.1111/j.1467-9531.2008.00203.x

Butts, C. T., & Marcum, C. S. (2017). A Relational Event Approach to Modeling Behavioral Dynamics. In A. Pilny & M. S. Poole (Eds.), *Group Processes. Data-Driven Computational Approaches* (pp. 51–92). Cham: Springer International Publishing. http://doi.org/10.1007/978-3-319-48941-4_4

Cappella, J. N. (1985). Controlling the Floor in Conversations. In A. W. Siegman & S. Feldstein (Eds.), *Multichannel Integrations of Nonverbal Behavior* (pp. 69–104). New York: Lawrence Erlbaum Associates, Inc.

Cattuto, C., Van den Broeck, W., Barrat, A., Colizza, V., Pinton, J.-F., & Vespignani, A. (2010). Dynamics of person-to-person interactions from distributed RFID sensor networks. *PloS One*, *5*(7), e11596. http://doi.org/10.1371/journal.pone.0011596

Chaffin, D., Heidl, R., Hollenbeck, J. R., Howe, M., Yu, A., Voorhees, C., & Calantone, R. (2015). The Promise and Perils of Wearable Sensors in Organizational Research. *Organizational Research Methods*, 1094428115617004-. http://doi.org/10.1177/1094428115617004

Chen, H.-E., & Miller, S. R. (2017). Can Wearable Sensors be Used to Capture Engineering Design Team Interactions? An Investigation into the Reliability of Wearable Sensors. In *Proceedings of the ASME 2017 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*.

Cornwell, B. (2015). *Social Sequence Analysis. Methods and Applications*. New York: Oxford University Press.

Cronin, M. A., Weingart, L. R., & Todorova, G. (2011). Dynamics in Groups: Are We There Yet?

The Academy of Management Annals, *5*(1), 571–612. http://doi.org/10.1080/19416520.2011.590297

Derntl, B., Finkelmeyer, A., Eickhoff, S., Kellermann, T., Falkenberg, D. I., Schneider, F., & Habel, U. (2010). Multidimensional assessment of empathic abilities: neural correlates and gender differences. *Psychoneuroendocrinology*, *35*(1), 67–82. http://doi.org/10.1016/j.psyneuen.2009.10.006

Do, T. M. T., Kalimeri, K., Lepri, B., Pianesi, F., & Gatica-Perez, D. (2013). Inferring social activities with mobile sensor networks. In *Proceedings of the 15th ACM on International conference on multimodal interaction - ICMI '13* (pp. 405–412). New York, New York, USA: ACM Press. http://doi.org/10.1145/2522848.2522894

Donges, U.-S., Kersting, A., & Suslow, T. (2012). Women's greater ability to perceive happy facial emotion automatically: gender differences in affective priming. *PloS One*, *7*(7), e41745. http://doi.org/10.1371/journal.pone.0041745

DuBois, C., Butts, C. T., McFarland, D., & Smyth, P. (2013). Hierarchical models for relational event sequences. *Journal of Mathematical Psychology*, *57*(6), 297–309. http://doi.org/10.1016/j.jmp.2013.04.001

Ellyson, S. L., Dovidio, J. F., & Brown, C. E. (1992). The Look of Power: Gender Differences and Similarities in Visual Dominance Behavior. In C. L. Ridgeway (Ed.), *Gender, Interaction, and Inequality* (pp. 50–80). New York: Springer.

Fu, E. Y., Leong, H. V., Ngai, G., & Chan, S. (2015). Automatic Fight Detection Based on Motion Analysis. In *2015 IEEE International Symposium on Multimedia (ISM)* (pp. 57–60). IEEE. http://doi.org/10.1109/ISM.2015.98

George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big Data and Data Science Methods for Management Research. *Academy of Management Journal*, *59*(5), 1493–1507. http://doi.org/10.5465/amj.2016.4005

Gibson, D. R. (2003). Participation Shifts: Order and Differentiation in Group Conversation. *Social Forces*, *81*(4), 1335–1380. http://doi.org/10.1353/sof.2003.0055

Gloor, P. A., Fischbach, K., Fuehres, H., Lassenius, C., Niinimäki, T., Olguín-Olguín, D., … Putzke, J. (2011). Towards "Honest Signals" of Creativity – Identifying Personality Characteristics Through Microscopic Social Network Analysis. *Procedia - Social and Behavioral Sciences*, *26*, 166–179. http://doi.org/10.1016/j.sbspro.2011.10.573

Gloor, P. A., Grippa, F., Putzke, J., Lassenius, C., Fuehres, H., Fischbach, K., & Schoder, D. (2012). Measuring Social Capital in Creative Teams Through Sociometric Sensors.

International Journal of Organisational Design and Engineering, *2*(4), 380–401. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.404.8130

Grant, A. M., & Berry, J. W. (2011). The Necessity of Others is The Mother of Invention: Intrinsic and Prosocial Motivations, Perspective Taking, and Creativity. *Academy of Management Journal*, *54*(1), 73–96. http://doi.org/10.5465/AMJ.2011.59215085

Hall, J. A., Coats, E. J., & Smith LeBeau, L. (2005). Nonverbal behavior and the vertical dimension of social relations: a meta-analysis. *Psychological Bulletin*, *131*(6), 898–924. http://doi.org/10.1037/0033-2909.131.6.898

Humphrey, S. E., & Aime, F. (2014). Team Microdynamics: Toward an Organizing Approach to Teamwork. *The Academy of Management Annals*, *8*(1), 443–503. http://doi.org/10.1080/19416520.2014.904140

Hung, H., Englebienne, G., & Kools, J. (2013). Classifying social actions with a single accelerometer. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing - UbiComp '13* (p. 207). New York, New York, USA: ACM Press. http://doi.org/10.1145/2493432.2493513

Ilgen, D. R., Hollenbeck, J. R., Johnson, M., & Jundt, D. (2005). Teams in Organizations: From Input-Process-Output Models to IMOI Models. *Annual Review of Psychology*, *56*(1), 517–543. http://doi.org/10.1146/annurev.psych.56.091103.070250

Jayagopi, D. B., Hung, H., Yeo, C., & Gatica-Perez, D. (2009). Modeling Dominance in Group Conversations Using Nonverbal Activity Cues. *IEEE Transactions on Audio, Speech, and Language Processing*, *17*(3), 501–513. http://doi.org/10.1109/TASL.2008.2008238

Kayhan, V. O., Chen, Z., French, K. A., Allen, T. D., Salomon, K., & Watkins, A. (2018). How honest are the signals? A protocol for validating wearable sensors. *Behavior Research Methods*, 1–27. http://doi.org/10.3758/s13428-017-1005-4

Kennedy, L. S., & Ellis, D. P. W. (2003). Pitch-based emphasis detection for characterization of meeting recordings. In *2003 IEEE Workshop on Automatic Speech Recognition and Understanding* (pp. 243–248). IEEE. http://doi.org/10.1109/ASRU.2003.1318448

Koenig, A. M., & Eagly, A. H. (2005). Stereotype Threat in Men on a Test of Social Sensitivity. *Sex Roles*, *52*(7–8), 489–496. http://doi.org/10.1007/s11199-005-3714-x

Kozlowski, S. W. J. (2015). Advancing research on team process dynamics: Theoretical, methodological, and measurement considerations. *Organizational Psychology Review*, *5*(4), 270–299. http://doi.org/10.1177/2041386614533586

Lee, D. H., Seo, I. W., Choe, H. C., & Kim, H. D. (2012). Collaboration network patterns and research performance: the case of Korean public research institutions. *Scientometrics*, *91*(3), 925–942. http://doi.org/10.1007/s11192-011-0602-8

Leenders, R. T. A. J., Contractor, N. S., & DeChurch, L. A. (2016). Once upon a time. Understanding team processes as relational event networks. *Organizational Psychology Review*, *6*(1), 92–115. http://doi.org/10.1177/2041386615578312

Lepri, B., Staiano, J., Rigato, G., Kalimeri, K., Finnerty, A., Pianesi, F., … Pentland, A. (2012). The SocioMetric Badges Corpus: A Multilevel Behavioral Dataset for Social Behavior in Complex Organizations. In *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing* (pp. 623–628). IEEE. http://doi.org/10.1109/SocialCom-PASSAT.2012.71

Lepri, B., Staiano, J., Shmueli, E., Pianesi, F., & Pentland, A. (2016). The role of personality in shaping social networks and mediating behavioral change. *User Modeling and User-Adapted Interaction*, *26*(2–3), 143–175. http://doi.org/10.1007/s11257-016-9173-y

Marcum, C. S., & Butts, C. T. (2015). Constructing and Modifying Sequence Statistics for relevent Using informR in R. *Journal of Statistical Software*, *64*(5), 1–36. http://doi.org/10.18637/jss.v064.i05

Mast, M. S. (2001). Gender Differences and Similarities in Dominance Hierarchies in Same-Gender Groups Based on Speaking Time. *Sex Roles*, *44*(9–10), 537–556. http://doi.org/10.1023/A:1012239024732

Mast, M. S. (2002). Dominance as Expressed and Inferred Through Speaking Time. *Human Communication Research*, *28*(3), 420–450. http://doi.org/10.1111/j.1468-2958.2002.tb00814.x

Mast, M. S., & Cousin, G. (2013). Power, Dominance and Persuasion. In J. A. Hall & M. L. Knapp (Eds.), *Nonverbal Communication.* (pp. 613–635). Berlin, Boston: Walter de Gruyter.

Onnela, J.-P., Waber, B. N., Pentland, A., Schnorf, S., & Lazer, D. (2014). Using sociometers to quantify social interaction patterns. *Scientific Reports*, *4*. http://doi.org/10.1038/srep05604

Palmer, M. T. (1989). Controlling conversations: Turns, topics and interpersonal control. *Communication Monographs*, *56*(1), 1–18. http://doi.org/10.1080/03637758909390246

Pentland, A. (2007). Social Signal Processing. *IEEE Signal Processing Magazine*, *24*(4), 108–111. http://doi.org/10.1109/MSP.2007.4286569

Pentland, A. (2008). *Honest signals how they shape our world*. Cambridge, Mass. : MIT Press,.

Pentland, B. T., Pentland, A. P., & Calantone, R. J. (2017). Bracketing off the actors: Towards an action-centric research agenda. *Information and Organization*, *27*(3), 137–143. http://doi.org/10.1016/j.infoandorg.2017.06.001

Pilny, A., Schecter, A., Poole, M. S., & Contractor, N. (2016). An illustration of the relational event model to analyze group interaction processes. *Group Dynamics: Theory, Research, and Practice*, *20*(3), 181–195. http://doi.org/10.1037/gdn0000042

Quintane, E., Conaldi, G., Tonellato, M., & Lomi, A. (2014). Modeling Relational Events: A Case Study of an Open Source Software Project. *Organizational Research Methods*, *17*(1), 23–50. http://doi.org/10.1177/1094428113517007

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*(1), 203–212. http://doi.org/10.1016/j.jrp.2006.02.001

Ridgeway, C. L. (Ed.). (1992). *Gender, Interaction, and Inequality*. New York: Springer Science+Business Media.

Ridgeway, C. L. (2007). Gender as a Group Process: Implications for the Persistence of Inequality. In S. J. Correll (Ed.), *Social Psychology of Gender* (Vol. 24, pp. 311–333). Bingley: Emerald Group Publishing Limited.

Ridgeway, C. L., & Smith-Lovin, L. (1999). The Gender System and Interaction. *Annual Review of Sociology*, *25*(1), 191–216. http://doi.org/10.1146/annurev.soc.25.1.191

Roe, R. A., Gockel, C., & Meyer, B. (2012). Time and change in teams: Where we are and where we are moving. *European Journal of Work and Organizational Psychology*, *21*(5), 629–656. http://doi.org/10.1080/1359432X.2012.729821

Sanchez-Cortes, D., Aran, O., Jayagopi, D. B., Schmid Mast, M., & Gatica-Perez, D. (2013). Emergent leaders through looking and speaking: from audio-visual data to multimodal recognition. *Journal on Multimodal User Interfaces*, *7*(1–2), 39–53. http://doi.org/10.1007/s12193-012-0101-0

Sapru, A., & Bourlard, H. (2015). Automatic Recognition of Emergent Social Roles in Small Group Interactions. *IEEE Transactions on Multimedia*, *17*(5), 746–760. http://doi.org/10.1109/TMM.2015.2408437

Schecter, A., & Contractor, N. (2017). Understanding and Assessing Collaborative Processes Through Relational Events. In A. A. von Davier, M. Zhu, & P. C. Kyllonen (Eds.), *Innovative Assessment of Collaboration* (pp. 223–231). Springer International Publishing. http://doi.org/10.1007/978-3-319-33261-1_14

Schecter, A., Pilny, A., Leung, A., Poole, M. S., & Contractor, N. (2017). Step by step: Capturing the dynamics of work team process through relational event sequences. *Journal of Organizational Behavior*. http://doi.org/10.1002/job.2247

Schulte-Rüther, M., Markowitsch, H. J., Shah, N. J., Fink, G. R., & Piefke, M. (2008). Gender differences in brain networks supporting empathy. *NeuroImage*, *42*(1), 393–403. http://doi.org/10.1016/j.neuroimage.2008.04.180

Sociometric Solutions. (2014). Preliminary User Guide version 1.21. Boston, MA: Sociometric solutions.

Stadtfeld, C., Hollway, J., & Block, P. (2017). Dynamic Network Actor Models: Investigating Coordination Ties through Time. *Sociological Methodology*, 8117501770929. http://doi.org/10.1177/0081175017709295

Starnini, M., Lepri, B., Baronchelli, A., Barrat, A., Cattuto, C., & Pastor-Satorras, R. (2017). Robust Modeling of Human Contact Networks Across Different Scales and Proximity-Sensing Techniques (pp. 536–551). Springer, Cham. http://doi.org/10.1007/978-3-319-67217-5_32

Stehlé, J., Charbonnier, F., Picard, T., Cattuto, C., & Barrat, A. (2013). Gender homophily from spatial behavior in a primary school: A sociometric study. *Social Networks*, *35*(4), 604–613. http://doi.org/10.1016/j.socnet.2013.08.003

Tonidandel, S., King, E. B., & Cortina, J. M. (2016). Big Data Methods: Leveraging Modern Data Analytic Techniques to Build Organizational Science. *Organizational Research Methods*. http://doi.org/10.1177/1094428116677299

Tranmer, M., Marcum, C. S., Morton, F. B., Croft, D. P., & de Kort, S. R. (2015). Using the relational event model (REM) to investigate the temporal dynamics of animal social networks. *Animal Behaviour*, *101*, 99–105. http://doi.org/10.1016/j.anbehav.2014.12.005

Tripathi, P., & Burleson, W. (2012). Predicting creativity in the wild: experience sample and sociometric modeling of teams. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work - CSCW '12* (p. 1203). New York, New York, USA: ACM Press. http://doi.org/10.1145/2145204.2145386

Turban, S., Freeman, L., & Waber, B. (2017, October). A Study Used Sensors to Show That Men and Women Are Treated Differently at Work. *Harvard Business Review*. Retrieved from https://hbr.org/2017/10/a-study-used-sensors-to-show-that-men-and-women-are-treated-differently-at-work

Ubaldi, E., Vezzani, A., Karsai, M., Perra, N., & Burioni, R. (2017). Burstiness and tie activation strategies in time-varying social networks. *Scientific Reports*, *7*. http://doi.org/10.1038/srep46225

Varni, G., Camurri, A., Coletta, P., & Volpe, G. (2009). Toward a Real-Time Automated Measure of Empathy and Dominance. In *2009 International Conference on Computational Science and Engineering* (pp. 843–848). IEEE. http://doi.org/10.1109/CSE.2009.230

Vinciarelli, A., Pantic, M., & Bourlard, H. (2009). Social signal processing: Survey of an emerging domain. *Image and Vision Computing*, *27*(12), 1743–1759. http://doi.org/10.1016/j.imavis.2008.11.007

Vu, D., Pattison, P., & Robins, G. (2015). Relational event models for social learning in MOOCs. *Social Networks*, *43*, 121–135. http://doi.org/10.1016/j.socnet.2015.05.001

Waber, B. (2014, January 31). Gender Inequality in the Workplace: What Data Analytics Says - Bloomberg. *Bloombert*. Retrieved from https://www.bloomberg.com/news/articles/2014-01-30/gender-inequality-in-the-workplace-what-data-analytics-says

Wang, W., Precoda, K., Hadsell, R., Kira, Z., Richey, C., & Jiva, G. (2012). Detecting leadership and cohesion in spoken interactions. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 5105–5108). IEEE. http://doi.org/10.1109/ICASSP.2012.6289069

Wickham, H., & Grolemund, G. (2017). *R for data science : import, tidy, transform, visualize, and model data*. Sebastopol, CA: O'Reilly Media.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a Collective Intelligence Factor in the Performance of Human Groups. *Science*, *330*(October), 686–688. Retrieved from http://science.sciencemag.org/content/early/2010/10/01/science.1193147.abstract

Wu, L., Waber, B. N., Aral, S., Brynjolfsson, E., & Pentland, A. (2008). Mining Face-to-Face Interaction Networks using Sociometric Badges: Predicting Productivity in an IT Configuration Task. *SSRN Electronic Journal*. http://doi.org/10.2139/ssrn.1130251

Yin, R. (1994). *Case study research : design and methods*. Thousand Oaks: Sage Publications.

Yu, D., Blocker, R. C., Sir, M. Y., Hallbeck, M. S., Hellmich, T. R., Cohen, T., … Pasupathy, K. S. (2016). Intelligent Emergency Department: Validation of Sociometers to Study Workload. *Journal of Medical Systems*, *40*(3), 53. http://doi.org/10.1007/s10916-015-0405-1

ANNEX I – Comparative Case Study Table

	Leadership	Team Collaboration	Role based Hierarchy	Gender Differences
T1	Speech: highest speaking, highest listening Body movement: 4 synchronized activities. F2F-Stats: Eigen, Degree, Shannon together with 2 others. Interviews: dual leadership.	Diverse profile. Eigen sd: .25 Degree sd: 1.46 3 influential individuals with similar high Eigen, Degree + Shannon network statistics. Shared lab: .25 Mean F2F: 427	Highest F2F- Eigen: Senior, PhD, Leader Speaking: Leader, Senior, PhD Horizontal interaction across roles, largely conditioned by three influential members.	Mean F2F: 481 (women), 385 (men) Mean speaking: 2.98 (women), 2.18 (men) Mean listening: 1.67 (women), 0.74 (men) Mean G. Stereotype: 2.036
T2	Speech: highest speaking, lowest listening Body movement: 3 synchronized activities. F2F-Stats: Eigen (.92), Degree (8), High Shannon together with 7 others. Interviews: Participatory / caring leadership	Uniform profile. Eigen sd: .09 Degree sd: 1.05 Shared high Eigen, Degree + Shannon network statistics for 7 out of 10 members, including leader. Shared lab: .9 Mean F2F: 1735	Highest F2F-Eigen: Senior + Postdoc + PhD + PhD, Leader, Senior, PhD. Speaking: Leader, Senior, PhD, Postdoc Horizontal among Senior, Postdoc, PhD. Leader limited to Senior + Postdocs.	Mean F2F: 1585 (women), 2085 (men) Mean speaking: 3.4 (women), 1.4 (men) Mean listening: 4.43 (women), 4.37 (men) Mean G. Stereotype: 3.000
T3	Speech: medium speaking, medium listening Body movement: 2 synchronized activities. F2F-Stats: Eigen (.88), Degree (5). Low Shannon. Interviews: Constrained leadership.	Diverse profile. Eigen sd: 0.32. Degree sd: 1.90 3 influential individuals with similar high Eigen, Degree. Most interactions between Leader + PhD and PhD + MA. Shared lab: .25 Mean F2F: 1870	Highest F2F-Eigen: Postdoc + PhD, Leader + Senior. Speaking: Admin, Postdoc, Leader Individualized between Leader + PhD; PhD + MA. Other roles absent.	Mean F2F: 2040 (women), 1446 (men) Mean speaking: 3.09 (women), 3.61 (men) Mean listening: 2.92 (women), 125 (men) Mean G. Stereotype: 2.844

	Leadership	Team Collaboration	Role based Hierarchy	Gender Differences
T4	Speech: Low speaking and listening duration Body movement: 1 synchronized activity. F2F-Stats: Eigen (.55), Degree (4). Second highest Gini-C concentration Interviews: Legacy leadership. Absence of leader.	Diverse profile. Eigen sd: 0.17. Degree sd: 1.58 for PhD/MA. Most frequent: PhD+PhD, PhD + MA. Shared lab: 1.0 Mean F2F: 569	Highest F2F-Eigen: Senior + PhD, MA + PhD, PhD. Speaking: MA, PhD, Senior Delegation from Leader to Senior, but also Leader to PhD and Senior to PhD. No further cascading.	Mean F2F: 453 (women), 774 (men) Mean speaking: 3.3 (women), 3.97 (men) Mean listening: 3.99 (women), 4.91 (men) Mean G. Stereotype: 2.50
T5	Speech: Highest speaking; Medium listening. Body movement: 7 synchronized activities. F2F-Stats: Eigen (1), Degree (10). High Shannon. Interviews: Professional leadership; strong steering.	Uniform profile: Eigen sd: .05. Degree sd: 0.67 Most frequent: PhD+PhD, PhD+Postdoc. Shared lab: .72 Mean F2F: 2322	Highest F2F-Eigen: 8 equally highest across all roles. Speaking: Leader, Postdoc, PhD. Pyramidal, cascading delegation: from leader to Postdocs to PhDs/MAs. Consistent in direction and frequency.	Mean F2F: 2247 (women), 2384 (men) Mean speaking: 2.6 (women), 3.38 (men) Mean listening: 7.04 (women) 6.28 (men) Mean G. Stereotype: 2.675
T6	Speech: Lowest speaking, Medium Listening. Body movement: 1 synchronized activities. F2F-Stats: Eigen (.42), Degree (3). Lowest Shannon. Interviews: Engaged, transformational leadership.	Diverse profiles: Eigen sd: 0.35 Degree sd: 1.61 Most frequent: Postdoc + Leader, Postdoc + Postdoc. Shared lab: .2 Mean F2F: 845	Highest F2F-Eigen: Postdoc, Senior, PhD Speaking: Assistant, Postdoc, Senior Individualized: Postdoc + Leader. No delegation.	Mean F2F: 404 (women), 992 (men). Mean speaking: 2.62 (women), 2.28 (men) Mean listening: 1.09 (women), 0.98 (men) Mean G. Stereotype: 3.031

	Leadership	Team Collaboration	Role based Hierarchy	Gender Differences
T7	Speech: Low speaking, Medium Listening Body movement: 5 / 2 synchronized activities F2F-Stats: Eigen (.97), Degree (9), High Shannon (1.25 vs. max: 1.53) Interviews:	Diverse profiles: Eigen sd: 0.32 Degree sd: 2.9 (over two weeks!) Most frequent: Senior-Senior, Senior-Admin, Postdoc-Leader Shared lab: .68 Mean F2F: 2418	Highest F2F-Eigen: Admin, Leader, Senior Speaking: Senior, Postdoc, Admin Horizontal interaction among most roles.	Mean F2F: 2933 (women), 2096 (men) Mean speaking: 2.09 (women), 2.72 (men) Mean listening: 6.27 (women), 7.46 (men) Mean G. Stereotype: 2.732
T8	Speech: No leader, low speaking for senior Body movement: 2 synchronized activities F2F-Stats: No leader, Senior, Postdoc Interviews: -	Diverse profiles: Eigen sd: 0.21 Degree sd: 1.48 Most frequent: Senior + Admin, Postdoc + Senior and among Seniors. Shared lab: 1.0 Mean F2F: 237	Highest F2F-Eigen: Senior, Postdoc, Senior Speaking: Admin, Postdoc, Senior Horizontal interactions between most roles.	Mean F2F: 86 (Women) 287 (Men). Mean speaking: 2.29 (women), 1.79 (men) Mean listening: 7.58 (women) 6.68 (men) Mean G. Stereotype: 2.929

ANNEX II – Selected Graphs / Charts

Team 1

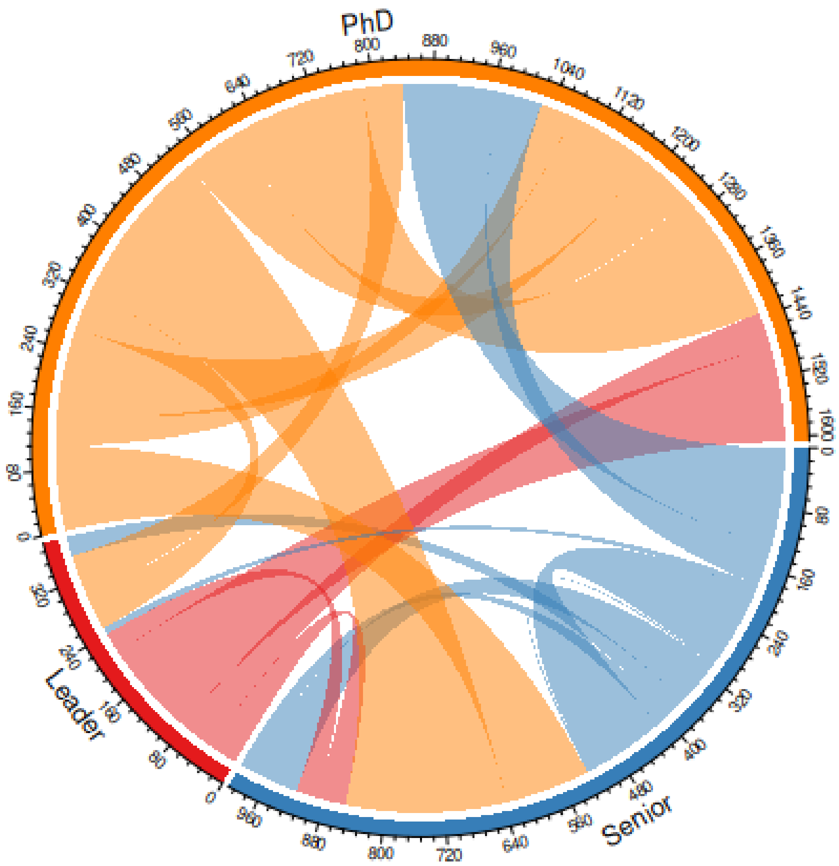


Illustration 41: Chord diagram of face-to-face interactions by role for team 1

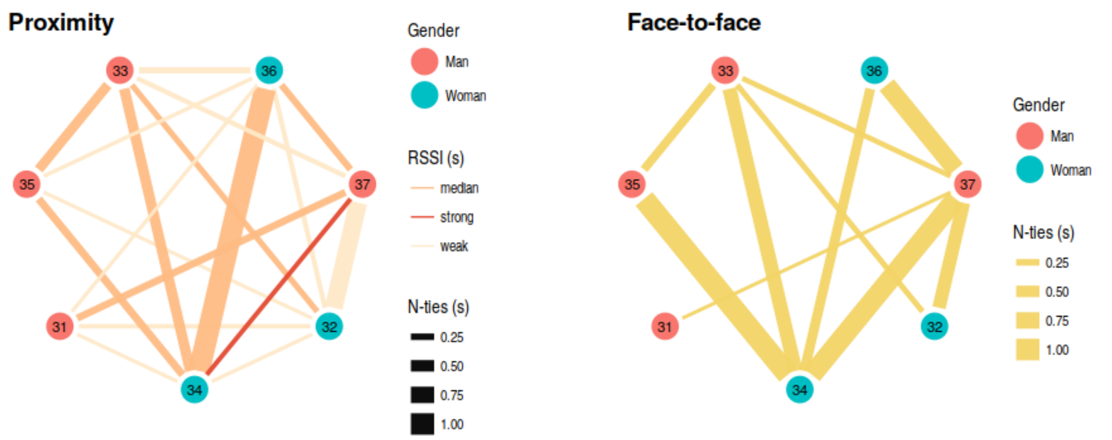


Illustration 42: Network graph of aggregated proximity and face-to-face interactions T1

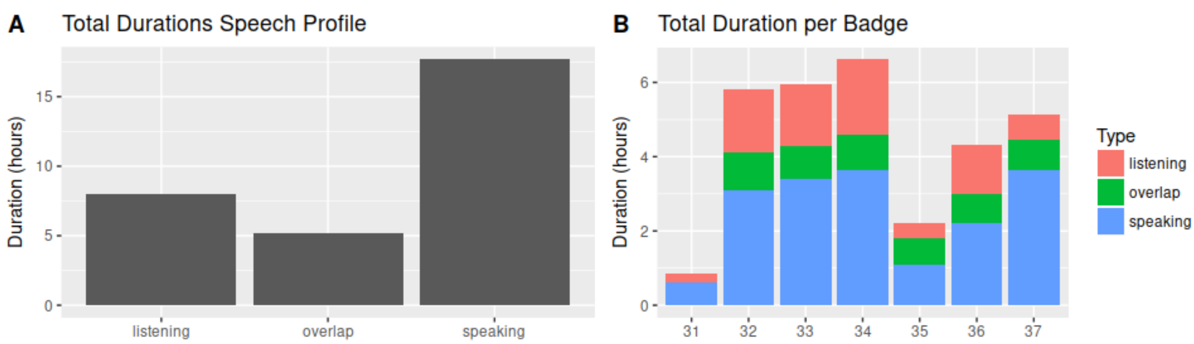


Illustration 43: Speech profile for T1 by speech type and badges

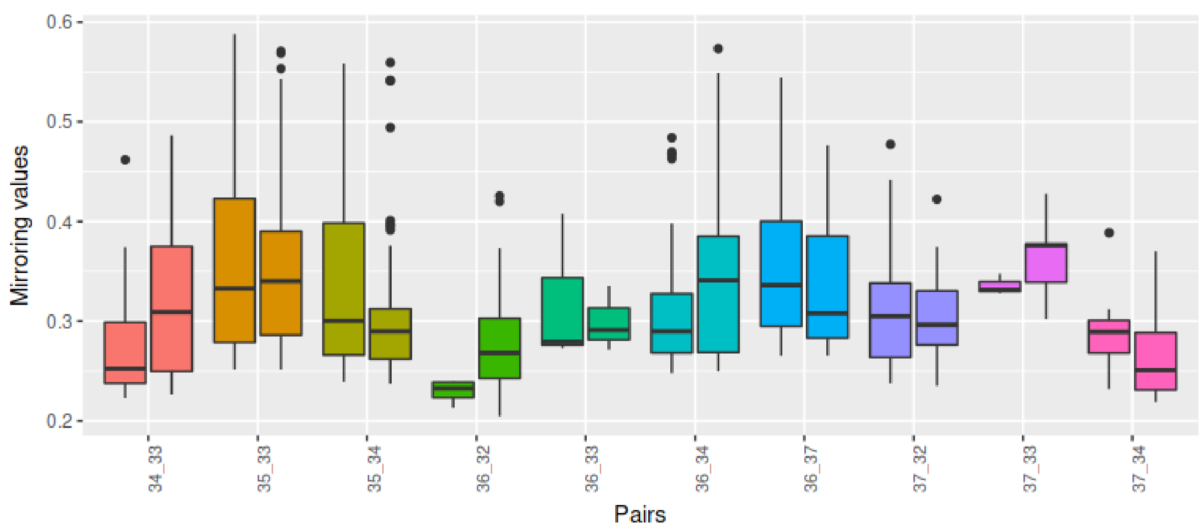


*Illustration 44: Nonverbal body language mirroring values > mean + 2 * sd*

Team 2

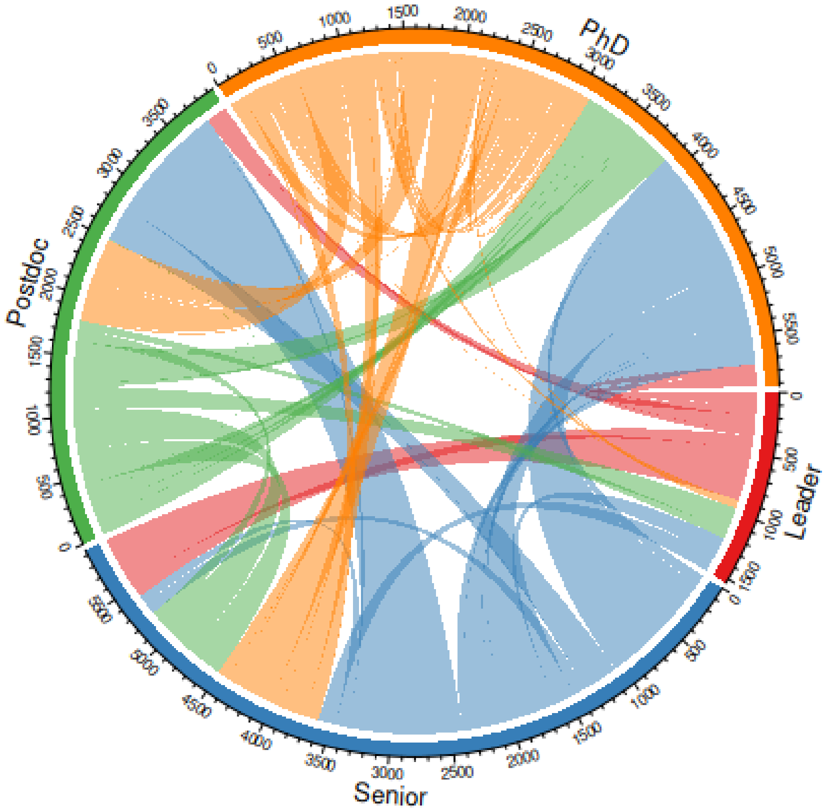


Illustration 45: Chord diagram face-to-face interactions by role Team 2

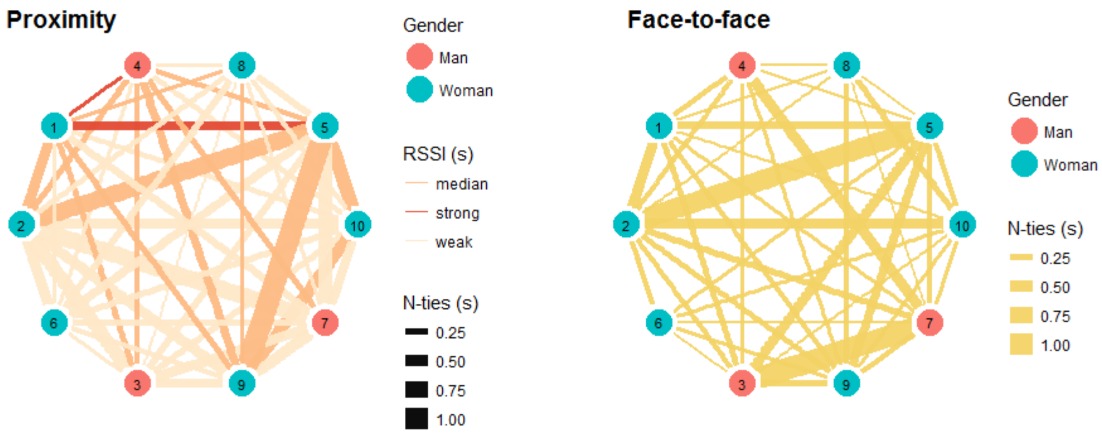


Illustration 46: Network graph of aggregated proximity and face-to-face interactions T2

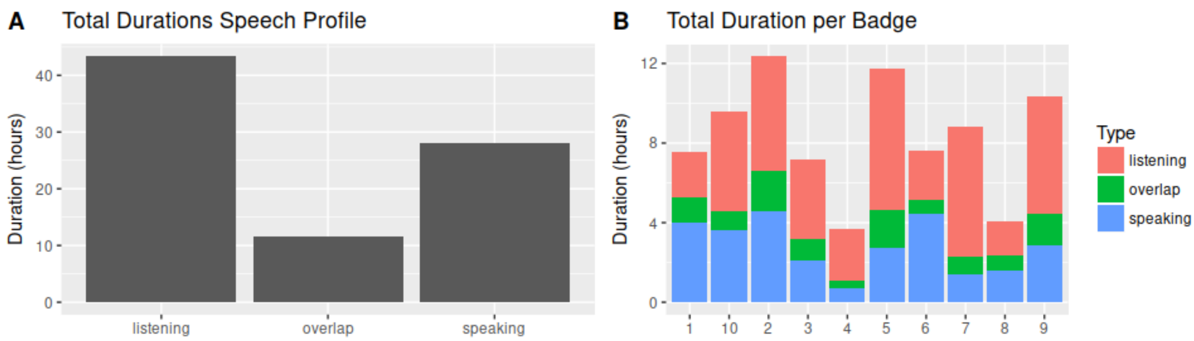


Illustration 47: Speech profile for T2 by speech type and badges

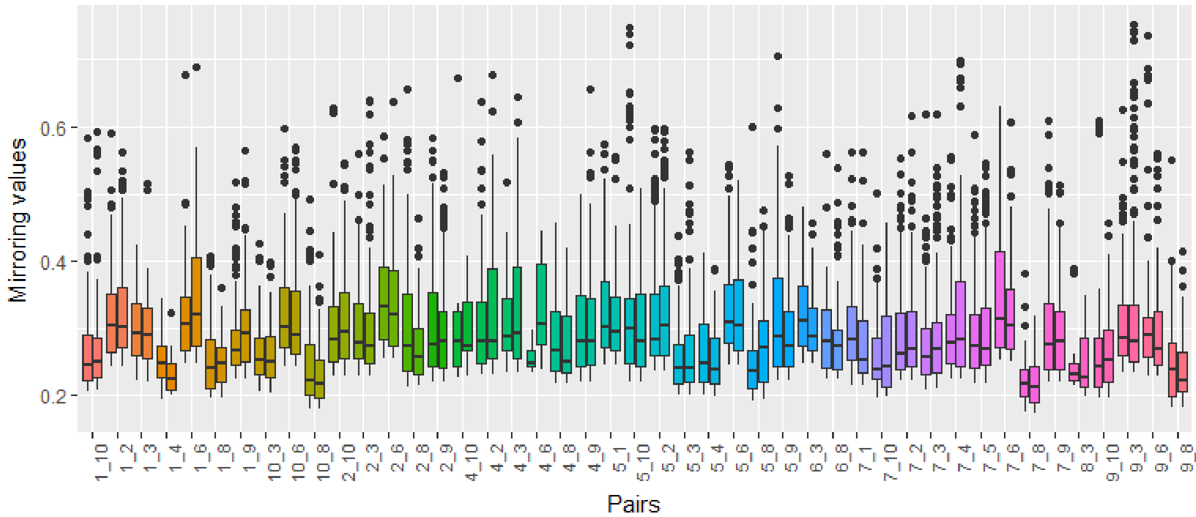


*Illustration 48: Nonverbal body language mirroring values > mean + 2 * sd*

Team 3

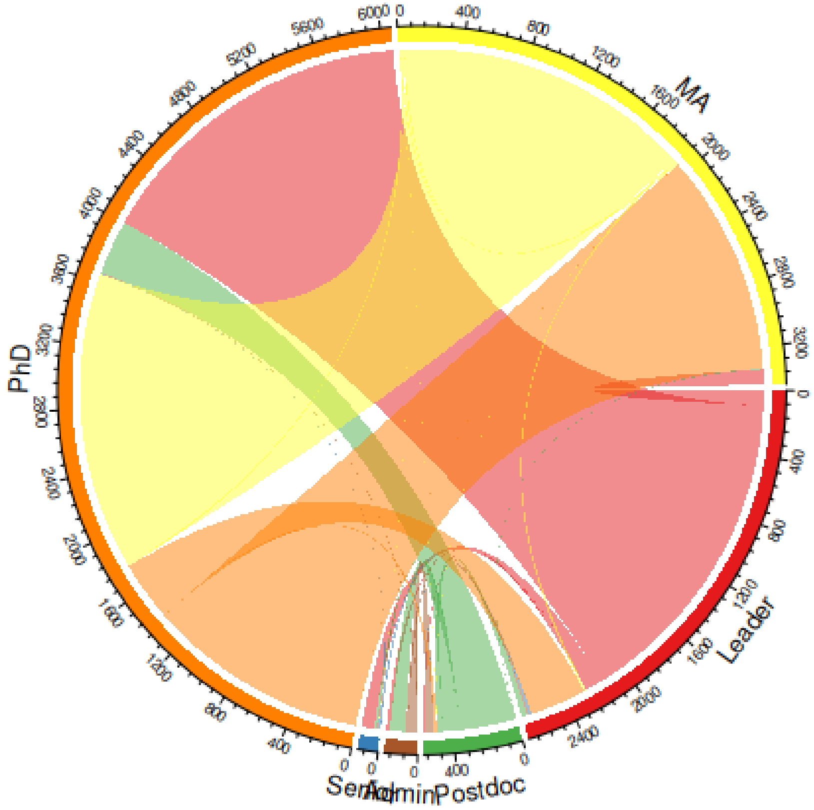


Illustration 49: Chord diagram face-to-face interactions by role Team 3

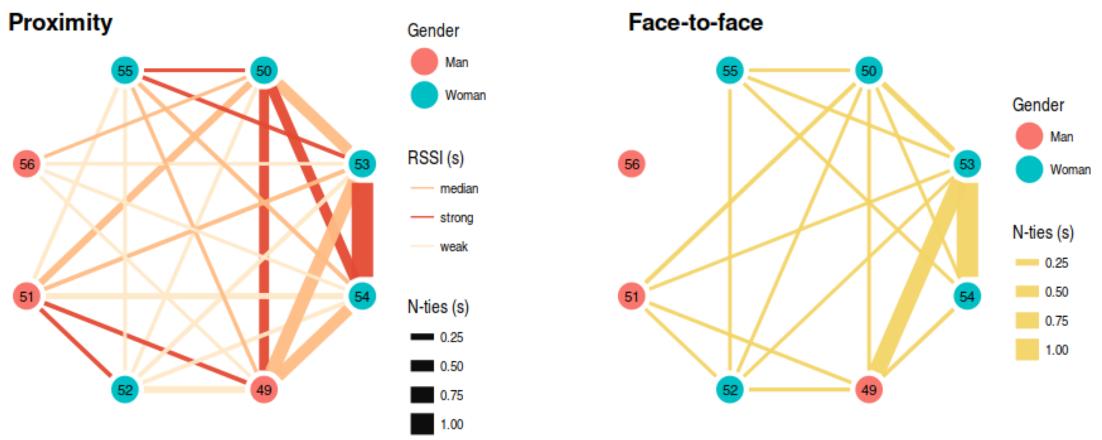


Illustration 50: Network graph of aggregated proximity and face-to-face interactions T3

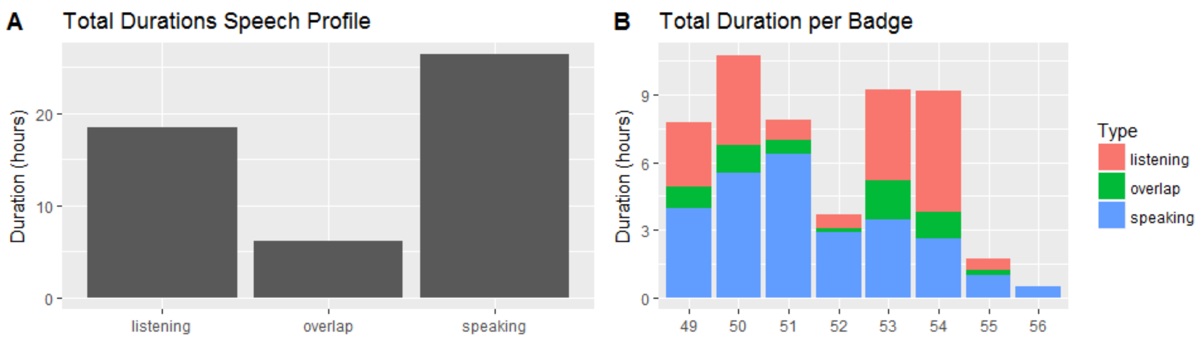


Illustration 51: Speech profile for T3 by speech type and badges

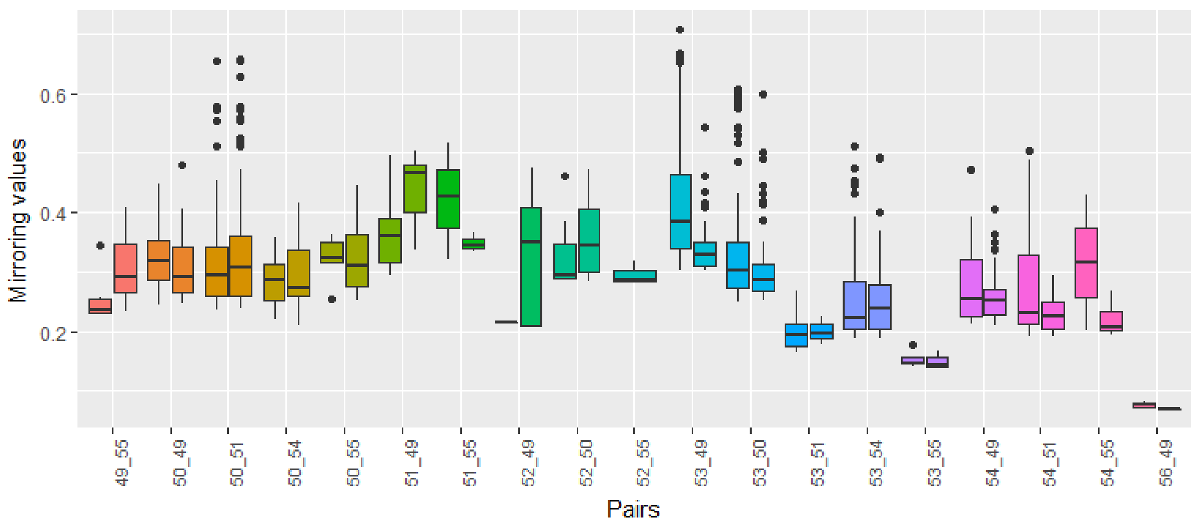


*Illustration 52: Nonverbal body language mirroring values > mean + 2 * sd*

Team 4

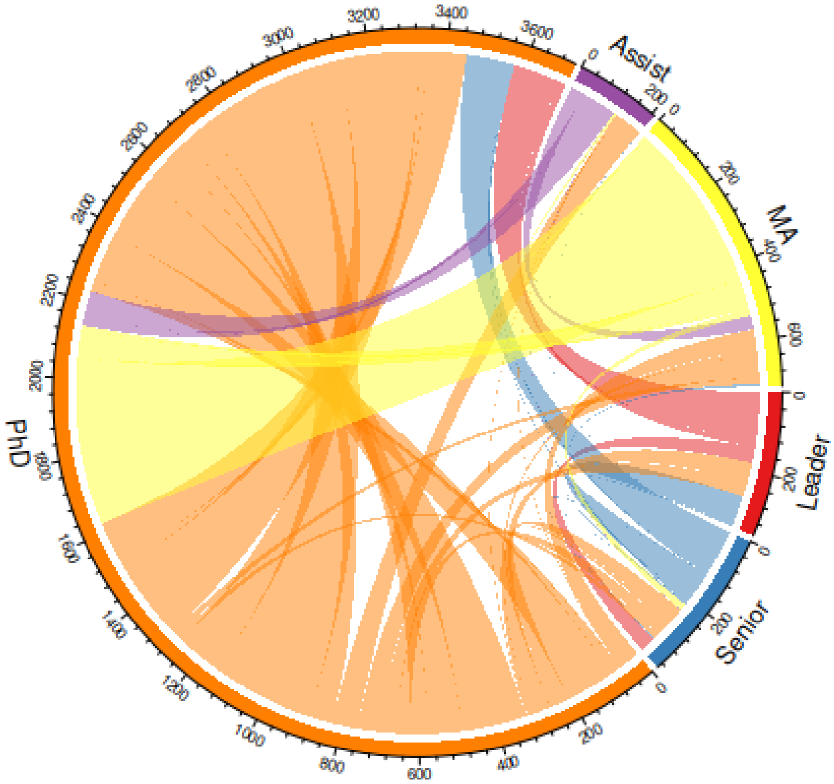


Illustration 53: Chord diagram face-to-face interactions by role Team 4

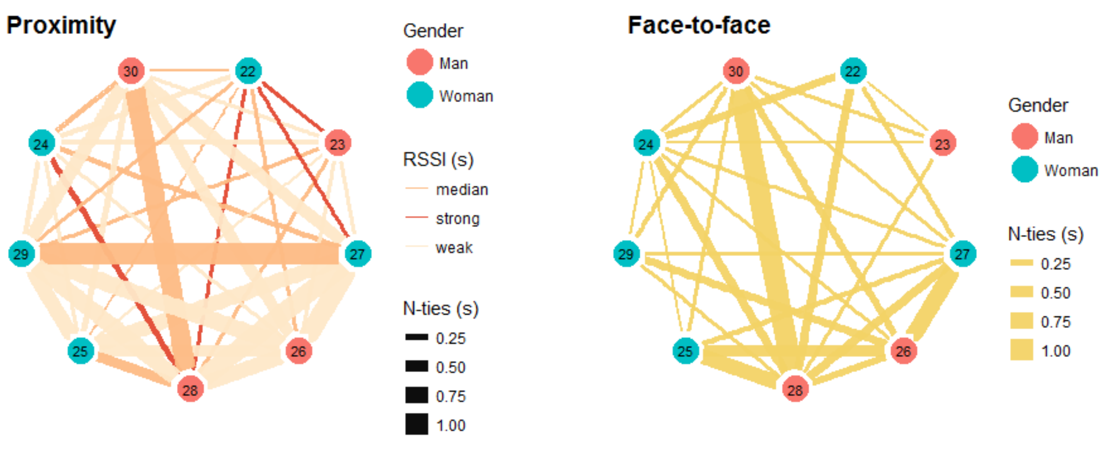


Illustration 54: Network graph of aggregated proximity and face-to-face interactions T4

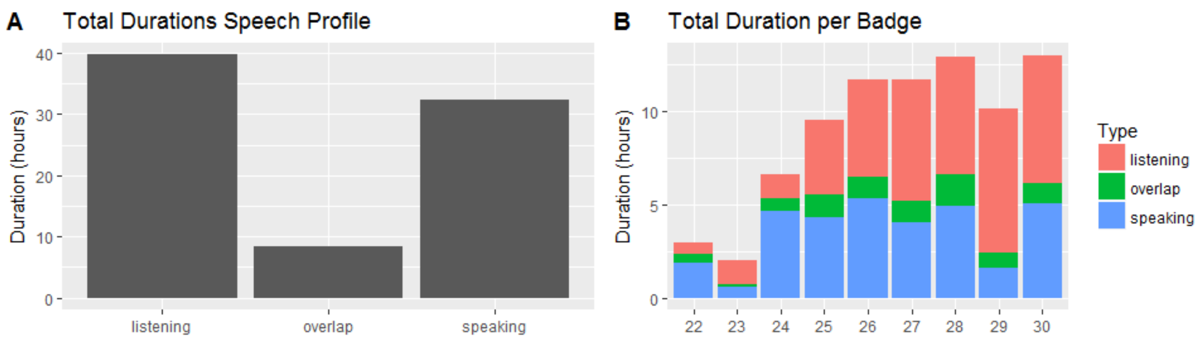


Illustration 55: Speech profile for T4 by speech type and badges

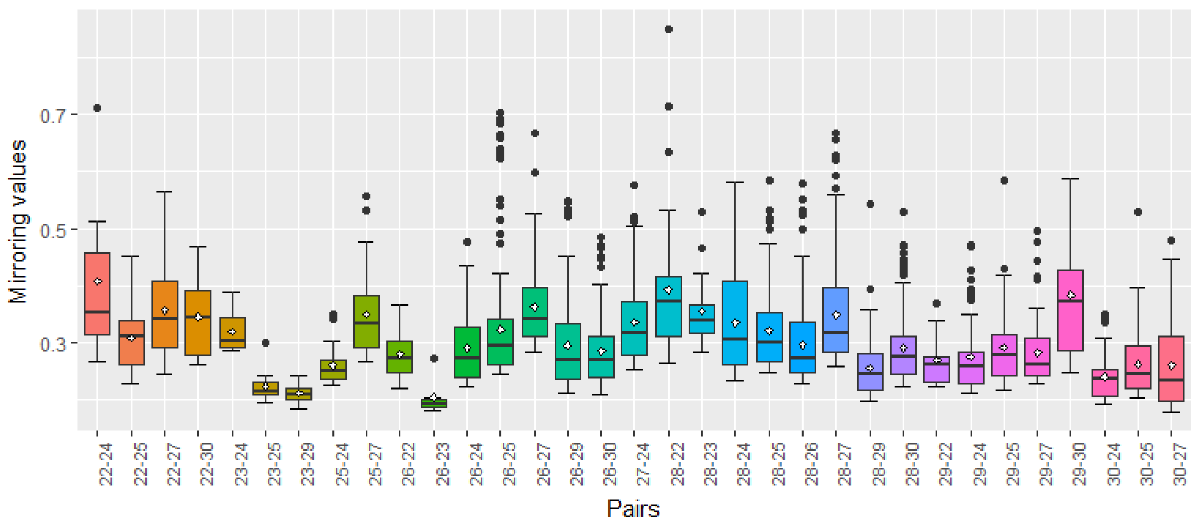


*Illustration 56: Nonverbal body language mirroring values > mean + 2 * sd*

Team 5

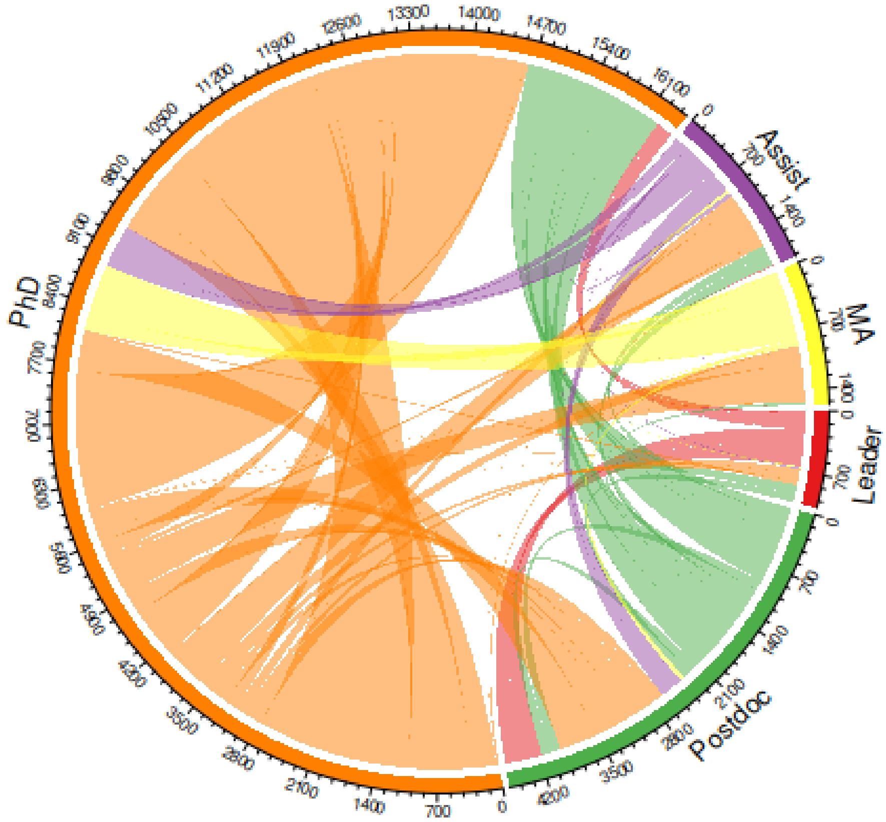


Illustration 57: Chord diagram face-to-face interactions by role Team 5

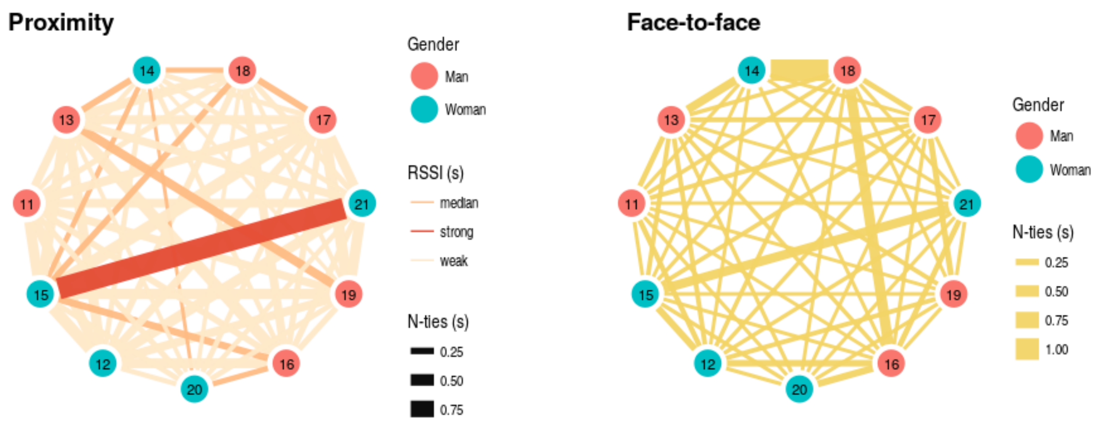


Illustration 58: Network graph of aggregated proximity and face-to-face interactions T5

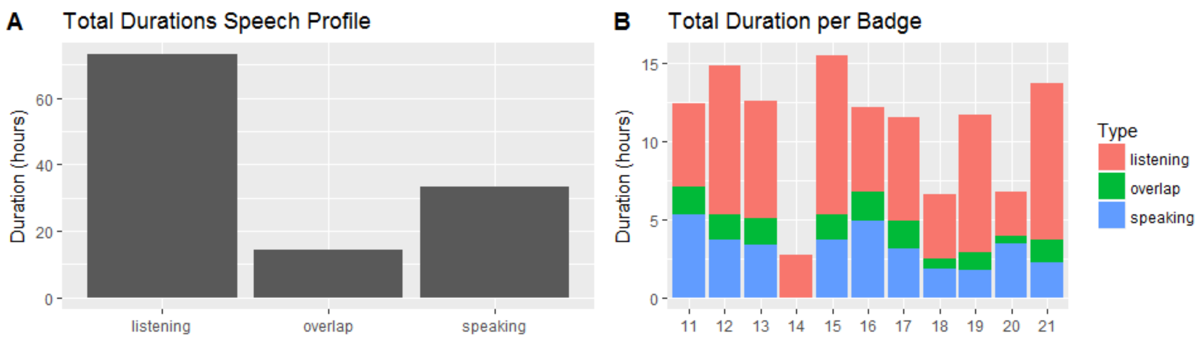


Illustration 59: Speech profile for T5 by speech type and badges

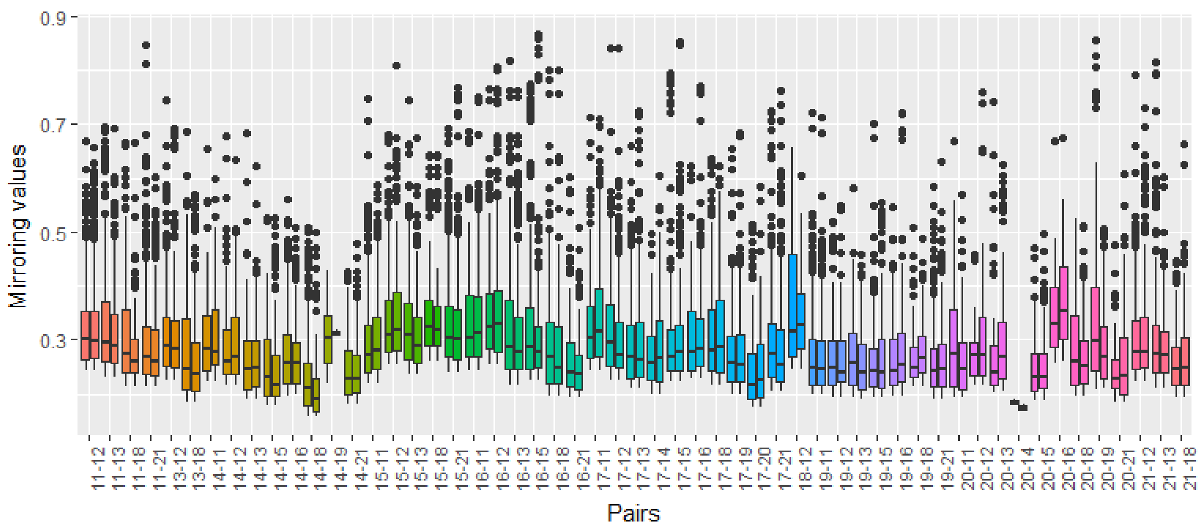


*Illustration 60: Nonverbal body language mirroring values > mean + 2 * sd*

Team 6

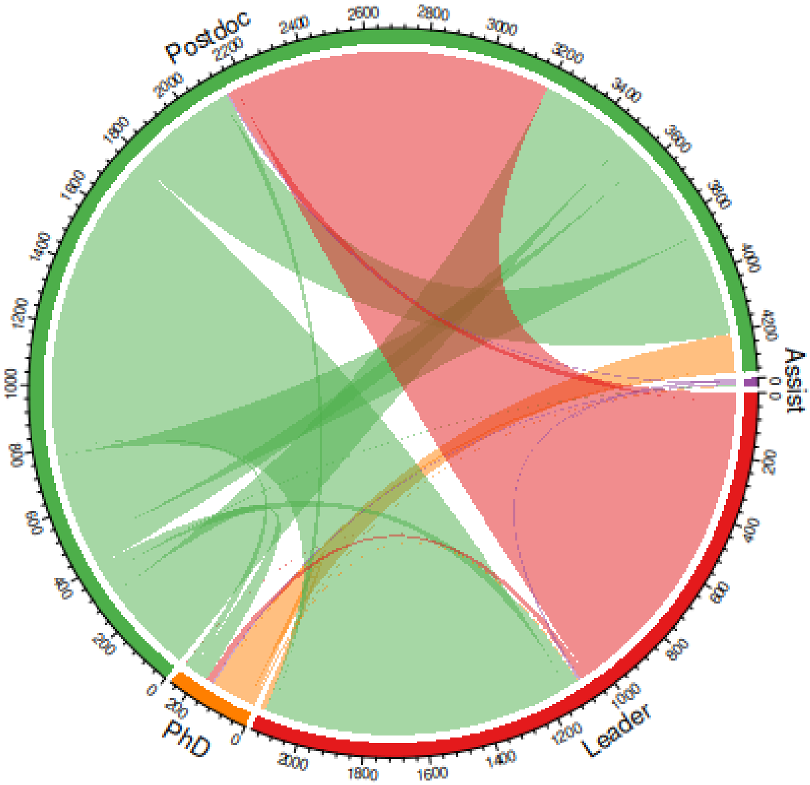


Illustration 61: Chord diagram face-to-face interactions by role Team 6

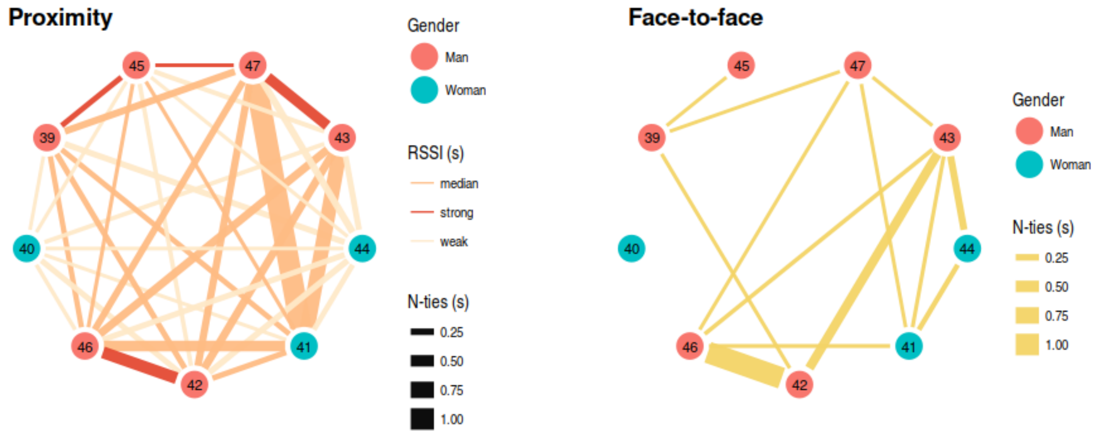


Illustration 62: Network graph of aggregated proximity and face-to-face interactions T6

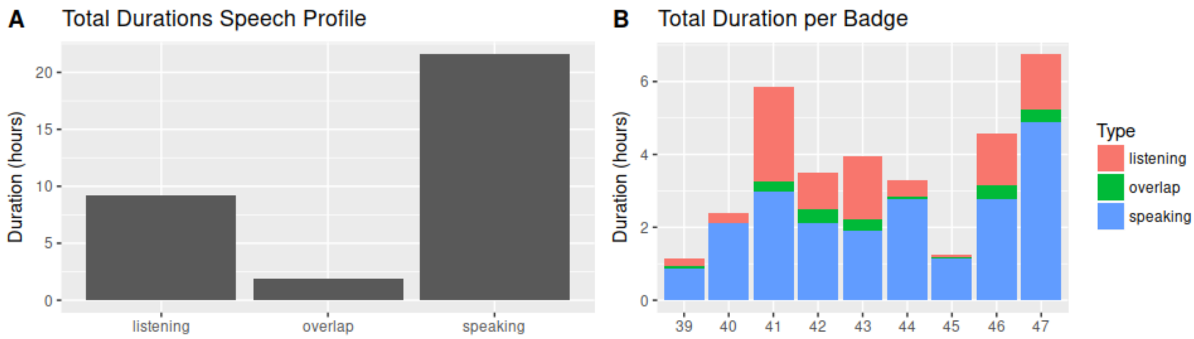


Illustration 63: Speech profile for T6 by speech type and badges

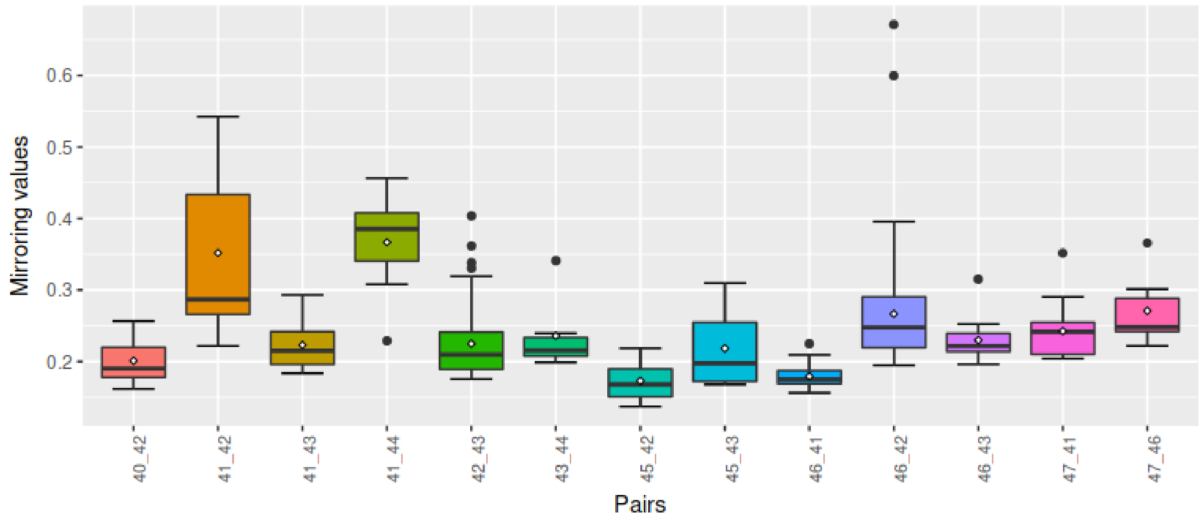


*Illustration 64: Nonverbal body language mirroring values > mean + 2 * sd*

Team 7

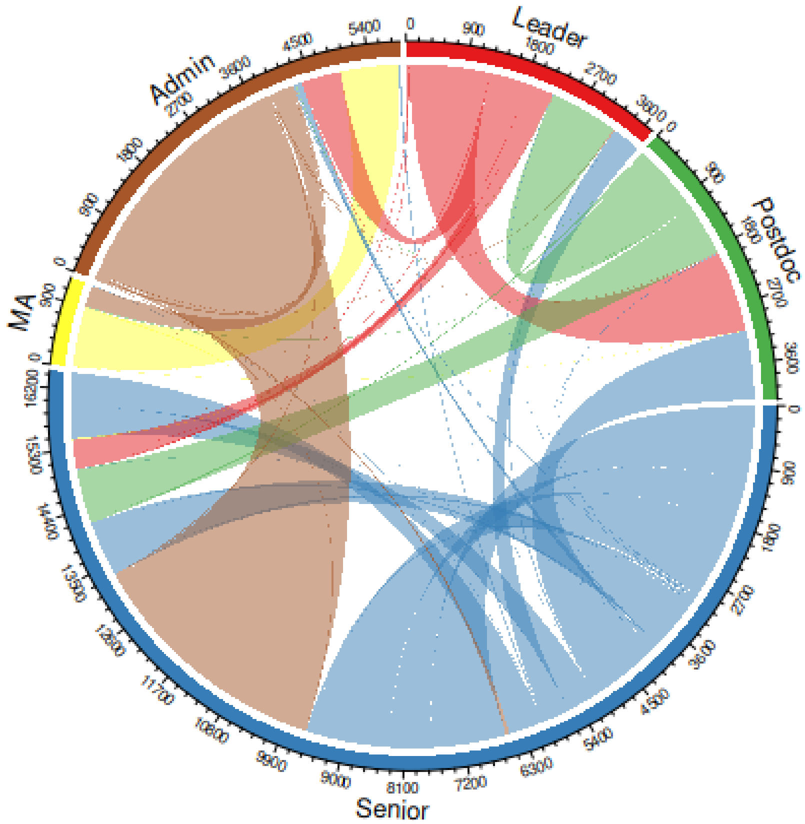


Illustration 65: Chord diagram face-to-face interactions by role Team 7

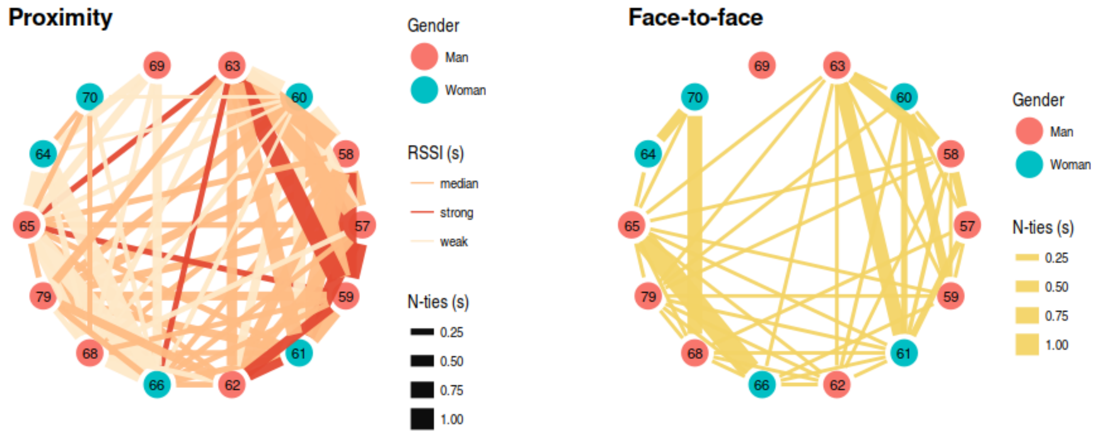


Illustration 66: Network graph of aggregated proximity and face-to-face interactions T7

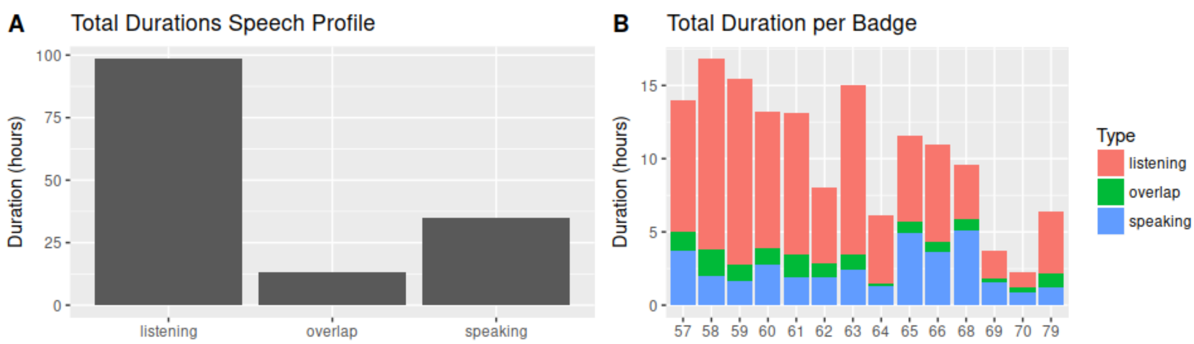


Illustration 67: Speech profile for T7 by speech type and badges

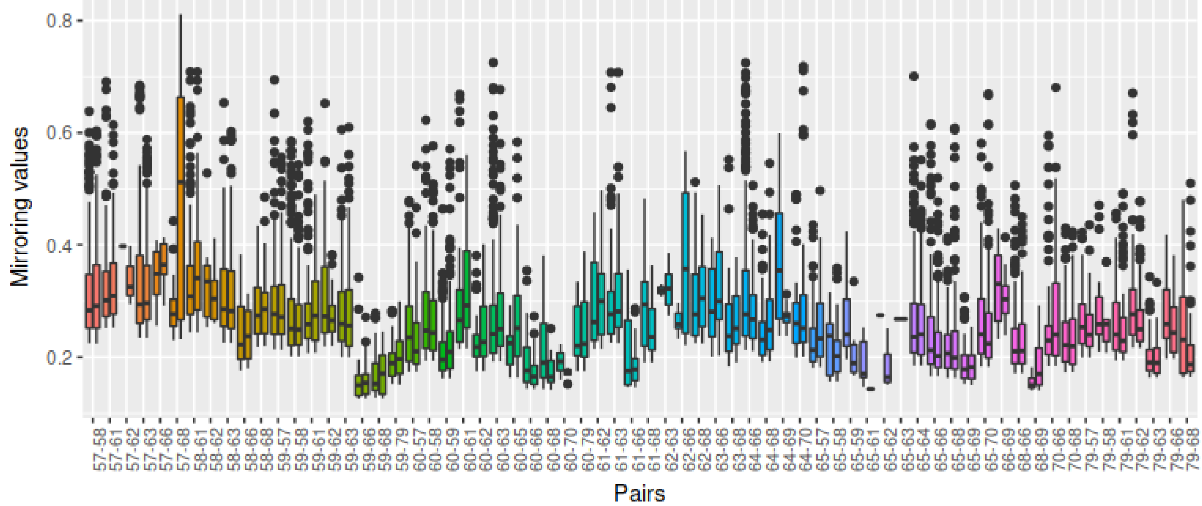


*Illustration 68: Nonverbal body language mirroring values > mean + 2 * sd*

Team 8

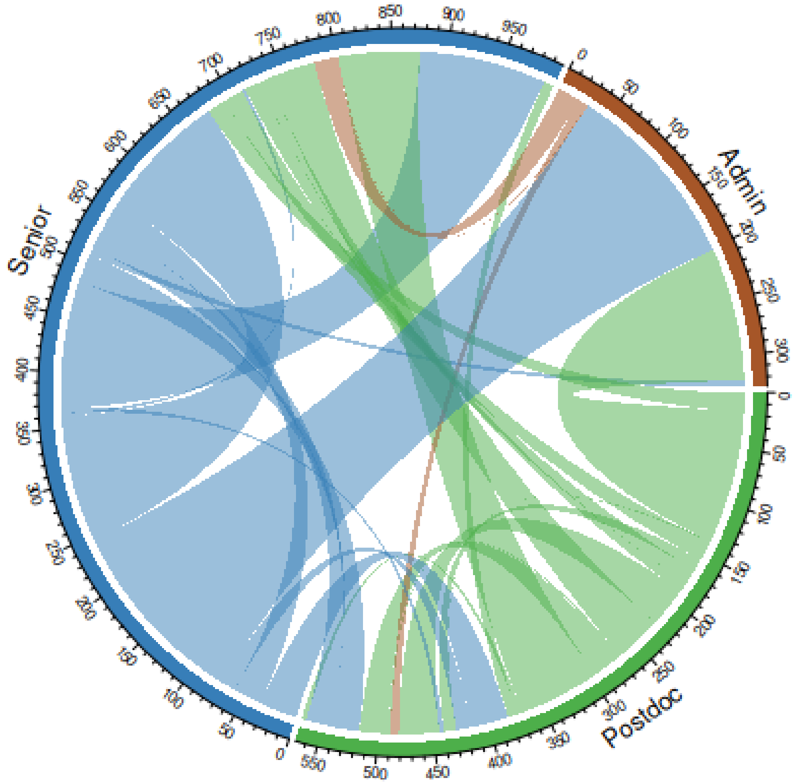


Illustration 69: Chord diagram face-to-face interactions by role Team 8

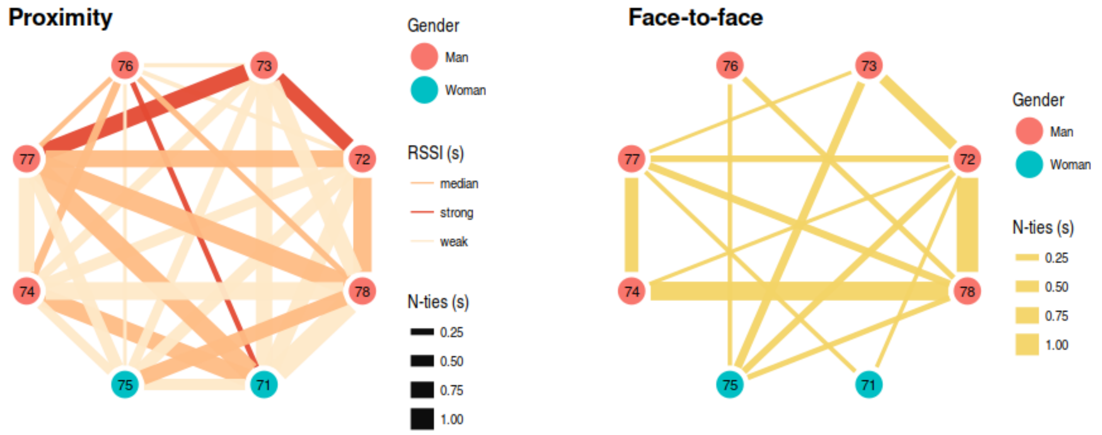


Illustration 70: Network graph of aggregated proximity and face-to-face interactions T8

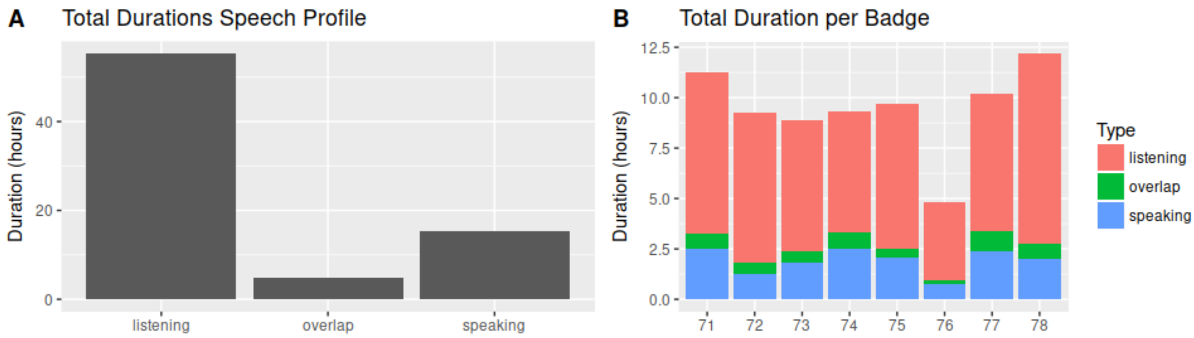


Illustration 71: Speech profile for T8 by speech type and badges

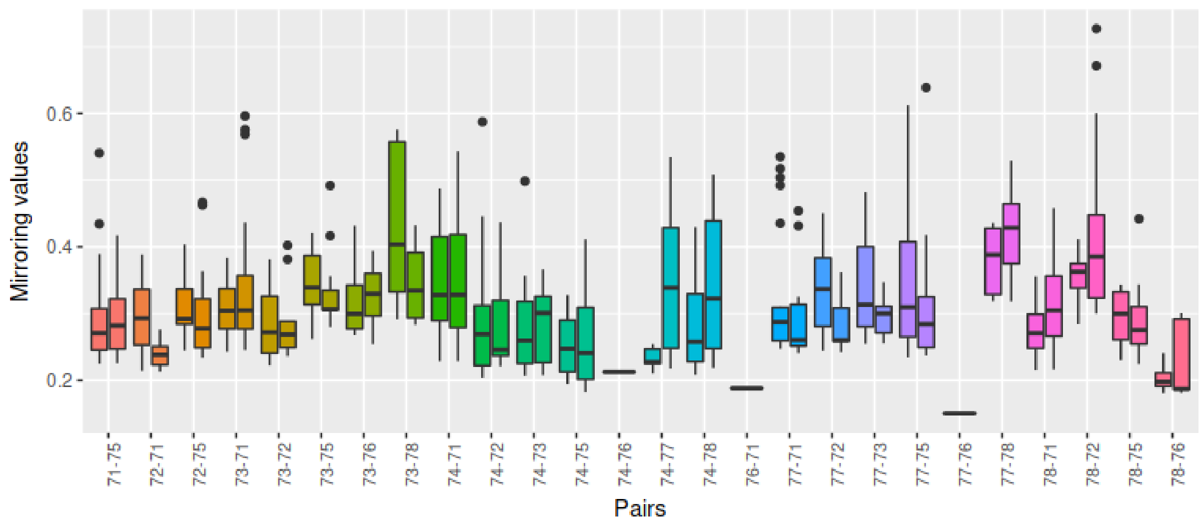


*Illustration 72: Nonverbal body language mirroring values > mean + 2 * sd*

ANNEX III – Basic Badge Measures & Network Statistics

Team 1

	Proximity			Face-to-face		F2F Network Measures		
Badge	Count	Percent	Badge	Count	Percent	Betweenness	Eigen	Degree
34	4572	0.23	**34**	859	0.29	2.5	0.92	4
36	3815	0.19	**37**	753	0.25	7.0	1.00	5
37	3545	0.18	**35**	388	0.13	0.0	0.56	2
33	2898	0.14	**33**	379	0.13	2.5	0.92	4
32	2844	0.14	**36**	360	0.12	0.0	0.58	2
35	1611	0.08	**32**	224	0.08	0.0	0.58	2
31	711	0.04	**31**	23	0.01	0.0	0.30	1

Table 22: Total proximity and face-to-face detects; network statistics of aggregated 5 day face-to-face interactions T1

Team 2

	Proximity			Face-to-face		F2F Network Measures		
Badge	Count	Percent	Badge	Count	Percent	Betweenness[26]	Eigen[27]	Degree
5	21961	0.16	**2**	3274	0.19	0.85	1.00	9
2	21413	0.16	**5**	3007	0.17	0.85	1.00	9
9	17949	0.13	**3**	2667	0.15	0.51	0.91	8
7	16766	0.13	**7**	2647	0.15	0.85	1.00	9
10	13741	0.10	**1**	1534	0.09	0.29	0.92	8
3	11602	0.09	**9**	1480	0.09	0.85	1.00	9
1	9108	0.07	**10**	1044	0.06	0.51	0.91	8
6	9004	0.07	**4**	941	0.05	0.14	0.82	7
4	6591	0.05	**6**	386	0.02	0.00	0.72	6
8	5005	0.04	**8**	368	0.02	0.14	0.82	7

Table 23: Total proximity and face-to-face detects; network statistics of aggregated 5 day face-to-face interactions T2

26 Betweenness centrality quantifies the number of times a node acts as a bridge along the shortest path between two other nodes.

27 Eigencentrality measure of the influence of a node on a network. It assigns relative scores to all nodes in the network based on the concept that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes.

Team 3

	Proximity			Face-to-face			F2F Network Measures	
Badge	Count	Percent	Badge	Count	Percent	Betweenness	Eigen	Degree
53	12990	0.26	**53**	6065	0.46	1.17	1.00	6
54	12540	0.25	**54**	3399	0.26	0.25	0.73	4
49	9994	0.20	**49**	2696	0.21	0.58	0.88	5
50	8587	0.17	**50**	583	0.04	1.17	1.00	6
51	3235	0.06	**51**	195	0.01	0.00	0.76	4
52	2306	0.05	**52**	118	0.01	0.58	0.88	5
55	1001	0.02	**55**	34	0.00	0.25	0.73	4
56	19	0.00	**56**	-	-	0.00	0.00	0

Table 24: Total proximity and face-to-face detects; network statistics of aggregated 5 day face-to-face interactions T3

Team 4

	Proximity			Face-to-face			F2F Network Measures	
Badge	Count	Percent	Badge	Count	Percent	Betweenness	Eigen	Degree
27	18507	0.17	**28**	1496	0.27	2.70	1.00	8
29	18133	0.17	**26**	926	0.17	0.20	0.85	6
28	18065	0.17	**30**	717	0.13	1.45	0.92	7
26	15616	0.15	**27**	714	0.13	1.45	0.92	7
30	14299	0.13	**25**	647	0.12	0.00	0.75	5
25	11887	0.11	**22**	349	0.07	0.25	0.55	4
24	5141	0.05	**24**	348	0.06	2.70	1.00	8
23	2822	0.03	**29**	208	0.04	0.00	0.75	5
22	2672	0.02	**23**	27	0.01	0.25	0.55	4

Table 25: Total proximity and face-to-face detects; network statistics of aggregated 5 day face-to-face interactions T4

Team 5

	Proximity			Face-to-face		F2F Network Meaures		
Badge	Count	Percent	Badge	Count	Percent	Betweenness	Eigen	Degree
15	34534	0.15	**18**	5390	0.21	0.25	1.00	10
21	29615	0.12	**14**	4306	0.17	0.00	0.83	8
13	27132	0.11	**16**	3001	0.12	0.25	1.00	10
12	26882	0.11	**13**	2362	0.09	0.25	1.00	10
17	24958	0.10	**12**	2243	0.09	0.25	1.00	10
19	22681	0.10	**15**	2206	0.09	0.25	1.00	10
11	22483	0.09	**21**	1581	0.06	0.25	1.00	10
16	19818	0.08	**17**	1555	0.06	0.25	1.00	10
18	14866	0.06	**11**	1056	0.04	0.25	1.00	10
14	7582	0.03	**19**	942	0.04	0.00	0.92	9
20	7533	0.03	**20**	902	0.04	0.00	0.92	9

Table 26: Total proximity and face-to-face detects; network statistics of aggregated 5 day face-to-face interactions T5

Team 6

	Proximity			F2F		F2F Network Measures		
Badge	Count	Percent	Badge	Count	Percent	Betweenness	Eigen	Degree
39	1723	0.07	**39**	25	0	6.33	0.42	3
40	451	0.02	**40**	0	0	0.00	0.00	0
41	5366	0.21	**41**	255	0.04	1.67	0.87	4
42	3322	0.13	**42**	2653	0.39	2.67	0.62	3
43	3961	0.15	**43**	1124	0.17	4.17	1.00	5
44	1897	0.07	**44**	554	0.08	0.00	0.50	2
45	620	0.02	**45**	4	0	0.00	0.11	1
46	3539	0.14	**46**	2120	0.31	0.83	0.89	4
47	4931	0.19	**47**	27	0	5.33	0.85	4

Table 27: Total proximity and face-to-face detects; network statistics of aggregated 5 day face-to-face interactions T6

Team 7

Proximity			Face-to-face			F2F Network Measures		
Badge	Count	Percent	Badge	Count	Percent	Betweenness	Eigen	Degree
63	4150	0.13	**64**	1491	0.05	0.00	0.15	2
65	3958	0.13	**70**	3701	0.12	0.70	0.24	3
66	3864	0.12	**57**	1700	0.05	0.94	0.53	5
61	3735	0.12	**68**	2439	0.08	4.83	0.70	7
70	3701	0.12	**59**	1243	0.04	2.03	0.76	7
58	2957	0.09	**60**	1874	0.06	1.01	0.78	7
68	2439	0.08	**66**	3864	0.12	0.76	0.81	7
60	1874	0.06	**58**	2957	0.09	1.89	0.89	8
57	1700	0.05	**62**	191	0.01	1.55	0.91	8
64	1491	0.05	**63**	4150	0.13	1.11	0.92	8
59	1243	0.04	**65**	3958	0.13	16.03	0.96	10
62	191	0.01	**61**	3735	0.12	3.03	0.97	9
79	135	0.00	**79**	135	0.00	2.11	1.00	9

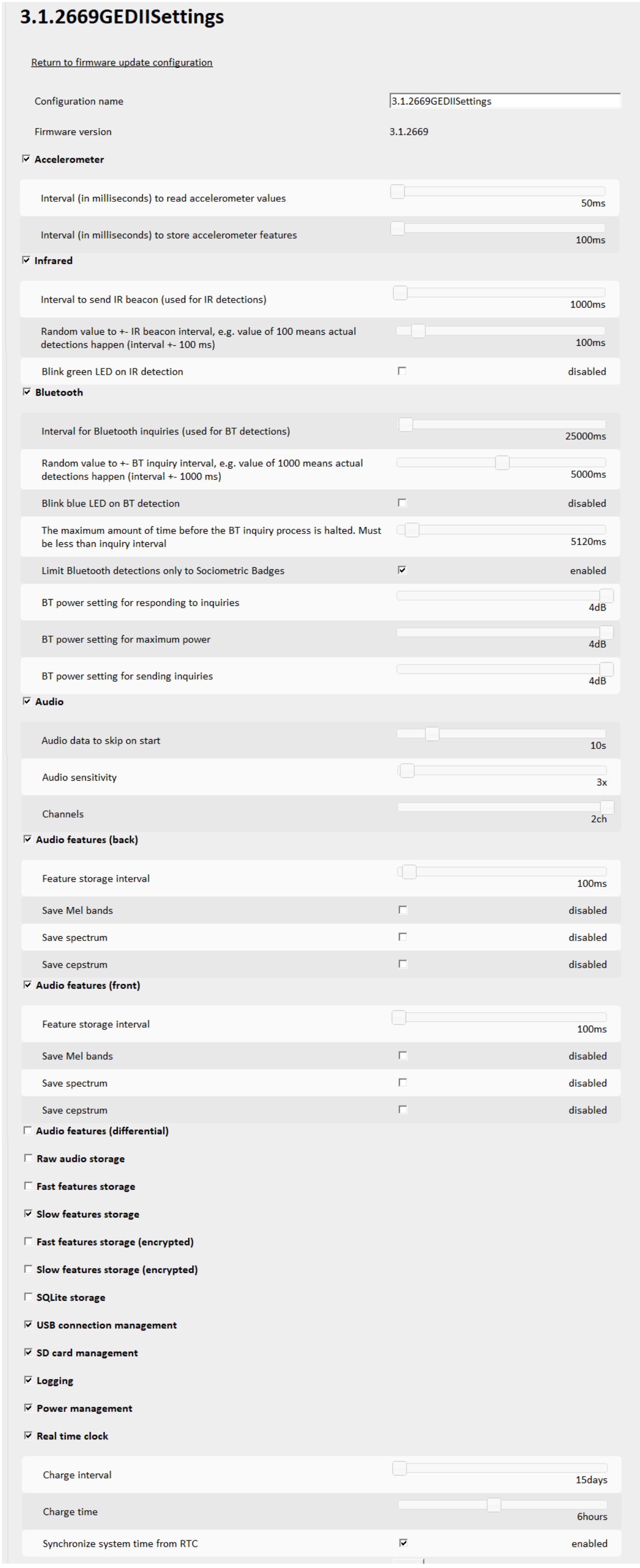
Table 28: Total proximity and face-to-face detects; network statistics of aggregated 5 day face-to-face interactions T7

Team 8

Proximity			F2F			F2F Network Measures		
Badge	Count	Percent	Badge	Count	Percent	Betweenness	Eigen	Degree
78	42797	0.16	**76**	42	0.02	0.00	0.37	2
71	40194	0.15	**71**	23	0.01	0.00	0.45	2
77	39711	0.15	**73**	160	0.08	0.33	0.62	3
72	36005	0.13	**74**	328	0.17	0.00	0.66	3
74	34100	0.13	**75**	149	0.08	2.17	0.69	4
73	33667	0.12	**78**	530	0.28	4.00	0.87	5
75	29195	0.11	**77**	237	0.12	2.67	0.87	5
76	16545	0.06	**72**	427	0-23	4.83	1.00	6

Table 29: Total proximity and face-to-face detects; network statistics of aggregated 5 day face-to-face interactions T8

ANNEX IV - Sociometric Badges Firmware Setting



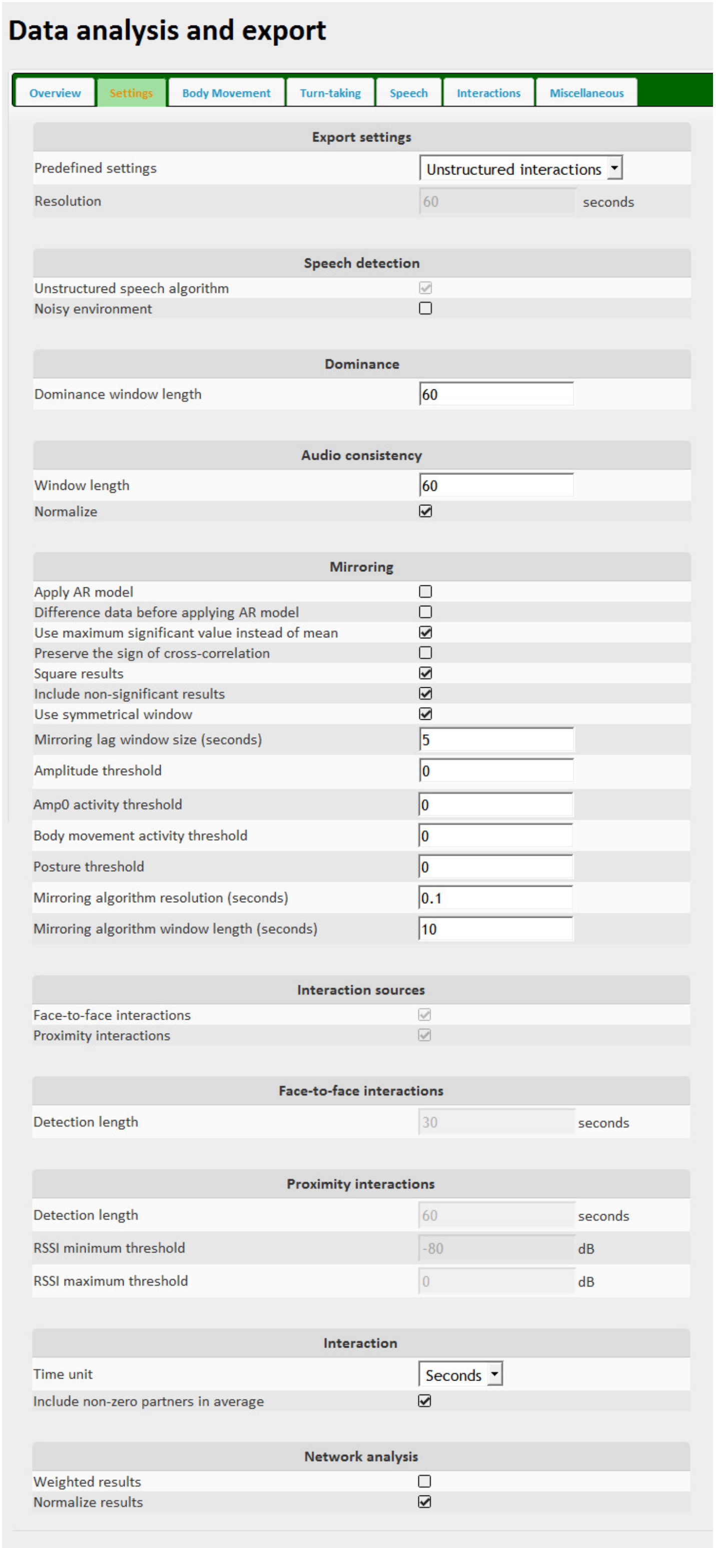
3.1.2669GEDIISettings

Return to firmware update configuration

Configuration name	3.1.2669GEDIISettings	
Firmware version	3.1.2669	
☑ **Accelerometer**		
Interval (in milliseconds) to read accelerometer values		50ms
Interval (in milliseconds) to store accelerometer features		100ms
☑ **Infrared**		
Interval to send IR beacon (used for IR detections)		1000ms
Random value to +- IR beacon interval, e.g. value of 100 means actual detections happen (interval +- 100 ms)		100ms
Blink green LED on IR detection	☐	disabled
☑ **Bluetooth**		
Interval for Bluetooth inquiries (used for BT detections)		25000ms
Random value to +- BT inquiry interval, e.g. value of 1000 means actual detections happen (interval +- 1000 ms)		5000ms
Blink blue LED on BT detection	☐	disabled
The maximum amount of time before the BT inquiry process is halted. Must be less than inquiry interval		5120ms
Limit Bluetooth detections only to Sociometric Badges	☑	enabled
BT power setting for responding to inquiries		4dB
BT power setting for maximum power		4dB
BT power setting for sending inquiries		4dB
☑ **Audio**		
Audio data to skip on start		10s
Audio sensitivity		3x
Channels		2ch
☑ **Audio features (back)**		
Feature storage interval		100ms
Save Mel bands	☐	disabled
Save spectrum	☐	disabled
Save cepstrum	☐	disabled
☑ **Audio features (front)**		
Feature storage interval		100ms
Save Mel bands	☐	disabled
Save spectrum	☐	disabled
Save cepstrum	☐	disabled
☐ **Audio features (differential)**		
☐ **Raw audio storage**		
☐ **Fast features storage**		
☑ **Slow features storage**		
☐ **Fast features storage (encrypted)**		
☐ **Slow features storage (encrypted)**		
☐ **SQLite storage**		
☑ **USB connection management**		
☑ **SD card management**		
☑ **Logging**		
☑ **Power management**		
☑ **Real time clock**		
Charge interval		15days
Charge time		6hours
Synchronize system time from RTC	☑	enabled

ANNEX V – Sociometric Badges Export Settings

The standard export settings of the Sociometric Datalab have been used. Depending on the type of data exported, "Unstructured" vs "Structured Meetings" was selected which basically changes the resolution from 60 to 1 second.



ANNEX VI – Sociometric Fieldwork Recommendations

Here is a small collection of helpful points for conducting field research with Sociometric badges.

- Recruitment of R&D groups for the current study was difficult. Personal contacts to research institutes were key. However, after most team members did not share “our” (the researchers) primary concern, namely the highly intrusive character of the sociometric badges tracking. Data privacy issues were seldom the most pressing concern during the presentation of the study to the teams; for most participants curiosity regarding the badges outweighed by far any worries about potential negative impacts. The most repeated question concerned the possible effects of manipulating the data: to which degree people would behave differently because they were wearing the badges. However, as most participants recognized during and after the study, usually they were not aware that they were wearing the badges. Usually participants forget that they do wear them.
- Sociometric badges should be charged the night before being used in the field. If several badges are connected one and the same (USB) charger, one has to be careful that the performance is sufficiently high. Different models of USB chargers deliver different levels of power that range from 0.2A to 2 A for some models. Using a 2A model for 4 badges simultaneously was sufficient to charged the badges a 100% over 6 hours.
- The badges internal clock has to be synchronized to a computer clock. This is necessary to accurately timestamp events across the badges deployed. For the GEDII case studies this was done the night before when discharging the data.
- Sociometric Badges are connected via the USB port to a computer. The Sociometric Data Lab is only available for the Windows Operating system. Windows XP and Windows 7 and 8 recognize without problems the necessary driver (RNDIS Gadget) as described in the user guide. However, Windows 10 does not. A work around consists of installing and using the Amazon Kindl driver as described in the following forum thread: https://www.mobileread.com/forums/showthread.php?p=3283986
- Field logistics for distributing the badges to each research team included agree on a centralized (lab) place where individuals can pick-up and drop-off the badges. Badges were deposited in the morning and picked-up in the afternoon. A list with team members names and badge number was provided to make sure that participants wear the same badge each day. Downloading the data in the evening and charging the badges over night was not problematic.
- In order to facilitate the analysis of pitch scores, a speech sample should be recorded for each participant. A standard text passage such as the Rainbow passage can be

used for English speakers.[28] The sound should be recorded with the badge as well as a standard recording device.

- A useful “trick” to easily identify segments of data in the sociometric recordings is by producing a loud beep. The high volume is easily identifiable in the audio profile and facilitates to determine the exact timestamps when a certain event occurred. This technique was used especially during the execution of the controlled experiments in order to mark off the various speech segments.
- The firmware settings used for the Sociometric badges were configured to store body movement activity at 0.1 second interval. Although this setting is recommended only for structured meetings, it was used for the entire field period. The higher sampling rate does not provide any problems in terms of data storage nor battery life – during one day. On average, one day of data per participate produced between 200 and 350 MB of data.
- Exporting data from the Sociometric Datalab to Excel sheets is a time consuming process depending on the chosen settings and variables. Interaction data can be exported relatively quickly. However, body mirroring values or speech profiles can take a very long time. It is recommended to export only one dimension at a time for selected badges and sessions. Even then, the export process can take several hours. For example, for a team of 11, exporting the speech profile for one day at 60 seconds aggregation level took 5-6 hours on an Intel Core i5 3.2 Ghz processor.
- In order to work with the Excel sheets of “raw” data exported from the Sociometric Datalab, many custom functions have been written in R. These include scripts to convert the Excel sheets into the “tidy” data format, quickly filtering timestamped data, clustering timestamps, anonymize badge ids, extract session data, or quickly generate network graphs and other visualizations. These scripts will be made available in the near future on Github and the projects website. Anybody interested is also encourage to get directly in contact with Jörg Müller (jmuller@uoc.edu).

28 http://www.dialectsarchive.com/the-rainbow-passage