Understanding how Teams Interact - An Analysis of Teams in an International Science and Engineering Competition

Rathin J
Roll No. 15PT30

DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

FIVE YEAR INTEGRATED
M.Sc THEORETICAL COMPUTER SCIENCE

OF ANNA UNIVERSITY

MAY 2020

DEPARTMENT OF APPLIED MATHEMATICS AND COMPUTATIONAL SCIENCES

PSG COLLEGE OF TECHNOLOGY
(Autonomous Institution)

COIMBATORE – 641 004

PSG COLLEGE OF TECHNOLOGY
(Autonomous Institution)

COIMBATORE – 641 004

Tenth Semester
Project work

Understanding how Teams Interact - An Analysis of Teams in an International Science and Engineering Competition

Bona fide record of work done by

Rathin J
Roll No. 15PT30

Submitted in partial fulfillment of the requirements for the degree of

FIVE YEAR INTEGRATED
M.Sc THEORETICAL COMPUTER SCIENCE
of Anna University

MAY 2020

______________________ ______________________

Faculty Guide **Head of the Department**

Submitted for the Viva-Voce Examination held on ______________________

Internal Examiner **External Examiner**

______________________ ______________________

Contents

Abstract

With a rise in collaborative research [1], the science of how teams conduct science is becoming increasingly relevant. Successful teams share several features which help them develop and sustain their efforts towards a productive outcome [2]. Taking the iGEM (International Genetically Engineered Machine) competition as a testbed for the study, the report aims to translate interaction data collected through series of surveys from 19 participating teams into insights on organisation, learning and factors driving collaborative activity within teams. In particular, the (i) team interaction hierarchy, (ii) attributes driving work and social interactions and (iii) skill contagion are discussed with a view to build hypotheses central to the dynamics of teams practising science. The work presented in this report was done during my time as a *stagiaire* with Dr. Marc Santolini, long term research fellow at the Centre de Recherches Interdisciplinaires (CRI), Paris.

Keywords Team Science . Surveys . Interaction Networks . Assortativity . Multilayer Networks . Skill Contagion . iGEM

Acknowledgements

I am very grateful to Dr. Marc Santolini for inviting me as an intern to his team. It has been a continuous learning arc over the past six months and I owe it primarily to his mentoring and support. I'd also like to thank the Network TIES team and the CRI for making my experience a pleasant and rewarding one. I look forward to continuing my association here and doing more science.

I offer my sincerest gratitude to my internal guide Dr. R.S. Lekshmi, Dr. Nadarajan and the Applied Mathematics and Computational Sciences department for their belief and constant words of encouragement. I cherish the memories and invaluable lessons I learned along the way immensely.

I am much obliged to my hosts, Piramaladevi and St Hubert Chrysostome, for helping me in every aspect of arriving and settling in Paris - especially during the COVID19 pandemic.

Finally, but not the least, I want to send love to my parents, brother and friends - who have kept me sane with their regular calls and interesting conversations.

Chapter 1

Introduction

1.1 Motivation

Science is done predominantly by teams [1]. With a push towards interdisciplinary research efforts for scientific discovery, collaboration in research has become critical now more than ever [2]. But organising science into research groups has its own implications. There have been studies on how the science of teams effects in improving research and practices of team science [3], and how collaborative organisation of teams impact team creativity [4] and productivity.

The primary question being phrased in the context of team science is - “What makes teams successful in their endeavours?” - and the key approach is to dissect and decipher activity within teams. From physical interactions in a lab setting, to online in the case of open source contribution, the collaboration *per se* can be measured through different means. With the advent of digital tools to collect large-scale interaction data [5], data-driven approaches to study team dynamics have become significant [6].

With this question in mind, the iGEM Competition is chosen as the *testbed* to study collaborative science in a team setting. In particular, the report focuses on how interactions within participating teams help understand processes underlying teamwork, performance, organisation and learning. The competition structure is detailed in the following section with reasoning on why iGEM is a unique candidate for the study of team science.

1.2 iGEM - Competition and Structure

International Genetically Engineered Machine, or henceforth referred to as iGEM, is a global scale team based synthetic biology and engineering competition. Hosted yearly by the iGEM foundation (based in Boston), the competition primarily caters to undergraduate and graduate students, though community biolabs and high school students are also eligible to participate. Since its inception in 2004 with 5 teams, iGEM has grown to encompass over 300 teams from several countries across the world in 2019.

The crux of the competition is for participating teams to design synthetic biology parts - or *biobricks* to tackle open real-world challenges. Teams work collaboratively to design, build and test these biobricks - with a growing repository maintained every year for reuse and further development. The competition is organised into three broad sections

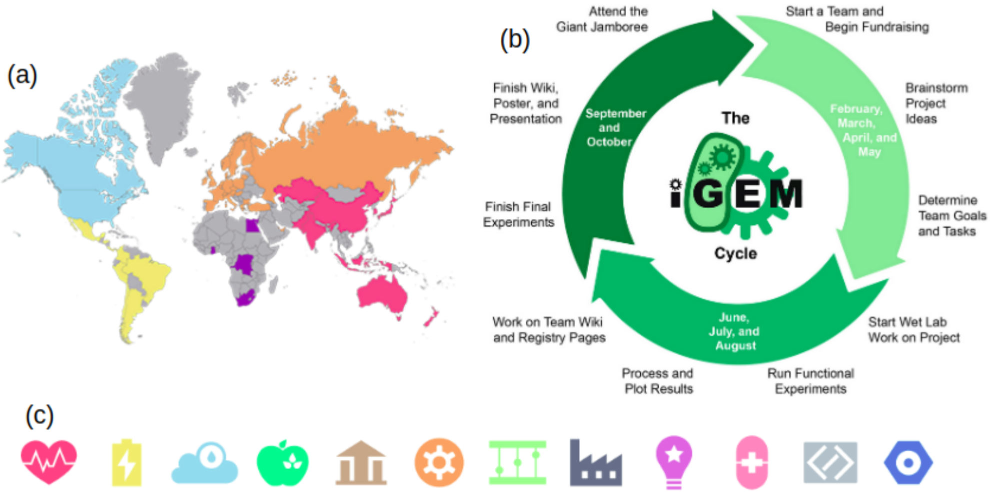


Figure 1.1: (a) Geographical distribution of iGEM teams over the course of the competition. (b) iGEM Competition cycle. (c) Participation Tracks. Images from iGEM website

(undergraduate, overgraduate and high school) and multiple tracks (environment, energy, manufacturing, diagnostics etc.) based on the design challenge that the team has chosen. Over the last few editions, iGEM culminates in a giant jamboree held at Boston where teams are encouraged to present their work and to also serve as a networking event for the entire iGEM community.

The competition cycle (documented in Figure 1.1 (b)) shows the workflow of iGEM from a participating teams' standpoint. Teams register for the competition, usually in early March. The project cycle starts at May - where teams begin ideating and run functional experiments. They wrap their work around September/October and the competition is concluded at the jamboree in November.

Participating teams maintain a detailed documentation of their biobrick designs, team practices and other aspects of the competition on an open collaborative wiki page. Other competition deliverables include a poster, presentation showcased in the jamboree, judging forms and evolving elements over the years - such as standards for measurement, and engaging in safety and responsible practices.

Based on the progress teams' have made on the deliverables, they are awarded medals (gold, silver, bronze or none). Teams also compete with other teams for track and section specific awards in addition to an overall prize. They are adjudicated by a panel of peer experts for the award of the prizes and all award details are documented in a competition wiki page. [7] links to the iGEM website - offering a more comprehensive account of the competition and its evolution over the years.

Several attributes of iGEM point it be a unique platform to study interactions, learning and organisational attributes leading to success in the context of team-based science. Firstly, the competition cycle and deliverables are controlled - establishing a notion of progress and success.

Teams are hierarchically organised, with team members assigned specific roles within their team. This is defined by iGEM and participating teams are required to specify member roles during registration. The composition of teams is as follows:

- **PI (Principal Investigator):** Primary and Secondary PIs are established researchers who lead the team on their research direction, organise fundraising and serve as points of contact with iGEM.

- **Advisors/Instructors:** They play a day-day role within the iGEM team - providing support and technical assistance.
- **Student Leaders:** Bridge the PIs, Advisors with the student members and take additional leadership responsibilities within the team.
- **Student Members:** Work with the other roles in brainstorming, planning and executing the team project.

Secondly, this provides a framework to study the effects of interaction amongst members of a team with varying roles and what effects those interactions. Diversity within a team can be leveraged for analysis - with no restrictions on team size, member skills, choice of participating track and teams free to organize themselves aligning to their project goals. Crucially, competition data is openly accessible, which is significant in iGEM's candidature to study team interactions in science.

1.3 Report Organisation

The report is organised into three further chapters. In chapter two, the study to collect data from iGEM teams for the year 2019 is described. Emphasis is provided on the processing and representation of interaction data between responding members of teams. In addition, methods for analysis and summary statistics of the survey responses are presented.

The main results - highlighting the effect of the team organisational hierarchy, participant skill and social attributes in driving interactions and learning is described in chapter three. The report is concluded by discussing the key takeaways and inferences to help better facilitate study of scientific teams, in the particular context of iGEM, for the later studies of the competition. References and materials for further and related reading are enlisted, along with supplementary information, which are cited in due course of the report.

Chapter 2

Survey and Data Collection

This chapter outlines the study and associated survey design for collecting interaction, team and individual metadata from the participants in iGEM 2019. Data cleaning and network construction are described along with an exploratory analysis of the data to better facilitate the analysis presented in the forthcoming chapters.

2.1 iGEM TIES 2019

The motivation of iGEM TIES (Team IntEraction Study) 2019 was to dissect and decipher organizational structure of participating iGEM teams. The study also explores how interactions within teams map to the learning and team performance in iGEM. This was facilitated through a series of surveys and questionnaires to examine participant attributes, demographics and how diversity effectively influences the team and their outcome. In addition, participants were also asked to self-report their feeling of belonging within the team - with questions pertaining to social and work organization (see table 2.1).

More elaborately, the surveys were administered to consenting participants thrice over the course of the iGEM competition cycle. The surveys are organized as:

- First round of questionnaires (June) - with questions focusing on static individual attributes (demographics, skills), team attributes (team selection, role, subgroups) and personality questionnaires.
- Middle questionnaire (August) - asking dynamic individual (learning) and dynamic team attributes (team interactions)
- Concluding questionnaire (October) - dynamic individual (learning) and dynamic team attributes (team interactions) and questions regarding team problem resolution.

Over the course of the rounds of surveys, the participants were asked about their team interactions according to 17 data fields. Answering them involved nominating members of their team for each interaction type question. This facilitates the construction of *ego-centric networks* focused on each respondent and offering quantitative data to study interaction patterns at the team level. Table 2.1 describes the different interaction questions posed to the participants in the survey.

Interaction Question	**Key**
Who is in your work subgroup?	Subgroup
Who is a leader in your team?	Leader
Who is an advisor in your team?	Advisor
Whom did you receive mentorship from?	Mentor
With whom did you work more closely in the team?	Work Closely
Whom did you know previously?	Known Before
Who do you trust in your team?	Trust
Who do you think trusts you?	Trust You
Who do you empathise with in your team?	Empathy
Who do you think empathises with you?	Empathy You
You understand the team goal contributions of?	Work Understanding
Who shares knowledge with you?	Share Knowledge
Who helps you improve your skills?	Help You
You value the contributions of?	Value Work
Who enjoy working with you?	Enjoy Working
Who helps perform your tasks effectively?	Help You Efficient
Who helps everybody?	Helps Everybody

Table 2.1: Interaction questions described in the survey. The key presents a shorthand notation to refer to the corresponding question henceforth. Multiple nominations were allowed for the participants to members within their team. The responses were not shared with other participants.

Skills are iGEM specific, with participants given a multiple choice between a previously compiled list of technical (experimental biology, coding, writing etc.) and soft skills (communication, teamwork etc.). The questions focused on the skills they possess and what was required for the technical role within their team. The learning surveys aimed to report on skill improvement of respondents during the middle and at the end of the competition.

Participating iGEM teams were contacted through email, Skype and social media and directed to a web link for the completion of surveys at different stages of the competition. Teams were made aware of the privacy and data storage policies while onboarding to the study, which was approved by the IRB (Institutional Review Board). Analysis and visualizations of the team data was made available to the participating teams, albeit anonymized.

2.2 Data Cleaning

Participating members and teams are indexed throughout the study and different surveys through their iGEM and team IDs respectively, which are unique and provided by iGEM upon registration to the competition. This along with other team and user metadata (section, track, composition, demography, user role) are scraped from the iGEM database and mapped with the survey results. Primary tasks in cleaning the survey data included:

- Handling multiple responses from a participant to the same survey. Survey responses were prefixed with the timestamp of submission and in such cases, the latest submission of a participant was considered for analysis.

- Homogeneity in attribute representation (username etc.) through uniform formatting and matching between the surveyed participant information and data scraped from iGEM.
- Transforming individual survey responses to network data (elaborated below).

Each nomination for an interaction question is a directed edge from the respondent to the nominee. The interaction network for a given question is the aggregation of all ego networks of the respondents to that question. The suite of questions and ego networks now facilitate representation as a *multilayer network* - with 17 layers focusing on each interaction question with actors (nodes) preserved across layers. For the case of eliminating response bias - members who were nominated, but haven't participated in the study themselves have been removed from the final analysis.

2.3 Methods

Having discussed the construction and analysis of the survey data as a multilayer network, this section describes the metrics and measures which are used over the course of the report.

Jensen-Shannon distance: With density matrices ρ and σ of order N (in the context of graphs - the Laplacian of the adjacency matrix of dimension N), the Jensen-Shannon distance between them is computed as:

$$D_{JS}(\rho||\sigma) = h(\mu) - \frac{h(\rho) - h(\sigma)}{2} \tag{2.1}$$

where $\mu = \frac{\rho+\sigma}{2}$ and $h(\rho)$ the *von-Neumann* entropy of ρ, defined as:

$$h(\rho) = -\sum_{i=1}^{N} \lambda_i log_2(\lambda_i) \tag{2.2}$$

λ_i corresponds to i^{th} eigenvalue of ρ. [8]

Network Assortativity: The assortativity of a network evaluates the preference of nodes to be connected to *similar* nodes - here nodes with matching attributes. The coefficient of assortativity, introduced in [9], measures this level of homophyly through the categorical labels of vertices:

$$r = \frac{\sum_i e(i,i) - \sum_i a(i)b(i)}{1 - \sum_i a(i)b(i)} \tag{2.3}$$

with $a(i) = \sum_i e(i,j)$ and $b(i) = \sum_j e(i,j)$. i, j are variables for representing the node attributes. The assortativity score ranges from $[-1, +1]$ with -1 indicating complete heterophily (no edges between nodes with a matching attribute) and +1 showing complete homophily.

Graph/Network Rewiring: Rewiring the edges of a graph involves arbitrarily picking an edge and reassigning endpoints to a new pair of vertices chosen randomly. The graph considered here is directed and the rewiring performed is *degree-preserving*, where the degree sequence of the parent and rewired graph are the same, with degree referring to the *in-degree*. This is effected by picking two directed edges (a, b) and (c, d)

arbitrarily and replace them with edges (a,d) and (c,b), in the case these new edges do not exist in the parent graph. This way, multiple edges are avoided [10]

Hypergeometric Enrichment Test: The hypergeometric distribution describes the probability of events:

$$h(x,N,n,k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \tag{2.4}$$

where N is the population size, k the successes in the population, n the size of a random sample from the population and x the number of successes in it. This gives the probability of x successes in a random draw of n items from the defined population.

In the context of enrichment tests, the interest is to compute the probability of more than x successes over random samples of size n. Assuming there is a predefined sample of interest of size n with x successes (from a population of size N with k successes), a low probability (based on the level of confidence), argues that the sample of interest is *enriched* with successful objects.

Wilcoxon Signed Rank Test: A non-parametric alternative to the *t test*, when the conditions for the latter aren't met (t-test requires the differences between variables to be normally distributed). It is used to compare shifts of one population with respect to another - with a null hypothesis that the mean difference between the two pairs is zero.

Correlation: Unless mentioned otherwise, correlation scores mean the Pearson correlation coefficient - which measures the linear association between two variables. The value ranges from -1 (negatively associated) to +1 (perfect linear association).

Bonferroni Correction: A simple and conservative correction method applied when several statistical tests are performed simultaneously. This aims at minimizing the rate of false positives. In this method, the significance cutoff is set at α/n, where n is the number of tests being performed. Since this is done at the expense of false negatives, the cutoff is extremely conservative.

Hierarchical Clustering: Agglomerative clustering is used here, where initially each object is its own cluster. Iteratively, similar clusters are merged until there is a single cluster and at each stage, the distance between clusters are computed using *Ward's* method (refer R documentation of function *hclust* for more information).

System Configuration: As a footnote to this section, processed data were stored in .CSV format and all the analysis presented in this report were done using R and markdown on a 64 bit system running on Linux environment. The memory of the system was capped at 8 GB, with Intel® Core™ i5-4300U CPU processing capabilities and a disk size of 512 GB.

2.4 Data Summary

As defined in the network construction, participants of the study are nodes in the network and interactions are directed edges between the participant and the members of their team whom they have nominated. Participants to the study are alternately referred to as respondents to the survey in due course.

19 teams participated in the iGEM TIES study for 2019, with 79 respondents. Supplementary image 1 summarizes the key attributes of the respondents to the survey. A majority of participants have no prior iGEM participation ($\sim$ 76%), with their gender ratio at 55% to 45% - female to male. Participants are primarily from the country in which their iGEM team is based, though there is some diversity in the native language of participants ((d) and (e)). Supplementary image 3(b) establishes a relatively fair sampling from the pool of members of participating teams, with student leaders over represented in comparison to other team roles.

Supplementary image 2 highlights the attributes of participating teams. Key points include more successful teams (11 gold medal winners) and equal representation of under and overgrad sections. Teams *IIT-Madras* and *Munich* have a high ratio of respondents (Supplementary image 3(c)) and are hence studied in particular to understand skill contagion and learning over the course of the competition.

A total of 1151 unique interactions were recorded through the survey between various members of participating teams. All interactions were amongst actors belonging to the same team. Eliminating interactions with members whom haven't completed the survey themselves gives a result of 575 interactions. The unique interactions exist in atleast one of the 17 network layers

Supplementary image 4(a) shows the cumulative network of these interactions, with the edge weight being the number of network layers they are present in. The two significant components correspond to the teams with relatively complete information - IIT-Madras and Munich. 4(b) shows the log-log plot of the cumulative degree distribution of the network layers. The networks aren't completely *scale-free* - where the node degrees follow a power law distribution and offers key insights on network properties like the presence of hubs [11]. This is attributed to lack of completeness in responses within teams.

An initial observation on the responses to different interaction questions is the correlation of answers between them. The correlation scores are computed between layer edge activity vectors - vectors for each layer with the encoded information being the presence/absence of an edge in that layer [12] [13]. Supplementary figure 5(a) shows the heatmap of the correlations (with insignificant correlations cutoff using Bonferroni's correction). There is a signal for questions with similar responses and a potential clustering of questions - implying merging network layers. An entropy based method for clustering layers of a multilayer network is presented in [8], where distance between networks is computed using the *Jensen-Shannon* distance.

Hierarchical clustering of the distance matrix is performed (5(b)) with aggregation of the closest layers at each step. A quality function computes how distinguishable the clustering is from the entire aggregated network (5(c)). The maximum value highlights the best clustering scheme, represented by the dotted line in figure 5(b).

Key observation from the clustering of layers is the formation of broad question groups: Social, Work and Mentoring. This takeaway can be leveraged to augment survey structure and composition for the study in the forthcoming editions of iGEM. In addition, this enables the categorization of respondent activity, portrayed by the cluster in question:

- **Social**: Empathy
- **Work**: Work understanding, enjoy working, value work, Helping you, helping everybody, work closely, share knowledge, trust
- **Mentorship**: Mentor, advisor

This is particularly interesting, as mentioned previously in identifying categories of interactions as well as to contrast participant activity between said categories. This is explored further on in the report.

Finally, The skill attributes of participants are shown in Supplementary image 6 as an affiliation network. Skills and role specific skills are primarily technical with experimental biology being the most cited skill. Of the qualities, communication and teamwork are the most reported. Participants could cite multiple skills for each category. Supplementary Image 7 shows the distribution of skill by the participant role. Clearer insights on the role specific skills, where mentorship, public engagement are some skills which are primarily attributed to senior roles. This is interesting in the context of understanding enrichment of skills and knowledge transmission within a team.

Chapter 3

Results

The main results from analysis of the survey data are reported in this chapter. The methods used for analysis are described with some key observations. The chapter is divided into four sections - focusing on the effects of participant role, experience, social and skill attributes in understanding the collaborative activity within iGEM teams.

3.1 Interaction Patterns by Team Role

The predefined hierarchy within iGEM teams, structured through team member roles, raises the question to measure the rate of interaction amongst them. This is computed as a ratio between the observed number of interactions between members of different categories to the expected number of such interactions.

First, for each layer of the interaction networks, a node compression is performed by mapping nodes having the same role into one. The compressed network now has four nodes corresponding to PIs (Primary and Secondary PIs are identified as a single node), Advisors (Advisors and Instructors), Student Leaders and Student Members. The weight of edges between these mapped nodes correspond to the number of interactions in the actual network. It is to note that interactions between same role classes are preserved as weighted loops in the compressed network. For two given roles i and j, the strength of directed interaction from i to j is computed as:

$$Strength(i,j) = \frac{Weight(i,j)}{P(j)Out(i)}$$

where $Weight(i,j)$ is the weight of interactions between i and j in the compressed network, $P(j)$ the probability of a node being in class j and $Out(i)$ the outdegree of node i. The logarithm of the *strength* of an edge in the compressed network is an indicator of the over or under representation of interactions between different classes of nodes - here their role. The observations are summarized in Figure 3.1 (a).

There is clear evidence of the existence of a hierarchy in interaction between members of different roles within the team:

- **Mentorship:** Typical idea of mentorship would be with junior members nominating senior members of the team. There are some refutes though - with over-representation of interaction between student members to advisors and PIs, but student leaders only with PIs. Advisors and PIs tend to nominate other advisors and PIs.

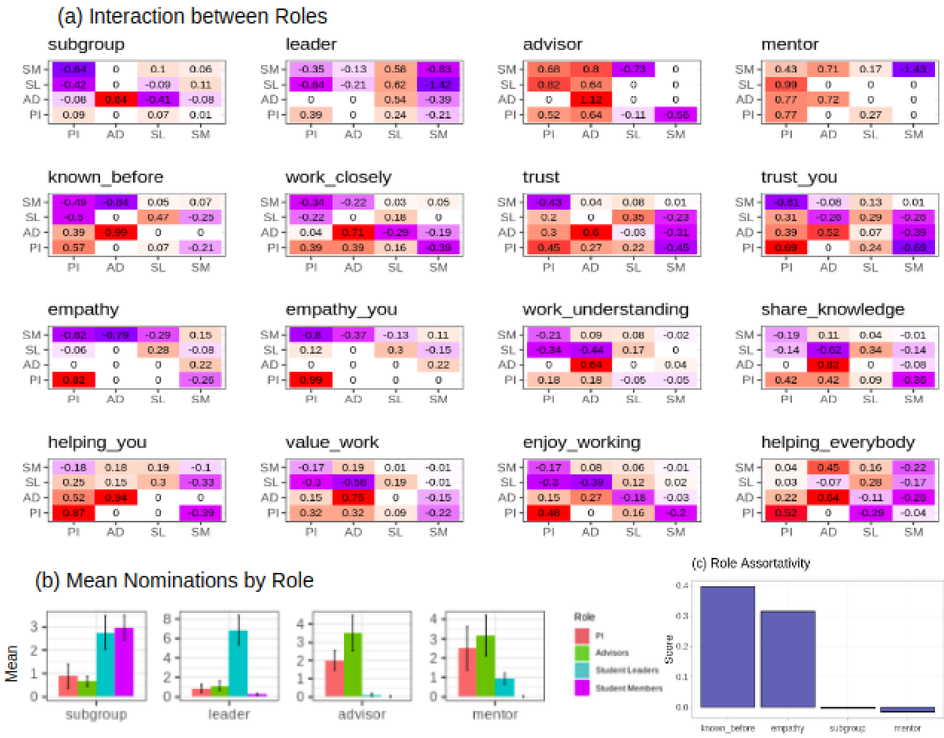


Figure 3.1: (a) Heatmap of Ratio of actual to expected interactions between participants of different roles. (b) Mean Nominations by role for select network layers with one standard error. (c) Network assortativity with participant role as key attribute.

- **Social Interactions:** Social layers like *empathy* offer insights on the circles of comfort within teams. One such circle is amongst the students (leaders and members). This agrees with the over representation of interactions between them and the opposite on interactions with superior members. An interesting observation is the over-representation of interactions between advisors and student members - which can be explained to a fair extent by close association between the two roles at the mentorship level.

- **Workflow:** In the network layers - *value work, work understanding* and *enjoy working* - interactions between student leaders to PIs and advisors are under-represented. This suggests at a level of autonomy in the functioning of student leaders' role within the team. PIs are increasingly involved in the activity with Advisors and student leaders, while student members increasingly nominate advisors in comparison to other role categories.

- **Helping and knowledge sharing:** The patterns in this layer follows the hierarchy of superiority - for each role, there is an over-representation of interactions with immediate superior roles - albeit with student leaders being more involved with PIs than advisors. The general consensus of over-representation of superior roles in the helping everybody layer also supports this claim.

With insights on team hierarchical organisation, the next study is on the effect of team role in driving interactions. The mean in-degree is computed for each role in the iGEM team. This is summarized for significant networks in Figure 3.1 (b). The observations agree with basic consistency checks in the form that on average, mentorship is primarily provided by advisors and then PIs, whereas student leaders take up leadership roles within the team. It can also be noted that subgroups are primarily comprised of students. The effect of participant role in driving interactions in these layers is well established.

Figure 3.1 (c) shows the assortativity of the interaction networks based on role as the associated node attribute. Extending with the sanity checks, mentor and subgroup networks show low role based assortativity - highlighting the heterogeneity of such interactions. The possibility students knowing other students and PIs/advisors knowing other participants of the same role explains the high role based assortativity in the *known before* layer and to some extent, the shared social status associated with having a similar role with the team.

3.2 Interaction Patterns by Experience

Prior experience in iGEM is an important attribute surveyed from the study participants - as a numeric data point determining the number of times they have participated in iGEM before. This prompts a measure of interaction amongst participants factoring in their experience in the competition. Since the majority of the respondents have no prior participation (Supplementary image 1(b)), the data is folded into a binary variable - 1 for prior participation and 0 for none.

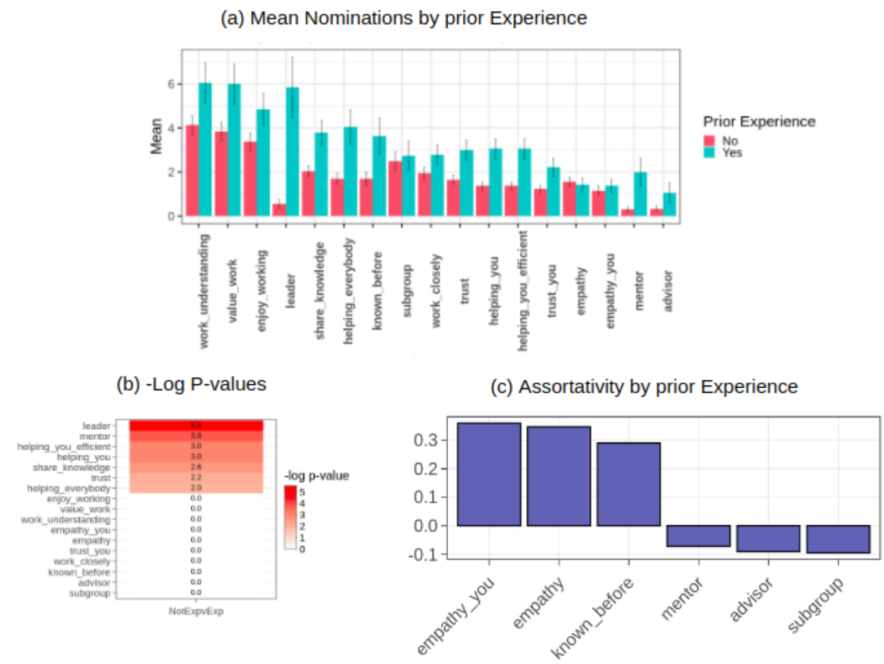


Figure 3.2: (a) Mean in-Degree by prior iGEM experience, (b) -Log P values (base 10) and (c) Network Assortativity based on iGEM Experience

The mean in-degree of respondents is calculated across the interaction network layers, between participants having and not having any prior iGEM experience. The results are summarized in Figure 3.2 (a). The differences in the in-degree of respondents of the two classes are compared using the Wilcoxon signed rank test. Negative log (base 10) of the p-value (with a cutoff of 2) is computed and shown in Figure 3.2 (b).

The effect of experience in being a *driving* factor for interactions amongst participants within a team is hypothesized. Senior members of a team are typically more experienced in the competition (Supplementary Image 3(a)) and this explains the difference in the *leader* and *advisor* layers in favour of the experienced participants. However, it is evident in the *helping you* and *share knowledge* layers that prior experience drives interactions - again with experienced members gaining more nominations on average. On the other hand, social layers like *empathy* and *empathy˙you* have an indistinguishable difference between nominations based on experience - asserting that this is not a factor for inducing social interactions.

Network Assortativity based on iGEM experience is presented in figure 3.2 (c). The higher score for the social layers (*empathy*, *known before*) could be due to students members forming a bulk of the *inexperienced* iGEM participants and the senior role members the *experienced* campaigners. The negative score for *advisor* and *mentor layers* shows a signal (though not strong) for inexperienced members seeking mentorship from the more experienced members within the team. This further strengthens the claim for experience in driving mentorship associations.

3.3 Social and Skill Assortativity

Revisiting network assortativity, a key question is to test the effect of shared social attributes in driving interactions amongst participants. The surveyed attributes considered here are gender, native language and nationality (Supplementary Image 1(c,d,e)). Figure 3.3 shows the assortativity score for the network layers based on the corresponding attributes.

Gender offers a clear signal towards the question being tested. The *social* layers - empathy and trust show evidence on a higher assortativity score,

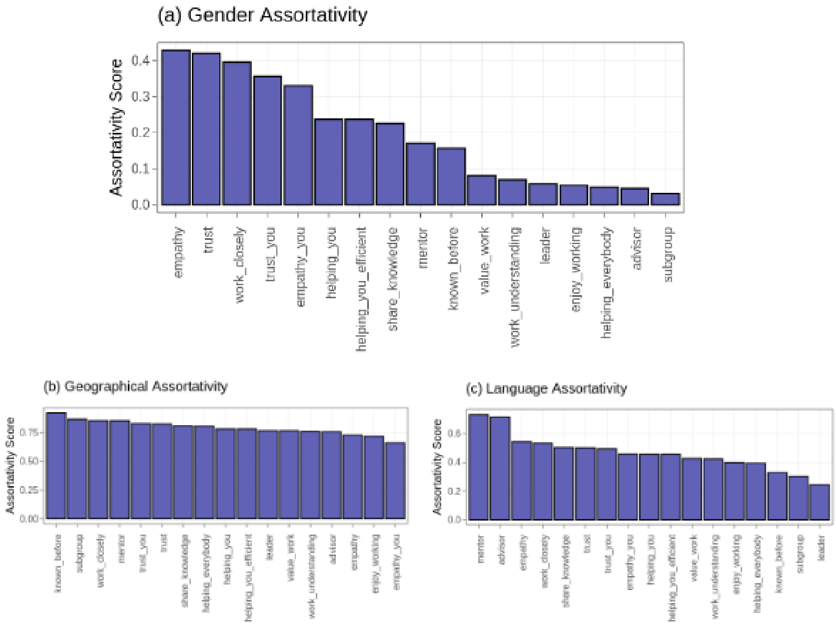


Figure 3.3: Assortativity score for the network layers based on (a) Participant Gender, (b) Geographic location and (c) Mother Tongue (Native Language).

highlighting gender being a driving factor in emotional interactions. On the other end of the spectrum, subgroups show little or no assortativity, indicating work subgroups are heterogeneous and agnostic to the gender of its members.

The geographic location of respondents conveys little information, with participating teams primarily composed of members based on the same locale as the team. On the other hand, the participants' native language shows a key observation - with higher assortativity in the mentor and advisor layers. This is particularly interesting, hinting that the shared attribute of a common tongue drives mentorship interactions.

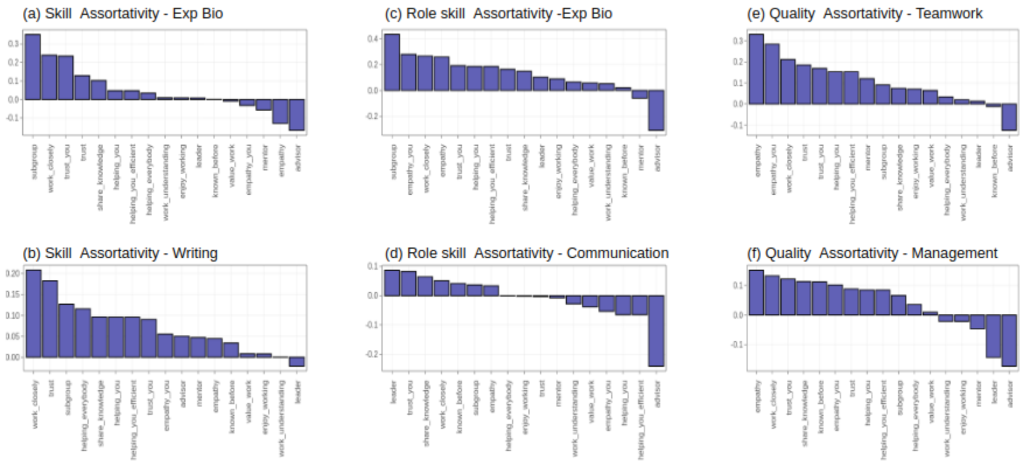


Figure 3.4: Network assortativity score for Skills - (a) Experimental Biology and (b) Writing; Role specific skills (c) Experimental Biology and (d) Communication; Qualities (e) Teamwork and (f) Management

The focus is now on skill attributes. The top skills of participants are identified (technical and role-specific skills) and the network assortativity is computed to check the hypothesis of whether having similar skills drive interactions. The assortativity scores are shown in Figure 3.4

The assortativity scores do not report clear observations on skills being factors for driving work interactions. This could be because of multiple skill nominations from a majority of participants, which makes it hard to properly account for the assortativity scores. But on a crude level, skill based assortativity scores are low for social interaction layers. Subgroup layer has a high score for the skill experimental biology ((a) and (c)) - indicating signs of subgroup task division - which is reported further in the next section. It is also to note that experimental biology is the most cited skill by the participants as well as key skill in iGEM.

Skill	P-Val Cluster 1	P-Val Cluster 2
Experimental Biology	*0*	*0.998*
Modelling	*0*	*0.568*
Data Analysis	*0*	*0.568*
Writing	*0.439*	*0.175*
Coding	*0*	*0.568*
Public Engagement	*0.883*	*0.012*
Communication	*0.880*	*0.011*
Graphic Design	*0.754*	*0.024*
Safety and Security	*0*	*0.730*

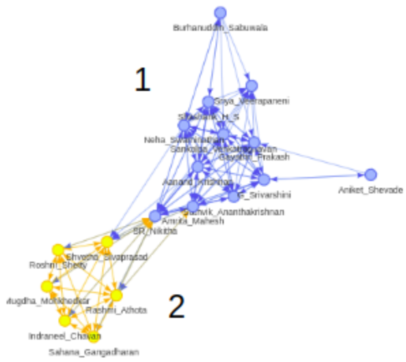


Figure 3.5: Hypergeometric p-values for skill enrichment in the subgroup structure of team IIT-Madras. Right is the identifier for the cluster code.

Mentor and advisor layers have a slight negative assortativity score (a), (d), (f). A subset of these skills such as communication and public engagement are focused skills for senior team members (refer supplementary image 7), explaining this observation. For the qualities Teamwork (e) and Management (f), there is a marginal higher score for the empathy and work closely layers. This suggests at evidence for soft skills playing a role in driving emotional interactions and work associations.

3.4 Skill Enrichment and Learning

One of the early questions in the study with ties to organisation is whether team structure is effected by technical capabilities of its members. With *IIT-Madras* being the team with the highest proportions of respondents, their subgroup structure is focused on for the analysis presented here.

The subgroup structure for team IIT-Madras shows a clear division of nodes into two groups. With this in mind, the hypergeometric p-value is computed to check for enrichment of skills within either subgroup. In the current context, for each subgroup in the network, x denotes the number of skilled individuals within that subgroup, the sample size n the size of the subgroup, N the size of the entire team and k the total number of skilled individuals.

Figure 3.5 shows the p-value scores for the different skills (Refer supplementary image 8 for a more visual representation). Subgroup 1 is enriched with technical skills or rather *hands-on skills* - mainly experimental biology, data analysis, modelling and coding, whereas the primary enriched skills in subgroup 2 are public engagement, communication, graphic design and to some extent writing. This suggests a division of iGEM specific tasks within subgroups - with group 1 focusing on wet lab work and associated analysis (factoring in safety and security practices). On the other hand, group 2 can be hypothesized to associate themselves with tasks like maintaining the wiki. This division of workflow may be fluid, but offers a fundamental understanding of work group separation in team IIT Madras, piquing the interest to make relevant analysis to other iGEM teams.

Learning surveys were handed out during the middle and at end of the iGEM competition cycle. They focused on identifying the skills and qualities the respondents had learned/developed over the course of the competition.

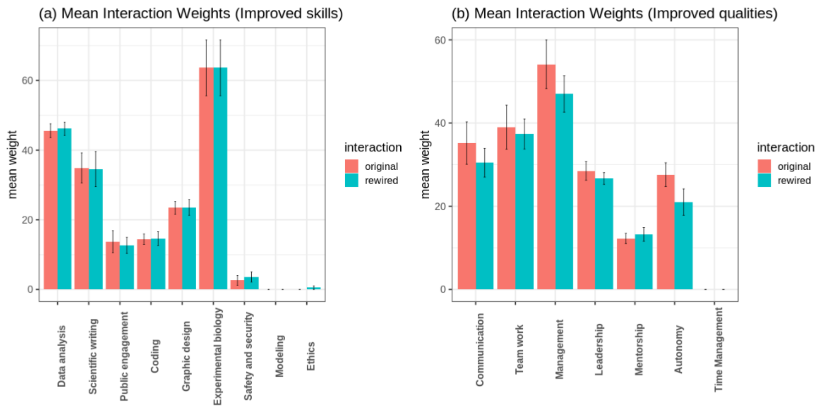


Figure 3.6: Mean interaction with skilled individuals for respondents who reported a learning - (a) Technical Skills and (b) qualities when compared to a rewired network of team Munich.

Due to a lack of responses for the middle learning survey, the responses from the end of cycle survey are considered for the analysis. This time, the network structure of team *Munich* is studied - due to their higher response rate for these surveys in comparison to other teams.

Figure 3.6 shows the weighted mean interactions to a skilled member within team Munich. The aggregate network of interactions is considered with the weight of each interaction being the number of layers in which the edge exists. This is measured for participants who have claimed that they had an improvement for the skill in consideration. To facilitate a comparison, a rewired cumulative network is constructed for team Munich. There is no reported significant difference in the interaction with skilled individuals in comparison to the rewired network ((a) and (b)).

Another question that was posed to the participants along with the learning surveys is whether they would have liked to work with a member of their team more (referred henceforth as *worked more*). Supplementary image 9(a) shows the degree correlation between the different network layers and the *worked more* network. There is a higher score for the social layers like *empathy* and *trust*. Emphasis can be placed on the effect of *social comfort* in driving a potential future interaction. This is discussed further in the next chapter.

9(b) shows the edge overlap between *worked more* and the other network layers. With a higher ratio of overlap with *enjoy work*, *value work* and *work understanding* layers, there is a driving factor to work with someone with whom there already is a positive affiliation with regard to work.

3.5 Questions Perception

One aspect of the 2019 study was to analyse and understand the survey reception by the participants and aim at optimising its structure. One method of testing this is by computing correlations of edge activity vectors for the two teams with a high fraction of respondents - *IIT-Madras* and *Munich*. Hierarchical clustering of the correlation scores is performed with the observations summarised in Figure 3.7. The insignificant correlations are cut off using the Bonferroni correction (shown in white).

The way network layers are clustered highlights differences in the perception of questions at a team scale. In the case of team IIT-Madras, *trust* is close with layers *helping you* while *empathy* is closer to *work closely*. On the other hand for team Munich, *trust* is close to *work closely*

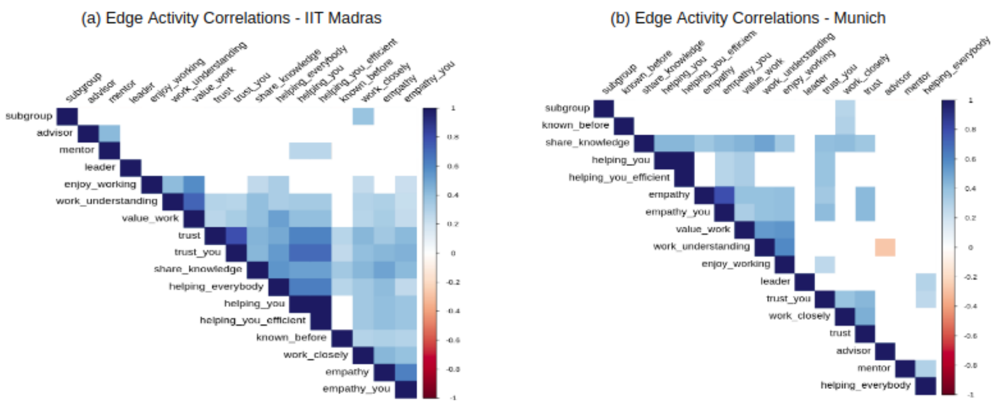


Figure 3.7: Edge activity correlations of the teams (a) IIT Madras and (b) Munich across the different network layers.

whereas *empathy* is closer to *value work, work understanding* layers.

On one hand, while this offers insights on which interactions are key in forming social ties within a team, it also suggests at a variance in the perception of the survey questions at a team level. Questions such as *Whom do you trust in your team?* could be inferred either on an academic or social setting. With this in mind, one of the main takeaways is to target at redesigning focused survey questions.

Chapter 4

Discussion

This chapter tries to piece together the hints and evidences presented in Results chapter to try and streamline the conveyed story. With the data not definite enough to come to strong conclusions on categorising the dynamics of collaborative activity within scientific teams, the focus is to summarize the primary observations of the iGEM TIES study 2019 to better facilitate the study of scientific teams for the upcoming editions of the competition.

4.1 Hierarchy in Teams

The first set of observations to highlight are surrounding the hierarchy in the workflow within the team. Krackhardt [14] suggested on the dimensions of organisational structure - with connectedness and graph hierarchy relevant in the current context. With an inherent hierarchy defined on iGEM teams, the review was on the former - connectedness amongst different team roles, which was presented in section 3.1.

The organisational system of iGEM teams provide insights on how they function and work on tasks. On a broader scale - evidence points to under representation of ties between student members and PIs. Student members look to advisors for mentorship and help, establishing a close association of advisors in the regular functioning of iGEM teams - since student members comprise a bulk of the team workforce.

Student leaders on the other hand have a notable under representation of interactions with advisors, suggesting a level of autonomy in their activity, which can be attributed to the fact that most student leaders in the study have prior iGEM experience. Advisors have close work associations with the PIs, and a lack of represented interactions with student leaders, with more hints on their role in effecting the activity of student members. PIs primarily have links with other PIs and advisors, but the link between PIs and students is directed through the student leaders.

The birds view of the system suggests a lateral structure of activity - PIs at the root, performing a supervisory role with workflow paths between PI $\rightarrow$ Advisor $\rightarrow$ Student Members and PI $\rightarrow$ Student Leaders. The key extension to this would be how the organisational structure differs for successful teams in comparison to unsuccessful ones, with success defined in the context of medals and awards in the competition. With the observations of iGEM 2019 as a base, the aim is to expand on the analysis of interaction hierarchy and teamwork to model how effectively it determines success in the context of team science.

4.2 Social Homophily

There have been research on understanding what makes work groups creative. Paulus [15] cites *social inhibition* as a primary factor for low creativity in brainstorming groups. This social inhibition is a product of a multitude of reasons - with lack of recognition, negative comparisons and anxiety being some. Edmondson [16] highlights that psychological safety, is crucial for the learning behaviour of an individual in a group setting. With the definition of psychological safety closely associated with whether an individual is comfortable taking decisions within their work environment, there is a motivation to investigate the interpersonal dynamic of interaction amongst teams.

The analysis hints at the existence of a *safe space* within a team. First is the formation of social interactions which are primarily between members of the same role. This circle of comfort is further strengthened by the high assortativity of participants towards the social attributes tested here - gender, geography and language. With evidence that the shared social attributes drive emotional, and to some extent, work interactions, it is interesting to contrast how the formation of comfort circles within a team effects the way they collaborate across other interaction dimensions.

[17] suggests the collective intelligence of a group is positively correlated with the *average social sensitivity* of the group members. The target would be to quantify this in the context of iGEM teams - and test its effect on the team outcome.

4.3 Skill Contagion

The crux of iGEM is to inspire innovation with collaborative problem solving. This is emphasized in their evaluation criteria (reusing biobricks, partnering with other teams amongst others). The learning curve of participants in the study is significant, with a focus of the iGEM ties study to understand this through interactions amongst team members. Section 3.2 showcases the significance of experience in driving mentorship and interactions resulting in information contagion.

The enrichment of skills in subgroups - in addition to showcasing potential work distribution and functioning - hints at a platform for learning. With skills enriched in a work subgroup, it can be hypothesized on the effect of regular interaction with skilled members, resulting in learning process. The skill based network assortativity, suggests (loosely) at signals for affiliation amongst participants - with a positive score hinting at enriched work groups and negative scores at learning interactions.

Despite this, there is neither significant evidence of improvement of the learning outcome (the case of team Munich), nor, the observations of shared skills in driving interactions at different dimensions. Nevertheless, this remains an important question to be tested with more comprehensive data sourced from several iGEM teams.

4.4 Concluding Remarks

The iGEM TIES study for 2019, mainly a pilot study, has shown some relevant signals to understanding the effect of interactions in team organisation, learning and success. The key results - on team hierarchy, effective driving factors for interactions, existence of social sensitivity and skill enrichment provides a platform to build upon with regard to the study of team science. One aspect of iGEM teams not leveraged in the analysis is the contributions of participants to their team wiki - which is a significant indicator of collaborative activity within a team. This is a key immediate objective.

The study for the upcoming edition for iGEM targets at collecting physical interaction data through a Bluetooth based mobile application. In addition, participant activity journalling is touted as the primary source of collecting nomination based interaction data. Both methods introduce a temporal factor to the recorded interactions and this is an intriguing prospect, to test the observed hypotheses on richer data source.

Pruning of the survey questionnaire based on the prior responses is another proposed outcome. Lower response rate for a majority of participating teams, coupled with clustering and categorization of questions into significant groups offer a way to obtain qualitative data through redesign of interaction questions and making it less cumbersome for participants to report.

Extending with the survey design and observations on question perception, structuring the questions posed is also important. This is something that is being worked on in parallel with the design of the journalling module of the mobile application. Putting all together, there is a significant target and scope to extend the analysis presented in this report to study team science using iGEM as a rich, fine-grained testbed.

Bibliography

[1] S. Wuchty, B. F. Jones, and B. Uzzi, "The increasing dominance of teams in production of knowledge," *Science*, vol. 316, no. 5827, pp. 1036–1039, 2007.

[2] L. M. Bennett and H. Gadlin, "Collaboration and team science: From theory to practice," *Journal of Investigative Medicine*, vol. 60, no. 5, pp. 768–775, 2012.

[3] S. M. Fiore, "Interdisciplinarity as teamwork: How the science of teams can inform team science," *Small Group Research*, vol. 39, no. 3, pp. 251–277, 2008.

[4] Y. N. Lee, J. P. Walsh, and J. Wang, "Creativity in scientific teams: Unpacking novelty and impact," *Research Policy*, vol. 44, no. 3, pp. 684–697, 2015.

[5] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-l. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. V. Alstyne, "Computational Social Science," vol. 323, no. February, pp. 721–724, 2009.

[6] M. Klug and J. P. Bagrow, "Understanding the group dynamics and success of teams," *Royal Society Open Science*, vol. 3, no. 4, 2016.

[7] "International genetically engineered machine (igem) - main page." https://igem.org/Main˙Page.

[8] M. De Domenico, V. Nicosia, A. Arenas, and V. Latora, "Structural reducibility of multilayer networks," *Nature Communications*, vol. 6, pp. 1–9, 2015.

[9] M. E. Newman, "Mixing patterns in networks," *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, vol. 67, no. 2, p. 13, 2003.

[10] "R igraph package - documentation." .

[11] A.-L. Barabási *et al.*, *Network science*. Cambridge university press, 2016.

[12] V. Nicosia and V. Latora, "Measuring and modeling correlations in multiplex networks," *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 92, no. 3, pp. 1–20, 2015.

[13] F. Battiston, V. Nicosia, and V. Latora, "Structural measures for multiplex networks," *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 89, no. 3, pp. 1–16, 2014.

[14] D. Krackhardt, "Graph theoretical dimensions of informal organizations," 1994.

[15] P. Paulus, "Groups, teams and creativity," *Journal of the Academy of Marketing Science*, vol. 28, pp. 109–119, 2000.

[16] A. C. Edmondson, “Psychological safety, trust, and learning in organizations: AGroup-level lens,” *Trust and Distrust in Organizations : Dilemmas and Approaches*, no. December, pp. 239–272, 2004.

[17] A. W. Woolley, C. F. Chabris, A. Pentland, N. Hashmi, and T. W. Malone, “Evidence for a collective intelligence factor in the performance of human groups,” *Science*, vol. 330, no. 6004, pp. 686–688, 2010.

[18] I. Kontro and M. Génois, “Combining surveys and sensors to explore student behaviour,” *Education Sciences*, vol. 10, no. 3, 2020.

[19] D. Barkoczi and M. Galesic, “Social learning strategies modify the effect of network structure on group performance,” *Nature Communications*, vol. 7, pp. 1–8, 2016.

[20] S. Deri, J. Rappaz, L. M. Aiello, and D. Quercia, “Coloring in the links: Capturing social ties as they are perceived,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 2, no. CSCW, 2018.

[21] M. T. Rivera, S. B. Soderstrom, and B. Uzzi, “Dynamics of Dyads in Social Networks: Assortative, Relational, and Proximity Mechanisms,” *Annual Review of Sociology*, vol. 36, no. 1, pp. 91–115, 2010.

[22] “What google learned from its quest to build the perfect team.” https://www.nytimes.com/2016/02/28/magazine/what-google-learned-from-its-quest-to-build-the-perfect-team.html.

Supplementary Images

Supplementary Image 1

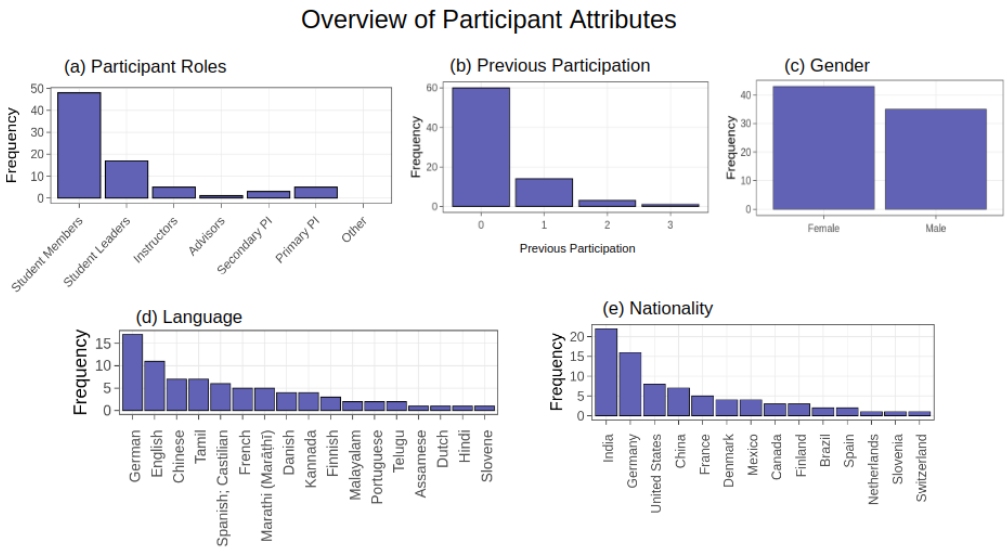


Figure 4.1: Attributes of the survey respondents. (a) Role in iGEM team. (b) Prior participation in iGEM. (c) Gender. (d) Mother Tongue (Language with native proficiency of participants). (e) Nationality.

Supplementary Image 2

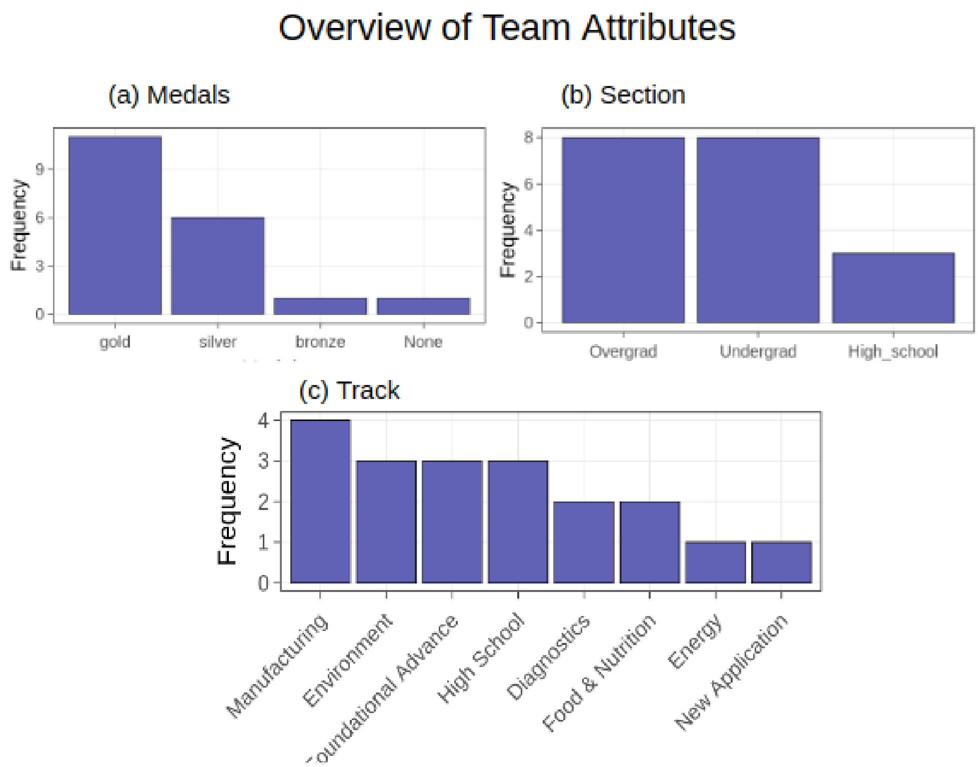


Figure 4.2: Attributes of responding teams. (a) Medals achieved in iGEM 2019. (b) Section and (c) Track of participation in iGEM 2019.

Supplementary Image 3

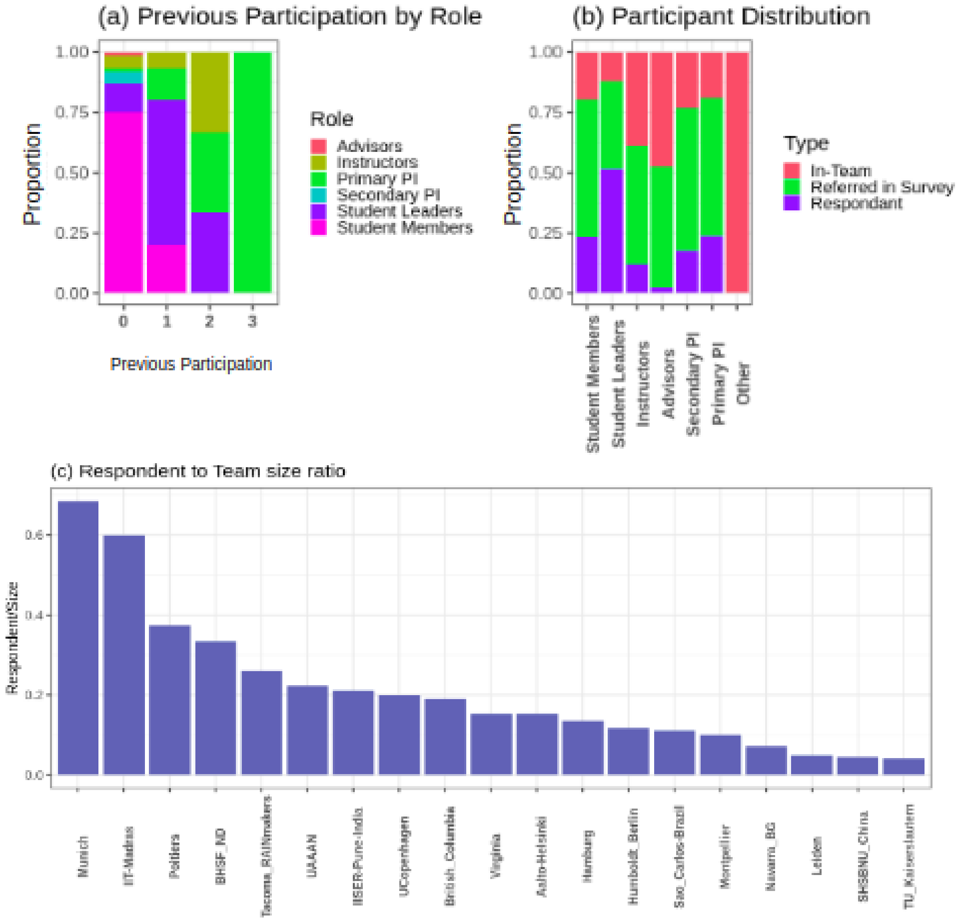


Figure 4.3: (a) Which team roles are the one with prior iGEM experience? The ratio of previous participation by members' role in the team. (b) By role, what fraction of a team have responded to the survey (Respondent), didn't respond but were referred by someone who responded (Referred in Survey) and belonged to the team but were not referred (In-Team).(c) Ratio of team size to responding members.

Supplementary Image 4

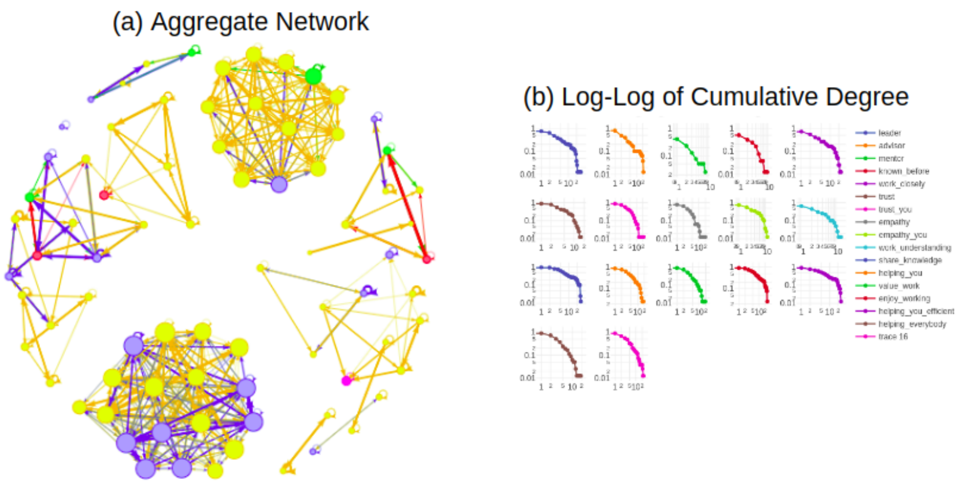


Figure 4.4: (a) Aggregate network over 17 network layers. Node size corresponds to in-degree and color to the iGEM team role (Yellow - Student Members, Blue - Student Leaders, , Pink - Advisors, Red - Secondary PI, Green - Primary PI). Edge weight is the number of interactions over the network layers. (b) Log-Log plot of cumulative degree distribution of each network layer.

Supplementary Image 5

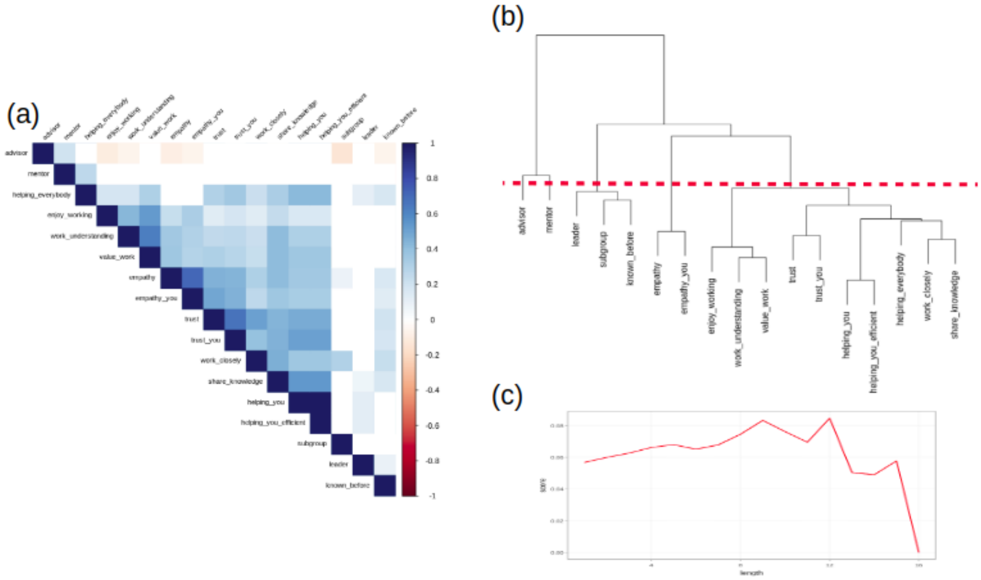


Figure 4.5: (a) Correlation between network layers based on edge presence. (b) Dendogram for clustering network layers (c) Entropy of the network throughout stages of hierarchical clustering. Dotted line corresponds to the point of maximum entropy.

Supplementary Image 6

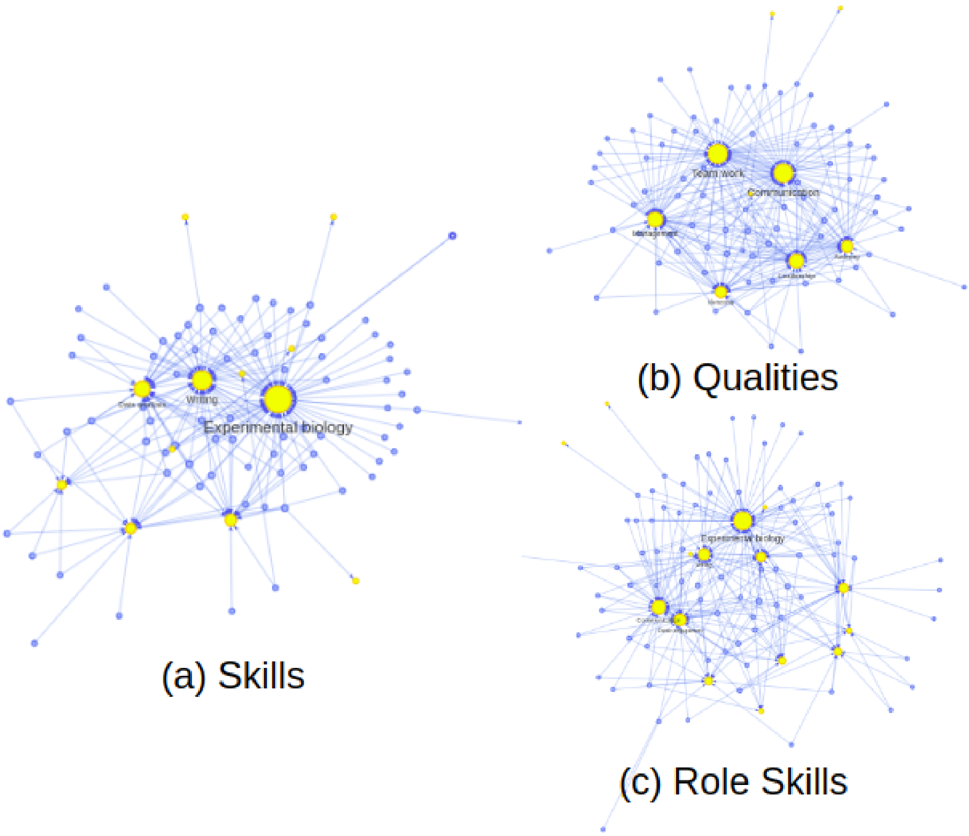


Figure 4.6: (a) Skill affiliation network (b) Qualities (Soft Skills) affiliation and (c) Role specific skill affiliation. Node size is the number of nominations for that particular skill in question, with skill nodes shown in yellow.

Supplementary Image 7

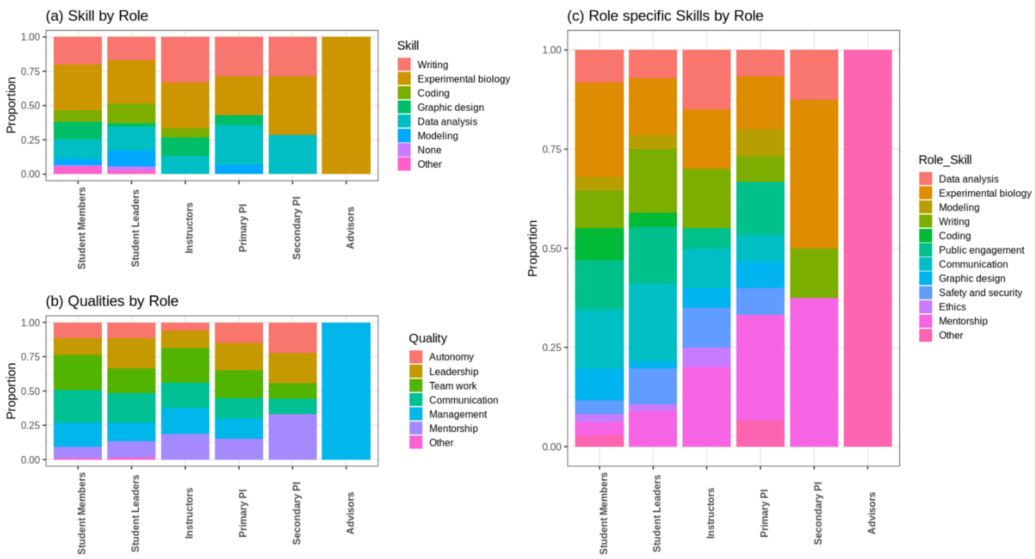


Figure 4.7: Distribution of skills by participant role in the iGEM team. (a) Technical Skills, (b) Qualities and (c) Role specific skills.

Supplementary Image 8

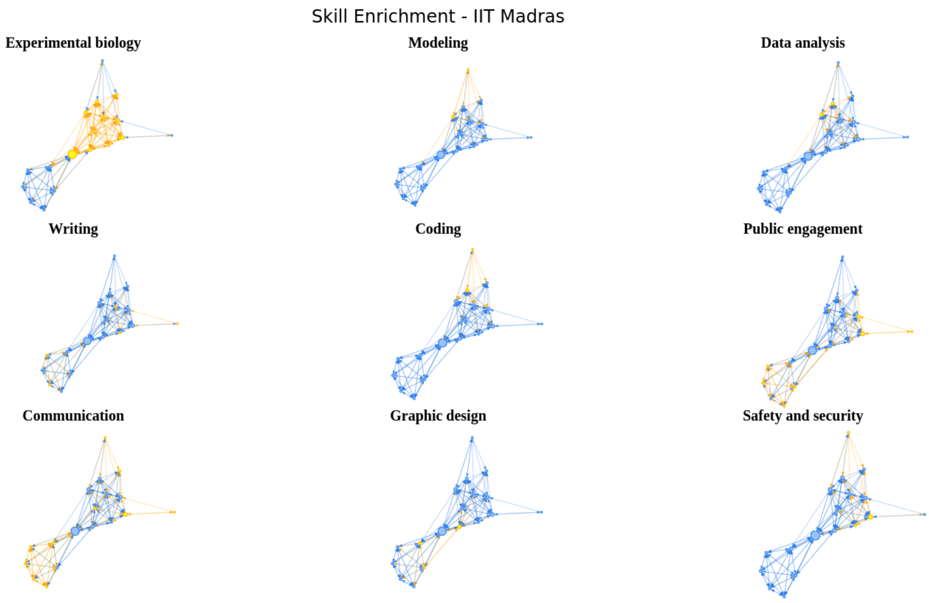
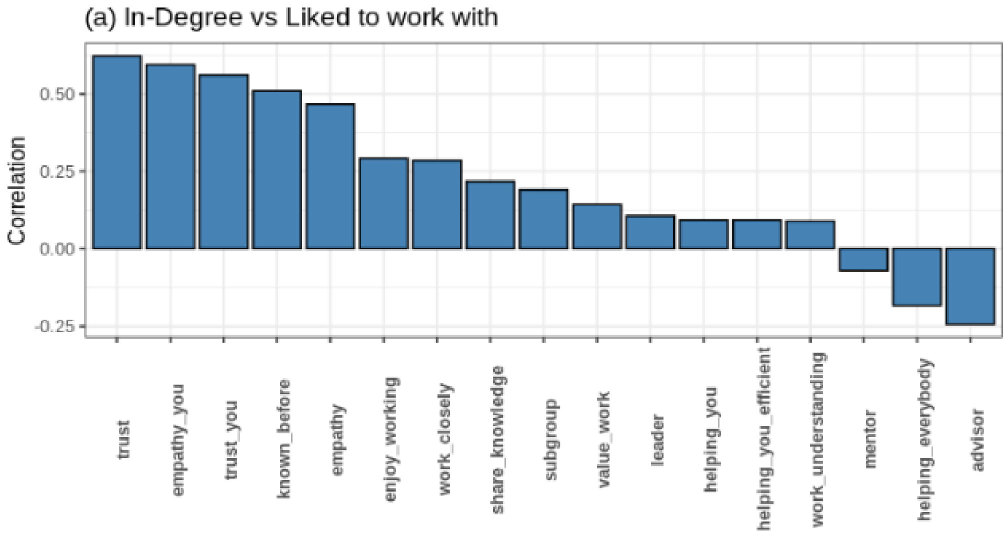


Figure 4.8: *Subgroup* Network layer for team IIT-Madras for select role-specific skills. Highlighted nodes (yellow) at each layer depicts the nodes possessing the mentioned skill.

Supplementary Image 9



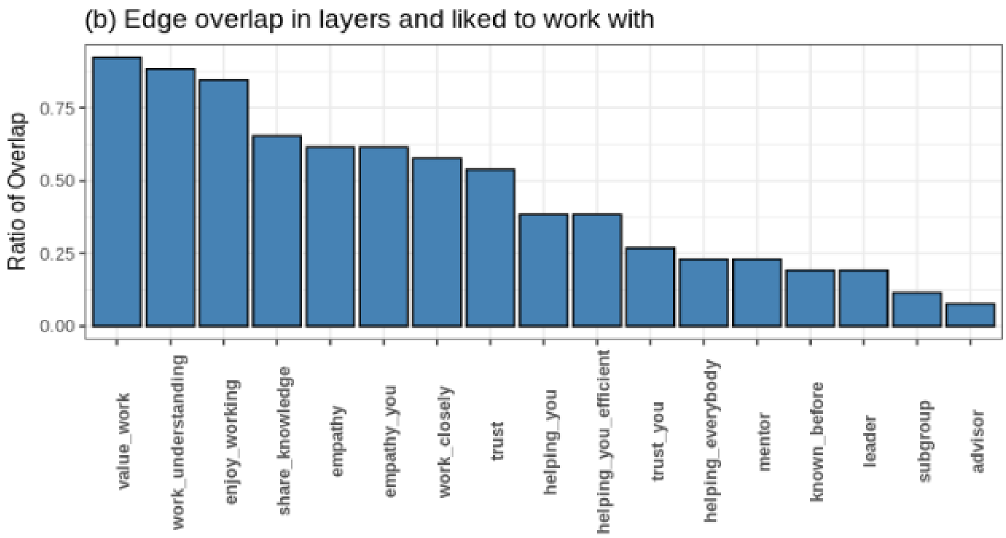


Figure 4.9: Responses to the question "Who would you have liked to work more with?" for team Munich. (a) Correlation between node degree in the interaction network layers and nominations for this question. (b) Nomination (edge) overlap between the layers and the question in hand.