



HAL
open science

Network Approaches to Reconstruct and Analyse Team Collaborative Structures in Open Science and Innovation

Rathin Jeyaram

► **To cite this version:**

Rathin Jeyaram. Network Approaches to Reconstruct and Analyse Team Collaborative Structures in Open Science and Innovation. Computer Science [cs]. Université Paris Cité, 2024. English. NNT : . tel-04847989

HAL Id: tel-04847989

<https://theses.hal.science/tel-04847989v1>

Submitted on 19 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Université
Paris Cité

UNIVERSITÉ PARIS CITÉ

Frontière de l'Innovation en Recherche et Education (ED474)

Learning Planet Institute (ex. CRI) UMR1284

Network Approaches to Reconstruct and Analyse Team Collaborative Structures in Open Science and Innovation

Par RATHIN JEYARAM

Thèse de doctorat de MATHÉMATIQUES ET SCIENCES INFORMATIQUES

Dirigée par MARC SANTOLINI

Co-supervisé par MEGAN PALMER

Présentée et soutenue publiquement le 16/12/2024

Devant un jury composé de :

PRIMAVERA DE FILIPPI	DR, Université Paris - Pantheon Assas	Examinatrice
FLORIANA GARGIULO	CR, Sorbonne Université	Examinatrice
OLGA KOKSHAGINA	Associate Professor, EDHEC Business School	Membre invité
ARIEL LINDNER	DR, Université Paris Cité	Examinateur
THOMAS MAILLART	Senior Lecturer, Université de Genève	Rapporteur
MEGAN PALMER	Adjunct Professor, Stanford University	Membre invité
MARC SANTOLINI	Researcher, Learning Planet Institute	Directeur
HENRY SAUERMAN	Professor, ESMT Berlin	Rapporteur

Abstract

Title: Network Approaches to Reconstruct and Analyse Team Collaborative Structures in Open Science and Innovation

Modern science and innovation challenges require the integration of diverse perspectives alongside specialised knowledge, prompting a shift towards interdisciplinary and collaborative team-based research. This transformation towards collective knowledge production continues to uphold high-quality outputs and research impact, as evidenced by frameworks like challenge-based innovation and crowd science. Understanding how larger and more diverse teams collaborate, divide labour and coordinate tasks requires a mixed-method approach, combining qualitative and quantitative data to capture their structural properties. Yet, a lack of longitudinal and granular datasets has limited our understanding of how collaborative structures evolve and influence team performance. In this thesis, I leverage network science methods to study collaborative structures in challenge-based, open science and innovation contexts. First, I study citizen science teams addressing Sustainable Development Goal challenges, identifying collaboration and engagement patterns through surveys and digital traces from Slack communications. I demonstrate how these patterns inform team performance, emphasising the importance of sustained engagement with individuals across diverse network locations. Second, I investigate the organisation dynamics of 1,200+ unique teams participating in the iGEM interdisciplinary science and engineering competition across 12 years, analysing their division of labour and coordination using self-reported attribution statements and digital traces from unstructured, collaboratively edited, online wiki pages. I introduce a novel computational method for inferring task allocation networks of teams at scale using Large Language Models, demonstrating its ability to recover local and global network properties with high precision. I argue that network-based indicators such as modularity and nestedness capture specialisation and coordination within division of labour structures, and show that they significantly impact team performance in the competition. Furthermore, I show that teams adapt their division of labour over time, converging towards optimal structures mediated by their size, experience and task complexity. Broadly, this thesis contributes to research on team science in open innovation and science ecosystems by introducing data- and network-driven approaches to capture organisational structures, offering insights for researchers and practitioners into team performance, engagement, and adaptation, thereby supporting the development of strategies to foster innovation in collaborative research.

Keywords: Team Science, Network Science, Collaboration Networks, Task Allocation Structures, Bipartite Networks, Challenge-based Innovation

Résumé

Titre: Approches de Réseau pour Reconstruire et Analyser les Structures de Collaboration d'Equipe dans la Science Ouverte et l'Innovation

Les défis modernes en matière de science et d'innovation exigent l'intégration de perspectives diverses parallèlement à des connaissances spécialisées, ce qui entraîne une évolution vers une recherche interdisciplinaire et collaborative basée sur le travail d'équipe. Cette transformation vers la production collective de connaissances continue à soutenir des résultats de haute qualité et l'impact de la recherche, comme en témoignent des cadres tels que l'innovation basée sur les défis et la science des foules. Pour comprendre comment des équipes plus grandes et plus diversifiées collaborent, répartissent le travail et coordonnent les tâches, il faut adopter une approche mixte, combinant des données qualitatives et quantitatives pour saisir leurs propriétés structurelles. Cependant, le manque d'ensembles de données longitudinales et granulaires a limité notre compréhension de la manière dont les structures de collaboration évoluent et influencent les performances de l'équipe. Dans cette thèse, je m'appuie sur les méthodes de la science des réseaux pour étudier les structures de collaboration dans des contextes d'innovation et de science ouverte basés sur des défis. Tout d'abord, j'étudie les équipes de science citoyenne qui s'attaquent aux défis des Objectifs de développement durable, en identifiant les modèles de collaboration et d'engagement par le biais d'enquêtes et de traces numériques provenant des communications Slack. Je démontre comment ces modèles influencent la performance de l'équipe, en soulignant l'importance d'un engagement soutenu avec des individus à travers divers emplacements du réseau. Deuxièmement, j'étudie la dynamique organisationnelle de plus de 1 200 équipes uniques participant au concours interdisciplinaire de science et d'ingénierie iGEM sur 12 ans, en analysant leur division du travail et leur coordination à l'aide de déclarations d'attribution autodéclarées et de traces numériques provenant de pages wiki en ligne non structurées et éditées en collaboration. J'introduis une nouvelle méthode informatique pour déduire les réseaux de répartition des tâches des équipes à l'échelle à l'aide de grands modèles de langage, démontrant sa capacité à récupérer les propriétés locales et globales du réseau avec une grande précision. Je soutiens que les indicateurs basés sur les réseaux, tels que la modularité et l'imbrication, rendent compte de la spécialisation et de la coordination au sein des structures de division du travail, et je montre qu'ils ont un impact significatif sur les performances de l'équipe lors de la compétition. En outre, je montre que les équipes adaptent leur division du travail au fil du temps, convergeant vers des structures optimales en fonction de leur taille, de leur expérience et de la complexité de la tâche. D'une manière générale, cette thèse contribue à la recherche sur la science des équipes dans les écosystèmes d'innovation ouverte et de science en introduisant des approches basées sur les données et les réseaux pour saisir les structures organisationnelles, offrant aux chercheurs et aux praticiens un

aperçu de la performance, de l'engagement et de l'adaptation des équipes, soutenant ainsi le développement de stratégies visant à favoriser l'innovation dans la recherche collaborative.

Mots Cles: Science des équipes, Science des réseaux, Réseaux de collaboration, Structures de répartition des tâches, Réseaux bipartites, Innovation basée sur les défis

Résumé Substantiel

Synthèse:

Le processus moderne de la science et de l'innovation est complexe et nécessite l'intégration d'une expertise et d'expériences à la fois spécialisées et interdisciplinaires. L'intérêt croissant pour les paradigmes de la science ouverte, qui mettent l'accent sur la transparence et la participation, démocratise le processus de production des connaissances scientifiques, renforce la nécessité d'une action collective et encourage les approches collaboratives. La science ouverte, collaborative et fondée sur le travail d'équipe pour résoudre les problèmes sociétaux mondiaux bénéficie des diverses compétences de ses membres et de l'accès à des données et à des ressources qui sont difficiles à obtenir. Toutefois, la coordination entre ces parties prenantes internes et externes entraîne des coûts.

Pour favoriser une recherche plus collaborative et plus efficace, il faut mieux comprendre comment les équipes s'organisent pour résoudre des problèmes complexes. Ces connaissances peuvent ensuite être utilisées pour aider les facilitateurs à soutenir les équipes. La science des réseaux est largement utilisée pour représenter et étudier l'organisation des équipes et la manière dont leurs structures de collaboration sont associées à leurs résultats. Cependant, l'identification des structures organisationnelles des équipes et leur modélisation en tant que réseaux nécessitent des données précises sur la manière dont les équipes répartissent les tâches entre elles et sur la manière dont elles se coordonnent avec les parties prenantes externes. La collecte et le traitement de ces données constituent un défi majeur, nécessitant des méthodes qualitatives et quantitatives pour recueillir les données des équipes de manière longitudinale et à grande échelle.

Les méthodes de la science des réseaux peuvent être mises à profit pour obtenir un aperçu complet de l'organisation des équipes, de la diversité, de la collaboration et de la manière dont ces structures sous-tendent les performances et l'apprentissage au fil du temps. Ces connaissances peuvent ensuite être utilisées pour soutenir les équipes de science ouverte et d'innovation dans leurs efforts de recherche scientifique. Ces stratégies de facilitation, qui visent à surmonter les pièges de la communication et de la coordination au sein de l'équipe et à produire une science percutante, nécessitent la création de boîtes à outils numériques pour aider les praticiens à maintenir l'engagement et à identifier les moments où des interventions et un soutien sont nécessaires.

Dans cette thèse, je compile mes recherches doctorales en intégrant les réseaux et les approches de la science des données pour étudier les structures de collaboration dans les équipes. des données pour étudier les structures de collaboration dans la science ouverte

et l'innovation basées sur le travail d'équipe. Afin de relever le défi de la conservation des données, j'utilise une approche mixte - en collectant des données qualitatives à partir d'enquêtes et de rapports personnels et des données quantitatives à partir de traces numériques. Je présente des pipelines pour traiter ces données mixtes et les représenter sous forme de réseaux. Pour étudier comment la structure de l'organisation de l'équipe sous-tend la performance et l'apprentissage, je construis des mesures de réseau identifiant les propriétés structurelles significatives et les associe aux résultats de l'équipe. Pour montrer comment les inférences de réseau peuvent être employées pour faciliter les équipes dans un contexte d'innovation ouverte, je construis un tableau de bord pour informer sur la collaboration, l'engagement et discuter de la façon dont ils peuvent soutenir la conception de l'intervention.

Contributions:

Les principales contributions de cette thèse sont les suivantes :

- Développer des pipelines de données pour collecter, traiter et conserver les collaborations d'équipes internes et externes à partir de traces numériques. Ces traces numériques sont collectées à partir des structures de contribution des membres de l'équipe travaillant en collaboration sur une ressource partagée (Wikis) et des modèles de communication sur des plateformes en ligne partagées (telles que Slack).
- Développer des méthodologies pour traiter des données qualitatives non structurées et autodéclarées (collectées à partir d'enquêtes, de wikis) provenant d'équipes, afin d'identifier leur structure organisationnelle interne et les collaborations entre les équipes - en utilisant des avancées récentes dans le traitement du langage et les grands modèles de langage (LLM).
- L'utilisation de la science des réseaux pour élaborer des mesures permettant de décrire les structures de collaboration des équipes dans les domaines de la science ouverte et de l'innovation. la science ouverte et l'innovation. J'explore l'interaction entre la spécialisation et la coordination dans l'organisation interne des équipes. dination dans l'organisation interne des équipes et j'étudie comment les équipes s'appuient sur des collaborations externes pour combler les lacunes techniques et favoriser l'innovation. collaborations externes pour combler les lacunes techniques et favoriser l'innovation.
- Construire des tableaux de bord de visualisation qui présentent des données clés, des observations et des résultats (en particulier sur leur diversité, leur collaboration et leur engagement) afin d'aider les praticiens des communautés de science ouverte à surveiller l'évolution de la science et de l'innovation. Les praticiens des communautés scientifiques ouvertes peuvent ainsi surveiller et soutenir les équipes.
- Discuter de la manière dont ces méthodes de conservation et de traitement des collaborations inter- et intra-équipes peuvent être étendues à d'autres contextes scientifiques et d'innovation et présenter les futurs champs d'application de la recherche.

Contexte

Dans cette thèse, j'étudie la science ouverte et les équipes d'innovation dans les contextes suivants.

L'Incubateur de Projets Scientifiques Citoyens Crowd4SDG

Crowd4SDG est un projet Horizon 2020 qui vise à former des équipes d'étudiants à construire un projet de science urbaine qui s'attaque aux défis des Objectifs de Développement Durable (ODD). Le projet vise à former des équipes d'étudiants à la construction d'un projet scientifique urbain qui s'attaque aux défis des objectifs de développement durable (ODD). Le projet y parvient en organisant un cycle annuel d'innovation en quatre étapes, appelé GEAR (Gather, Evaluate, Accelerate and Refine). Les étudiants sont regroupés en équipes, encadrés et dotés d'outils techniques et numériques pour conceptualiser et prototyper leur projet de science citoyenne. À la fin de chaque étape, les équipes sont évaluées et sélectionnées pour passer à l'étape suivante. Les équipes qui passent à l'étape finale sont invitées à présenter leurs projets aux communautés de recherche, aux sponsors potentiels et aux parties prenantes.

Le projet Crowd4SDG s'est déroulé sur trois cycles avec des objectifs multiples. Tout d'abord, le cycle GEAR est un cadre de formation et de facilitation dont l'un des principaux objectifs est d'évaluer la méthodologie et la manière dont le développement de nouvelles boîtes à outils de science citoyenne et de stratégies de formation favorise la mise en œuvre de projets innovants. Outre l'efficacité du cycle GEAR, il est essentiel de comprendre comment les équipes restent engagées dans le programme et quels sont les facteurs prédictifs de l'innovation au sein de l'écosystème d'innovation collaborative basé sur les défis. Cela permet aux facilitateurs de reconfigurer leurs stratégies et d'offrir des perspectives d'avenir pour la conception de cadres similaires. Pour ce faire, nous suivons et étudions les équipes tout au long de leur progression dans le programme GEAR à l'aide d'une approche à multiples facettes - y compris des enquêtes qualitatives, des communications sur des plateformes web et des traces d'outils numériques.

L'écosystème d'Apprentissage Basé sur le Défi iGEM

L'International Genetically Engineered Machines (iGEM) est un concours de sciences et d'ingénierie dans le cadre duquel des équipes d'étudiants (hébergées dans des universités ou des écoles secondaires) travaillent à la conception d'un projet de biologie synthétique. Au cours du cycle du concours (qui se déroule pendant l'été et dont l'événement final a lieu en novembre), les équipes doivent entreprendre plusieurs tâches dans le cadre de leur projet, telles que la création de pièces de biologie synthétique, la documentation de leur travail sur un cahier de laboratoire numérique, la collaboration avec d'autres équipes et l'engagement auprès du public et de diverses parties prenantes.

La structure globale de l'iGEM, l'accent mis sur les principes de la science ouverte, la rigueur scientifique et les données fines issues des traces numériques et d'une stratégie d'évaluation formalisée pour les équipes en font un excellent banc d'essai pour étudier les équipes, en mettant l'accent sur la manière dont les structures d'organisation des équipes

sous-tendent leurs performances. Les équipes iGEM ayant participé au concours pendant plusieurs éditions, le banc d'essai offre une perspective unique sur la manière dont l'organisation des équipes évolue avec l'expérience en matière de science et d'innovation.

Ces études de cas offrent des perspectives contrastées sur les équipes scientifiques et d'innovation. L'incubateur Crowd4SDG a des équipes plus petites (2-4 membres) qui sont principalement auto-organisées. Les principaux objectifs de l'incubateur sont de faciliter l'apprentissage et de soutenir les équipes dans la mise en place de projets scientifiques citoyens innovants. Nous utilisons ces résultats pour comprendre les facettes des équipes qui réussissent, en nous concentrant sur leur collaboration, leur composition et leurs modèles d'engagement, et la structure cyclique de Crowd4SDG permet d'intégrer les idées des équipes par l'intermédiaire des organisateurs et des facilitateurs.

L'iGEM, quant à lui, est composé d'équipes plus importantes (en moyenne 15-18 membres) avec une hiérarchie d'équipe définie. Les équipes sont dirigées par des chercheurs seniors (PI), mais la plupart des recherches sont menées par les étudiants. La structure du concours iGEM, les résultats bien définis et le fait que les équipes participent à des éditions successives, le banc d'essai permet d'explorer en profondeur la dynamique de l'évolution de l'organisation des équipes avec l'expérience.

Organisation de la Thèse

La thèse est organisée en quatre chapitres et se termine par des remarques finales :

- **Chapitre 1:** Présente une introduction générale à la thèse, en particulier sur les processus de production de connaissances scientifiques, l'émergence de pratiques collaboratives ouvertes dans la science, l'utilisation de réseaux pour modéliser ces collaborations et la facilitation de l'innovation dans des contextes de collaboration et de science ouverte. Présente l'organisation de la thèse.
- **Chapitre 2:** Présente l'incubateur de science citoyenne Crowd4SDG et décrit les méthodologies utilisées pour collecter les collaborations internes et externes des équipes participantes, par le biais d'enquêtes et de communications Slack. J'utilise les interactions recueillies auprès des équipes participantes pour construire des indicateurs de collaboration, de diversité et d'engagement, et je démontre comment ces indicateurs sous-tendent leurs performances et leurs résultats. Je construis un tableau de bord de visualisation pour aider les praticiens de l'incubateur à suivre et à soutenir les équipes de science citoyenne participantes.
- **Chapitre 3:** Présente les méthodes de collecte et de traitement des données des équipes participant au concours de science et d'innovation iGEM. Je décris les pipelines pour traiter les données quantitatives provenant des traces numériques et les données qualitatives provenant des auto-rapports pour récupérer les réseaux de collaboration inter- et intra-équipes.

-
- **Chapitre 4:** Présente les méthodes de la science des réseaux pour étudier la structure organisationnelle interne des équipes iGEM au niveau global et au niveau des tâches. Je me concentre sur la modularité et l'imbrication en tant qu'indicateurs de la spécialisation et de la coordination dans les équipes et je démontre comment elles prédisent le succès de l'équipe. J'exploite l'aspect longitudinal de la participation des équipes pour comprendre comment elles apprennent avec l'expérience et comment cet apprentissage se reflète dans l'organisation de leur équipe.

Acknowledgements

This thesis is a culmination of the last three years (four counting my time as a research engineer) of work and I am able to present it to the readers only through the support and encouragement from people around me.

First and foremost, I would like to thank my advisor Marc, who has been an integral part of my research journey. Ever since I joined his team as an intern, he has been a constant source of motivation and guidance. His positivity is infectious and I have felt as strong a sense of belonging as I have ever felt. Being a part of the Interaction Data Lab has been an experience filled with joy and discovery, from learning about hypergeometric tests to meditative and contemplative practices. This is a chapter of my life, teeming with core memories that I look forward to cherish.

My co-advisor Megan has backed my progress every step along the way and I am forever grateful for her support. Thank you Robbie for always helping me craft ideas (and sharing a truckload of decade old research papers). I am indebted to Liubov, Raph, Chakresh, Camille, Leo, Olga, Fabrice, Stephanie, Savvy and all the members of the research team for making my time infinitely more enjoyable.

Thank you to my thesis advisory committee - Henry, Sasha and Jean-Philippe. Our checkpoint meetings were crucial in aligning my research directions and your motivation, feedback and support has unlocked several possibilities that I was unaware of.

My exchange visit to Georgia Tech was an important experience during my PhD, introducing me to the world of social science and public policy. Thank you to Dr. Cassidy Sugimoto for enabling this visit and for the discussions with your research team. Thank you to Dr. Walsh and Dr. Lariviere for your insightful comments and feedback which were crucial in shaping my research.

The Center for Research and Interdisciplinarity (and the Learning Planet Institute) has been my *home away from home* for the last five years. I am grateful for the welcome and support from the entire community. Thank you to the FIRE doctoral school for facilitating my PhD and to Beatrice and the admin team for your immeasurable help navigating paperwork. Thank you to Chetan, Vaibhav, the HOME association and my fellow PhD cohort for being a point of reference with your friendship and encouragement.

Thank you to Dr. Lekshmi, Dr. Nadarajan and all the faculty at the AMCS department for providing the shoulders atop which I am able to reach new heights. Thank you to my

internship advisors - Adam and Dr. Zhang for your motivation and guidance.

Nothing will ever suffice the incredible warmth I have felt from my close friends. Thank you to Priyanka for completely transforming the experience of being under lockdown in a foreign country into an emotionally and intellectually stimulating one. I am grateful for the regular calls and check-ins from PK, Prasanth, Pavi and Swetha, whom even in distance, made me feel closer (and very much sane).

And finally, I dedicate this thesis to my family. To Mom, Dad, Chandru and my grandparents. Your love has no bounds and I only wish to give back twofold.

Contents

Abstract	i
Résumé	iii
Résumé Substantiel	v
Acknowledgements	xi
1 Introduction	1
1.1 Scientific knowledge Production as a Process	2
1.2 The Evolution of Modern Science and Innovation Processes	3
1.2.1 Collaborative Knowledge Production in Science	3
1.2.2 Collaboration in Open Science and Innovation	4
1.2.3 Structure and Impact	6
1.3 Networks and Team Collaborations	9
1.3.1 Networks of Internal Organisation in Teams	9
1.3.2 Networks of Collaboration across Teams	11
1.3.3 Evolution of Structure over time	12
1.4 Facilitating Outcomes in Open Science and Innovation	13
1.4.1 Monitoring and Evaluation	14
1.4.2 Challenges	15
1.5 Presentation of the Thesis	15
1.5.1 Contributions	16
1.5.2 Contexts	17
1.5.3 Thesis Organisation	17
2 Monitoring Citizen Science Teams using Digital Traces	29
2.1 Collaboration and Performance of Citizen Science Projects Addressing the Sustainable Development Goals	31
2.1.1 Introduction	33
2.1.2 Methods	34
2.1.3 Results	39
2.1.4 Discussion	46
2.1.5 Conclusion	50
2.2 Statistical Model of the Association Between Collaboration Dynamics and Project Performance	58
2.2.1 Introduction	60

2.2.2	Methods	60
2.2.3	Results	65
2.2.4	Discussion of the results	70
2.2.5	Conclusion and perspectives	71
2.3	Visualisation Dashboards to Monitor Citizen Science Teams	73
2.3.1	Methods	74
2.3.2	Snapshots of the Dashboard	76
2.3.3	Discussion	79
2.4	Perspectives	81
3	Network science and computational methods for processing collaboration networks from digital lab notebooks	89
3.1	Team Collaboration Dynamics from Digital Traces	91
3.1.1	The iGEM Synthetic Biology Competition	92
3.1.2	Data	93
3.1.3	Methods	97
3.1.4	Discussion and Perspectives	102
3.2	Team Collaboration Dynamics from Self-Reported Data	103
3.2.1	Introduction	105
3.2.2	Related Work	106
3.2.3	The iGEM Dataset	107
3.2.4	Methods	108
3.2.5	Results	111
3.2.6	Discussion	115
3.2.7	Conclusion	117
3.3	Perspectives	132
4	Evolving Structures of Division of Labour and Coordination in Interdisciplinary Teams	137
4.1	Introduction	139
4.2	Background	141
4.2.1	Division of Labour in Teams	141
4.2.2	Measuring Division of Labour	142
4.2.3	Recovering the Task Structure	143
4.2.4	Team Adaptation and Learning	144
4.2.5	Objectives	145
4.3	Data	146
4.4	Methods	149
4.4.1	Team Division of Labour from Attribution Statements	149
4.4.2	Task Structure and Similarity in Teams	149
4.4.3	Organisation from the Wiki Co-edit Structure	150
4.4.4	Measuring Division of Labour Structures	151
4.4.5	Outcomes	152
4.5	Results	153
4.5.1	Nestedness and Modularity in Networks of Division of Labour	153
4.5.2	Division of Labour and Performance	154
4.5.3	Co-edit Structure and Performance	155

4.5.4	Evolution of Task Structures	155
4.6	Discussion and Perspectives	157
	Conclusion	169
	List of Figures	176
	List of Tables	177

Chapter 1

Introduction

1.1 Scientific knowledge Production as a Process

In his seminal book, "Laboratory Life: The Social Construction of Scientific Facts" in 1979, Bruno Latour (with Steve Woolgar) [noa, a] presents the discovery of scientific facts as a social process, shaped through interactions, negotiations and routines. The laboratory is presented as the epicenter of this process of discovery and is a highly social setting. Performing activities like discussions, experiments and writing drive the process of knowledge production.

Extending his research in [noa, b], Latour details these processes that transform scientific knowledge into irrevocable fact. Building scientific facts requires networking across people, resources and institutions - making science a highly collective endeavour shaped by these interactions. Scientific knowledge is more often than not, a response to a collective social need. Although knowledge is produced through collective social processes at the level of a laboratory - they also include negotiations with actors that drive the societal need, whom also play an essential role in its acceptance and dissemination. Science, zoomed out, is a dynamic process strongly embedded in social interactions between heterogeneous actors. The fusing of human and material actors into social and technical interactions makes the scientific process dynamic and highly interconnected (referred to as the *mangle*, highlighting the complex intertwined nature of doing science [Pickering, 1996])

The particulars of the processes that shape scientific knowledge production change by field. Each scientific discipline develop their own *epistemic cultures*, capturing the distinct facets that shape the knowledge production processes [Cetina, 1999]. A laboratory in molecular biology is traditionally smaller, and composing members perform experiments and work hands on with organic matter. By contrast, high energy physics inherently relies on large scale collaborations (across countries, funders and stakeholders) to manage the high costs of building and running complex machines. Knowledge produced are often indirect, relying on computational simulations and statistical analysis. Collaboration, division of labour and integrating knowledge from interdisciplinary backgrounds is foundational in statistical physics, whereas molecular biology laboratories emphasize specialisation of researchers in performing experiments. These contrasting cultures are also reflected through a laboratory's social organisation, with the former having bureaucratic and hierarchical structures and the latter more decentralised.

The field of *Science and Technology Studies* has emerged from the need to understand more about these complex interconnected processes that drive scientific knowledge production. Despite the differences across fields, knowledge production processes have common underlying facets - that are a response to the modern and evolving scientific landscape.

1.2 The Evolution of Modern Science and Innovation Processes

The core contemporary challenges that we face together as a community are inherently complex and interdisciplinary. Producing scientific output to tackle these societal challenges - like global warming or sustainable development - needs coordinated and collective efforts from individuals from diverse backgrounds and experiences. This is reflected in the shift of the scientific ecosystem towards a collaborative team-based organisation and the evolution of more *open* forms of knowledge production.

1.2.1 Collaborative Knowledge Production in Science

Scientific research is increasingly team-based and collaborative [Wuchty et al., 2007]. Traditionally small-scaled research fields like mathematics have increased their size through the number of authors per publication, whereas research in science and engineering domains has seen a steady growth in the size of authors per publication over last decades [Fortunato et al., 2018].

Increasing complexity at the field and task levels are a significant contributor for the shift towards collaborative science. Complexity at the task level is underlined by the challenges in accomplishing distinct components within each task, coordination across these components and their dynamic nature [Wood, 1986, Campbell, 1988]. The process of scientific discovery inherently pushes boundaries and bridges gaps across disciplines, increasing the task complexities across these dimensions. This emerges the need to integrate diverse expertise and adopt effective organisation strategies to tackle this complexity and generate scientific knowledge.

More and more scientific knowledge is produced every day. This exponential growth presents challenges at the individual level. Producing novel ideas requires efficiently recombining existing knowledge, but the scale of knowledge production greatly transgresses the cognitive capacities of an individual scientist [Jones, 2009]. This presents the need to collaborate to maximise the knowledge base required in producing interdisciplinary and innovative science.

Technical and economic needs also foster this shift towards collaborative research. Reusing data, methods and high-cost technologies from prior works form a basis for scientific collaboration. Furthermore, the growing reliance of funding as support mechanisms for universities and the skewed nature of strategies to assign funds (biasing towards large universities and geographical hotspots) incentivises collaboration across institutional and economic boundaries [Stephan, 2015, Adams et al., 2005].

Collaborative science also offers significant benefits for research impact. Despite the

costs to coordination, team based science is associated with higher impact - measured through citations [Larivière et al., 2015, Milojević, 2020]. This impact is realised through the diversity in technical and material expertise and experiences that were efficiently recombined to produce knowledge. Collaborative science is also significant to advance individual career trajectories - through access to resources and connections [Lamain et al., 2024, Piezunka and Grohsjean, 2023], which still remains the key strategy for hiring in science.

This shift towards collaborative science has been significantly supercharged by the increased geographic and digital connectivity across the globe. From Paul Erdos, adopting a nomadic lifestyle - dedicated to mathematical collaboration and problem solving - to now geographic mobility supported by funding institutions aiming to diversify their investment in science and technology [Catalini, 2017, Adams, 2013, Bercovitz and Feldman, 2011] and accelerated through transportation and technological advances. But the main transformation is driven by the development of the internet and online tools [Nielsen, 2012]. Digital protocols - as simple as the email to more complex frameworks for collaboration like Git and version control - have significantly enabled the sharing and dissemination of data, knowledge and widened the access to key resources.

Despite the benefits to collaboration, namely the distributing technical efforts and costs - collaborative knowledge production in science is limited by institutional structures that are hard to reform. Scientific knowledge produced collaboratively within a laboratory setting rely on securing funding through grant proposals, being hosted within an academic or R&D ecosystem, communicating science through peer-review and justifying scientific investments, collaborations through reports and outreach initiatives. However, many modern societal issues are localised to specific contexts, require rapid prototyping and sharing solutions, while fundraising to tackle these issues are significantly limited by the lack of equitable support across the geographic and diverse knowledge bases. This shifting need to generate scientific knowledge relying less on the typically rigid institutional structures, have blossomed into new open science initiatives, emphasising on participation and sharing of knowledge - supported by the transformative power of online digital tools.

1.2.2 Collaboration in Open Science and Innovation

Open science, as the name suggests, *opens* or unlocks the processes are performed in producing scientific knowledge. The shift towards open science practices stem from the complexities of tackling modern science and innovation challenges, the crisis of reproducing existing scientific methodology and the need to have equitable access to resources and knowledge around the world.

Open science in academic research aims to achieve this by emphasising on transparency as a key foundational philosophy. Challenges to reproduce previously peer-reviewed and accepted scientific fact are often attributed to a lack of openly available data and in using questionable study design, analysis methodology and communication strategies [Ioannidis,

2005, Baker, 2016, Molloy, 2011]. Making data, methods and protocols public and transparent, facilitates the validation and sharing of these resources indiscriminately, alleviating some of the reservations in the acceptance and dissemination of scientific fact, while also encouraging its reuse across disciplinary boundaries.

However, open science practices are traditionally not incentivized in research. This rigidity arises from the additional costs that are incurred within the scientific knowledge production process in making research open. Moving towards reproducible scientific research requires time and retraining efforts in using computational tools and platforms (like Git, data repositories like Zenodo, software etc.) and clear documentation strategies [Stodden et al., 2018]. Publishing peer-reviewed manuscripts and data through open access often incurs disproportionate costs from journals, and until recently had minimal impact on researchers' career progress in academia. Despite these challenges, the significant value of open practices in democratising science and knowledge production makes promoting this culture a key challenge for the modern global scientific community [Nielsen, 2012, Nosek et al., 2015].

Despite this push towards openness in academic research, the fundamental processes that take place in producing knowledge are closed. In most cases, peer-reviewed knowledge is publicly communicated while intermediate methods, outputs and results are privy to members within the research team [Franzoni and Sauermann, 2014]. This along with institutional challenges prevent public participation across different facets of the research process. With the complexities of addressing global challenges requiring collective efforts beyond the traditional boundaries of academic research, the open science movement has led to crowd science initiatives.

Crowd science initiatives lean on the strength of the global public in accomplishing several foundational and research tasks in producing knowledge [Poetz and Sauermann, 2024]. They contrast with scientific research in the communication of intermediate results and facilitating more open participation [Franzoni and Sauermann, 2014]. Despite the heterogeneity in science training, crowd science has resulted in the production of novel and innovative research output. Crowd initiatives such as the Polymath project [Cranshaw and Kittur] have successfully used a forum-style collaborative platform to solve open problems in mathematics. Wikipedia is a case study for collective curation of content, performing key tasks such as information gathering, fact checking, citation and presentation with no monetary reward [Puranam et al., 2014]. Platforms leveraging on the strength of the crowd for data curation and annotation have produced comparable performances to experts and used in research - such as the cases of Zooniverse [Cox et al., 2015] and Amazon Mechanical Turk [noa, c, Yetisgen-Yildiz et al., 2010]. Open source projects (OSS) on platforms like Github lay the foundation for key computational and software advancements on the last decades. OSS projects have their software code and architecture public, allowing for contributors to provide their expertise through code or customise functionalities, which ultimately can be integrated into the main project.

At the smallest scales, crowd and citizen science initiatives call for participation and action in tackling the Sustainable Development Goals (SDGs), which encompass the key challenges we face together as a community [Haklay, 2018, Fritz et al., 2019, Fraisl et al., 2020]. From small-group participatory research in monitoring acute medical conditions, to initiatives that prototype solutions addressing global issues like the pandemic [Masselot et al., 2022], crowd science initiatives harnessing the latent transformative power at individual and collective scales and drives impact through efficiently integrating knowledge and experiences.

Both crowd and open science strongly relies on collaboration and collective efforts to organise the processes of knowledge production. The focus on reproducibility, accessibility, using digital resources and auxiliary tasks such as coordinating with public, stakeholders and science communication (beyond writing papers) increase its organisational complexities. Efficient organisation of teams in open science is a key facet that underlies their propensity in producing knowledge and impact.

1.2.3 Structure and Impact

The structure of a team reflects how they organise themselves in accomplishing their tasks. Henry Mintzberg, in his book *The Structuring of Organizations* [Mintzberg, 1989], describes this structure as a composite of interactions (or flows) between team members, corresponding to authority, social communications and formal work relationships. These interactions have been captured through surveys, identifying hierarchies, social communication and division of labour structures. In collaborative knowledge production, the team organisational structure is influenced by several factors, such as composition, task complexity and this structure offers insights on the novelty and impact of the produced output.

The scientific output is typically measured through units of publications. Increased productivity is indicated by publishing more peer-reviewed articles, technical reports or patents. However, measuring the quality and impact of a publication is a key field of research in science and policy. The traditional measures of impact are derived from citations. Innovative research recombines references across disciplinary boundaries [Uzzi et al., 2013] and its impact is realised through future citations from research spanning diverse fields. Future citations also indicate the reproducibility of data and methods. However, disambiguating the specific context of each citation in a publication (whether they rely on theoretical advancements or methodological practices) is a challenge.

Impact has a different scope in crowd science initiatives. They integrate outcome-based measures of impact (such as project feasibility) with process based measures like participation and learning [Jaeger et al.]. Hackathons and OSS project impact can be measured through its (re)use by people and institutions (captured through stars, forks on GitHub repositories). Polymath projects measure output through the mathematical problems solved (and new questions that prompts). Open data curation platforms measure impact through

the quality of the data and its propensity for reuse in science and research. Crowd science project initiatives are often vessels to train students and the global public on research methodology - and in addition to impact, learning and maintaining engagement is an outcome that generates benefits that are realised over the long term.

This subsection presents a short overview of the interplay between team structure and impact in open science and innovation.

Team Composition and Diversity

Team composition describes attributes of a team and its constituent members. These attributes can be:

- **Size:** With most science research now team-based, larger teams help realise higher citation impact [Milojević, 2020, Larivière et al., 2015]. Larger teams are more interdisciplinary and have a higher division of labour amongst their members [Haeussler and Sauermann, 2020, Adams et al., 2005].
- **Diversity:** Increase in team size often corresponds to a higher diversity on geographic and technical backgrounds of members. Increased technical diversity has an inverse 'U' shaped relationship with impact in science and R&D [Yegros-Yegros et al., 2015, Hoisl et al., 2017]. Despite the coordination costs to bridge people and institutions across geographic regions, increased diversity realises additional outcomes in terms of access to resources and procurement of grants [Bercovitz and Feldman, 2011, Larivière et al., 2015]. Diversity in terms of experience also indicates a higher rate of success and impact [Taylor and Greve, Klug and Bagrow, 2016].
- **Team Assembly:** Team Assembly describes the methodologies used to construct a team and how team members assign themselves into tackling different tasks. Team assembly takes size, diversity and task complexity into consideration, and informs on teams organisational structure [Raveendran et al., 2022] and their project impact [Guimerà et al., 2005, Twyman and Contractor, 2019]. Teams in open and crowd science are often self-assembled, with digital communication tools leveraged to provide recommendations and spark interactions with new(er) members based on project needs.

Decisions on team assembly, size and diversity impact how teams organise their structure in performing different tasks in open science and innovation.

Specialisation and Integration

Teams play an optimisation game, balancing the benefits of specialisation in performing particular tasks with the costs of coordination across these specialised groups. Increased

specialisation, reflected by an extensive division of labour in teams, positively effects productivity and supports the training of users into a narrow set of skills [Becker and Murphy, 1992]. This however inversely relates to coordination costs [Lawrence and Lorsch, 1967] - and in complex task environments, the need for coordination across specialists is high and require robust integration structures within the team division of labour [Kretschmer and Puranam, 2008].

Specialisation and coordination structures inform on teams' global organisational patterns. Flat teams are often smaller in size and diversity, have higher coordination between composing members, with decreased individual specialisation to tasks. But, flat teams explore more atypical solutions and realise higher innovation and impact [Xu et al., 2022]. Whereas higher specialisation increases the need for coordination across specialist groups, reflecting a more modular and hierarchical structure. Hierarchies benefit from their ability to converge to a stable organisational structure over time, allowing exploitation and rapid prototyping of solutions [Koçak et al., 2023].

The interplay between specialisation and integration is also of importance in open innovation settings. Successful OSS projects evolve into narrow organisation structures, where a *core* set of contributors coordinate across overlapping subgroups, which provide specialist input [Klug and Bagrow, 2016, Palazzi et al., 2019]. The significance of specialist contributors who remain at the *periphery* of the contribution structure cannot be understated. Interactions from contributors at the periphery spark increased engagement and develop new ideas in Polymath projects [Gargiulo et al., 2022].

With open science and innovation initiatives predominantly relying on integrating expertise from diverse actors in performing scientific and technical tasks, the emergence of specialisation and coordination structures to produce output is a key research direction.

Internal and External Collaborative Processes

To produce novel and innovative research, in addition to integrating diverse expertise internally, teams collaborate with and access external inputs from other teams, institutions and resources.

External collaboration allows teams in science access to knowledge, resources and opportunities (such as funding) which are otherwise hidden. This results in differences in the organisation structure of teams, which need to harmonise to efficiently integrate external input with their internal processes. Collaboration with external stakeholders suffer from coordination challenges and misunderstandings, which are caused due to group diversity, geographical dispersion, interconnected tasks, cultural barriers, competition or economic gains [Walsh and Maloney, 2007]. However, these can be bridged using modern communication tools, as simple as the email or mobile phone calls. The impact realised by allowing for specialised external input through collaboration often times exceed the costs, especially

in tackling complex problems.

Crowd science opens for participation from diverse actors at different stages of the knowledge generation process. By design, they foster collaboration as a fundamental strategy to overcome challenges in realising impactful solutions for community-facing problems and facilitating learning along the way.

The organisational structure of teams in open science and innovation settings underlie their main outcome of producing impactful research. Modelling and studying this structure is key to understanding the optimal practices and designing methodologies to implement them in diverse settings. The next section presents networks and network science as a mathematical and computational framework to represent team organisation and derive inferences about their structure and how it associates with their outcomes.

1.3 Networks and Team Collaborations

Network science is a powerful tool to model interactions between diverse entities. The organisation structure in scientific teams are captured through interactions - internal task allocation between members and tasks which highlight the division of labour, coordination - and external collaboration between teams and other heterogeneous actors.

This section discusses the use of networks in modelling interactions and the organisational structures of teams in open science and innovation. I present the key methodologies from network science and social networks literature that are used to study these structures and provide motivations for my specific focus on network approaches in this thesis.

1.3.1 Networks of Internal Organisation in Teams

The internal division of labour networks capture how teams allocate different tasks that need to be performed in the knowledge production process, amongst their constituent members. With interactions only occurring between team members and tasks, the resulting network has two categories of vertices - with no interactions between vertices of the same class. These are denoted as *bipartite* networks in network science literature.

Bipartite networks are alternatively represented as an *incidence matrix*, where the rows correspond to all vertices of one category (users) and columns to that of the other (tasks), with each cell indicating the strength (or frequency) of the interaction between them (1.1a). Studies often use the adjacency matrices to describe the division of labour and team organisation structures [Haeussler and Sauermann, 2020], but the definitions can be reframed in bipartite network terminologies.

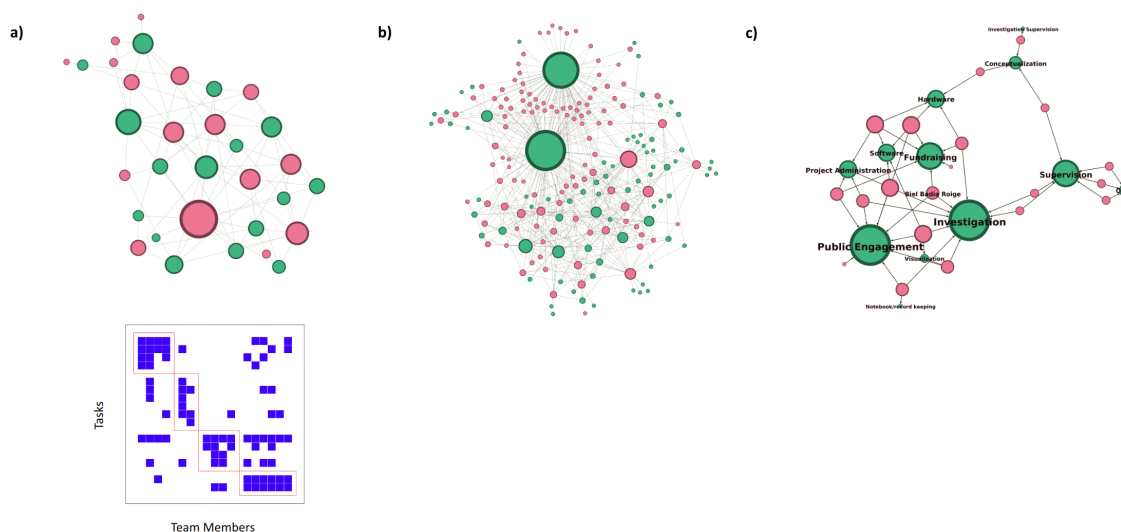


Figure 1.1: Bipartite networks a) Interactions between team members (green) and tasks (red). Bottom shows the incidence matrix of the bipartite network. b) Contribution structure of the *django-money* OSS project on Github. Green indicates the contributors and red the files. Vertex size highlights the degree, showing the presence of core contributors. c) Member (red) - Task (green) network with the tasks corresponding to the CREDIT categories.

Constructing the bipartite network of division of labour requires capturing the complete task allocation structure in a team. In academic research, this task allocation structure is declared by the authors of a manuscript in the *Contribution* section. To track academic contributions consistently, several taxonomies have been designed describing the specific roles that authors undertake in realising their research work. The CREDIT contributor role taxonomy [noa, d], has 14 roles, where each role involves performing specific functional tasks in producing scientific knowledge. Recent academic publications report their contribution structure consistent with the CREDIT roles [Sauermann and Haeussler, 2017, Larivière et al., 2020], a processed contribution statement can be represented as a bipartite network - with interactions between authors and roles (1.1c).

Digital traces captured from online tools used in open science and innovation, provide a tractable method to identify the contribution structure in teams. Here, interactions occur between members and the components that they co-create. In OSS projects - these components correspond to specific code or files (1.1b) [Lima et al., 2014, Palazzi et al., 2019]. In Wikipedia, interactions are between editors and articles (or sections for a more fine-grained view) [Yasseri et al., 2012, Klein et al., 2015, Keegan et al., 2012]. In forums, they are between members and posts [Poquet et al., 2020, Gargiulo et al., 2022].

Bipartite networks of a teams' internal organisation help identify several structural properties, corresponding to their composition, diversity, specialisation and coordination. Clustering bipartite networks helps identify modules (consisting of members and tasks)

who work closely, indicating specialised task sub-groups. This modular organisation allows to quantify the level of differentiation of members into tasks and understand the structures of coordination - by looking at interactions across different modules [Guimerà et al., 2007, Palazzi et al., 2019, Pasquaretta and Jeanson, 2018]. The modular structure also allows to understand the diversity of a teams' structure using member level attributes such as their experience and background. This has been a base to explore how different self-organisation strategies (assigning members to tasks) enable different organisational structures and higher team performance, through simulations and rewiring interactions in bipartite networks [Wax et al., 2017, Twyman and Contractor, 2019, Raveendran et al., 2022].

Bipartite networks are also projected onto either vertex category to generate a unipartite network. Two team members are connected in the unipartite network, if they work together on the same task. And conversely, two tasks are connected if the same user contributes to them both. The unipartite network on tasks provide a proxy of the interdependences between them within a team project. These task interdependences have been a focus to identify member roles (looking at how they are assigned to different tasks), the hierarchies and the complexities of coordination Xu et al. [2022], Haeussler and Sauermann [2020].

Both bipartite and unipartite networks are used to capture the structural properties of internal collaboration and division of labour in teams.

1.3.2 Networks of Collaboration across Teams

Collaborative activities of teams are identified through interactions connecting them to other teams and institutions. This external collaborations allows teams access to much needed resources and technical expertise which are not available internally.

Collaboration networks in academic research are constructed using co-authorship data. The institutional affiliations of authors of a paper are set to be the vertices and an interaction indicates a collaboration between researchers from two institutions. Collaborations that bridge across disciplinary and geographic boundaries are more successful in producing innovative research, securing funding [Bercovitz and Feldman, 2011] and in overcoming creative roadblocks [Lamain et al., 2024]. Beyond academic collaborations, crowd science initiatives also encourage collaborations between teams and projects. The open-ended participation strategies encourage involvement from diverse stakeholders who are otherwise often decoupled from science and innovation processes, in developing a collective vision [Senabre Hidalgo et al., 2021].

External collaborations connecting teams to diverse actors has been at the focus of identifying novel ideas and methods. Burt [Burt, 2004] uses a network science method to quantify the gaps that a collaboration connects between two diverse actors, with wider the gap, the more novel the recombination. Simpler proxies, such as the closing of triangles (a new

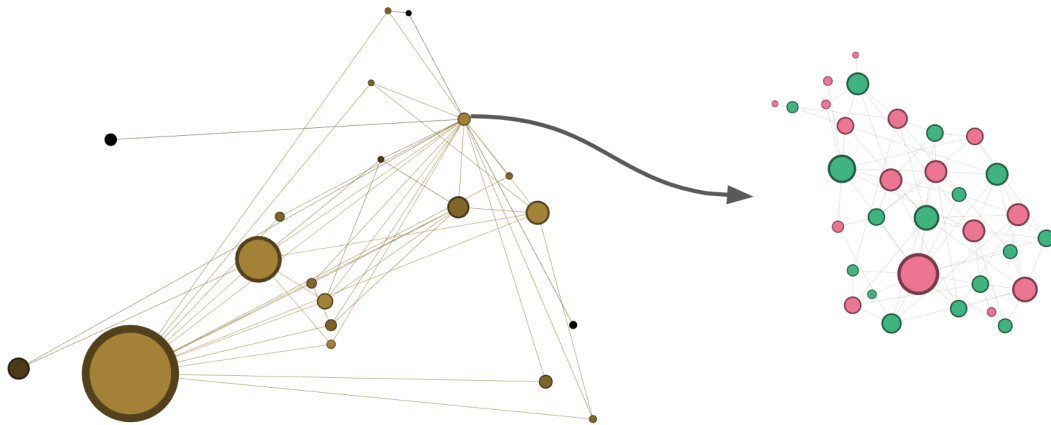


Figure 1.2: Multilayer networks of Collaborations. (Left) External collaborations between teams and (Right) Zoomed in version of the internal organisation between members (red) and tasks (green)

interaction between two actors who already have an actor they mutually interacted with) have shown to spark novelty and even commercial success in a case study of jazz recording sessions [Vedres, 2017, 2021].

The internal division of labour and external collaboration networks of teams can be leveraged together to gain a comprehensive view of organisation in science and knowledge production. Lazega [Lazega et al.], proposed linking the inter-individual collaborations with inter-organisational networks, allowing the identification of strategies used to navigate local and global interdependences to gain a competitive advantage (focusing on individual scientist careers). This is also true for teams, who need to integrate input from external collaboration while balancing their internal division of labour, signifying the need for a multi-level network approach to capture organisational complexities (figure 1.2).

1.3.3 Evolution of Structure over time

Collaborations change over time. Internal organisation of teams can be reconfigured, depending on emerging needs, challenges or conflicts. These can also prompt teams to seek out external support. Beyond the dynamic reconfiguration, teams learn to conserve optimal strategies and organisational routines with experience. These allow teams to specialise in different tasks and strengthen internal and external social ties - having positive effects on their productivity and impact [Becker, 2020, Puranam and Maciejovsky, 2020]. These routines and the evolution of the corresponding organisational structures can be captured through networks, using snapshots of team collaboration networks [Borge-Holthoefter et al., 2017] over time or modelling temporal networks [Paranjape et al., 2017] to identify the emergence of repeating structural patterns.

The main limitation to leveraging the strength of temporal network methods is the lack of granular data. OSS projects, Wikipedia articles or forums collect digital traces of contributions over time, which allow to identify emergent patterns longitudinally. However, in science, only a snapshot of the author contributions is recovered at a given time. Surveys can help capture snapshots of collaborations at different times of administering, but it is a time and resource intensive process.

In summary, network science literature provides opportunities and methods to represent inter- and intra- team collaborative processes, and offers measures to identify key structural properties that underlie impact in open science and innovation.

1.4 Facilitating Outcomes in Open Science and Innovation

Collective knowledge production, especially in open science and innovation settings require team members to take up the mantle of multiple key responsibilities. Core members of a research team perform roles such as fundraising, public engagement, promoting *open science* in addition to making decisions on team organisation, collaboration, coordination structures and undertaking technical project tasks [noa, d, Larivière et al., 2021]. Despite the perceived benefits to collective and open science methods, implementing them in practice is a complex endeavour - mainly requiring specialised training for students and researchers coming out of the scientific research ecosystem [Chesbrough, 2003, Borycz et al., 2023]. This is a foundational challenge to usher in the next wave of open participation and collaborative research to tackle real-world challenges

Open science and innovation initiatives are a vessel to aid in tackling this challenge. Structured programs are designed where participants work towards building a science and innovation project, are trained on various technical skills - where they, at the end, gain first-hand experience about the knowledge production processes, particularly emphasising on open practices. The focus on these dual outcomes - of learning by doing and producing impact - requires support structures that ensure participating teams meet them.

Building effective support structures within collaborative open innovation settings requires robust methodologies to monitor the teams' collective processes, measure and evaluate their outcomes which then help design interventions that facilitate learning and producing impact. This section discusses the motivations behind the strategies employed to monitor and facilitate open science and innovation teams.

1.4.1 Monitoring and Evaluation

Realising the outcomes in open science and innovation requires teams and collectives to efficiently balance the challenges in organising their internal structure (balancing specialisation and coordination), collaborate with external actors and resources and disseminate the knowledge produced. Understanding better about how different collaboration structures enhance learning and impact is a key question that will foster designing strategies to enhance open science and innovation activities.

Gaining a fine-grained view of how these teams organise and do science is a start. The increased costs to collaboration have been overcome by the use of digital tools, which capture (through networks) how teams build their projects, share data and collaborate with other teams [Nielsen, 2012]. Digital traces from these tools can be leveraged to capture structured and fine grained team quantitative data. Using version control software (such as Git while building software) and collaborative editing toolkits (such as forums, wikis, online editors to document their project and progress) allow tracking the contributions of team members in collectively building their project. Communication platforms like email, Slack also allow to track collaborations between individuals, teams and other stakeholders, allowing to measure external collaborations and engagement. Despite the scalability and longitudinal nature of the collected data, digital trace data are approximate proxies for team collaboration and engagement.

Surveys are a powerful instrument to collect qualitative data from teams. Surveys can be structured, enabling teams to regularly report on their internal and external collaborations [noa, e, Walsh and Lee, 2015], diversity [Van Der Vegt and Bunderson, 2005, Guzzo and Dickson, 1996] and satisfaction levels [Spector, 1997]. Surveys can also be used to collect unstructured data, allowing teams to self-report specifics about their methodology, attributions and detailing potential roadblocks. These surveys are administered through public digital tools or proprietary software, allowing to implement privacy preserving mechanisms for sensitive data. However, surveying is a resource intensive task - requiring additional time and manpower to validate and process collected data. Surveys are also challenging to scale to longitudinal settings and to larger teams or communities.

Team progress is usually measured by the impact their work has in cultivating new research. Measures like citations (in science) and reuse (of code in open innovation) are classic examples. Structured open science and innovation initiatives measure progress and impact by evaluating team projects and their activities. This measure of project impact is an amalgamation of project feasibility, novelty, development of open data and frameworks, collaborations and engagement [Chesbrough, 2003, Chesbrough et al., 2017].

Engagement is a key predictor of performance and learning in collective settings [Cox et al., 2015, Haklay, 2018]. Participant engagement is measured by their involvement with their project and within the initiative, using proxies such as interacting with people, materials or responding to queries and surveys. Sustaining engagement facilitates individual

and collective learning of skills and strategies used to tackle complex tasks. Tracking participant engagement furthers the need for fine grained strategies to monitor and support teams.

An ideal monitoring toolkit leverages both qualitative and quantitative data to gain a comprehensive overview of a teams' collective efforts, impact and engagement. This allows practitioners to identify moments where progress is halted due to an emerging need, lack of resources or poor collective organisation. Interventions can be designed to support teams during these points and facilitate future learning and impact.

1.4.2 Challenges

Administering surveys at designed intervals and prompting teams to use open tools capture fine-grained and longitudinal collaboration and coordination structures internal and external to teams in open innovation settings. Implementing this robust monitoring and evaluation framework and leveraging a multi-faceted approach for data collection requires significant effort costs. This necessitates a specialised set of organisers or practitioners who track teams during their participation using digital tools and administer surveys.

In addition, identifying team collaboration, organisation, diversity and engagement from both quali- and quantitative data sources requires data processing and pipelines using networks and data science methods to curate and build measures of team organisation and outcomes. Furthermore, digital platforms need to be designed to disseminate key observations to practitioners, which aid them in identifying specific team needs that require interventions and support.

1.5 Presentation of the Thesis

The modern science and innovation process is complex, requiring the integration of both specialised and interdisciplinary experiences. The growing focus on open science paradigms, which emphasise on transparency and participation, democratise the knowledge production process, and furthers the need for collective approaches. Open, collaborative and team-based science to tackle global societal problems benefit from the diverse skills of its members and access to data and resources that are scarce to obtain. However, they incur costs to coordinate across these internal and external stakeholders.

The underlying objective of fostering more collaborative and impactful research requires understanding more about how teams organise to solve complex problems and how can we support them in this process. Network science is widely used to represent and study the collaborative structures in teams and how they associate with their outcomes. But,

identifying team organisational structures requires fine-grained data of how teams allocate tasks amongst themselves and how they coordinate with external stakeholders. Gathering and processing this data is a key challenge, needing qualitative and quantitative methods to curate data from teams longitudinally and in scale.

Leveraging network science methods, gives a comprehensive insight into team organisational structures and how they underlie performance and learning over time. Facilitating open science and innovation teams to overcome the pitfalls in communication and coordination to produce impactful science requires building digital toolkits to help practitioners in sustaining engagement and identifying moments where interventions and support are necessary.

In this thesis, I compile my doctoral research integrating networks and data science approaches to study collaborative structures in team-based open science and innovation. In order to address the challenge of data curation, I use a mixed approach - collecting qualitative data from surveys and self reports and quantitative data from digital traces. I present pipelines to process this mixed data and represent them as networks. To study how team organisation structure underlies performance and learning, I construct network measures identifying the significant structural properties and associate them with team outcomes. To show how network inferences can be employed to facilitate teams in an open-innovation setting, I build a dashboard to inform on collaboration, engagement and discuss how they can support intervention design.

1.5.1 Contributions

The key contributions of this thesis are:

- Developing pipelines to collect, process and curate internal and external team collaborations from digital traces, using the contribution structures of team members collaboratively working on a shared resource (Wikis) and communication patterns on shared online platforms (Slack).
- Developing methodologies to process unstructured and self-reported qualitative data to identify team collaborations and organisation structures - using recent advances in language processing and Large Language Models (LLMs).
- Using network science to build measures to describe team collaborative structures in open science and innovation. I explore the interplay between specialisation and coordination in the internal organisation of teams and study how teams rely on external collaborations to bridge technical gaps and foster innovation.
- Building visualisation dashboards presenting data and findings to support practitioners in open science communities to monitor and support teams.

1.5.2 Contexts

In the thesis, I study open science and innovation teams in the following contexts.

The Crowd4SDG Citizen Science Project Incubator

Crowd4SDG is a citizen science project incubator program, with the key objective of facilitating student teams in designing novel and impactful citizen science projects addressing the Sustainable Development Goals (SDGs). Participants in the program are selected and formed into teams (based on their skills and backgrounds). They are provided with contextual training on citizen science and in using digital resources to help them conceptualise a citizen science project, which are then evaluated. Selected projects are passed onto the next round where they are given additional training and resources to build prototypes. Final projects, after an additional round of evaluations and selection, are connected with funders and stakeholders. With data collected from citizen science teams in real-time, this presents a case study to understand how participant engagement and team collaborations associate with performance and in designing toolkits to directly support the incubator framework.

The iGEM Challenge based Learning Ecosystem

International Genetically Engineered Machines (iGEM) is a science and engineering competition where student teams (hosted at universities or high schools) work on designing a synthetic biology project. Teams are expected to undertake several tasks such as creating synthetic biology parts, documenting their work on a digital lab notebook, collaborating with other teams and engaging with the public and diverse stakeholders. The global structure of iGEM, its emphasis on open science principles and the fine grained data curated from digital traces and a formalised evaluation strategy for teams make it an excellent testbed to study how team organisation structures underlie performance and evolve with experience in science and innovation.

1.5.3 Thesis Organisation

The thesis is organised into 3 further chapters, ending with some concluding remarks:

- **Chapter 2:** Presents the Crowd4SDG citizen science incubator and describes the methodologies used to collect internal and external collaborations of participating teams, through surveys and Slack communication. I use interactions collected from participating teams to build indicators of collaboration, diversity and engagement, and demonstrate how these indicators underlie their performance and outcomes. I build a visualisation dashboard to aid practitioners in the incubator to monitor and support participating citizen science teams.

- **Chapter 3:** Presents data collection and processing methods from teams participating in the iGEM science and innovation competition. I describe the pipelines to process quantitative data from digital traces and qualitative data from self-reports to recover inter- and intra- team collaboration networks.
- **Chapter 4:** Presents network science methods to study the internal organisational structure of iGEM teams at the global and task levels. I focus on modularity and nestedness as indicators of specialisation and coordination in teams and demonstrate how they predict team success. I leverage the longitudinal aspect of team participation to understand how they learn with experience and how this learning is reflected onto their team organisation.

Bibliography

- Laboratory Life: the Social Construction of Scientific Facts | bruno-latour.fr, a. URL <http://www.bruno-latour.fr/node/218>.
- Science in Action, How to Follow Scientists and Engineers through Society | bruno-latour.fr, b. URL <http://www.bruno-latour.fr/node/130.html>.
- Andrew Pickering. The Mangle of Practice: Time, Agency, and Science. *Bibliovault OAI Repository, the University of Chicago Press*, 38, November 1996. doi: 10.2307/3106908.
- Karin Knorr Cetina. *Epistemic Cultures: How the Sciences Make Knowledge*. Harvard University Press, 1999.
- Stefan Wuchty, Benjamin F. Jones, and Brian Uzzi. The Increasing Dominance of Teams in Production of Knowledge. *Science*, 316(5827):1036–1039, May 2007. doi: 10.1126/science.1136099. URL <https://www.science.org/doi/10.1126/science.1136099>. Publisher: American Association for the Advancement of Science.
- Santo Fortunato, Carl T. Bergstrom, Katy Börner, James A. Evans, Dirk Helbing, Staša Milojević, Alexander M. Petersen, Filippo Radicchi, Roberta Sinatra, Brian Uzzi, Alessandro Vespignani, Ludo Waltman, Dashun Wang, and Albert-László Barabási. Science of science. *Science*, 359(6379):eaao0185, March 2018. doi: 10.1126/science.aao0185. URL <https://www.science.org/doi/10.1126/science.aao0185>. Publisher: American Association for the Advancement of Science.
- Robert E Wood. Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes*, 37(1):60–82, February 1986. ISSN 0749-5978. doi: 10.1016/0749-5978(86)90044-0. URL <https://www.sciencedirect.com/science/article/pii/0749597886900440>.
- Donald Campbell. Task Complexity: A Review and Analysis. *The Academy of Management Review*, 13:40–52, October 1988. doi: 10.5465/AMR.1988.4306775.
- Benjamin F. Jones. The Burden of Knowledge and the “Death of the Renaissance Man”: Is Innovation Getting Harder? *The Review of Economic Studies*, 76(1):283–317, January 2009. ISSN 0034-6527. doi: 10.1111/j.1467-937X.2008.00531.x. URL <https://doi.org/10.1111/j.1467-937X.2008.00531.x>.

- Paula Stephan. *How Economics Shapes Science*. Harvard University Press, September 2015. ISBN 978-0-674-26755-8. Google-Books-ID: b5svEAAAQBAJ.
- James D. Adams, Grant C. Black, J. Roger Clemmons, and Paula E. Stephan. Scientific teams and institutional collaborations: Evidence from U.S. universities, 1981–1999. *Research Policy*, 34(3):259–285, April 2005. ISSN 0048-7333. doi: 10.1016/j.respol.2005.01.014. URL <https://www.sciencedirect.com/science/article/pii/S0048733305000132>.
- Vincent Larivière, Yves Gingras, Cassidy R. Sugimoto, and Andrew Tsou. Team size matters: Collaboration and scientific impact since 1900. *Journal of the Association for Information Science and Technology*, 66(7):1323–1332, 2015. ISSN 2330-1643. doi: 10.1002/asi.23266. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.23266>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.23266>.
- Staša Milojević. Towards a More Realistic Citation Model: The Key Role of Research Team Sizes. *Entropy*, 22(8):875, August 2020. ISSN 1099-4300. doi: 10.3390/e22080875. URL <https://www.mdpi.com/1099-4300/22/8/875>. Number: 8 Publisher: Multidisciplinary Digital Publishing Institute.
- C. Lamain, S. Brugman, M. Boes, C. Schoevaars, D. Tetteroo, M. D. Veldhuizen, J. P. Wijnen, D. Lakens, F. Albronda, S. Hofmann, S. Knittel, J. Duncan, and CUCo. Finding joy, creativity and meaning through unusual interdisciplinary collaborations. *Humanities and Social Sciences Communications*, 11(1):1–7, September 2024. ISSN 2662-9992. doi: 10.1057/s41599-024-03634-w. URL <https://www.nature.com/articles/s41599-024-03634-w>. Publisher: Palgrave.
- Henning Piezunka and Thorsten Grohsjean. Collaborations that hurt firm performance but help employees’ careers. *Strategic Management Journal*, 44(3): 778–811, 2023. ISSN 1097-0266. doi: 10.1002/smj.3447. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/smj.3447>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/smj.3447>.
- Christian Catalini. Microgeography and the Direction of Inventive Activity. *Management Science*, July 2017. doi: 10.1287/mnsc.2017.2798. URL <https://pubsonline.informs.org/doi/abs/10.1287/mnsc.2017.2798>. Publisher: INFORMS.
- Jonathan Adams. The fourth age of research. *Nature*, 497(7451):557–560, May 2013. ISSN 1476-4687. doi: 10.1038/497557a. URL <https://www.nature.com/articles/497557a>. Publisher: Nature Publishing Group.
- Janet Bercovitz and Maryann Feldman. The mechanisms of collaboration in inventive teams: Composition, social networks, and geography. *Research Policy*, 40(1):81–93, February 2011. ISSN 00487333. doi: 10.1016/j.respol.2010.09.008. URL <https://linkinghub.elsevier.com/retrieve/pii/S0048733310001988>.

- Michael Nielsen. *Reinventing Discovery: The New Era of Networked Science*. Princeton University Press, 2012. ISBN 978-0-691-16019-1. URL <https://www.jstor.org/stable/j.ctt7s4vx>.
- John P. A. Ioannidis. Why Most Published Research Findings Are False. *PLoS Medicine*, 2(8):e124, August 2005. ISSN 1549-1676. doi: 10.1371/journal.pmed.0020124. URL <https://dx.plos.org/10.1371/journal.pmed.0020124>.
- Monya Baker. 1,500 scientists lift the lid on reproducibility. *Nature*, 533(7604):452–454, May 2016. ISSN 1476-4687. doi: 10.1038/533452a. URL <https://www.nature.com/articles/533452a>. Publisher: Nature Publishing Group.
- Jennifer C. Molloy. The Open Knowledge Foundation: Open Data Means Better Science. *PLoS Biology*, 9(12):e1001195, December 2011. ISSN 1545-7885. doi: 10.1371/journal.pbio.1001195. URL <https://dx.plos.org/10.1371/journal.pbio.1001195>.
- Victoria Stodden, Friedrich Leisch, and Roger D. Peng, editors. *Implementing Reproducible Research*. Chapman and Hall/CRC, New York, December 2018. ISBN 978-1-315-37346-1. doi: 10.1201/9781315373461.
- B. A. Nosek, G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, C. D. Chambers, G. Chin, G. Christensen, M. Contestabile, A. Dafoe, E. Eich, J. Freese, R. Glennerster, D. Goroff, D. P. Green, B. Hesse, M. Humphreys, J. Ishiyama, D. Karlan, A. Kraut, A. Lupia, P. Mabry, T. Madon, N. Malhotra, E. Mayo-Wilson, M. McNutt, E. Miguel, E. Levy Paluck, U. Simonsohn, C. Soderberg, B. A. Spellman, J. Turitto, G. VandenBos, S. Vazire, E. J. Wagenmakers, R. Wilson, and T. Yarkoni. Promoting an open research culture. *Science (New York, N.Y.)*, 348(6242):1422–1425, June 2015. ISSN 0036-8075. doi: 10.1126/science.aab2374. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4550299/>.
- Chiara Franzoni and Henry Sauermann. Crowd science: The organization of scientific research in open collaborative projects. *Research Policy*, 43(1):1–20, February 2014. ISSN 0048-7333. doi: 10.1016/j.respol.2013.07.005. URL <https://www.sciencedirect.com/science/article/pii/S0048733313001212>.
- Marion K. Poetz and Henry Sauermann. Citizen Science and Crowd Science. In *Oxford Research Encyclopedia of Business and Management*. March 2024. ISBN 978-0-19-022485-1. doi: 10.1093/acrefore/9780190224851.013.385. URL <https://oxfordre.com/business/display/10.1093/acrefore/9780190224851.001.0001/acrefore-9780190224851-e-385>.
- Justin Cranshaw and Aniket Kittur. The Polymath Project: Lessons from a Successful Online Collaboration in Mathematics.
- Phanish Puranam, Oliver Alexy, and Markus Reitzig. What's "New" About New Forms of Organizing? *The Academy of Management Review*, 39(2):162–180, 2014. ISSN 0363-7425. URL <https://www.jstor.org/stable/43699235>. Publisher: Academy of Management.

- J. Cox, E. Oh, B. Simmons, C. Lintott, K. Masters, A. Greenhill, G. Graham, and K. Holmes. Defining and Measuring Success in Online Citizen Science: A Case Study of Zooniverse Projects. *Computing in Science and Engineering*, 17(4), 2015. ISSN 1521-9615. URL <https://ora.ox.ac.uk/objects/uuid:5885ed3d-a5ea-4208-bbd2-48f9b12a2575>. Publisher: IEEE.
- An MTurk Crisis? Shifts in Data Quality and the Impact on Study Results - Michael Chmielewski, Sarah C. Kucker, 2020, c. URL <https://journals.sagepub.com/doi/full/10.1177/1948550619875149>.
- Meliha Yetisgen-Yildiz, Imre Solti, Fei Xia, and Scott Halgrim. Preliminary Experiments with Amazon’s Mechanical Turk for Annotating Medical Named Entities. In Chris Callison-Burch and Mark Dredze, editors, *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon’s Mechanical Turk*, pages 180–183, Los Angeles, June 2010. Association for Computational Linguistics. URL <https://aclanthology.org/W10-0728>.
- Muki Haklay. Participatory citizen science. In Muki Haklay, Susanne Hecker, Anne Bowser, Zen Makuch, Johannes Vogel, and Aletta Bonn, editors, *Citizen Science, Innovation in Open Science, Society and Policy*, pages 52–62. UCL Press, 2018. ISBN 978-1-78735-235-3. URL <https://www.jstor.org/stable/j.ctv550cf2.11>.
- Steffen Fritz, Linda See, Tyler Carlson, Mordechai (Muki) Haklay, Jessie L. Oliver, Dilek Fraisl, Rosy Mondardini, Martin Brocklehurst, Lea A. Shanley, Sven Schade, Uta Wehn, Tommaso Abrate, Janet Anstee, Stephan Arnold, Matthew Billot, Jillian Campbell, Jessica Espey, Margaret Gold, Gerid Hager, Shan He, Libby Hepburn, Angel Hsu, Deborah Long, Joan Masó, Ian McCallum, Maina Muniafu, Inian Moorthy, Michael Obersteiner, Alison J. Parker, Maike Weisspflug, and Sarah West. Citizen science and the United Nations Sustainable Development Goals. *Nature Sustainability*, 2(10):922–930, October 2019. ISSN 2398-9629. doi: 10.1038/s41893-019-0390-3. URL <https://www.nature.com/articles/s41893-019-0390-3>. Publisher: Nature Publishing Group.
- Dilek Fraisl, Jillian Campbell, Linda See, Uta Wehn, Jessica Wardlaw, Margaret Gold, Inian Moorthy, Rosa Arias, Jaume Piera, Jessica L. Oliver, Joan Masó, Marianne Penker, and Steffen Fritz. Mapping citizen science contributions to the UN sustainable development goals. *Sustainability Science*, 15(6):1735–1751, November 2020. ISSN 1862-4057. doi: 10.1007/s11625-020-00833-7. URL <https://doi.org/10.1007/s11625-020-00833-7>.
- Camille Masselot, Bastian Greshake Tzovaras, Chris L. B. Graham, Gary Finnegan, Rathin Jeyaram, Isabelle Vitali, Thomas Landrain, and Marc Santolini. Implementing the Co-Immune Open Innovation Program to Address Vaccination Hesitancy and Access to Vaccines: Retrospective Study. *Journal of Participatory Medicine*, 14(1):e32125, January 2022. doi: 10.2196/32125. URL <https://jopm.jmir.org/2022/1/e32125>. Company: Journal of Participatory Medicine Distributor: Journal of Participatory Medicine Institution: Journal of Participatory Medicine Label: Journal of Participatory Medicine Publisher: JMIR Publications Inc., Toronto, Canada.

- Henry Mintzberg. The Structuring of Organizations. In David Asch and Cliff Bowman, editors, *Readings in Strategic Management*, pages 322–352. Macmillan Education UK, London, 1989. ISBN 978-1-349-20317-8. doi: 10.1007/978-1-349-20317-8_23. URL https://doi.org/10.1007/978-1-349-20317-8_23.
- Brian Uzzi, Satyam Mukherjee, Michael Stringer, and Ben Jones. Atypical Combinations and Scientific Impact. *Science*, 342(6157):468–472, October 2013. doi: 10.1126/science.1240474. URL <https://www.science.org/doi/10.1126/science.1240474>. Publisher: American Association for the Advancement of Science.
- Johannes Jaeger, Camille Masselot, Bastian Greshake Tzovaras, Enric Senabre Hidalgo, Mordechai (Muki) Haklay, and Marc Santolini. An epistemology for democratic citizen science. *Royal Society Open Science*, 10(11):231100. ISSN 2054-5703. doi: 10.1098/rsos.231100. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10646465/>.
- Carolin Haeussler and Henry Sauermann. Division of labor in collaborative knowledge production: The role of team size and interdisciplinarity. *Research Policy*, 49(6):103987, July 2020. ISSN 00487333. doi: 10.1016/j.respol.2020.103987. URL <https://linkinghub.elsevier.com/retrieve/pii/S0048733320300676>.
- Alfredo Yegros-Yegros, Ismael Rafols, and Pablo D’Este. Does Interdisciplinary Research Lead to Higher Citation Impact? The Different Effect of Proximal and Distal Interdisciplinarity. *PLOS ONE*, 10(8):e0135095, August 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0135095. URL <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0135095>. Publisher: Public Library of Science.
- Karin Hoisl, Marc Gruber, and Annamaria Conti. R&D team diversity and performance in hypercompetitive environments. *Strategic Management Journal*, 38(7):1455–1477, 2017. ISSN 1097-0266. doi: 10.1002/smj.2577. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/smj.2577>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/smj.2577>.
- Alva Taylor and Henrich R Greve. SUPERMAN OR THE FANTASTIC FOUR? KNOWLEDGE COMBINATION AND EXPERIENCE IN INNOVATIVE TEAMS. *Academy of Management Journal*.
- Michael Klug and James P. Bagrow. Understanding the group dynamics and success of teams. *Royal Society Open Science*, 3(4):160007, April 2016. ISSN 2054-5703. doi: 10.1098/rsos.160007. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4852640/>.
- Marlo Raveendran, Phanish Puranam, and Massimo Warglien. Division of Labor Through Self-Selection. *Organization Science*, 33(2):810–830, March 2022. ISSN 1047-7039. doi: 10.1287/orsc.2021.1449. URL <https://pubsonline.informs.org/doi/10.1287/orsc.2021.1449>. Publisher: INFORMS.

- Roger Guimerà, Brian Uzzi, Jarrett Spiro, and Luís A. Nunes Amaral. Team Assembly Mechanisms Determine Collaboration Network Structure and Team Performance. *Science*, 308(5722):697–702, April 2005. doi: 10.1126/science.1106340. URL <https://www.science.org/doi/full/10.1126/science.1106340>. Publisher: American Association for the Advancement of Science.
- Marlon Twyman and Noshir Contractor. Team Assembly. In Kara L. Hall, Amanda L. Vogel, and Robert T. Croyle, editors, *Strategies for Team Science Success: Handbook of Evidence-Based Principles for Cross-Disciplinary Science and Practical Lessons Learned from Health Researchers*, pages 217–240. Springer International Publishing, Cham, 2019. ISBN 978-3-030-20992-6. doi: 10.1007/978-3-030-20992-6_17. URL https://doi.org/10.1007/978-3-030-20992-6_17.
- Gary S. Becker and Kevin M. Murphy. The Division of Labor, Coordination Costs, and Knowledge*. *The Quarterly Journal of Economics*, 107(4):1137–1160, November 1992. ISSN 0033-5533. doi: 10.2307/2118383. URL <https://doi.org/10.2307/2118383>.
- Paul R. Lawrence and Jay W. Lorsch. Differentiation and Integration in Complex Organizations. *Administrative Science Quarterly*, 12(1):1–47, 1967. ISSN 0001-8392. doi: 10.2307/2391211. URL <https://www.jstor.org/stable/2391211>. Publisher: [Sage Publications, Inc., Johnson Graduate School of Management, Cornell University].
- Tobias Kretschmer and Phanish Puranam. Integration Through Incentives Within Differentiated Organizations. *Organization Science*, 19:860–875, December 2008. doi: 10.1287/orsc.1070.0352.
- Fengli Xu, Lingfei Wu, and James Evans. Flat teams drive scientific innovation. *Proceedings of the National Academy of Sciences*, 119(23):e2200927119, June 2022. doi: 10.1073/pnas.2200927119. URL <https://www.pnas.org/doi/abs/10.1073/pnas.2200927119>. Publisher: Proceedings of the National Academy of Sciences.
- Özgecan Koçak, Daniel A. Levinthal, and Phanish Puranam. The Dual Challenge of Search and Coordination for Organizational Adaptation: How Structures of Influence Matter. *Organization Science*, 34(2):851–869, March 2023. ISSN 1047-7039. doi: 10.1287/orsc.2022.1601. URL <https://pubsonline.informs.org/doi/full/10.1287/orsc.2022.1601>. Publisher: INFORMS.
- María J. Palazzi, Jordi Cabot, Javier Luis Cánovas Izquierdo, Albert Solé-Ribalta, and Javier Borge-Holthoefer. Online division of labour: emergent structures in Open Source Software. *Scientific Reports*, 9(1):13890, December 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-50463-y. URL <http://www.nature.com/articles/s41598-019-50463-y>.
- Floriana Gargiulo, Maria Castaldo, Tommaso Venturini, and Paolo Frasca. Distribution of labor, productivity and innovation in collaborative science. *Applied Network Science*, 7(1):1–15, December 2022. ISSN 2364-8228. doi: 10.1007/s41109-022-00456-0. URL <https://appliednetsci.springeropen.com/articles/10.1007/s41109-022-00456-0>. Number: 1 Publisher: SpringerOpen.

- John P. Walsh and Nancy G. Maloney. Collaboration Structure, Communication Media, and Problems in Scientific Work Teams. *Journal of Computer-Mediated Communication*, 12(2):712–732, January 2007. ISSN 1083-6101. doi: 10.1111/j.1083-6101.2007.00346.x. URL <https://doi.org/10.1111/j.1083-6101.2007.00346.x>.
- CRedit, d. URL <https://credit.niso.org/>.
- Henry Sauermann and Carolin Haeussler. Authorship and contribution disclosures. *Science Advances*, 3(11):e1700404, November 2017. ISSN 2375-2548. doi: 10.1126/sciadv.1700404.
- Vincent Larivière, David Pontille, and Cassidy R. Sugimoto. Investigating the division of scientific labor using the Contributor Roles Taxonomy (CRedit). In *Quantitative science studies*, volume 2, pages 111–128. MIT Press, 2020. doi: 10.1162/qss_a_00097. URL <https://papyrus.bib.umontreal.ca/xmlui/handle/1866/25787>. Accepted: 2021-11-03T11:49:58Z Issue: 1.
- Antonio Lima, Luca Rossi, and Mirco Musolesi. Coding Together at Scale: GitHub as a Collaborative Social Network. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1):295–304, May 2014. ISSN 2334-0770. doi: 10.1609/icwsm.v8i1.14552. URL <https://ojs.aaai.org/index.php/ICWSM/article/view/14552>. Number: 1.
- Taha Yasseri, Robert Sumi, András Rung, András Kornai, and János Kertész. Dynamics of Conflicts in Wikipedia. *PLoS ONE*, 7(6):e38869, June 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0038869. URL <https://dx.plos.org/10.1371/journal.pone.0038869>.
- Maximilian Klein, Thomas Maillart, and John Chuang. The Virtuous Circle of Wikipedia: Recursive Measures of Collaboration Structures. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, CSCW '15*, pages 1106–1115, New York, NY, USA, February 2015. Association for Computing Machinery. ISBN 978-1-4503-2922-4. doi: 10.1145/2675133.2675286. URL <https://doi.org/10.1145/2675133.2675286>.
- Brian Keegan, Darren Gergle, and Noshir Contractor. Do editors or articles drive collaboration?: multilevel statistical network analysis of wikipedia coauthorship. 2012.
- Oleksandra Poquet, Liubov Tupikina, and Marc Santolini. Are forum networks social networks? a methodological perspective: 10th International Conference on Learning Analytics and Knowledge: Shaping the Future of the Field, LAK 2020. *LAK 2020 Conference Proceedings - Celebrating 10 years of LAK*, pages 366–375, March 2020. doi: 10.1145/3375462.3375531. URL <http://www.scopus.com/inward/record.url?scp=85082391489&partnerID=8YFLogxK>. Publisher: Association for Computing Machinery.
- Roger Guimerà, Marta Sales-Pardo, and Luís A. Nunes Amaral. Module identification in bipartite and directed networks. *Physical Review E*, 76(3):036102, September 2007. doi: 10.1103/PhysRevE.76.036102. URL <https://link.aps.org/doi/10.1103/PhysRevE.76.036102>. Publisher: American Physical Society.

- Cristian Pasquaretta and Raphaël Jeanson. Division of labor as a bipartite network. *Behavioral Ecology*, 29(2):342–352, March 2018. ISSN 1045-2249. doi: 10.1093/beheco/arx170. URL <https://doi.org/10.1093/beheco/arx170>.
- Amy Wax, Leslie A. DeChurch, and Noshir S. Contractor. Self-Organizing Into Winning Teams: Understanding the Mechanisms That Drive Successful Collaborations. *Small Group Research*, 48(6):665–718, December 2017. ISSN 1046-4964. doi: 10.1177/1046496417724209. URL <https://doi.org/10.1177/1046496417724209>. Publisher: SAGE Publications Inc.
- Enric Senabre Hidalgo, Josep Perelló, Frank Becker, Isabelle Bonhoure, Martine Legris, and Anna Cigarini. Participation and Co-creation in Citizen Science. In Katrin Vohland, Anne Land-Zandstra, Luigi Ceccaroni, Rob Lemmens, Josep Perelló, Marisa Ponti, Roeland Samson, and Katherin Wagenknecht, editors, *The Science of Citizen Science*, pages 199–218. Springer International Publishing, Cham, 2021. ISBN 978-3-030-58278-4. doi: 10.1007/978-3-030-58278-4_11. URL https://doi.org/10.1007/978-3-030-58278-4_11.
- Ronald S. Burt. Structural Holes and Good Ideas. *American Journal of Sociology*, 110(2):349–399, 2004. URL <http://www.jstor.org/stable/10.1086/421787>.
- Balazs Vedres. Forbidden triads and creative success in jazz: the Miles Davis factor. *Applied Network Science*, 2(1):31, December 2017. ISSN 2364-8228. doi: 10.1007/s41109-017-0051-2. URL <http://appliednetsci.springeropen.com/articles/10.1007/s41109-017-0051-2>.
- Balazs Vedres. Network mechanisms in innovation: borrowing and sparking ideas around structural holes. *SSRN Electronic Journal*, 2021. ISSN 1556-5068. doi: 10.2139/ssrn.3878902. URL <https://www.ssrn.com/abstract=3878902>.
- Emmanuel Lazega, Marie-Thérèse Jourda, Lise Mounier, and Rafaël Stofer. Catching up with big fish in the big pond? Multi-level network analysis through linked design.
- Markus C. Becker. Organizational Routines and Organizational Learning. In Linda Argote and John M. Levine, editors, *The Oxford Handbook of Group and Organizational Learning*, page 0. Oxford University Press, January 2020. ISBN 978-0-19-026336-2. doi: 10.1093/oxfordhb/9780190263362.013.4. URL <https://doi.org/10.1093/oxfordhb/9780190263362.013.4>.
- Phanish Puranam and Boris Maciejovsky. Organizational Structure and Organizational Learning. In Linda Argote and John M. Levine, editors, *The Oxford Handbook of Group and Organizational Learning*, page 0. Oxford University Press, January 2020. ISBN 978-0-19-026336-2. doi: 10.1093/oxfordhb/9780190263362.013.35. URL <https://doi.org/10.1093/oxfordhb/9780190263362.013.35>.
- Javier Borge-Holthoefer, Raquel A. Baños, Carlos Gracia-Lázaro, and Yamir Moreno. Emergence of consensus as a modular-to-nested transition in communication dynamics. *Scientific Reports*, 7(1):41673, January 2017. ISSN 2045-2322. doi: 10.1038/srep41673. URL <https://www.nature.com/articles/srep41673>. Publisher: Nature Publishing Group.

- Ashwin Paranjape, Austin R. Benson, and Jure Leskovec. Motifs in Temporal Networks. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pages 601–610, February 2017. doi: 10.1145/3018661.3018731. URL <http://arxiv.org/abs/1612.09259>. arXiv:1612.09259 [physics, stat].
- Vincent Larivière, David Pontille, and Cassidy R. Sugimoto. Investigating the division of scientific labor using the Contributor Roles Taxonomy (CRediT). *Quantitative Science Studies*, 2(1):111–128, April 2021. ISSN 2641-3337. doi: 10.1162/qss_a_00097. URL https://doi.org/10.1162/qss_a_00097.
- Henry William Chesbrough. *Open Innovation: The New Imperative for Creating and Profiting from Technology*. Harvard Business Press, 2003. ISBN 978-1-57851-837-1. Google-Books-ID: 4hTRWStFhVgC.
- Joshua Borycz, Robert Olendorf, Alison Specht, Bruce Grant, Kevin Crowston, Carol Tenopir, Suzie Allard, Natalie M. Rice, Rachael Hu, and Robert J. Sandusky. Perceived benefits of open data are improving but scientists still lack resources, skills, and rewards. *Humanities and Social Sciences Communications*, 10(1):1–12, June 2023. ISSN 2662-9992. doi: 10.1057/s41599-023-01831-7. URL <https://www.nature.com/articles/s41599-023-01831-7>. Publisher: Palgrave.
- A Multi-Level Systems Perspective for the Science of Team Science, e. URL <https://www.science.org/doi/10.1126/scitranslmed.3001399>.
- John P. Walsh and You-Na Lee. The bureaucratization of science. *Research Policy*, 44(8):1584–1600, October 2015. ISSN 00487333. doi: 10.1016/j.respol.2015.04.010. URL <https://linkinghub.elsevier.com/retrieve/pii/S0048733315000700>.
- Gerben S. Van Der Vegt and J. Stuart Bunderson. Learning and Performance in Multidisciplinary Teams: The Importance of Collective Team Identification. *The Academy of Management Journal*, 48(3):532–547, 2005. ISSN 0001-4273. URL <https://www.jstor.org/stable/20159674>. Publisher: Academy of Management.
- Richard A. Guzzo and Marcus W. Dickson. TEAMS IN ORGANIZATIONS: Recent Research on Performance and Effectiveness. *Annual Review of Psychology*, 47(Volume 47, 1996): 307–338, February 1996. ISSN 0066-4308, 1545-2085. doi: 10.1146/annurev.psych.47.1.307. URL <https://www.annualreviews.org/content/journals/10.1146/annurev.psych.47.1.307>. Publisher: Annual Reviews.
- Paul E. Spector. *Job Satisfaction: Application, Assessment, Causes, and Consequences*. SAGE, March 1997. ISBN 978-0-7619-8923-3. Google-Books-ID: nCkXMZjs0XcC.
- Henry Chesbrough, Wim Vanhaverbeke, Joel West, Henry Chesbrough, Wim Vanhaverbeke, and Joel West, editors. *New Frontiers in Open Innovation*. Oxford University Press, Oxford, New York, March 2017. ISBN 978-0-19-880399-7.

Chapter 2

Monitoring Citizen Science Teams using Digital Traces

Preface

Open Science and Innovation initiatives focus on training students and the general public in collaborative knowledge production and facilitate them in tackling global societal problems. Organising initiatives with these dual targets necessitates building robust strategies to monitor teams, understand how they interact with other actors and engage with different facets of the initiative. This chapter explores how monitoring and evaluation strategies can be implemented to curate qualitative and quantitative interaction data, informing on collaboration, diversity and engagement of teams and how these inferences can be feedback into designing interventions to support teams. I focus on the Crowd4SDG citizen science incubator as the testbed for the study and analysis.

Crowd4SDG is a Horizon 2020 project that aims to train student teams in building a citizen science project addressing the Sustainable Development Goal (SDG) challenges. The project achieves this by organising a yearly innovation cycle with 4 stages - called GEAR (Gather, Evaluate, Accelerate and Refine). Students are assembled into teams, coached and are provided with technical and digital toolkits to conceptualise and prototype their citizen science project. At the end of each stage, teams are evaluated and selected into the next one. Teams that progress to the final stage are invited to present their projects to research communities, prospective sponsors and stakeholders.

The Crowd4SDG project ran over the course of 3 cycles with multifaceted objectives. Primarily, the GEAR cycle is a training and facilitation framework, a key objective is to evaluate the methodology, and how developing new citizen science toolkits and training strategies foster the building of innovative projects. In addition to the efficacy of the GEAR cycle, it is key to understand how teams remain engaged with the program and what are the predictors of innovation within the collaborative challenge-based innovation ecosys-

tem. This allows facilitators to reconfigure their strategies and offer future perspectives on designing similar frameworks. We achieve this by monitoring and studying teams through their progress within the GEAR framework using a multifaceted approach - including qualitative surveys and digital traces from web tools.

The chapter is split into three broad sections. In the first section, I present our work in studying teams participating in the GEAR cycle framework. The second *Evaluate* phase of GEAR is the stage where participants are grouped into teams and offered training over the course of 5 weeks. The section presents the monitoring strategies leveraged during this phase to study participating citizen science teams and we highlight how team composition, collaboration and activity patterns are associated with their engagement within the framework and the quality of their project. In the second section, I present our work building statistical models associating collaboration dynamics and project performance - leveraging data from the *Evaluate* phase of GEAR cycles 2 and 3 (which had the formalised monitoring strategies, excluding the pilot GEAR cycle 1). Finally in the third section, I discuss the needs of practitioners in organising the GEAR cycles and how we design visualisation tools to help them monitor participants on a regular basis.

2.1 Collaboration and Performance of Citizen Science Projects Addressing the Sustainable Development Goals

Abstract

Measuring the progress towards the Sustainable Development Goals (SDGs) requires the collection of relevant and reliable data. To do so, Citizen Science can provide an essential source of non-traditional data for tracking progress towards the SDGs, as well as generate social innovations that enable such progress. At its core, citizen science relies on participatory processes involving the collaboration of stakeholders with diverse standpoints, skills, and backgrounds. The ability to measure these participatory processes is therefore key for the monitoring and evaluation of citizen science projects and to support the decisions of their coordinators. Here, we show that the monitoring of social interaction networks provides unique insights on the participatory processes and outcomes of citizen science projects. We studied fourteen early-stage citizen science projects that participated in an innovation cycle focused on SDG 13, Climate Action, as part of the Crowd4SDG project. We implemented a monitoring strategy to measure the collaborative profiles of citizen science teams. This allowed us to generate dynamic interaction networks across complementary dimensions, making visible both formal and informal interactions associated with the division of labor, collaborations, advice seeking, and communication processes of the projects during their development. Leveraging jury evaluation data, we showed that while team composition and communication are associated with project quality, measures of collaboration and activity are associated with engagement quality. Overall, monitoring social interaction dynamics helps build a more comprehensive picture of participatory processes, which is of importance for guiding citizen science projects and for designing initiatives leveraging citizen science to address the SDGs.

Contributions and Publication

The work presented here was published in the journal *Citizen Science: Theory and Practice* in 2023. The following section reports the contents of this article as is, reformatted to the style of the thesis.

MASSELOT, C.*, **JEYARAM, R.***, TACKX, R., FERNANDEZ-MARQUEZ, J.L., GREY, F. AND SANTOLINI, M. (2023) 'COLLABORATION AND PERFORMANCE OF CITIZEN SCIENCE PROJECTS ADDRESSING THE SUSTAINABLE DEVELOPMENT GOALS', *CITIZEN SCIENCE: THEORY AND PRACTICE*, 8(1), p. 45.

I took lead on collection (with the help of Raphael Tackx and Camille Masselot) and curation of the data, performing the analysis and visualisation. Credit to Marc Santolini for conceptualising and supervising the study and to Camille Masselot and Marc Santolini for writing the manuscript.

2.1.1 Introduction

The United Nations Sustainable Development Goals (SDGs) are a series of development targets designed to address the world’s most pressing societal, environmental, and economic challenges by 2030. Measuring their progress requires the obtention of timely, relevant, reliable data across a multitude of stakeholders. By engaging in scientific activities, citizens can foster the progress towards the resolution of the SDGs [Fritz et al., 2019], for example by generating evidence to identify gaps in their monitoring [Franzoni et al., 2022], collecting and analyzing data to support the decisions taken by local and national stakeholders [Ballerini and Bergh, 2021, Fraisl et al., 2020], and accelerating the development of solutions [Kokshagina, 2022, Masselot et al., 2022].

Compared with traditional scientific work, citizen science requires the definition of processes of engagement and coordination, from simple data collection to co-design strategies [Franzoni and Sauermaun, 2014, Haklay, 2018, Senabre Hidalgo et al., 2021]. Frameworks have recently been developed to assess the impact of citizen science projects towards the SDGs [Parkinson et al., 2022], to understand modes of co-production for sustainability [Chambers et al., 2021] and to evaluate success of online teams in terms of scientific contribution and public engagement [Cox et al., 2015]. Yet those frameworks often apply to large teams or to advanced projects, and organizers of initiatives such as Crowd4SDG [noa, a] lack supporting evidence to guide their practice in forming and coordinating successful citizen science projects at early stages. The evaluation of participatory processes such as those involved in citizen science emphasizes measures of diversity, engagement, collaboration, and learning [Jaeger et al., Schaefer et al., 2021]. The ability to measure these participatory processes is therefore key for the monitoring and evaluation of citizen science projects.

Mixed methods involving digital traces and questionnaires are traditionally used in social studies describing collaborative activities, for example to understand how social networks shape individual performance in collaborative learning [Poquet et al., 2020], or to describe how team interactions and community organization shape collective performance within open research programs [Kokshagina, 2022, Masselot et al., 2022] or open source communities [Gargiulo et al., 2022, Klein et al., 2015, Klug and Bagrow, 2016]. Complementing digital traces, the collection of self-reported data yields qualitative insights across perceived interactions [Deri et al., 2018]. However, building a comprehensive group-scale network requires the engagement of a large proportion of the participants involved in the self-report activity in order to accurately represent the social network, calling for specific survey instruments that allow for collection of the social ties of a participant while minimizing survey burden. The recent availability of such survey instruments has rendered collaborative network data collection easier and scalable, allowing researchers to capture temporal organizational networks within groups of various sizes [Tackx et al., 2021].

In this study, we tackle the question of how participatory processes shape project performance within early citizen science projects addressing the SDGs. To do so, we focused on Crowd4SDG, a European project that guides young individuals from pitching an idea

on a social platform to the design of prototypes of citizen science projects via a one-year cycle of innovation. We developed and implemented a framework to monitor the activity and collaborations across 14 citizen science projects from the Crowd4SDG project. We highlight how this framework generated complementary interaction networks informing on the division of labor, collaborations, advice seeking, and communication processes of the citizen science projects. Finally, we show the usefulness of these measures for monitoring engagement and supporting the evaluation process and discuss how this framework could be used in future programs

2.1.2 Methods

Description of the Gear Cycle

The Crowd4SDG project organizes three one-year cycles of innovation, aimed at coaching teams of young social entrepreneurs through the steps of building a citizen science project. Each project tackles a challenge related to Climate Action or involves crowdsourcing tools that can generate data relevant for tracking progress towards the SDGs. The innovation methodology used follows a “GEAR Methodology” to coach teams through the innovation process required to develop new citizen science projects (Figure 2.1 a). Each GEAR cycle includes several phases of online coaching and in-person support: Gather, Evaluate, Accelerate and Refine. The Gather phase is promoted as a global crowdsourcing of ideas, called the Open17 Challenge (Figure 2.1 b), on the social network Goodwall. Some participants entered the phase with their own team, others were assigned teammates by a teaming algorithm (see Supplemental File 2: Teaming Algorithm for the parameters used in the algorithm). At the start of the Evaluate phase, 30 to 50 participants are selected to enroll in the Open17 weekly coaching during which they learn about developing and pitching their citizen science project. The best teams then benefit from a Challenge-Based Innovation Workshop (CBIW), which focuses on building a working prototype for the project, using crowdsourcing tools developed by the Crowd4SDG consortium partners and other relevant ones. The most promising projects are invited to participate in the Geneva Dialogue, an opportunity to meet sponsors and potential partners amongst the international organizations in Geneva. Each phase of the GEAR methodology filters projects based on their novelty, relevance, feasibility, and appropriate use of crowdsourcing tools, and helps participants advance towards practical deployment.

The citizen science projects developed in the three GEAR cycles of Crowd4SDG aim to address the nexus between Climate Action (SDG 13) and several other key SDGs: Sustainable Cities (SDG 11 in 2020), Gender Equality (SDG 5 in 2021), and Peace, Justice and Strong Institutions (SDG 16 in 2022).

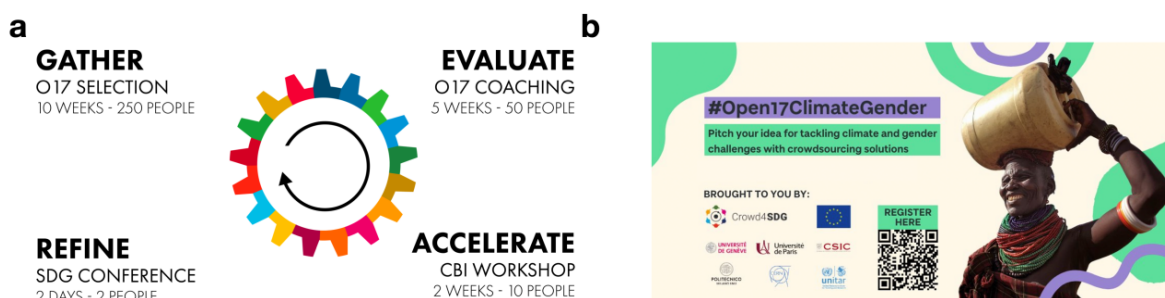


Figure 2.1: Description of the challenge. (a) Schematics of a GEAR cycle. (b) Visual for the GEAR Cycle 2.

Description of the Team Projects

In this work, we study the “Evaluate” phase of the GEAR cycle 2, the coaching program where teams ideate their project and engage in interactions with their peers and mentors. This focus allowed us to gather data on a large enough sample of projects corresponding to 14 teams. We document in Table 1 the objectives of the projects, along with team size and the final stage of the challenge they achieved.

Communication Data

A Slack workspace was used by the teams during the GEAR cycle as a means to communicate with other teams and with the organizing team.

The data was extracted in JSON format using the export function available to the owners/admins of the Slack workspace. This allowed us to gather a data frame containing across all public channels the messages (post contents), their time stamp, sender, and channel it was sent in. The raw data was then processed to obtain mentions. A mention occurs when a Slack user types in a message the Slack username of a target user prefixed by “@”. Each recorded mention has information on the source (who wrote the message), target (who is being mentioned), and the timestamp (when the message was sent). Slack also allows users to broadcast messages by citing all users in a channel or a workspace by using specific commands (@all, @here, @channel_name). These were not included as mentions in order to focus on direct interactions only.

Survey Data

We used two types of surveys: those related to participant attributes (e.g., their background or country of origin), and those related to participant interactions (e.g., who they collaborated with or sought advice from).

The initial survey was related to attributes only and was disseminated using a Google

Team Name	Team size	Project objectives	Furthest stage achieved
WOMER	2	Assess the effect of climate change on women in indigenous communities in Colombia Empower indigenous women via their engagement in citizen science	Refine
DonateWater	4	Map the operating status of water handpumps in rural communities of Nigeria using crowdsourcing techniques. Discuss changes in water access solutions	Refine
Andapé Institute	3	Map the accessibility of sidewalks and their conditions in the urban environment in Brazil. Design an ecological sidewalk from recycled materials	Accelerate
Climate Gender Justice	2	Generate area-specific data related to climate change impact on sexual and reproductive health and rights	Accelerate
Water Warriors	3	Organize and classify emergencies and the responses adopted to solve them	Accelerate
Women & Technology	3	Map women participation in climate change projects connected to technology	Evaluate
Against Climate Change	3	Map women participation in climate change projects connected to technology	Evaluate
SDesiGn	2	Increase gender equality in architecture and product designs that prioritize environment protection	Evaluate
UpGet app - CitiCERN Project	2	Build a digital wallet to get eco-friendly rewards	Evaluate
Eco Winners	3	Build Digital Platform where farmers in Nigeria - of which 70% are women, can rent farming tools and equipments at low cost on a peer to peer basis	Evaluate
Women 4 Sustainable World	2	Set up meteorological data collection kits that send alerts to users on their mobile phones on favorable periods for the seed, make the treatments with fertilizers or pesticides, the type of seeds adapted to the soil followed by an estimate	Evaluate
TEAM FOILED	3	Allow affected communities to crowdsource damages caused by floods and access help faster	Evaluate
PAM	3	Map and classify climate disasters and give a peculiar attention to women's voices	Evaluate
Rights of Climate	2	Assess the link between Women Land Rights and their vulnerability to Climate change in Malawi	Evaluate
Flood Rangers	2	Generate gender segregated data related to the impact of flood on men and women in Nigeria	Evaluate

Table 2.1: Gear Cycle 2 project description

Form at registration to the Evaluate phase.

We then disseminated 4 weekly surveys related to social interactions and activities using the CoSo platform [Tackx et al., 2021] (Supplemental File 2: Figure S1). The CoSo platform is designed to collect self-reported interaction data with a simple, reactive interface, and an analysis-ready database. To document their interactions, the users could select target users across all other participants and organizers. The interactions spanned prior ties in the first survey (“Which of these people did you know personally before?”), and on a weekly basis their advice-seeking behavior (“Who did you seek advice from last week?”) and collaborations (“Who did you work with last week?”). To document their activity, they could select across 26 activities encompassing routine activities within research teams inspired from the CRediT contribution taxonomy, as well as specific questions regarding Crowd4SDG, for example specific tool usage. Activities encompassed different levels of complexity in their realization. They ranged from tasks that could be performed in a distributed fashion such as preparing the final pitch and analyzing data, to tasks involving higher levels of collaboration such as brainstorming.

The surveys were advertised through Slack, and the organizing team dedicated 10 minutes for participants to fill them during weekly sessions, ensuring a high engagement: Except for the team “Flood Rangers,” who answered only one CoSo survey before dropping out from the program, at least one member of each team answered surveys each week (Supplementary File 2: Table 1).

Team Features

To document measures of participation, we monitored features related to team composition, communication, collaboration, and activity (features from Figure 2.7 are in italics below).

For team composition, we built measures of size, diversity, education level, and prior experience. *Team size* was assessed using the number of members of a team. Background diversity was assessed by computing the *background span* – that is, the number of unique academic backgrounds in the team as declared in the registration form. The *education level* was computed by taking the average level of education in a team based on the response to the question “What is your current or highest level of education,” to which we attributed the following score based on the answer: 0 for secondary school, 1 for high school, 2 for undergraduate, and 3 for graduate. Finally, *prior experience* was computed as the average answer to the question, “Have you participated in data projects or contributed as citizen scientist to data production before?” (yes = 1 and no = 0) within each team.

For communication, we leveraged the activity and interactions on Slack public channels. The *Slack activity* was assessed as the total number of messages posted by team members. For interactivity, we measure *Slack interaction intra team* as the number of mentions among members of a team, and *Slack interaction organizing team* as the number of mentions between members of a team with the organizing team. We counted mentions regardless of their directionality.

For collaborations, we focused on the number of collaborations within the teams, as well as the centrality of the teams within the advice network. For the intra-team collaborations (*coso interaction intra-team*), we summed for each team the weights of the intra-team edges in the “work with” collaboration network. For the centrality in the advice network, we computed the Burt constraint [Burt, 2004], a measure of social capital that takes low values when a neighborhood is diverse (ties to separate neighborhoods), and higher values when the neighborhood is constrained (dense ties to the same neighborhood). *Advice diversity* was computed by taking the negative of Burt constraint, with higher values indicating higher levels of diversity (more structural holes). This quantifies the ability of a team to leverage diverse sources of information for advice seeking.

Finally, for the activity, we focused on measures of diversity and engagement of activities performed. For diversity, we computed the *activity span* as the proportion of activities performed by a team among the 26 listed. For engagement, we considered the *activity regularity* by first computing the number of activities reported by a team each week, and then computing the negative of the Gini index on the resulting vector. The Gini index ranges from 0 (perfectly regular) to 1 (perfectly irregular). 1-Gini is higher if activities are regularly conducted across weeks. Finally, we quantified for each team the *survey engagement* as the proportion of survey responses per team across all CoSo surveys, a measure of engagement to the study.

Team Performance Data

To quantify team performance, we used the scores that teams obtained in their assessment by the jury and the Crowd4SDG organizing team (features from Figure 2.7 are in italics below).

Performance was assessed through 4 team outcome metrics (crowdsourcing component, feasibility, relevance, novelty) as judged by a panel of experts selected by the Crowd4SDG consortium, and by 4 process parameters (project documentation, members attendance, commitment, and weekly evaluation) as assessed by the organizing team.

More precisely, *crowdsourcing* was assessed using the mean score attributed to the question “Is there an effective crowdsourcing component?” (yes = 1 and no = 0). We measured the *feasibility*, *relevance*, and *novelty* by computing the mean score attributed by the jury on a scale from 0 to 5 to the questions “Feasibility: Is the project implementable with reasonable time and effort from the team?,” “Novelty: Is the pitch based on a new idea or concept or using existing concepts in a new context?,” and “Relevance: Is the solution proposed relevant to the challenge or potentially impactful?”

In terms of process, all variables were integer values with scores ranging from 0 to 5 for deliverables and attendance, 0 and 1 for commitment. For *weekly evaluation* the score was a continuous value ranging from 0 to 10 scoring the overall quality of their weekly pitch sessions. *Deliverables* measured the total number of deliverables submitted and documented on the platform Innprogress (<https://innprogresstest.unige.ch/>) among the expected ones. *Attendance* was estimated by the proportion of sessions attended by team members. *Commitment* was scored 1 if teams were willing to continue their project after the end of Evaluate, or 0 otherwise.

Network Construction

The SDG and background networks in Figure 2 are built using the co-occurrence of answers in multiple choice questions across participants in the first survey. In these networks, two nodes are linked if they are co-cited in an answer, and the weight of the link is the number of participants who reported this co-occurrence. The SDGs were declared in questions related to past projects: “Have you contributed to projects on SDGs before,” and “Which of the SDGs was the project addressing? Select all that apply.” The backgrounds were selected across a multiple-choice question asking “What are your main fields of work or study? Select all that apply.”

The Slack mention network links a user A to a user B if they mention them, with a weight corresponding to the number of times A has mentioned B. When aggregating at the team level, intra-team mentions are encoded as self-loops with an edge weight given by the sum of the intra-team links weights.

The CoSo networks are directly inferred from the surveys. In Figures 2.5 and 2.6, we aggregated the networks over all timepoints collected, yielding weighted interaction networks where edge weights correspond to the number of times an interaction was reported.

Figure 2.6 further aggregates the individual networks at the team level. In Figure 2.6 c, centralization is computed as the Gini coefficient of the degree distribution. More precisely, we compute for each team its undirected, total degree (number of neighbors). We then compute the Gini coefficient of the degrees. Its value ranges from 0 (all degrees equal) to 1 (one team dominates the degree distribution), and indicates the degree to which interactions are concentrated towards one “hub” node in the network. Network centrality measures were computed using the *igraph* library in R. Network visualizations were produced with Gephi 0.9.7 with a force layout.

Statistical Analysis

Data were analyzed using the R software. The association between performance and team features was assessed using Pearson’s correlations. The p value for the correlation is calculated by computing the t statistic (*cor.test* function in R), with the null hypothesis that the correlation between the dependent and independent variable is 0. The level of significance was set at $p = 0.1$

2.1.3 Results

In this study, we provide an analysis of teamwork and collaborations during the second GEAR cycle of innovation from the Crowd4SDG project. The citizen science projects developed in this GEAR cycle aim to address the nexus between Climate Action (SDG 13) and Gender Equality (SDG 5). To study the process of generation of citizen science projects in this context, we focus on the “Evaluate” phase (see Methods), a coaching program in which teams build their project and engage in interactions with their peers and mentors.

Cohort Description

The cohort was made up of a total of 38 participants covering 17 nationalities (Supplemental File 2: Figure S2). A total of 14 teams were formed with sizes varying from 1 to 4 members and showing diversity in terms of age and gender (Supplemental File 2: Figure S3b). All teams comprised students, mostly at the university level (12 out of 14), with some high school–level teams (2/14). Overall, 68% of participants were younger than 25 years old, with an age range of 16 to 32 years old (Supplemental File 2: Figure S3a).

Some participants had prior experience with citizen science, with 26% (10/38) answering positively to the question “Have you participated in data projects or contributed as citizen scientist to data production before?” Moreover, the prior experience of participants with SDGs covered most goals, with a primary focus on Climate Action and Gender Equality, as expected from the topic of the GEAR cycle (Figure 2.2 a). Interestingly, the participants (and the teams themselves) displayed a high level of interdisciplinarity, with backgrounds spanning natural sciences, technology, and humanities (Figure 2.2 band S2).

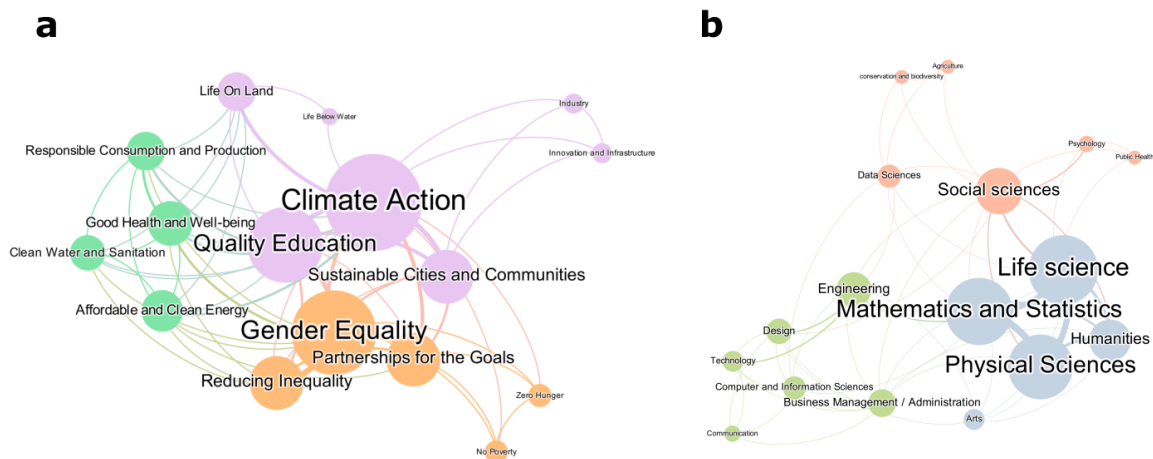


Figure 2.2: **Description of the cohort.** Co-occurrence networks across participants highlighting (a) their previous experience with Sustainable Development Goals and (b) their main background or field of study. In these networks nodes are linked by the number of times they are co-reported by a participant, and colors correspond to denser sub-networks as determined by the modularity algorithm. Participants had prior experience with Climate Action and Gender Equality, and came from interdisciplinary backgrounds.

Team Communication

During the Evaluate phase, teams used a Slack workspace to discuss with other team members from their own or from another team, as well as with the organizing team. Since the challenge was fully conducted online, this workspace was a central repository for communications at the cohort level. We analyzed the data from the public channels of the Slack workspace to study the patterns of engagement of participants within and across teams, as well as with organizers.

We first observed that the activity of the Slack workspace, measured by the number of posts per week, closely follows the phases of the GEAR cycle, with low activity outside of the phases (Figure 2.3 a). This might be because teams would work solely during the program, or because they would synchronize on other communication channels outside of these phases, such as Whatsapp, e-mails, or private Slack conversations (Supplemental File 2: Figure S4).

To examine the interaction dynamics between participants, we used a network approach. This allows for representation of the flow of information characterizing this phase, in particular highlighting the interactions with the organization team. We computed the number of mentions of a “target” participant B from a “source” participant A as an indication of a directed interaction from A to B. Mapping participants to their respective teams, we derived a directed, weighted network indicating the interaction strength between and within teams (Figure 2.3 b).

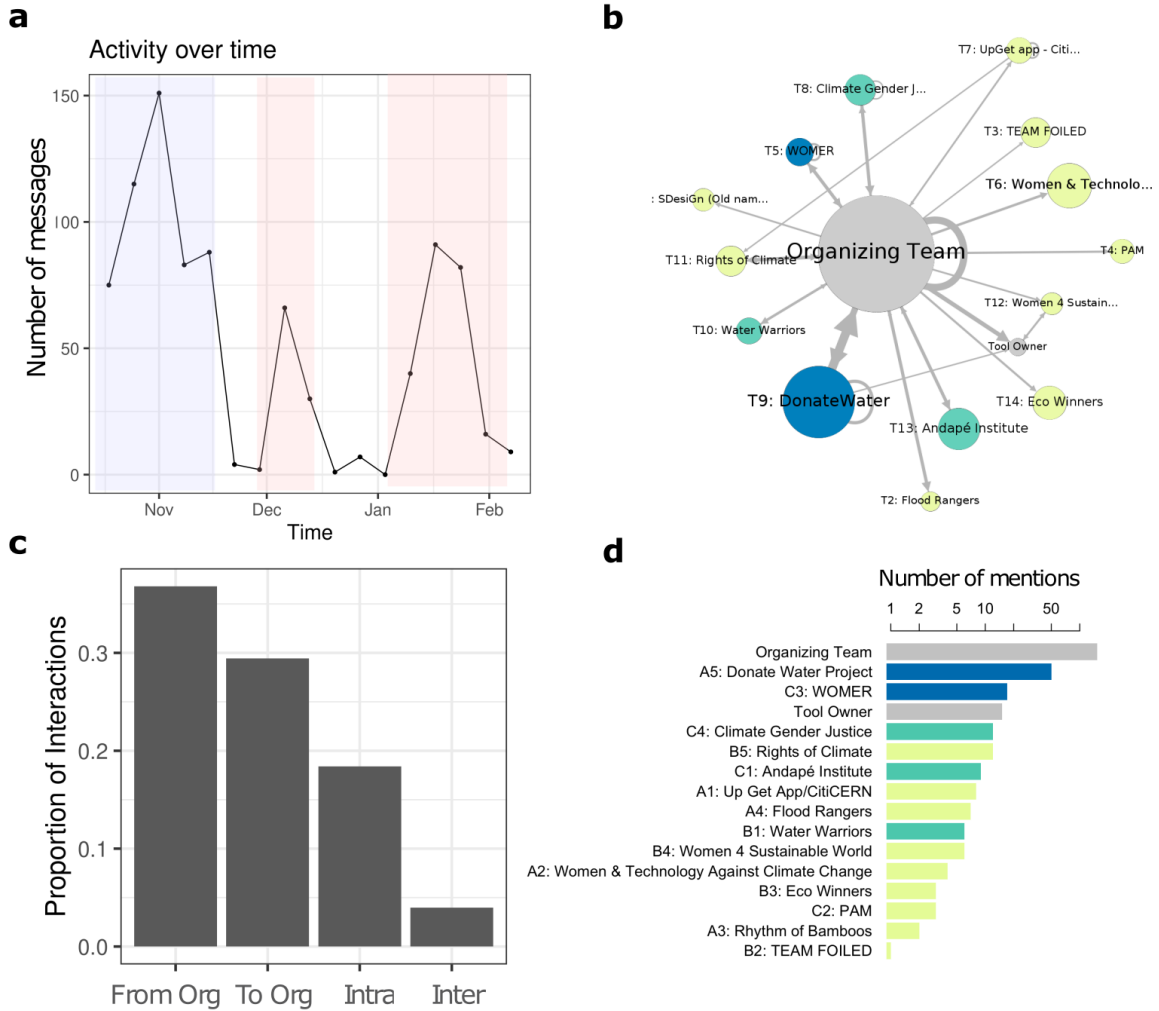


Figure 2.3: **Communication activity.** a) Total number of posts on Slack per week. The Evaluate phase is highlighted in blue, and the Accelerate phase, consisting of two periods, is highlighted in red. (b) Mention network extracted from Slack during the Evaluate phase. Nodes represent aggregated individuals at the team level. Teams are linked by weighted edges quantifying the number of times an individual from one team mentions an individual from another team. Self-loops denote intra-team interactions. Grey color denotes the organizing team, and the color denotes the stage achieved in the program: in order, Evaluate (yellow), Accelerate (green), Refine (blue). (c) Proportion of Slack mentions that are from the organization team, towards the organization team, intra-team or inter-team.(d) Number of mentions (sent or received) per team, following the color code from b.

We observe a high centralization of interactions to the organizing team (Figure 2.3 c), both in terms of incoming (teams reaching out to organizers) and outgoing links (organizers reaching out to teams). While the workspace was also used for within-teams interactions, there were very few inter-team interactions (Figure 2.3 c), confirming that the workspace was mostly used as a means to interact with organizers.

Beyond the organizing team, we found that the two teams that were eventually selected as finalists of the program, Donate Water Project and WOMER, had the highest network centrality teams when considering their weighted degree (i.e., the total number of incoming and outgoing mentions they partake in), suggesting that team level of engagement early in the program is important for project success.

Team Activities

While Slack informs on participant engagement and their interactions with organizers, it does not provide information on what activities teams perform, or what type of (informal and formal) interactions occur. Such information can help guide coordinators in managing citizen science communities. To gather deeper insights into team dynamics, we performed weekly surveys on activities performed and on collaborations during the four weeks of the Evaluate phase preceding the presentations to the jury.

The activities most performed were consistent with the purpose of the Evaluate phase: coaching teams into generating a feasible, novel citizen science project. As such, the main activity performed across the 4 weeks was the preparation of the final pitch (Figure 2.4 a). The early weeks were enriched in activities related to brainstorming and ideation, task planning, team building and literature review, while later weeks showed activities related to the preparation of documentation material and result interpretation. Moreover, it is interesting to note a significant number of participants declared “Meeting with people affected by the problem you are trying to solve” during the 4 weeks, a marker of engagement with stakeholders. The number of activities and their regularity varied widely across teams (Supplemental File 2: Figure S5), with an overall stronger push at the last week, suggesting a deadline effect (Figure 2.4b).

Collaboration Dynamics

Beyond activities performed, the surveys enquired about formal (“who did you work with?”) and informal (“who did you know before?” and “who did you seek advice from?”) interactions (Figure 2.5 a,b,c). These surveys were aimed at investigating the collaborative dynamics during the GEAR, its evolution in time, and eventual impact on team performance.

In the GEAR cycle, participants could join as a team, or as individuals. The latter were

assigned to a team using a matching algorithm (see Supplemental File 2: Teaming Algorithm). The existence of pre-formed teams is revealed in the “prior ties” network (Figure 2.5 a). Yet, beyond intra-team links, we found that several participants acted as bridges between teams in the prior ties network. This is probably because the Gather phase was able to tap into already existing communities, in particular through the social platform Goodwall.

Work collaborations occurred mostly within teams, as well as with organizers (Figure 2.4 b,d), while only few inter-team interactions were observed. However, advice seeking interactions, in which participants report having asked for advice from another participant, showed more inter-team interactions, with around 10% of them being inter-team ties (Figure 2.4 c,e). Moreover, while participants sought primarily advice within their own team in the first week, they gradually increased their outreach to the organizers, eventually constituting 55% of interactions. In both networks, organizers occupied the most central position, acting as bridges between teams.

Comparison of the Interaction Networks

The collected data allowed us to infer 4 interaction networks: communications from Slack mentions, and prior ties, collaborations, and advice seeking from surveys. When aggregated at the team level, these constitute a “multiplex” network, with the same nodes (the teams) having different types of links. Here, we question whether these networks provide similar or complementary information to inform on team behavior.

We show in Figure 2.6 a the networks at the team level. One can observe that the networks have similar densities (Figure 2.6 c), but different structures: The Slack mentions network is much more centralized than the surveyed interaction networks (Figure 2.6 c), indicating that Slack usage was mostly used to exchange with the organizing team who acted as a strong hub. When aggregating the networks, one obtains a more comprehensive interaction network (Figure 2.6 b), doubling the density of links compared with any single network (Figure 2.6 c).

To further assess the topological similarity between the networks obtained, we computed the Jaccard similarity between any pair of networks, that is, the ratio of the number of links in common (intersection) to the total number of links present in both networks (union). Completely dissimilar networks would have a Jaccard of $J = 0$, while identical networks have $J = 1$. We find that the collaboration (“work with”) and advice seeking networks are the most similar ($J = 0.74$), while their similarity with the Slack mention network is much smaller ($J = 0.2$). Prior ties are predictive of collaboration and advice seeking ($J = 0.2$) but not of Slack mentions ($J = 0$), which is probably due to the fact that most interactions on Slack were with the organizer team. Finally, the networks show a similarity to the full, aggregated networks ranging between $J = 0.4$ – 0.5 , indicating that a network measured with a single method encapsulates less than half of the information about formal and informal social interactions.

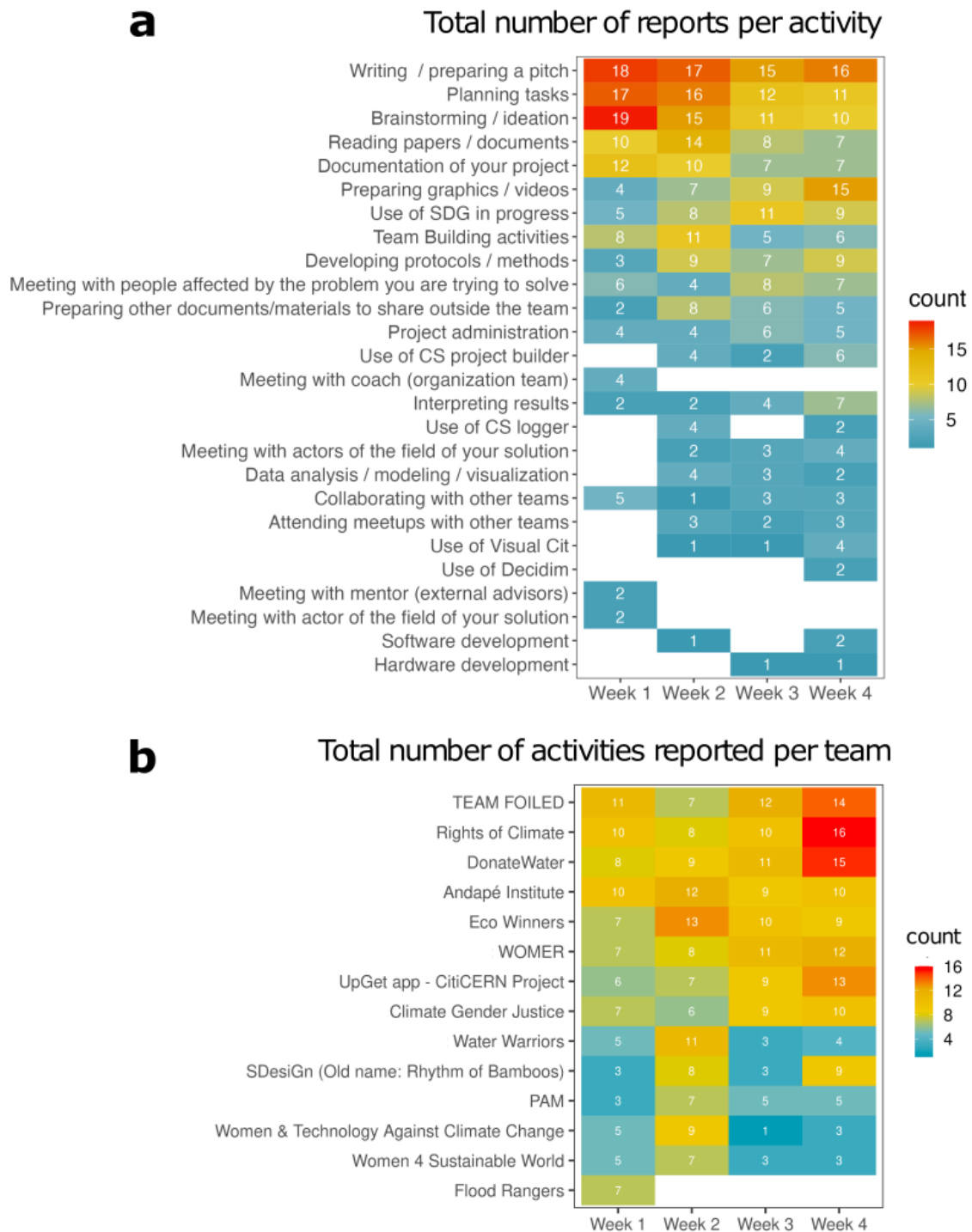


Figure 2.4: **Activities during the Evaluate phase.** (a) Total number of reports of an activity per week. We see a switch from brainstorming/ planning activities to the documentation and preparation of the final presentation. (b) Total number of activities reported per team per week.

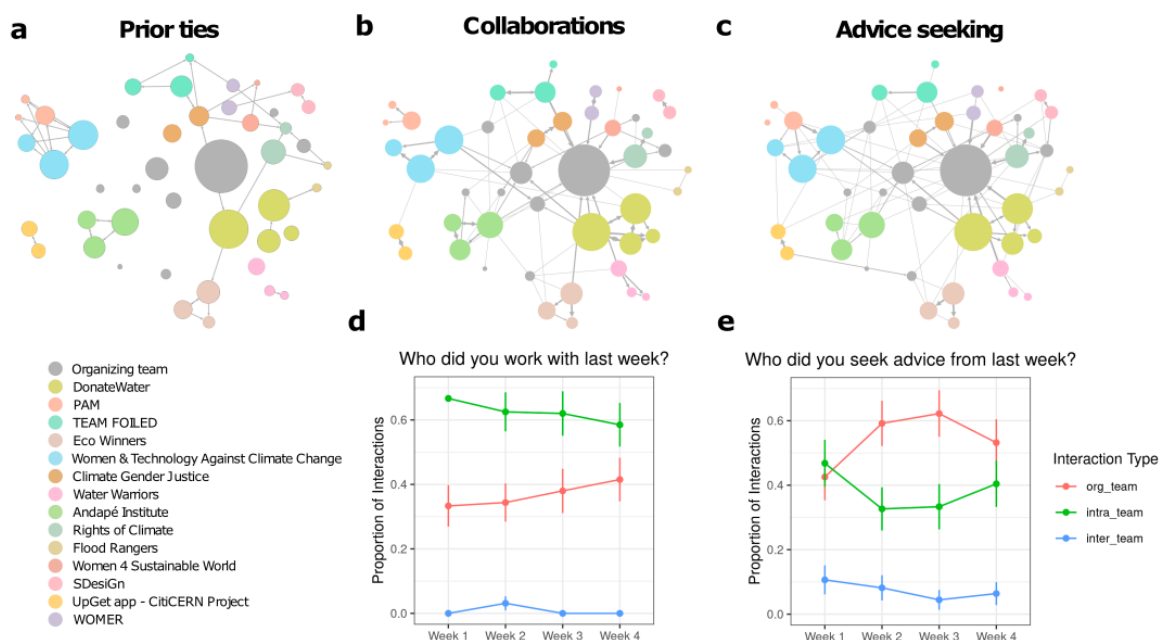


Figure 2.5: **Collaboration activity.** (a–c) Participant interaction networks constructed from self-report data from CoSo, using prompts:(a) “Which of these people did you know personally before?” (b) “Who did you work with last week?” and (c) “Who did you seek advice from last week?” The size of a node is proportional to the total number of interactions of a node across the 3 networks. (d) Proportion of interactions in the collaboration network that involve the organization team (red), or that are intra- (green) or inter- (blue) team, in time. Error bars denote the standard error of the estimate given a number of interactions observed, assuming a binomial statistics. (e) Same than d, for the advice seeking network.

Overall, we find that the collected interaction profiles from digital traces and from surveys highlight different aspects of the social interactions, providing complementary insights to inform community management.

Team Performance

Finally, we analyzed whether features of team composition, communication, teamwork and collaboration were associated with team performance at the Evaluate phase. The performance was measured using various features that can be grouped into two overarching categories: outcome, that is, the evaluation of the project itself; and process, that is, the assessment of engagement within the program (see Methods and Supplemental File 2: Figure S6a). Given the small number of teams from which we can compute an association with performance (14 data points), we use a correlation analysis with a soft significance threshold at $p = 0.1$. We present the results of this analysis in Figure 2.7, where we highlight that the quality of the outcome is generally associated with team profiles and communication

activity from Slack, while the level of engagement in the program (*process*) as judged by the organizing team is associated with self-reported measures of collaborations and activity.

More precisely, for the team composition, we find that team size is associated with the use of at least a tool in the Citizen Science Solution Kit (*crowdsourcing*), suggesting a need for human power to set up a crowdsourcing infrastructure. The diversity of backgrounds in the team is associated with the novelty of the project, supporting findings that interdisciplinarity begets innovative work [Singh et al., 2022]. Prior experience with citizen science is important for the relevance and novelty of the project, indicating the importance of past work in related areas to achieve well-defined, innovative projects in this short time span. Similarly, we find that the average education level in a team is associated with the *novelty*, *feasibility*, and *relevance* of the project.

In the case of communication activity, we find that the overall *Slack activity* (which is very correlated with the number of interactions with the organizing team, see Supplemental File 2: Figure S6b) is associated with the *relevance* of the project, highlighting the role of mentoring for helping teams craft a relevant project. Intra-team interactions from Slack mentions are associated with *relevance*, *novelty*, and *crowdsourcing* aspects of the project, as well as with the quality of *deliverables*. Interestingly, we find similar results when measuring the intra-team collaborations with CoSo surveys, indicating that the digital traces do capture relevant qualitative information about team interactions.

In contrast, we find that team engagement in activities and advice seeking is associated with the quality of the process, as judged by the organizing team, encompassing team commitment, attendance, weekly evaluation, and their ability to produce qualitative deliverables. Beyond engaging in diverse activities in a regular manner, survey engagement was found to be a strong predictor of program engagement. Moreover, we note the importance of the ability of teams to engage in advice seeking from diverse network neighborhoods, as measured by (lower) Burt constraint [Burt, 2004] in the advice-seeking network. These results may indicate that the organizing team, who was responsible for judging these criteria, was particularly sensitive to the ability of teams to engage and collaborate throughout the cycle, information that was not readily available to other experts.

2.1.4 Discussion

Processes of engagement and coordination are fundamental to citizen science projects [Schaefer et al., 2021, Jaeger et al.]. Here, we showcase a framework to measure indicators of participation, contribution, and collaboration during the elaboration of citizen science projects. We show that surveys of social interactions collected at several points in time provide information otherwise invisible from digital traces obtained from a Slack workspace that can be leveraged by practitioners who guide citizen science projects at their early stage of development.

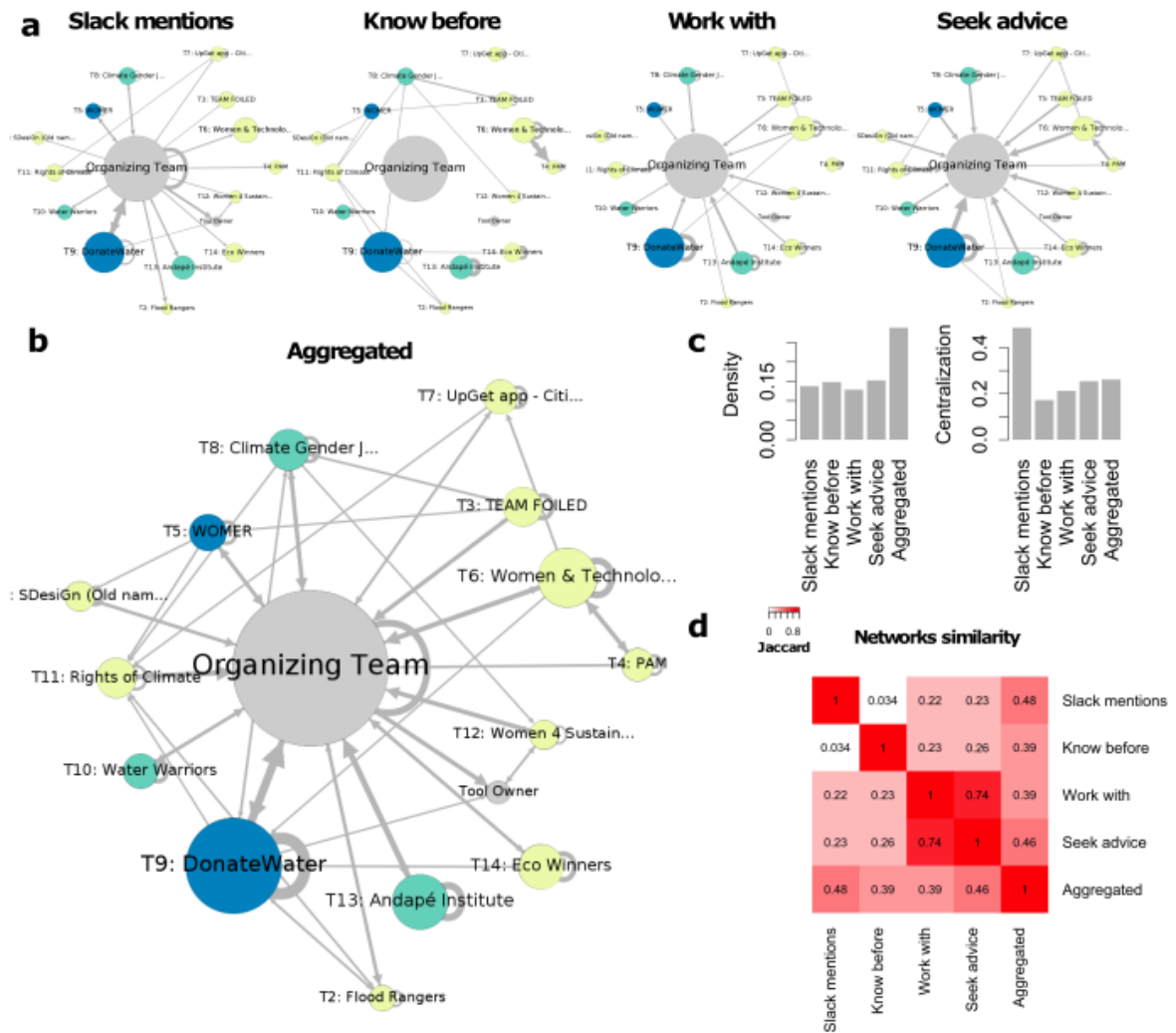


Figure 2.6: **Comparison of the interaction networks.** (a) Team-level networks for the different interaction networks collected. (b) Corresponding aggregate network, where edge weights are the sum over weights across the individual networks shown in a. (c) Network density and centralization (see Methods) across the 4 considered networks. (d) Jaccard similarity between the networks in a. The similarity measures the number of edges shared between any two networks (intersection), divided by the total number of edges present in both networks (union), and ranges from 0 (most dissimilar) to 1 (most similar).

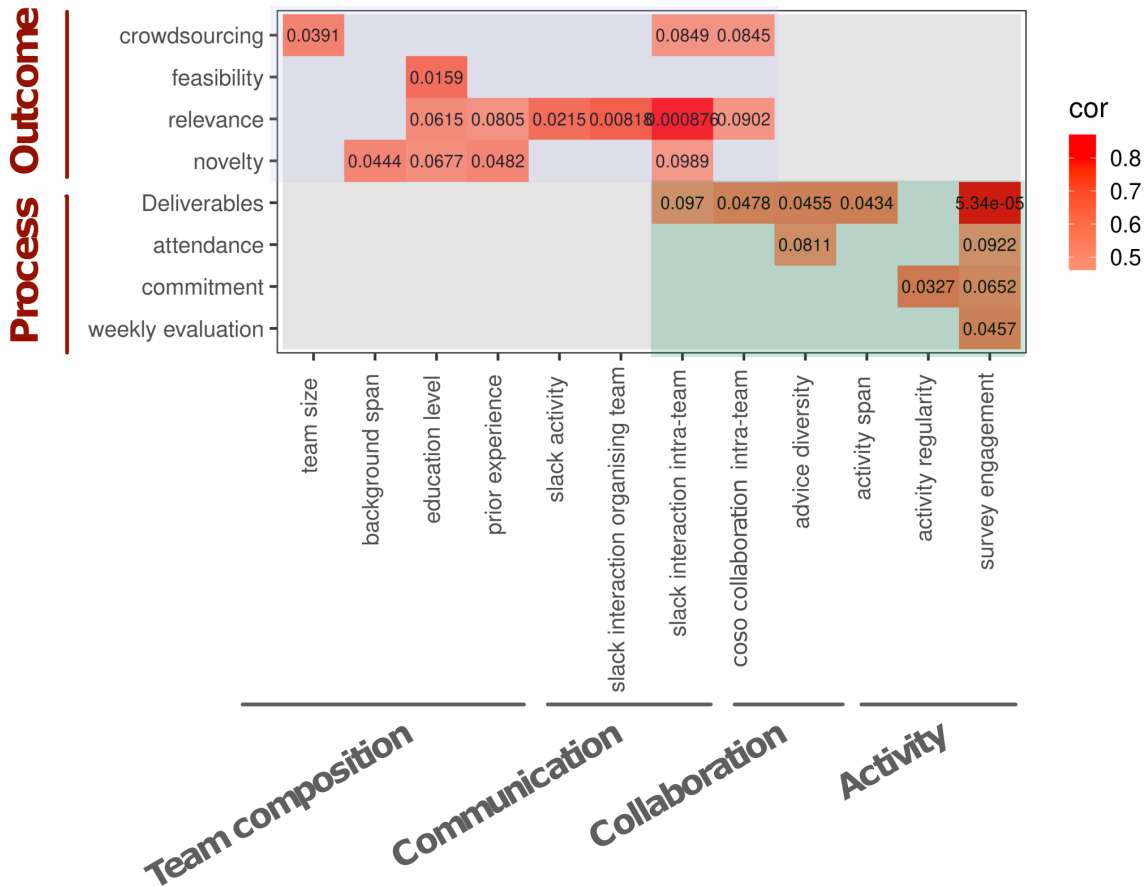


Figure 2.7: **Association with performance.** Correlations between performance assessment (rows) and team features (columns). The correlation value is indicated when the correlation is significant at the $p = 0.1$ level. Shaded areas correspond to sets of features associated with metrics related to outcome (blue) or to process (green).

Given the nature of the program, time could be set aside by the organizers for engaging participants in surveys on a weekly basis, as part of the curriculum. As such, the engagement with the survey instrument was particularly good, allowing to obtain a near-complete coverage. In other contexts where regular meetups with participants would not be feasible, the method could be adapted to incentivize participants to build and analyze their collaboration and stakeholder network and learn from it, for example by providing a dashboard for visual feedback [Tackx et al., 2021].

Our framework is particularly suited to investigate measures related to teamwork. The organizational literature shows that the effectiveness of traditional teams depends on their composition, the collaboration of their members, the task allocation, and the activity level [Hackman, 1987]. Here, we showed that we could monitor proxies for these features, and that they were associated with the ability of teams to produce well-defined deliverables, an indicator of team performance to a standardized task. Beyond small-scale teamwork, the proposed framework can be interesting for quantifying contributions within larger projects. This would allow fine-grain recognition of the different activities achieved, acting both as an incentivization mechanism for monitoring, as well as a reward system for the (usually volunteer) work done.

Beyond team dynamics of early-stage projects, leveraging social networks measurements within citizen science programs offers opportunities to document and understand the build-up of a community around a citizen science project, the engagement patterns of participants, and the contribution to different tasks. This is particularly useful to facilitate the coordination processes of potentially large communities [Kokshagina, 2022, Santolini, 2020], allowing the core team to react and assess whether certain individuals or sub-projects would need help.

Yet this work has limitations. First, the case study could offer only a small sample size, and more data will be needed in further studies to validate the associations with performance. Moreover, during the Evaluate phase, the citizen science projects are at a very early stage of ideation, which did not allow to investigate interactions between teams and citizens. Future work could investigate more mature projects. Finally, the activities performed have different levels of complexity, and do not require the same levels of engagement or collaboration. To highlight these fine-grained aspects would require more qualitative insights from participants, as well as an adaptative strategy to integrate these insights as new (sub)tasks. Beyond the specificity of our project, such an adaptative co-design strategy to account for the varied activities performed is also an important step to be conducted by citizen science projects, as it renders visible the contribution structure and affordances of engagement.

2.1.5 Conclusion

One challenge that organizers of programs like Crowd4SDG face is to support with evidence their decisions related to the formation of citizen science teams and their management, as well as the directions they give to participants to maximize the relevance of the data they generate, their ability to develop innovative solutions, and eventually their impact on the problem they are addressing.

Here, we implemented a monitoring framework leveraging digital traces as well as self-reports to gather compositional and social interaction data during the makeup of citizen science projects. This approach complements traditional outcome-driven metrics in the evaluation of science [Fortunato et al., 2018] by emphasizing the importance of the participation process [Jaeger et al., Schaefer et al., 2021]. We reconstructed a multi-layer social network with interactions of various types, from informal social ties to formal collaborations. We showed that these layers obtained from various means (passive digital traces and active self-reports) cover multiple complementary facets of the interaction dynamics, informing both on interactions with coordinators from the organizing team, as well as intra-team and inter-team interactions. We showed that network centrality measures can be leveraged to quantitatively assess the relative centralization within a given layer, informing on the reliance over a few central nodes. In particular, we found that the ability of a team to manage their social capital by forming interactions across diverse neighborhoods in the network is important for the success of their project, a finding in line with the literature on innovation [Burt, 2004]. Furthermore, we showed that measures of team composition, intra-team collaborations, and communication with the organizers are associated with the quality of the projects, in particular the relevance and novelty of their solutions to the SDGs. Measures of engagement in activities and advice seeking are on their end associated with the elaboration process, in particular the ability of teams to provide timely deliverables.

Overall, we introduced a framework to monitor the evolution of participatory processes in citizen science projects. The obtained interaction networks reveal both formal and informal relational networks that underlie the collective learning and performing, making visible structural patterns that are otherwise invisible to coordinating teams. Network measures of centrality, peripherality, or diversity can then be leveraged to quantify the embeddedness (or lack thereof) of participants in the ecosystem, informing on concrete interventions to improve engagement and project outcomes. Such insights can therefore prove useful to support practitioners in the design and coordination of programs aiming at fostering engagement, inclusion, and diversity in citizen science projects.

Supplementary Information

Tables

Survey	Data collection platform	Respondents (% of total participants)
Evaluate registration Form	Google Form	38 (100%)
Evaluate weekly 1	CoSo	26 (68%)
Evaluate weekly 2	CoSo	26 (68%)
Evaluate weekly 3	CoSo	20 (53%)
Evaluate weekly 4	CoSo	22 (58%)
Evaluate Final Form	Google Form	22 (58%)

Table 2.2: List of surveys of the GEAR cycle 2

Methods

Teaming Algorithm Participants who joined as individuals were assigned to a novel team using the teaming algorithm Edu2Com [Georgara et al., 2024]. Edu2Com is an heuristic algorithm that generates team allocation based on a certain strategy, which were in this case, competence, preference and personality of the participants. The participants were asked to fill in a survey answering questions related to the competencies, skills and personality and a preference survey, where they ranked the pitches of all the selected ideas from 1 to 5, based on how interesting they found the idea. These surveys were needed so that the algorithm could propose possible options for team formations. Eight of the twenty pitches were team pitches, and twelve were individual pitches. A majority favored fourteen of the ideas. The Algorithm proposed six combinations of teams retaining the existing teams and six combinations with a completely new proposal of teams. The team profiling algorithm proposed six alternatives for team formations altering the weightage between competence, personality and preferences. From the six alternatives provided, The final selected team profiling was based on a weightage that had 10% match of their 1 competencies, 20% match of their personalities and 70% of their preferred choices. This particular alternative was chosen since it gave an ideal combination of teaming up individuals as a team along with the pre-formed teams.

Figures



Who are we?

At Crowd4SDG, we are using surveys for two purposes: improving your experience with the programme, and exploring how activity, interactions, and diversity of teams impact their success. Thank you for helping us to do so! You can find more about it in this [information form](#).

To access the survey, please login using your email and select Participants as a team name.

Team

Select your team

Email (the mail email you gave in the registration form)

Enter your email

Submit

How helpful did you find the session on [name of the session] ?

Did not attend & did not watch the Zoom recording

Very unhelpful

Unhelpful

Neither unhelpful nor helpful

Helpful

Very helpful

Who did you seek advice from during the previous block?

Select the persons by name from the list below. You may need to scroll down to see all of them.

Search the participants

Who did you work with since the previous block?

Select collaborators within your team or from other teams. You may need to scroll down to see all of them.

Search the participants

Which task(s) did you perform since the previous block?

- Brainstorming / ideation
- Planning tasks
- Developing protocols / methods
- Reading papers / documents
- Attending meetups with other teams
- Collaborating with other teams
- Interpreting results
- Preparing other documents/materials to share outside the team
- Preparing graphics / videos
- Writing / preparing a pitch
- Project administration
- Software development
- Hardware development
- Data analysis / modeling / visualization
- Meeting with people affected by the problem you are trying to solve
- Meeting with actors of the field of your solution
- Team Building activities
- Documentation of your project
- Use of CS project builder
- Use of CS logger
- Use of Decidim
- Use of Visual Cit
- Use of SDG in progress

Figure 2.8: Screenshots of the CoSo interface

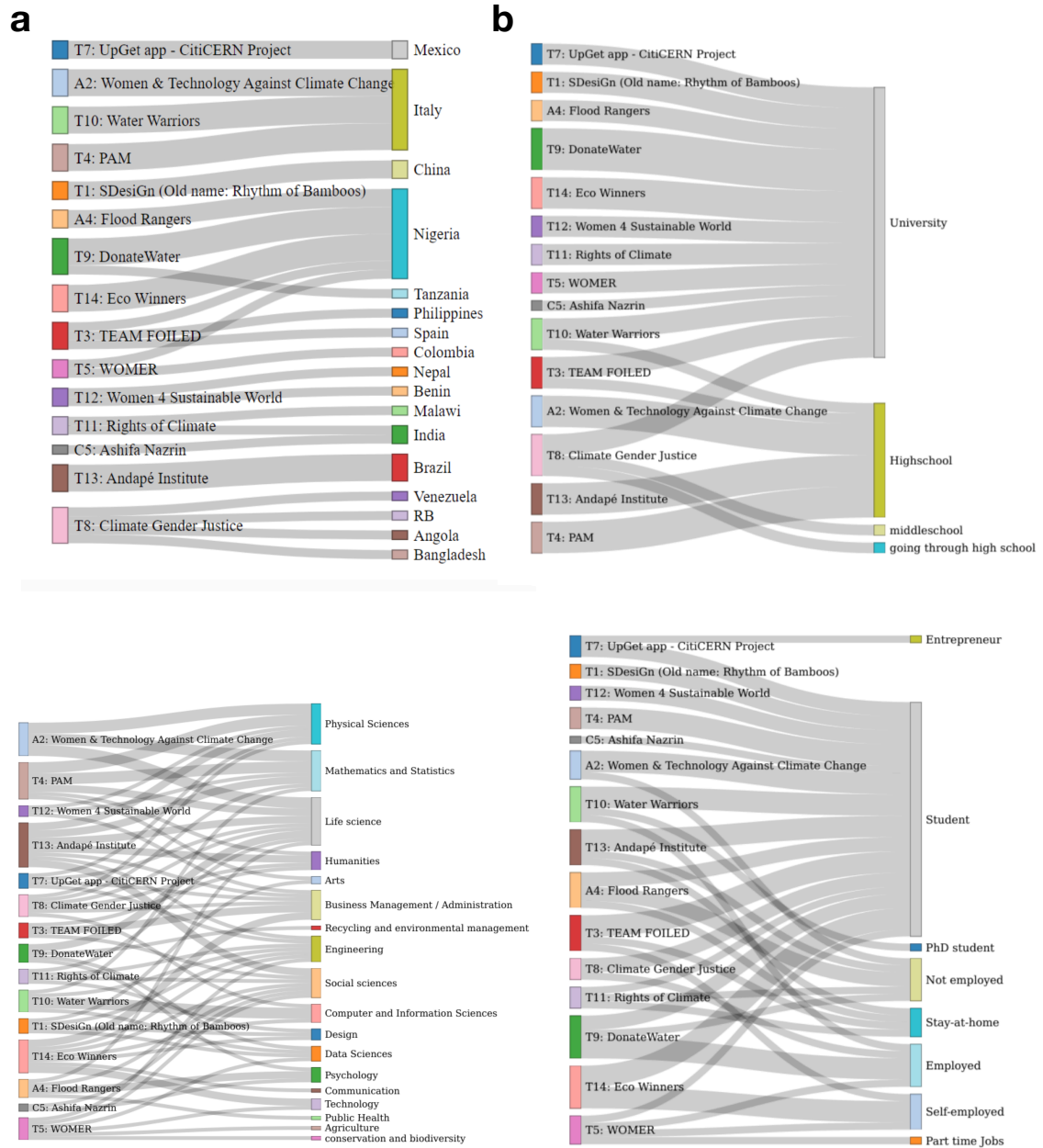


Figure 2.9: Sankey diagrams of teams current or highest level of study (a) and disciplinary backgrounds (b)

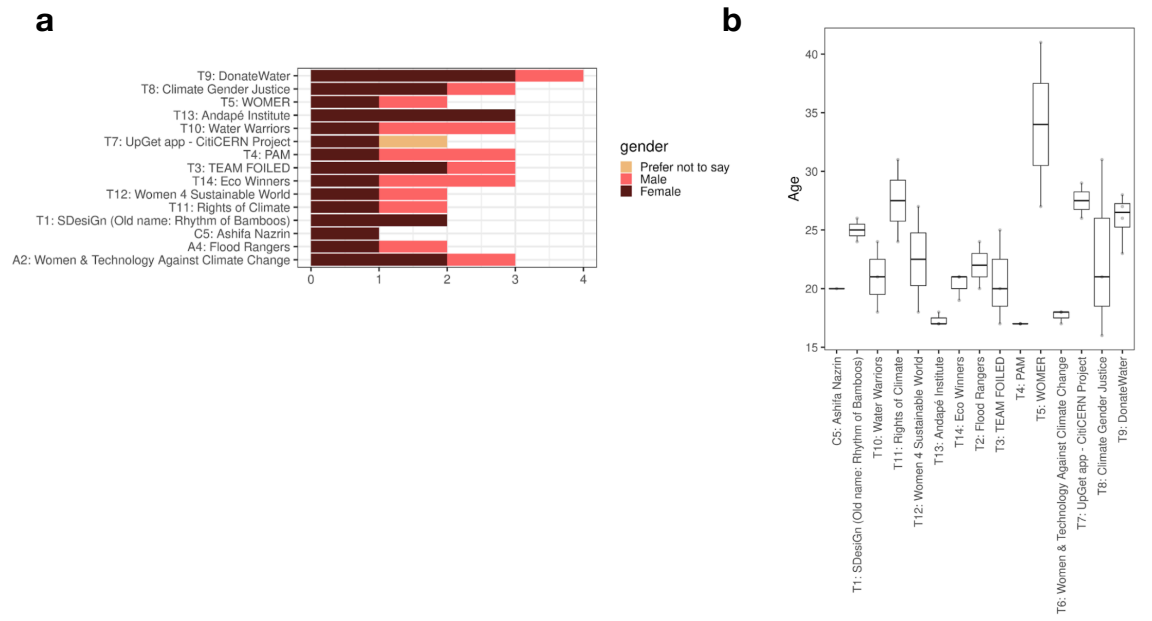


Figure 2.10: Gender (a) and age (b) distributions across teams

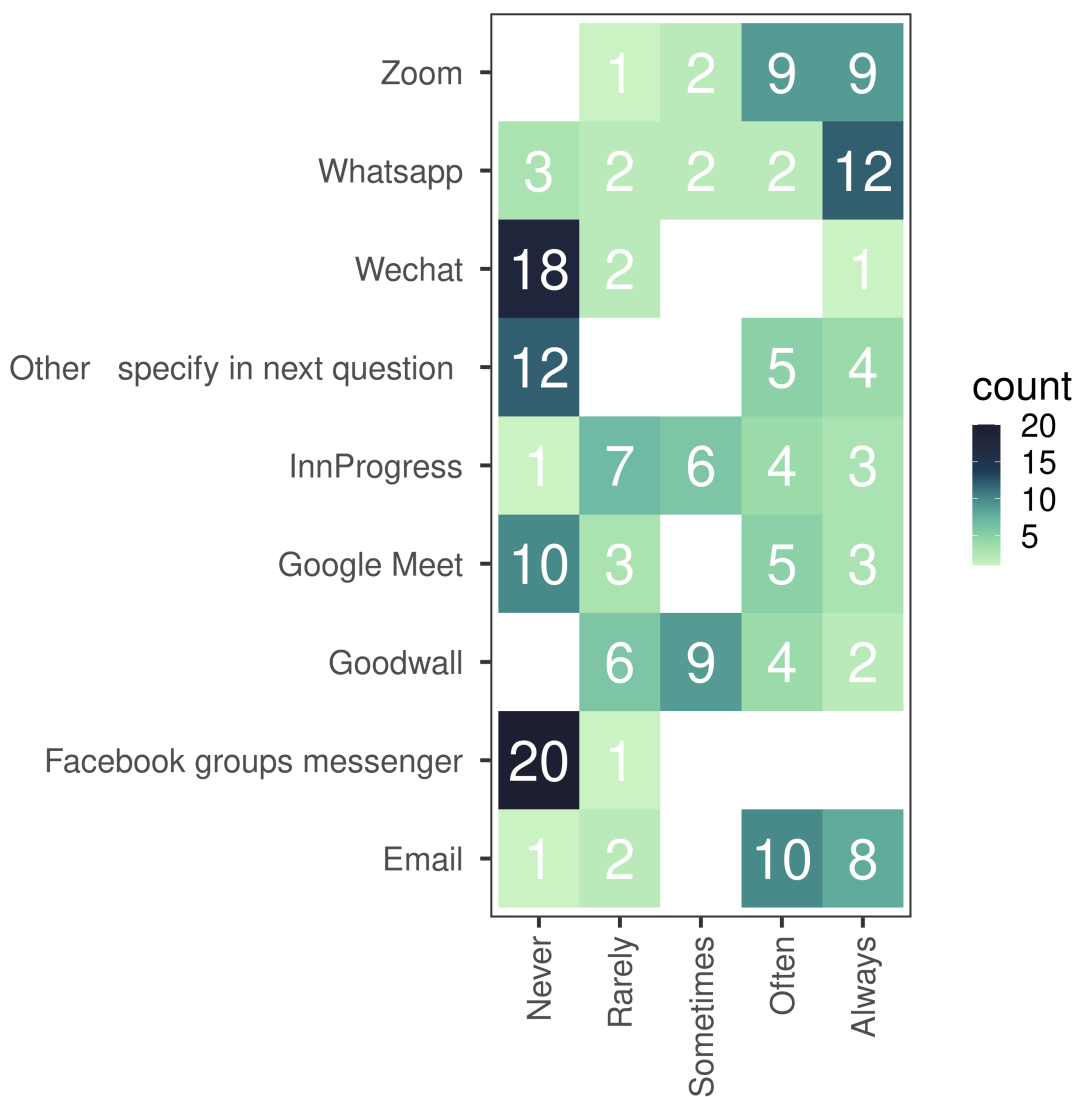


Figure 2.11: Communication tools reported to be used by teams to communicate. Number indicates number of answers across participants (total N=22 participants).

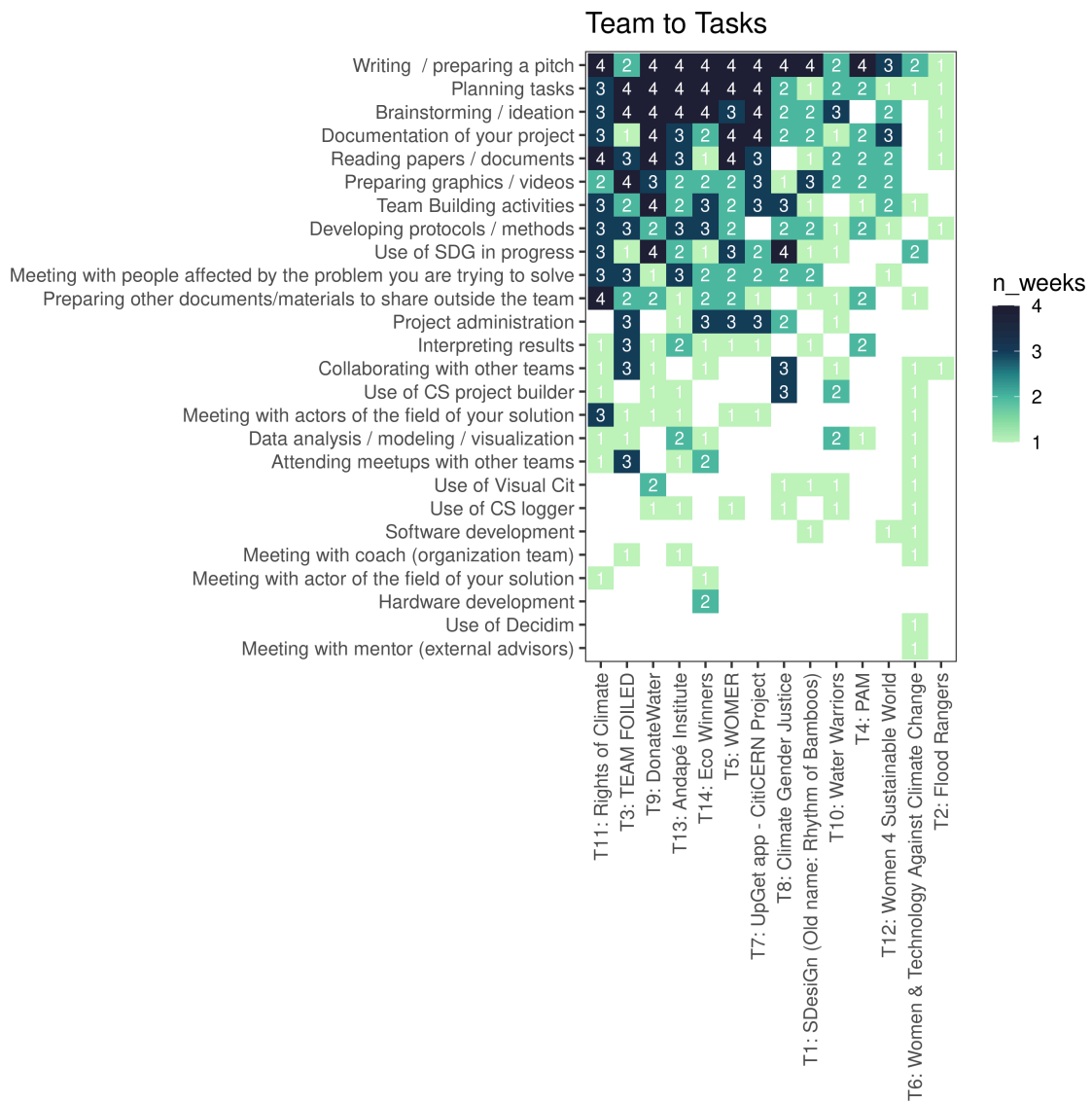


Figure 2.12: Heatmap indicating the number of weeks each activity has been reported by a given team, across 4 weeks. Activities and teams are ranked by row and column sums respectively.

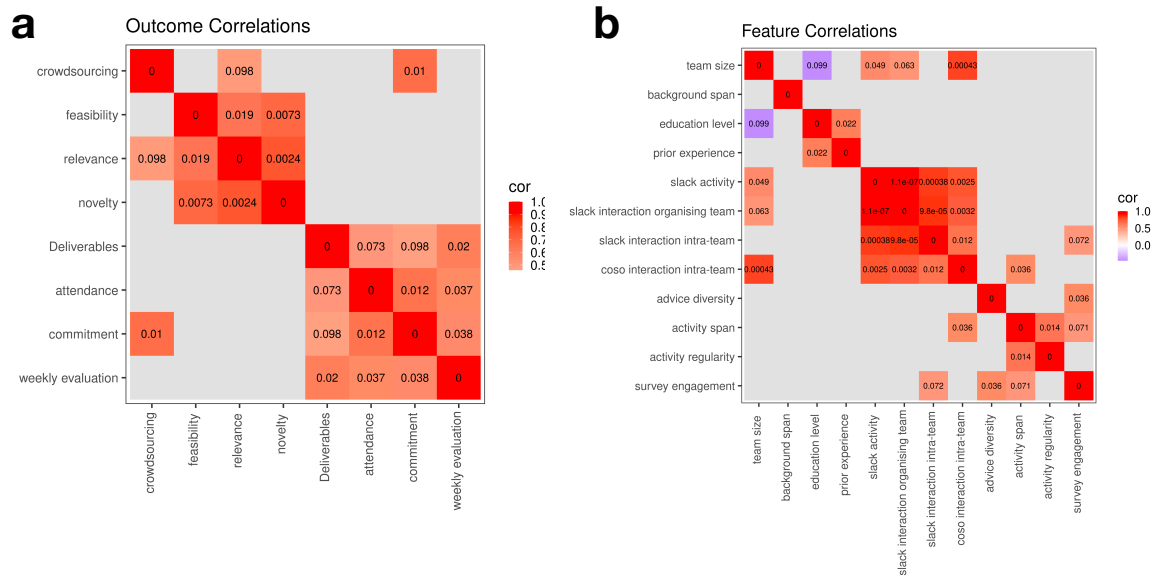


Figure 2.13: Correlation matrices of the features shown in Fig 7, for evaluations (a) and team features (b). Numbers correspond to p-values of the correlations. We grayed out cells with a p-value $p > 0.1$. We find two groups of evaluations: outcomes (top left) and processes (bottom right). For team features, we find that Slack activity is correlated with the intra-team collaborations measured with CoSo, highlighting that digital traces can capture qualitative insights on team work.

2.2 Statistical Model of the Association Between Collaboration Dynamics and Project Performance

Preface

The previous section describes the pipeline to collect and curate digital trace and survey data from teams participating in the Crowd4SDG citizen science consortium. I describe the network science measures used to capture collaboration, diversity and engagement of teams and show their correlation with their evaluation scores in the GEAR cycle.

This section further investigates the association between collaboration dynamics and project performance of citizen science teams in the Crowd4SDG consortium's GEAR 2 and GEAR 3 cycles, using self-reported data collected through surveys in the CoSo platform and digital traces from Slack. The analysis reveals the importance of team engagement, diversity in composition, activity span and effective collaboration strategies in determining project outcomes. The findings also indicate compositional and structural aspects of the Evaluate phase serve as early predictors of the teams' eventual performance. Based on these insights, it is recommended to foster robust team engagement, assemble diverse teams and implement efficient collaboration strategies to enhance the success of future GEAR cycles or similar programs.

Contributions

The contents of this section are derived from a technical report submitted to the European Union as a part of the deliverables for the Crowd4SDG Horizon 2020 project. The report is presented 'as is' (with slight adaptations to maintain coherence with the previous sections of this chapter) and reformatted to fit the style of the thesis. The full original report can be found in the following citation:

SANTOLINI, M., MASSELOT, C. AND **JEYARAM, R.**, 'STATISTICAL MODEL OF THE ASSOCIATION BETWEEN COLLABORATION DYNAMICS AND PROJECT PERFORMANCE', TECHNICAL REPORT, EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME.

I performed the data collection, curation and analysis. Credit to Marc Santolini for conceptualising the study, performing analysis, supervision and writing of the report (along with Camille Masselot).

2.2.1 Introduction

A key deliverable of Crowd4SDG focuses on creating and monitoring new metrics and statistical models of team engagement and collaboration, which contribute to the diverse outcomes of citizen science projects within the Crowd4SDG consortium over its 3-year duration. It has two primary objectives: 1) Develop standardised metrics and descriptors for assessing the diversity, originality, effectiveness, sustainability/robustness, and adaptation/appropriateness of solutions and insights obtained from citizen science projects; and 2) Implement these metrics and descriptors as tools for analysing digital records of citizen science collaborations and their generated solutions and insights. As a result, this supports Crowd4SDG’s specific objectives of enhancing citizen science skills, producing high-quality scientific outcomes, and generating economic and social outputs relevant to achieving SDGs through challenge-based citizen science events, particularly focusing on climate change resilience.

In this report, we present a statistical modelling framework for identifying predictors of performance and impact metrics for citizen science projects. Prior research has identified key characteristics of high-performing teams [Pentland, 2012], such as Team Energy (interaction quantity and frequency), Team Engagement (closing conversation loops, assessed using network clustering), and Team Exploration (seeking external interactions and information). By analysing digital traces from the Slack workspace, demographic data, and self-report surveys collected in GEAR 2 (Santolini 2022) and GEAR 3 (Santolini 2023a), we extract various team organisational features related to these characteristics. We then utilise social network analysis to investigate centrality measures in communication processes and informal advice networks. Ultimately, we evaluate their association with the success and quality of citizen science projects using regression analyses on the performance metrics defined in our initial report on the epistemology of citizen science in (Jaeger 2021).

2.2.2 Methods

Communication Data

A Slack workspace was used by the teams during the GEAR cycle as a means to communicate with other teams and with the organising team. The data was extracted in JSON format using the export function available to the owners/admins of the Slack workspace. This allowed us to gather, across all public channels, a data frame containing the messages (post contents) and information on each message’s timestamp, sender, and target channel. The raw data was then processed to obtain mentions. A mention occurs when a Slack user types in a message the Slack username of a target user prefixed by “@” (e.g. @John). Each recorded mention has information on the source (who wrote the message), target (who is being mentioned) and the timestamp (when the message was sent). Slack also allows users to broadcast messages by citing all users in a channel or a workspace by using specific commands (@all, @here, @channel_name). The messages containing these built-in commands were not included as mentions in order to focus on direct interactions only.

Using the available Slack data, we employed the number of posts and number of reactions of a user as a marker of individual engagement, or team engagement when aggregated over team members. Furthermore, for each GEAR cycle we built social interaction networks where a user is linked to another user if he/she mentions him/her, with a weight corresponding to the number of mentions. When aggregating at the team level, intra-team mentions are encoded as self-loops, and the weights of the intra-team links are summed to create a final team-level network on which to compute centralities such as weighted degree. This allows to represent the flow of information characterising this phase, in particular highlighting the interactions with the organisation team.

CoSo Self-Reported Interaction Data

During GEAR cycles, we conducted two types of surveys: those related to participant attributes (e.g. their background, country of origin, etc), and those related to participant interactions (e.g. who they collaborated with, sought advice from, etc).

The initial survey was related to attributes only and was disseminated using a Google Form at registration to the Evaluate phase. We then disseminated 4 weekly surveys related to social interactions and activities using the CoSo platform [Tackx et al., 2021]. The CoSo platform is designed to collect self-reported interaction data with a simple, reactive interface, and an analysis-ready database (Santolini 2023b). To document their interactions, the users select target users across all other participants and organisers. The interactions span prior ties in the first survey (“Which of these people did you know personally before?”), and on a weekly basis their advice seeking interactions (“Who did you seek advice from last week?”) and work collaborations (“Who did you work with last week?”). To document their activity, they could also select across 26 activities encompassing routine activities within research teams inspired from the CRediT contribution taxonomy [noa, b], as well as specific questions regarding Crowd4SDG, for example specific tool usage. Activities encompassed different levels of complexity in their realisation. They ranged from tasks that could be performed in a distributed fashion such as preparing the final pitch and analysing data, to tasks involving higher levels of collaboration such as brainstorming.

The surveys were advertised through Slack and the organising team dedicated 10 minutes for participants to fill them during weekly sessions, ensuring a high engagement (Santolini 2023a, p12).

CoSo networks were directly inferred from the surveys. For each GEAR, we aggregated the networks over all time points collected, yielding weighted interaction networks where edge weights correspond to the number of times an interaction was reported. When considering team-level network centrality measures, that is, measures that indicate how strategic the position of the team is in the network of interactions, we further aggregated the individual networks at the team level. Network centrality measures were computed using the

igraph library in R [Csárdi and Nepusz, 2006].

Team Characteristics

The ability of teams to develop their project depends on compositional features such as who is in the team, as well as how the team operates, such as their collaboration activity and division of labour. Here we used the digital traces and survey data to derive and monitor features related to team composition, communication, collaboration, and activity which we detail below.

For team composition, we built measures of size, diversity, education level, and prior experience with SDGs. Team size was assessed using the number of members of a team. Background diversity was assessed by computing the background span, that is the number of unique academic backgrounds in the team as declared in the registration form. The education level was computed by taking the average level of education in a team based on the response to the question "What is your current or highest level of education" to which we attributed the following score based on the answer: 0 for secondary school, 1 for high school, 2 for undergraduate and 3 for graduate. Finally, prior experience with SDGs was computed as the average answer to the question "Have you participated in data projects or contributed as citizen scientist to data production before?" (yes = 1 and no = 0) within each team. For communication, we leveraged the activity and interactions on Slack public channels. The Slack activity was assessed as the total number of messages posted by team members. For interactivity, we measure Slack interaction intra-team as the number of mentions among members of a team, and Slack interaction organising team as the number of mentions between members of a team with the organising team. We counted mentions regardless of their directionality.

In studying team collaborations, we looked at both the number of partnerships within teams and the position of these teams in the broader network. The interactions span priorities ("Which of these people did you know personally before?"), their advice seeking interactions ("Who did you seek advice from last week?") and work collaborations ("Who did you work with last week?"). We measured internal (intra-team) interactions by adding up the connections within each team. To understand the team's place in the interaction network, we used a social capital indicator called Burt constraint [Burt, 2004]. The Burt constraint measures how diverse a team's network is, with lower values indicating a more varied network and higher values showing a concentrated network with many connections to the same group. It essentially gauges how connected a team is to other teams that are also connected to its neighbours. A higher constraint means the team has fewer or more similar (redundant) contacts. To assess network diversity, we took the negative of the Burt constraint, with higher values signifying greater diversity (more structural gaps). This helps us quantify a team's ability to access different sources of information for advice or collaborations.

Finally, for the activity, we focused on measures of diversity and engagement of activ-

ities performed, as measured by CoSo (see previous section). For diversity, we computed the activity span as the proportion of activities performed by a team among the 26 listed activities. For engagement, we considered the activity regularity by first computing the number of activities reported by a team each week, and then computing the negative of the Gini coefficient on the resulting vector. The Gini index ranges from 0 (perfectly regular) to 12 (perfectly irregular). The $(1 - \text{Gini})$ value is higher if activities are regularly conducted across weeks. Finally, we quantified for each team the survey engagement as the proportion of survey responses per team across all CoSo surveys, a measure of engagement to the study.

Team Performance Data

To quantify team performance, we used the scores that teams obtained in their assessment by the jury and the Crowd4SDG organising team, which were co-constructed using the results from (Jaeger 2021, pp 32-33).

At the end of each phase, experts composing a jury scored each team from 0 to 5 on the following criteria. We indicate the weight of each score between squared brackets. The sum of these scores constitute the final jury score, with a maximum value of 50.

- Novelty: Is the pitch based on a new idea or concept or using existing concepts in a new context? [10]
- Relevance: Is the solution proposed relevant to the challenge or potentially impactful? [10]
- Feasibility: Is the project implementable with reasonable time and effort from the team? [10]
- Crowdsourcing: Is there an effective crowdsourcing component? [10]
- Overall: How would you rate this team's overall presentation skills during this pitch? [10]

Between the Evaluate and Accelerate phases, additional criteria presented below were used by the organisation team. We indicate the weight of each score between squared brackets, summing to a maximum possible jury score of 40.

- Appropriateness of Methodology [5] (only for GEAR 2)
- Weekly Evaluation [10]
- Use of Toolkit [5]
- Data Collection and NSO [5]

- Commitment [5] (only for GEAR 2)
- Attendance [5]
- Deliverables [5]

The final score accounted for 60% of the jury score and 40% of the organisation team score:

$$\text{Final Score} = \text{jury score} * (60/50) + \text{organisation team score}$$

More precisely, crowdsourcing was assessed using the mean score attributed by judges to the question “Is there an effective crowdsourcing component?” (yes = 1 and no = 0). We measured the feasibility, relevance, and novelty by computing the mean score attributed by the jury on a scale from 0 to 5 to the questions “Feasibility: Is the project implementable with reasonable time and effort from the team?”, “Novelty: Is the pitch based on a new idea or concept or using existing concepts in a new context?”, and “Relevance: Is the solution proposed relevant to the challenge or potentially impactful?”.

All variables were integer values with scores ranging from 0 to 5 for deliverables and attendance, 0 and 1 for commitment. For weekly evaluation, the score was a continuous value ranging from 0 to 10 scoring the overall quality of their weekly pitch sessions. Deliverable score was measured by the total number of deliverables submitted and documented on the platform Innprogress (<https://innprogresstest.unige.ch/>) among the expected ones. Attendance was estimated by the proportion of sessions attended by team members. Commitment was scored 1 if teams were willing to continue their project after the end of the Evaluate phase, or 0 otherwise.

LASSO regression and Statistical Model

Statistical and network analyses were conducted using the R software. We leveraged libraries `glmnet` [Friedman et al., 2010], `MASS` [Venables and Ripley, 2002] and `jtools` [Long, 2024] for statistical modelling, and `igraph` for network centralities. Associations between team characteristics and performance measures were done as follows:

First, since the data originated from two different GEAR cycles, we considered the possible variation in overall values of both team features and performance by normalising the data. To do so, the features were centred (mean of 0) and rescaled (variance of 1) within each GEAR cycle using the `scale()` function in R, and concatenated into an overall dataframe.

Then, each performance variable was defined as a dependent (outcome) variable, and the data frame of team features was used as independent variables. Missing data was handled by imputation using the means of the nonmissing values (by the `makeX()` function of the `glm` package). We then conducted a LASSO (least absolute shrinkage and selection

operator) regression [Tibshirani, 1996] in order to eliminate team features that are not statistically contributing to the outcome, and select only the relevant features. We note that the LASSO regression has the desired characteristic that features that are not significantly contributing to the outcome are eliminated, i.e. their weight in the linear regression is set to be exactly 0, allowing for a strong filtering of weak signals. This differs from other methods, such as Bayesian linear regression, where the weights would be weak but have a non-zero value. Given the low number of data points, the LASSO therefore appeared as a relevant tool for drastically reducing the feature space to a reasonable dimension for downstream analysis. To select the shrinkage parameter (i.e. the strength of feature reduction), we first conducted a 10-fold cross-validation to find the optimal penalty value that minimises the Mean Squared Error of the regression to the outcome (by the `cv.glmnet()` function of the `glm` package). A final model was run for this optimal penalty value on the whole dataset to derive regression coefficients for all team features. Any feature with a coefficient equal to 0 was then discarded. A standard regression (by the `lm()` function) was then run using the remaining features to obtain standardised regression coefficients, 95% confidence intervals and p-values. Features with p-values less or equal to 10% were finally kept for the final figures shown in this report.

Overall, we considered for each outcome the features that i) are selected during a cross-validation step of the LASSO regression and ii) have less than 10% chance to be contributing to the outcome in a randomised setting. This stringent selection process ensures a significant reduction of the noise in the estimator considering the relatively small (N=26) number of data points.

Pseudo-anonymization and ethical approval

The data collection tools and research questions received the ethical approval of the INSERM committee attached to the University of Paris team (IRB00003888), in charge of collecting the data. Participants gave their consent to the collection of data as they registered to the Evaluate phase (see D4.5). Data was pseudo-anonymized by our team before the analysis.

2.2.3 Results

We report the results of the statistical modelling of the association between collaboration dynamics and project performance. Because of the low number of data points (N=26 teams), we leverage a stringent analysis in order to i) combine both GEAR 2 and GEAR 3 (batch correction) and ii) select relevant features (LASSO regression) for regression analysis (see Methods). We consider two main outcomes: the team performance at the Evaluate phase, and the advancement in the GEAR cycle. The former is directly related to the team characteristics measured at the Evaluate phase, while the latter interrogates whether early monitoring at the Evaluate phase informs on the ultimate stage achieved in the GEAR cycle (the Accelerate or Refine stage). In addition, we explore several fine-grain performance

measures that are aggregated to compute the Evaluate performance, such as the novelty, relevance, or feasibility of the projects.

Performance at Evaluate phase

We first focus on the performance at the Evaluate phase, which accounts for 60% of the jury score and 40% of the organisation team score (see Methods). Results of the LASSO feature selection and linear regression method are shown in 2.14. Features are ordered by decreasing significance (i.e. higher p-values), with all features having $p < 0.1$.

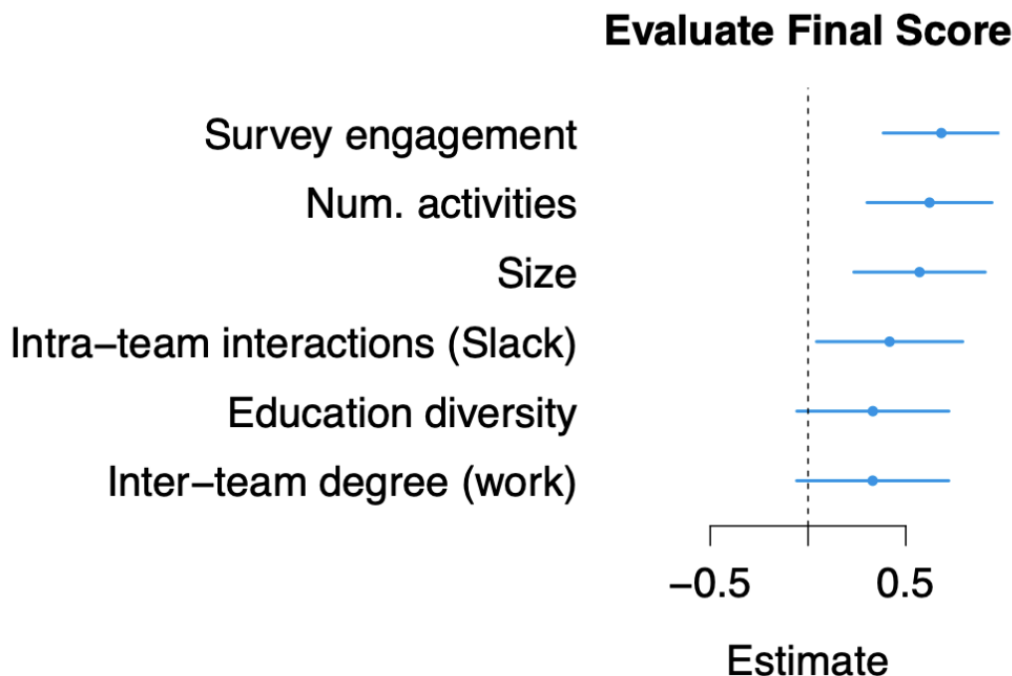


Figure 2.14: Standardised regression coefficients for the team characteristics associated with the Evaluate final score, selected through the LASSO regression (see Methods). Error bars denote 95% confidence intervals. Positive estimates denote a positive association between the feature and the outcome. For network measures, we show in brackets the type of network it is measured from. These consist of Slack network, or CoSo network: priorities (“Which of these people did you know personally before?”), advice seeking (“Who did you seek advice from last week?”) and work collaborations (“Who did you work with last week?”).

Firstly, our analysis reveals that a team’s engagement in the CoSo survey (mean answers per week) is the most significant predictor. This finding suggests that, beyond its data collection function, the engagement in the self-report survey serves as an indicator

of the team's dedication to participating in the program, and that these efforts impact the quality of their project (jury score) and the engagement perceived by the organising team. This is supported by the subsequent feature, the total number of activities performed during the phase, which is positively linked to performance. Activities ranged from tasks that could be performed in a distributed fashion such as preparing the final pitch and analysing data, to tasks involving higher levels of collaboration such as brainstorming. Overall, these two measures demonstrate that engagement in Evaluate activities influences performance at the end of the phase.

We also discover that team composition plays a role in performance, with a positive correlation between the number of team members (size) and the diversity of education levels within the team (education Shannon index). This implies that larger, more diverse teams have a performance advantage.

Additionally, our findings show that a team's position within the interaction network is crucial. Teams that collaborate with a higher number of teams (degree of inter-team collaboration), and have members who communicate more frequently (intra-team Slack interactions) perform better.

In summary, these results indicate that both composition and structural features are important in determining the outcome at the Evaluate phase. However, these are aggregated outcomes, and we will now shift our focus to specific fine-grained outcomes to delve deeper into which features are crucial for their success.

Aspects of project quality

In the Crowd4SDG project, teams have to design and pitch early-stage citizen science projects. As such, these projects must hold certain properties: they have to be relevant for the topic of the GEAR cycle, feasible, innovative, and involve a crowdsourcing component. We used the fine-grained data from the jury scores to compute relevant performance variables and explore team features that underlie them. Results are shown in figure 2.15.

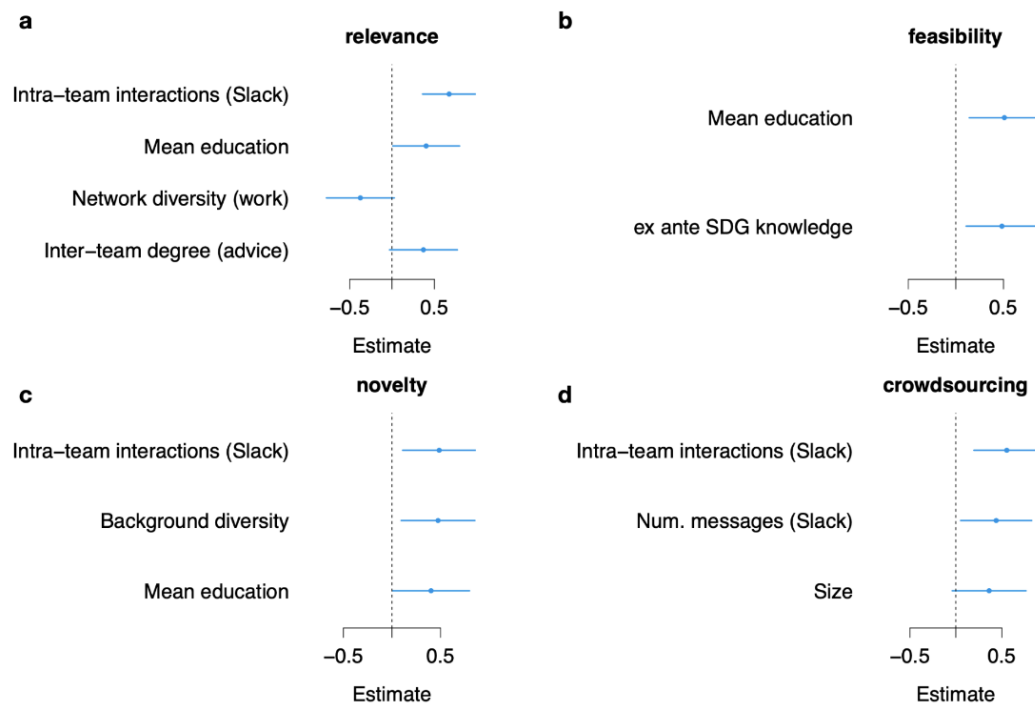


Figure 2.15: Same as figure 2.14, for the outcomes shown in bold.

Various aspects of team composition were found to be important. First, higher education levels (mean education) within teams correlated with more relevant, feasible, and novel projects, emphasising the importance of advanced academic skills for developing innovative yet realistic projects. Second, team size played a role in crafting crowdsourcing components, highlighting the benefits of a larger number of individuals to accomplish this task. Third, the diversity of backgrounds in the team, indicative of interdisciplinarity, was linked to novelty, a finding consistent with scientometrics research showing that interdisciplinarity fosters innovation (Singh 2022). Lastly, the average level of prior experience with SDGs (*ex ante* SDG knowledge) was associated with project feasibility, suggesting that participants draw on their SDG experience (possibly within the Goodwall platform, from which the majority of participants originated) to refine their ideas into viable projects.

Finally, team interactions proved to be crucial for project relevance. Teams that sought advice from a larger number of teams (Inter-team degree) and collaborated within a focused, tight network (low network diversity for “work with”) were more likely to achieve high relevance scores. This reflects a balance between seeking advice (gathering information from the network) and exploiting advice (collaborating with a more limited set of actors).

In summary, these findings demonstrate that team composition features (size, education level, and diversity of backgrounds), internal communication (engagement on Slack), and collaboration strategy (advice seeking and work interactions) are associated with distinct aspects of project quality.

Advancement in the GEAR cycle

Beyond the results from the Evaluate phase, we asked whether the obtained data at the Evaluate phase, which encompasses the largest number of teams (compared with Accelerate or Refine), could be used as an early predictor of the final stage achieved by teams during the GEAR cycle.

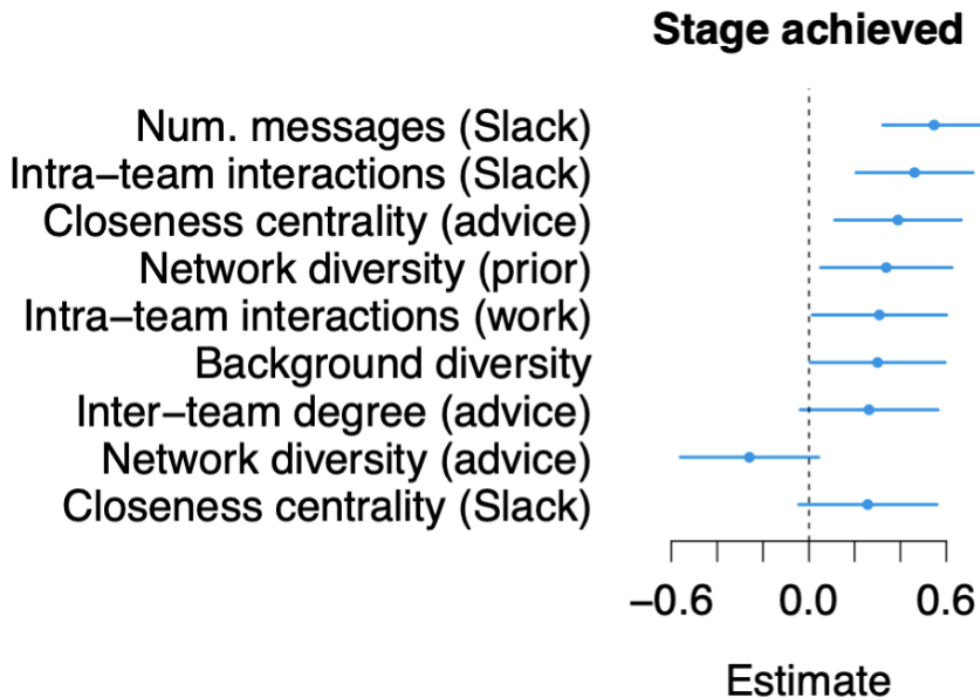


Figure 2.16: Same as figure 2.14, for the final stage achieved in the GEAR cycle. For network measures, we show in brackets the type of network it is measured from. These consist of Slack network, or CoSo network: prior ties (“Which of these people did you know personally before?”), advice seeking (“Who did you seek advice from last week?”) and work collaborations (“Who did you work with last week?”).

Figure 2.16 presents the results of feature selection and regression analysis for the stage achieved. These findings are consistent with previous insights and can be summarised as follows.

First, team activity in the Evaluate phase is associated with the final stage reached in several ways: the number of messages shared on Slack (both within the team and overall) and the number of self-reported work interactions within the team. In essence, hard work plays a significant role in ultimate success.

Second, we find that several diversity measures are associated with success: the di-

versity of backgrounds (background span), which suggests that team interdisciplinarity is essential to address the global challenges at hand, and network diversity of prior ties, indicating a broader reach within the informal network.

Lastly, advice-seeking behaviour is identified as important on multiple levels. In fact, we find that both local (inter-team degree, a measure of the number of immediate neighbours of a node) and global (closeness centrality, a measure of how close a node is to all other nodes in the network through shortest paths) centrality in the advice-seeking network are important, while maintaining strong connections with a focused, tightly knit neighbourhood (low network diversity).

In summary, these results demonstrate that compositional and structural aspects during the Evaluate phase serve as early indicators of the teams' eventual performance in the GEAR cycle.

2.2.4 Discussion of the results

The GEAR cycle analysis provides valuable insights into the factors that contribute to project performance and advancement in the context of citizen science. Specifically, aspects of team composition, internal communication, and collaboration strategy are crucial determinants of success, highlighting the interplay between individual and collective factors.

Engagement in the CoSo survey and the number of activities performed during the Evaluate phase significantly impact team performance, reflecting the importance of commitment and dedication. This finding echoes the social psychological concept of group cohesion, which is known to positively affect group performance [Carron et al., 1985].

Larger teams with diverse education levels and interdisciplinary backgrounds have a performance advantage, consistent with theories that emphasise the benefits of interdisciplinary collaboration for innovation (Singh, 2022). This resonates with research on the benefits of diverse teams in science, which shows that heterogeneous groups can bring different perspectives and expertise to bear on complex problems (Page, 2007).

Internal communication through Slack proves critical for project relevance, novelty, and crowdsourcing components, demonstrating that digital platforms can facilitate effective collaboration, especially during remote work scenarios. This aligns with prior studies examining the role of digital tools in fostering collaborative research networks [Lazer et al., 2009].

The study found that advanced academic skills, team size, and prior experience with SDGs correlate with distinct aspects of project quality. Teams that sought advice from a

larger number of teams and collaborated within a focused network achieved higher relevance scores. This balance between information gathering and effective collaboration aligns with Granovetter's "strength of weak ties" theory [Granovetter, 1973], which posits that weak ties provide access to novel information, while strong ties foster trust and collaboration. This balance is also consistent with Burt's structural hole theory [Burt, 2004], which suggests that individuals who bridge gaps in networks can access diverse information and resources, leading to improved performance and innovation.

In summary, the GEAR cycle analysis offers valuable insights into the interplay between individual and collective factors that contribute to project performance in citizen science initiatives. These findings emphasise the importance of fostering interdisciplinary teams, effective communication, and strategic collaboration, which are supported by existing social theories on networks and collaboration in science. This research provides a foundation for further exploration into the dynamics of collaborative networks in citizen science and the development of strategies to optimise project outcomes.

2.2.5 Conclusion and perspectives

WP4 aims to develop and monitor new metrics and develop statistical models of team engagement and collaboration that contribute to the many-faceted outcomes of the CS projects developed within the Crowd4SDG consortium. In this report, we presented a data-driven approach to develop a statistical model of the association between collaboration dynamics and project performance during GEAR cycles 2 and 3. We leveraged the CoSo platform for collecting self-reported data on collaborations and task allocation structure of participating teams, allowing us to measure characteristics of team composition, activity and interaction dynamics.

In this report, we demonstrated how different data sources on teamwork, effort, communication and collaborations inform on various measures of performance of their project. Given the relatively small number of teams ($N=26$), we leveraged a LASSO regression analysis in order to perform feature selection. We then investigated the association between collaboration dynamics and project performance in the context of the GEAR cycles, focusing on team performance at the Evaluate phase and advancement in the GEAR cycle. Results show that team composition and structural features are equally important in determining the outcome at the Evaluate phase. Key factors include team engagement, activity span, team size, diversity of education levels, and embeddedness in the interaction network. Further analysis of fine-grained outcomes reveals that team composition features (size, education level, and diversity of backgrounds), internal communication (engagement on Slack), and collaboration strategy (advice seeking and work interactions) are associated with different aspects of project quality.

We also examined whether data from the Evaluate phase can serve as an early predictor of the final stage achieved by teams during the GEAR cycle. Findings indicate that com-

positional and structural aspects at the Evaluate phase are indeed early predictors of the eventual performance of teams. Specifically, team activity in the Evaluate phase, diversity measures, and advice-seeking behaviour were found to be important for final success.

Overall, the study highlights the significance of team engagement, composition, and collaboration strategy for project performance in the GEAR cycle. The self-reported and surveyed data offer an opportunity to operationalise metrics and descriptors underlying the quality and novelty of citizen science projects. Our contribution extends beyond the Crowd4SDG project to the general evaluation of CS by informing project leaders, citizen scientists, and decision makers on what can be assessed online to perform high-quality citizen science based on the criteria provided in D4.2 and operationalised in this report.

In light of the findings presented in this report, we put forth the following recommendations to enhance the success of future GEAR cycles or comparable programs. Coordinators should prioritise cultivating robust team engagement, assembling teams with diverse compositions, and implementing efficient collaboration strategies. It is advisable to motivate participants to actively partake in activities and maintain frequent communication via platforms like Slack, which has proven beneficial for idea generation and project refinement. Forming teams with a diverse mix of education levels, backgrounds, and experiences can foster innovation and improve project quality. Additionally, establishing a collaborative atmosphere in which teams can access advice from an extensive network of peers while sustaining strong connections with a select group of collaborators is essential. By emphasising these aspects, coordinators can contribute to a more favourable environment for achieving successful project outcomes in GEAR cycles or similar initiatives.

2.3 Visualisation Dashboards to Monitor Citizen Science Teams

The goal of the previous sections is to identify the key predictors of co-designing an innovative citizen science project within the Crowd4SDG incubator framework. Over the course of the 3 GEAR cycles, we identify the importance of:

- *Engagement*: Response to surveys and activity on Slack are predictors of the final score (in the Evaluate phase) in addition to the stage of progress within the GEAR cycle. Interaction with the organising team is positively correlated with indicators of project quality.
- *Collaboration*: Messages between team members on Slack are significant predictors of their final performance, as well as project quality indicators like novelty and feasibility. Having a local and focused network of work partners in addition to having more diversity for seeking advice (with organisers and other teams) are predictors of higher project relevance.
- *Composition*: Teams with diverse backgrounds and education levels correlate with project quality (such as novelty) and are predictors of their advancement within the GEAR framework

These observations highlight core perspectives that are important to be in consideration while designing team-based innovation incubators. Organisers have to monitor and support teams to foster interdisciplinary collaboration amongst participants, but also maintain team engagement to accelerate their process of learning and project design.

Monitoring teams engagement and collaborative activities also critical to organisers on a day-day basis. The organising team within the GEAR cycles, in addition to structuring the training program, offer technical and logistical support - connecting teams to information, resources and personnel, and also help them troubleshoot their usage of digital tools (Slack, CoSo etc.). With teams composed of members with interdisciplinary backgrounds and based in different geographic regions (figure 2.8a and b), organisers play a vital role within the various stages of the Crowd4SDG GEAR cycles.

An immediate need for the organisers to support their facilitation activities is to have a streamlined method to monitor team activities and engagement. Each survey highlights collaborative activities across different dimensions (such as working together, sought advice), between various entities (teams, organisers, resource owners) while their Slack activity indicates their overall engagement, who teams are reaching out to, in addition to recording their queries. The Crowd4SDG consortium spans 6 institutions across 4 European nations, and having a platform to track responses to surveys, Slack activity and usage of program toolkits aids the organisers to coordinate their facilitation actions.

This gave birth to the need to build a dashboard to visualise team and participant engagement and collaboration. I present the digital resources used to construct the dashboard in the Methods section, showcase snapshots and discuss their utility to the organisers in the forthcoming sections.

2.3.1 Methods

Survey Data

Participants and teams respond to surveys in different stages of the GEAR cycle.

- *Gather*: The first phase collects participants meta data - including their demographic information (age, geographic location), educational background, email IDs and their prior knowledge about SDG related topics and themes. Participants also pitch a citizen science project idea and are selected into the next phase by judges.
- *Evaluate*: This phase groups participants into teams and provides them training across 5 weeks to co-design their project. The first survey in the evaluate phase asks about prior relationships between team members - i.e who in the team they knew before. This is followed by weekly surveys tracking the work and advice seeking relationships between participants. The final survey queries their satisfaction with various training and support aspects of the Evaluate phase.
- *Accelerate*: The surveys mimic the structure of the Evaluate phase, with the initial survey (querying teams' project details), 3 weekly surveys and the final satisfaction survey.
- *Refine*: Teams in the final phase present their project live at the Geneva Trialogue in front of organisers and stakeholders. There are no surveys administered at this stage.

The dashboard presents the replies to the surveys in the Evaluate phase (across GEAR cycles 2 and 3), with this being the stage where participants require maximum support from the organising team. Identifiers of participants are pseudo-anonymised using SHA hashing functions (SHA256 from the *digest* library in R).

Slack Data

Team messages on Slack are captured across different phases of the GEAR cycle. The previous sections detail the construction of collaboration networks with Slack messages (through tagging a workspace member with the "@" symbol, replying to messages in a thread and reacting to messages with an emoticon). The data from Slack is exported in a .json format by the administrators of the workspace. I used an R script to process the messages to construct the following data frames.

- *Users*: Details the identifiers of members on the workspace, including their names, geographic location (timezone), their summary description and their email IDs.
- *Channels*: The Slack workspace consists of different channels. This identifies each channel, its description, members and when they joined.
- *Mentions*: Lists messages where one or many workspace members were mentioned. Includes the source, cited target members, the contents of the message and its timestamp.
- *Reactions*: Lists messages and the members who reacted to that message. Includes the source, members who reacted, message contents, reaction type and the timestamps.
- *Replies*: Tracks message threads in the workspace. Due to a lack of texts in the reply format - they are not presented on the dashboard.

Private messages between workspace members are not included in the Slack export by default. All identifiers of users are pseudo-anonymised using SHA hashing functions (SHA256 from the *digest* library in R).

Correspondence between Surveys and Slack

The hashed email IDs are used to match survey responses to the corresponding user in Slack. Participants are identified across the different surveys through the same email ID (they have to login to the CoSo platform using the same ID to respond to each survey). However, participants do not necessarily use the same ID to register to Slack. Since the size of the cohort is small, around 60 participants and organisers, I manually identified and corrected the mismatches.

Once all users are correctly mapped between the surveys and Slack, it is possible to identify individual teams and their composing members on both sources. In addition, certain Slack channels are specific to a team and these channels are identified and matched manually.

Interaction Networks

Networks from the surveys correspond to each interaction question - "Who do you know personally before?", "Whom did you work with?" and "Who did you seek advice from?". Vertices are the responding participant and an edge indicates the corresponding interaction between them. The networks have the participant meta data and their team as vertex attributes and survey information (timestamp, program week, GEAR stage) as edge attributes.

The networks on Slack (text mentions, emoticon reactions) have vertices matching to each Slack user (matched to their survey identifier) and edge corresponding to the interaction type. The vertex attributes include their slack meta data and their team while edge

attributes include the timestamp and the channel where the text appears.

The networks are saved as *graphml* files using the *igraph* library in R.

Building the Dashboard

The dashboard was built using *flexdashboard* library in R [ade, 2024]. The library helps building interactive dashboards using R Markdown - with options to customise layouts, use html components and construct interactive filtering elements.

Visualisations presented on the dashboard were designed using the *ggplot2* and *plotly* libraries in R. The library *countrycode* is used to map country names to a three letter code - acceptable to plot onto a map on *plotly*. The *igraph* library is used to model networks and their interactive plots are constructed using a custom function built atop the *visNetwork* library in R.

2.3.2 Snapshots of the Dashboard

The final dashboard has four pages, each containing specific interactive visualisation elements.

Page 1: Registration

The registration page (figure 2.17) showcases the background information of participants - their geographic location, their age distribution and their background diversity.

The objective behind the registration page is to set the base of the demographic and educational diversity of the participants in the program.

Page 2: Survey Interactions

The survey interactions page (figure 2.18) showcases survey responses to the interaction questions - "Who do you know personally before?", "Whom did you work with?" and "Who did you seek advice from?".

The networks can be filtered by an individual team and explore the interactions between its members and across other teams in the programme. The networks are structured to have a *forceAtlas* layout - highlighting the core vertices from the periphery. The networks serve as a visual aid to identify individuals and teams' position within the network, showcase local and global clusters.

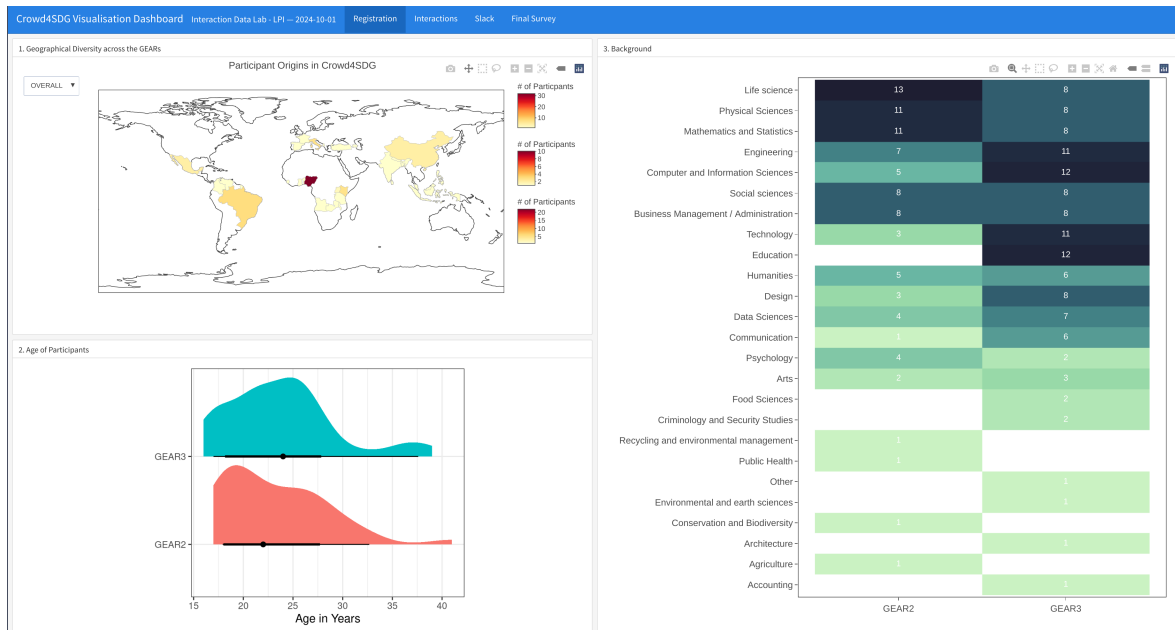


Figure 2.17: Page 1 of the dashboard. The panels on the left shows the nationalities (top, filtered by GEAR cycle) and age distributions (bottom). The panel on the right shows the distribution of academic background as a heatmap.

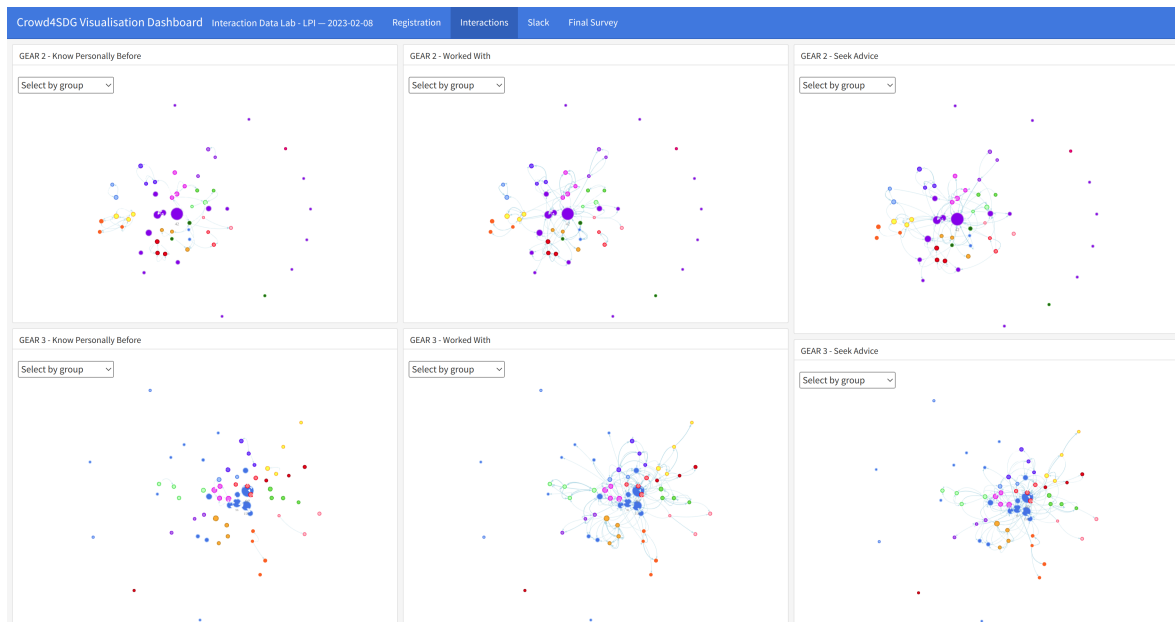


Figure 2.18: Interactions captured from surveys. Panels in each column plot an interactive network for the respective survey question for GEAR cycles 2 and 3. Nodes in the network are each individual participant and the network can be filtered by selecting a specific team - highlighted by a unique color. The node size is determined by its degree. The width of an edge indicates the number of mentions (or reactions) between the same pair of vertices.

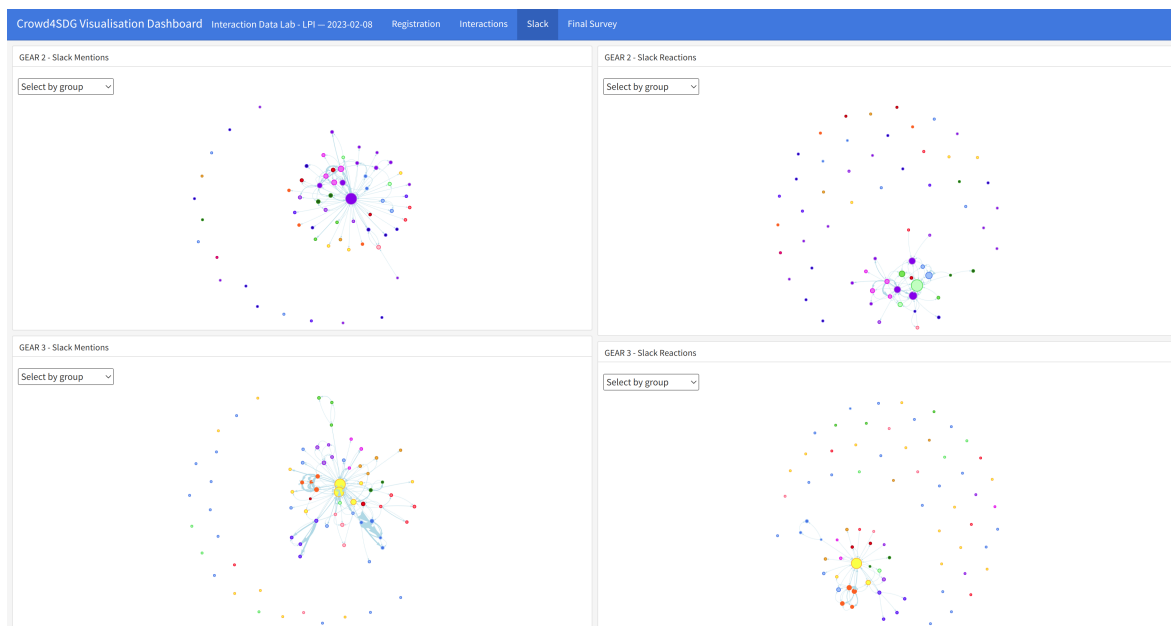


Figure 2.19: Interactions captured from Slack. The panels to the left show the network of mentions and the ones on the right the network of reactions in the GEAR cycles 2 and 3. Vertices correspond to an individual Slack user and the color highlights their team. The vertex size is the degree. The width of an edge indicates the number of mentions (or reactions) between the same pair of vertices.

Page 3: Slack Interactions

Figure 2.19 displays the slack interaction networks (mentions and reactions) of participants in the GEAR cycles 2 and 3. Filtering networks by individuals and teams demonstrate the centrality and connectedness in the interaction network - serving as a visual aid to identify the key predictors of engagement and collaboration.

Page 4: Final Survey

Digital tools form a significant component in the training and facilitation framework in the different stages of Crowd4SDG. The final surveys first capture the level of utility of each resource by the participants. Teams need to be further incentivised to use citizen science toolkits such as Goodwall and in-house tool InnProgress. While Slack, Zoom and Whatsapp are used by participants significantly - specifically among virtual and hybrid teams. The high usage on the Slack workspace further supports its utility as an ecosystem for teams to share information and receive support from administrators.

The feedback on participants on different facets of the program is a visual indicator for organisers to evaluate specific initiatives post-hoc.

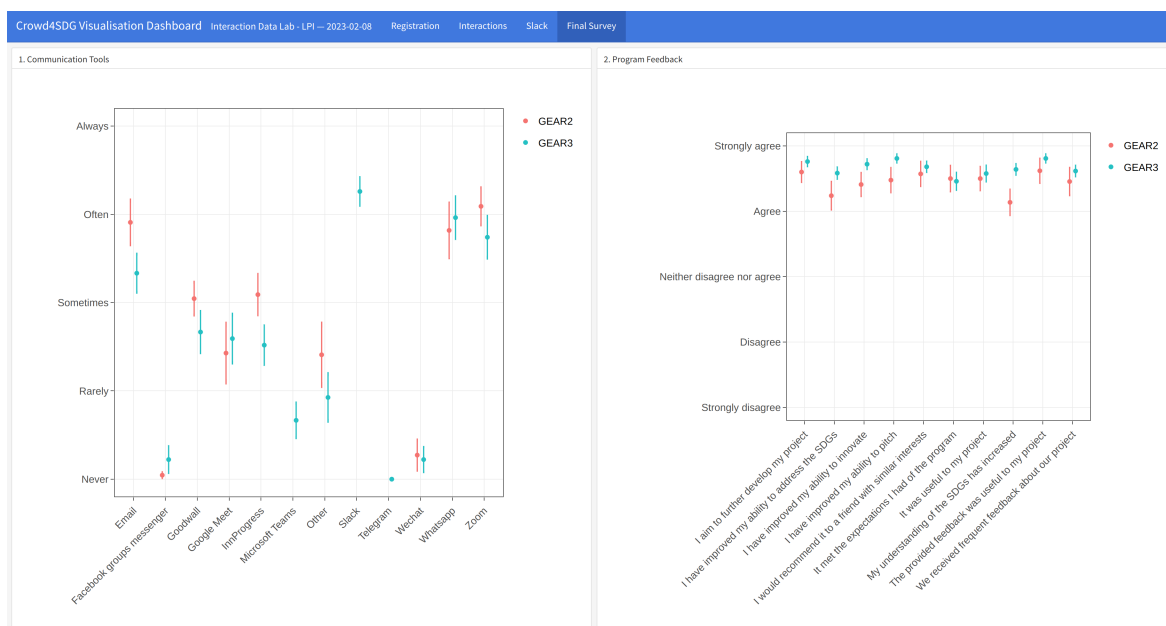


Figure 2.20: Participant responses to the digital platforms they used during the Evaluate phase (left panel) and their feedback on different facets of the incubators' training and support structures (right panel). The dots indicate the average with standard error bars.

2.3.3 Discussion

The three key dimensions that predict novel and innovative citizen science projects within the Crowd4SDG incubator framework are captured through the visualisation dashboard. The *composition* diversity of teams are represented through participants' age, geographical and educational background. The *collaboration* within and between teams is represented through the survey and Slack interaction networks. The interactive nature of the presented visualisations effect the identification of global and local properties like clustering and centrality with a simple overall view. *Engagement* of participants within the framework is demonstrated through interactions between participants and the organisers and the usage of participants and different digital toolkits.

The primary function of the dashboard is to showcase these predictors visually for organisers to identify and provide necessary support. Inactive teams (on Slack or not answering surveys) are reached out to inquire about their specific needs. Peripheral users and teams can be identified after each program week to better understand if they require further collaboration or connect them with relevant resources. Feedback on tool usage and program structure can be used to augment their administration in the following weeks.

Beyond its use case during the program, the dashboard provides feedback into co-creating future challenge-based and citizen science incubator initiatives. From streamlining surveys from GEAR 2 to 3 (by removing redundant questions, re-framing to suit monitoring needs) to aiding the re-design of participant training activities, a platform supporting

monitoring and evaluation strategies is a significant component.

Despite the interactive visualisations, the dashboard has its limitations. Newer data from surveys and Slack are manually added and the Markdown code is re-run. Hosting the dashboard online, with pipelines directly feeding new data makes the presentation dynamic. It is however a challenge to work with Slack data - as typically the export of communication data is performed manually. Building automated solutions using the API from Slack is workaroud. Furthermore, the current dashboard does not include a filter by timestamps (specifically for Slack data), which is primarily due to a lack of sufficient interaction data each week. Incentivising participants in using the workspace to voice their technical and project-related concerns increases the collaboration and engagement patterns of teams.

In summary, the visualisation dashboard provides necessary visual aid for making ad-and post-hoc decisions regarding organisation and facilitation. Building simple dashboards (in this case using R markdown) requires minimal time and cost efforts, further advocating for the usage of structured case by case visuals relevant to the corresponding context.

2.4 Perspectives

Open science and innovation initiatives aim to train participants in collective knowledge production processes and facilitate them towards building solutions and prototypes to address global societal challenges. The Crowd4SDG incubator was designed in this spirit - to support student teams to learn about and build citizen science projects addressing the SDGs. The iterative nature of the project (across the 4 GEAR steps) allows selecting innovative participant teams and helping them fine-tune their project and prototypes. This was mainly facilitated through the monitoring and evaluation strategies that were implemented across the different stages of the incubator.

Administering surveys at regular frequencies, collecting meta data about teams, their interactions within the team and to other participants, and their satisfaction levels allowed to gain a structured and customised view of teams across the programme. This was significantly aided by the administrators who regularly communicated with teams gathered survey responses used an in-house tool. While the latter was used to comply with data management regulations, several open source surveying tools can be alternately used. However, continuous prompting by the organisers is crucial to maintaining participant engagement, even in small cohorts as Crowd4SDG.

Although participants and teams were encouraged to use Slack as way to communicate with other participants, organisers and resource owners within the incubator, they used other platforms such as Whatsapp or email (Fig 2.20 left). Despite this, Slack activity informed on team engagement (with organisers), their collaboration with other participating teams and significantly associated with project quality indicators. Leveraging a mixed method approach integrating both qualitative and quantitative data sources can be crucial to capture varied dimensions of team collaborations in open science and innovation programs.

Network science measures, capturing engagement (survey responses, Slack activity and communication with organisers), collaboration (centrality in work and advice interactions from surveys, communication internal and external to teams on Slack) and diversity (backgrounds, education) are significantly associated with dimensions of project success, quality and their progress within the Crowd4SDG GEAR cycle. This is a key takeaway from this study

The pipeline to collect and process data, presented through the visualisation dashboard is an important aspect of the facilitation process. Organisers gain a rapid insight into participant and team diversity, their interactions and the toolkits that they use through visual aids (and network measures). This allows them to promptly identify potential blocking points and provide personalised support.

Overall, the study offers several perspectives towards designing open innovation and

crowd science incubators and more importantly recommendations for data collection, processing, visualisation and in the use of network science to model and study team diversity, engagement from different facets of team interactions. Despite the effort and time costs in implementing a comprehensive monitoring and evaluation framework, the inferences from this approach are crucial in designing interventions to foster success in future open science and innovation initiatives.

Bibliography

Steffen Fritz, Linda See, Tyler Carlson, Mordechai (Muki) Haklay, Jessie L. Oliver, Dilek Fraisl, Rosy Mondardini, Martin Brocklehurst, Lea A. Shanley, Sven Schade, Uta Wehn, Tommaso Abrate, Janet Anstee, Stephan Arnold, Matthew Billot, Jillian Campbell, Jessica Espey, Margaret Gold, Gerid Hager, Shan He, Libby Hepburn, Angel Hsu, Deborah Long, Joan Masó, Ian McCallum, Maina Muniafu, Inian Moorthy, Michael Obersteiner, Alison J. Parker, Maike Weisspflug, and Sarah West. Citizen science and the United Nations Sustainable Development Goals. *Nature Sustainability*, 2(10):922–930, October 2019. ISSN 2398-9629. doi: 10.1038/s41893-019-0390-3. URL <https://www.nature.com/articles/s41893-019-0390-3>. Publisher: Nature Publishing Group.

Chiara Franzoni, Marion Poetz, and Henry Sauermann. Crowds, citizens, and science: a multi-dimensional framework and agenda for future research. *Industry and Innovation*, 29(2):251–284, February 2022. ISSN 1366-2716. doi: 10.1080/13662716.2021.1976627. URL <https://doi.org/10.1080/13662716.2021.1976627>. Publisher: Routledge _eprint: <https://doi.org/10.1080/13662716.2021.1976627>.

Laura Ballerini and Sylvia I. Bergh. Using citizen science data to monitor the Sustainable Development Goals: a bottom-up analysis. *Sustainability Science*, 16(6):1945–1962, November 2021. ISSN 1862-4057. doi: 10.1007/s11625-021-01001-1. URL <https://doi.org/10.1007/s11625-021-01001-1>.

Dilek Fraisl, Jillian Campbell, Linda See, Uta Wehn, Jessica Wardlaw, Margaret Gold, Inian Moorthy, Rosa Arias, Jaume Piera, Jessica L. Oliver, Joan Masó, Marianne Penker, and Steffen Fritz. Mapping citizen science contributions to the UN sustainable development goals. *Sustainability Science*, 15(6):1735–1751, November 2020. ISSN 1862-4057. doi: 10.1007/s11625-020-00833-7. URL <https://doi.org/10.1007/s11625-020-00833-7>.

Olga Kokshagina. Open Covid-19: Organizing an extreme crowdsourcing campaign to tackle grand challenges. *R&D Management*, 52(2):206–219, 2022. ISSN 1467-9310. doi: 10.1111/radm.12470. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/radm.12470>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/radm.12470>.

Camille Masselot, Bastian Greshake Tzovaras, Chris L. B. Graham, Gary Finnegan, Rathin Jeyaram, Isabelle Vitali, Thomas Landrain, and Marc Santolini. Implementing the Co-Immune Open Innovation Program to Address Vaccination Hesitancy and Access to

- Vaccines: Retrospective Study. *Journal of Participatory Medicine*, 14(1):e32125, January 2022. doi: 10.2196/32125. URL <https://jopm.jmir.org/2022/1/e32125>. Company: Journal of Participatory Medicine Distributor: Journal of Participatory Medicine Institution: Journal of Participatory Medicine Label: Journal of Participatory Medicine Publisher: JMIR Publications Inc., Toronto, Canada.
- Chiara Franzoni and Henry Sauermann. Crowd science: The organization of scientific research in open collaborative projects. *Research Policy*, 43(1):1–20, February 2014. ISSN 0048-7333. doi: 10.1016/j.respol.2013.07.005. URL <https://www.sciencedirect.com/science/article/pii/S0048733313001212>.
- Muki Haklay. Participatory citizen science. In Muki Haklay, Susanne Hecker, Anne Bowser, Zen Makuch, Johannes Vogel, and Aletta Bonn, editors, *Citizen Science, Innovation in Open Science, Society and Policy*, pages 52–62. UCL Press, 2018. ISBN 978-1-78735-235-3. URL <https://www.jstor.org/stable/j.ctv550cf2.11>.
- Enric Senabre Hidalgo, Josep Perelló, Frank Becker, Isabelle Bonhoure, Martine Legris, and Anna Cigarini. Participation and Co-creation in Citizen Science. In Katrin Vohland, Anne Land-Zandstra, Luigi Ceccaroni, Rob Lemmens, Josep Perelló, Marisa Ponti, Roeland Samson, and Katherin Wagenknecht, editors, *The Science of Citizen Science*, pages 199–218. Springer International Publishing, Cham, 2021. ISBN 978-3-030-58278-4. doi: 10.1007/978-3-030-58278-4_11. URL https://doi.org/10.1007/978-3-030-58278-4_11.
- Stephen Parkinson, Sasha Marie Woods, James Sprinks, and Luigi Ceccaroni. A Practical Approach to Assessing the Impact of Citizen Science towards the Sustainable Development Goals. *Sustainability*, 14(8):4676, January 2022. ISSN 2071-1050. doi: 10.3390/su14084676. URL <https://www.mdpi.com/2071-1050/14/8/4676>. Number: 8 Publisher: Multidisciplinary Digital Publishing Institute.
- Josephine M. Chambers, Carina Wyborn, Melanie E. Ryan, Robin S. Reid, Maraja Riechers, Anca Serban, Nathan J. Bennett, Christopher Cvitanovic, María E. Fernández-Giménez, Kathleen A. Galvin, Bruce E. Goldstein, Nicole L. Klenk, Maria Tengö, Ruth Brennan, Jessica J. Cockburn, Rosemary Hill, Claudia Munera, Jeanne L. Nel, Henrik Österblom, Angela T. Bednarek, Elena M. Bennett, Amos Brandeis, Lakshmi Charli-Joseph, Paul Chatterton, K. Curran, Pongchai Dumrongrojwatthana, América Paz Durán, Salamatu J. Fada, Jean-David Gerber, Jonathan M. H. Green, Angela M. Guerrero, Tobias Haller, Andra-Ioana Horcea-Milcu, Beria Leimona, Jasper Montana, Renee Rondeau, Marja Spierenburg, Patrick Steyaert, Julie G. Zaehring, Rebecca Gruby, Jon Hutton, and Tomas Pickering. Six modes of co-production for sustainability. *Nature Sustainability*, 4(11):983–996, November 2021. ISSN 2398-9629. doi: 10.1038/s41893-021-00755-x. URL <https://www.nature.com/articles/s41893-021-00755-x>. Publisher: Nature Publishing Group.
- J. Cox, E. Oh, B. Simmons, C. Lintott, K. Masters, A. Greenhill, G. Graham, and K. Holmes. Defining and Measuring Success in Online Citizen Science: A Case Study of Zooniverse Projects. *Computing in Science and Engineering*, 17(4),

2015. ISSN 1521-9615. URL <https://ora.ox.ac.uk/objects/uuid:5885ed3d-a5ea-4208-bbd2-48f9b12a2575>. Publisher: IEEE.
- Citizen Science for Monitoring Climate Impacts and Achieving Climate Resilience | CROWD4SDG Project | Results | H2020, a. URL <https://cordis.europa.eu/project/id/872944/results>.
- Johannes Jaeger, Camille Masselot, Bastian Greshake Tzovaras, Enric Senabre Hidalgo, Mordechai (Muki) Haklay, and Marc Santolini. An epistemology for democratic citizen science. *Royal Society Open Science*, 10(11):231100. ISSN 2054-5703. doi: 10.1098/rsos.231100. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10646465/>.
- Teresa Schaefer, Barbara Kieslinger, Miriam Brandt, and Vanessa van den Bogaert. Evaluation in Citizen Science: The Art of Tracing a Moving Target. In Katrin Vohland, Anne Land-Zandstra, Luigi Ceccaroni, Rob Lemmens, Josep Perelló, Marisa Ponti, Roeland Samson, and Katherin Wagenknecht, editors, *The Science of Citizen Science*, pages 495–514. Springer International Publishing, Cham, 2021. ISBN 978-3-030-58278-4. doi: 10.1007/978-3-030-58278-4_25. URL https://doi.org/10.1007/978-3-030-58278-4_25.
- Oleksandra Poquet, Liubov Tupikina, and Marc Santolini. Are forum networks social networks? a methodological perspective: 10th International Conference on Learning Analytics and Knowledge: Shaping the Future of the Field, LAK 2020. *LAK 2020 Conference Proceedings - Celebrating 10 years of LAK*, pages 366–375, March 2020. doi: 10.1145/3375462.3375531. URL <http://www.scopus.com/inward/record.url?scp=85082391489&partnerID=8YFLogxK>. Publisher: Association for Computing Machinery.
- Floriana Gargiulo, Maria Castaldo, Tommaso Venturini, and Paolo Frasca. Distribution of labor, productivity and innovation in collaborative science. *Applied Network Science*, 7(1):1–15, December 2022. ISSN 2364-8228. doi: 10.1007/s41109-022-00456-0. URL <https://appliednetsci.springeropen.com/articles/10.1007/s41109-022-00456-0>. Number: 1 Publisher: SpringerOpen.
- Maximilian Klein, Thomas Maillart, and John Chuang. The Virtuous Circle of Wikipedia: Recursive Measures of Collaboration Structures. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, CSCW '15*, pages 1106–1115, New York, NY, USA, February 2015. Association for Computing Machinery. ISBN 978-1-4503-2922-4. doi: 10.1145/2675133.2675286. URL <https://doi.org/10.1145/2675133.2675286>.
- Michael Klug and James P. Bagrow. Understanding the group dynamics and success of teams. *Royal Society Open Science*, 3(4):160007, April 2016. ISSN 2054-5703. doi: 10.1098/rsos.160007. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4852640/>.
- Sebastian Deri, Jeremie Rappaz, Luca Maria Aiello, and Daniele Quercia. Coloring in the Links: Capturing Social Ties as They are Perceived. *Proc. ACM Hum.-Comput. Interact.*,

2(CSCW):43:1–43:18, November 2018. doi: 10.1145/3274312. URL <https://doi.org/10.1145/3274312>.

Raphael Tackx, Leo Blondel, and Marc Santolini. Quantified Us: A group-in-the-loop approach to team network reconstruction. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers, UbiComp/ISWC '21 Adjunct*, pages 502–507, New York, NY, USA, September 2021. Association for Computing Machinery. ISBN 978-1-4503-8461-2. doi: 10.1145/3460418.3479363. URL <https://dl.acm.org/doi/10.1145/3460418.3479363>.

Ronald S. Burt. Structural Holes and Good Ideas. *American Journal of Sociology*, 110(2):349–399, 2004. URL <http://www.jstor.org/stable/10.1086/421787>.

Chakresh Kumar Singh, Emma Barme, Robert Ward, Liubov Tupikina, and Marc Santolini. Quantifying the rise and fall of scientific fields. *PLOS ONE*, 17(6):e0270131, June 2022. ISSN 1932-6203. doi: 10.1371/journal.pone.0270131. URL <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0270131>. Publisher: Public Library of Science.

J. R Hackman. The design of work teams. In J. Lorsch, editor, *Handbook of organizational behavior*. Prentice-Hall, Englewood Cliffs, NJ, 1987. Backup Publisher: Prentice-Hall.

Marc Santolini. Covid-19: the rise of a global collective intelligence?, April 2020. URL <http://theconversation.com/covid-19-the-rise-of-a-global-collective-intelligence-135738>.

Santo Fortunato, Carl T. Bergstrom, Katy Börner, James A. Evans, Dirk Helbing, Staša Milojević, Alexander M. Petersen, Filippo Radicchi, Roberta Sinatra, Brian Uzzi, Alessandro Vespignani, Ludo Waltman, Dashun Wang, and Albert-László Barabási. Science of science. *Science*, 359(6379):eaa0185, March 2018. doi: 10.1126/science.aa0185. URL <https://www.science.org/doi/10.1126/science.aa0185>. Publisher: American Association for the Advancement of Science.

Athina Georgara, Marc Santolini, Olga Kokshagina, Camila Justine Jacinta Haux, Desmé Jacobs, Gloria Biwott, Marcela Correa, Carles Sierra, Jose Luis Fernandez-Marquez, and Juan A. Rodriguez-Aguilar. Optimising Team Dynamics: The Role of AI in Enhancing Challenge-Based Learning Participation Experience and Outcomes. September 2024. URL <https://hal.science/hal-04691209>.

Alex "Sandy" Pentland. The New Science of Building Great Teams. *Harvard Business Review*, April 2012. ISSN 0017-8012. URL <https://hbr.org/2012/04/the-new-science-of-building-great-teams>. Section: Collaboration and teams.

CRedit, b. URL <https://credit.niso.org/>.

Gábor Csárdi and T. Nepusz. The igraph software package for complex network research. 2006. URL <https://www.semanticscholar.org/paper/>

The-igraph-software-package-for-complex-network-Cs%
C3%A1rdi-Nepusz/1d2744b83519657f5f2610698a8ddd177ced4f5c.

Jerome Friedman, Trevor Hastie, and Rob Tibshirani. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(1):1–22, 2010. ISSN 1548-7660.

W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Statistics and Computing. Springer, New York, NY, 2002. ISBN 978-1-4419-3008-8 978-0-387-21706-2. doi: 10.1007/978-0-387-21706-2. URL <http://link.springer.com/10.1007/978-0-387-21706-2>.

Jacob A. Long. jtools: Analysis and Presentation of Social Scientific Data, August 2024. URL <https://cran.r-project.org/web/packages/jtools/index.html>.

Robert Tibshirani. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996. ISSN 0035-9246. URL <https://www.jstor.org/stable/2346178>. Publisher: [Royal Statistical Society, Oxford University Press].

A. V. Carron, W. N. Widmeyer, and L. R. Brawley. The development of an instrument to assess cohesion in sport teams: The Group Environment Questionnaire. *Journal of Sport Psychology*, 7(3):244–266, 1985. ISSN 0163-433X. Place: US Publisher: Human Kinetics.

David Lazer, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne. Computational Social Science. *Science*, 323(5915):721–723, February 2009. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1167742. URL <https://www.science.org/doi/10.1126/science.1167742>.

Mark S. Granovetter. The Strength of Weak Ties. *American Journal of Sociology*, 78(6): 1360–1380, 1973. ISSN 0002-9602. URL <https://www.jstor.org/stable/2776392>. Publisher: The University of Chicago Press.

flexdashboard: R Markdown Format for Flexible Dashboards, 2024. URL <https://pkgs.rstudio.com/flexdashboard/>.

Chapter 3

Network science and computational methods for processing collaboration networks from digital lab notebooks

Preface

In the previous chapter, I presented methods to monitor and evaluate citizen science teams in a collaborative open-innovation incubator ecosystem using structured surveys and traces from digital tools like Slack. The monitoring strategies leveraged has in part, a lot to owe to the methodology design of the Crowd4SDG incubator. Each stage of the innovation cycle was carefully facilitated by organisers, allowing the administering of surveys and tracking participants on the Slack workspace. Teams in various open-innovation settings, however do not have this level of fine-grained monitoring efforts. Recovering network-based measures to represent collaborative structures in teams requires high human and resource costs and is a challenge to scale up with increasing team size and participant cohorts.

However, a significant component of team-based science and innovation projects is documentation. Researchers present their work by writing a scientific article, integrating information with their collaborators. Publications include self-reported attributions [Larivière et al., 2020, Sauermann and Haeussler, 2017] and acknowledgement statements to discern collaborations between authors and the tasks they performed. Digital traces from how members collaboratively edit documents or code can help identify specific contributor roles and patterns in open-innovation settings, such as in Forums [Poquet et al., 2019, 2023], Wikipedia [Klein et al., 2015] and GitHub [Palazzi et al., 2019]. A combination of self-reports and digital traces can offer a more comprehensive view of collaborative structures in teams - but the heterogeneity in reporting styles across different fields and contexts make it a challenging endeavour.

In this chapter, I focus on teams participating in the iGEM science and engineering competition and show how their self-reports on an online wiki page can be leveraged to understand their a) collaborative editing structures, b) how they organise tasks amongst the team members and c) how they collaborate with other teams. The chapter is divided into three broad sections. The first section details the iGEM context and dataset and describes the methods to clean and process digital traces from members co-editing the wiki while the second section leverages the contents of the wiki to build a pipeline to recover inter- and intra- team collaborations from unstructured text. I conclude by offering perspectives for studying teams in open science and innovation contexts, using their self-reported data.

3.1 Team Collaboration Dynamics from Digital Traces

Abstract

This section presents the International Genetically Engineered Machines (iGEM) synthetic biology competition as a resource to study team collaboration dynamics in science and innovation. Teams self-report their project and document their progress on an online wiki website, which reveals organisational structures in over 2200 teams across 11 years. Coupled with meta data of team composition, output and performance, this section aims to showcase this fine-grained and longitudinal setting and discuss network science and computational methods leveraged to identify team collaboration and organisational structures in iGEM.

Contributions

The contents of this section are derived from the following pre-print and data resource.

SANTOLINI, M., BLONDEL, L., PALMER, M. J., WARD, R. N., **JEYARAM, R.**, BRINK, K. R., KRISHNA, A. AND BARABÁSI, A.-L. (2023), 'IGEM: A MODEL SYSTEM FOR TEAM SCIENCE AND INNOVATION', ARXIV

BLONDEL, L., **JEYARAM, R.**, KRISHNA, A., AND SANTOLINI, M. (2024), 'IGEM: A MODEL SYSTEM FOR TEAM SCIENCE AND INNOVATION' [DATA SET]. ZENODO.
[HTTPS://DOI.ORG/10.5281/ZENODO.11072818](https://doi.org/10.5281/zenodo.11072818)

3.1.1 The iGEM Synthetic Biology Competition

International Genetically Engineered Machines (iGEM) is a yearly science and engineering competition that incentivises student teams to build projects integrating synthetic biology with real-life challenges. At its heart, iGEM is a challenge-based learning initiative - student teams are composed of interdisciplinary backgrounds, are expected to build synthetic biology parts (also known as biobricks), providing safety and ethical considerations in addition to undertaking public outreach activities [Stemerding, 2015]. Over the course of the iGEM competition cycle, participating students learn technical and experimental skills and gain valuable experience in project co-creation and management.

A typical iGEM competition cycle is described in figure 3.1. Teams are formed and begin funding acquisition and ideation early in the year. Members carry out essential tasks - like performing experiments and outreach over the summer. Teams present their project in the final jamboree (usually conducted in November), and prepare their slides, posters and their wikis in the weeks before. This overall structure of an iGEM project mimics the processes of academic research. To produce a publication, the Principal Investigator (PI) needs to secure funding, identify technical profiles who can support performing the different research tasks and finally communicate their findings in the form of a manuscript and/or technical talks. PIs of iGEM teams, usually senior researchers, take the lead on sourcing funds, assembling the team and providing supervision. The fundraising process is key for iGEM - registration and jamboree participation costs are significant. Team assembly is guided by iGEM and project requirements - teams usually are composed of experimental biologists, modellers, computer scientists, engineers among others - with this distribution varying across different competition tracks (figure 3.1a). PIs supervise teams and assist with project conceptualisation, but student members usually take responsibility in performing experiments, processing and analysing data and building biobricks. iGEM projects are heavily linked with real-world problems, with participating teams undertaking significant public engagement initiatives to craft their problem and solution. This makes iGEM an incubator framework - which serves as a microcosm for science and innovation activities.

iGEM also emphasises on collaboration and open science philosophies. Biobricks designed by teams are available on a public repository for reuse by other teams and teams are motivated to share resources and provide technical help to their peers as a part of the competition outcomes. These key dimensions of iGEM presents a unique framework to study teams partaking in collaborative science and innovation. Despite participating teams being a student-led, albeit with support from PIs, they produce high quality output. iGEM projects have led to over 110 publications and the founding of 150 startups [Jainarayanan et al., 2021] in the 20 years since its inception. This makes iGEM an interesting testbed to study team collaborative structures in a more modern science and innovation setting.

However, studying collaborative structures suffer from the complexities in recovering granular and longitudinal data from teams. iGEM as a science and innovation setting papers over this gap - with the competition requiring teams to maintain a comprehensive

digital lab notebook (henceforth referred as the *Wiki*) reporting their project, team and attributions. The digital traces from the wiki co-edit structure paired with its contents offer a comprehensive, zoomed in version of how teams organise and perform tasks in science.

This section focuses on the former - using computational and network science approaches to represent the co-edit structure from the wiki digital traces. The next parts describe the particulars of data collection, the methods used to clean and process the digital traces and provide perspectives for extending to other contexts.

3.1.2 Data

The iGEM data presented in this and the following chapters include teams participating between 2008-2018.

Team and Participant Metadata

Teams register to iGEM at the beginning of each competition year. During the registration, teams provide information such as their name, affiliation, roster and are assigned a unique identifier each year. The registration is later validated, and based on teams providing complete information and their successful payment of fees - they are designated as *Accepted* teams and allowed to participate in iGEM. A total of 2296 teams participated in iGEM between 2008 and 2018

The team roster specifies the roles of each composing member. They can be a Principal Investigator (PI - primary or secondary), advisor, instructor, student leader or a student member. Each team member has a unique iGEM identifier, which is consistent across years. Teams are usually affiliated to the department of an academic institution, university, high school or community laboratories. This institutional information is recorded during registration, along with their section (undergrad, overgrad or high school) and their geographic location (figure 3.1b).

The teams' project information is also crucial component of the registration. Teams participate in one of many technical competition tracks, and are expected to declare this along with a brief summary of their project. The competition tracks evolve each year, with 12 unique tracks in 2018 (figure 3.1c). Teams detail specific information about their projects on their wiki pages.

The team and participant meta data are extracted from the public iGEM web page (<https://teams.igem.org/list>).

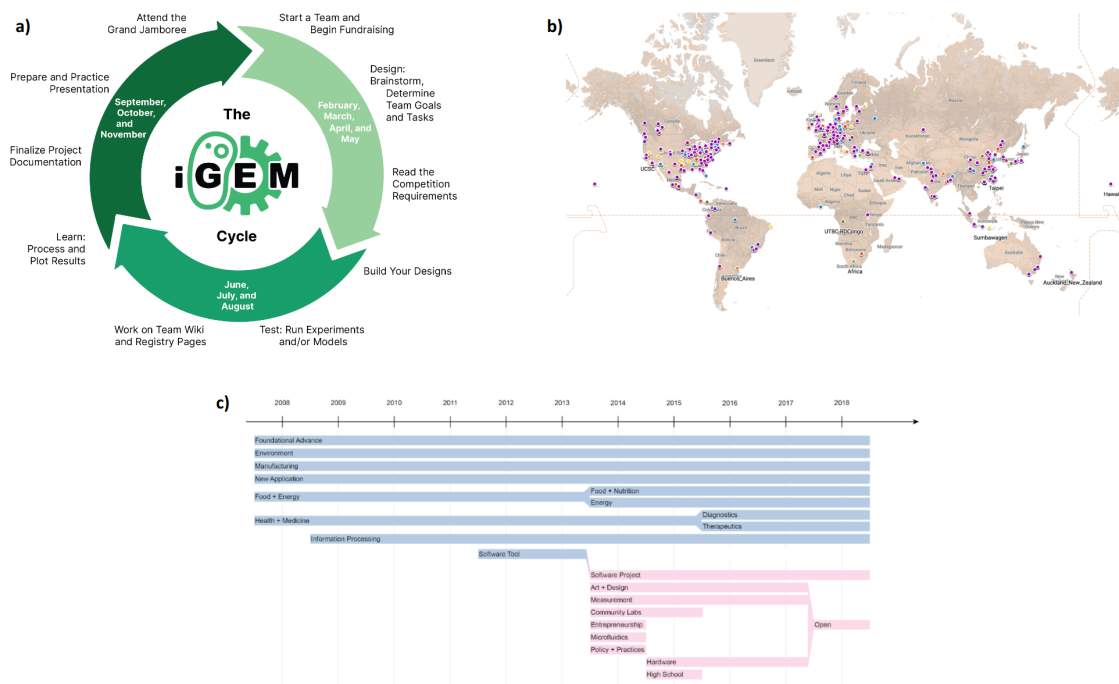


Figure 3.1: a) Competition Cycle in iGEM. b) Geographic Dispersion of iGEM Teams 2008-2018. Each dot corresponds to a team, and the color of the dot represents the last year the team participated in the competition. c) iGEM competition tracks and their evolution. All figures taken from [Santolini et al., 2023]

Team Wikis

Teams maintain an online wiki website, collaboratively edited by its members - where they document the multitude of aspects of their iGEM project and associated activities. These typically include:

- **Team:** Descriptions of the team roster, roles and their attributions - who performed which tasks in the team.
- **Project:** Broad and specific description of the team project - including information such as experimental design, models, technical results and a detailed lab notebook of day-day activities.
- **Parts:** Specifications of the the synthetic biology parts, or Biobricks. Includes designs of new parts, and details of composite parts.
- **Human Practices:** Outreach, education and public engagement activities that teams undertook during the course of their iGEM project. Collaborations between teams are also reported.
- **Safety:** Details about the safety and ethical considerations of the teams' project, in accordance to the safety forms defined by iGEM.
- **Other:** Other miscellaneous information linked to how teams are evaluated - such as digital materials and supporting documents.

Team wikis are created at the beginning of each competition year by the iGEM headquarters. They include template pages for each of the above categories which contain their descriptions and brief instructions on what to include. Teams use this as a start to design their wiki. Despite a consistent overall structure of the wiki template - teams can choose to edit heterogeneously, as long as they report the necessary team and project information.

The team wikis are publicly accessible, through links provided on the iGEM website (<https://teams.igem.org/list>). The full text of the team wiki is scraped by first identifying all the pages in a team wiki and then extracting the complete text content from each of these pages.

The digital traces of how team members co-edit their wikis were available through the iGEM MediaWiki API - which was opened for access specifically to collect this data between 2008 and 2018. The digital traces detail the IDs of the contributing team member, the IDs of the wiki page and section, the size (in bytes) and the nature (create, edit or delete) of each action [Santolini et al., 2023].

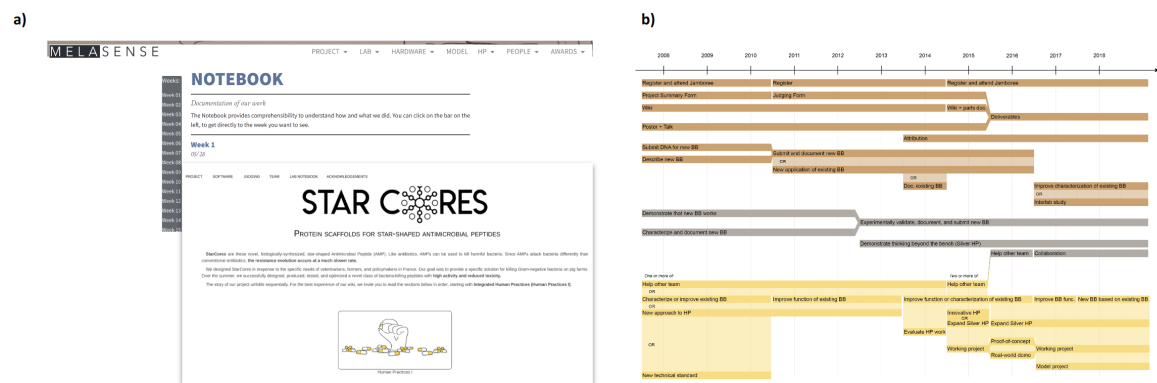


Figure 3.2: a) Snapshots from Team wikis. b) Criteria for Medals and Awards over the years [Santolini et al., 2023].

Medals and Awards

The performance of teams is assessed through the medals they achieve and the awards they win during each iGEM competition cycle.

Teams can win one of Gold, Silver, Bronze or no medals. This is determined by teams accomplishing specific targets on the iGEM project evaluation checklist (figure 3.2b). The medals are represented in an ordinal scale (0, 1, 2, 3) - for no medal, bronze, silver and gold respectively.

Awards winners are determined through their exceptional performance during the final jamboree - through their project quality and their presentation to the jury. Teams compete against other teams in the same section - undergrad, undergrad or high school. There are two broad award classes. Track specific awards are granted to teams with exceptional projects in a given track. Based on jury scores, top teams are nominated as a finalist for the track, of which, winners are selected. Teams are also awarded other special technical awards - such as the best projects in education, entrepreneurship, human practices, model, new part, composite part, presentation, wiki, sustainability and inclusivity. Furthermore, teams are selected to be the best overall prize winner in each section.

Biobricks

Information about the synthetic biology parts (biobricks) that teams design as a part of their iGEM project is also documented on the iGEM website (<https://parts.igem.org/Catalog>), and collected by parsing its contents. This includes the unique identifier of each biobrick, the author who logged its information on the website (their member ID), the team, year and whether the part is a new or a composite. The entire collection of parts made by teams, including their technical details such as sequence data, are reported. We identify 33,687

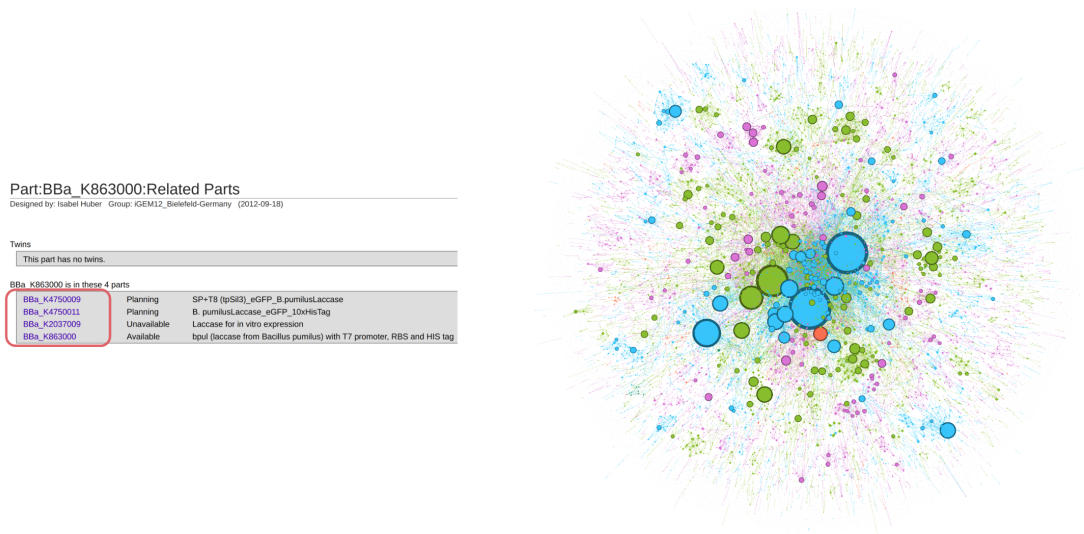


Figure 3.3: Example of the Wiki page showing biobrick reuse (left) and the network of biobrick reuse from 2008-2018, with colors indicating geographic regions of the teams that designed the biobrick (right) There are a core set of biobricks, that teams often reuse - indicated by vertex size and their central position in the network.

biobricks designed by iGEM teams between 2008 and 2018 .

Composite parts are engineered by integrating one or more pre-existing parts along with additional synthetic biology elements. Teams are required to build composite parts as a part of their iGEM project to incentivize reuse of previously designed biobricks. We record 13,170 composite parts with 11,478 reusing atleast one part designed by an iGEM team. The specific information about the composition of these parts are also logged. For these composite parts, we extract the identifiers of the biobricks that are used in its construction - creating a network of *biobrick reuse*, reporting 24,899 interactions between biobricks designed by iGEM teams.

3.1.3 Methods

Scraping of iGEM Team and Wiki Pages

The scraping of websites from iGEM was done using a custom HTML parser, written in Python [Blondel et al., 2024]. iGEM has several public websites with fixed structures - which report the team roster, project details, team performance (medals, awards), biobricks and team wikis. The key data from each of these pages is extracted and stored into .csv files.

The full text of each team wiki is extracted using *BoilerPy3*, based on the *Boilerpipe* li-

brary in python. The library allows several text *Extractors* - which are interfaces to recover text from HTML content. The *ArticleExtractor* is typically used to identify sentences from articles, but it fails to completely recover all text content from the wiki pages. We use the *KeepEverythingExtractor*, which recovers all text content from a wiki page - including text from buttons, menu bars and footers.

Once the text content of each page is extracted, they are saved into separate text files. The code for the extractor was written on python by Leo Blondel [Blondel et al., 2024], and executed for team wikis from 2008-2018 after making minor changes. These reflect changes on the iGEM website and URLs.

The scripts for accessing the MediaWiki API were also written in python by Leo Blondel [Santolini et al., 2023].

Identifying Teams across Participation

Teams are assigned a unique ID each year. This makes identifying teams across years a slightly challenging task. A majority of teams preserve their name across multiple participation. However, teams may change their names each year - despite them being led by the same PIs and hosted in the same institution. Institutional information is documented as free text, which makes it hard to compare consistently across years. Team names may change slightly, like team *EPF-Lausanne* in 2012 to *EPF_Lausanne* in 2013, but may also change significantly, like team *Lanzhou* in 2017 to team *LZU-CHINA* in 2018. This makes using a consistent sub string matching cutoff to match names a complex challenge. We hence use the unique member IDs and their inter-team mobility across years as a proxy to identify name changes.

To match teams across years - we use the following identifying strategy. Using the unique member IDs, we track the teams that a member is a part of across different years in iGEM. If at least 2 members move from one team to the same future team, and the teams have different names, we record the instance. This gives a directed network where the source vertex is an iGEM team and the target vertex is a future iGEM team, but with a different name.

We use this directed network to assign a new *matching ID*, consistent across team name changes. Teams are sorted by decreasing year, and for each team, its past versions are identified using the directed network of team name changes. If the current team name is present in the network, we identify all other vertices (teams) that are connected to the current team - indicating prior name changes. All iGEM teams with any of these names are now assigned a common matching ID. If a team is not present in the network, then all prior teams that only have the same name as the current team are given the common matching ID.

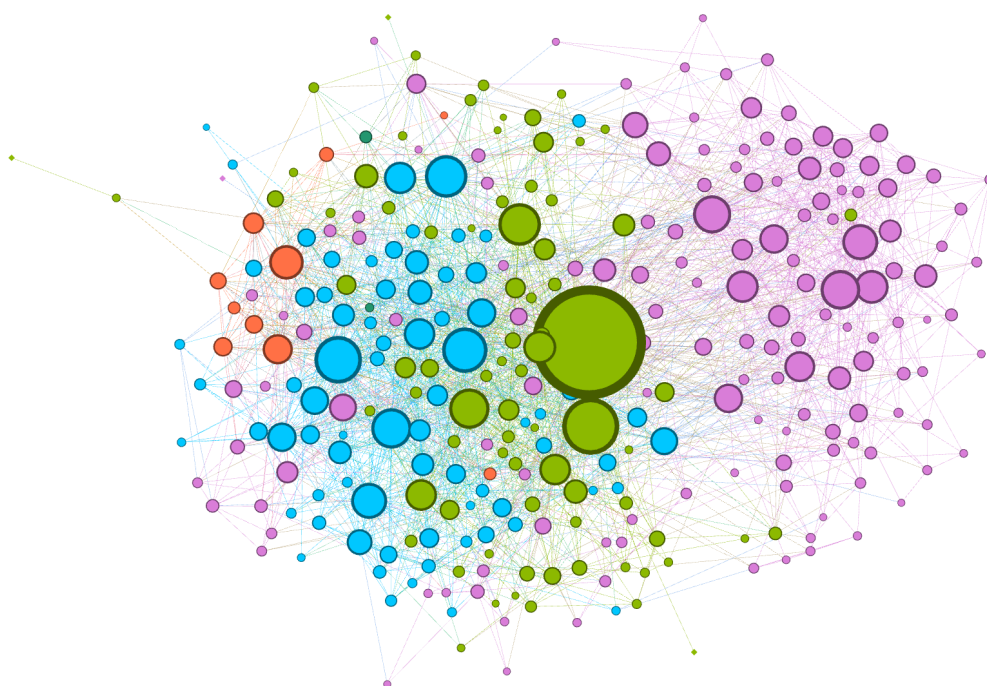


Figure 3.4: Collaboration network between teams participating in iGEM 2018. Teams are coloured based on their geographic region and the size of the vertex corresponds to the degree (number of collaborations that the team was involved in).

Inter-team Collaborations

Teams report their outreach and collaborative activities on iGEM on their wiki. Collaboration with other teams usually takes the form of socialising at a meetup, working together on each others' project, sharing resources and providing advice and support. The full text of a team wiki can be leveraged to identify these inter-team collaborations.

Once the full text has been extracted using parsers from the *BoilerPy3* library in python, we remove html content and tags using the *justext* library in python and then tokenize text into sentences using the English pickle tokenizer ('tokenizers/punkt/english.pickle') from the *nlTK* library. For each sentence in the wiki text, we search for the mention of an iGEM team name that participated that year. To account for slight variations in spelling, we use an approximate fuzzy matching - using the *partial_ratio* function of the *fuzzywuzzy* library in python. This allows for comparing substrings in each sentence with the name of an iGEM team, keeping the Levenshtein distance cutoff at 0.9. We handle the edge case where a tokenized sentences shorter than the team name, by using regular fuzzy matching (with a cutoff of 0.7) instead of partial matching. The latter generates false positives by matching team names to common page titles - like matching software teams (USTC_Software, Michigan_Software, SYSU-Software etc.) to the recurring page title SOFTWARE. This gives a dataframe with the name of the source team, the names of teams mentioned on their wiki and the sentence where the collaboration is reported.

The fuzzy matching method is susceptible to include "false positives" - where team names are matched to text which do not describe a collaborative activity. We do filtering to reduce the presence of these false positives. The first filtering step aims to remove the influence of *template pages*. These pages are created by iGEM HQ as a part of the template wiki, and sometimes remain an unedited part of their wiki (but not accessible publicly due to them being unlinked to the main wiki pages). These template pages include contextual information and links to past iGEM teams who are exceptional examples for reference. These are falsely identified as collaborations during fuzzy matching. To filter these out, we identify the number of times a sentence reports a collaboration to the same target team. If a sentence is repeated multiple times (with a conservative cutoff at 40), we claim that the only explanation for this high frequency is that they are from the template pages, and are thus present in several team wikis (40 in this case). Collaborations reported by these sentences are removed. To achieve a more precise filter - we use regular expressions. In a template page, referencing another iGEM team always follow the same format: [Year][][Team Name]. We build a regular expression to identify sentences which contain phrases of this format. If that same sentence is repeated by over 10 teams that year, we exclude the collaborations reported.

Another common false positive arises with the MIT open source license statements, which are mismatched with the iGEM team *MIT*. We build a dictionary to identify common phrases used to identify open source licenses and remove sentences that include said phrases. The dictionary of phrases include:

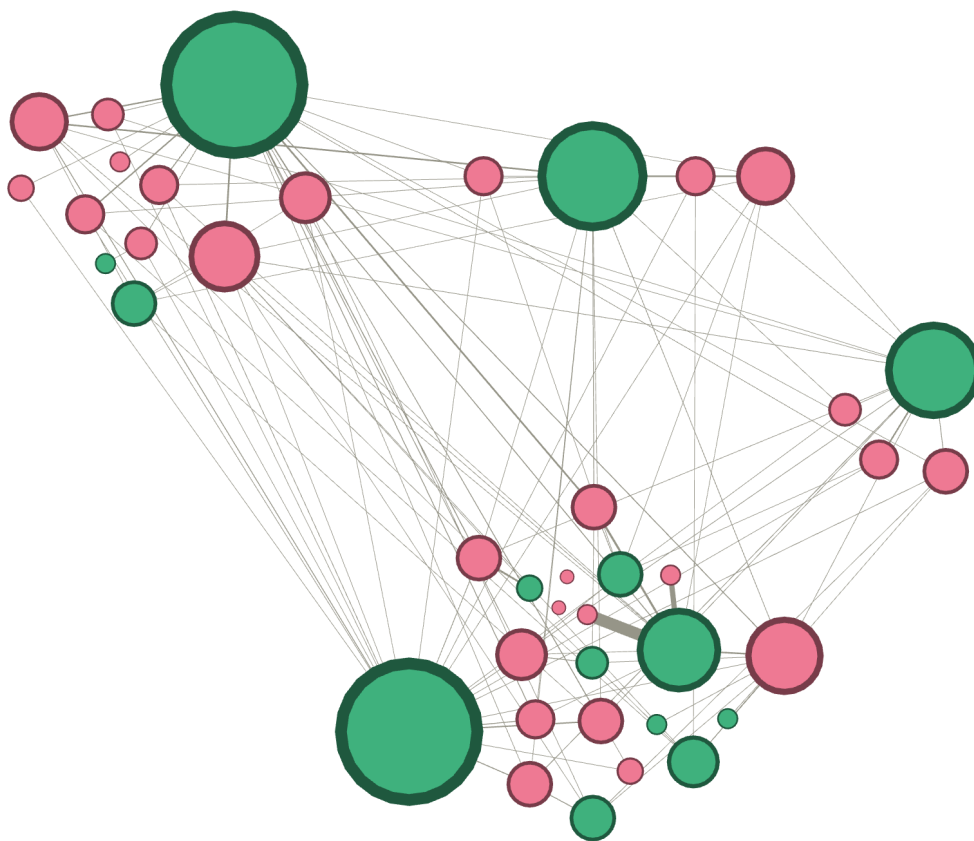


Figure 3.5: Example of a Wiki Co-edit network. Vertices coloured green correspond to the team members and red to the pages of the wiki. The network here shows a modular organisation of the teams' wiki editing process.

"Licensed under MIT", "Licensed under the MIT", "MIT License", "MIT license", "COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE", "PROVIDED BY THE COPYRIGHT HOLDER", "(Custom Build)", "WITHOUT WARRANTY OF ANY KIND", "Licensed MIT", "MIT/GPLv2 Lic", "MIT/GPL2", "Released under the MIT", "Dual licensed under", "License: MIT", "MIT Press", "MIT_License", "LICENSE-MIT", "MIT/BSD license", "MIT @license", "MIT Technology Review", "THIS CODE IS PROVIDED ON AN *AS IS* BASIS" and "MIT (c)"

The final collaboration network has 15,423 interactions between teams from 2008-2018. An example is presented in figure 3.4.

Co-edit structures of Team Wikis

The digital traces from the MediaWiki API includes the IDs of the editor (who is a registered iGEM team member), the identifier of the wiki page, nature of the action (create, edit, delete) and its size. Since all wiki pages are initially created as a template by iGEM HQ - we discard those actions. The editing structure of the wiki is represented as interactions between team members and wiki pages. An example is presented in figure 3.5.

3.1.4 Discussion and Perspectives

The section presents the data from the iGEM team-based science and innovation ecosystem and presents the methods to construct a consistent dataset of teams, participant and project meta data and most importantly, digital traces capturing intra- and inter-team interactions.

The co-edit structures from team wikis provide insights into how members organised to communicate their project output in iGEM. However, editing the wiki is one task, with team members performing several additional roles while developing their project. This specific task allocation between team members is not captured by the wiki edit structures. However, team wikis have an attribution page where they report on who in the team did what task. This attribution statement is unstructured, requiring language processing tools to identify the network of task allocation.

The fuzzy matching method to curate the inter-team collaborations also has its limits. It only identifies a collaboration if the target team is reported in text using its exact iGEM name or a slightly similar variation. In the next section of this chapter, I present methods leveraging Large Language Models to recover both intra-team collaborations from the attribution statements and the inter-team collaborations reported on team wikis.

3.2 Team Collaboration Dynamics from Self-Reported Data

Abstract

Science is a collaborative endeavor. Yet, unlike co-authorship, interactions within and across teams are seldom reported in a structured way, making them hard to study at scale. We show that Large Language Models (LLMs) can solve this problem, vastly improving the efficiency and quality of network data collection. Our approach iteratively applies filtering with few-shot learning, allowing us to identify and categorize different types of relationships from text. We compare this approach to manual annotation and fuzzy matching using a corpus of digital laboratory notebooks, examining inference quality at the level of edges (recovering a single link), labels (recovering the relationship context) and at the whole-network level (recovering local and global network properties). Large Language Models perform impressively well at each of these tasks, with edge recall rate ranging from 0.8 for the highly contextual case of recovering the task allocation structure of teams from their unstructured attribution page to 0.9 for the more explicit case of retrieving the collaboration with other teams from direct mentions, showing a 32% improvement over a fuzzy matching approach. Beyond science, the flexibility of LLMs means that our approach can be extended broadly through minor prompt revision.

Contributions and Publication

The work presented here is published in the journal *Applied Network Science* in 2024. The following section reports the contents of this article as is, reformatted to the thesis.

JEYARAM, R.*, WARD, R. AND SANTOLINI, M. (2024) 'LARGE LANGUAGE MODELS RECOVER SCIENTIFIC COLLABORATION NETWORKS FROM TEXT', *APPLIED NETWORK SCIENCE*, 2024

I took the lead on data curation, writing software code, performing the analysis and visualisation and writing the manuscript. Credit to Robert Ward and Marc Santolini for conceptualising and supervising the study. The manuscript was collaboratively reviewed by the three authors.

3.2.1 Introduction

A large amount of social network data is recorded in unstructured text: nodes are mentioned with non-standard labels, and the edges between them, which may constitute one or more kinds of interaction, are described with idiosyncratic, contextually dependent phrases. This makes it difficult to accurately reconstruct networks in an automated way. While manual annotation can be performed for small datasets, this approach becomes prohibitively costly at scale. In particular, leading scholars to either rely on low quality data, or to direct attention elsewhere.

Research on scientific collaboration is a good example. Identifying who works with whom, and what they contribute is essential for the allocation of resources and credit, as well as for improving the organization of scientific work [Fortunato et al., 2018]. Existing research focuses almost exclusively on co-authorship, in part because it can be easily observed in an article’s metadata. However, sociologists have long described the invisible college that underlies research production: helpful scientists who share resources and advice without being included as authors [De Solla Price and Beaver, 1966, Oettl, 2012]. These informal collaborations are described in the acknowledgements that accompany tens-of-millions of articles. They are accessible to researchers, but unused because they would be prohibitively difficult to extract with existing methods. Similarly, the division of labor amongst co-authors is central to the quality and reliability of their work. Journals are increasingly adopting contribution statements to describe the division of labor. But only a fraction of them are reported in a structured format, and even those that do typically lack fine-grained descriptions of the tasks [Larivière et al., 2020, Sauermann and Haeussler, 2017]. Other extensive qualitative approaches such as surveys [Walsh and Lee, 2015], interviews and manual annotation (Lazega et al., 2008) or self-reported statements [Masselot et al., 2023] have been used, and despite ensuring high quality in the curated data, are severely limited by time, granularity and resource constraints. As a result, we have a rather narrow and uncertain view of how scientific collaborations are formed, structured, and affect knowledge production [Hall et al., 2018].

In this study, we focus on a context-rich, large-scale text dataset of wiki-based digital laboratory notebooks from 2,000+ scientific teams participating in the international Genetically Engineered Machines (iGEM) synthetic biology competition [Santolini et al., 2023]. In their wiki page, teams document their scientific project and outcomes, along with team member attributions – who did what – and collaborations with other teams. Prior work using the iGEM dataset reconstructs a proxy of the task allocation network from how members of a team co-edit different sections of the wiki. However, moving beyond digital traces, the rich textual information provides a more elaborated view on task allocation structure in scientific teams. As such, obtaining these inter-team collaboration networks and intra-team organizational structures across this large number of teams requires extracting specific information from heterogeneous and unstructured text.

To address these challenges in measurement, we present a semi-supervised approach

leveraging on Large Language Models (LLMs). We evaluate the performance of using LLMs on two network retrieval tasks from unstructured text of increasing difficulty. First, we aim to identify the inter-team collaboration network, where direct interactions between teams are encoded within the wiki text. In this task, we have some prior knowledge about team names, and are interested in retrieving team collaborations along with the type of interaction. Second, we identify the task contribution structure of iGEM teams from the self-reported attribution statements of team members. Here, we retrieve a member-to-task bipartite network where we have prior knowledge about team members, and are inferring task allocation from contextual data. For both cases, we use a manual labeling test set to evaluate the accuracy of LLMs in retrieving relationships and their contexts, and investigate their ability to accurately reproduce local and global network properties. Beyond the quantitative results, we aim to provide the reader with a guide on the good practices and pitfalls we experienced with using LLMs for these information retrieval tasks.

The paper is organized into 4 further sections. We talk about recent approaches in using LLMs in information retrieval and present some background for the iGEM competition and dataset. We then describe in the Methods section the pipeline to curate and validate the intra-team contribution structure and the inter-team collaboration network. Finally, we present the results and discuss perspectives for future work in leveraging the iGEM dataset and using LLMs for network reconstruction in computational social science research.

3.2.2 Related Work

Text data encodes relationships between heterogeneous entities. Recognizing these relations is crucial across various scientific disciplines, and has benefitted from the advent of natural language processing and machine learning methods [Detroja et al., 2023, Pawar et al., 2017]. For example, biologists might be interested in retrieving interactions between proteins from the published literature to build a comprehensive protein interaction network that can serve medical insights (Rolland et al., 2014). For this, researchers will have to navigate the different ways these proteins can be referred to, and the different phrase structures that can evoke a similar relation. In the humanities, scholars might be interested in retrieving character networks from fiction works to study narrative patterns across stories (Labatut & Bost, 2020). In social sciences, the extraction of relationships from online data, free text surveys or unstructured self-reports can recover multimodal social relationships at scale [Deri et al., 2018] (Deri et al., 2018; Irfan et al., 2015).

Ensuring a high quality of information retrieval from text is challenging and requires tedious manual annotation and validation with experts. The use of information coding software, such as MAXQDA, has provided a strong support for the analysis of qualitative and mixed data, such as surveys with both standardized and open-ended questions (Kuckartz & Rädiker, 2021). Yet, with the increasing availability of large-scale digital text data collections, the complexity of extracting and annotating these relationships increases further. To cope with scale, a popular method has been the use of Amazon’s Mechanical Turk (MTurk) service. By leveraging non-expert crowd workers, the MTurk platform presents a low-cost

and scalable method for text annotation tasks, showcasing significant correlations with expert annotation [Snow et al., 2008]. However, recent studies have shown a significant decrease in data quality and reliability, requiring additional response validity indicators and data screening (Chmielewski & Kucker, 2020).

Against that background, Large Language Models (LLMs) present an opportunity for assisting mixed methods research at scale [Karjus, 2023]. Leveraging their ability to understand linguistic contexts and address low to medium complexity tasks, studies have quantified the efficacy of using LLMs for text annotation, and showed they improve over manual curation using Amazon Mechanical Turk [Gilardi et al., 2023, Törnberg, 2023](Alizadeh et al., 2023; Gilardi et al., 2023; Törnberg, 2023). By refining prompts using manually curated validation data (Reiss, 2023), annotation tasks can be customized for each use case to yield more relevant results. These encouraging findings have led to the development of open-source tools such as GraphGPT [Shenoy, 2023] that extract heterogeneous information networks out of text data, paving the way for the application of LLMs in the context of network reconstruction. Preliminary insights using the textbook example “Les Misérables” character network reconstruction from Hugo’s eponym fictional work show a promising increase in richness of information retrieval compared to the smaller manually obtained version, including non-plot characters discussed by Hugo in tangential sections of the book [Karjus, 2023].

As such, LLMs are a promising tool for reconstructing social networks from unstructured text data. Focusing on the digital lab notebooks from teams participating in the iGEM competition, we aim to address the gap in evaluating the efficacy in using them as an adaptive solution for curating social networks and present the framework as a low-effort, yet reliable solution to address similarly structured tasks in the information and the social sciences.

3.2.3 The iGEM Dataset

Launched in 2003, the iGEM competition has become a cornerstone in the Synthetic Biology field, promoting an open, collaborative approach to solving real-world problems using standardized DNA elements, or “BioBricks.” It encourages a bottom-up, community-based learning method, fostering dialogue and transparency around ethical and safety concerns. From its inception with just 5 teams, iGEM has expanded globally, involving over 4,300 university and high school teams from 40 countries. Annually, teams use BioBricks to create synthetic biology solutions, culminating in a Jamboree where their projects are judged. The competition emphasizes collaboration and high-quality documentation, with each team’s wiki serving as a key evaluative tool.

The team wiki website serves as an extended scientific article that includes technical specifications of the project, experimental and modeling methods, project results, and outreach practices the team members undertook. In addition to describing the content of the

team project, the website contains context-rich information on the team members, their background and contribution to the project, and the collaborations the team maintained with other iGEM teams. This rich data source enables analysis of collaboration networks and project outcomes, offering insights into the practice and impact of collaborative science in interdisciplinary contexts [Santolini et al., 2023].

While wikis have an overall topical structure dictated by the needs of the competition – i.e. they must showcase collaborations with other teams and describe the contributions of team members –, the presentation of the content within each page varies significantly across teams. With over 2,200 participating teams in 2008 to 2018 and an average 17 members per team, manually curating and labeling individual contributions within teams and collaborations across teams is a time and resource consuming process. Here we use LLMs and specifically the GPT family of models to tackle this challenge and evaluate their performance.

To reconstruct the team collaboration and task allocation networks, we first extract the raw text from all teams public wiki pages. The text content of the pages was extracted using the "KeepEverythingExtractor" option in the boilerpipe.extract library for processing and removing boilerplate content after web scraping. This amounts to 106,207 pages from 2,265 teams over 11 years, from 2008 to 2018. The raw html and processed text of the wikis is available in Zenodo, along with team metadata from the iGEM website [Blondel et al., 2024].

3.2.4 Methods

Network Reconstruction Pipeline

We show in 3.6 an overview of the workflow to reconstruct the collaboration and task allocation networks in iGEM teams using GPT models. The workflow is similar for both the intra- and inter-team networks, with specific changes in the implementation.

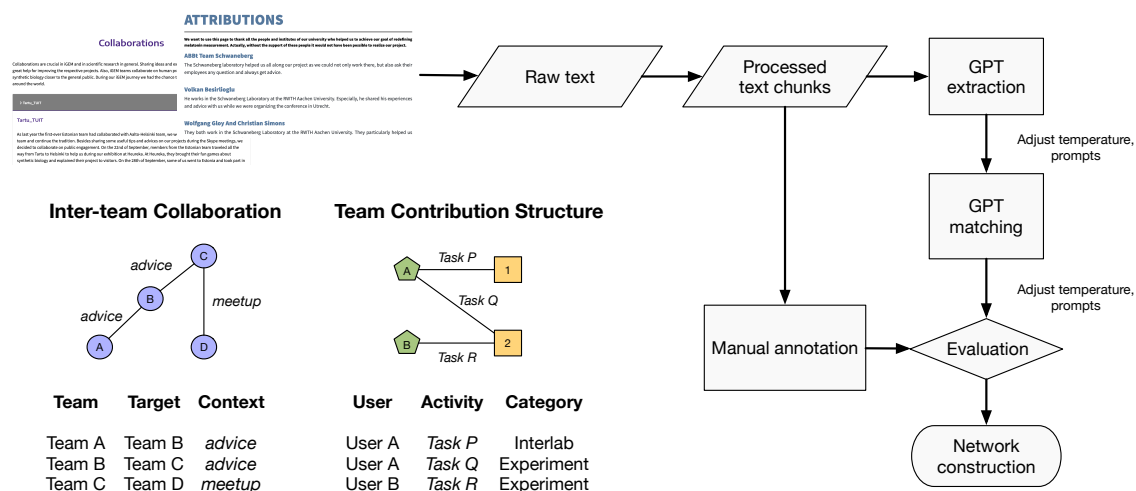


Figure 3.6: Overview of the pipeline to reconstruct social networks from text

Pipeline to extract the team contribution and inter-team collaboration structure from the text data of wiki pages from iGEM teams. The text from the wiki page is processed and passed to the language model along with the prompt. The output is a table with the decorated relationships, i.e. the user identifier, the activity they performed and the matched category for the contribution network and the team, target and the context of collaboration for the inter-team collaboration network. Schematic networks are shown in the bottom left, where circular blue nodes denote teams in the inter-team collaboration network, green pentagons denote team members and yellow rectangles the categories of tasks they performed in the intra-team contribution structure.

The first step is to process the text from the wiki pages into chunks. This is to primarily remove pages that do not contain information about either inter-team collaborations or about intra-team contributions. This step also ensures that input size limits (tokens) for the GPT models are preserved.

The curation using the GPT models is split into two steps. The first step extracts all relationships between entities present in the provided input text along with their described context. The objective of this step is to maximize recall and to ensure that there are minimal false negatives. The second step aims to match the extracted entities with their corresponding identifiers. In the inter-team case - it is to match team names to their official team IDs and the context of their collaboration to one of five standard collaboration categories in iGEM - “Work”, “Material Transfer”, “Meetup”, “Advice” or “Other”. For the intra-team case, it is to match team members to their member IDs and the tasks they undertook to a standard list of tasks that teams are expected to perform during the iGEM competition. This list is inspired by the CRediT Contributor Roles Taxonomy [Larivière et al., 2020], and constructed to encompass all tasks and deliverables that teams work on and are evaluated on in iGEM. The categories are: “Design”, “Experiments”, “Documentation”, “Interlab”, “Modeling”, “Analysis”, “Parts”, “Safety”, “Entrepreneurship”, “Hardware”, “Software”, “Human Practices”, “Public Engagement”, “Collaboration”, “Fundraising”, “Creative Contributions”,

“Administration”, “Material Supply”, “Supervision”, “Training” and “Other”. In this step, we eliminate the false positives and increase the precision of the final curated network.

To evaluate the curated networks, we create a validation set by manually curating relationships from teams and text chunks selected at random. For the inter-team case, we also use fuzzy matching to construct an alternate reference set - although this approach does not identify the contexts of inter-team collaborations. As a part of their iGEM project in 2016, team Waterloo constructed a database of collaborations between teams in 2015 - which we use as a further benchmark to evaluate the quality of the approach.

Specifications of the intra- and inter-team curation - such as the text processing, the prompts used, constructing the validation sets and the implementation details using the GPT models are detailed in the Supplementary Information.

Statistical Analysis

Standard errors for the precision and recall scores of the inter-team collaboration networks in Fig 3.7.a are computed using the jackknife method over chunks. In the case of the intra-team contribution networks, precision and recall are computed at the team level and we show in Fig. 3.8a the average and standard error across the manually curated teams. The confusion matrices between relationship contexts for the intra- and inter team networks in Figures 3.7b and 3.8b are row normalised using Z-scores: each observation x is centred and standardised using $Z = (x - \mu) / \sigma$, where μ is the row average and σ its standard deviation.

We use the pROC library in R for computing the ROC curves and Area under the ROC curve (AUC) of Fig. 3.7c.

Hierarchical clustering in Supplementary Fig. 3.2.7. is performed by computing the Euclidean distances between predicted categories from the confusion matrix (i.e. the rows) with the R `dist` function, and then using the “`hclust`” function in R.

All network analyses were performed with the `igraph` library in R. The betweenness and closeness centrality measures are unweighted and computed using their namesake functions. The coreness of the network was computed using the “`coreness`” function and the local clustering coefficient using the “`transitivity`” function setting the `type` argument to be “`local`”. The network assortativity measures were computed for the categorical vertex attributes region, country and section using the “`assortativity_nominal`” function in `igraph`. The degree assortativity is computed using the “`assortativity_degree`” function. Standard errors for the network properties are computed using the jackknife method by removing one team from the network in each iteration, recomputing the corresponding network property, and calculating the standard deviation across the obtained values. The degree distributions of the intra-team bipartite networks are computed using the in-built func-

tion `degree_distribution` from the `igraph` library with the “cumulative” option set to true. To quantify nestedness, we compute the standardized NODFc nestedness metric based on overlap and decreasing fill, described in [Song et al., 2017], using the `maxnodf` library in R [Hoeppke and Simmons, 2021].

3.2.5 Results

Team Collaboration Networks

Accuracy of Inferred Relations

We first investigate the accuracy of GPT models and a fuzzy matching approach in retrieving the collaboration relations between iGEM teams. Our approach is described in Methods. We manually curate 200 text chunks from the Collaboration wiki pages across teams to extract the relationships present in the text. The same text chunks are then passed through the GPT and fuzzy matching pipelines to retrieve predicted relationships. We compare the performance of the GPT and fuzzy models against manually curated data in Fig. 3.7a. We focus on precision and recall, quantifying respectively the proportion of predicted relations that are in the manually curated set, and the proportion of manually curated relations that are retrieved. We find that newer GPT models consistently increase in precision and recall. Amongst the various models tested, `gpt-4` has a higher recall while `gpt-3.5-turbo` and `gpt-4-1025-preview` models have a comparable performance, with the former being 10x cheaper. The turbo instruct models allow for inbuilt parallel request calls, but their performance is a detrimental tradeoff for information extraction and curation tasks. While precision flattens after GPT3.5-Turbo-16k, recall shows a significant improvement with GPT4, with a recall of 0.91. Interestingly, while fuzzy matching does not yield too many false positives – it has competing precision with GPT3.5 models –, it does not provide a good coverage of manual edges, with a recall of 0.69, giving GPT4 a 32% improvement over the fuzzy approach.

Inference of the Collaboration Context

Beyond edge retrieval, we evaluate the ability of the GPT models to infer the collaboration context from the surrounding text. In the following we focus on `gpt-3.5-turbo16k` as it has the higher performance-to-cost ratio. This harder problem, that the fuzzy approach does not tackle, is evaluated by comparing the results to manual labeling. We show in Fig. 3.7b the comparison between the categories inferred with `gpt-3.5-turbo-16k` and the manually labeled ones. We find an overall accuracy of 0.26 – i.e. the predicted category matches the manually labeled category 26% of the time. However, there is variation across categories. ‘Meetups’ are identified with a high accuracy of 0.65, while categories ‘Advice’ and ‘Material transfer’ have accuracies of 0.46 and 0.45 respectively, and are sometimes confused with one another. On the other end of the spectrum, most of the relationships identified as a ‘Work’ relationship by GPT are manually labeled as ‘Other’, leading to a poor accuracy of 0.16 for this category. This may be due to a lot of non-work relationships still using phrases like ‘we worked with’, ‘we collaborated with’ in their descriptions. Overall, these results show that while there is some degree of confidence in extracting categories from

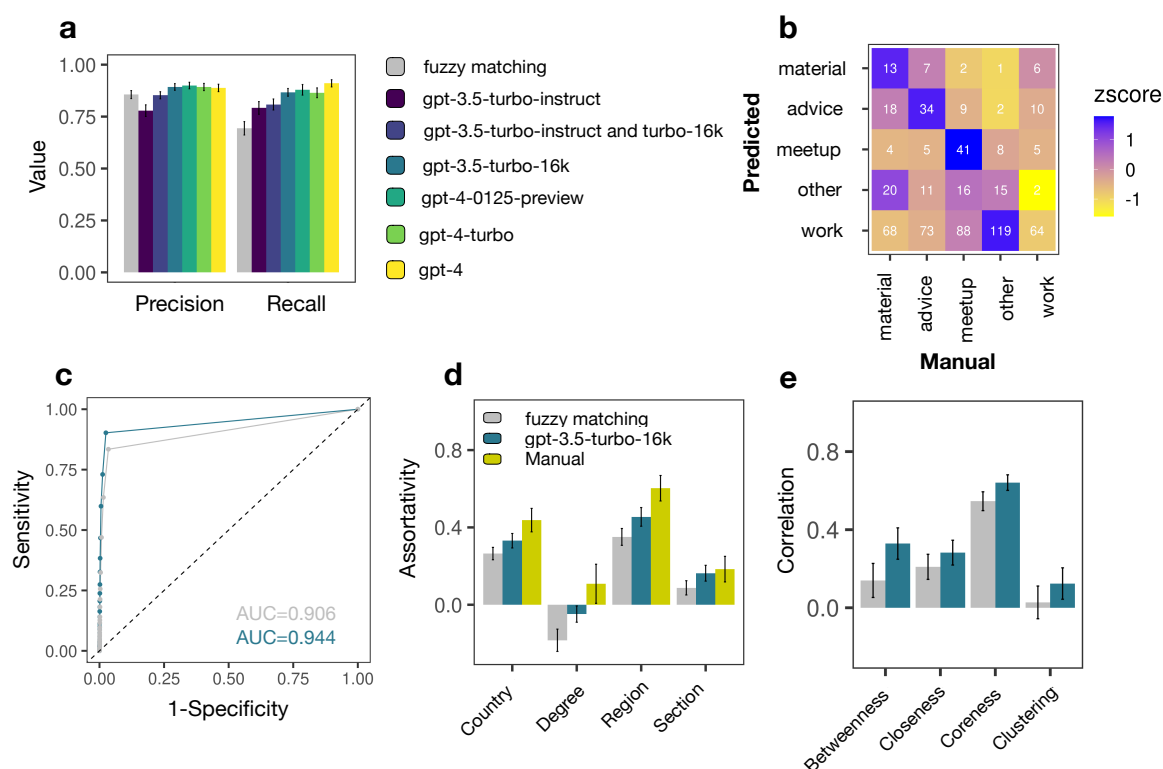


Figure 3.7: Reconstruction of inter-team collaboration networks (a) Precision and recall of GPT models and fuzzy matching approach against manually curated chunks. (b) Confusion matrix showing the accuracy of category prediction for relationship contexts. The color coding indicates row-standardized Z scores values. (c) Receiver Operating Curves (ROC curves) for weighted interactions from 2015. Dashed line indicates a line of slope 1 and intercept 0. (d) Comparison between network assortativity values for reconstructed and manually obtained 2015 inter-team network. (e) Correlations of network properties between the 2015 reconstructed and manual network. In all plots, the error bars are computed using the jackknife method

context, there is a strong variation between categories, with some being very accurate and others very hard to predict. We note that this exercise depends on the taxonomy chosen for the categorization, and the poor results in some categories might as well be seen as a poor definition of the category itself. Supplementary Fig. 3.7 shows the confusion matrix between categories for the other GPT models, indicating a similar pattern with the other models.

Accuracy of the Reconstructed Network

While previous measures focus on the ability to infer an edge and its context, here we investigate the ability of GPT models to infer accurate network properties in the reconstructed network. To do so, we evaluate the gpt-3.5-turbo model and fuzzy approach against a manually curated network of the significant inter-team collaborations in 2015, extracted by the 2016 Waterloo iGEM team (see Supplementary Information). This curated

network does not comprise all team mentions that can be retrieved from the pages, but only the ones deemed significant by the curators. For both inference methods, the edges are weighted by the number of unique text chunks where the corresponding collaboration is reported. These weights can then be used to predict the occurrence of a significant edge in the manually curated network. We show the resulting ROC curve in Fig. 3.7c, along with corresponding Areas under the ROC curve (AUCs). We find that GPT outperforms fuzzy matching with an AUC of 0.944 (edge recall of 0.90) compared to 0.906 (recall of 0.83) for the fuzzy method. Finally, we investigate the extent to which the reconstructed networks, though still somehow noisy at the edge level, recover some key network properties of the manually reconstructed network. We show in Fig. 3.7d that GPT3.5 consistently improves over fuzzy matching to recover assortativities of the team attributes. We finally investigate if GPT can recover the relative importance of nodes in the network. For this we use Spearman correlation as a measure of the similarity of ranks under various network centrality measures between the inferred networks and the manually curated one. We find larger Spearman correlations across several local (clustering) and global (betweenness, closeness, and coreness) centrality measures for GPT compared to fuzzy matching, showing the ability of the former to build a more accurate overall representation of the network. We note poor results for clustering, which might be due to curation decisions – e.g. meetups between several teams can create large cliques that might not be considered significant by the Waterloo team.

Team Contribution Structure

Accuracy of Inferred Relations

In this second part, we focus on the question of identifying the tasks that team members have done in their project from the attribution page. Identifying the contribution structure of a team is a more complex task as only the name of the member is encoded in the text, while the category of the relationship they were involved in is inferred from the description of their activities. We first compare the precision and recall scores for the task categories retrieved using GPT models against the manually curated teams. Figure 3.8a shows a significant performance improvement in using the newer gpt-4-turbo class of models, with an edge-level recall across the manually labeled teams of 0.79, and a precision of 0.62. The lower precision means that descriptive task contexts are often matched to both the manually assigned category (high recall), but also to other categories, decreasing the precision.

Task Level Accuracy

To investigate this mismatch, we compare the manual and GPT matching of the description of the individual members' task attributions to the defined task categories. Figure 3.8b shows the confusion matrix between the predicted and manual labels. Overall, we find an accuracy of 0.66, with variation between categories (see Supplementary Fig. 3.12 for precision and recall values across categories). We observe higher accuracy for categories with specific descriptions - such as being involved in interlab (accuracy=0.92), performing human practices (0.80) or modeling work (0.79). However, categories with high similarities in their descriptions are either misidentified, or identified to all similar alternatives, especially amongst experiments, performing analyses, design or between supervision and

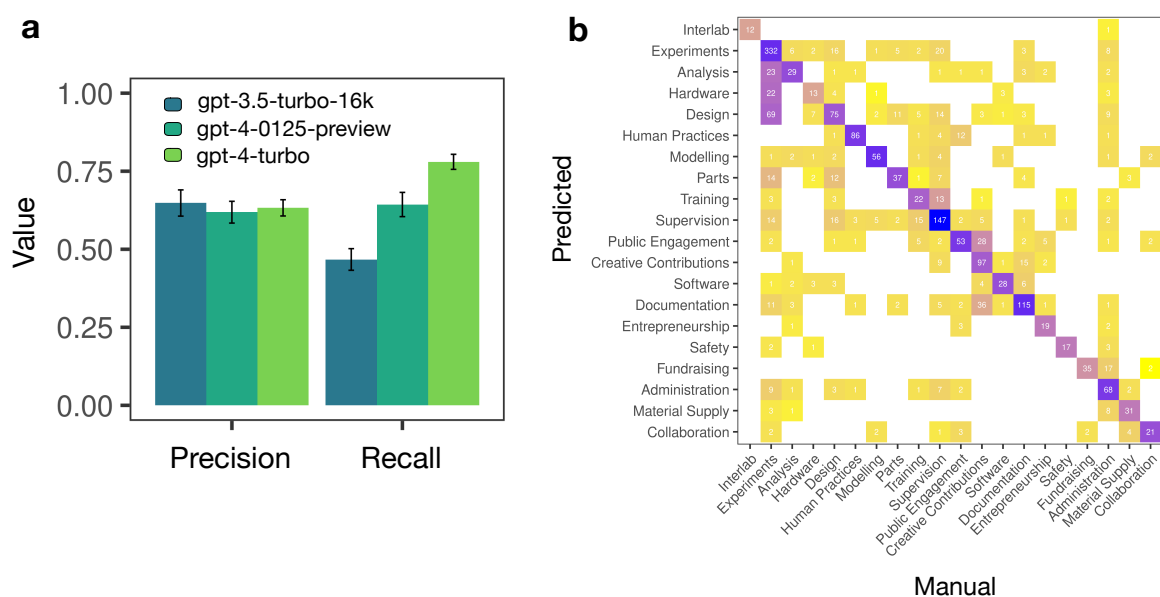


Figure 3.8: Reconstruction of team contribution structure (a) Precision and recall of GPT models against manually curated contribution networks. (b) Confusion matrix of inferred category labels

training. The clustering of categories based on the similarity of their labeling profile in the confusion matrix (Supplementary Fig. 3.2.7) identifies groups of strongly interrelated tasks (e.g. supervision and training), corresponding to similar tasks when working within an iGEM team. Information from these clusters could be used to increase the specificity of each subcategory description and reduce their similarity, or merge similar categories to improve the overall accuracy.

Accuracy of Inferred Network Properties

Finally, we go beyond the edge level to investigate the ability of GPT to reproduce key properties of the team bipartite networks. First, we show in Figure 3.9a the degree distributions for each layer in the inferred and manual networks. We find similar distributions, with large Pearson correlations between (the log10 of the) degrees of $r=0.76$ ($p<2e-16$) for users and $r=0.85$ ($p<2e-16$) for tasks. In addition to degree distributions, bipartite networks are often characterized by a correlation between the degrees of the layers, leading to nested structures that can be quantified with the NODFc value [Mariani et al., 2019]. We show in Fig. 3.9b the predicted and observed values of nestedness, with a high Pearson correlation of $r=0.66$ ($p=1e-4$). We find the presence of an outlier team with a high manually curated NODFc and a low predicted one. Under inspection it is due to the specificity format of the team page, involving images of team members rather than their name. Such outlier events can be taken care of using preprocessing of high-confidence pages that can offer a good inference, for example by evaluating the proportion of names of team members present in the page.

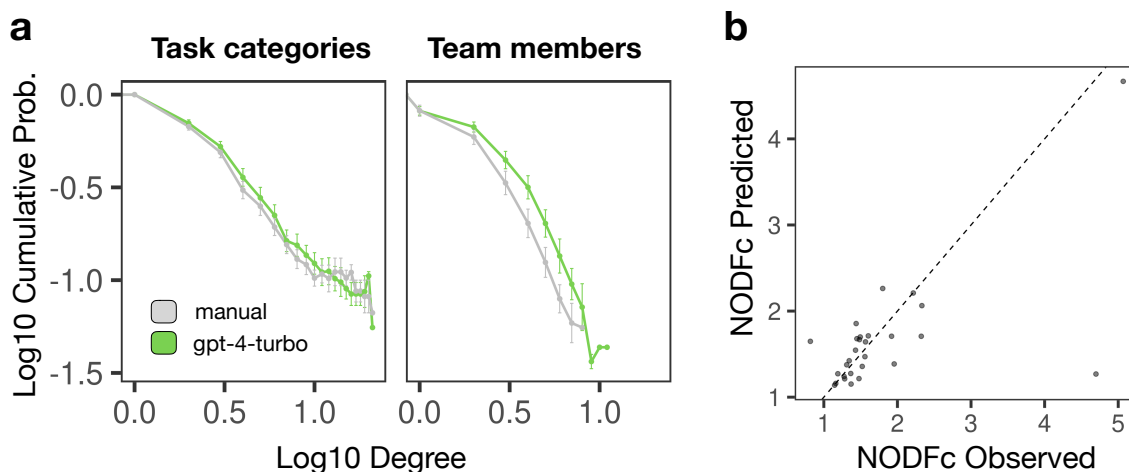


Figure 3.9: Recovering bipartite network properties. (a) Cumulative degree distributions for the task categories and team members layers of the bipartite network, comparing manually curated networks (grey) with gpt-4 inferred networks (green). (b) Nestedness values, computed using NODFc [Song et al., 2017], are shown between the predicted and manual networks. Dashed line indicates a line of slope 1 and intercept 0

3.2.6 Discussion

This study aimed at showcasing the use of GPT models to infer team collaboration networks from text, providing a quantified account of what had been scoped in previous work [Karjus, 2023]. Overall, we found that the use of GPT models has clear benefits for extracting such information at scale. In extracting direct relationships between teams encoded within the text, it performs significantly better than a fuzzy approach, and can recover key aspects of the network structure. For example, fuzzy matching fails in cases where teams mention the names of members using their first name or nicknames, or in cases where they describe teams by the names of the universities they are based in rather than their team name. Similarly, the partial substring matching of the fuzzy approach has difficulty disambiguating team names with similar acronyms, such as UCL and UCLA. GPT models overcome these challenges through their larger inbuilt contexts which allows better disambiguation. Extracting the contexts of the relationships is another advantage that GPT models have over fuzzy matching methods, which would require additional language processing methods such as topic modeling to extend to other similar settings.

The similar performance between gpt-3.5-turbo and the more expensive gpt-4 class of models, unlike the lower performance observed in the second, more complex task, suggests a higher feasibility in reconstructing these direct relationships with simple few shot prompts and minimal post-processing. We also note that further pre-processing of the dataset aiming to provide higher quality text through filtering out pages with low abundance of certain words (such as team names or team member names), would probably increase the overall precision and recall, while decreasing the amount of data that can be treated with the method.

Matching the contexts of the relationships to a list of categories has certain caveats. With the pipeline first extracting the relationship contexts and then performing the matching, some information beyond the sentence describing the relationship is lost. Attempting to perform both the extraction and the matching in a single step preserves most contextual information from the text, but performs poorly in precision and recall, especially when tested with the cheaper gpt-3.5 class of models. This could potentially be caused by an individual request now performing a more complex task, which current advances in model complexity might resolve.

Evaluating the performance of reconstructing the indirect relationships of the intrateam contribution structure shows promising initial results, with drastic improvements observed with newer GPT models. This might reflect the task complexity that heavily relies on context-awareness for the task allocation to team members. The similarity between the properties of the manual and inferred networks – degree distributions and nestedness values – indicates that the inferred networks can be leveraged for future studies comparing their structure across the large number of teams of the iGEM competition.

We also note that the process of manual curation is itself prone to inter-individual variation. Decisions on assigning a description to a specific category relies on personal judgment, and quantifying this variation would require a more extensive investigation using platforms like Mechanical Turk. As such, it is yet not clear whether GPT models already outperforms human annotators or not in the context of this study. The high recall rates and comparatively low precision have been interpreted as a larger richness of GPT models [Karjus, 2023], capturing more peripheral relations than immediately catches the eye. Future work could investigate whether this richness helps capture more subtle phenomena that can be useful in downstream analyses.

Finally, the ability to extract and label relationships from text in bulk saves on human effort, but does not eliminate it. There are several drawbacks in using proprietary LLMs - including but not limited to reproducibility, sensitivity to temperature, prompt adjustment and model updates. This creates the necessity to carefully evaluate model performance, ideally with high quality manually annotated data [Ollion et al., 2023, 2024](Reiss 2023). An additional concern is working with surveys and other confidential data sources. A potential workaround is by leveraging open source LLMs for network reconstruction. Open source LLMs have shown to outperform crowd workers, but lag behind the GPT models in text annotation (Alizadeh et al. 2023). However, this gap can potentially be reduced by fine-tuning open source LLMs with contextual data to improve on performance. The approach requires further human work in curating a training set, but allows customizability and executing requests in-house. This is a potential extension of the study.

3.2.7 Conclusion

We assessed the performance of using OpenAI’s GPT family of models in curating and annotating team collaboration networks from unstructured text. We leveraged digital laboratory notebooks from scientific teams participating in the iGEM competition to infer inter-team collaborations and the team contribution structure from heterogeneous self-reported data. We show that despite wide differences in page structures, the networks can be extracted with Large Language Models by using prompts that include minimally supervised team-specific information (list of users, team names, and types of contributions), resulting in structured graph data showing precision in par with manually curated data. We find that recall rates consistently increase with each new model, indicating that future improvements are expected with upcoming releases. We also showcase that LLMs can partially retrieve contextual information to infer edge types, with future improvements possible with new models and improved prompt design and category definition. This work has implications for the study of division of labor in scientific teams that often relies on CREDIT contribution reports [Xu et al., 2022], providing both a Method and a fine-grained dataset [Blondel et al., 2024] for studying fine-grained task allocation structure. More generally, this study suggests that LLMs can be a useful, scalable and efficient approach to network reconstruction for assisting manual curation work in computational social science and digital humanities.

Supplementary Information

Inter-Team Collaboration Networks

Text Processing

To reconstruct the inter-team collaboration network in a given year, we need to extract mentions to iGEM teams from each team wiki. Since most content of the wiki pages is not related to reporting collaborations between teams, we do an initial filter to remove pages that do not contain potential collaborations to reduce the number of API requests. We first keep the pages that contain the name of at least one other iGEM team participating that year. The lookup for other team names is performed by fuzzy matching using the fuzzy-wuzzy library in python. We used the `partial_ratio` function for substring matching, setting the Levenshtein distance cutoff at 0.9. If a page contains formatting elements like in CSS or Javascript, we remove it. This second step aims to remove pages whose contents are purely stylistic. This reduces the number of pages from 106,207 to 20,459.

In order to ensure that each request to the GPT model is a small modular task and to effect parallel execution, the text from these pages is split into overlapping chunks. Each chunk is of size 500 tokens with an overlap of 30 tokens with the previous chunk. A token is the unit of text clustering in GPT models and is approximately 4 characters long (<https://platform.openai.com>). This gives a total of 103,766 chunks, split into batches of 100 chunks each.

Curation with GPT

The entire process is split into an extraction and two matching steps as shown in Figure 3.6 of the main text.

Extraction The first step aims to extract teams from a chunk along with the context of their interaction with the team from which the chunk originates. In order to ensure a higher recall, the prompt instructs the GPT model to extract relationships between entities along with their context, where an entity is loosely defined to be a team, a university, a lab or similar. Teams that are matched to “Other” are eventually filtered out in post processing. Each request includes the prompt and the text from the corresponding chunk. The output is a table with the names of the source and target teams, and the context of their collaboration.

The prompt for the first extraction step is few-shot, providing contextual examples. The relationships are instructed to be output as a list of lists. Here is the prompt used:

“Extrapolate as many relationships between teams as possible from the text and provide a list of relationships. Teams may be affiliated with universities, labs or colleges. If a relationship is found, provide [“TEAM 1”, “LABEL”, “TEAM 2”]. “TEAM 1” is the team who wrote the text. “LABEL” is a description of the

relationship. "TEAM 2" is a team that has a relationship with TEAM 1.

example: We are team Greendale. We collaborated with team City College on our project this year and mentored iGEM Harmon College. relationships: [{"Greendale", "collaborated", "City College"}, {"Greendale", "mentored", "Harmon College"}]

example: We are team Mideastern. We received financial and material support from Addgene, and plasmids were generously gifted to us by the Baker lab at Southwestern University. We co-organized a meetup with Branch Western. We also participated in a survey run by Barkley. relationships: [{"Mideastern", "support from", "Addgene"}, {"Southwestern University", "gifted", "Mideastern"}, {"Mideastern", "co-organized meetup", "Branch Western"}, {"Mideastern", "participated in survey", "Barkley"}]

If there are no relationships in the text, provide [].

Matching The matching is performed in two steps. In the first matching step, the names of the source and target teams are matched to the actual team names from iGEM. Team names may vary by year, so the matching is done separately for each year. Each request includes the prompt, the list of all teams participating in iGEM that year and the list of names extracted in the previous step. Team names that don't have a close match are instructed to be matched to "other". We employ GPT for the matching task over conventional approaches like fuzzy matching to overcome several instances where the latter could fail. For example, a lot of teams report a collaboration by describing the name of the university the target team is based in, instead of the teams' iGEM identifier - like "Armed Forces Medical College" instead of "AFCM-Egypt" and "Chung Cheng University" for "CCU-Taiwan". Matching using GPT handles these cases better and reduces the chances of teams with similar names being matched with each other - like "UCL" to "UCLA". After matching all the names to their actual identifier, relationships where teams are matched to "other" are removed.

The prompt for matching the team names is one-shot and provides an example and prescribes the output to be a list of lists.

"Given a new list of names - find the closest match of each element in the new list with the first list. Some names could be matched with abbreviations or expansions - so please select the closest possible match. If there is no possible good match - please mention 'other'. Provide the output as a list of lists: [{"team name from the new list", "best matching from the first list"}]

example: list of teams: 'Gifu', 'CCU_Taiwan', 'Peking_China' list of names: 'Team Gifu', 'Chung Cheng University', 'pectinase' matches: [{"Team Gifu", "Gifu"}, {"Chung Cheng University", "CCU_Taiwan"}, {"pectinase", "other"}]

Here is the new list of names: "

Team names which exactly match the corresponding teams' identifier are excluded to reduce the number of requests. Since we use a comma to separate each name in the input list, each comma contained within an input team name is replaced with a single space. Team identifiers do not contain a comma by default.

In the second matching step, we use GPT to match the context of the filtered relationships to a general collaboration category. Collaboration relationships between teams in iGEM are usually a result of both teams working together, helping each others' project, providing/sharing resources or meeting and discussing in a social setting. Parsing several descriptions of collaboration activities on teams' wikis, we define a list of five categories: "Work", "Material Transfer", "Meetup", "Advice" or "Other" that encompass these relationships.

The prompt is few-shot, with its overall structure similar to the previous matching step. We also include short descriptions of each category we want the relationship to be matched with. Here is the prompt used:

"Match each relationship name to the closest possible category listed below:

Provide each matching as [[RELATIONSHIP, MATCHING CATEGORY]]

The possible categories are:

"work": Teams worked together or collaborated on aspects of their projects

"material transfer": One team shared information, data, synthetic biology parts or laboratory materials with the other.

"advice": One team gave advice or support to the other team concerning their project or about the competition.

"meetup": Teams met each other at a meetup or in a social setting and discussed their project.

"other": Contexts not fitting to the categories listed above.

example: "provided substrates to, participated in synthetic biology day with, did pcr amplification for" Matching: [["provided substrates to", "material transfer"], ["participated in synthetic biology day with","meetup"], ["did pcr amplification for","work"]]

example: "provided thoughts and suggestions to" Matching: [["provided thoughts and suggestions to","advice"]]

The following are the list of relationships to match: "

Similarly to the previous matching step, commas within the context description are replaced with spaces. The final curated relationships are weighted based on the number of unique chunks where a source-target pair is reported.

Evaluation

Extraction with Fuzzy Matching We compare the relationships curated using the GPT pipeline against the simpler fuzzy matching method. This method reports relationships between teams that are present in the text, but not their context. We use the same text chunks from the selected pages, split into sentences. In each sentence, we search for the mention of another iGEM team participating in the same year. Fuzzy matching is implemented using the `fuzzywuzzy` library in python. We used the `partial_ratio` function for substring matching, setting the Levenshtein distance cutoff at 0.9. For edge cases where the sentence is smaller than the length of the team name we are comparing with, we use a regular fuzzy matching with a cutoff of 0.7. This is because the partial ratio generates false positives in these edge cases by matching teams to common title substrings, like the postfix `Software` matching to all `Software` track teams (`USTC-Software`, `Michigan_Software`, `XMU_Software` etc.). Each relationship is weighted by the number of unique text chunks where the same pair of teams are observed.

Manual Curation of Text Chunks In order to evaluate the edge-level retrieval performance of the GPT models and fuzzy approach, we construct a random sample of 200 chunks containing text from the “Collaboration” wiki pages of teams, where the inter-team relationships are usually reported. These chunks were manually annotated by the authors to identify the source, target teams along with the context of collaboration. This context is then manually matched to one of the categories defined earlier - “Work”, “Material Transfer”, “Advice”, “Meetup” or “Other”.

Manual curation of the 2015 inter-team collaboration network In 2016, the iGEM team Waterloo¹ curated a database of team collaborations in 2015 as a part of their iGEM project. This database was constructed using the content of the 2015 iGEM wiki pages to manually select the significant collaborations between participating teams. To identify whether a collaboration is significant, the Waterloo team evaluated if the relationship helped advance each team’s project. This database thus contains a subset of high-confidence inter-team collaborations in 2015, serving as a benchmark to evaluate the quality of the GPT extraction pipeline.

Network Construction

Once all the relationships between teams are curated, we construct the inter-team collaboration networks for each competition year. Although the relationships mention the source and target teams, determining the true direction of collaboration requires further detailed specification in the prompt. For example, “We collaborated with team B. They have given us several reagents for our experiments” determines a direction of collaboration flowing from team B, but is often mislabeled by GPT as from team A. We hence construct the network as undirected. Each edge is weighted by the number of text chunks where the same pair of teams are reported to have a collaboration relationship. The networks are constructed using the `igraph` library in R.

¹https://2016.igem.org/Team:Waterloo/Integrated_Practices/Networks

Intra-Team Contribution Structure

Text Processing

In their wiki page, teams report a description of each member of the team and of the role and tasks they undertook during the course of the competition. Here is an excerpt of a self reported attribution of a team member - “Leading the project, arranged team meetings, user interviews and handled internal/external communication and was responsible for updating team website and survey.” The text clearly outlines the specific tasks the person was responsible for - namely team administration, maintaining the website, communication and outreach activities. The aim is to reconstruct these member-to-tasks relationships, thus describing the teams’ contribution structure.

To reduce the number of GPT requests, we filter the dataset to only include wiki pages that contain contribution statements and team member information. This filter is performed based on the presence of specific keywords in the page title, such as "Attribution", "Team", "Acknowledgement", "Member", "Contribution", "People" and "About Us". This dictionary was constructed by the authors by parsing through the most frequently repeated wiki page titles. This amounts to 8,184 pages across 2,173 teams.

Curation with GPT

Extraction The first step is to extract all the raw relationships between team members and tasks from the text. The prompt specifies instructions to extract names of people and the descriptions of the tasks they were involved in. This aims to parse the text completely and curate all possible member-task relations. The filtering out of false positives, such as people who helped the team, but aren’t in their official roster, or tasks that aren’t specific for the competition is done in post-processing. The request to GPT includes the prompt and the entire text content of the wiki page. Text is not chunked into smaller sizes as the structure of the attribution wiki pages with titles and small paragraphs meant a higher likelihood of mislabelling relationships (see Supplementary Figure 3.10). The output is a table with the name of the individual, the description of the activity and the identifier of the team.

The prompt is few-shot, instructing the output to be a list of lists.

“The following text describes the relationships between people and the activities they performed as a part of a synthetic biology team.

Extrapolate all these relationships, and if a relationship is found, provide ["MEMBER", "ACTIVITY"].

"MEMBER" is the name of the team member. "ACTIVITY" is the activity they performed as described in the text.

If a member is involved in multiple activities, report each activity as a separate relationship.

example: Marcus North - project conceptualization, outreach. relationships: `[["Marcus North", "project conceptualization"],["Marcus North","outreach"]]`

example: We are team Mideastern. Our team members Brad Hogg and Sophie Devine performed wet lab experiments. Andrew Symonds was responsible for collaborating with other iGEM teams. relationships: `[["Brad Hogg","wet lab experiments"],["Sophie Devine","wet lab experiments"],["Andrew Symonds","collaborating with other teams"]]`

Please make sure there is no additional text in the response other than the relationships in the prescribed format.

The text: ”

Matching The matching is performed in two steps. First, we match each individual name extracted in the previous step with the team’s roster from the competition website. All names are processed to remove accents by translating into their closest ASCII representation. The prompt instructs the GPT model to match all names extracted in the previous step with the closest match in the team’s corresponding roster list. The names that don’t have a close match are assigned “other”.

The prompt used is a modified version from the matching team names in the inter-team relationships.

“List 1 contains the names of members of a team. List 2 contains a list of names which may or may not be a member of the team. Find the closest match of each name in list 2 with list 1. If there is no possible good match - please mention ‘other’. Provide the output as a list of lists: `[["name from the list 2", "closest match from list 1"]]`”

example: List 1: "Andrew Scott', 'Thomas Alves', 'Min Jang'" List 2: "'Andrew', 'Tommy Alves', 'Charles Jang', 'Sue Perkins'" matches: `[["Andrew", "Andrew Scott"],["Tommy Alves", "Thomas Alves"],["Charles Jang", "Min Jang"], ["Sue Perkins","other"]]`

Please make sure there is no extra text or explanation before or after the formatted output. ”

Commas are replaced with spaces in member names.

In the second matching step, the description of each task is matched with a standardized list of activities that teams are expected to undertake in iGEM. This list is inspired by the CRediT Contributor Roles Taxonomy [Larivière et al., 2020], and constructed to encompass all tasks and deliverables that teams work on and are evaluated on in iGEM. The prompt provides a short description of each category and is few-shots, providing examples of matches. The set of unique tasks are batched into groups of 10 for each request, so as to provide each GPT request a small enough matching task. It is specified that some task descriptions could be matched with multiple categories and that descriptions not related to

any category should be matched to “other”.

The description for each category is constructed by summarizing the text iGEM provides to participating teams on their template wiki pages. The template pages correspond to the task-related categories described here, and contain a few paragraphs to hint at what teams’ are expected to document when they perform the said tasks. Descriptions of other categories like “Supervision”, “Training”, “Material Supply” were defined by the authors using contextual information from iGEM.

“Each phrase describes activities performed as a part of a team participating in the synthetic biology iGEM competition.

Given a list of phrases, match each phrase with the categories provided below. Some phrases can be matched to multiple categories.

Provide the output as a list of lists - `[["PHRASE", "MATCHING CATEGORY"]]`.

The possible categories, along with their description are:

Design: Conceptualising, doing background research and/or designing the teams’ project idea.

Experiments: Performed synthetic biology experiments in the wet laboratory as a part of the teams’ project.

Documentation: Creating, managing and editing the teams’ wiki website, report writing and scientific documentation

Interlab: Performed the interlab study.

Modeling: Performed mathematical models, computer simulations and/or used engineering principles to model their project.

Analysis: Performed dry lab work, data curation, data analysis and/or bio informatics as a part of the teams’ project.

Parts: Was responsible for creating, characterizing and documenting basic or composite synthetic biology parts, also called biobricks.

Safety: Was responsible for addressing safety considerations of the teams’ project.

Entrepreneurship: Was involved with building a business case and/or commercializing the teams’ project.

Hardware: Worked with or built hardware components for their teams’ project.

Software: Created computational tools and/or software as a part of their teams’ project.

Human Practices: Was responsible to understand the ethical, legal, economic and social considerations of the teams’ project. Sometimes abbreviated as ‘HP’.

Public Engagement: Established a public dialogue through outreach, educational tools and/or social media to discuss their project, science and synthetic biology with people outside the lab.

Collaboration: Was responsible for collaborating with other teams participating in the iGEM competition.

Fundraising: Was responsible in fundraising and/or finding sponsors for the teams’ project.

Creative Contributions: Making presentations, designing team logos and suits,

creating art pieces, and producing promotional materials.

Administration: Was responsible for management, organization and coordinating the laboratory and/or activities of the teams' project.

Material Supply: Providing laboratory space, equipment, supplies and providing technical materials or reagents.

Supervision: Provided advice, feedback, support, guidance, assistance or help in various aspects of a teams' project by being a Principal Investigator (PI), advisor or instructor.

Training: Conducting educational workshops, courses and/or teaching lab techniques.

Other: For descriptions that are not matched to any of the categories above.

example: "project conceptualization and working in the wetlab", "outreach"

Matching: [{"project conceptualization and working in the wetlab", "Design"}, {"project conceptualization and working in the wetlab", "Experiments"}, {"outreach", "Public Outreach"}]

example: "flux balance analysis", "Advisor for modeling tasks"

Matching: [{"flux balance analysis", "Experiments"}, {"Advisor for modeling tasks", "Supervision"}, {"Advisor for modeling tasks", "Modeling"}]

example: "assisted the team on many technical concepts", "made drawn images on the wiki", "absolutely love food and sleep"

Matching: [{"assisted the team on many technical concepts", "Supervision"}, {"made drawn images on the wiki", "Documentation"}, {"made drawn images on the wiki", "Creative Contributions"}, {"absolutely love food and sleep", "other"}]

The following are the list of phrases: "

Evaluation

Manual Curation of Text Chunks To evaluate the accuracy of the intra-team relationships curated using GPT, we construct a sample of teams whose member contribution relationships were curated manually. We select 10 teams randomly from each year between 2015 and 2018, as the attribution pages were made compulsory on the wiki only since 2015. Some teams from this sample do not describe their member attributions on their wiki and are excluded. This leaves a total of 26 teams across 4 years. The manual curation process mirrors the same steps as with GPT, wherein the description of each task is extracted as defined in the wiki text and then matched with the list of categories.

Network Construction

We construct for each team a bipartite network of team members to tasks representing the team contribution structure. Team members consist of all those present in the team's official roster, and the tasks consist in the category list from above. The bipartite networks are constructed using the `igraph` library in R.

GPT Model Parameters

Inter-Team Collaboration Networks We tested the pipeline for extraction of the collaboration networks with both the “completions” and “chatCompletion” endpoints from the OpenAI API. The “completions” endpoint works with the older generation of GPT models, currently grouped under the umbrella of GPT-3.5-turbo-instruct (completions-api). This endpoint allows for a maximum of 20 parallel requests processed per call, not exceeding token rate limits. This inbuilt parallelisation greatly optimizes runtime, albeit at the expense of poorer results. The newer GPT models, under the umbrella of GPT-3.5-turbo, GPT-4 and GPT-4-preview work with the “chatCompletion” endpoint. They follow a similar structure to using chatGPT, where each request allows the user to set a system state, and alternate between user and model responses. This however does not allow inbuilt batch processing. To improve on processing time, we use the multiprocessing library in python to create a pool of workers to execute calls in parallel.

The model request allows to tune several parameters based on the type of task. A key parameter is the temperature, which determines how deterministic or random/creative the model output is. The temperature value ranges between 0 and 2, with a lower score indicating a more deterministic response. Lower temperatures have been used in annotation tasks previously, showing higher performance, with a temperature of 0 in extracting and annotating data from medical documents [Goel et al., 2023] and 0.2 in annotating Twitter posts [Törnberg, 2023]. We observe a slightly better performance with a temperature of 0.3 for both extraction and matching tasks. This could be attributed to the heterogeneity amongst the text chunks in describing relationships. However, we acknowledge the caveat that the best temperature value could change with upgrades to and across various GPT models.

Other parameters like `max_tokens` allow setting a limit to the size of the model response. This is particularly useful in cases where the model generates a false output of maximum length permissible, increasing the running cost. This is more prevalent with the older 3.5-turbo-instruct models. In our case, we set a limit of 1,000 tokens to the output size per request. The `chatCompletion` endpoint now allows for a JSON response mode - where all output is formatted in JSON. This however still requires explicit specification in the prompt that the output be in JSON along with providing the schema. In our case, we instruct each prompt to output responses as a list of lists. This works well with both the endpoints, while requiring slightly fewer tokens in both the input and the output. The accuracy of the output being an actual list of lists varies based on the model. For example, in some cases the output provides `[["teamA", "teamB"]] [{"teamA", "teamC"}]` or `[["teamA", "teamB"]] [{"teamA", "teamC"}]` instead of a proper list of lists `[["teamA", "teamB"], [{"teamA", "teamC"}]]`. Also, in spite of explicitly instructing in the prompt to furnish the output with no additional text - the model response often includes filler sentences such as “Here are the relationships found in the text:”. To extract the relevant information from the model output, we use a python regular expression to identify the structure of a list, namely starting and ending with square braces with two elements enclosed within quotation marks and separated by a comma. This solely extracts each ‘list’ irrespective of potential errors in the output format.

For the two matching steps, the output is formatted as a list of lists, similar to the extraction step. The temperature is again set at 0.3. We split the matching tasks into batches of 10 - to enable a small modular task for the model to complete. In matching the team names, the prompt already includes the list of all teams participating that year - which averages around 300 for the more recent years. We realized that providing a larger second list of raw team names decreases the accuracy of the matching output while matching each raw name as a separate request is expensive in terms of time and cost. The reasoning is similar for matching each context with the predefined categories.

Sometimes, either with the multiprocessing approach or with runtime errors on the API side, text chunks or matching requests may not be evaluated. These exceptions are caught and retried.

The API requests using the Completions endpoint were executed between December 2023 and January 2024. The requests using the chatCompletions endpoint and the newer GPT models were executed between March and April 2024.

Team Contribution Structure With the older GPT-3.5-turbo-instruct models being archived earlier this year, we observed a drop in performance in running API requests. Hence, we extract the contribution structure with the models accessible through the chatComplete endpoint. The calls are executed in parallel using the multiprocessing library in python. These API calls were executed between March and April 2024.

The temperature for extraction is set at 0.3 and the output similarly specified to be a list of lists. For matching the members names to the teams' roster, each request handles all raw names extracted from the wiki of that team. For matching the task description to the categories, they are grouped into batches of 10, which are then processed in separate requests.

Figures

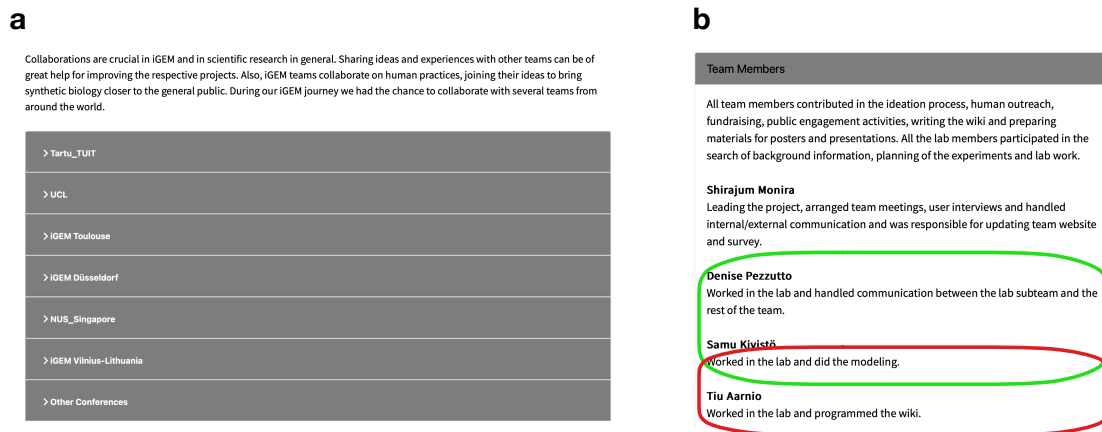


Figure 3.10: a-b. Examples of team wiki pages for collaboration and attribution. b. We showcase an example of why chunking attribution pages can result in misleading networks. The chunk highlighted in green presents a case where the inference works properly, while in the red chunk, the GPT model assigns the text included in the overlap to the next member name.

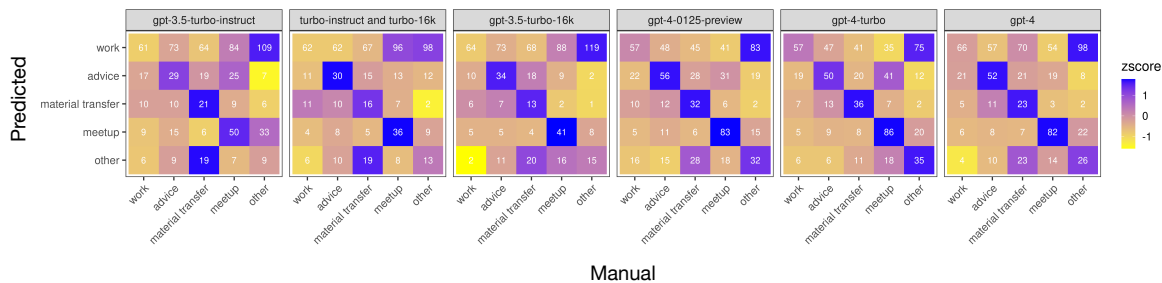


Figure 3.11: Confusion matrices like in Fig 3.7b, for the different models of Fig 3.7a.

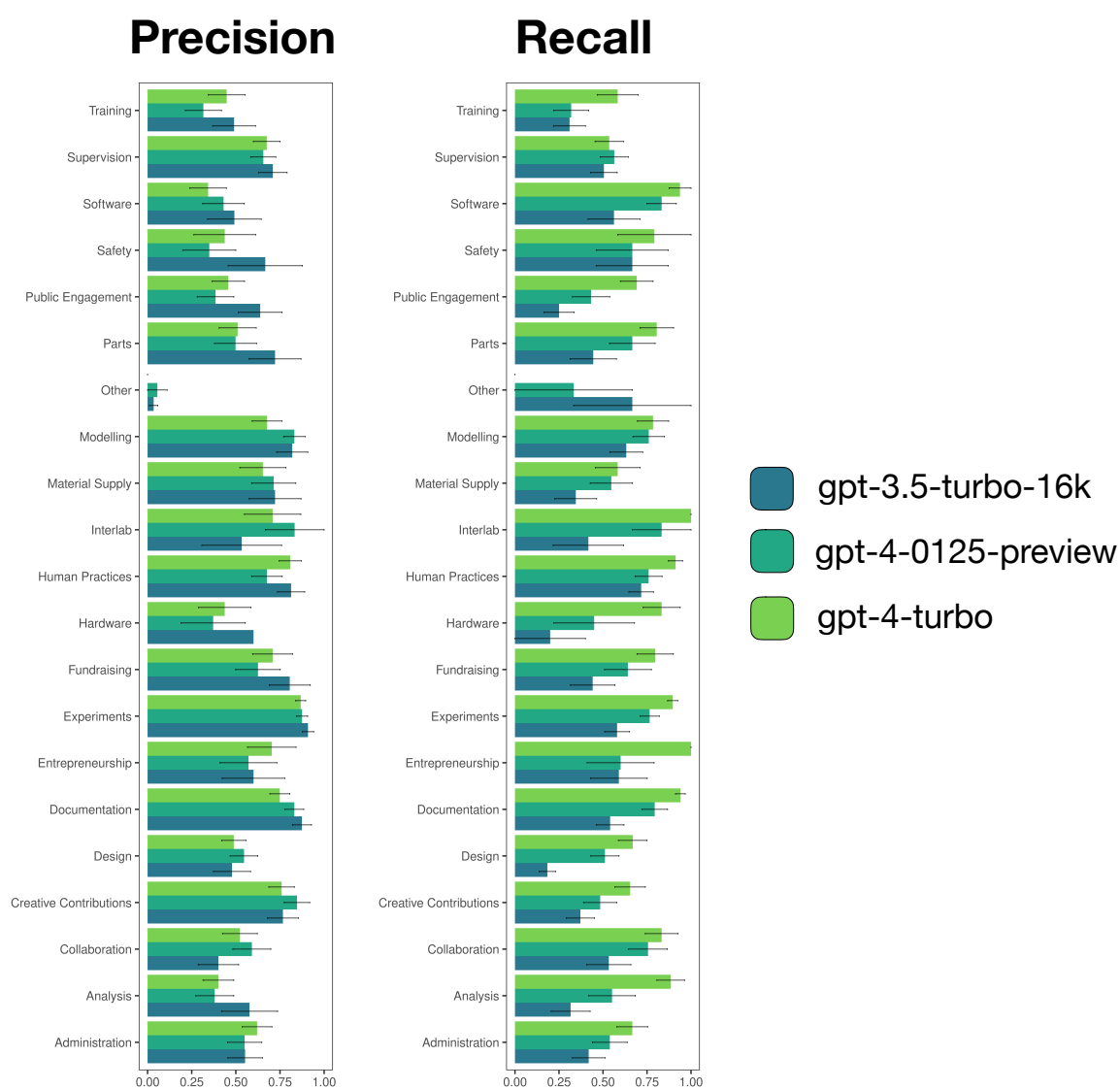


Figure 3.12: Precision and recall values, as in Fig 3.8a, but for each category of Fig 3.8b.

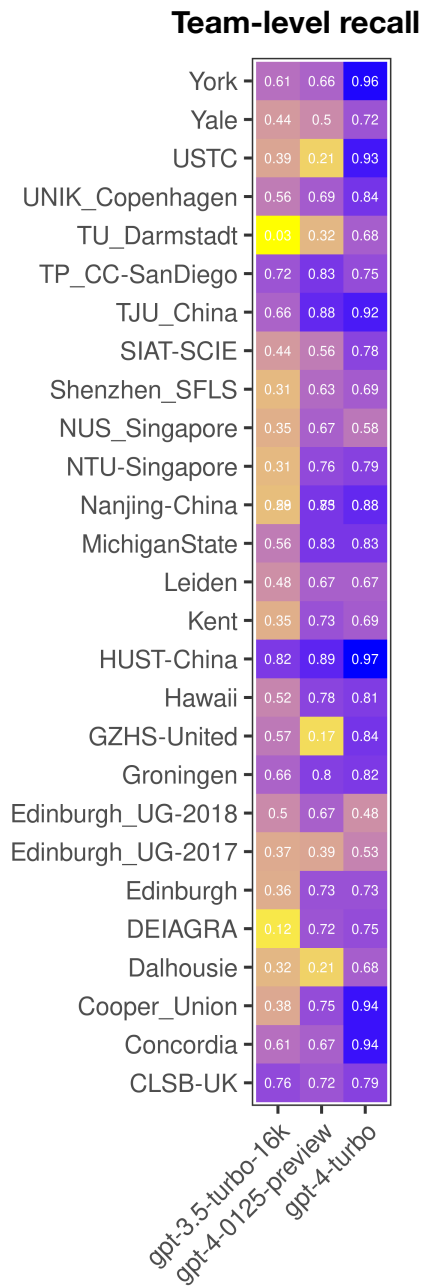


Figure 3.13: Recall values, as in Fig 3.8a, for each team.

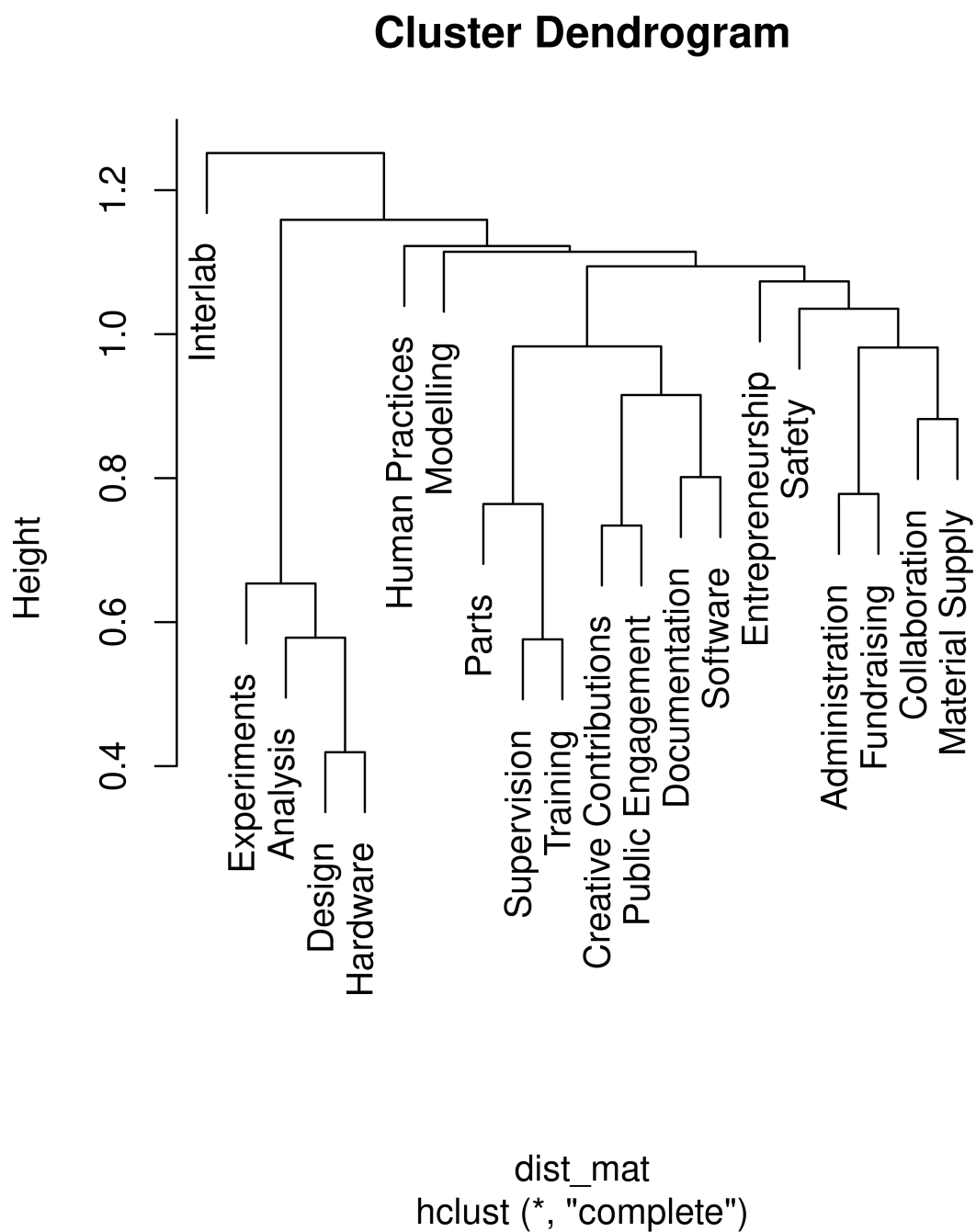


Figure 3.14: Cluster dendrogram for the confusion matrix of Fig 3.8b.

3.3 Perspectives

This chapter presents the iGEM synthetic biology competition as a testbed to study collaboration dynamics of teams in open science and innovation. I specifically focus on reconstructing collaboration networks from both qualitative (digital traces) and quantitative (self-reported text) data sources, addressing the challenge of a lack of fine-grained data to study team collaboration structures.

Recent advancements in natural language processing, especially Large Language Models (LLMs) enables building data pipelines to rapidly process and curate key information from unstructured text. We used models from OpenAI (GPT-3.5-Turbo, GPT4o) in this chapter, however, they are still proprietary. The relative ease in constructing the pipeline and high quality of the output is balanced by little knowledge about the inner workings of the model and the considerably high costs to process requests (about 1200 USD to curate inter- and intra- team collaborations in iGEM, with the total size of requests sent in order of < 1GB). In addition, working with sensitive data makes it a challenge to use these models across different case settings. Using open source models is an alternative, but they require additional effort (technical and computational resources) in fine-tuning them to function on par with the proprietary state of the art models.

The method to curate collaborations from unstructured text has potential use cases to study teams in science and innovation. The use of the CREDIT taxonomy to homogenise attribution statements in publications is a recent advancement. There is a wealth of old(er) articles whose attributions follow an unstructured format - which can be captured using LLMs. Furthermore, collaborations between fields in science are identified through citations. The contexts of each citation in a research article - such as whether the citation reinforces (or disputes) existing knowledge, whether they reuse methodology or whether they link knowledge across fields - can be reasonably identified using LLMs. This adds to the granularity in studying collaborations and knowledge integration in science.

In the next chapter, I use the networks curated from both the attribution statements and the co-edit of the team wikis to study organisation structures at the team level and within the task level.

Bibliography

Leo Blondel, Rathin Jeyaram, Abhijeet Krishna, and Marc Santolini. iGEM: a model system for team science and innovation, April 2024. URL <https://zenodo.org/records/11072818>.

Derek J. De Solla Price and Donald Beaver. Collaboration in an invisible college. *American Psychologist*, 21(11):1011–1018, 1966. ISSN 1935-990X. doi: 10.1037/h0024051. Place: US Publisher: American Psychological Association.

Sebastian Deri, Jeremie Rappaz, Luca Maria Aiello, and Daniele Quercia. Coloring in the Links: Capturing Social Ties as They are Perceived. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW):43:1–43:18, November 2018. doi: 10.1145/3274312. URL <https://doi.org/10.1145/3274312>.

Kartik Detroja, C. K. Bhensdadia, and Brijesh S. Bhatt. A survey on Relation Extraction. *Intelligent Systems with Applications*, 19:200244, September 2023. ISSN 2667-3053. doi: 10.1016/j.iswa.2023.200244. URL <https://www.sciencedirect.com/science/article/pii/S2667305323000698>.

Santo Fortunato, Carl T. Bergstrom, Katy Börner, James A. Evans, Dirk Helbing, Staša Milojević, Alexander M. Petersen, Filippo Radicchi, Roberta Sinatra, Brian Uzzi, Alessandro Vespignani, Ludo Waltman, Dashun Wang, and Albert-László Barabási. Science of science. *Science*, 359(6379):eaa0185, March 2018. doi: 10.1126/science.aao0185. URL <https://www.science.org/doi/10.1126/science.aao0185>. Publisher: American Association for the Advancement of Science.

Fabrizio Gilardi, Meysam Alizadeh, and Maël Kubli. ChatGPT Outperforms Crowd-Workers for Text-Annotation Tasks. *Proceedings of the National Academy of Sciences*, 120(30):e2305016120, July 2023. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2305016120. URL <http://arxiv.org/abs/2303.15056>. arXiv:2303.15056 [cs].

Akshay Goel, Almog Gueta, Omry Gilon, Chang Liu, Sofia Erell, Lan Huong Nguyen, Xiaohong Hao, Bolous Jaber, Shashir Reddy, Rupesh Kartha, Jean Steiner, Itay Laish, and Amir Feder. LLMs Accelerate Annotation for Medical Information Extraction. In *Proceedings of the 3rd Machine Learning for Health Symposium*, pages 82–100. PMLR, December 2023. URL <https://proceedings.mlr.press/v225/goel23a.html>. ISSN: 2640-3498.

- Kara L. Hall, Amanda L. Vogel, Grace C. Huang, Katrina J. Serrano, Elise L. Rice, Sophia P. Tsakraklides, and Stephen M. Fiore. The science of team science: A review of the empirical evidence and research gaps on collaboration in science. *American Psychologist*, 73(4):532–548, 2018. ISSN 1935-990X. doi: 10.1037/amp0000319. Place: US Publisher: American Psychological Association.
- Christoph Hoepfke and Benno I. Simmons. maxnodf: An R package for fair and fast comparisons of nestedness between networks. *Methods in Ecology and Evolution*, 12(4):580–585, 2021. ISSN 2041-210X. doi: 10.1111/2041-210X.13545. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13545>. [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.13545](https://onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.13545).
- Ashwin K. Jainarayanan, Anastasios Galanis, Athira Sreejith, Sourav Suresh, Amattullah Mustafa Nakara, Guilherme E. Kundlatsch, and Roger Rubio-Sánchez. iGEM comes of age: trends in its research output. *Nature Biotechnology*, 39(12):1599–1601, December 2021. ISSN 1546-1696. doi: 10.1038/s41587-021-01152-7. URL <https://www.nature.com/articles/s41587-021-01152-7>. Number: 12 Publisher: Nature Publishing Group.
- Andres Karjus. Machine-assisted mixed methods: augmenting humanities and social sciences with artificial intelligence. Technical Report arXiv:2309.14379, arXiv, September 2023. URL <http://arxiv.org/abs/2309.14379>. arXiv:2309.14379 [cs] type: article.
- Maximilian Klein, Thomas Maillart, and John Chuang. The Virtuous Circle of Wikipedia: Recursive Measures of Collaboration Structures. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '15, pages 1106–1115, New York, NY, USA, February 2015. Association for Computing Machinery. ISBN 978-1-4503-2922-4. doi: 10.1145/2675133.2675286. URL <https://doi.org/10.1145/2675133.2675286>.
- Vincent Larivière, David Pontille, and Cassidy R. Sugimoto. Investigating the division of scientific labor using the Contributor Roles Taxonomy (CRediT). In *Quantitative science studies*, volume 2, pages 111–128. MIT Press, 2020. doi: 10.1162/qss_a_00097. URL <https://papyrus.bib.umontreal.ca/xmlui/handle/1866/25787>. Accepted: 2021-11-03T11:49:58Z Issue: 1.
- Manuel Sebastian Mariani, Zhuo-Ming Ren, Jordi Bascompte, and Claudio Juan Tesone. Nestedness in complex networks: Observation, emergence, and implications. *Physics Reports*, 813:1–90, June 2019. ISSN 0370-1573. doi: 10.1016/j.physrep.2019.04.001. URL <https://www.sciencedirect.com/science/article/pii/S037015731930119X>.
- Camille Masselot, Rathin Jeyaram, Raphael Tackx, Jose Luis Fernandez-Marquez, François Grey, and Marc Santolini. Collaboration and Performance of Citizen Science Projects Addressing the Sustainable Development Goals. *Citizen Science: Theory and Practice*, 8(1), June 2023. ISSN 2057-4991. doi: 10.5334/cstp.565. URL <https://theoryandpractice.citizenscienceassociation.org/articles/10.5334/cstp.565>.

- Alexander Oettl. Reconceptualizing Stars: Scientist Helpfulness and Peer Performance. *Management Science*, 58(6):1122–1140, June 2012. ISSN 0025-1909. doi: 10.1287/mnsc.1110.1470. URL <https://pubsonline.informs.org/doi/abs/10.1287/mnsc.1110.1470>. Publisher: INFORMS.
- Etienne Ollion, Rubing Shen, Ana Macanovic, and Arnault Chatelain. Chatgpt for Text Annotation? Mind the Hype! *SocArXiv*. October, 4, 2023. URL <https://files.osf.io/v1/resources/x58kn/providers/osfstorage/651d60731bc8650a79f376cf?action=download&direct&version=1>.
- Étienne Ollion, Rubing Shen, Ana Macanovic, and Arnault Chatelain. The dangers of using proprietary LLMs for research. *Nature Machine Intelligence*, 6(1):4–5, January 2024. ISSN 2522-5839. doi: 10.1038/s42256-023-00783-6. URL <https://www.nature.com/articles/s42256-023-00783-6>. Number: 1 Publisher: Nature Publishing Group.
- María J. Palazzi, Jordi Cabot, Javier Luis Cánovas Izquierdo, Albert Solé-Ribalta, and Javier Borge-Holthoefer. Online division of labour: emergent structures in Open Source Software. *Scientific Reports*, 9(1):13890, December 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-50463-y. URL <http://www.nature.com/articles/s41598-019-50463-y>.
- Sachin Pawar, Girish K. Palshikar, and Pushpak Bhattacharyya. Relation Extraction : A Survey. Technical Report arXiv:1712.05191, arXiv, December 2017. URL <http://arxiv.org/abs/1712.05191>. arXiv:1712.05191 [cs] type: article.
- Oleksandra Poquet, Liubov Tupikina, and Marc Santolini. *Are Forum Networks Social Networks? A Methodological Perspective*. December 2019. doi: 10.1145/3375462.3375531.
- Oleksandra Poquet, Sven Trenholm, and Marc Santolini. Forum posts, communication patterns, and relational structures: A multi-level view of discussions in online courses. *Educational technology research and development*, June 2023. ISSN 1556-6501. doi: 10.1007/s11423-023-10262-9. URL <https://doi.org/10.1007/s11423-023-10262-9>.
- Marc Santolini, Leo Blondel, Megan J. Palmer, Robert N. Ward, Rathin Jeyaram, Kathryn R. Brink, Abhijeet Krishna, and Albert-Laszlo Barabasi. iGEM: a model system for team science and innovation, October 2023. URL <http://arxiv.org/abs/2310.19858>. arXiv:2310.19858 [physics].
- Henry Sauermann and Carolin Haeussler. Authorship and contribution disclosures. *Science Advances*, 3(11):e1700404, November 2017. ISSN 2375-2548. doi: 10.1126/sciadv.1700404.
- Varun Shenoy. GraphGPT, September 2023. URL <https://github.com/varunshenoy/GraphGPT>. original-date: 2023-01-31T18:31:32Z.

- Rion Snow, Brendan O'Connor, Daniel Jurafsky, and Andrew Ng. Cheap and Fast – But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks. In Mirella Lapata and Hwee Tou Ng, editors, *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 254–263, Honolulu, Hawaii, October 2008. Association for Computational Linguistics. URL <https://aclanthology.org/D08-1027>.
- Chuliang Song, Rudolf P. Rohr, and Serguei Saavedra. Why are some plant–pollinator networks more nested than others? *Journal of Animal Ecology*, 86(6):1417–1424, 2017. ISSN 1365-2656. doi: 10.1111/1365-2656.12749. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/1365-2656.12749>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/1365-2656.12749>.
- Dirk Stemerding. iGEM as laboratory in responsible research and innovation. *Journal of Responsible Innovation*, 2(1):140–142, January 2015. ISSN 2329-9460. doi: 10.1080/23299460.2014.1002171. URL <https://doi.org/10.1080/23299460.2014.1002171>. Publisher: Routledge _eprint: <https://doi.org/10.1080/23299460.2014.1002171>.
- Petter Törnberg. ChatGPT-4 Outperforms Experts and Crowd Workers in Annotating Political Twitter Messages with Zero-Shot Learning. April 2023. URL <https://arxiv.org/pdf/2304.06588.pdf>.
- John P. Walsh and You-Na Lee. The bureaucratization of science. *Research Policy*, 44(8):1584–1600, October 2015. ISSN 00487333. doi: 10.1016/j.respol.2015.04.010. URL <https://linkinghub.elsevier.com/retrieve/pii/S0048733315000700>.
- Fengli Xu, Lingfei Wu, and James Evans. Flat teams drive scientific innovation. *Proceedings of the National Academy of Sciences*, 119(23):e2200927119, June 2022. doi: 10.1073/pnas.2200927119. URL <https://www.pnas.org/doi/abs/10.1073/pnas.2200927119>. Publisher: Proceedings of the National Academy of Sciences.

Chapter 4

Evolving Structures of Division of Labour and Coordination in Interdisciplinary Teams

Contributions

The work presented here is from the following working paper.

JEYARAM, R.* AND SANTOLINI, M. (2024) 'EVOLVING STRUCTURES OF TASK ALLOCATION AND COORDINATION IN INTERDISCIPLINARY TEAMS: A NETWORK ANALYSIS', *WORKING PAPER*

The following section reports the contents of this article as is, reformatted to the thesis. I conceptualised the study, curated the data, performed the analysis and wrote the draft. Credit to Marc Santolini for conceptualisation, supervision and helping with the writing and review.

4.1 Introduction

Modern scientific and innovation challenges are inherently complex, requiring not only specialized knowledge but also the integration of diverse perspectives and expertise across disciplines. As the scale and scope of scientific problems have expanded, single-authored publications have become insufficient to address these challenges, prompting a shift towards more interdisciplinary, team-based research models [Wuchty et al., 2007]. The rise of collaborative research signals a fundamental transformation in how scientific knowledge is created, shared, and applied, as scientific progress becomes more reliant on collective efforts. This shift highlights the need for coordinated efforts and expertise-sharing, and it has introduced new opportunities for tackling increasingly complex questions. Importantly, this transition has not reduced research impact, signaling a transformation in how scientific knowledge is produced and disseminated [Hagstrom, 1964, Larivière et al., 2015].

At the same time, this evolving landscape of knowledge production has led to the development of new organizational frameworks that support innovation. Examples from different domains, such as the open-source software movement and citizen science, demonstrate how collective efforts can successfully produce high-quality outputs. Open-source projects, in particular, have proven that large, distributed teams can generate complex, reliable software through collective collaboration. Similarly, project-based and citizen science initiatives have been instrumental in addressing global challenges such as those outlined by the United Nations' Sustainable Development Goals [Franzoni and Sauermann, 2014, Masselot et al., 2023, 2022].

However, managing such complexity presents its own set of challenges. Effective team functioning relies on how well tasks are divided among members, how these members work together to accomplish each task and how seamlessly different member groups coordinate to integrate their respective outputs. Complex tasks are often composed of interconnected sub-tasks, requiring the formation of specialized sub-groups. While these sub-groups can improve efficiency and focus, they introduce the challenge of information integration, as members must align their contributions to produce a unified output. This balance between specialization and integration is critical, as it affects both team dynamics and the overall performance of knowledge production efforts [Lawrence and Lorsch, 1967].

Scientists have examined team specialization and integration through the framework of task allocation structures, that represents which members are responsible for specific tasks. In bibliometric studies, contribution statements from publications are used to identify the roles of authors, typically utilizing standardized attribution taxonomies like CREDIT [Larivière et al., 2020, Haeussler and Sauermann, 2020]. Similarly, digital traces documenting individual contributions are analyzed to understand the contribution structures of open source software projects [Klug and Bagrow, 2016]. Alternatively, this structure in teams has also been recovered qualitatively, through surveys [Walsh and Lee, 2015, Masselot et al., 2023].

The contribution structures represent the division of labour of how tasks are distributed across team members. They inform on how specialist members are grouped and assigned to core tasks while generalists integrate across these task groups. However, beyond the static attributions, they do not reflect how coordination occurs within each task. The coordination structure for each task differs based on the technical and social costs required to accomplish it. Brainstorming in teams integrates several social, compositional and cognitive factors [Paulus and Kenworthy, 2019] and often requires specialised practices to capture its conceptual and iterative nature. Experimental tasks however are often performed methodologically, with member roles and its interdependencies defined in advance. However, identifying these interdependencies requires in-depth knowledge on a case by case basis. Writing, in an interesting case-study of coordination and integration in performing a key task in contemporary science. Latour [Latour et al.] emphasises the importance of *Inscription devices* in science, which transform observable phenomena in analysable and communicable knowledge. Writing a manuscript requires integrating contributions across team members and coordinating with the designated authors in accomplishing the construction of a coherent text. Although attribution statements only capture the members who undertook this responsibility, modern collaborative writing tools can shed light on the dynamic structures of coordination [Erkens et al., 2005, Sardo et al., 2023].

Bipartite networks are used to represent the the contribution and coordination structures in a team. The networks are constructed with two categories of vertices - team members and the tasks they work on, with no edges between vertices of the same type. This approach enables the analysis of structural properties at the individual, group and team levels. Simple measures, such as the degree, task spread of a user indicates the level of individual specialisation while correlations between user-task incidences reveal the interdependencies of user activities within the task structure at the pairwise level [Haeussler and Sauermann, 2020]. However, these measures fail to capture mesoscale structures - such as sub-groups or modules and the interconnections between them. To identify these mesoscale structures - bipartite networks are projected onto their user or task categories, where users are connected by an edge if they work on the same task. These projections have been used to identify user and task modules [Xu et al., 2022], but they are highly sensitive to the projection method used [Coscia and Rossi, 2019] and often require edge filtering in dense networks. Projecting the bipartite division of labour networks on the tasks, where tasks are linked if they are acted upon by a common user, gives the overall structure of task interdependencies and complexity.

Exploring the internal division of labor and task complexity structures are essential to understand how teams coordinate to produce high-quality output. There is growing evidence that these structures influence team performance, with flatter team organisation often producing more innovative results [Xu et al., 2022], while smaller, more effective work groups have been linked to higher efficiency in task completion [Klug and Bagrow, 2016]. An important question remains: how do these structures evolve over time, particularly as teams gain experience and adjust their organizational patterns. While organizations and teams learn by producing, retaining, and transferring knowledge [Argote, 2013], there is a notable gap in the literature concerning longitudinal studies that track the evolution of task and division of labor structures in teams.

In this working paper, we leverage a longitudinal, large-scale dataset of synthetic biology teams participating in the iGEM competition to address these gaps by 1) developing granular, network-based indicators for division of labor structures in teams, 2) exploring the relationship between division of labour at the global and task levels with team performance, and 3) assessing how these organizational structures evolve with experience.

The paper is organised into 4 further sections. In the *Background* section, We describe the motivations for using new measures of division of labour, strongly influenced from bipartite network literature. We also substantiate the key hypothesis we test with its motivations. In the *Methods* section, we expand more on the dataset and tools used for its processing and analysis. We present the *Results* and provide a *Discussion* of the findings.

4.2 Background

4.2.1 Division of Labour in Teams

Mintzberg [noa, a] defines the structure of an organisation to be the ways in which they divide labour into tasks and how they achieve coordination amongst these tasks. In scientific teams, the internal division of labour describes how different tasks that are expected to be performed are allocated amongst the team members. The CREDIT taxonomy is a widely used for describing task classes in scientific research [noa, b]. These categories range from conceptual tasks (like ideation, study design), empirical tasks (analysis, data curation) to other allied tasks like administration and supervision [Haeussler and Sauermann, 2020, Larivière et al., 2020]. This taxonomy helps clarify the distinct yet interdependent roles that contributors play in producing a peer-reviewed scientific publication.

Contribution statements highlight how these key tasks, which are essential for realizing an academic manuscript, are distributed amongst the authors. This division of labor naturally varies across fields, publications and teams. Last authors (who in most cases are senior amongst the list of authors) are responsible for supervision and administrative tasks, while being involved in different aspects of the study in a general capacity. Smaller teams have multiple members designing the methodology, performing the investigation, analysing and presenting the results. Larger teams however, have a core set of authors performing conceptual tasks, but have subgroups focusing on design and investigation while others performing analysis and building software tools. This division of labour relies on specialists (or technicians) performing niche tasks, while generalists supervise and coordinate across different activities. [Larivière et al., 2020, Xu et al., 2022, De Solla Price and Beaver, 1966]. In parallel, open source software projects also have a similar organisation. There is often a rapid turnover in the contributors for an Open Source Software (OSS) project, although, a core set of contributors offer overall support while the project is split into overlapping

sub-groups led by specialists [Klug and Bagrow, 2016, Palazzi et al., 2019a].

This balance between specialisation and coordination is a key challenge in efficient division of labour in teams. Specialisation increases the technical propensity of members accomplishing tasks, but also increases the costs for coordination - especially when the tasks are strongly interdependent. These coordination costs are reduced by generalists, but they have the complex challenge of integrating information at the task level and coordinating across team levels. Teams seek this balance to maximise their productivity and performance payoffs [Kretschmer and Puranam, 2008].

4.2.2 Measuring Division of Labour

Bipartite networks have been used to model division of labour interactions in teams and organisations. Contribution statements from bibliometric data are represented as a bipartite network with edges between authors and tasks. Bipartite networks are a regular fixture in ecological studies to represent plant-pollinator interactions and other food webs. Here division of labour, more specifically, the allocation of mutualistic or antagonistic relationships, speaks to the global stability of the natural system [Pasquaretta and Jeanson, 2018]. Contributions to open source software projects and answering on forums are also modelled as bipartite interactions between contributor-code and user-post respectively [Palazzi et al., 2019a, Poquet et al., 2023]. These networks help identify activity profiles of users and their relative position within the community as a whole.

Network measures can be used to capture properties of the networks of division of labour at different scales. At the individual scale, the ratio of generalists versus specialists in performing different tasks and the propensity in which tasks are performed by a few versus many users are captured through aggregate degrees [Haeussler and Sauermann, 2020] or through information theoretic measures inspired by diversity and entropy [Blüthgen et al., 2006]. Interdependencies (or similarity in activity) between users (and tasks) are measured through pairwise correlations between rows and columns of the bipartite incidence matrices.

At the network scale, two important properties that provide insights into the organisation of division of labour structures are nestedness and modularity. Nestedness is a characteristic property of mutualistic ecological networks which highlights a *nested* organisation between generalists and specialists - wherein specialists significantly overlap their incidences with vertices of a higher degree, cascading until the generalists of the system. Nestedness is also a characteristic property of socio-economic networks, such as trade and inter-organisational communication [Saavedra et al., 2009, Mariani et al., 2019]. The nested organisation of generalists and specialists in networks of division of labour speak to the overall structure of coordination. High nestedness suggests an organisation with hierarchy - represented by the heterogeneity in interactions.

Modularity in networks measures the propensity of vertices grouped being into clusters (or modules) where intra-modular interactions are more prevalent than inter-modular. Modularity in networks of division of labour, focuses on identifying sub-groups in a team, which focus on performing specialised tasks, and how these sub-groups are linked. With a great focus on specialisation in teams to maximise productivity and output while tackling complex problems [Becker and Murphy, 1992, Jones, 2009], modularity is an intuitive measure to quantify specialisation at the network scale.

While nested and modular structures can coexist, they have an antagonistic relationship: higher modularity often reduces nestedness and vice versa [Palazzi et al., 2018, Fortuna et al., 2010]. Despite the advantages to a modular organisation, nested structures have been studied in their role towards stability. Online social networks adapt from a more modular to a more nested organisation over time [Borge-Holthoefer et al., 2017], which is attributed to a development of consensus of opinion against that of competition over time - mimicking the stability of mutualistic ecological systems.

However, this interplay between nested and modular organisation is mediated by changes at the mesoscale level - by the presence of repetitive motifs and the structure within each module. This highlights the organisation of members performing the same task, whose coordination structures differ when compared to the whole team. Open source projects self-organise into a modular organisation. However, the structure within each module is nested - indicating compartmentalisation in addition to emerging hierarchies between generalists and specialists in the team [Palazzi et al., 2019a,b].

Leveraging measures of division of labour at the individual (vertex), meso- (intra-modular) and network scales provide a comprehensive insight into team organisation to tackle complex and interconnected challenges. While aggregate measures at the vertex level like the degree or pairwise correlations provide proxies for specialisation, they can indicate contrasting global structural organisation when computing nestedness or modularity 4.1. This motivates the need for a set of composite measures to describe division of labour network structures.

4.2.3 Recovering the Task Structure

While the bipartite networks of task allocation between users highlight the division of labour in teams, they also provide indicators for the overall task interdependencies and structure. Melvin Conway [Conway, 1968] hypothesised that the interactions between people and sub-groups in a team are constrained by the inter-dependency structure across the subsystems that these groups work on. The interdependence between tasks can be recovered using the mutual incidence between users and tasks as a proxy. This helps identify tasks that are statistically more likely to be performed by a common user, highlighting their

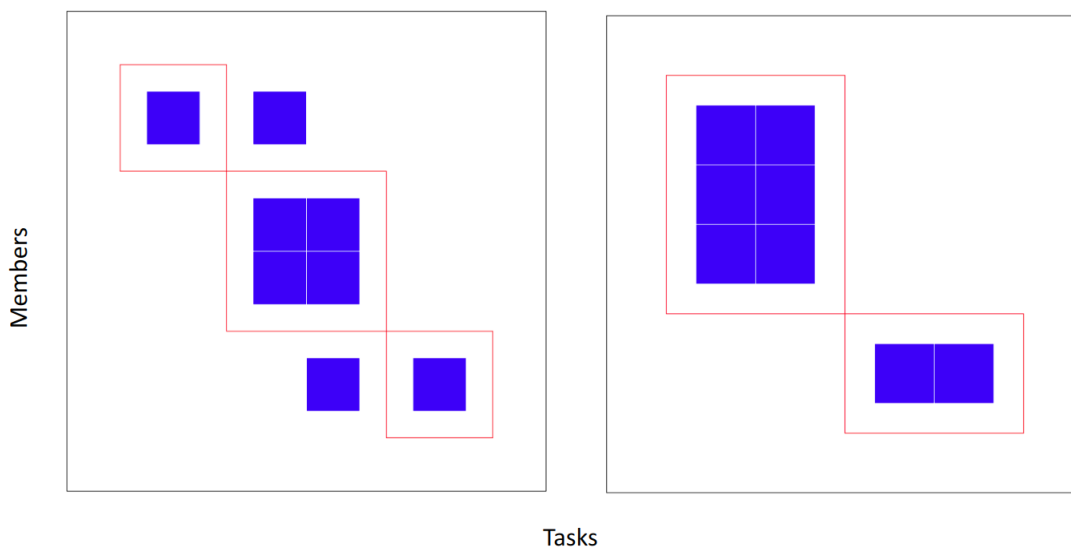


Figure 4.1: Example of two bipartite networks, represented by incidence matrices, where the aggregate measures (degree correlations between users and tasks) are the same. But the matrix on the left has intermediate modularity while the matrix on the right has high modularity.

dependence to each other in addition to helping identify clusters of broad task roles and hierarchies [Xu et al., 2022].

This interdependence between tasks vary across fields. Taking the CREDIT taxonomy [noa, b] example, different fields emphasise on specific categories. Software is not always a significant component of biological research, while mathematical studies rarely focus on data curation and analysis. These field level differences in the task structure conversely constrain the division of labour structures - biological research teams are traditionally smaller and more specialised while high-energy physics is more interdisciplinary and distributed amongst large teams [Cetina, 1999]. With the task structure constraining team organisation and division of labour, identifying these interdependencies offer insights for balancing collaboration and coordination costs in complex research environments

4.2.4 Team Adaptation and Learning

Organisational learning has been studied extensively in the last decades, specifically looking at how they increase their productivity and output with experience. Teams and organisations learn by creating, retaining and disseminating knowledge [Argote, 2013]. Internally, this learning is reflected through individual members by them specialising in performing different tasks, and through *routines* embedded within the teams' organisational processes [Argote et al., 2021]. These routines are repetitive interaction and coordination patterns that are enabled by specific organisational structures, and are key in teams accomplish-

ing complex tasks consistently [Becker, 2020, Nelson and Winter, 2004, Kogut and Zander, 1992]. With dynamically evolving task ecosystems, developing routines ensures teams' capacity to navigate uncertainty and change [Wiese and Burke, 2019].

Linking routines in teams with their evolving organisational structure is not straightforward. Routines can be reflected through social and tacit interaction patterns that are not always captured through the teams' division of labour. However, the significance of specific interaction patterns and their role in supporting learning cannot be understated. Clustering of members to tasks impedes the spread of best practices across subgroups in a team, but facilitates higher specialisation [Lazer and Friedman, 2007, Puranam and Maciejovsky, 2020]. Hierarchical structures adapt towards optimal organisational and coordination patterns over time, but at the detriment of exploring novel solutions [Koçak et al., 2023], which may penalise their efforts in competitive task environments.

Studying the evolution of team and task level organisational structures, in response to experience and task complexities offer significant insights into team learning processes supports the design of interventions to optimise team productivity. However, this remains a complex challenge, mainly limited by a lack of fine-grained longitudinal data.

4.2.5 Objectives

In this working paper, we study the specialisation and coordination structures in scientific teams from their networks of division of labour at the project and task levels, and highlight how they underlie team performance and learning.

Interdisciplinary teams have a higher division of labour [Haeussler and Sauermaun, 2020]. This is in response to having multiple interconnected sub-tasks which requires diversity in expertise. Teams aim to maximise their output by specialising on sub-tasks and having optimal coordination structures. This results in a modular division of labour organisation, with specialists in each module. However, the generalist team members coordinate across modules, reflecting on the task complexities and interdependences.

Hypothesis 1: The division of labour in scientific teams, working in interconnected task environments, have a highly modular organisation. These specialised modules are coordinated by generalists, who overlap across modules, giving rise to an intermediate nested structure at the global level.

Performing each task requires specialist team members to efficiently recombine their knowledge, background and experiences in producing an innovative output. Technical tasks, such as performing experiments or data analysis involves specialist members working as a flat sub-group, with generalists then coordinating with members from other interconnected tasks. Capturing the granular interactions of within each task group is complex,

requiring specific domain knowledge on a case by case basis. However, tasks like documentation and writing, which are an essential component of the scientific research process [Latour and Woolgar, 1986], can be tracked extensively by using traces from digital collaborative writing tools. At this task level, specialised contributions from team members are integrated by generalists, presenting organisational structure with high nestedness

Hypothesis 2: The task level organisation in teams present structures with high nestedness and intermediate modularity.

Division of labour structures have been studied with their link to performance - with flat teams in science producing more innovative research [Xu et al., 2022] and an inverse-U relationship between network connectivity and performance [Lazer and Friedman, 2007]. Successful teams organise their structure efficiently, balancing the interplay between specialisation and integration. Modularity and nested organisation are indicators of this interplay at the network level and capture structural properties beyond aggregated measures at the vertex.

Hypothesis 3: Nestedness and modularity are stronger predictors of team performance and productivity compared to aggregate network measures.

Teams retain members across successive participation. This allows teams to adapt and change their network structure over time. Teams select into similar task interdependence structures over time compared to other successful teams.

Hypothesis 4: Teams task allocation structures evolve into optimal network structures over time, indicating teams search for optimal solutions, which are mediated by factors like size and prior experience.

4.3 Data

Since its inception in 2003, the iGEM (International Genetically Engineered Machines) competition incentivizes student teams to tackle real-world problems and explore the space of synthetic biology by designing standardised DNA elements, also known as biobricks. Over 4000 university, high school and community lab teams from 40+ countries have participated in this challenge-based ecosystem for innovation, producing several interdisciplinary publications and projects scaled up as startups [Jainarayanan et al., 2021]. But beyond co-creating innovative synthetic biology projects, teams are expected to address the safety, ethical considerations of their work, as well as to undertake outreach initiatives and dialogue with multiple stakeholders who would be influenced by their work. Each year, iGEM culminates with the grand jamboree, where teams present their project and are evaluated by juries on the different facets of the competition. Teams are awarded medals (gold, silver, bronze) based on them completing the checklist that iGEM participants are required to fulfill - the more items completed, the higher the medal. Exceptional teams in different

competition tracks are also awarded prizes based on the jury evaluations.

One of the key evaluation criterion is that teams extensively document their project, its outcomes and their progress on an online wiki style website. This website, referred to henceforth as the *wiki*, presents as a rare intersection between a scientific article and a blog - detailing key technical aspects of a team project in addition to elaborating on their outreach initiatives, their collaborative activities and member attributions. This richness of data from team wikis presents a unique opportunity to study team-based collaborative science and obtain insights on performance and impact [Santolini et al., 2023].

Wikis follow a general template - designed by iGEM at the beginning of the competition year to highlight the various project aspects that teams are expected to report on. The wiki template contains pages, whose titles correspond to these different aspects - such as *Experiments*, *Parts*, *Human Practices*. But teams can choose to report in a non-standardised manner, as long as they fulfill the evaluation criterion for reporting their complete project. As an example, teams can document their experiments in a single *Project* page or in several pages, each highlighting a significant sub-part. Teams wikis also include an *Attribution* page which reports who did what in the team - albeit heterogeneously. This makes manually annotating and curating information about team organisation from the wiki a challenging task.

Beyond the attribution pages, the team wikis serve as a report of their iGEM project, with regular lab notebook entries, specifics of their synthetic biology project and details of their outreach initiatives. Writing this wiki is a collaborative task - with teams designating editors who take point on populating the different pages. The digital traces of who edited which page offer insights on how members of an iGEM team integrate information across these different project aspects, in addition to the division of labour inferred from the attributions.

The extensive iGEM dataset [Blondel et al., 2024], reports team meta information - such as their roster, the track they participated in, their performance (medal, awards), the wiki full text and digital traces of how team members co-edited their wiki. The wiki traces and text are collected for teams participating in iGEM from 2008-2018. Reporting the attributions of team members and clearly highlighting the work done by them against the mentors and collaborators became a bronze medal criterion in 2014. Teams after 2014 are expected to report the attributions on the wiki, while it was optional for teams before.

We use iGEM teams as a testbed for collaboration and innovation in complex task environments, studying their division of labour and wiki contribution structures.



Figure 4.2: a) Constructing the bipartite network of co-edits (members to pages) of the team wikis b) Pipeline to construct the bipartite division of labour network from unstructured wiki attribution pages. Large language models from the GPT family are used to curate the interactions between users and the tasks they performed, matching users to their identifiers and description of tasks to a standardised list of tasks performed by iGEM teams. [Jeyaram et al., 2023] describes the pipeline and methods to evaluate its performance in more detail. c) Example of the task structure of a team, identified by projecting the bipartite division of labour network on the task layer.

4.4 Methods

4.4.1 Team Division of Labour from Attribution Statements

Attribution statements from the team wiki report on the members of the team and what tasks they worked on during the project. With these statements reported as free text, we use Large Language Models (LLMs) to design a pipeline to extract the member-task relationships with high confidence [Jeyaram et al., 2023]. Figure 4.1.b describes the pipeline to perform this process - which involves first extracting all person-activity relationships and then matching each person to their iGEM member ID and each activity to a list of standard tasks. These tasks are inspired from the CREDIT author-contribution roles frequently used in bibliometric studies [Larivière et al., 2020]. Further details on the standard tasks and their descriptions are documented in the supplementary information.

The resulting member-task bipartite network are consistent across teams and accurately represent division of labour structures from the self reported attribution statements. As a simple sanity check to ensure the inferred networks are accurate - we compare the list of team members reported in the attribution statements against the team roster. Teams where atleast 70% of the roster are recaptured in the attributions are of relatively high confidence.

The attribution statements were made a compulsory part of iGEM team wikis in 2014 3.2b and were designated as a criterion to win a bronze medal (and thus required to win silver and gold). Teams before 2014 still report their attributions, with the GPT pipeline recovering 40% of teams having mentioned to atleast 3 members and 3 tasks in their attribution statements. This ratio increases to 70% for teams after 2014. We specify the filters used to in each analysis explicitly.

4.4.2 Task Structure and Similarity in Teams

The task structure in iGEM can be recovered from the member attribution statements. Projecting the bipartite member-task networks onto the task layer creates a network of tasks, where two tasks are linked if they are performed by a common user. The projected network provides a proxy for interdependences between tasks (if the same user performs two distinct tasks, it is likely that these tasks have an inherent associations).

Weights of the task projected bipartite network highlight key information about the strength of links between tasks. The traditional method of weighting projections uses the number of mutual links to the same actor in the bipartite network. If task A and task B are performed by three users - 1,2 and 3, the weight of the interaction between tasks A and B in the projected network is 3. This weight is then normalised at the team level to enable comparison across different task structures. A larger weight signifies a stronger interdependence between tasks. An example is presented in figure 4.2c.

The edges in a bipartite projection can also be weighted by other methods, contrasting weights based on the inverse degrees of bipartite vertices [Newman, 2001] or by computing similarity between two projected vertices by the likelihood that random walkers pass through sequentially [Yildirim and Coscia, 2014, Coscia and Rossi, 2019]. Bipartite networks often result in a projection with high density and varying edge weights which often do not identify key structural properties and are hard to analyse. Backboning [Neal, 2022], is a method to extract the significant edges (or the backbone) of a projected bipartite network. In this working paper, we use the classic bipartite projection to weight interactions between tasks. Studying the other weighting regimes, the backbone and contrasting them against the task structure are upcoming directions.

The bipartite network of each team provides a different task interdependence structure. This structure changes based on the team project tracks - software teams use different toolkits and perform different significant tasks compared to an experimental team. To identify the task environment for each track, we use the projected task networks of award-winning teams. Multiple teams win awards every year, indicating their exceptional iGEM performance in their track. We use the task projections of these successful teams as a reference to compare the projections of other teams in each track.

We define a distance measure - which compares the task projection networks of each team with that of the successful teams in the given iGEM competition year. The distance between two projected networks is computed using:

- **Euclidean Distance:** Using the vectors of vertex strengths in the task projections. A lower distance indicates closer similarity between the two vectors. However, this measure hides the interactions between vertices in computing the distance.
- **Jensen-Shannon Divergence:** Uses information theory measures to compare the distance between two networks from their density matrices. Method detailed in [De Domenico et al., 2015]. The lower the divergence, the more similar the two networks.

The average distance of a team's task structure is computed from its distances to the best teams from the same track. A lower average distance indicates a structural interdependence organisation closer to successful teams.

4.4.3 Organisation from the Wiki Co-edit Structure

Wikis are collaboratively edited by a subset of members from the corresponding iGEM team. The typical wiki contains pages where teams document various components of their project, like experiments, outreach activities, synthetic biology parts and presenting their team. The bipartite network between team members and the wiki pages follows from the

digital traces of the wiki, which identifies each member by their identifiers, the page they edited, the timestamp and the size of each edit action.

Teams however have different ways to construct their wiki notebook. Some teams have a small set of core editors, who take point on designing and constructing the entire wiki. Or they have multiple editors documenting pages corresponding to the different iGEM tasks and activities they performed. We use cutoffs - ratio between team size and the number of editors - highlight the wiki editing structures. We select teams where atleast 20% of the team have edited the wiki.

4.4.4 Measuring Division of Labour Structures

Bipartite network measures of division of labour capture the level of generalisation or specialisation in different scales and offer trends about the interdependencies between users and tasks. We focus on modularity and nestedness as two key measures in this study.

Modularity

Modularity measures the organisation of nodes into groups in way that maximises interactions within a group compared to across. In a bipartite network, the modularity can be evaluated in two ways. The bipartite modularity identifies modules of users and tasks which have more edges between them. We compute the bipartite modularity using the *compute-Modules* function from the *bipartite* library in R. This outputs a module object highlighting the module of each node and the modularity of the network using the algorithm from [Beckett, 2016].

Alternatively, a bipartite network can be projected onto the user or the task layer - where an edge between users (or tasks) indicates them co-working on a common task (or tasks having a common user doing them). Modularity can be computed on these projected networks, identifying clusters of users (or tasks). The classic projection method involves weighting each edge in the projected network to account for the number of tasks co-worked by the two users (in case of projection on the users). We use this method to create projected networks on the user level and the task level. The modularity is computed using a greedy modularity maximisation method such as *Cluster Louvain*.

Nestedness

NODF is among the widely used measure to quantify nestedness in a bipartite system. Introduced in [Almeida-Neto et al., 2008], the *NODF* measure calculates the extent to which rows with lower degree overlap with the rows of a higher degree. A perfect nested structure has high Node Overlap and an ordering with Decreasing Fill (*NODF*). The measure ranges

between 0 and 100, with higher the score, the more nested the bipartite system. The measure, despite being among the most widely used, has its share of criticisms in quantifying nestedness. NODF has shows to be positively correlated with matrix density and negatively with size [Payrató-Borràs et al., 2020].

Other measures for nestedness capture the deviation of the observed incidence matrix from a perfectly nested structure - usually falling under the umbrella of measures denoted as *nested temperature*. These measures however suffer from a much stronger bias to size and density [Payrató Borràs et al., 2020].

We use the *nest.smdm* function from the *Bipartite* library in R to make the computation. The function allows computing of the NODF measure for a given bipartite incidence matrix - but also at the modular level. The outputs of interest include the NODF measure for the whole matrix and the NODF measure for nodes within the same modules in the network (averaged over all modules). The modular organisation of the bipartite network is computed using the bipartite modularity function *computeModules*, whose output is fed into the nestedness computation to identify global and module specific nestedness measures.

Alternative Measures

Beyond indicators of specialisation and co-organisation into modules, measures like that capture the representation of bipartite motifs offers information about meso-scale collaborative structures [Paranjape et al., 2017]. This is a working extension to the current draft.

4.4.5 Outcomes

Team outcomes in iGEM are measured through their evaluation in the final jamboree. Teams are awarded medals (gold, silver, bronze or none) based on them achieving specific targets on the iGEM checklist. These targets include designing new biobricks, reusing existing biobricks, collaborating with other teams, documenting their progress on the wiki and so on. Teams are also awarded with prizes, indicating exceptional projects for each track.

Focusing on writing as a specific task that teams work on, we use the size of the wiki (in bytes) as a proxy for outcome. Larger wikis are a predictor of team performance (medals and awards) [Santolini et al., 2023] and explore how different network structures associate with the wiki edit size as an outcome.

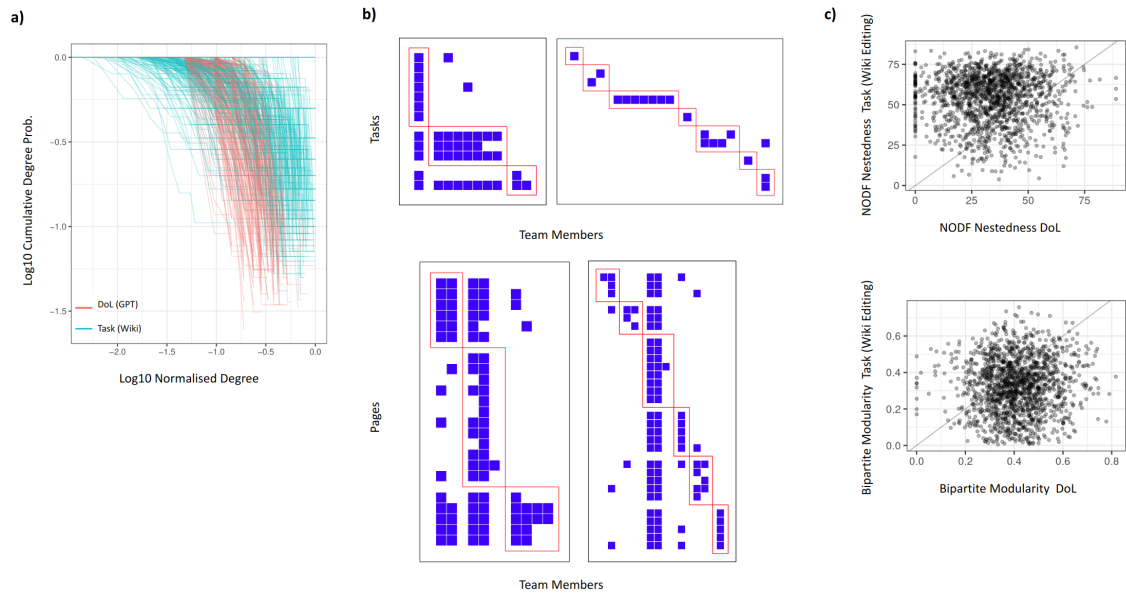


Figure 4.3: Comparing Bipartite networks of division of labour (curated from attribution statements using GPT) and organisation at the task level (co-editing the wiki) a) Cumulative degree distribution of users in both division of labour and co-edit networks in iGEM 2017. b) Example team incidence matrices. (top left) Team with high nestedness in their division of labour network. (top right) Team with high modular structure in their division of labour network. (bottom left) Team with a nested wiki co-edit structure. (bottom right) Team with a modular wiki co-edit structure. Matrices plotted using the *plotModuleWeb* function in the *noa* [2024] library in R. c) Scatter plot of nestedness (top) and modularity (bottom) comparing teams' division of labour network to their respective wiki co-edit network.

4.5 Results

4.5.1 Nestedness and Modularity in Networks of Division of Labour

Figure 4.3 compares the bipartite division of labour networks and the co-edit structure of the team wiki. The degree distribution, shown in figure 4.3a, indicates a long tailed structure in the co-edit network, indicating the presence of generalist users who take up a majority of the core responsibility in completing the wiki writing. The more flatter shape of the division of labour networks indicate a relative spread of different tasks across team members. The organisation of these structures are reflected in figure 4.3c, with division of labour networks having a less nested and more modular than the co-edit structure in wiki editing.

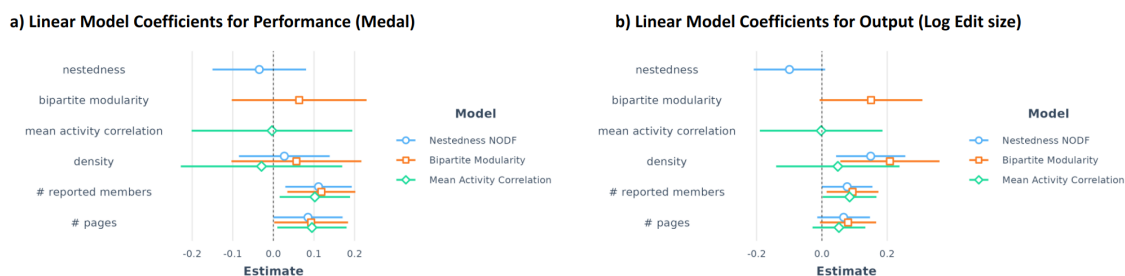


Figure 4.4: a) Linear model coefficients comparing the effects of nestedness, bipartite modularity and mean correlation between activity-user incidence vectors in predicting the medal achieved. The subset teams participated in iGEM since 2014 and have atleast 70% of the team roster captured from the attribution statements. Teams also report atleast 10 users and 10 tasks - filtering for large teams b) Linear model coefficients comparing the effects of nestedness, bipartite modularity and mean correlation between activity-user incidence vectors in predicting log edit size of the team wiki. The subset teams participated in iGEM since 2014 and have atleast 70% of the team roster captured from the attribution statements. Teams also report atleast 10 users and 10 tasks - filtering for large teams.

4.5.2 Division of Labour and Performance

Figure 4.4 estimates how nested and modular organisation in the division of labour networks associate with performance measures - a) the medal achieved by the team and b) the size of the team wiki. Teams are filtered to ensure they have atleast 70% of the roster recovered through GPT curation and mention atleast 10 users and tasks. Teams after 2014 are presented here ($n = 816$ teams). The linear models have the density, number of users, number of tasks and the competition year as controls.

Figure 4.4a compares the linear model coefficients predicting the medal achieved by iGEM teams. While neither nestedness nor modularity are significant in predicting the medal achieved, their effect is visible when looking at the wiki size as an outcome (Figure 4.4b). The edit size is an indicator of productivity of the team and positively correlates with the medal achieved (Correlation coefficient 0.38). Modularity is positively associated with the wiki edit size while nestedness is negatively associated, highlighting the importance of forming specialised subgroups tackling the distinct, yet interconnected tasks during iGEM.

For both outcomes, the aggregate measure comparing the similarity between tasks using correlations between user incidences (higher correlation indicating tasks are performed by the same set of users) is not a significant predictor for productivity or output.

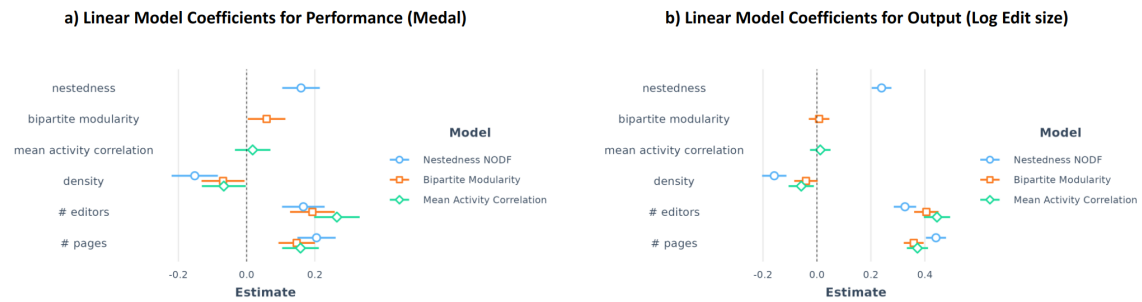


Figure 4.5: a) Linear model coefficients comparing the effects of nestedness, bipartite modularity and mean correlation between activity-user incidence vectors in predicting the medal achieved. At least 20% of the team roster are editors. Teams also report at least 10 editors and 10 tasks - filtering for large teams. b) Linear model coefficients comparing the effects of nestedness, bipartite modularity and mean correlation between activity-user incidence vectors in predicting log edit size of the team wiki. At least 20% of the team roster are editors of the wiki. Teams also report atleast 10 editors and 10 tasks - filtering for large teams

4.5.3 Co-edit Structure and Performance

Figure 4.5 estimates the association between nested and modular organisation within a task (here that of editing the wiki) and performance. Teams are filtered to ensure they have atleast 20% of the roster editing the wiki and have atleast 10 editors and pages ($n = 612$ teams). The linear models have the density, number of users, number of tasks and the competition year as controls.

Nestedness is strongly associated with a higher medal achieved as well as in creating a larger wiki (log edit size). The nested organisation highlights a set of core editors who are responsible to integrate information across the other *specialised* editors in the task. With the selected teams here having atleast 10 editors, a modular structure is also positively associated to the medal achieved, where each module takes point on editing specific wiki pages (corresponding to technical tasks), who are then coordinated across modules and tied together by the core editors. These core editors, in addition to integrating input from different members, also perform specialised auxiliary tasks, such as website design - further highlighting the significance of a nested organisation within the writing task.

In both figure 4.5 a and b., the aggregate measure comparing the similarity between tasks using correlations is not significant in predicting performance.

4.5.4 Evolution of Task Structures

Figure 4.6 shows how the task interdependence structure evolves with multiple participation. The task structure of teams gets closer to that of successful teams in each iGEM track. This decrease is not driven by teams that participated a few times and not again (highlighted by each trace - indicating the maximum participation of the team - distance

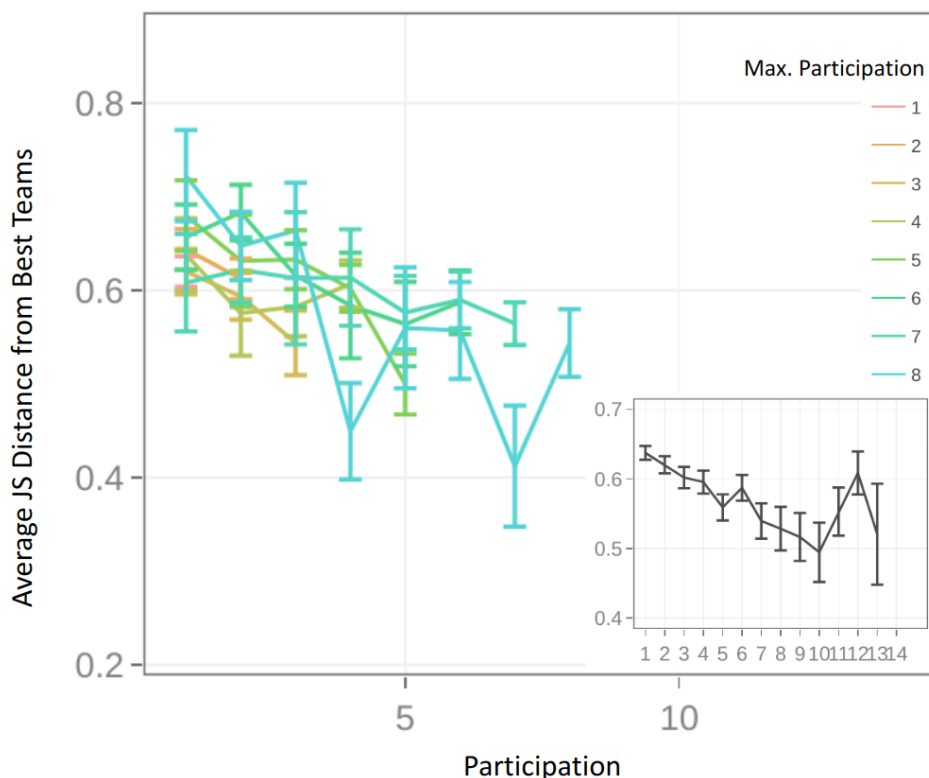


Figure 4.6: Mean distance of each teams’ task structure (constructed by projecting their bipartite division of labour networks on the task layer), from that of the best teams for each track that participated that year. The colour of each trace is grouped by the maximum number of times teams have participated in iGEM (with the inlay figure showing the aggregated trend). The distance is computed using the Jensen-Shannon divergence between networks (see methods).

decreasing with experience). This indicates the presence of specific structural patterns that teams evolve to organise their division of labour and task structure towards.

4.6 Discussion and Perspectives

This chapter presents preliminary results in studying the organisation structures of science and innovation teams at the global level (using networks of division of labour) and at the task level (focusing on the task of collaboratively editing the team wiki). We present the use of nestedness and modularity as network structural properties that capture specialisation and coordination in teams.

Teams emphasise on a more modular organisation while dividing tasks amongst their members, taking into advantage the benefits of specialisation. With interconnected yet distinct tasks, some members play generalist roles, who coordinate across different modules. Modularity is significantly associated with team performance, providing evidence supporting modular organisation in addressing interdisciplinary challenges.

Collaboratively editing the wiki is a key task that iGEM teams undertake and are evaluated on. It is strongly linked to the other tasks within iGEM, as the wiki reports on the outcomes across all distinct tasks. The co-edit structures of the wiki have a highly nested organisation, highlighting the importance of core editors who coordinate the contributions of other editors in the team. Large teams also organise their writing into modules - with editors in each module contributing to documenting the outcomes of specific tasks - and core editors coordinating across these modules.

Increased modularity at the team level and increased nestedness at the task level mirror observations across different networks of division of labour. Open Source Software (OSS) projects compartmentalise into modules, but with each module having a highly nested internal organisation Palazzi et al. [2019a] which signifies cooperation in mutualistic settings Mariani et al. [2019], Alves et al. [2019].

Nestedness and modularity are significantly associated with team performance and are more informative than aggregate measures such as degree and correlations. Both measures are at the level of the network. At the meso-scale level, bipartite motifs have shown to have facilitating or limiting effects in organisational structures Blau and McKinley [1979], Paranjape et al. [2017]. Exploring the over/under representation of these motifs at the team and task level organisation and associating them with performance is a working extension of the study.

Comparing nestedness and modularity across teams with different dimensions is not a

straightforward task. Structural properties are compared against NULL models to identify the Z-scores of over and under-representation of nestedness [Payrató-Borràs et al., 2020, Ulrich and J. Gotelli, 2007] or modularity [Pasquaretta and Jeanson, 2018, Fortuna et al., 2010]. Studying how the normalised measures underlie team performance is a working direction.

With repeated participation, teams evolve their task structure to be closer to the organisation of successful teams in their field. This provides a base to study how the evolution of task structure is reflected onto structural patterns at the team and task levels and identify classes of teams who successfully rewire their organisation to increase performance and impact.

Beyond the internal division of labour in iGEM teams, we curate the networks of collaborations between teams (with the context of their collaboration) and the networks of biobrick reuse. This presents a unique opportunity to explore how teams' external division of labour influences how they organise internally. External collaboration is primarily beneficial to gain access to knowledge and resources that are otherwise hidden. This allows teams to delegate tasks externally, and reorient their internal organisation to balance the increase in coordination costs.

iGEM changed its regulations in 2015 to incentivise teams to perform more outreach initiatives and collaborate with other teams as a part of their project. This was enabled by making collaboration as a category in the silver medal checklist from gold. This is reflected in a sharp change in the collaboration network size and density prior to and after 2015 4.7. Using this regulation change as a *shock*, a future direction is to study how external collaborations influenced teams' internal organisation.

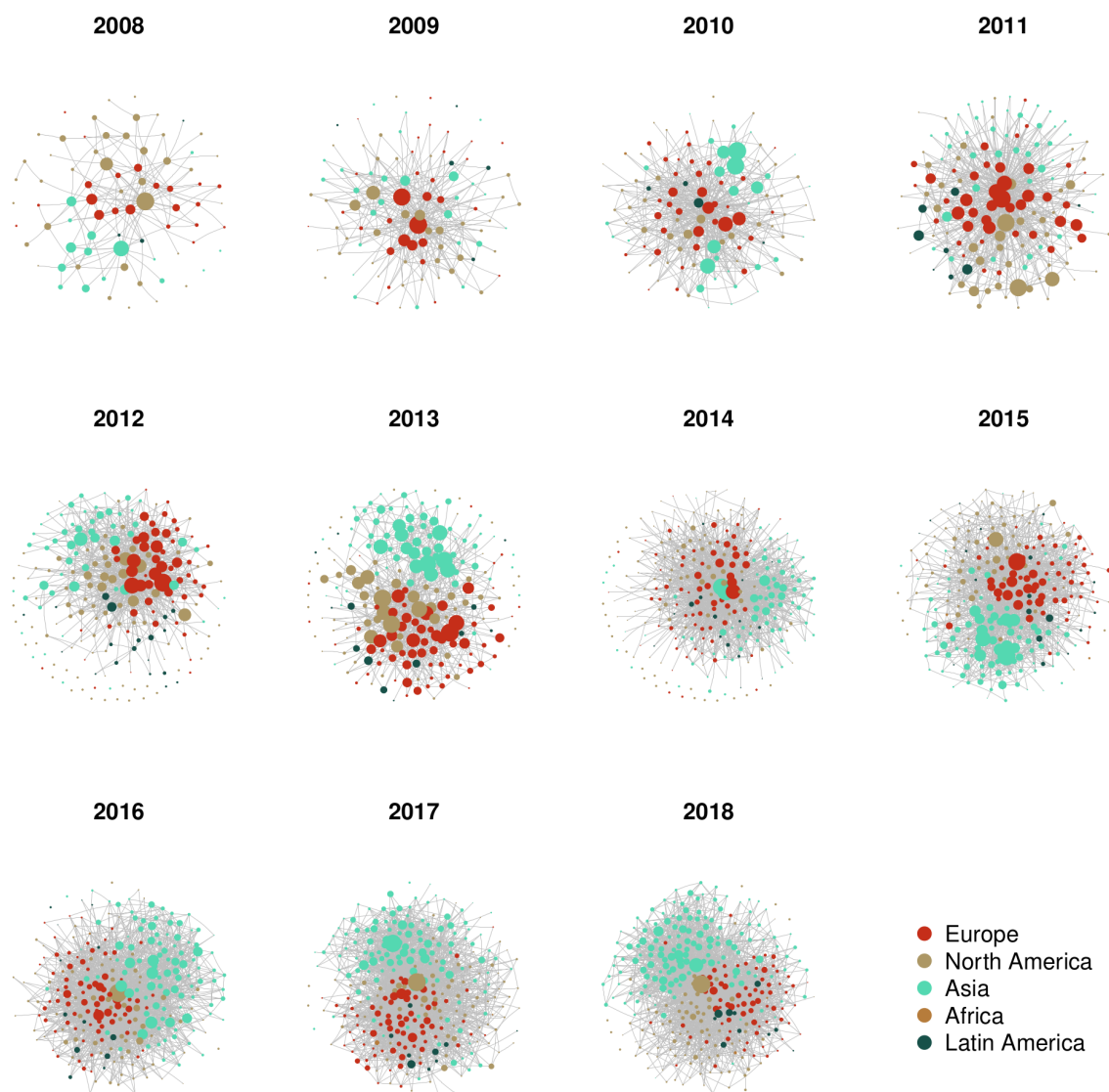


Figure 4.7: Inter-team collaboration networks in iGEM.

Bibliography

- Stefan Wuchty, Benjamin F. Jones, and Brian Uzzi. The Increasing Dominance of Teams in Production of Knowledge. *Science*, 316(5827):1036–1039, May 2007. doi: 10.1126/science.1136099. URL <https://www.science.org/doi/10.1126/science.1136099>. Publisher: American Association for the Advancement of Science.
- Warren O. Hagstrom. Traditional and Modern Forms of Scientific Teamwork. *Administrative Science Quarterly*, 9(3):241–263, 1964. ISSN 0001-8392. doi: 10.2307/2391440. URL <https://www.jstor.org/stable/2391440>. Publisher: [Sage Publications, Inc., Johnson Graduate School of Management, Cornell University].
- Vincent Larivière, Yves Gingras, Cassidy R. Sugimoto, and Andrew Tsou. Team size matters: Collaboration and scientific impact since 1900. *Journal of the Association for Information Science and Technology*, 66(7):1323–1332, 2015. ISSN 2330-1643. doi: 10.1002/asi.23266. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.23266>.
_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.23266>.
- Chiara Franzoni and Henry Sauermann. Crowd science: The organization of scientific research in open collaborative projects. *Research Policy*, 43(1):1–20, February 2014. ISSN 0048-7333. doi: 10.1016/j.respol.2013.07.005. URL <https://www.sciencedirect.com/science/article/pii/S0048733313001212>.
- Camille Masselot, Rathin Jeyaram, Raphael Tackx, Jose Luis Fernandez-Marquez, François Grey, and Marc Santolini. Collaboration and Performance of Citizen Science Projects Addressing the Sustainable Development Goals. *Citizen Science: Theory and Practice*, 8(1), June 2023. ISSN 2057-4991. doi: 10.5334/cstp.565. URL <https://theoryandpractice.citizenscienceassociation.org/articles/10.5334/cstp.565>.
- Camille Masselot, Bastian Greshake Tzovaras, Chris L. B. Graham, Gary Finnegan, Rathin Jeyaram, Isabelle Vitali, Thomas Landrain, and Marc Santolini. Implementing the Co-Immune Open Innovation Program to Address Vaccination Hesitancy and Access to Vaccines: Retrospective Study. *Journal of Participatory Medicine*, 14(1):e32125, January 2022. doi: 10.2196/32125. URL <https://jopm.jmir.org/2022/1/e32125>. Company: Journal of Participatory Medicine Distributor: Journal of Participatory

Medicine Institution: Journal of Participatory Medicine Label: Journal of Participatory Medicine Publisher: JMIR Publications Inc., Toronto, Canada.

- Paul R. Lawrence and Jay W. Lorsch. Differentiation and Integration in Complex Organizations. *Administrative Science Quarterly*, 12(1):1–47, 1967. ISSN 0001-8392. doi: 10.2307/2391211. URL <https://www.jstor.org/stable/2391211>. Publisher: [Sage Publications, Inc., Johnson Graduate School of Management, Cornell University].
- Vincent Larivière, David Pontille, and Cassidy R. Sugimoto. Investigating the division of scientific labor using the Contributor Roles Taxonomy (CRediT). In *Quantitative science studies*, volume 2, pages 111–128. MIT Press, 2020. doi: 10.1162/qss_a_00097. URL <https://papyrus.bib.umontreal.ca/xmlui/handle/1866/25787>. Accepted: 2021-11-03T11:49:58Z Issue: 1.
- Carolin Haeussler and Henry Sauermann. Division of labor in collaborative knowledge production: The role of team size and interdisciplinarity. *Research Policy*, 49(6): 103987, July 2020. ISSN 00487333. doi: 10.1016/j.respol.2020.103987. URL <https://linkinghub.elsevier.com/retrieve/pii/S0048733320300676>.
- Michael Klug and James P. Bagrow. Understanding the group dynamics and success of teams. *Royal Society Open Science*, 3(4):160007, April 2016. ISSN 2054-5703. doi: 10.1098/rsos.160007. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4852640/>.
- John P. Walsh and You-Na Lee. The bureaucratization of science. *Research Policy*, 44(8):1584–1600, October 2015. ISSN 00487333. doi: 10.1016/j.respol.2015.04.010. URL <https://linkinghub.elsevier.com/retrieve/pii/S0048733315000700>.
- Paul B. Paulus and Jared B. Kenworthy. Effective Brainstorming. In Paul B. Paulus and Bernard A. Nijstad, editors, *The Oxford Handbook of Group Creativity and Innovation*, page 0. Oxford University Press, June 2019. ISBN 978-0-19-064807-7. doi: 10.1093/oxfordhb/9780190648077.013.17. URL <https://doi.org/10.1093/oxfordhb/9780190648077.013.17>.
- Bruno Latour, Philippe Maugin, and Geneviève Teil. A NEW METHOD TO TRACE THE PATH OF INNOVATIONS.
- Gijsbert Erkens, Jos Jaspers, Maaïke Prangma, and Gellof Kanselaar. Coordination processes in computer supported collaborative writing. *Computers in Human Behavior*, 21(3):463–486, May 2005. ISSN 0747-5632. doi: 10.1016/j.chb.2004.10.038. URL <https://www.sciencedirect.com/science/article/pii/S0747563204001566>.
- Donald Ruggiero Lo Sardo, Pietro Gravino, Christine Cuskley, and Vittorio Loreto. Exploitation and exploration in text evolution. Quantifying planning and translation flows during writing. *PLOS ONE*, 18(3):e0283628, March 2023. ISSN 1932-6203. doi: 10.1371/journal.pone.0283628. URL <http://arxiv.org/abs/2302.03645>. arXiv:2302.03645 [cs].

- Fengli Xu, Lingfei Wu, and James Evans. Flat teams drive scientific innovation. *Proceedings of the National Academy of Sciences*, 119(23):e2200927119, June 2022. doi: 10.1073/pnas.2200927119. URL <https://www.pnas.org/doi/abs/10.1073/pnas.2200927119>. Publisher: Proceedings of the National Academy of Sciences.
- Michele Coscia and Luca Rossi. The Impact of Projection and Backboning on Network Topologies. In *2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 286–293, August 2019. doi: 10.1145/3341161.3342862. URL <https://ieeexplore.ieee.org/document/9073533>. ISSN: 2473-991X.
- Linda Argote. *Organizational Learning: Creating, Retaining and Transferring Knowledge*. Springer US, Boston, MA, 2013. ISBN 978-1-4614-5250-8 978-1-4614-5251-5. doi: 10.1007/978-1-4614-5251-5. URL <https://link.springer.com/10.1007/978-1-4614-5251-5>.
- The Structuring of Organizations - Henry Mintzberg. a.
CRedit, b. URL <https://credit.niso.org/>.
- Derek J. De Solla Price and Donald Beaver. Collaboration in an invisible college. *American Psychologist*, 21(11):1011–1018, 1966. ISSN 1935-990X. doi: 10.1037/h0024051. Place: US Publisher: American Psychological Association.
- María J. Palazzi, Jordi Cabot, Javier Luis Cánovas Izquierdo, Albert Solé-Ribalta, and Javier Borge-Holthoefer. Online division of labour: emergent structures in Open Source Software. *Scientific Reports*, 9(1):13890, December 2019a. ISSN 2045-2322. doi: 10.1038/s41598-019-50463-y. URL <http://www.nature.com/articles/s41598-019-50463-y>.
- Tobias Kretschmer and Phanish Puranam. Integration Through Incentives Within Differentiated Organizations. *Organization Science*, 19:860–875, December 2008. doi: 10.1287/orsc.1070.0352.
- Cristian Pasquaretta and Raphaël Jeanson. Division of labor as a bipartite network. *Behavioral Ecology*, 29(2):342–352, March 2018. ISSN 1045-2249. doi: 10.1093/beheco/arx170. URL <https://doi.org/10.1093/beheco/arx170>.
- Oleksandra Poquet, Sven Trenholm, and Marc Santolini. Forum posts, communication patterns, and relational structures: A multi-level view of discussions in online courses. *Educational technology research and development*, June 2023. ISSN 1556-6501. doi: 10.1007/s11423-023-10262-9. URL <https://doi.org/10.1007/s11423-023-10262-9>.
- Nico Blüthgen, Florian Menzel, and Nils Blüthgen. Measuring specialization in species interaction networks. *BMC Ecology*, 6(1):9, 2006. ISSN 14726785. doi: 10.1186/1472-6785-6-9. URL <http://bmcecol.biomedcentral.com/articles/10.1186/1472-6785-6-9>.

- Serguei Saavedra, Felix Reed-Tsochas, and Brian Uzzi. A simple model of bipartite cooperation for ecological and organizational networks. *Nature*, 457(7228):463–466, January 2009. ISSN 1476-4687. doi: 10.1038/nature07532.
- Manuel Sebastian Mariani, Zhuo-Ming Ren, Jordi Bascompte, and Claudio Juan Tessone. Nestedness in complex networks: Observation, emergence, and implications. *Physics Reports*, 813:1–90, June 2019. ISSN 0370-1573. doi: 10.1016/j.physrep.2019.04.001. URL <https://www.sciencedirect.com/science/article/pii/S037015731930119X>.
- Gary S. Becker and Kevin M. Murphy. The Division of Labor, Coordination Costs, and Knowledge*. *The Quarterly Journal of Economics*, 107(4):1137–1160, November 1992. ISSN 0033-5533. doi: 10.2307/2118383. URL <https://doi.org/10.2307/2118383>.
- Benjamin F. Jones. The Burden of Knowledge and the “Death of the Renaissance Man”: Is Innovation Getting Harder? *The Review of Economic Studies*, 76(1):283–317, January 2009. ISSN 0034-6527. doi: 10.1111/j.1467-937X.2008.00531.x. URL <https://doi.org/10.1111/j.1467-937X.2008.00531.x>.
- María J. Palazzi, Javier Borge-Holthoefer, Claudio Tessone, and Albert Solé-Ribalta. Antagonistic Structural Patterns in Complex Networks. Technical Report arXiv:1810.12785, arXiv, November 2018. URL <http://arxiv.org/abs/1810.12785>. arXiv:1810.12785 [physics] type: article.
- Miguel A. Fortuna, Daniel B. Stouffer, Jens M. Olesen, Pedro Jordano, David Mouillot, Boris R. Krasnov, Robert Poulin, and Jordi Bascompte. Nestedness versus modularity in ecological networks: two sides of the same coin? *Journal of Animal Ecology*, 79(4):811–817, 2010. ISSN 1365-2656. doi: 10.1111/j.1365-2656.2010.01688.x. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2656.2010.01688.x>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2656.2010.01688.x>.
- Javier Borge-Holthoefer, Raquel A. Baños, Carlos Gracia-Lázaro, and Yamir Moreno. Emergence of consensus as a modular-to-nested transition in communication dynamics. *Scientific Reports*, 7(1):41673, January 2017. ISSN 2045-2322. doi: 10.1038/srep41673. URL <https://www.nature.com/articles/srep41673>. Publisher: Nature Publishing Group.
- M. J. Palazzi, J. Borge-Holthoefer, C. J. Tessone, and A. Solé-Ribalta. Macro- and mesoscale pattern interdependencies in complex networks. *Journal of The Royal Society Interface*, 16(159):20190553, October 2019b. doi: 10.1098/rsif.2019.0553. URL <https://royalsocietypublishing.org/doi/10.1098/rsif.2019.0553>. Publisher: Royal Society.
- Melvin E Conway. HOW DO COMMITTEES INVENT? 1968.
- Karin Knorr Cetina. *Epistemic Cultures: How the Sciences Make Knowledge*. Harvard University Press, 1999.

- Linda Argote, Sunkee Lee, and Jisoo Park. Organizational Learning Processes and Outcomes: Major Findings and Future Research Directions. *Management Science*, 67(9):5399–5429, September 2021. ISSN 0025-1909, 1526-5501. doi: 10.1287/mnsc.2020.3693. URL <https://pubsonline.informs.org/doi/10.1287/mnsc.2020.3693>.
- Markus C. Becker. Organizational Routines and Organizational Learning. In Linda Argote and John M. Levine, editors, *The Oxford Handbook of Group and Organizational Learning*, page 0. Oxford University Press, January 2020. ISBN 978-0-19-026336-2. doi: 10.1093/oxfordhb/9780190263362.013.4. URL <https://doi.org/10.1093/oxfordhb/9780190263362.013.4>.
- Richard R. Nelson and Sidney G. Winter. *An evolutionary theory of economic change*. The Belknap Press of Harvard Univ. Press, Cambridge, Mass., digitally reprinted edition, 2004. ISBN 978-0-674-27228-6.
- Bruce Kogut and Udo Zander. Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science*, 3(3):383–397, 1992. ISSN 1526-5455. doi: 10.1287/orsc.3.3.383. Place: US Publisher: Institute for Operations Research & the Management Sciences (INFORMS).
- Christopher W. Wiese and C. Shawn Burke. Understanding Team Learning Dynamics Over Time. *Frontiers in Psychology*, 10, June 2019. ISSN 1664-1078. doi: 10.3389/fpsyg.2019.01417. URL <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2019.01417/full>. Publisher: Frontiers.
- David Lazer and Allan Friedman. The Network Structure of Exploration and Exploitation. *Administrative Science Quarterly*, 52(4):667–694, December 2007. ISSN 0001-8392, 1930-3815. doi: 10.2189/asqu.52.4.667. URL <http://journals.sagepub.com/doi/10.2189/asqu.52.4.667>.
- Phanish Puranam and Boris Maciejovsky. Organizational Structure and Organizational Learning. In Linda Argote and John M. Levine, editors, *The Oxford Handbook of Group and Organizational Learning*, page 0. Oxford University Press, January 2020. ISBN 978-0-19-026336-2. doi: 10.1093/oxfordhb/9780190263362.013.35. URL <https://doi.org/10.1093/oxfordhb/9780190263362.013.35>.
- Özgecan Koçak, Daniel A. Levinthal, and Phanish Puranam. The Dual Challenge of Search and Coordination for Organizational Adaptation: How Structures of Influence Matter. *Organization Science*, 34(2):851–869, March 2023. ISSN 1047-7039. doi: 10.1287/orsc.2022.1601. URL <https://pubsonline.informs.org/doi/full/10.1287/orsc.2022.1601>. Publisher: INFORMS.
- Bruno Latour and Steve Woolgar. *Laboratory Life: The Construction of Scientific Facts*. Princeton University Press, September 1986. ISBN 978-0-691-02832-3. Google-Books-ID: XTcjm0flPdYC.

- Ashwin K. Jainarayanan, Anastasios Galanis, Athira Sreejith, Sourav Suresh, Amatullah Mustafa Nakara, Guilherme E. Kundlatsch, and Roger Rubio-Sánchez. iGEM comes of age: trends in its research output. *Nature Biotechnology*, 39(12):1599–1601, December 2021. ISSN 1546-1696. doi: 10.1038/s41587-021-01152-7. URL <https://www.nature.com/articles/s41587-021-01152-7>. Number: 12 Publisher: Nature Publishing Group.
- Marc Santolini, Leo Blondel, Megan J. Palmer, Robert N. Ward, Rathin Jeyaram, Kathryn R. Brink, Abhijeet Krishna, and Albert-Laszlo Barabasi. iGEM: a model system for team science and innovation, October 2023. URL <http://arxiv.org/abs/2310.19858>. arXiv:2310.19858 [physics].
- Leo Blondel, Rathin Jeyaram, Abhijeet Krishna, and Marc Santolini. iGEM: a model system for team science and innovation, April 2024. URL <https://zenodo.org/records/11072818>.
- Rathin Jeyaram, Robert N. Ward, and Marc Santolini. Reconstructing networks from text using Large Language Models (LLMs). In *The 12th International Conference on Complex Networks and their Applications*, Menton, France, November 2023. URL <https://hal.science/hal-04514924>.
- M. E. J. Newman. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2):404–409, January 2001. doi: 10.1073/pnas.98.2.404. URL <https://www.pnas.org/doi/10.1073/pnas.98.2.404>. Publisher: Proceedings of the National Academy of Sciences.
- Muhammed A. Yildirim and Michele Coscia. Using Random Walks to Generate Associations between Objects. *PLoS ONE*, 9(8):e104813, August 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0104813. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4143196/>.
- Zachary P. Neal. backbone: An R package to extract network backbones. *PLOS ONE*, 17(5):e0269137, May 2022. ISSN 1932-6203. doi: 10.1371/journal.pone.0269137. URL <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0269137>. Publisher: Public Library of Science.
- Manlio De Domenico, Vincenzo Nicosia, Alexandre Arenas, and Vito Latora. Structural reducibility of multilayer networks. *Nature Communications*, 6(1):6864, April 2015. ISSN 2041-1723. doi: 10.1038/ncomms7864. URL <https://www.nature.com/articles/ncomms7864>. Number: 1 Publisher: Nature Publishing Group.
- Stephen J. Beckett. Improved community detection in weighted bipartite networks. *Royal Society Open Science*, 3(1):140536, January 2016. ISSN 2054-5703. doi: 10.1098/rsos.140536.
- Mário Almeida-Neto, Paulo Guimarães, Paulo R. Guimarães Jr, Rafael D. Loyola, and Werner Ulrich. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos*, 117(8):1227–1239, 2008. ISSN 1600-0706. doi: 10.1111/j.0030-1299.2008.16644.x. URL <https://onlinelibrary>.

- wiley.com/doi/abs/10.1111/j.0030-1299.2008.16644.x.
_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.0030-1299.2008.16644.x>.
- Clàudia Payrató-Borràs, Laura Hernández, and Yamir Moreno. Measuring nestedness: A comparative study of the performance of different metrics. *Ecology and Evolution*, 10(21):11906–11921, 2020. ISSN 2045-7758. doi: 10.1002/ece3.6663. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.6663>.
_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/ece3.6663>.
- Clàudia Payrató Borràs, Yamir Moreno Vega, and Laura Hernández. Complexity in the entangled bank: On the structural and dynamical properties of empirical mutualistic networks. 2020. URL <https://zaguan.unizar.es/record/101124>.
- Ashwin Paranjape, Austin R. Benson, and Jure Leskovec. Motifs in Temporal Networks. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pages 601–610, February 2017. doi: 10.1145/3018661.3018731. URL <http://arxiv.org/abs/1612.09259>. arXiv:1612.09259 [physics, stat].
- biometry/bipartite, July 2024. URL <https://github.com/biometry/bipartite>. original-date: 2014-10-06T13:26:44Z.
- Luiz G. A. Alves, Giuseppe Mangioni, Isabella Cingolani, Francisco Aparecido Rodrigues, Pietro Panzarasa, and Yamir Moreno. The nested structural organization of the world-wide trade multi-layer network. *Scientific Reports*, 9(1):2866, February 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-39340-w. URL <https://www.nature.com/articles/s41598-019-39340-w>. Number: 1 Publisher: Nature Publishing Group.
- Judith R. Blau and William McKinley. Ideas, Complexity, and Innovation. *Administrative Science Quarterly*, 24(2):200–219, 1979. ISSN 0001-8392. doi: 10.2307/2392494. URL <https://www.jstor.org/stable/2392494>. Publisher: [Sage Publications, Inc., Johnson Graduate School of Management, Cornell University].
- Werner Ulrich and Nicholas J. Gotelli. Disentangling community patterns of nestedness and species co-occurrence. *Oikos*, 116(12):2053–2061, 2007. ISSN 1600-0706. doi: 10.1111/j.2007.0030-1299.16173.x. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2007.0030-1299.16173.x>.
_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.2007.0030-1299.16173.x>.

Conclusion

Issues that require urgent action, like reaching the Sustainable Development Goal (SDG) targets, affects the global public indiscriminately, yet its challenges are more pronounced in settings that constantly suffer due to the disparity in resources and support. The foundational goal of open science and innovation practices is to democratise the access to scientific knowledge and participation in the processes that lead to its production, with the aim of bridging this gap. This however has its obstacles. SDG challenges are interdisciplinary, often requiring expert knowledge across diverse fields to design solutions. Scientists in these settings have to wear many hats - like seeking out funding and resources, coordinating inputs from domain experts, organise the scientific process amongst the participating team, communicate knowledge at each step of the way - all in addition to producing rigorous and impactful outputs. These obstacles block the entry of new interests, whether be it academic students or the general public, into addressing SDG goals, despite open innovation encouraging participation.

The goal of practitioners is to overcome these obstacles through initiatives that train participants in collective and open science. Events like hackathons, summer schools, incubators train participants in tackling these challenges while producing scientific knowledge with actionable impact. This is an ever changing landscape - with in depth knowledge required on what enables collaborative knowledge production - to facilitate the next generation of scientists and the global public to participate in collective action.

Collaborative knowledge production, or doing science in teams, has been a focus of research in organisation theory, team science and Science and Technology Studies (STS). Research in these fields focus on understanding the benefits to diversity (background, experiences), internal organisation, external collaboration in producing impact. Extending key observations from these fields to study open science and innovation settings requires a comprehensive understanding of team organisation processes, measure outcomes and understand how the two are associated with each other, helping practitioners in designing initiatives that foster innovation over the short and long term.

This thesis attempts to move the needle ever so slightly in this direction. Positioned at the interface of network science, computer science and inspired from team science and organisation theory, my main contributions are:

-
- Curating team meta data, internal and external collaborations using a mixed approach of both quantitative (digital traces) and qualitative (self-reports) data sources. Developing a pipeline to recover team collaborations with contextual information from unstructured text, using Large Language Models.
 - Representing team collaborations as networks and building measures that capture diversity, engagement, internal and external collaborations. Focusing specifically on how teams organise their members to specialise on different tasks and how they coordinate to integrate inputs across the team, through the lens of nested and modular network structures. Studying how teams evolve their organisational structure with experience and offering perspectives for future research on how internal organisation of teams adapts with external collaboration.
 - Presenting team observations through dashboards to support practitioners in providing facilitation in open science and innovation settings.

With new(er) complex challenges emerging globally with every passing second, the need for collective action increases exponentially. I hope the work presented in this thesis can be a valuable step in the ladder used to reach greater heights and enable collaborative research in an open and democratic scientific community.

List of Figures

1.1	Bipartite networks a) Interactions between team members (green) and tasks (red). Bottom shows the incidence matrix of the bipartite network. b) Contribution structure of the <i>django-money</i> OSS project on Github. Green indicates the contributors and red the files. Vertex size highlights the degree, showing the presence of core contributors. c) Member (red) - Task (green) network with the tasks corresponding to the CREDIT categories.	10
1.2	Multilayer networks of Collaborations. (Left) External collaborations between teams and (Right) Zoomed in version of the internal organisation between members (red) and tasks (green)	12
2.1	Description of the challenge. (a) Schematics of a GEAR cycle. (b) Visual for the GEAR Cycle 2.	35
2.2	Description of the cohort. Co-occurrence networks across participants highlighting (a) their previous experience with Sustainable Development Goals and (b) their main background or field of study. In these networks nodes are linked by the number of times they are co-reported by a participant, and colors correspond to denser sub-networks as determined by the modularity algorithm. Participants had prior experience with Climate Action and Gender Equality, and came from interdisciplinary backgrounds.	40
2.3	Communication activity. a) Total number of posts on Slack per week. The Evaluate phase is highlighted in blue, and the Accelerate phase, consisting of two periods, is highlighted in red. (b) Mention network extracted from Slack during the Evaluate phase. Nodes represent aggregated individuals at the team level. Teams are linked by weighted edges quantifying the number of times an individual from one team mentions an individual from another team. Self-loops denote intra-team interactions. Grey color denotes the organizing team, and the color denotes the stage achieved in the program: in order, Evaluate (yellow), Accelerate (green), Refine (blue). (c) Proportion of Slack mentions that are from the organization team, towards the organization team, intra-team or inter-team.(d) Number of mentions (sent or received) per team, following the color code from b.	41

2.4	Activities during the Evaluate phase. (a) Total number of reports of an activity per week. We see a switch from brainstorming/ planning activities to the documentation and preparation of the final presentation. (b) Total number of activities reported per team per week.	44
2.5	Collaboration activity. (a–c) Participant interaction networks constructed from self-report data from CoSo, using prompts:(a) “Which of these people did you know personally before?” (b) “Who did you work with last week?” and (c) “Who did you seek advice from last week?” The size of a node is proportional to the total number of interactions of a node across the 3 networks. (d) Proportion of interactions in the collaboration network that involve the organization team (red), or that are intra- (green) or inter- (blue) team, in time. Error bars denote the standard error of the estimate given a number of interactions observed, assuming a binomial statistics. (e) Same than d, for the advice seeking network.	45
2.6	Comparison of the interaction networks. (a) Team-level networks for the different interaction networks collected. (b) Corresponding aggregate network, where edge weights are the sum over weights across the individual networks shown in a. (c) Network density and centralization (see Methods) across the 4 considered networks. (d) Jaccard similarity between the networks in a. The similarity measures the number of edges shared between any two networks (intersection), divided by the total number of edges present in both networks (union), and ranges from 0 (most dissimilar) to 1 (most similar).	47
2.7	Association with performance. Correlations between performance assessment (rows) and team features (columns). The correlation value is indicated when the correlation is significant at the $p = 0.1$ level. Shaded areas correspond to sets of features associated with metrics related to outcome (blue) or to process (green).	48
2.8	Screenshots of the CoSo interface	52
2.9	Sankey diagrams of teams current or highest level of study (a) and disciplinary backgrounds (b)	53
2.10	Gender (a) and age (b) distributions across teams	54
2.11	Communication tools reported to be used by teams to communicate. Number indicates number of answers across participants (total N=22 participants).	55
2.12	Heatmap indicating the number of weeks each activity has been reported by a given team, across 4 weeks. Activities and teams are ranked by row and column sums respectively.	56

2.13	Correlation matrices of the features shown in Fig 7, for evaluations (a) and team features (b). Numbers correspond to p-values of the correlations. We grayed out cells with a p-value $p > 0.1$. We find two groups of evaluations: outcomes (top left) and processes (bottom right). For team features, we find that Slack activity is correlated with the intra-team collaborations measured with CoSo, highlighting that digital traces can capture qualitative insights on team work.	57
2.14	Standardised regression coefficients for the team characteristics associated with the Evaluate final score, selected through the LASSO regression (see Methods). Error bars denote 95% confidence intervals. Positive estimates denote a positive association between the feature and the outcome. For network measures, we show in brackets the type of network it is measured from. These consist of Slack network, or CoSo network: prior ties (“Which of these people did you know personally before?”), advice seeking (“Who did you seek advice from last week?”) and work collaborations (“Who did you work with last week?”).	66
2.15	Same as figure 2.14, for the outcomes shown in bold.	68
2.16	Same as figure 2.14, for the final stage achieved in the GEAR cycle. For network measures, we show in brackets the type of network it is measured from. These consist of Slack network, or CoSo network: prior ties (“Which of these people did you know personally before?”), advice seeking (“Who did you seek advice from last week?”) and work collaborations (“Who did you work with last week?”).	69
2.17	Page 1 of the dashboard. The panels on the left shows the nationalities (top, filtered by GEAR cycle) and age distributions (bottom). The panel on the right shows the distribution of academic background as a heatmap.	77
2.18	Interactions captured from surveys. Panels in each column plot an interactive network for the respective survey question for GEAR cycles 2 and 3. Nodes in the network are each individual participant and the network can be filtered by selecting a specific team - highlighted by a unique color. The node size is determined by its degree. The width of an edge indicates the number of mentions (or reactions) between the same pair of vertices.	77
2.19	Interactions captured from Slack. The panels to the left show the network of mentions and the ones on the right the network of reactions in the GEAR cycles 2 and 3. Vertices correspond to an individual Slack user and the color highlights their team. The vertex size is the degree. The width of an edge indicates the number of mentions (or reactions) between the same pair of vertices.	78
2.20	Participant responses to the digital platforms they used during the Evaluate phase (left panel) and their feedback on different facets of the incubators’ training and support structures (right panel). The dots indicate the average with standard error bars.	79

3.1	a) Competition Cycle in iGEM. b) Geographic Dispersion of iGEM Teams 2008-2018. Each dot corresponds to a team, and the color of the dot represents the last year the team participated in the competition. c) iGEM competition tracks and their evolution. All figures taken from [Santolini et al., 2023]	94
3.2	a) Snapshots from Team wikis. b) Criteria for Medals and Awards over the years [Santolini et al., 2023].	96
3.3	Example of the Wiki page showing biobrick reuse (left) and the network of biobrick reuse from 2008-2018, with colors indicating geographic regions of the teams that designed the biobrick (right) There are a core set of biobricks, that teams often reuse - indicated by vertex size and their central position in the network.	97
3.4	Collaboration network between teams participating in iGEM 2018. Teams are coloured based on their geographic region and the size of the vertex corresponds to the degree (number of collaborations that the team was involved in).	99
3.5	Example of a Wiki Co-edit network. Vertices coloured green correspond to the team members and red to the pages of the wiki. The network here shows a modular organisation of the teams' wiki editing process.	101
3.6	Overview of the pipeline to reconstruct social networks from text	109
3.7	Reconstruction of inter-team collaboration networks (a) Precision and recall of GPT models and fuzzy matching approach against manually curated chunks. (b) Confusion matrix showing the accuracy of category prediction for relationship contexts. The color coding indicates row-standardized Z scores values. (c) Receiver Operating Curves (ROC curves) for weighted interactions from 2015. Dashed line indicates a line of slope 1 and intercept 0. (d) Comparison between network assortativity values for reconstructed and manually obtained 2015 inter-team network. (e) Correlations of network properties between the 2015 reconstructed and manual network. In all plots, the error bars are computed using the jackknife method	112
3.8	Reconstruction of team contribution structure (a) Precision and recall of GPT models against manually curated contribution networks. (b) Confusion matrix of inferred category labels	114
3.9	Recovering bipartite network properties. (a) Cumulative degree distributions for the task categories and team members layers of the bipartite network, comparing manually curated networks (grey) with gpt-4 inferred networks (green). (b) Nestedness values, computed using NODFc [Song et al., 2017], are shown between the predicted and manual networks. Dashed line indicates a line of slope 1 and intercept 0	115

3.10	a-b. Examples of team wiki pages for collaboration and attribution. b. We showcase an example of why chunking attribution pages can result in misleading networks. The chunk highlighted in green presents a case where the inference works properly, while in the red chunk, the GPT model assigns the text included in the overlap to the next member name.	128
3.11	Confusion matrices like in Fig 3.7b, for the different models of Fig 3.7a. . .	128
3.12	Precision and recall values, as in Fig 3.8a, but for each category of Fig 3.8b.	129
3.13	Recall values, as in Fig 3.8a, for each team.	130
4.1	Example of two bipartite networks, represented by incidence matrices, where the aggregate measures (degree correlations between users and tasks) are the same. But the matrix on the left has intermediate modularity while the matrix on the right has high modularity.	144
4.2	a) Constructing the bipartite network of co-edits (members to pages) of the team wikis b) Pipeline to construct the bipartite division of labour network from unstructured wiki attribution pages. Large language models from the GPT family are used to curate the interactions between users and the tasks they performed, matching users to their identifiers and description of tasks to a standardised list of tasks performed by iGEM teams. [Jeyaram et al., 2023] describes the pipeline and methods to evaluate its performance in more detail. c) Example of the task structure of a team, identified by projecting the bipartite division of labour network on the task layer.	148
4.3	Comparing Bipartite networks of division of labour (curated from attribution statements using GPT) and organisation at the task level (co-editing the wiki) a) Cumulative degree distribution of users in both division of labour and co-edit networks in iGEM 2017. b) Example team incidence matrices. (top left) Team with high nestedness in their division of labour network. (top right) Team with high modular structure in their division of labour network. (bottom left) Team with a nested wiki co-edit structure. (bottom right) Team with a modular wiki co-edit structure. Matrices plotted using the <i>plotModuleWeb</i> function in the <i>noa</i> [2024] library in R. c) Scatter plot of nestedness (top) and modularity (bottom) comparing teams' division of labour network to their respective wiki co-edit network.	153

4.4	a) Linear model coefficients comparing the effects of nestedness, bipartite modularity and mean correlation between activity-user incidence vectors in predicting the medal achieved. The subset teams participated in iGEM since 2014 and have atleast 70% of the team roster captured from the attribution statements. Teams also report atleast 10 users and 10 tasks - filtering for large teams b) Linear model coefficients comparing the effects of nestedness, bipartite modularity and mean correlation between activity-user incidence vectors in predicting log edit size of the team wiki. The subset teams participated in iGEM since 2014 and have atleast 70% of the team roster captured from the attribution statements. Teams also report atleast 10 users and 10 tasks - filtering for large teams.	154
4.5	a) Linear model coefficients comparing the effects of nestedness, bipartite modularity and mean correlation between activity-user incidence vectors in predicting the medal achieved. At least 20% of the team roster are editors. Teams also report at least 10 editors and 10 tasks - filtering for large teams. b) Linear model coefficients comparing the effects of nestedness, bipartite modularity and mean correlation between activity-user incidence vectors in predicting log edit size of the team wiki. At least 20% of the team roster are editors of the wiki. Teams also report atleast 10 editors and 10 tasks - filtering for large teams	155
4.6	Mean distance of each teams' task structure (constructed by projecting their bipartite division of labour networks on the task layer), from that of the best teams for each track that participated that year. The colour of each trace is grouped by the maximum number of times teams have participated in iGEM (with the inlay figure showing the aggregated trend). The distance is computed using the Jensen-Shannon divergence between networks (see methods).	156
4.7	Inter-team collaboration networks in iGEM.	159

List of Tables

2.1	Gear Cycle 2 project description	36
2.2	List of surveys of the GEAR cycle 2	51
