



Collaboration and Performance of Citizen Science Projects Addressing the SDGs

Camille Masselot, Rathin Jeyaram, Raphael Tackx, Jose Luis
Fernandez-Marquez, François Grey, Marc Santolini

► To cite this version:

Camille Masselot, Rathin Jeyaram, Raphael Tackx, Jose Luis Fernandez-Marquez, François Grey, et al.. Collaboration and Performance of Citizen Science Projects Addressing the SDGs. 2022. hal-03808080

HAL Id: hal-03808080

<https://hal.science/hal-03808080>

Preprint submitted on 10 Oct 2022

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.

Collaboration and Performance of Citizen Science Projects Addressing the SDGs

Camille Masselot^{*1,2}, Rathin Jeyaram^{*1,2}, Raphael Tackx^{1,2}, Jose Luis Fernandez-Marquez³, François Grey³, Marc Santolini^{1,2,+}

Affiliations:

(1) Université Paris Cité, Inserm, System Engineering and Evolution Dynamics, F-75004 Paris, France

(2) Learning Planet Institute, F-75004 Paris, France

(3) University of Geneva, Carouge, CH-1227, Switzerland

* Contributed equally

+ Correspondances: marc.santolini@cri-paris.org

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. In this article, we ask whether the monitoring of social interaction networks can inform on the participatory processes and outcomes of citizen science projects addressing the SDGs. We studied fourteen early-stage citizen science projects that participated in an innovation cycle focused on 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 multiple complementary dimensions, informing on 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 engagement are associated with participation processes. Overall, we exhibit the relevance of monitoring social interaction dynamics for guiding early-stage citizen science projects, which is of importance for initiatives leveraging Citizen Science to address the SDGs.

Keywords

Citizen science, Network Science, Monitoring and Evaluation, Sustainable Development Goals, Team science, Collaboration

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., 2021), 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, 2021; Masselot et al., 2022).

Compared with traditional scientific work, citizen science requires to define processes of engagement and coordination, from simple data collection to co-design strategies (Franzoni and Sauermann, 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), understand modes of co-production for sustainability (Chambers et al., 2021) and 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 advanced projects, and organizers of initiatives such as Crowd4SDG (Crowd4SDG consortium, 2020) 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., 2022; 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, 2021; Masselot et al., 2022) or open source communities (Gargiulo et al., 2021; 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 to collect 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.

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 1a). 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 1b), 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, thirty to fifty 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 Trialogue, 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.

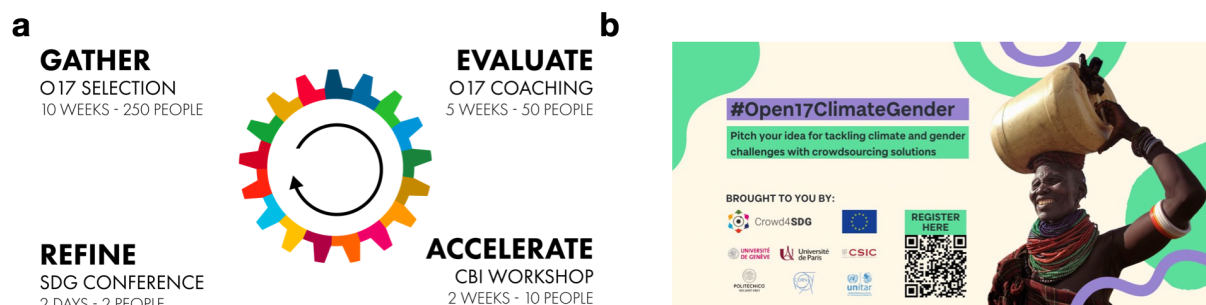


Figure 1: Description of the challenge (a) Schematics of a GEAR cycle. (b) Visual for the GEAR Cycle 2.

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).

Description of team projects

In this study 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.

Table 1. GEAR cycle 2 project description

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 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

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, 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) (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 (“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.

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: with the exception of 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

In order to document measures of participation, we monitored features related to team composition, communication, collaboration and activity (features from Figure 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 amount 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 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.”, “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, and the weights of the inter-team links are summed.

The CoSo networks are directly inferred from the surveys. In Figure 5 and 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 6 further aggregates the individual networks at the team level.

In Figure 6c, 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 analyses

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 $p = 0.1$.

Results

In this study, we provide an analysis of team work 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 where teams build their project and engage in interactions with their peers and mentors.

Cohort description

The cohort was constituted 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 highschool 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).

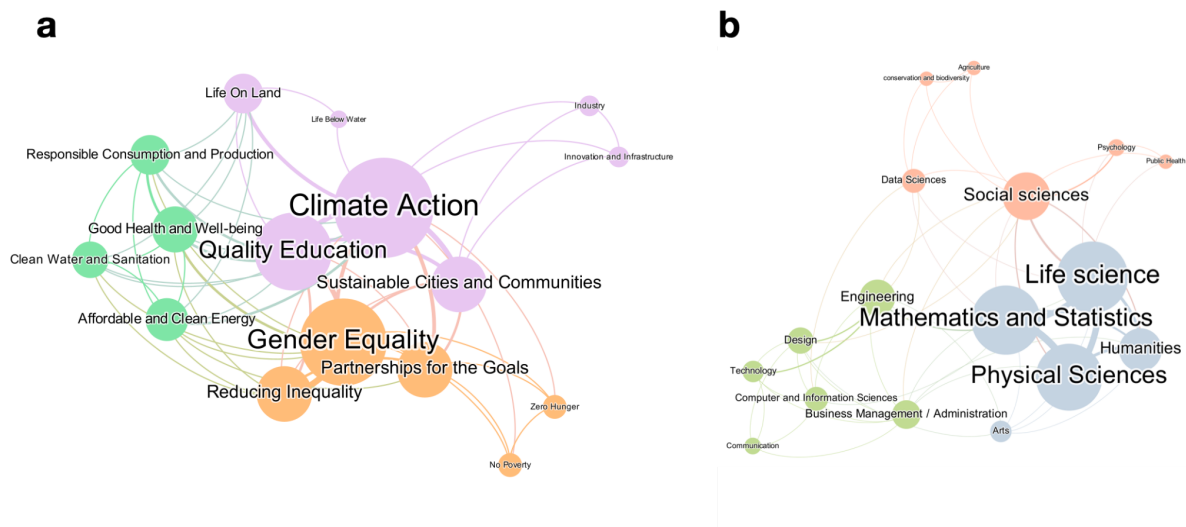


Figure 2: Description of the cohort. Co-occurrence networks across participants highlighting (a) their previous experience with SDGs and (b) their main background or field of study. In these networks nodes are linked by the number of times they are co-reported by a participant, and colors correspond to denser subnetworks as determined by the modularity algorithm. Participants had prior experience with Climate Action and Gender Equality, and came from interdisciplinary backgrounds.

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 particular SDGs covered most goals, with a primary focus on Climate Action and Gender Equality, as expected from the topic of the GEAR cycle (Figure 2a). Interestingly, the participants (and the teams themselves) displayed a high level of interdisciplinarity, with backgrounds spanning natural sciences, technology, and humanities (Figure 2b and S2).

Team communication

During the Evaluate phase, teams used a Slack workspace in order 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.

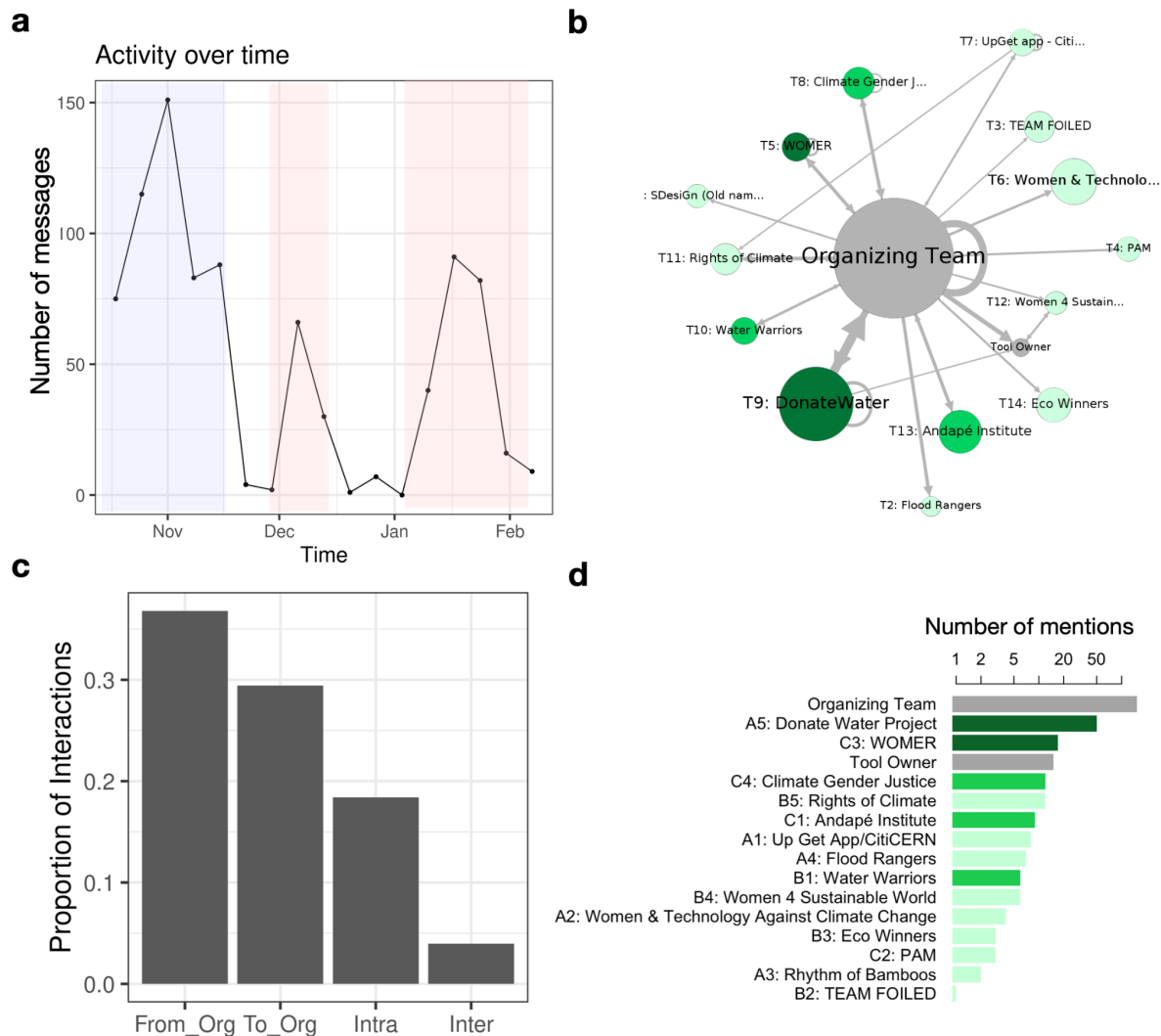


Figure 3: Communication activity. (a) Total number of posts on Slack per week. The Evaluate phase is highlighted in blue, and the Accelerate phase, consisting of two periods, is highlighted in red. (b) Mention network extracted from Slack during the Evaluate phase. Nodes represent aggregated individuals 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 green color denotes

the stage achieved in the program: in order, Evaluate (light), Accelerate (medium), Refine (dark). (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 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 3a). This might be due to the fact that teams would work solely during the program, or that 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 to represent 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 3b).

We observe a high centralization of interactions to the organizing team (Figure 3c), 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 3c), 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. In order 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 4a). 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 4b).

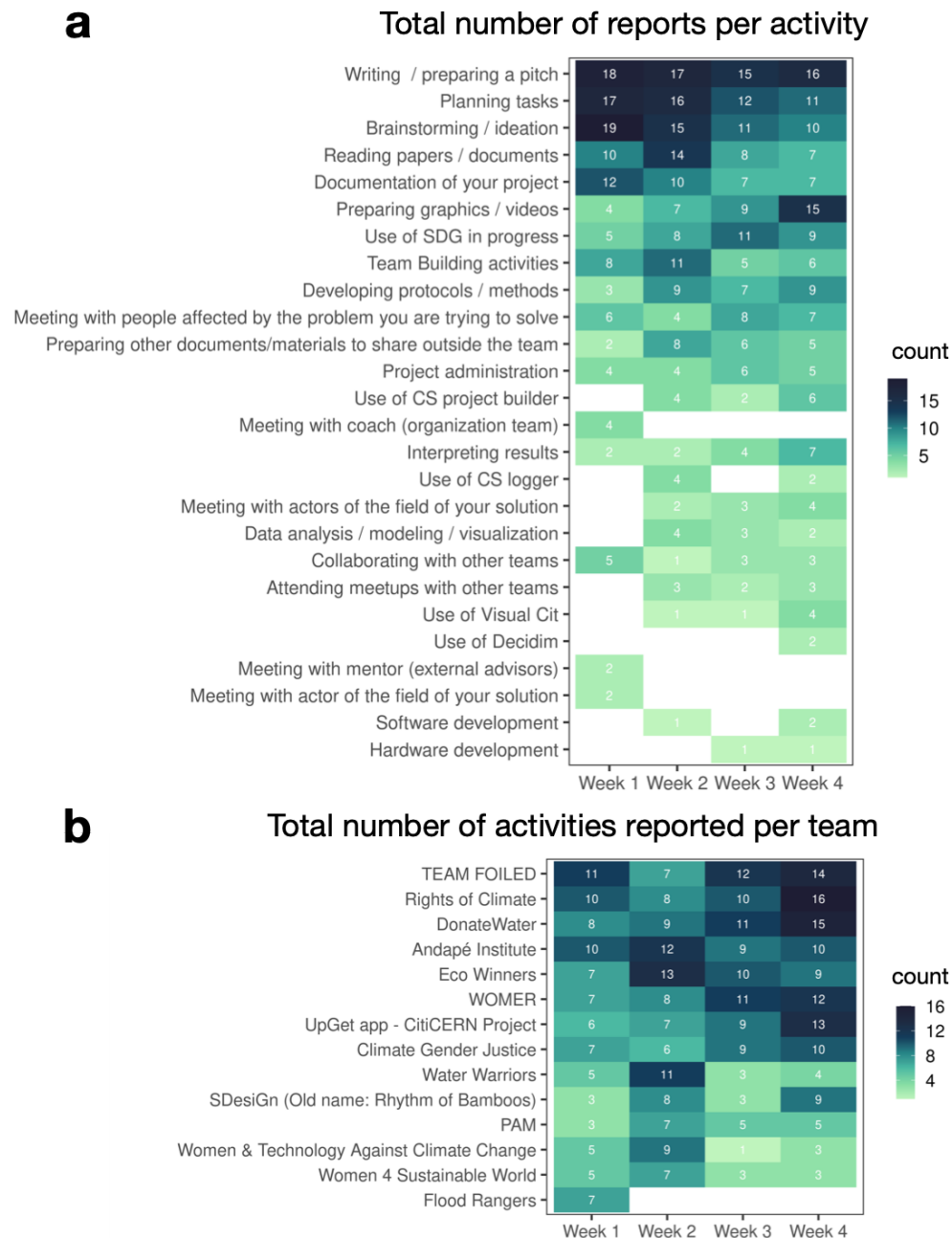


Figure 4: Activities during the Evaluate phase. (a) Total number of reports of an activity per week. We see a switch from brainstorming/planning activities to the documentation and preparation of the final presentation. (b) Total number of activities reported per team per week.

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 5a,b,c). These surveys were aimed at investigating the collaborative dynamics during the GEAR, its evolution in time and eventual impact on team performance.

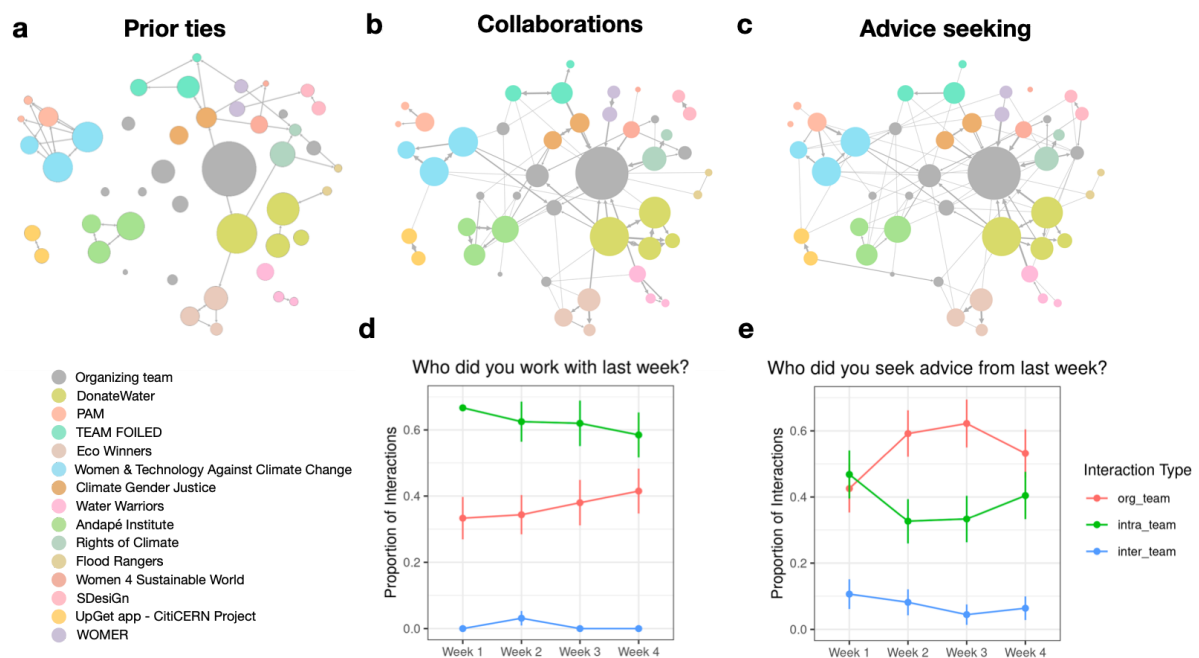


Figure 5: Collaboration activity. (a-c) Participant interaction networks constructed from self-report data from CoSo, using prompts: (a) “Which of these people did you know personally before?”, (b) “Who did you work with last week?”, and (c) “Who did you seek advice from last week?”. The size of a node 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.

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 5a). Yet, beyond intra-team links, we found that several participants acted as bridges between teams in the prior ties network. This is probably due to the fact that 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 4b,d), while only few inter-team interactions were observed. On the other hand, advice seeking interactions, where participants report having asked for advice from another participant, showed more inter-team interactions, with around 10% of them being inter-team ties (Figure 4c,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 6a the networks at the team level. One can observe that the networks have similar densities (Figure 6c), but different structures: the Slack mentions network is much more centralized than the surveyed interaction networks (Figure 6c), indicating that Slack usage was mostly used to exchange with the organizing team who acted as a strong hub. When aggregating the networks together, one obtains a more comprehensive interaction network (Figure 6b), doubling the density of links compared with any single network (Figure 6c).

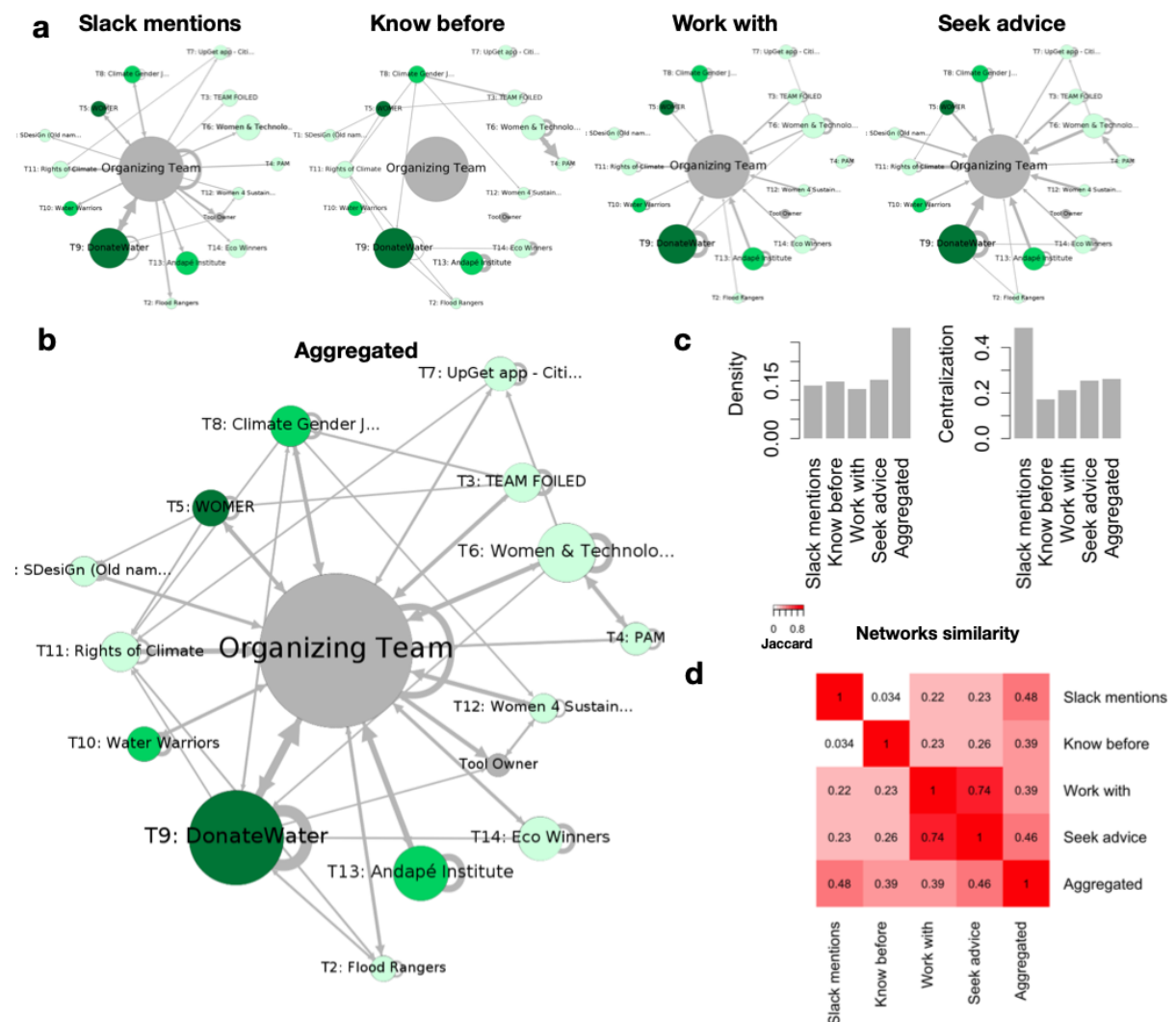


Figure 6: Comparison of the interaction networks. (a) Team-level networks for the different interaction networks collected. (b) Corresponding aggregate network, where

edge weights are the sum over weights in individual networks 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).

To further assess the topological similarity between the networks obtained, we computed the Jaccard similarity between any pair of networks, i.e. 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 ($J\sim 0.2$). Prior ties are predictive of collaboration and advice seeking ($J=0.2$) but not of Slack mentions ($J\sim 0$), 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.

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, i.e. the evaluation of the project itself, and process, i.e. 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 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., 2021). 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.

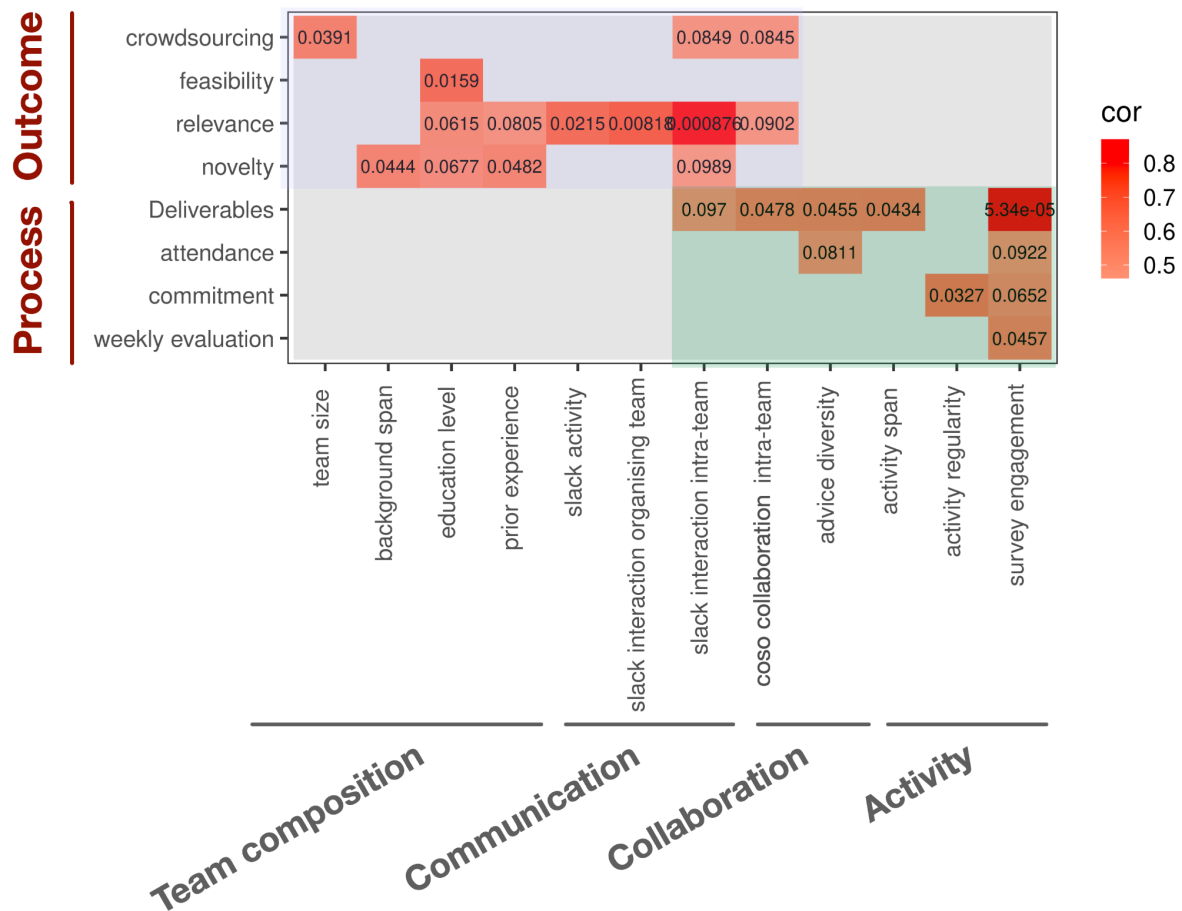


Figure 7: Association with performance. Correlations between performance assessment (rows) and team features (columns). The correlation value is indicated when the correlation is significant at the $p=0.1$ level. Shaded areas correspond to sets of features associated with metrics related to outcome (blue) or to process (green).

In the case of communication activity, we find that the overall *Slack activity* (which is very correlated with the amount 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 on 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 to judge these criteria, was particularly sensitive to the ability of teams to engage and collaborate throughout the cycle, an information that was not readily available to other experts.

Discussion

Processes of engagement and coordination are fundamental to citizen science projects (Jaeger et al., 2022; Schaefer et al., 2021). 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.

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 in particular associated with the ability of teams to produce well-defined deliverables, an indicator of team performance to a standardized task. Beyond small-scale team work, 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 incentivisation 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, 2021; 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 only offer 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.

Conclusion

One challenge that organizers of programs similar to Crowd4SDG face is to support with evidence their decision 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 course of 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., 2022; 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 measure 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 activity and collaboration networks in citizen science projects. We showed its usefulness to document the participatory processes and outcomes of the teams, demonstrating its potential to support practitioners in the design and coordination of programs aiming at fostering citizen science projects.

Supplemental Materials

Supplemental file 1: Consent form

Supplemental file 2: Tables, Methods and Figures

Supplemental file 3: Slack mention network

Supplemental file 4: CoSo self-reported interactions

Supplemental file 5: CoSo self-reported activities

Supplemental file 6: Team features and performance

Data Availability

Anonymised data can be found in the Supplemental files 3, 4, 5, 6.

Ethics and Consent

The surveys and research questions received the ethical approval of the Inserm committee (IRB00003888). Participants gave their consent to the collection of data as they registered to the Evaluate phase (see Supplemental File 1). Data was pseudo-anonymized before analysis.

Acknowledgements

We acknowledge Barbara Pernici and Jesus Cerquides for their comments on an early version of this draft, and Amudha Ravi Shankar for her important help in disseminating surveys.

Funding Information

This work was supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No 872944.

Competing Interests

The authors are part of the Crowd4SDG consortium.

Author's contribution statement

FG, JLF and MS acquired funding for the study. FG conceptualized the GEAR cycle, FG and JLF coordinated the implementation of GEAR 2. MS conceptualized and

supervised the study. MS and CM designed the monitoring methodology. RT developed the CoSo interface for survey collection. CM, RJ and RT collected data. RJ curated data. RJ and MS analyzed and visualized the data. CM and MS wrote the manuscript.

References

- Ballerini, L., Bergh, S.I., 2021. Using citizen science data to monitor the Sustainable Development Goals: a bottom-up analysis. *Sustain. Sci.* <https://doi.org/10.1007/s11625-021-01001-1>
- Burt, R.S., 2004. Structural Holes and Good Ideas. *Am. J. Sociol.* 110, 349–399. <https://doi.org/10.1086/421787>
- Chambers, J.M., Wyborn, C., Ryan, M.E., Reid, R.S., Riechers, M., Serban, A., Bennett, N.J., Cvitanovic, C., Fernández-Giménez, M.E., Galvin, K.A., Goldstein, B.E., Klenk, N.L., Tengö, M., Brennan, R., Cockburn, J.J., Hill, R., Munera, C., Nel, J.L., Österblom, H., Bednarek, A.T., Bennett, E.M., Brandeis, A., Charli-Joseph, L., Chatterton, P., Curran, K., Dumrongrojwattana, P., Durán, A.P., Fada, S.J., Gerber, J.-D., Green, J.M.H., Guerrero, A.M., Haller, T., Horcea-Milcu, A.-I., Leimona, B., Montana, J., Rondeau, R., Spierenburg, M., Steyaert, P., Zaehring, J.G., Gruby, R., Hutton, J., Pickering, T., 2021. Six modes of co-production for sustainability. *Nat. Sustain.* 4, 983–996. <https://doi.org/10.1038/s41893-021-00755-x>
- Cox, J., Oh, E.Y., Simmons, B., Lintott, C., Masters, K., Greenhill, A., Graham, G., Holmes, K., 2015. Defining and Measuring Success in Online Citizen Science: A Case Study of Zooniverse Projects. *Comput. Sci. Eng.* 17, 28–41. <https://doi.org/10.1109/MCSE.2015.65>
- Crowd4SDG consortium, 2020. Citizen Science for Monitoring Climate Impacts and Achieving Climate Resilience [WWW Document]. CORDIS Eur. Comm. URL <https://cordis.europa.eu/project/id/872944/fr> (accessed 9.21.22).
- Deri, S., Rappaz, J., Aiello, L.M., Quercia, D., 2018. Coloring in the Links: Capturing Social Ties as They are Perceived. *Proc. ACM Hum.-Comput. Interact.* 2, 1–18. <https://doi.org/10.1145/3274312>
- Fortunato, S., Bergstrom, C.T., Börner, K., Evans, J.A., Helbing, D., Milojević, S., Petersen, A.M., Radicchi, F., Sinatra, R., Uzzi, B., Vespignani, A., Waltman, L., Wang, D., Barabási, A.-L., 2018. Science of science. *Science* 359, eaao0185. <https://doi.org/10.1126/science.aao0185>
- Fraisl, D., Campbell, J., See, L., Wehn, U., Wardlaw, J., Gold, M., Moorthy, I., Arias, R., Piera, J., Oliver, J.L., Masó, J., Penker, M., Fritz, S., 2020. Mapping citizen science contributions to the UN sustainable development goals. *Sustain. Sci.* 15, 1735–1751. <https://doi.org/10.1007/s11625-020-00833-7>
- Franzoni, C., Poetz, M., Sauermann, H., 2021. Crowds, citizens, and science: a multi-dimensional framework and agenda for future research. *Ind. Innov.* 1–34. <https://doi.org/10.1080/13662716.2021.1976627>
- Franzoni, C., Sauermann, H., 2014. Crowd science: The organization of scientific

- research in open collaborative projects. *Res. Policy* 43, 1–20.
<https://doi.org/10.1016/j.respol.2013.07.005>
- Fritz, S., See, L., Carlson, T., Haklay, M. (Muki), Oliver, J.L., Fraisl, D., Mondardini, R., Brocklehurst, M., Shanley, L.A., Schade, S., Wehn, U., Abrate, T., Anstee, J., Arnold, S., Billot, M., Campbell, J., Espey, J., Gold, M., Hager, G., He, S., Hepburn, L., Hsu, A., Long, D., Masó, J., McCallum, I., Muniafu, M., Moorthy, I., Obersteiner, M., Parker, A.J., Weisspflug, M., West, S., 2019. Citizen science and the United Nations Sustainable Development Goals. *Nat. Sustain.* 2, 922–930. <https://doi.org/10.1038/s41893-019-0390-3>
- Gargiulo, F., Castaldo, M., Venturini, T., Frasca, P., 2021. A collaborative path to scientific discovery: Distribution of labor, productivity and innovation in collaborative science. *ArXiv210502303 Phys.*
- Hackman, J.R., 1987. The design of work teams, in: Lorsch, J. (Ed.), *Handbook of Organizational Behavior*. Prentice-Hall, Englewood Cliffs, NJ.
- Haklay, M., 2018. Participatory Citizen Science, in: Hecker, S., Haklay, M., Bowser, A., Makuch, Z., Vogel, J., Bonn, A. (Eds.), *Citizen Science: Innovation in Open Science, Society and Policy*. UCL Press, London, UK, pp. 52–62.
- Jaeger, J., Masselot, C., Tzovaras, B.G., Hidalgo, E.S., Haklay, M., Santolini, M., 2022. An Epistemology for Democratic Citizen Science.
<https://doi.org/10.31219/osf.io/j62sb>
- Klein, M., Maillart, T., Chuang, J., 2015. 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*. Association for Computing Machinery, New York, NY, USA, pp. 1106–1115. <https://doi.org/10.1145/2675133.2675286>
- Klug, M., Bagrow, J.P., 2016. Understanding the group dynamics and success of teams. *R. Soc. Open Sci.* 3, 160007. <https://doi.org/10.1098/rsos.160007>
- Kokshagina, O., 2021. Open Covid-19: Organizing an extreme crowdsourcing campaign to tackle grand challenges. *RD Manag. radm.* 12470.
<https://doi.org/10.1111/radm.12470>
- Masselot, C., Tzovaras, B.G., Graham, C.L.B., Finnegan, G., Jeyaram, R., Vitali, I., Landrain, T., Santolini, M., 2022. Implementing the Co-Immune Open Innovation Program to Address Vaccination Hesitancy and Access to Vaccines: Retrospective Study. *J. Particip. Med.* 14, e32125.
<https://doi.org/10.2196/32125>
- Parkinson, S., Woods, S.M., Sprinks, J., Ceccaroni, L., 2022. A Practical Approach to Assessing the Impact of Citizen Science towards the Sustainable Development Goals. *Sustainability* 14, 4676.
<https://doi.org/10.3390/su14084676>
- Poquet, O., Tupikina, L., Santolini, M., 2020. Are forum networks social networks? a methodological perspective, in: *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge, LAK '20*. Association for Computing Machinery, New York, NY, USA, pp. 366–375.
<https://doi.org/10.1145/3375462.3375531>
- Santolini, M., 2020. Covid-19: the rise of a global collective intelligence? [WWW Document]. *The Conversation*. URL <http://theconversation.com/covid-19-the-rise-of-a-global-collective-intelligence-135738> (accessed 4.2.21).

- Schaefer, T., Kieslinger, B., Brandt, M., van den Bogaert, V., 2021. Evaluation in Citizen Science: The Art of Tracing a Moving Target, in: Vohland, K., Land-Zandstra, A., Ceccaroni, L., Lemmens, R., Perelló, J., Ponti, M., Samson, R., Wagenknecht, K. (Eds.), *The Science of Citizen Science*. Springer International Publishing, Cham, pp. 495–514. https://doi.org/10.1007/978-3-030-58278-4_25
- Senabre Hidalgo, E., Perelló, J., Becker, F., Bonhoure, I., Legris, M., Cigarini, A., 2021. Participation and Co-creation in Citizen Science, in: Vohland, K., Land-Zandstra, A., Ceccaroni, L., Lemmens, R., Perelló, J., Ponti, M., Samson, R., Wagenknecht, K. (Eds.), *The Science of Citizen Science*. Springer International Publishing, Cham, pp. 199–218. https://doi.org/10.1007/978-3-030-58278-4_11
- Singh, C., Barme, E., Ward, R., Tupikina, L., Santolini, M., 2021. Quantifying the rise and fall of scientific fields. *ArXiv210703749 Phys.*
- Tackx, R., Blondel, L., Santolini, M., 2021. 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 '21*. Association for Computing Machinery, New York, NY, USA, pp. 502–507. <https://doi.org/10.1145/3460418.3479363>

Supplemental file 2: Tables, Methods and Figures

Table 1: List of surveys of the GEAR cycle 2

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%)

Methods

Teaming algorithm


Participants who joined as individuals were assigned to a novel team using the teaming algorithm Edu2Com (Georgara et al., 2020). 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

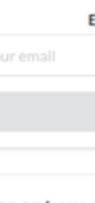
competence, personality and preferences. From the six alternatives provided, The final selected team profiling was based on a weightage that had 10% match of their competencies, 20% match of their personalities and 70% of their preferred choices. This particular alternative was chosen since it gave an ideal combination of teaming up individuals as a team along with the pre-formed teams.


competence	personality	preferences
5%	20%	75%
5%	35%	60%
5%	45%	50%
10%	20%	70%
10%	40%	50%
20%	40%	40%

Combination proposed by the Teaming Algorithm.

Supplementary figures


Crowd4SDG





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

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

Submit

How helpful you found 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

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

☐ Brainstorming / ideation
☐ Planning tasks
☐ Developing protocols / methods
☐ Reading papers / documents
☐ Attending meetings 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 Citi
☐ Use of SDG in progress

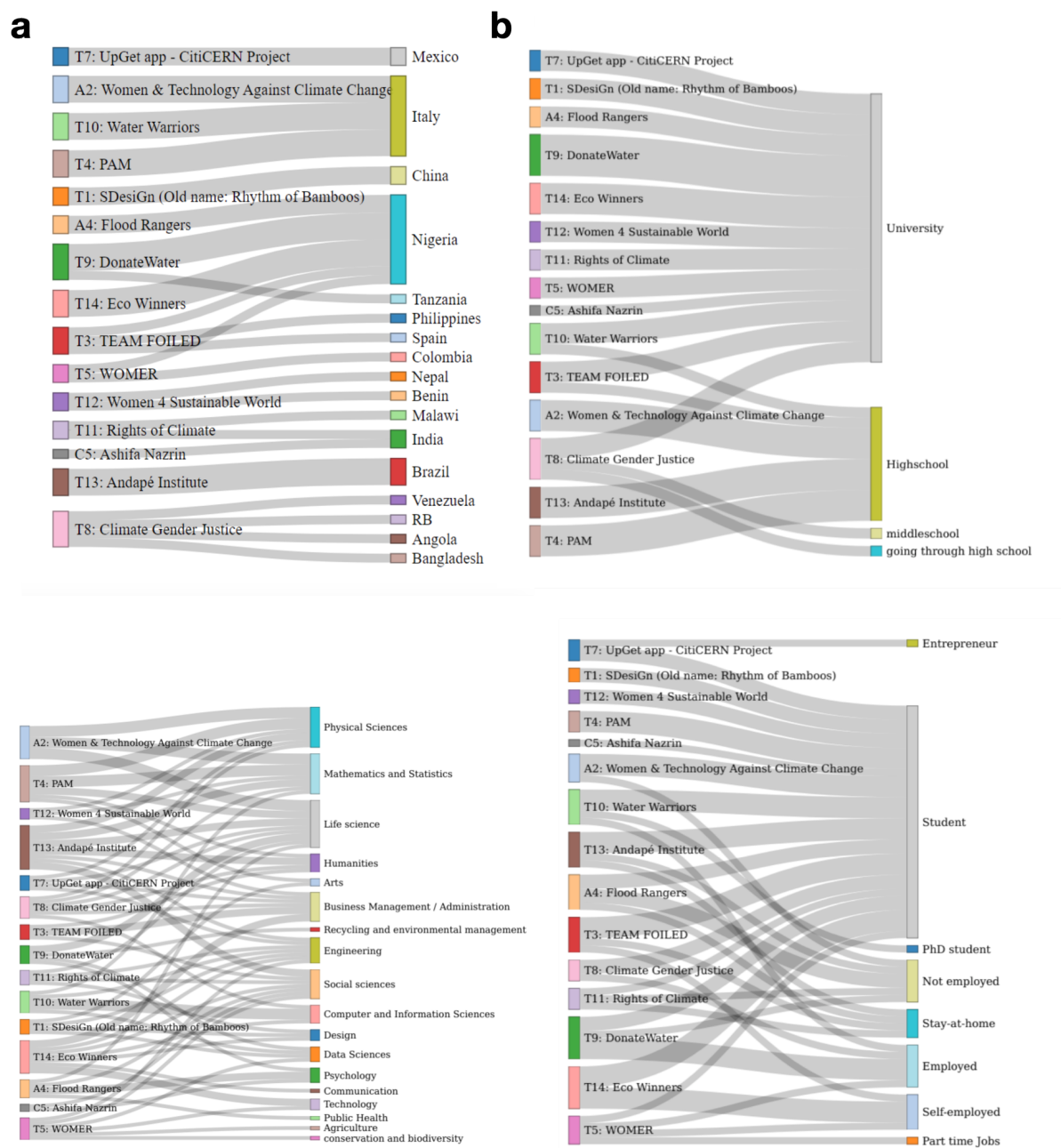
Who did you seek advice from during the previous block?

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

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.

Figure S1: Screenshots of the CoSo interface.



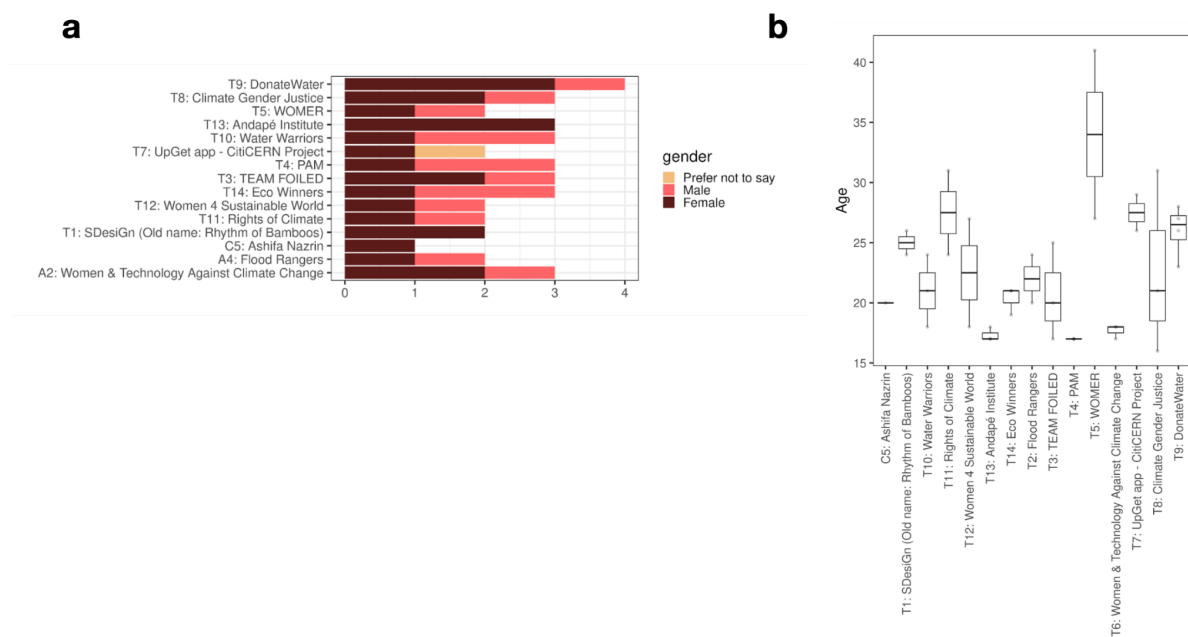


Figure S3: Gender (a) and age (b) distributions across teams.

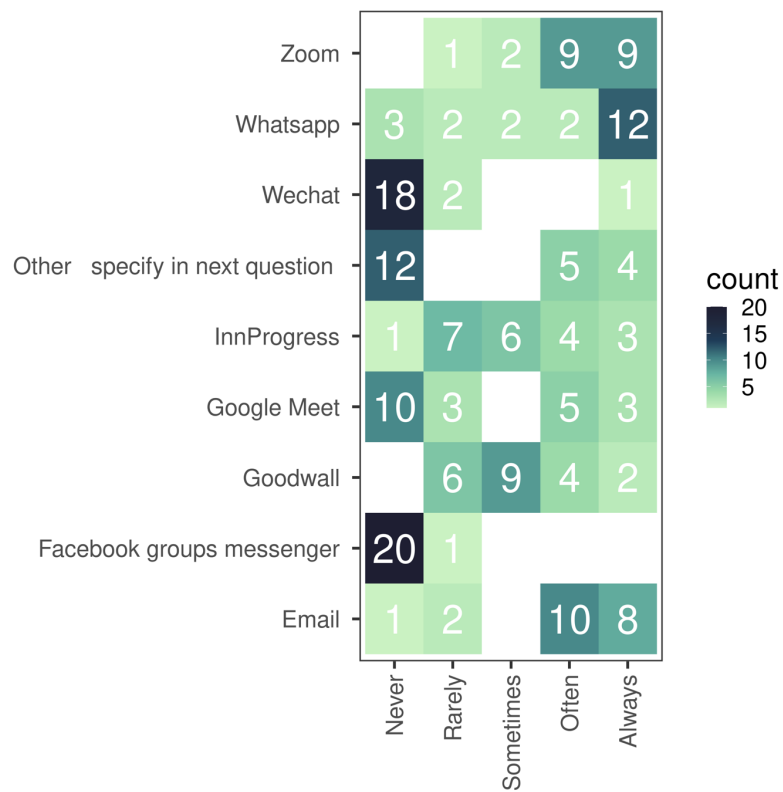


Figure S4: Communication tools reported to be used by teams to communicate. Number indicates number of answers across participants (total N=22 participants).

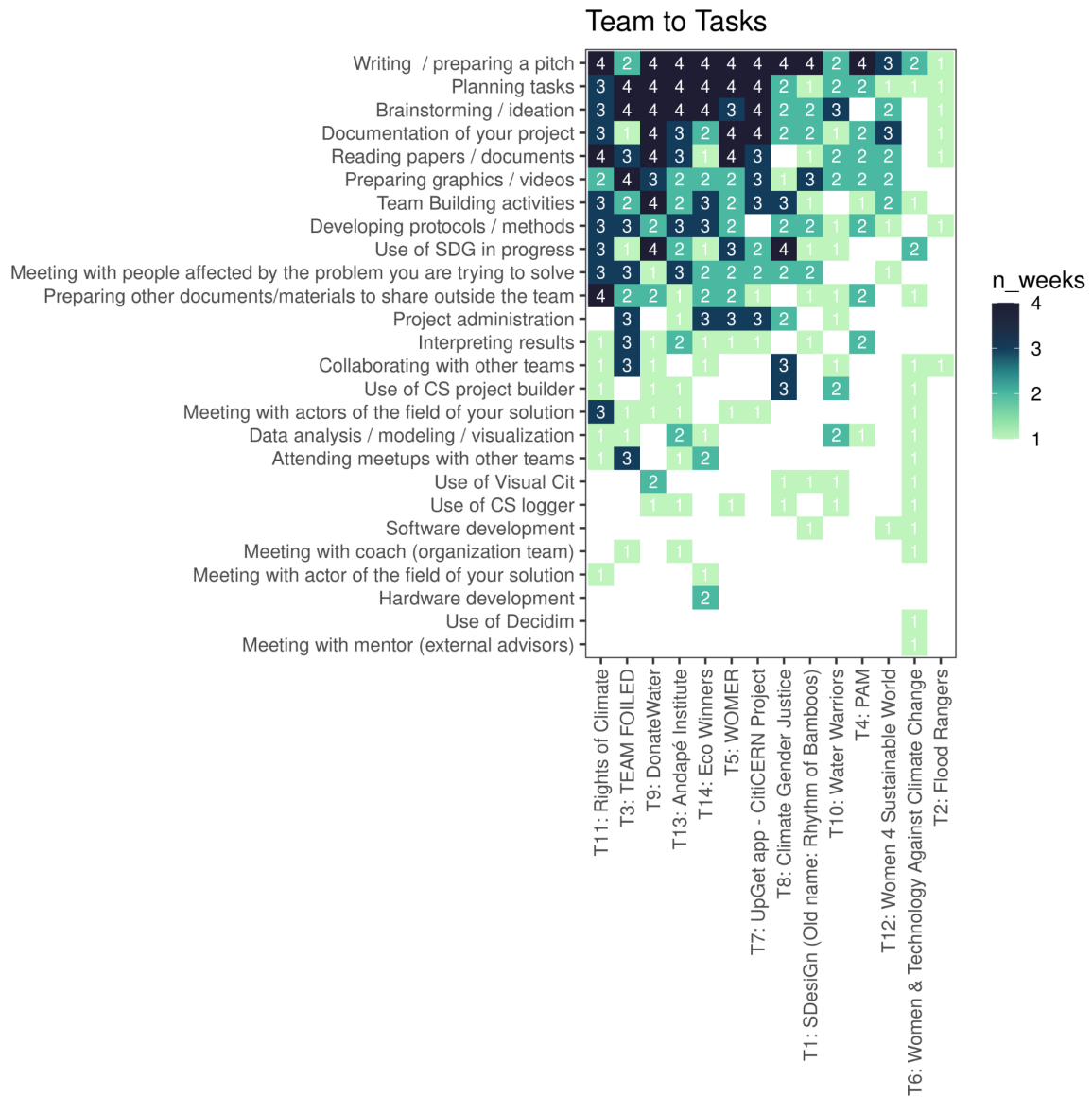


Figure S5: Heatmap indicating the number of weeks each activity has been reported by a given team, across 4 weeks. Activities and teams are ranked by row and column sums respectively.

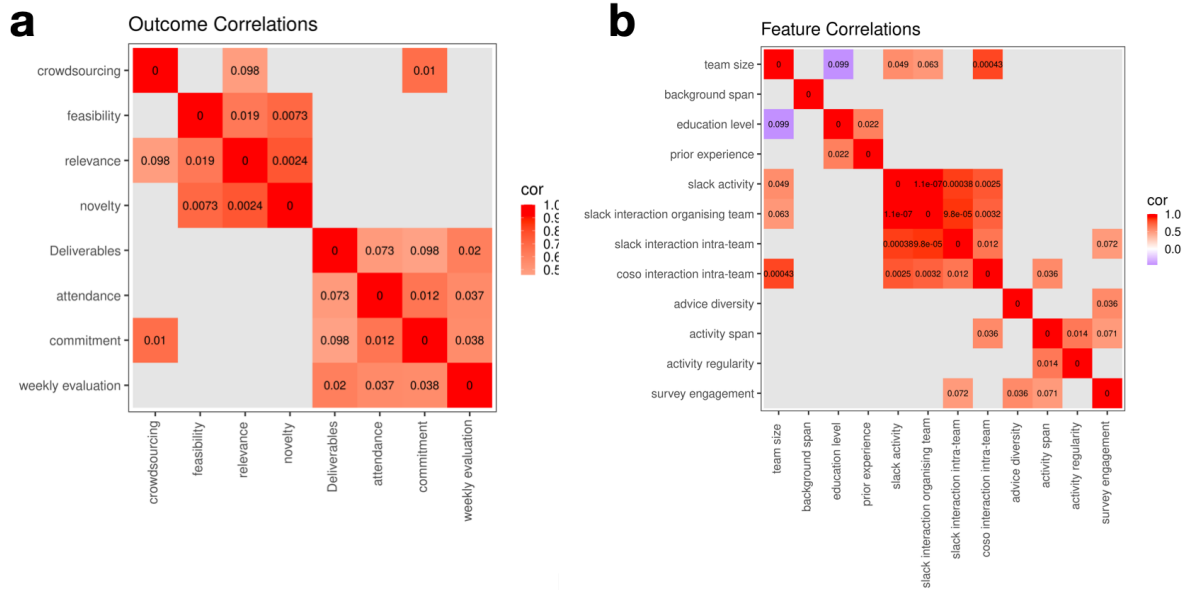


Figure S6: Correlation matrices of the features shown in Fig 7, for evaluations (a) and team features (b). Numbers correspond to p-values of the correlations. We grayed out cells with a p-value $p > 0.1$. We find two groups of evaluations: outcomes (top left) and processes (bottom right). For team features, 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.