Important declarations

Please remove this info from manuscript text if it is also present there.

Associated Data

Data supplied by the author:
Data is available at Dryad at the following link:
https://datadryad.org/stash/share/Gzddl9okhlq1RCRP9RP69qwmw39xZGSzTEyttFSaOkI

Required Statements

Competing Interest statement:
The authors declare that they have no competing interests.

Funding statement:
Funds from National Science Foundation Advanced Informal STEM Learning grant no. 1713562 supported this research.

Exploring connections among projects across the citizen science landscape: A social network analysis

Sara E Futch [1] , **Lincoln R Larson** [Corresp., 2] , **Caren B Cooper** [3] , **Bethany B Cutts** [2] , **Bradley Allf** [3] , **Maria V Sharova** [4] , **Darlene Cavalier** [5] , **Cathlyn Davis** [6]

[1] The Nature Conservancy, Maitland, Florida, United States

[2] Department of Parks, Recreation, & Tourism Management, North Carolina State University, Raleigh, North Carolina, United States

[3] Fisheries, Wildlife, and Conservation Biology Program, North Carolina State University, Raleigh, North Carolina, United States

[4] Thriving Earth Exchange, American Geophysical Union, Washington D.C., United States

[5] School for the Future of Innovation in Society, Arizona State University, Tempe, Arizona, United States

[6] Center for Environmental Science, University of Maryland, Frostburg, Maryland, United States

Corresponding Author: Lincoln R Larson
Email address: lrlarson@ncsu.edu

The rapid rise of citizen science is supported by web platforms that host and/or aggregate thousands of projects (e.g., SciStarter.org). While research on citizen science volunteers has historically focused on single projects, emerging research suggests many volunteers engage in multiple projects. Platforms like SciStarter enable exploration and characterization of volunteer activity across multiple projects. To learn more about the phenomenon of multi-project engagement, we carried out a descriptive social network analysis using digital trace data from the SciStarter platform depicting volunteer activity from Sept 2017- Dec 2018 on the SciStarter.org platform. During this time period, our sample included 624 citizen science projects, and 3,650 unique volunteers that engaged in these projects. On average, 73% of SciStarter volunteers joined multiple projects. For 94% of projects, at least three-fourths of the volunteers joined multiple projects. We used these data to visualize and analyze project connection networks that form when volunteers join multiple projects. Volunteers joined an average of 2.93 projects spanning many different scientific disciplines (e.g., topics such as Health & Medicine, Ecology & Environment, Astronomy) and modes of participation (e.g., online, offline). Volunteer engagement in citizen science produced a complex network of project connections with low network centrality, low levels of homophily and clustering, and ample evidence of boundary spanning (e.g. based on topic or mode). The projects most central in the network, which were also the most popular, were featured as affiliates on the website or in promotional email campaigns. By using a network approach to analyze digital trace data, our research illustrates the extent of multi-project, multi-disciplinary engagement on a third party platform, laying the groundwork for researchers and platform managers to explore and

facilitate multi-project engagement and its implications for the larger field of citizen science. Tools like social network analysis can help citizen science researchers identify pathways to deeper connections within citizen science and reveal new information about the important role that platforms play in the citizen science landscape.

Exploring connections among projects across the citizen science landscape: A social network analysis

Sara E. Futch[1], Lincoln R. Larson[2], Caren B. Cooper[3], Bethany B. Cutts[2], Bradley Allf[3], Maria V. Sharova[4], Darlene Cavalier[5], Cathlyn Davis[6]

[1] The Nature Conservancy, Florida, Maitland, FL, USA
[2] Department of Parks, Recreation & Tourism Management, North Carolina State University, Raleigh, NC, USA
[3] Fisheries, Wildlife, and Conservation Biology Program, North Carolina State University, Raleigh, NC, USA
[4] Thriving Earth Exchange, American Geophysical Union, Washington, DC, USA
[5] School for the Future of Innovation in Society, Arizona State University, Tempe, AZ, USA
[6] University of Maryland Center for Environmental Science, Frostburg, MD, USA

Corresponding Author:
Lincoln R. Larson[2]
4008L Biltmore Hall, NC State University, Raliegh, NC, 27695-8004, USA
Email address: lrlarson@ncsu.edu

Abstract

The rapid rise of citizen science is supported by web platforms that host and/or aggregate thousands of projects (e.g., SciStarter.org). While research on citizen science volunteers has historically focused on single projects, emerging research suggests many volunteers engage in multiple projects. Platforms like SciStarter enable exploration and characterization of volunteer activity across multiple projects. To learn more about the phenomenon of multi-project engagement, we carried out a descriptive social network analysis using digital trace data from the SciStarter platform depicting volunteer activity from Sept 2017- Dec 2018 on the SciStarter.org platform. During this time period, our sample included 624 citizen science projects, and 3,650 unique volunteers that engaged in these projects. On average, 73% of SciStarter volunteers joined multiple projects. For 94% of projects, at least three-fourths of the volunteers joined multiple projects. We used these data to visualize and analyze project connection networks that form when volunteers join multiple projects. Volunteers joined an average of 2.93 projects spanning many different scientific disciplines (e.g., topics such as Health & Medicine, Ecology & Environment, Astronomy) and modes of participation (e.g., online, offline). Volunteer engagement in citizen science produced a complex network of project connections with low network centrality, low levels of homophily and clustering, and ample evidence of boundary spanning (e.g. based on topic or mode). The projects most central in the network, which were

also the most popular, were featured as affiliates on the website or in promotional email campaigns. By using a network approach to analyze digital trace data, our research illustrates the extent of multi-project, multi-disciplinary engagement on a third party platform, laying the groundwork for researchers and platform managers to explore and facilitate multi-project engagement and its implications for the larger field of citizen science. Tools like social network analysis can help citizen science researchers identify pathways to deeper connections within citizen science and reveal new information about the important role that platforms play in the citizen science landscape.

Introduction

As a collaboration between the science community and members of the public, citizen science enables researchers to answer questions and collect data that would be difficult to answer and access with traditional research methods (e.g. Bonney et al., 2009; Chandler et al., 2017; Cooper et al., 2007; McKinley et al., 2017). Citizen science also functions as an outreach tool to accomplish educational goals (Bonney et al., 2016; Jordan, Ballard, & Phillips, 2012; Peter et al., 2021; Phillips et al., 2018) and may influence volunteer behavior (e.g. Toomey & Domroese, 2013; Peter, Diekötter, & Kremer, 2019). Over the past few decades, the number of citizen science projects has rapidly grown (Parrish et al., 2018), as has research on the practice of citizen science (Jordan et al., 2015). Although citizen science research does look across multiple citizen science projects, it can be difficult to access sufficient data to get a landscape view of the field. Because of this, researchers are still learning about citizen scientists themselves (Peter et al., 2021) and patterns of volunteer participation across different projects (Allf et al., 2022).

The conventional paradigm for managing citizen science volunteers is at the individual project level, focused on managing volunteers while they are in service to a particular project (Sharova, 2020). Additionally, most investigations of the practice of citizen science have generally focused on single projects, with a few notable exceptions that examined volunteer activity patterns across multiple projects on an online platform (Allf et al., 2022; Herodotou et al., 2020; Ponciano & Pereira, 2019). Compared to individual project-based management, platform-based management is more volunteer-centric, focused on managing volunteers holistically and beyond the bounds of a single project (Allf et al., 2022). And, although many citizen science projects operate in isolation (Bonney et al., 2014), many volunteers are already engaging in multiple projects (Allf et al., 2022; Hoffman et al., 2017; Ponciano & Pereira, 2019), a behavior that is associated with increased durations of engagement with citizen science compared to single-project volunteers (Ponciano & Pereira, 2019). Platform-based research therefore provides the landscape-level view needed to generate broader insights into the volunteer experience as well as the potential for coordinating the management of volunteers.

With widespread availability of information and community-oriented technologies, such as Web 2.0 and smartphones, the operation of modern citizen science projects requires a sufficient amount of digital infrastructure (Brenton er al, 2018). Web and phone application (app) platforms provide many citizen science projects with the necessary digital infrastructure for obtaining data from volunteers as well as managing the subsequent databases. Platforms that support project capacity in these ways include Citsci.org, Zooniverse.org, Anecdata, iNaturalist.org, and more. Thus, the landscape of citizen science includes projects using custom-created digital infrastructure as well as the shared infrastructure of various platforms and apps. Some prior studies have examined volunteers engaged in different projects across shared infrastructure of a platform (Herodotou, et al. 2022). The use of application programming interfaces (APIs) allows the secure transfer of user information between platforms and apps. SciStarter.org is a web platform structured as a searchable clearinghouse of citizen science projects, the majority of which operate with their own infrastructure or with the infrastructure of platform or app. SciStarter members navigate from SciStarter to the projects they choose to join. In addition, some citizen science projects use the SciStarter API to share information about the activity of SciStarter members in their projects. Thus, digital trace data from SciStarter contains information about volunteer engagement across a broad range of thousands of projects, irrespective of their source infrastructure.

Although online platforms already play an important role in the citizen science landscape, researchers are still defining what this role is, how it can be assessed, and how volunteers benefit from participating in citizen science through online platforms (Herodotou, et al. 2022). Our research seeks to understand the role platforms play by mapping and analyzing multi-project engagement on one platform, SciStarter.org. First, we mapped the landscape of projects connected by shared volunteers to visualize multi-project engagement on the website. Next, we assessed the attributes of projects within this landscape to understand how volunteer behavior differs between projects with different attributes across the entire SciStarter platform.

Social Network Analysis: A Tool for Exploring Project Connections. As with many online platforms, volunteer activity on SciStarter is measured through digital trace data, which is a digital record of the activity on the website. These data allow managers and researchers to map users' activity patterns over time. Using tools such as social network analysis, researchers can visualize and analyze the resulting network of engagement across projects within the large SciStarter.org repository. Social network analysis (SNA) is a research paradigm that allows for formalization and characterization of social units within a larger network (Borgatti et al., 2009; Wasserman & Faust, 1994). A short glossary of SNA terms relevant to this research is included in S1 Appendix. SNA has been used to examine connections and interactions among citizen science volunteers, but often within the context of a specific project (Amarasinghe et al., 2021). Few studies have examined volunteer networks across multiple projects (Herodotou et al., 2020).

Social networks are made up of actors (nodes) and the relationships between them (edges). In one-mode networks, all nodes in the network are of the same category; for example, families connected by marriage (Padgett, 1994). In two-mode networks, also called bipartite networks, nodes are of two different categories; for example, parties and the people who attend them (Davis et al., 1941; Freeman, 2003) or, in the current case, citizen science projects and the volunteers who join them. Bipartite networks can also be projected into one-mode networks (Herodotou et al., 2020; Pettey et al., 2016; Sankar et al., 2015) that, with thoughtful interpretation (Opsahl, 2013), can offer insights into patterns of co-incidence in a large network (Pettey et al., 2016). For example, in the citizen science landscape, we can project the bipartite network of project and volunteer connections into a network to show the project connections formed when a volunteer joins multiple projects. Descriptive analysis of this network can reveal what characteristics unite the different projects that a volunteer joins. Researchers can test for homophily (similar projects connected to each other) and clustering (distinct groups of interconnected projects) to determine if volunteers tend to join similar groups of projects. Researchers can also test for centrality to determine what projects are the most connected and show the most multi-project engagement. Exploring connections between projects is a critical new direction in citizen science research, where multi-project engagement by volunteers is already happening but is poorly understood (Allf et al., 2022; Herodotou et al., 2020; Hoffman et al., 2017).

Attributes that Might Influence Project Connections. When volunteers view projects on a platform, they may be attracted to intrinsic attributes of the project, such as the scientific topic of the project or the process for participating. Additionally, they may be affected by extrinsic attributes of the project, such as how the project is featured on the online platform. We were interested in the influence of intrinsic and extrinsic project attributes on volunteer activity across the platform.

We investigated two types of intrinsic project attributes (topic, mode) and two types of extrinsic project attributes (affiliate status, campaign feature status) to determine how volunteers are currently engaging with projects across and within attributes. Topic refers to the scientific discipline of the project. Projects on platforms such as SciStarter.org and Zooniverse.org are commonly grouped by topic or scientific disciplines, which might include themes such as the environment, public health, and astronomy. Other typologies of citizen projects often focus on methods of participation, although the specific words used to describe these different methods may vary. For instance, Parrish et al. (2018) refers to the division as “hands-on” and “virtual,” and Sharova (2020) refers to “outdoor” and “indoor” projects. In the current research, we chose online versus offline to reflect the online status of SciStarter.org. Topic and mode are attributes intrinsic to the project and are not manipulated by the SciStarter Platform. Alternatively, project affiliate status and campaign feature status refer to the statuses given to projects by the SciStarter platform and are thus extrinsic to the projects. Affiliate projects on SciStarter.org share data contribution information with SciStarter and in return are marked with an icon indicating their status as affiliates. A project’s campaign feature status is a result of whether it has been featured in an email campaign to SciStarter members. Collectively, these intrinsic and extrinsic attributes could influence overlap of volunteers across the larger citizen science landscape.

Research Objectives. The first objective in our exploratory study was to describe the structure of the project connection network from a snapshot in time on Scistarter.org (Obj. 1). We aimed to describe network structure by whole network measures, such as clustering and centralization, and the node-level measures of centrality (definitions provided in S1 Appendix). For our second objective, we were interested in quantifying the patterns of participation within and across project attribute groups to determine if homophily is present in the network and if intrinsic (topic, mode) and extrinsic (affiliate status, campaign status) project attributes influence volunteer overlap among different projects (i.e., multi-project participation; Obj 2). Our third and final objective was to discover the projects central to the network, as these central projects may be important stakeholders in the citizen science landscape because their volunteers are likely to be highly involved across projects. We explored project centrality in two ways: first determining which projects are central to the projection network (Obj. 3a) and then examining project attributes associated with centrality (Obj. 3b). By addressing these objectives, we sought to describe and map the patterns of multi-project participation on SciStarter.org, building important foundational knowledge about the current state of multi-project participation in citizen science.

Materials & Methods

SciStarter website. SciStarter.org hosts a web platform and database of thousands of citizen science projects that provides volunteers with a wide choice of projects and creates opportunities for volunteer-centric management. At the time of data collection, SciStarter managed a database of user-generated content about each citizen science project, a Project Finder application that provided advanced functionality for searching for projects, and a syndicated blog to promote citizen science projects, often facilitating volunteer connections to and participation in multiple projects (Allf, et al., 2022). The Project Finder application allowed interested volunteers to search for projects by various tags including keyword, topic (scientific discipline), mode of participation (“exclusively online”, “on a lunch break”, “at school”, etc.), location, and target age group. Volunteers can become SciStarter members by creating a free account, and making a user profile in which they click projects to “join.” These are saved, thus allowing for tracking of engagement in projects within or separate from the SciStarter web platform.

All projects on the site have a project page that shares information about the project and how to participate. Project pages are searchable through the project finder tool. However, not all projects on the SciStarter website are presented in the same way. Some projects are highlighted because they are SciStarter ‘affiliates’, which means that they use SciStarter’s API to share information about the frequency of volunteers’ data contributions to their project. This is then displayed on the volunteer’s profile on SciStarter. These affiliate projects are marked with an icon in the search results and on the project page and appear first in search queries. Another way that members find projects is through monthly emails that feature certain projects that fit the theme of that month’s campaign. For instance, an email in October might feature projects that align with a Halloween theme, such as projects about bats or spiders. These differences in how projects are viewed or appear on the website could affect volunteer choices of projects, and thus the project’s position in the larger project connection network.

Defining project attributes. As noted, we wanted to learn more about how project differences may affect their network position, and so we collected information related to four project attributes: topic, mode of participation, SciStarter affiliation status, and campaign feature status.

To code project topics (topic), we applied the typology of 14 project topics (S2 Appendix) used in Allf et al. (2022) and Sharova (2020). Postings that were determined to not be citizen science projects, including postings for tools (such as spectrometers) and postings for other citizen science platforms (such as for the website CitSci.org), were dropped from the sample. In some analyses, we grouped project topics into four larger groups (super-topics) so that we could more clearly test differences between these larger disciplines: Earth & Life Sciences, Behavior & Social Sciences, Engineering & Physical Sciences, and Health & Medicine. The compositions of these super-topics are also shared in S2 Appendix.

Mode of participation (mode) was coded by the authors using expert review with the following definitions. Offline projects are projects where the primary mode of data collection is offline. Projects in which data collection is offline but submitted through an online platform (e.g. iNaturalist) are considered offline projects. Online projects are projects where the primary mode of data collection or classification is online (e.g. Stall Catchers). For both topic and mode classification, we used project descriptions available on SciStarter.org. Additional information from project websites was used if needed. Projects with no further information and three or fewer instances of engagement (joins or bookmarks) during the year of data collection were excluded from the sample.

We coded a project's affiliate status and campaign feature status using records from SciStarter.org. Because of the longitudinal nature of data collection, we assigned different status levels based on when the project became an affiliate or when it was featured in monthly email campaigns. Affiliate status was separated into three groups: (1) projects that were never a SciStarter affiliate (Not Affiliate), (2) projects that became an affiliate sometime during the data collection period (Part-time Affiliate), and (3) projects that were an affiliate throughout the entire data collection period (Full-time Affiliate). A project's campaign feature status tells us when that project was featured in one of the monthly emails that go out to volunteers. The campaign feature status was separated into four categories: (1) projects that were never featured in a SciStarter promotional campaign, (2) projects featured only before data collection, (3) projects featured only during data collection, and (4) projects featured in promotional campaigns both before and during data collection.

SciStarter digital trace data. Digital trace data for this research was extracted from SciStarter.org (existed as SciStarter.com at the time data collection was initiated). On Dec 6th, 2018, our research team received anonymized digital trace data of SciStarter members' activity on the website between Sept 19th, 2017 and Dec 3rd, 2018. Sept 19th, 2017 marked the launch of SciStarter 2.0 (Hoffman et al., 2017), which introduced the added functionality of member accounts, dashboards, and profile pages. Use of these secondary data was approved by the NC State University Institutional Review Board (IRB Protocol # 20934) prior to analysis. There was very little demographic data included, as we only had access to what volunteers willingly entered into their volunteer profile; 6.87% of volunteers provided their birth year, and 6.25% provided a gender. Consequently, we were unable to investigate associations between volunteer demographic characteristics and multi-project participation in this analysis, though that remains an important direction for future research.

At the time of data collection, volunteers could click buttons to either "join" or "bookmark" projects of interest on the SciStarter website (Fig 1). Additionally, volunteers could check a box to indicate that they had previously joined a project. Clicking "join" sent the volunteer to the project's website (unless that project was exclusively hosted on SciStarter), and it automatically added the project to a list of joined projects on a volunteer's profile. Thus, the join function in SciStater could be considered "conversion" (Crall et al., 2017), or an expression of interest in joining a project. Clicking "bookmark" added the project to a list of bookmarked projects on the volunteers' profile. Volunteer activities like "joins" and "bookmarks" were recorded in the digital trace data, along with an anonymized participant ID number.

As noted, some citizen science projects are affiliates that are directly hosted on the SciStarter website or use SciStarter's API. Because of this, we were able to track individual contributions to these affiliate projects – a deeper level of project engagement than we could track in non-affiliate projects. However, because a record of data contributions was only available for 30 of the 624 projects that were active during data collection, we used volunteers selecting the "join project" button as an indication of volunteers' behavioral intent to join a project and a proxy for project engagement (Ajzen, 1991) rather than data contributions. Analysis of the data contributions showed that "joining" a project is a better predictor of subsequent contributions than "bookmarking" a project; we were able to confirm that at least 30% of volunteers who clicked "join" on an affiliate project later contributed to that project, while only 10% of volunteers who clicked "bookmark" ended up contributing.

Data analysis. We conducted analyses in R (R Core Team, 2019) using RStudio (RStudio Team, 2016) and in SPSS (IBM Corp, 2017). We used social network analysis to better understand the relationships between projects connected by shared volunteers, which reveal volunteers' patterns of engagement across the citizen science landscape. To get deeper insight into project connectivity across this landscape (Obj. 1), we used the R packages igraph (Csardi & Nepusz, 2006) and sna (Butts, 2019) to create a one-mode projection of citizen science projects (nodes) connected by shared volunteers (edges), with weighted connections in proportion to the number of volunteers that two projects share. Whole network descriptive statistics such as component analysis, clustering, and centralization revealed the structure of the network (definitions provided in S1 Appendix).

We used an exponential random graph model (ERGM) to test for homophily in the network. Homophily is the tendency in networks for similar nodes to be connected to each other. We test for homophily to determine if volunteers tend to join projects with similar attributes, such as the same topic or method of participation (Obj, 2). ERGM is an approach to data analysis that allows researchers to statistically test the impact of connectedness and project attributes on how volunteers move among projects. We used the ergm R package (Handcock et al., 2018; Hunter et al., 2008) to develop a predictive model that allowed for link prediction, predicting which projects would be connected in the future, based on node attributes (Robins et al., 2007; van der Pol, 2019). We followed this with a Chi-square analysis of connections between projects of varying attributes to determine if observed instances of connection differed from what would be expected in a random distribution. This demonstrates whether certain attributes are more or less connected than would be expected in a random network.

We then determined which projects were central to the network (Obj. 3a) by calculating the centrality of the nodes in the project connection network. Using degree centrality (the number of connections from one project to other project nodes) as the dependent variable, we used a negative binomial regression analysis to explore how project attributes influenced project centrality (Obj. 3b). We chose a negative binomial regression for this analysis because the data were over-dispersed, with the variance exceeding the mean (Hilbe, 2011).

Results

Dataset description and description of project-based network projection. During the period of data collection, 3,650 SciStarter members joined 624 projects, resulting in 9,678 instances of project joins. The 624 projects spanned all project attributes, representing 14 different topics (scientific disciplines). The breakdowns of the attributes are shown in Table 1.

In 98.4% of these projects, at least one volunteer also joined other projects and, in 93.9% of projects, at least three-fourths of the volunteers joined other projects. The 624 projects had an average of 17.1 volunteers, with a minimum of 1 and a maximum of 480 volunteers (note that these numbers do not represent the total number of volunteers within a given project but only those with SciStarter accounts). The distribution of projects was skewed, however, as 476 projects had 10 or fewer volunteers (Fig. 2). The digital trace data showed that volunteers joined an average of 2.9 projects, with a median of two joins, minimum of 1 and maximum of 34 project joins; 73% of all active volunteers joined 2 or more projects, and 46% joined three or more projects. These patterns align with self-reported participation among SciStarter members (Allf et al., 2022).

The project connection projection network (hereafter project network) shows relationships among the various citizen science projects on SciStarter (Fig. 3). In this network, the nodes represent the 624 projects, and an edge between two nodes represents the connection made when one or more volunteers joins both projects. Node color represents project topic, and node shape represents project mode. Only 10 of the projects in the project network are isolates, meaning that they are not connected to any other projects. The remaining 614 projects (98.34% of all projects) share volunteers with at least one other citizen science project. The 614 project nodes make up one connected component. The longest geodesic distance between any two projects is 7, meaning that there is a maximum of 7 edges, or different links representing shared volunteers, between projects. A modularity-optimization-based cluster analysis (Blondel et al., 2008) did not reveal significant clustering in the network. Clauset, Newman, & Moore (2004) suggests the result of such an analysis be at least 0.3 to indicate clustering, the present network has a modularity of 0.2, meaning that we are not able to detect distinct communities of densely connected nodes.

Table 2 displays descriptive statistics for the connected component in the project network. These descriptive statistics again highlight the high interconnectedness of the network. In particular, the relatively high transitivity compared to the density (S1 Appendix) emphasizes the interconnectedness at the center of the network.

Exponential random graph modeling. We used exponential random graph modeling (ERGM) to determine if links in the network could be predicted using project attributes. We aimed to determine if projects with similar attributes (topic, mode, affiliate status, and campaign feature) were more likely to share volunteers, which would result in homophily in the network. Neither the original bipartite (two-mode) nor the projection (project-mode alone) network produced a non-degenerate model (see S3 Appendix for a more in-depth explanation). The persistent degeneracy of the models suggests a lack of statistically significant homophily in the network, meaning either that volunteers are regularly joining projects across different attributes, or that volunteer behavior cannot be reliably predicted based on the project attributes investigated in this study (topic, mode, affiliate status, and campaign feature).

Chi-square analysis of distribution of direct connections. To dig deeper into the patterns of connections across and within different groups of attributes, we calculated the number of connections between projects with different attributes (e.g., connections between online and offline projects). Connections throughout the network mean that the two connected projects share at least one volunteer. For this analysis, we compared the observed percentage of connections between each attribute type with what would be expected in a random distribution of connections. Table 3 shows the differences in observed connections (two projects sharing at least one volunteer) between each type of attribute compared to what would be expected in a random distribution based on the number of projects in each attribute group.

Chi-square analyses revealed that the observed percentages differ from what would be expected in a random distribution for the project topics, X^2 (9, N = 19,356) = 304.21, $p < .001$, modes of participation, X^2 (1, N = 19,356) = 304.5, $p < .001$, affiliate status, X^2 (9, N = 19,356) = 59.03, $p < .001$, and campaign feature status, X^2 (9, N = 19,365) = 84.67, $p < .001$. This suggests that, although there is no statistically significant homophily at a whole network level (the tendency for similar projects to be connected to one another) there are patterns of connection between projects when looking attribute by attribute. For instance, online projects are connected to other online projects 24% percent more, on average, than what would be expected in a random distribution (Table 3). Negative percentages in Table 3 indicate fewer connections than would be expected in a random distribution. For the intrinsic attributes of project topic and project mode, we see that, in general, project connections between two projects with matching topics or matching modes are proportionately more likely than connections between projects with different topics or modes, with the exception of Earth & Life Sciences (likely because many

SciStarter project were offline Earth & Life Science projects with minimal connections to other projects). When looking at the matrices for the extrinsic attributes (affiliate status and campaign feature status), however, we see a different pattern. The largest positive numbers (i.e., connections) were associated with projects that had Full-time Affiliate status or those that were featured in campaigns Before and During Data Collection. In addition to these patterns, we continue to see connections across all attributes, underscoring the high level of boundary spanning (based on discipline, mode, etc.) seen in the project network.

Centralization and centrality in the project network. We analyzed the centralization and centrality of the large component of the project network (see Fig 2), excluding the ten isolate projects, to see if certain projects were more prominent nodes of connectivity than others. We first analyzed the centralization across the large component (network without isolates). These scores are normalized measures - the closer they are to 1, the more centrality is present around one node (S1 Appendix). The degree centralization for the project network was 0.43, demonstrating that degree centrality is not centered on only a few projects, and that many projects are well-connected in the network. Betweenness centralization in the project network was only 0.08, suggesting that most projects play an equal role in connecting other projects. This supports the lack of clustering we found, suggesting there are not disparate groups of projects in the network. The project network presented high eigenvector centrality at 0.84, meaning that a few highly connected projects are well connected to each other. Figure 4 highlights the strongest connections in the network, showing only edges with 10 or more shared volunteers. This network illustrates the high interconnectedness, and high degree of boundary spanning based on topic and mode, at the center of the network.

The most connected projects in this network were also the most popular projects. We found a significant correlation, $r(624) = .87$, $p < .001$, between project popularity, as measured by the number of volunteers that joined a project, and project degree centrality, as measured by the number of projects to which a project is directly connected (Fig. 5). In projects with 10 or more volunteers, more than 90% of volunteers joined other projects, showing that multi-project engagement is prevalent throughout the network, and especially common within these popular projects. This suggests that it is rare for a project to be both popular and isolated from the volunteers of other projects.

Project attributes associated with project centrality. We ran a negative binomial regression model to determine the influence of project topic, mode, affiliate status, and campaign promotion status on the degree centrality of projects in the network. Degree centrality, the dependent variable in this model, reflects the count of connections between a project and other project nodes in the network (i.e., projects with more connections to other projects had higher degree centrality). All attributes were significant in predicting degree centrality (Table 4). Online projects were more connected to other projects than offline projects ($B = -.788$, $p < .001$). Among the project topics, Health & Medicine focused projects were more connected than Earth

& Life Science projects (B = -.705, p < .001) or Engineering & Physical Sciences projects (B = -0.858, p < .001). Full-time Affiliates were more connected than projects that were Not an Affiliate (B = -1.193, p < .001), but there was no significant difference based on when the project became an affiliate (Fullt-time versus Part-time). Projects featured in campaigns were also more connected, with projects featured both Before and During data collection associated with more volunteer connections than projects only Featured Before data collection (B = -.570, p = .003) and projects only featured During data collection (B = -.540, p = .017). Degree centrality was lowest among projects Never Featured in a SciStarter campaign (B = -.949, p < .001).

Discussion

Using a social network analysis with digital trace data from the SciStarter platform, our study explored the degree to which projects are connected by shared volunteers and the key role that third-party platforms can play in the citizen science landscape. Our results support recent studies showing that citizen science volunteers engage across multiple projects (Allf et al., 2022; Herodotou et al., 2020; Hoffman et al., 2017; Ponciano & Pereira, 2019). In 94% of projects, three-fourths of the volunteers joined other projects. Volunteers on SciStarter joined more than two projects on average, and 73% of volunteers joined 2 or more projects.

The phenomenon of multi-project engagement in citizen science is important to consider for the sake of both the volunteer experience and project management. Previous research on platforms has shown that volunteers engaging with multiple projects tend to stay on platforms longer than those volunteers that engage with a single project (Ponciano & Pereira, 2019). Longer engagement with a platform could indicate more prolonged engagement with citizen science as a whole. This multi-project engagement presents more opportunities for volunteers to build skills across projects and advance learning outcomes (Phillips et al., 2018; Peter et al. 2019). Platforms also benefit project managers, providing cyber-infrastructure, access to volunteers, and a chance for collaborative learning to advance best practices across projects (Newman et al., 2012). For example, research suggests that third-party "enabler" organizations - whether in-person institutions or online platforms - can strengthen volunteer engagement with science in multiple forms (Salmon et al., 2021). Given the potential benefits of multi-project engagement, it is important to understand how projects are connected by shared volunteers.

Our projection depicting the network of citizen science projects connected by shared volunteers is an interconnected network with one large component. This network presented minimal clustering and homophily, meaning that at a whole network level, similar projects are not predictably sharing volunteers, and there are not clear groups of projects that all share the same volunteers. It appears that citizen science volunteers are open to exploring new project types and often cross over to participate in projects with different attributes such as topic, mode, affiliate status, and campaign feature status, opening the opportunity for transdisciplinary citizen science (Spasiano, et al., 2021). A cluster analysis on a similar dataset from the citizen science platform Zooniverse also found minimal evidence for clustering (Herodotou et al., 2020), suggesting that the lack of clustering among projects may be a common trait shared by many citizen science platforms.

Looking more closely at this intermixing, we do see some patterns of participation emerge on an individual attribute level (Table 3). For the intrinsic attributes of project topic and project mode, we see that project connections between two projects with matching topics or modes are proportionally more likely than connections between projects with different topics or modes. So, there is a slight tendency for volunteers to join multiple projects within the same general topic or mode of projects, but only within intrinsic attributes (topic and mode) at the individual attribute level. When looking at the whole network, this tendency fades away. More research is needed to investigate the strength of these patterns when multiple attributes are considered.

A different pattern emerges for the extrinsic attributes: affiliate status and campaign feature status. For all extrinsic attribute levels, the proportionally most likely connections are with projects that hold Full-time Affiliate status, or projects that were featured in campaigns Before and During Data Collection. We were not able to look into whether there was a causal relationship between connections and project features as part of this research. The high connectivity of featured projects may underscore the influence that platforms can have on multi-project engagement, potentially helping volunteers discover and connect with new types of science. However, it is also possible that SciStarter may only promote particularly popular projects. Future research should explore the patterns of participation as well as any causal relationships more.

The significant correlation between centrality and popularity (based on number of joins) showed that popular projects were central to the SciStarter network, and the skewed distribution of joins revealed that a few projects were responsible for most of the activity on SciStarter. High eigenvector centrality and high clustering coefficients compared to density measures also demonstrate that these central, popular projects were highly connected to each other. Regression results suggested that the centrality of these projects was in part due to their topic (health and medicine projects were most popular) and mode (online projects were popular). Project popularity was also linked to SciStarter featuring these projects as affiliates and in email campaigns. Projects that are featured by SciStarter in these ways received higher numbers of joins than projects that are not featured. Again, in this study we were not able to investigate whether project features led to popularity or popularity led to a project being featured. A more advanced understanding of what makes projects popular could address the “nibble and drop” problem in citizen science (Fischer, Cho, & Storksdieck, 2021). Additionally, since the collection of the data in this study, SciStarter has implemented an intelligent recommendation tool that recommends projects volunteers might be interested in based on past data, which has led to increased participation among volunteers shown these recommendations (Zaken, et al., 2021). Future research could explore the patterns of participation resulting from the new recommendation tool to learn more about the influence of third-party suggestions and features.

The interconnected network on SciStarter and the suggested influence of platform activity on project popularity demonstrates the role that platforms play in the citizen science landscape. Many of the projects connected by shared volunteers in this network are not featured together in any location other than the SciStarter website. By sharing the cyber-infrastructure, volunteer recruitment efforts, and volunteers themselves, platforms like SciStarter not only serve as “enabler” organizations (Salmon et al., 2021) but also foster shared management practices that allow for conservation of resources, including the resource of volunteer energy and maximization of scientific and volunteer outcomes (Brudney & Meijs, 2009; Newman et al., 2012; Sharova, 2020). By acknowledging and encouraging the boundary-spanning engagement revealed by our social network analysis, project and platform managers can effectively leverage the phenomenon of multi-project participation in citizen science, potentially extending the impacts of citizen science in society (Lynch-O’Brien, et al., 2021; McKinley et al., 2017).

From a project management perspective, our results suggest less of a need for competition among project leaders to recruit from a scarce pool of potential volunteers; rather there is evidence for abundant opportunities for coordinating the management of volunteers across projects in a more collectivist manner. This may be welcome news to the broader field of citizen science, which is currently struggling with recruitment challenges (Allf, et al., 2022; Fischer, Cho, & Storksdieck, 2021). Just as a project brings together volunteers for a shared goal, platforms such as SciStarter can bring together projects and overlapping volunteers for goals that span individual projects, such as supporting volunteer learning or building social capital. Additionally, managers who set project spanning goals could seek to facilitate retention of volunteers within the citizen science field, as coordination among projects could allow for differing designs that respond to volunteers’ shifting motivations over time (Aristeidou, Scanlon, & Sharples, 2017; Larson et al., 2020).

Limitations. Digital trace data is a helpful source of information, but it is “found” data, in that it was not created for this study and is thus not expressly matched to the intent of the study (Howison, Wiggins, & Crowston, 2011). As such, there are several limitations that should be acknowledged. We only had access to data from SciStarter members, which excluded the activity of an unknown number of volunteers that use the SciStarter website without a membership. Restriction to SciStarter members means that we excluded volunteers engaged in projects outside the SciStarter infrastructure, who represent the majority of volunteers in a given project. For example, large projects such as iNaturalist have thousands of volunteers, only a small portion of whom are also SciStarter members (our study only captured activity for 182 iNaturalist volunteers). Because we were limited to monitoring activity on the SciStarter website, our best measure of project engagement came through the use of the “join” button on SciStarter. Joining a project in this way does indicate behavioral intention to engage with a project, and behavioral pledges are associated with actual behavior in other contexts (Costa, Schaffner, & Prevost, 2018; Katzev & Wang, 1994). Thus, project joins could be considered a reliable path to conversion and ultimate participation (Crall et al., 2017). Nevertheless, for the majority of projects on SciStarter, we have no record of volunteers’ data contributions after they clicked the join button. Within the present digital trace data, we see that about 30% of joins in SciStarter lead to project contributions. This estimate may be low, as volunteers would have to be logged in through SciStarter for their contributions to be accurately recorded. However, across other citizen science projects, a significant portion of active participants do not contribute data on a regular basis, confounding efforts to estimate participation rates (Cooper, et al, 2017; Fischer, Cho, & Storksdieck, 2021). Given these challenges, using “joins” as a proxy for project engagement remains among the best options available.

We also have little data on volunteer demographics, as we only have access to what volunteers provided in their SciStarter profiles (most of which were incomplete). This prevented us from exploring demographic attributes in the bipartite network and from associating certain demographic information with the behaviors observed in the data. Other studies have demonstrated that, within single projects, information sharing within social networks varies based on the type of people participating (e.g., volunteers vs. moderators vs. scientists), underscoring the importance of accounting for heterogeneity across networks (Amarasinghe et al., 2021). Also of note, data for this project were collected from 2017-2018. Since this time, SciStarter has migrated to a new url (from .com to .org) and has expanded considerably, in terms of number of the registered users and projects, as well as platform capabilities. As of July 2022, SciStarter.org had approximately 135,000 members, growing from around 51,000 members in December 2018. Additionally, SciStarter.org has several new features that may influence the connections among projects, including AI-generated project recommendations(Zaken, et al., 2021; Zaken, et al., 2022), new microsites customized to feature combinations of citizen science projects for school systems or companies, and a feature called Lists which is free functionality to create a page with citizen science projects of one’s choosing, add personalized instructions, and access analytics to track progress of those accessing the page, such as for teachers to create assignments for their students and monitor their progress. SciStarter also updated the functionality to verify project joins and track more volunteer contributions among the rapidly growing number of affiliate projects. Patterns observed and described in our current social network analysis might therefore be expanded with the digital trace data from over 100,000 current SciStarter members. Lessons learned from these approaches, and similar analyses applied to other third party platforms (Herodotou, et al., 2020), could shed more light on the nature of multi-project, multi-disciplinary project participation across the broader citizen science landscape.

Conclusions

Although research is beginning to reveal the prevalence of multi-project participation among citizen science volunteers (Hoffman et al., 2017; Ponciano & Pereira, 2019; Allf et al., 2022), our study is among the first, along with Herodotou et al.'s (2020) study of project connections on Zooniverse.org, to use social network analysis techniques to describe connections of projects within the larger citizen science landscape. Unlike Herodotou et al’s (2020) study, our research is the first, to our knowledge, to use SNA to explore project connections across multiple apps, organizations, and platforms (via SciStarter), including numerous offline projects, simultaneously. These social network analyses based on digital trace data show that an interconnected landscape of citizen science projects exists, and that volunteers are active throughout it. Our work also demonstrates that online platforms can facilitate and influence connectivity and shared management of volunteers across the network of projects by strategically featuring and promoting certain groups of projects.

Future research could investigate the citizen science project connection network using data that accounts for a deeper level of engagement (beyond project “joins”). This might include self-reports of project engagement or digital trace data that includes project contributions. Additionally, other project attributes, such as location or type of engagement (contributory style, co-created, etc.; Shirk et al., 2012) could reveal additional influences on project popularity and centrality that influence volunteer recruitment and movement among projects. Research that is able to incorporate volunteer demographics would reveal who is currently engaging with online citizen science platforms (and who isn’t) and should be conducted to guide future project design and outreach efforts (Allf et al., 2022; Pateman, et al., 2021).

Additional research and practice could also incorporate the temporal component of citizen science project engagement, investigating the existence of “gateway” projects, or volunteers’ first projects that lead them into the world of multi-project engagement. Longitudinal research could also analyze individual learning trajectories to understand what multi-project engagement means in terms of sustained engagement with citizen science, skill development over time, and broadening connections to science (Ponciano & Pereira, 2019; Stylinski et al., 2020). In this vein, SciStarter has developed organization-specific portals that offer select projects that could guide citizen scientists along project pathways and emphasize particular learning outcomes (see the page built for Girl Scouts at https://scistarter.com/girlscouts/info). Experimental manipulation of platform management strategies (e.g., systematic selection of projects for campaigns, targeted features of selected projects in thematic portals) could reveal causal pathways that help to better explain volunteer recruitment and retention and create new multi-project engagement pathways.

Our work underscores the need for continued research with landscape-level projections to investigate how the larger network of projects and volunteers interact to influence broader participation outcomes. An enhanced understanding of connectivity across citizen science projects would help both project and platform managers leverage existing capacity and benefit from the resources (e.g. access to data, volunteer energy, and project infrastructure) this larger landscape provides, ultimately increasing the potential of citizen science to impact society.

Acknowledgements

We thank Joe Heimlich and Bruce Lewenstein for their input, Lisa Lundgren for contributing to and guidance in coding projects for this study, and we thank the SciStarter.org team for providing access to the anonymized digital trace data that made this project possible. Thanks are also due to citizen scientists on Scistarter.org for their contributions to hundreds of science projects that supported these projects and allowed us to better understand the landscape of projects.

References

Ajzen, I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50, 179-211.

Allf, B. C, Cooper, C. B., Larson, L. R., Dunn, R. R., Futch, S. E., Sharova, M., & cavalier, D. (2022). Citizen Science as an Ecosystem of Engagement: Implications for Learning and Broadening Participation, BioScience, biac035, https://doi.org/10.1093/biosci/biac035

Amarasinghe, I., Manske, S., Hoppe, H. U., Santos, P., & Hernandez-Leo, P. (2021). Using Network Analysis to Characterize Participation and Interaction in a Citizen Science Online Community. In Hernandez-Leo, D., Hishiyama, R., Zurita, G., Weyers, B., Nolte, A., & Ogata, H. (Eds.): Collaboration Technologies and Social Computing (LNCS 12856; pp. 67-82). Springer.

Aristeidou, M., Scanlon, E., & Sharples, M. (2017). Profiles of engagement in online communities of citizen science participation. Computers in Human Behavior, 74, 246-256. https://doi.org/10.1016/j.chb.2017.04.044

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 10. https://doi.org/10.1088/1742-5468/2008/10/P10008

Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., & Shirk, J. (2009). Citizen science: A developing tool for expanding science knowledge and scientific literacy. BioScience, 59(11), 977-984.

Bonney, R., Phillips, T. B., Ballard, H. L., & Enck, J. W. (2016). Can citizen science enhance public understanding of science? Public Understanding of Science, 25(1), 2-16.

Bonney, R., Shirk, J., Phillips, T. B., Wiggins, A., Ballard, H. L., Miller-Rushing, A. J., et al. (2014). Next steps for citizen science. Science, 343, 1436-1437.

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network Analysis in the Social Sciences. Science, 323, 892-895.

Brenton, P., von Gavel, S., Vogel, E., & Lecoq, M. E. (2018). Technology infrastructure for citizen science. In: Hecker, S., Haklay, M., Bowser, A., Makuch, Z., Vogel, J. & Bonn, A. 2018. Citizen Science: Innovation in Open Science, Society and Policy (pp. 63-80). UCL Press, London. https://doi.org/10.14324/111.9781787352339

Brudney, J. L., & Meijs, L. C. P. M. (2009). It ain't natural: Toward a new (natural) resource conceptualization for volunteer management. Nonprofit and Voluntary Sector Quarterly, 38(4), 564-581.

Butts, C. T. (2019). Tools for Social Network Analysis. In (Version 2.5)
Chandler, M., See, L., Copas, K., Bonde, A. M. Z., Lopez, B. C., Danielsen, F., . . . Turak, E. (2017). Contribution of citizen science towards international biodiversity monitoring. Biological Conservation, 213(B), 280-294.

Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. Physical Review E, 70(6). https://doi.org/10.1103/PhysRevE.70.066111
Cooper, C. B., Dickinson, J., Phillips, T., & Bonney, R. (2007). Citizen Science as a Tool for Conservation in Residential Ecosystems. Ecology and Society, 12(2). http://www.jstor.org/stable/26267884

Cooper, C., Larson, L., Holland, K. K., Gibson, R., Farnham, D., Hsueh, D., ... & McGillis, W. (2017). Contrasting the views and actions of data collectors and data consumers in a volunteer water quality monitoring project: implications for project design and management. Citizen Science: Theory and Practice, 2(1).

Costa, M., Schaffner, B. F., & Prevost, A. (2018). Walking the walk? Experiments on the effect of pledging to vote on youth turnout. PloS one, 13(5). https://doi.org/https://doi.org/10.7910/DVN/UTX95J

Crall A, Kosmala M, Cheng R, Brier J, Cavalier D, Henderson S, Richardson A. (2017). Volunteer recruitment and retention in online citizen science projects using marketing strategies: Lessons from Season Spotter. Journal of Science Communication, 16:01. DOI: https://doi.org/10.22323/2.16010201.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. In Complex Systems.

Davis, A., Gardner, B., Gardner, M., & Warner, W. L. (1941). Deep South: a sociological anthropological study of caste and class. University of Chicago Press.

Fischer, H., Cho, H., & Storksdieck, M. (2021). Going beyond hooked participants: The nibble-and-drop framework for classifying citizen science participation. Citizen Science: Theory and Practice, 6(1).

Freeman, L. C. (2003). Finding social groups: A meta-analysis of the southern women data. Dynamic Social Network Modeling and Analysis, Washington.

Handcock, M. S., Hunter, D. R., Butts, c. T., Goodreau, S. M., Krivitsky, P. N., & Morris, M. (2018). ergm: Fit, simulate and diagnose exponential-family models for networks. In https://CRAN.R-project.org/package=ergm

Herodotou, C., Aristeidou, M., Miller, G., Ballard, H., & Robinson, L. (2020). What do we know about young volunteers? An exploratory study of participation in Zooniverse. Citizen Science: Theory and Practice, 5(1). https://doi.org/http://doi.org/10.5334/cstp.248

Herodotou, C., Ismail, N., Aristeidou, M., Miller, G., Lahnstein, A. I. B., Ghadirim, M., ... & Ballard, H. (2022). Online Community and Citizen Science supports environmental science learning by young people. Computers & Education, 104515.

Hilbe, J. M. (2011). Negative binomial regression. Cambridge University Press.

Hoffman, C., Cooper, C., Kennedy, E. B., Farooque, M., & Cavalier, D. (2017). SciStarter 2.0: A digital platform to foster and study sustained engagement in citizen science. In Analyzing the Role of Citizen science in Modern Research (pp. 50-60). IGI Global.

Howison, J., Wiggins, A., & Crowston, K. (2011). Validity issues in the use of social network analysis with digital trace data. Journal of the Association for Information Systems, 12(12), 2.

Hunter, D. D., Handcock, M. S., Butts, c. T., M, G. S., & Morris, M. (2008). ergm: A package to fit, simulate and diagnose exponential-family models for networks. 24(3), 1-29.

IBM Corp. (2017). IBM SPSS Statistics for Windows. In (Version 25) IBM Corp.

Jordan, R. C., Ballard, H. L., & Phillips, T. B. (2012). Key issues and new approaches for evaluating citizen-science learning outcomes. Frontiers in Ecology, 10(6), 307-309.

Jordan, R. C., Crall, A., Gray, S., Phillips, T., & Mellor, D. (2015). Citizen science as a distinct field of inquiry. BioScience, 65(2), 208-211.

Katzev, R., & Wang, T. (1994). Can commitment change behavior? A case study of environmental actions. Journal of Social Behavior and Personality, 9(1), 13-26.

Larson, L. R., Cooper, C. B., Futch, S., Singh, D., Shipley, N. J., Dale, K., ... & Takekawa, J. Y. (2020). The diverse motivations of citizen scientists: Does conservation emphasis grow as volunteer participation progresses?. Biological Conservation, 242, 108428.

Lynch-O'Brien, L. I., Babchuk, W. A., Dauer, J. M., Heng-Moss, T., & Golick, D. (2021). Transference of Citizen Science Program Impacts: A Theory Grounded in Public Participation in Scientific Research. Diversity, 13(8), 339.

McKinley, D. C., Miller-Rushing, A. J., Ballard, H. L., Bonney, R., Brown, H., Cook-Patton, S. C., . . . Soukup, M. A. (2017). Citizen science can improve conservation science, natural resource management, and environmental protection. Biological Conservation, 208, 15-28.

Newman, G., Wiggins, A., Crall, A., Graham, E., Newman, S., & Crowston, K. (2012). The future of citizen science: Emerging technologies and shifting paradigms. Frontiers in Ecology and the Environment, 10(6), 298-304.

Opsahl, T. (2013). Triadic Closure in two-mode networks: Redefining the global and local clustering coefficients. Social Networks, 35(2), 159-167.

Padgett, J. F. (1994). Marriage and Elite Structure in Renaissance Florence, 1282-1500 Social Science History Association,

Parrish, J. K., Burgess, H., Weltzin, J. F., Fortson, L., Wiggins, A., & Simmons, B. (2018). Exposing the science in citizen science: Fitness to purpose and intentional design. Integrative and Comparative Biology, 58(1), 150-160.

Pateman, R. M., Dyke, A., & West, S. E. (2021). The diversity of participants in environmental citizen science. Citizen Science: Theory and Practice.

Peter, M., Diekötter, T., & Kremer, K. (2019). Participant outcomes of biodiversity citizen science projects: A systematic literature review. Sustainability, 11(10), 2780.

Peter, M., Diekötter, T., Höffler, T., & Kremer, K. (2021). Biodiversity citizen science: Outcomes for the participating citizens. People and Nature, 3(2), 294-311.

Pettey, W. B. P., Toth, D. J. A., Redd, A., Carter, M. E., Samore, M. H., & Gundlapalli, A. V. (2016). Using network projections to explore co-incidence and context in large clinical datasets: Application to homelessness among U.S. Veterans. Journal of Biomedical Informatics, 61, 203-213.

Phillips, T., Porticella, N., Constas, M., & Bonney, R. (2018). A framework for articulating and measuring individual learning outcomes from participation in citizen science. Citizen Science: Theory and Practice, 3(2).

Ponciano, L., & Pereira, T. E. (2019). Characterising volunteers' task execution patterns across projects on multi-project citizen science platforms Brazilian Symposium on Human Factors in Computing Systems, Vitoria, ES, Brazil.

R Core Team. (2019). R a language and environment for statistical computing. In R foundation for Statistical Computing. www.R-project.org

Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph models for social networks. Social Networks, 29, 173-191.
RStudio Team. (2016). Rstuido: Integrated Development for R. In RStudio, Inc. www.rstudio.com

Salmon, R. A., Rammell, S., Emeny, M. T., & Hartley, S. (2021). Citizens, Scientists, and Enablers: A Tripartite Model for Citizen Science Projects. Diversity, 13(7), 309.

Sankar, C. P., Asokan, K., & Kumar, K. S. (2015). Exploratory social network analysis of affiliation networks of Indian listed companies. Social Networks, 43, 113-120.

Sharova, M. V. (2020). Next steps in shared management of citizen scientists: Understanding project owner attitudes and volunteer connections to science North Carolina State University].

Shirk, J. L., Ballard, H. L., Wilderman, C. C., Phillips, T., Wiggins, A., Jordan, R., . . . Bonney, R. (2012). Public participation in scientific research: A framework for deliberate design. Ecology and Society, 17(2), 29.

Spasiano, A., Grimaldi, S., Braccini, A. M., & Nardi, F. (2021). Towards a Transdisciplinary Theoretical Framework of Citizen Science: Results from a Meta-Review Analysis. Sustainability, 13(14), 7904.

Stephens, B., Chen, W., & Butler, J. S. (2016). Bubbling up the good ideas: A two-mode network analysis of an intra-organizational idea challenge. Journal of Computer-Mediated Communication, 21, 210-299.

Stylinski, C. D., Peterman, K., Phillips, T., Linhart, J., & Becker-Klein, R. (2020). Assessing science inquiry skills of citizen science volunteers: A snapshot of the field. International Journal of Science Education, Part B. https://doi.org/10.1080/21548455.2020.1719288

Toomey, A. H., & Domroese, M. C. (2013). Can citizen science lead to positive conservation attitudes and behaviors? Research in Human Ecology, 20(1), 50-62.
van der Pol, J. (2019). Introduction to network modeling using Exponential Random graph Models (ERGM): Theory and an application using R-project. Computational Economics, 54(3), 845-875.

van der Pol, J. (2019). Introduction to network modeling using exponential random graph models (ergm): theory and an application using R-project. Computational Economics, 54(3), 845-875.

Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications. Cambridge University Press.

Zaken, D. B., Gal, K., Shani, G., Segal, A., & Cavalier, D. (2021). Intelligent Recommendations for Citizen Science. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 35, No. 17, pp. 14693-14701).

Zaken, D. B. , Segal, A., Cavalier, D., Shani, G., and Gal. K. (2022). Generating Recommendations with Post-Hoc Explanations for Citizen Science. In Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '22). Association for Computing Machinery, New York, NY, USA, 69–78. https://doi.org/10.1145/3503252.3531290

Figure 1

Example of project page on SciStarter.com

At the time of data collection, interested volunteers could either click to “join” or “bookmark” a project on SciStarter. When volunteers click the “join” feature depicted here, they are navigated to the project website and the project is added to the volunteer dashboard. In this study, clicking the “join” button was used as a proxy for project engagement.

Figure 2

Distribution of projects on SciStarter.org by number of volunteers who joined each project as registered SciStarter members

Percentages represent the percent of total projects that fall in each category

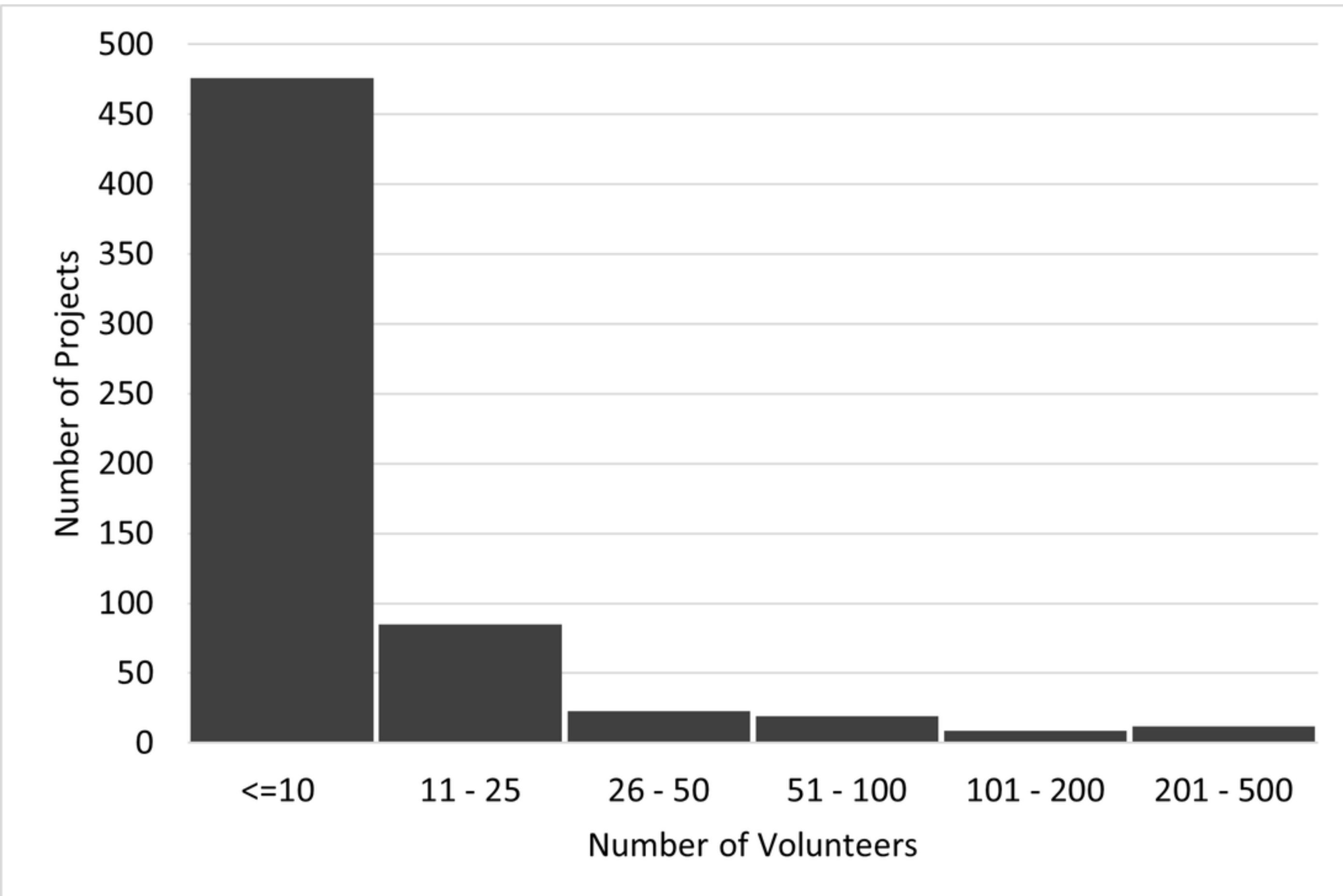

Figure 3

Project connection network for all citizen science projects on SciStarter.org

This is a one-mode projection of projects connected by shared volunteers. Node location was determined by default igraph plotting function, node size was determined by degree centrality. Disconnected nodes (isolates) share no volunteers with other groups

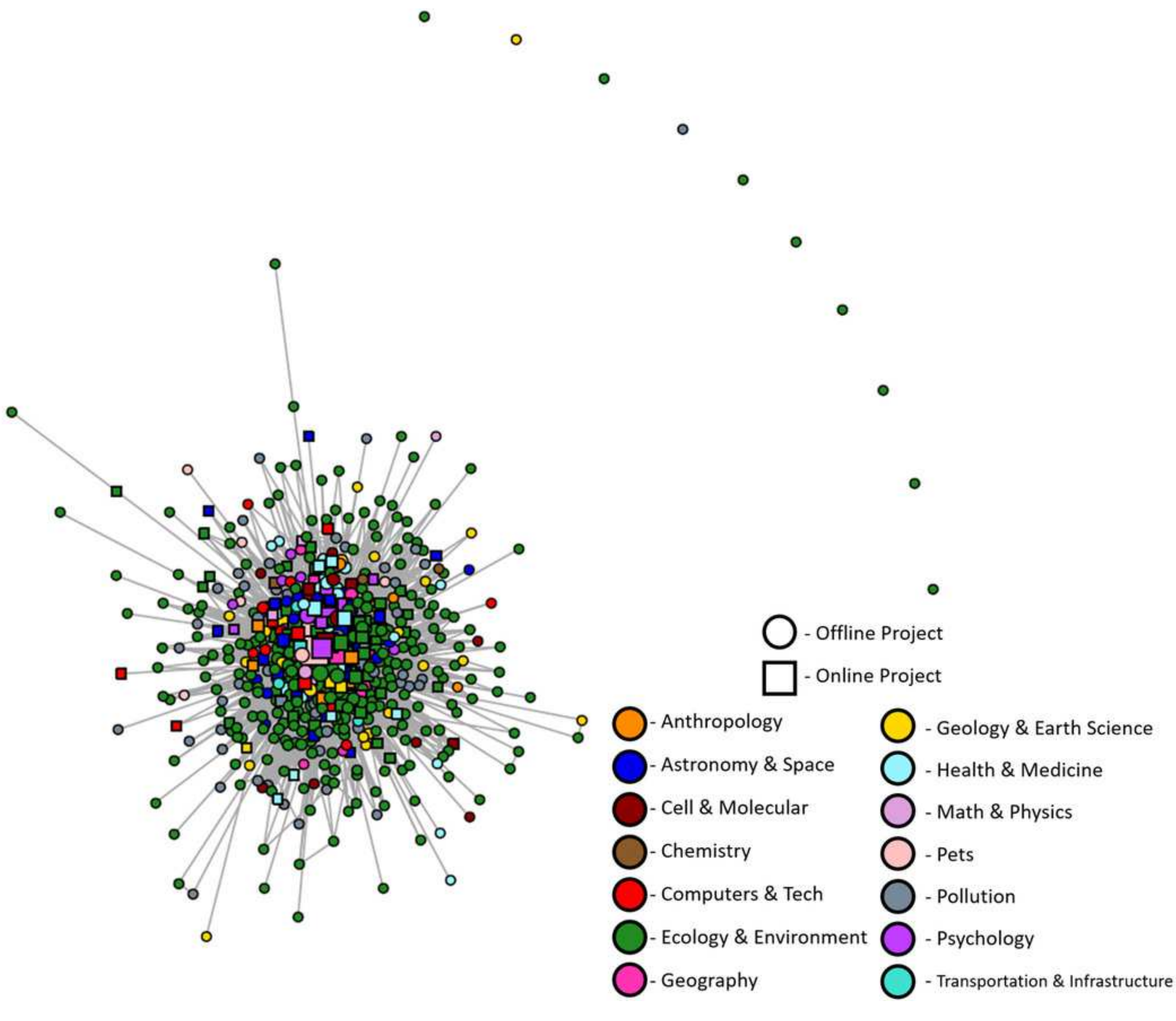

Figure 4

Project connection network for most highly connected citizen science projects on SciStarter.org

Projection showing the most highly connected citizen science projects across the SciStarter.org platform. Only projects that have 10 or more shared volunteers are included. Node location was determined by default igraph plotting function. Node size determined by degree centrality.

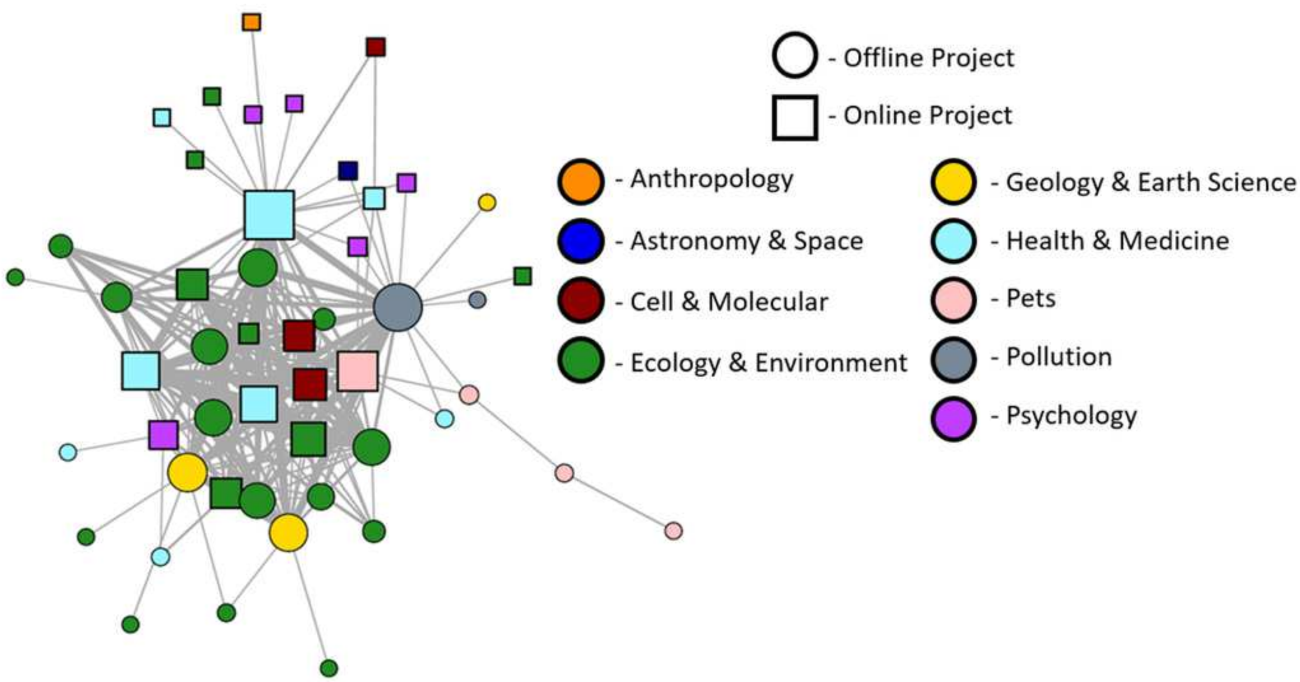

Figure 5

Correlation between Citizen Science Project Centrality and Popularity across the SciStarter.org Platform

The significant correlation between project degree popularity, as measured by the number of volunteers that joined a project, and project centrality, as measured by the number of projects to which a project is directly connected, shows that the most popular projects are those most likely to share volunteers with other projects.

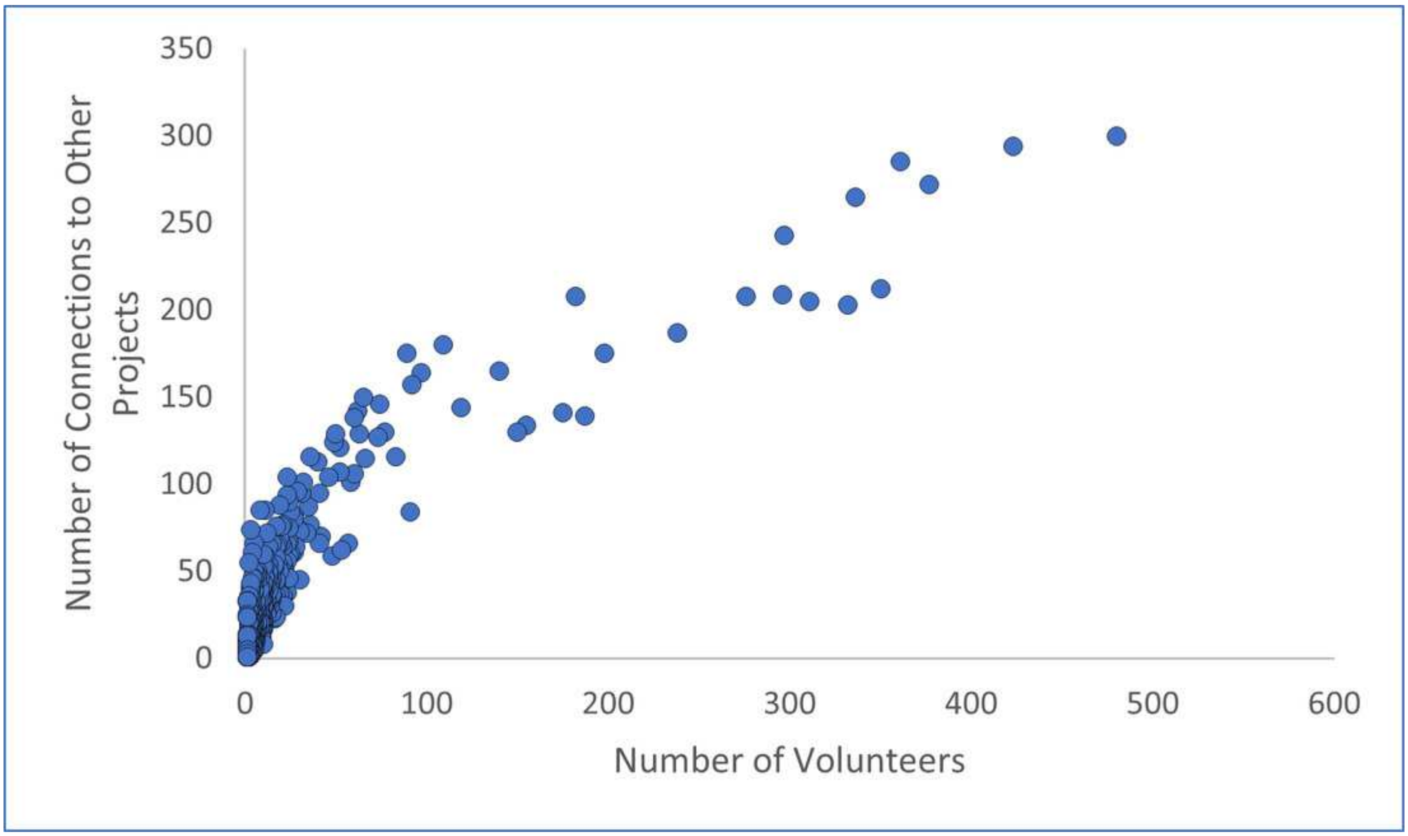

Table 1(on next page)

Counts and ratios of the number of citizen science projects displaying each attribute in the project connection network from SciStarter.org

Table 1:
Counts and ratios of the number of citizen science projects displaying each attribute in the project connection network from SciStarter.org

Attribute	Category	Number of Projects	Percentage
Topic	Cell & Molecular	21	3.3%
	Chemistry	4	0.6%
	Ecology & Environment	337	54%
	Pollution	57	9.1%
	Geography	9	1.4%
	Geology & Earth Science	36	5.8%
	Anthropology	13	2.1%
	Pets	13	2.1%
	Psychology	24	3.8%
	Transportation & Infrastructure	4	0.6%
	Astronomy & Space	39	6.3%
	Computers & Technology	17	2.7%
	Math & Physics	6	1.0%
	Health & Medicine	44	7.1%
Mode of Participation	Offline	455	73%
	Online	169	27%
Affiliate Status	Not an Affiliate	578	92.6%
	Part-time Affiliate	19	3.0%
	Full-time Affiliate	27	4.3%
Campaign Promotion Status	Never featured	387	62.0%
	Featured before data collection	153	24.5%
	Featured during data collection	42	6.7%
	Featured before and during data collection	42	6.7%

Table 2(on next page)

Descriptive statistics for the single large component in the project network

Table 2. Descriptive statistics for the single large component in the project network.

Measure	Measure Explanation	Project Network
Number of Nodes	Number of projects	614
Number of Edges	Number of connections made between projects by shared volunteers	9678
Diameter	Longest distance between two projects, as measured by the number of projects that connect them	7.00
Average Path Length	Average distance between projects	2.26
Density	Ratio of edges present to edges possible	0.05
Transitivity	Ratio of triangles (three nodes connected by all possible edges) present to triangles possible	0.32

Table 3(on next page)

(on next page)

Connections between and across citizen science project attributes

Based on percentage of observed connections (based on shared volunteers) relative to what would be expected in a random distribution (bottom row of each matrix). Positive percentages (above 0%) imply more connections than expected, and negative percentages (below 0%) indicate fewer connections than expected. Each row sums to 0%.

Table 3:

Connections between and across citizen science project attributes. Based on percentage of observed connections (based on shared volunteers) relative to what would be expected in a random distribution (bottom row of each matrix). Positive percentages (above 0%) imply more connections than expected, and negative percentages (below 0%) indicate fewer connections than expected. Each row sums to 0%.

a.		**Topic of Project 2**			
		Earth & Life Sciences	Behavior & Social Sciences	Engineering & Physical Sciences	Health & Medicine
Topic of Project 1	Earth & Life Sciences	-7%	5%	-2%	5%
	Behavior & Social Sciences	-17%	10%	0%	8%
	Engineering & Physical Sciences	-18%	8%	2%	7%
	Health & Medicine	-20%	8%	-1%	13%
	Percentage of Projects	74%	9%	10%	7%

b.		**Mode of Project 2**	
		Online	Offline
Mode of Project 1	Online	24%	-24%
	Offline	12%	-12%
	Percentage of Projects	28%	72%

c.		**Affiliate Status of Project 2**		
		Full-time Affiliate	Part-time Affiliate	Never an Affiliate
Affiliate Status of Project 1	Full-time Affiliate	8%	2%	-10%
	Part-time Affiliate	9%	3%	-12%
	Never an Affiliate	12%	4%	-16%
	Percentage of Projects	4%	3%	93%

d.		Campaign Feature of Project 2			
		Featured Before and During Data	Featured During Data	Featured Before Data	Never Featured
Campaign Feature of Project 1	Featured Before and During Data	14%	6%	3%	-23%
	Featured During Data	12%	5%	3%	-20%
	Featured Before Data	11%	4%	1%	-18%
	Never Featured	8%	5%	2%	-16%
	Percentage of Projects	7%	7%	25%	62%

Table 4 (on next page)

Regression results.

Negative binomial regression[a] results showing project attributes (topic, mode, affiliate status, campaign features) associated with degree centrality[b] for citizen science projects on the SciStarter.org platform.

Table 4:
Regression results. Negative binomial regression[a] results showing project attributes (topic, mode, affiliate status, campaign features) associated with degree centrality[b] for citizen science projects on the SciStarter.org platform.

Variable	Proportion of Projects with Attribute	B	SE	p-value
Intercept		6.265	0.268	<0.001
Topic: Health and Medicine[c]	0.07			
Topic: Behavior & Social Sciences	0.09	-0.228	0.208	0.273
Topic: Earth & Life Sciences	0.74	-0.705	0.162	<0.001
Topic: Engineering & Physical Sciences	0.10	-0.858	0.202	<0.001
Mode: Online[c]	0.73			
Mode: Offline	0.27	-0.788	0.100	<0.001
Affiliate: Full-time Affiliate[c]	0.04			
Affiliate: Part-time Affiliate	0.03	-0.158	0.315	0.616
Affiliate: Not Affiliate	0.92	-1.193	0.219	<0.001
Campaign: Featured Before and During[c]	0.07			
Campaign: Featured During	0.07	-0.540	0.226	0.017
Campaign: Featured Before	0.24	-0.570	0.190	0.003
Campaign: Never Featured	0.62	-0.949	0.181	<0.001

[a]Omnibus Test X2 (9, N=624) = 301.6, $p < .001$
[b]Degree centrality represents the number of connections from one project to other project nodes in the network (M = 31.02, range = 0 to 300).
[c]Indicates reference category for each attribute group