Why Won't Aliens Talk to Us? Content and Community Dynamics in Online Citizen Science

Markus Luczak-Roesch, Ramine Tinati, Elena Simperl, Max Van Kleek, Nigel Shadbolt
Web and Internet Science, Electronics and Computer Science
University of Southampton

Robert Simpson
Astrophysics
Department of Physics
Oxford University

Abstract

We conducted a quantitative analysis of ten citizen science projects hosted on the Zooniverse platform, using a data set of over 50 million activity records and more than $250,000$ users, collected between December 2010 and July 2013. We examined the level of participation of users in Zooniverse discussion forums in relation to their contributions toward the completion of scientific (micro-)tasks. As Zooniverse is home to a multitude of projects, we were also interested in the emergence of cross-projects effects, and identified those project characteristics, most importantly the subject domain and the duration of a project. We also looked into the adoption of expert terminology, showing that this phenomenon is dependent on the scientific domain which a project addresses but also affected by how the communication features are actually used by a community. This is the first study of this kind in this increasingly important class of online community, and its insights will inform the design and further development of the Zooniverse platform, and of citizen science systems as a whole.

Introduction

While the history of amateur involvement in scientific discovery originated long before the establishment of modern scientific institutions (Silvertown 2009), the Internet has fundamentally re-vitalized and expanded the ways and scale in which untrained citizens can participate in scientific investigations. These so-called *citizen science* projects (Bonney et al. 2009; Gray, Nicosia, and Jordan 2012; Irwin 1995) have thus far enlisted the help of millions of volunteers in a wide array of scientific inquiries, ranging from the taxonomic classification of galaxies (Fortson et al. 2011; Lintott and others 2008) and the creation of an online encyclopedia of all living species on Earth (Wilson 2003), to the derivation of solutions to protein folding problems (Khatib et al. 2011), and the tracking and measuring the population and migratory patterns of animals in the Serengeti national park.[1] Perhaps more significantly, the data produced by these projects has crucially enabled hundreds of scientific discoveries in various disciplines, demonstrating that, beyond being a *viable* technique to tackle large and difficult problems, citizen science has become accepted as an effective, and even a *preferred* method to pursue science in the 21^{st} century.

While most citizen science initiatives have a strong human computation character (Quinn and Bederson 2011), and thus have to put special emphasis on the design of the tasks to be completed by the crowd and on attracting and sustaining participation, they have also created an environment that encourages serendipitous scientific discovery. Just as amateur scientists in the 19^{th} century defined new lines of scientific inquiry based on their observations, users of modern citizen science platforms have more often than not adventured far beyond the human-computation framework of a given project in order to work on research questions they came up with themselves, typically based on something they noticed while performing the (routine) tasks they were intended to. Central to such citizen-led investigations have been online discussion forums. Usually established with the purpose of simply allowing participants to find answers to common questions, discussion forums have become conduits for a wide range of different kinds of information sharing and ideas exchange activities within the community. Such interactions, ranging from the purely social to goal-driven and collaborative, have been seen to ultimately benefit a citizen science project both directly, in their question-answering function and as enabler of novel discoveries, and indirectly, as a means to attract interest and long-term engagement (Mugar, Osterlund, and Hassman 2014).

In this paper we sought to understand the role and characteristics of discussion forums in Web-based citizen science projects. To do so, we conducted a quantitative analysis of an ecosystem of ten such projects hosted on the Zooniverse[2] platform, using a data set of over 50 million activity records and more than $250,000$ users, collected between December 2010 and July 2013. In our analysis we asked three fundamental research questions:

1. Is there a relationship between the level of participation of users in discussions, and their contributions toward the completion of scientific (micro-)tasks? In other words we were interesting in learning more about the users undertaking each type of activities and identifying typical user behavior patterns.

[1] http://www.snapshotserengeti.org/

[2] http://www.zooniverse.org

2. What kinds of projects are likely to appeal to Zooniverse users? In particular, we explored so-called *cross-project effects* and studied the typology of users who contributed to more than one project hosted by the platform.
3. Do discussions facilitate domain learning? More specifically, we looked into the increased adoption of a specialized vocabulary, assumed to be indicative of advanced domain knowledge, by users joining projects at different points in time.

Our research revealed that, similar to other crowdsourced problem-solving environments, only a small proportion of the overall user base of a citizen science project is responsible for a majority of the contributions. The subject domain of a project (i.e., Astrophysics, Nature, etc.), its duration, and the number of contributors it attracts seem to determine the willingness of this community to take on other scientific tasks published on the platform. This observation does not fully match the findings put forward by related literature in the area of online communities (Whittaker et al. 1998a; Zhongbao and Changshui 2003; Adamic et al. 2008), and leaves room for further research. To learn more about the properties of this distinctive form of online community, we looked into the use of expert terminology in forum posts, showing that such learning effects do exist, and that knowledge transferability depends primarily on the scientific domain that a project addresses. To the best of our knowledge, this is the first large-scale quantitative analysis of this kind in this increasingly important class of online community. While there exists a mature body of literature on both content analysis and collaboration in online communities (Butler, Joyce, and Pike 2008; Kittur et al. 2007a; 2007b), studies of citizen science systems have predominantly focused on specific aspects of crowdsourced task management, including user motivation, task design, and quality assurance, rather than exploring the full range of activities carried out as part of these systems (Raddick et al. 2010; Jordan et al. 2011; Rotman et al. 2012). The insights gained from our analysis will inform the design and further development of the Zooniverse platform, and of citizen science systems as a whole.

Preliminaries and Related Work

Zooniverse is a Web-based citizen science platform that hosts[3] 30 distinct citizen science projects spanning various scientific subjects and lines of inquiry. Participants contribute to projects by performing simple human computation tasks on digital artifacts such as images, videos, and audio recordings of any timeliness, from historic captures up to live data feeds. These artifacts are referred to as *subject*s and we refer to all the activities performed on them uniformly as `Tasks`. Whilst the primary function of the Zooniverse platform is to provide a human computation interface to perform `Tasks` (e.g., to recognize galaxies or classify whale sounds), users can also share information, ask questions, and discuss their views and ideas through an integrated communication mechanism called `Talk`. Specifically, `Talk` provides a space for peer question-answering support and serendipitous collaboration, and facilitates social community building. As a subcategory of `Talk`, Zooniverse supports a microblogging-like communication style limited to 140 characters per post. This feature is integrated into the `Task` interface, while the resulting posts can be accessed, just as any other discussion activities in the `Talk` area of a project. Altogether, `Talk` is comparable to discussion forums in other peer-production systems and online communities in general which we explore in the following as a first dimension of related research.

[3] as of January 2014

Online Communities

Online communities have been a recurrent research topic for many years, attracting great interest among computing scholars, social scientists, and economists. Researchers in fields as diverse as CSCW, Web technologies, crowdsourcing, social structures, or game theory, have long studied them from different perspectives, from the behaviour and level of participation of specific groups and individuals (Lampe and Johnston 2005; Arguello et al. 2006), to the characteristics of peer-production systems and information sharing repositories (Merkel et al. 2004; Krieger, Stark, and Klemmer 2009), and the emergence of novel social structures (Kumar, Novak, and Tomkins 2006; Backstrom et al. 2006).

There is a long list of studies of online information sharing environments, including newsgroups, bulletin board systems, discussion forums, question answering sites, and email networks, which investigate questions such as, answer quality (Agichtein et al. 2008; Harper et al. 2008); topics coverage and language (Danescu-Niculescu-Mizil et al. 2013; Rowe et al. 2013); user profiles, roles, and expertise (Campbell et al. 2003; Fisher, Smith, and Welser 2006; Zhang, Ackerman, and Adamic 2007; Welser et al. 2007); levels of engagement (Joyce and Kraut 2006; Rafaeli, Ravid, and Soroka 2004); and community structure and dynamics (Anderson et al. 2012; Adamic et al. 2008; Yang and Leskovec 2012; Kairam, Wang, and Leskovec 2012; Whittaker et al. 1998b). In a citizen science context many of these research questions and the methods and insights gained in prior work remain relevant and valuable, though their framing will naturally be different. We sought to understand if the use of discussion features affects user behavior and engagement in the science-related tasks she is asked to solve. In addition, given the role of importance of citizen-led scientific discoveries in crowdsourced science, we were ultimately interested in studying discussion forums as enablers of such phenomena. These topics have been analyzed in great detail in the context of online collaborative environments such as Wikipedia (Butler, Joyce, and Pike 2008; Kittur et al. 2007a; 2007b) By comparison to Wikipedia research, our work spanned across multiple projects, each addressing a different scientific problem in a different domain. Even more importantly, collaboration in Zooniverse projects is achieved only via discussion forums, as the actual tasks the users are invited to engage with are meant to be solved individually.

Citizen Science

Some of the most popular showcases of crowdsourced science have inspired researchers to analyze this emergent trend with respect to the quality of its outcomes, the workflows and interaction models that facilitated these outcomes, and the characteristics of the volunteering community. Descriptive accounts of specific initiatives become more and more pervasive both in natural sciences and the digital humanities, as professional scientists struggle to master the ever greater data collection and analysis challenges of modern times (Cohn 2008; Zook et al. 2010; Heinzelman and Waters 2010; Westphal and others 2005; Wiggins and Crowston 2011). Most of these papers offer an informative overview of the corresponding projects, and report on the most important design and community management decisions. A prominent topic in the literature documenting these projects is user engagement. Raddick *et al.* conducted interviews to understand the motivations of Zooniverse volunteers (Raddick et al. 2010); similar studies can be found in (Nov, Arazy, and Anderson 2011), (Rotman et al. 2012), and (Jordan et al. 2011). (Cooper 2013; Kawrykow et al. 2012) looked into the quality of volunteer contributions, proposing methods to validate the input created by the crowd. (Khatib et al. 2011) presented an algorithm that learns the best solution patterns in protein sequencing from players of the FoldIt game. (Yu, Wong, and Hutchinson 2010) proposed a method to assess the expertise of community members in the eBird bird watching project, while (Wiggins and Crowston 2010) introduced a model to capture community organization. These efforts are illustrative for a trend that can be distinctively observed in citizen science - science teams, as well as developers of citizen science technology platforms need richer analytical insight into how citizen science projects should be optimally run. Compared to prior work our paper provides the first in-depth analysis of the role of discussion forums in communities of citizen scientists, applied over a large data set representing ten different projects in multiple domains that are run on the same technology platform.

Study Set-up

Our analysis was based on the complete project histories for ten Zooniverse projects, comprising of data captured between December 2010 and July 2013. This data set includes information about the `Task` and `Talk` contributions of each user and each project. This adds up to a total of 250,071 users, 50,995,591 classifications, and 663,791 discussion posts. Table 1 gives an overview of the data set. Posts are organized into four different boards that are predefined by the Zooniverse platform: `Help`, `Science`, `Chat`, and 'Untyped', the latter standing for the microposts directly linked to `Task` subjects.

Methods for Analysis

Guided by the three research questions introduced in Section we chose the methods for the analysis of the aforementioned data set. For studying the participation of users in `Talk` and `Task` we chose a community model based on a standard method for computing social affinity given posts in online forums (Girvan and Newman 2002; Clauset, Newman, and Moore 2004). A thread represents a chronological list of posts and is linked to a board as well as a project. Thus, each thread constitutes a network of users contributing to it and the reference to the projects allows us to compute both *intra-* and *cross-* project community representations. This is backed up by a time series analysis for which we applied standard methods for seasonal decomposition and harmonisation on the `Task` and `Talk` logs, both independently for each of the two types of user contributions, and in combination. The analysis had only retrospective and no predictive purposes, hence we did not resort to any further regression or smoothing techniques. These methods were also the base for statistically analysing the participation across projects. To detect and analyse the shift of a common community and user vocabulary we adapted the framework presented in (Danescu-Niculescu-Mizil et al. 2013). We used a common technique to handle varying user activity over time by regarding the overall number of contributions in a community or of a user and representing them in equally sized and chronologically ordered subsets. More precisely, we applied a sliding window approach that extracts the 10 most frequently used terms for chronological slices of 10% of the posts. The terms were statistically derived from an uni-gram term-document-matrix based on the filtered (lower case stop words, numeric, punctuation), tokenized, and stemmed main body of all comments. The vocabulary shift was thus the difference of the top 10 terms of a slice compared to the slice before, generating a score between 0 (*no change*) to 10 (*all terms differ*).

Results

In this section we present the results of the various studies conducted for this broad system analysis as follows: first, we describe some general properties of the Zooniverse `Talk` feature; second, we focus on the multifaceted relationship between participation levels in `Talk` versus `Task`; third, we introduce observations pertaining to cross-project participation as well as benefits and consequences of hosting multiple citizen science projects around a central core community; and, finally, we contribute an analysis of expert vocabulary dissemination as an indicator of whether novice citizen scientists learn new concepts.

General Observations

We examined the structure and timespan of discussion threads along the four types of `Talk` boards mentioned in Section . The result is shown in Tables 1 and 2. A significant proportion of `Talk` contents (over 90%) result from microposts in direct relation to a classified subject. This post category generally features the highest proportion of single post threads and this share grows higher the larger the projects are in terms of the number of `Talk` entries (e.g., projects such as `GZ`, `PH`).[4] This dominance of micropost communication in the bigger projects means that most threads do not form

[4] Abbreviations in project names refer to Table 1.

Project	Time Period	Classifications	Posts	Chat (%)	Help (%)	Science (%)	Untyped (%)
Planet Hunters (PH)	12/10 - 07/13	19,179,696	427,917	1	1	3	**95**
Cyclone Center (CC)	09/12 - 06/13	218,317	1,615	6	9	9	**77**
Galaxy Zoo 4 (GZ)	09/12 - 07/13	7,751,825	89,956	1	1	**18**	80
Seafloor Explorer (SF)	09/12 - 07/13	1,682,511	33,367	1	1	2	**99**
Snapshot Serengeti (SG)	12/12 - 07/13	7,800,896	39,250	2	8	5	80
Andromeda Project (AP)	12/12 - 07/13	1,091,406	10,198	1	3	2	**94**
Planet Four (PF)	01/13 - 07/13	3,900,785	32,097	5	5	**11**	78
Notes from Nature (NN)	04/13 - 07/13	209,169	4,208	8	**24**	2	**65**
Space Warps (SW)	05/13 - 07/13	7,037,472	20,978	4	4	7	8
Worm Watch Lab (WS)	07/13 - 07/13	90,350	855	6	**18**	4	**72**
ZOO	12/10 - 07/13	50,995,591	663,791	2	2	5	91

Table 1: Number of classifications (Task) and discussion posts (Talk) per project

actual conversations. By contrast, two of the much smaller projects (CC and NN) have a significantly lower proportion of single post threads, and a higher average length of threads. These projects also obtained a higher proportion of posts on the Help board. Whilst thread duration for different projects and boards varied, the science board threads (notably so in CC, PH and SF) tended to last longer than other threads.

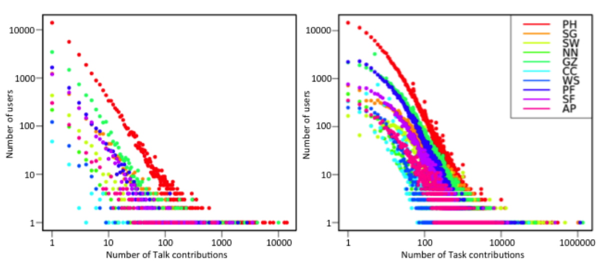

Figure 1: Distribution of Talk (left) and Task (right) contributions, for all users and projects

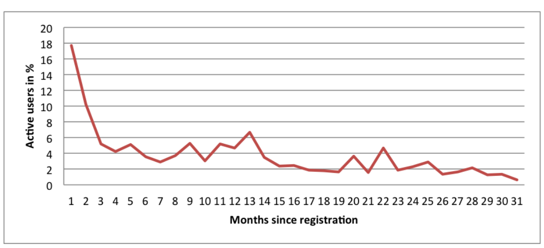

Figure 2: Retention of the cohort of users actively joining PH in the first month of the project. Users tend to come back after months of inactivity.

Talk and Task Participation

From $250,071$ users, a notable proportion of 40.5% had contributed to both classifications and discussions. Figure 1 illustrates the distribution of the absolute numbers of Talk and Task activities, which is interestingly almost identically in all projects and does not show much outlier noise except for Talk on the CC project. The overall trend of the relationship between classifications and posts of every individual user suggested that there are generally two profiles of contributing users: those being active on both Task and Talk and those contributing to the Task only. Most users belong to the latter category contributing a significantly lower amount of comments or do not use Talk at all. This is reflected by a median of 600 Task entries opposed by a median of 14 posts on Talk per user. We will regard these values as the threshold for the average user. The most prolific users (with activity levels bigger or equal the average) were responsible for 29.0% and 72.0% of the overall amount of Tasks and Talks.

In order to be able to perform a visual analysis of recurrent behavior patterns we sampled a set of highly active users as being those contributing to both Task and Talk by the factor 10 higher than the average user. For those users that created at least 140 posts and 6000 classifications we visualized the monthly activity. We found three core classes of Talk and Task activity: (i) users that consistently contribute; (ii) users alternating between months of activity and months of inactivity; and (iii) users who stop visiting the platform entirely after some time. By conducting a cohort analysis, which regards cohorts of participants performing their first activity in a particular month and project, we are able to confirm that these patterns hold beyond this group of highly active users. Figure 2 depicts the course of the user retention of the cohort of users actively joining PH in the first month of the project.

User Lifecycle Table 3 shows the number of users who performed their first classification via a particular project, as well as the average lifespan of a user in respect to Talk and Task. User lifecycle metrics are calculated by measuring the delta between a user's first and last entry on either Talk and Task. In order to overcome the skewing effects caused by the distribution of user contributions, the values shown in Table 3 refer solely to the sample of users responsible for at least 600 Tasks and 14 Talk posts.

Cross-Project Effects

In order to find out whether there was a notable cross-over between users participating in more than one of the ten different Zooniverse projects, we focused on those users that have contributed to both Talk and Task at least once.

Proj	Threads (C-H-U-S)	Avg. Thread (posts) (C-H-U-S)	Avg. Thread Dur. (days) (C-H-U-S)	Med. Response (Hours) (C-H-U-S)	SP(%) (C-H-U-S)
AP	28 — 29 — 5520 — 47	5 — 1 — 1 — 5	20.16 — 1.8 — 3.11 — 1.56	6.01 — 1.42 — 48.81 — 1.75	25.00 — 6.90 — 62.37 — 14.89
CC	17 — 22 — 450 — 12	5 — 6 — 3 — 12	35.00 — 20.33 — 3.56 — 141.32	44.29 — 20.17 — 8.68 — 372.34	0.00 — 4.55 — 24.89 — 0.00
GZ	73 — 76 — 50303 — 39	5 — 6 — 1 — 7	28.47 — 17.67 — 2.17 — 21.71	24.31 — 37.64 — 21.71 — 13.05	17.81 — 18.42 — 55.12 — 15.38
NN	34 — 105 — 1428 — 11	8 — 5 — 2 — 3	23.44 — 13.22 — 4.20 — 27.62	11.64 — 6.34 — 1.59 — 6.55	20.59 — 16.19 — 41.46 — 27.27
PF	90 — 109 — 18514 — 102	8 — 8 — 1 — 17	11.72 — 6.90 — 74.76 — 4.48	8.71 — 5.28 — 6.04 — 7.65	28.89 — 16.51 — 69.60 — 25.49
PH	605 — 256 — 244951 — 676	7 — 12 — 1 — 17	8.61 — 14.37 — 0.82 — 13.23	19.89 — 18.70 — 14.63 — 46.14	18.84 — 8.20 — 71.00 — 19.82
SF	62 — 88 — 24102 — 25	4 — 4 — 1 — 3	9.21 — 7.54 — 15.29 — 12.27	26.56 — 10.84 — 3.33 — 24.27	33.87 — 22.73 — 66.77 — 12.00
SG	111 — 243 — 26676 — 127	7 — 4 — 1 — 3	5.42 — 4.97 — 6.63 — 7.65	8.28 — 6.09 — 1.60 — 8.49	18.92 — 15.64 — 76.21 — 7.87
SW	42 — 105 — 9893 — 50	12 — 6 — 1 — 8	12.43 — 5.80 — 36.28 — 3.94	7.16 — 4.45 — 4.32 — 4.00	9.52 — 22.86 — 54.52 — 30.00
WS	7 — 35 — 531 — 2	5 — 3 — 1 — 7	3.96 — 9.19 — 1.66 — 4.26	1.73 — 2.07 — 0.22 — 1.07	14.29 — 31.43 — 0.60 — 0.00

Table 2: Discussion post and thread metrics of the ten Zooniverse projects. C - Chat, H - Help, U - Untyped, S - Science, ATL - Average Thread Length (posts), ATD - Average Thread Duration (days), SP - Single Posts

Project	Users	New Users	Avg. Talk Lifespan	Avg. Task Lifespan
PH	147,268	142,663	400	508.53
GZ	46,889	36,544	139.79	165.04
CC	4,351	1,693	263.91	267.54
SF	14,526	6,713	119.88	234.63
AP	5,471	3,324	52.63	65.82
SG	20,767	14,123	127.37	183.02
PF	36,551	30,030	74.45	91.07
NN	3,172	911	61.45	88.53
SW	9,184	4,395	49.36	210.27
WS	3,168	647	13.80	20.65
ZOO	-	-	324.83	416.06

Table 3: Number of users per project and number of users who were first active on a given project. The average Talk and Task lifespan (days) of a user

Table indicates that PH, which contains the largest number of users and posts, also has the highest proportion of cross-overs to other projects. The general user cross-over patterns are similar for Task and Talk, with a slightly stronger signal for the former. Projects roughly belonging to the same scientific discipline share a significant number of users. This is best exemplified by the SF and SG projects (both about animals and nature) and the large collection of astrophysics projects (SW, GZ, AP, PF, PH). The most important cross-domain overlay could be observed between SF and AP, which can be traced back to the Talk activities in SF at the launch of AP.

Community Vocabulary Change

To study learning effects as observed through the use of a specific terminology, we first computed the vocabulary shift for the whole duration of the projects and all four types of boards they host. The differences between the early and late stage vocabularies are shown in Table 5. The terms highlighted in this table, in particular for the projects AP, CC, NN, and WS, suggested the presence of recurring errors at the level of the Task interface, which were discussed in the forums. In AP we could notice a change from a specialized language, which refers to the Task to be completed, towards a dialogue that is rather centred around the term 'Andromeda', which was related most likely to issues with the use of the classification interface ('click', 'link'). A similar behavior could be observed in the CC project, while in WS the shift seemed to have happened in the opposite direction, from UI problem reporting to Task-specific posts. Finally, NN was 'plagued' by continuous problem reports from the launch of the project throughout its lifetime.

We computed the same metrics along the themes Chat, Help, Science and 'Untyped'. As illustrated in Figure 3 vocabulary use seemed to be much more stable when talking about subjects (i.e., the microposts classified as 'Untyped') and a core of astrophysics projects, namely PH, GZ and PF, features a significantly lower vocabulary shift, while the CC and NN projects trended into the opposite direction. This stability of the vocabulary in microblog-like posts is not only a result of the restriction to 140 characters but also due to the fact that the users established own hashtags for annotating subjects showing typical or untypical characteristics and guide others to discuss their hypothesis about the subject at hand. In some projects these tags are highly specific already in the early days of a project (e.g. PH and PF) while others start with a rather generic set of hashtags which evolve to become more specific over the course of the project (e.g. SF). By interweaving the distribution of posts across boards (cf. Table 1) we can determine that the PF project features one of the largest proportions of posts on the Science board (11% out of $32,097$) and with PH the lowest vocabulary shift in comparison to other projects' Science discussions.

Discussion

Is there a Connection Between Discussion and Task Completion?

Across all projects, we found that 40.5% of users had contributed to Talk, while 90.8% completed at least one Task. The lower proportion of Talk participation may be a result of interface design, little awareness of the users that this feature exists, or even barriers to entry that arise from lack of domain expertise. Considering only these $(47,141)$ participants, the number of Talk and Task entries exhibited power-law characteristics, as visible in Figure 1. Such a distribution is familiar to many who have studied online communities and crowdsourcing platforms, as

Prj. (Talk/Task Users)	NN	CC	SW	GZ	AP	PF	WS	PH	SG	SF
NN (393/3,126)	-	**11.70/13.38**	3.47/6.48	0.61/1.96	2.23/5.61	0.63/1.92	7.97/18.03	0.26/0.83	3.61/6.12	3.58/8.99
CC (94/4,343)	2.80/18.59	-	0.96/7.81	0.20/2.96	0.69/8.41	0.26/2.11	2.90/13.00	0.07/1.33	0.82/6.53	0.72/11.40
SW (836/13,245)	7.38/27.45	8.51/23.83	-	2.66/4.68	9.09/10.95	1.48/3.70	14.49/24.01	0.92/1.92	4.02/9.84	3.76/17.71
GZ (6,424/46,487)	9.92/29.21	13.83/31.66	**20.4516.41**	-	**19.73/30.46**	3.81/8.65	13.04/28.12	**3.12/7.32**	7.78/13.97	**12.30/28.53**
AP (583/5,433)	3.31/9.76	4.26/10.52	6.34/4.49	1.79/3.56	-	1.11/1.74	5.07/8.32	0.63/1.38	3.25/5.63	**14.32/6.47**
PF (2,704/36,469)	4.3322.36	7.45/17.71	4.785.73	1.60/6.79	5.15/11.67	-	7.25/20.09	1.32/4.83	2.78/8.31	2.95/15.00
WS (138/3,161)	2.80/18.23	4.26/9.46	2.39/10.19	0.28/1.91	1.20/4.84	0.37/1.74	-	0.13/0.90	1.55/5.35	1.12/6.69
PH (23,075/147,032)	15.52/39.12	15.96/45.15	**25.4821.37**	11.19/23.16	**25.04/37.33**	11.24/19.46	**21.74/42.01**	-	13.66/22.38	**17.54/38.15**
SG (1,940/20,742)	**17.81/40.60**	**17.02/31.18**	9.33/15.41	2.35/6.23	10.81/20.45	2.00/4.72	**21.74/35.08**	1.15/3.16	-	**12.13/30.08**
SF (2,235/14,333)	**20.36/41.20**	**17.02/37.62**	10.05/19.17	4.28/8.80	**54.89/17.06**	2.44/5.89	**18.12/30.34**	1.70/3.72	**13.97/20.79**	-

Table 4: The correlation between the Zooniverse Project's [Talk/Task] community. Note: this is read as XX % of users that have completed a [Talk/Task] in project [column] are users that also have completed a [Talk/Task] in project [row].

Prj.	First 10% slice vocabulary	Last 10% slice vocabulary	Mean shift	Median shift
AP	**cluster**, **galaxi**, imag, one, **star**, **synthet**, see, right, can, look	galaxi, cluster, imag, sdss, lonesomeblu, **andromeda**, mayb, **click**, detail, **link**	2.9	2.5
CC	storm, eye, one, look, like, imag, cloud, area, **embeddedcent**, **curvedband**	storm, imag, eye, **can**, **pictur**, center, one, like, think, **classifi**	3.1	3
SG	anim, one, like, look, imag, see, just, lion, can, zebra	look, like, wildebeest, see, one, bird, zebra, lion, right, anim	**2.3**	2
SF	like, look, **fish**, sea, **scallop**, thing, imag, right, star, left	**corallinealga**, **anemon**, object, **hermitcrab**, **bryozoan**, stalkedtun, shrimp, left, **cerianthid**, **sanddollar**	**2.5**	3
PF	imag, fan, look, like, ice, mark, interest, can, featur, one	fan, imag, frost, blotch, boulder, look, spider, blue, like, vent	**2.1**	2
NN	**field**, record, one, use, **enter**, get, work, can, specimen, **button**	like, **field**, record, date, name, **can**, **click**, look, get, label	4.4	5
SW	imag, len, galaxi, lens, look, one, like, left, simul, blue	sim, galaxi, blue, quasar, len, arc, imag, oclock, ring, einstein	**2.3**	2
PH	transit, star, day, aph, look, one, planet, like, possibl, dip	day, transit, httparchive…, possibl, star, kid, dip, look, planet, like	**2.1**	2
GZ	galaxi, star, object, look, classifi, like, imag, fhb, hard, veri	galaxi, star, spiral, object, imag, look, like, merger, one, starform	**1.5**	1.5
WS	worm, **video**, **click**, egg, **onli**, dot, one, side, **can**, **start**	present, **egg**, worm, **lay**, one, **move**, **larva**, activ, count, whi	4.8	5

Table 5: Top 10 terms of the common vocabularies used in Talk posts in (a) the first and (b) last 10% slice of the project lifetime and the mean vocabulary shift coefficient. We use stem words

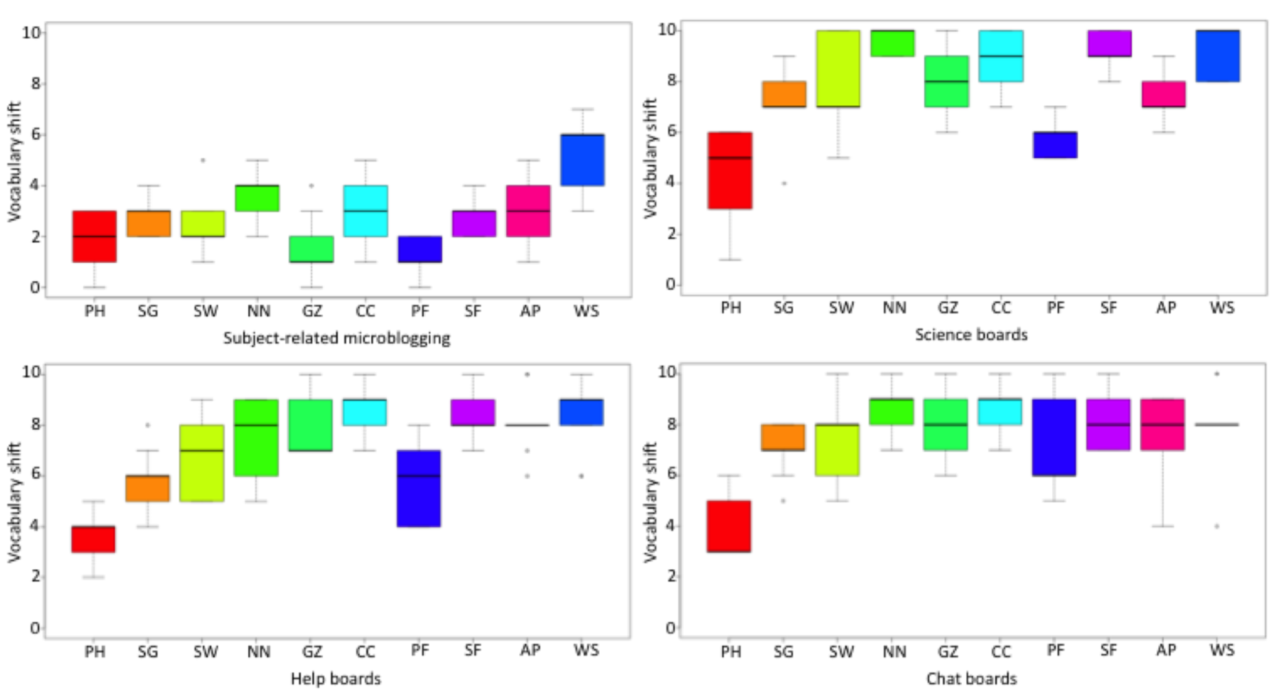

Figure 3: Vocabulary shift for each type of discussion board

is often reflective of engagement levels in these environments (e.g., (Ortega, Gonzalez-Barahona, and Robles 2008; Wang et al. 2013)).

Comparing `Talk` to `Task` participation, users were less likely to comment than performing `Task`. Despite this, a positive correlation between the number of `Talk` entries and the number of `Tasks` performed was identified. 1% of the total number of users who contributed above the average to both `Talk` and `Task` were responsible for over two thirds of the total number of `Talk` entries, and one third of the total `Tasks` performed. This behavior is common to crowdsourcing and collective intelligence systems (Adamic et al. 2008; Wang et al. 2013). However, we also noticed that this effect takes a stronger shape with respect to `Talk`. Advancing on this we identified three typical user profiles within this specific set of highly active users. While one of these profiles simply reflects the typical user loosing interest in an online community after the initial enthusiasm fades away, we concluded from the other two that the notion of an active user needs to be carefully designed. The profiles suggest that project launches and projects being marked as completed heavily affect the level of contributions of highly active users. Core users tend to develop an understanding about when their contributions are needed, and when they can refrain from contributing actively.

Furthermore, as the number of single and 'untyped' posts in Table 2 show, across all boards and projects, the microblogging-style service was the chosen mechanism for posting comments representing 91% of the total posts. As shown in Table 2, the average thread length of these microblogging boards ('Untyped') were only of one post in length, which suggests that this mechanism, was not used for discussion or dialog, but rather as a means to comment or remark on their classification. However, given that participants favour using this microblogging system, does its use reduce the chances of discussion and collaborative problem solving within the Zooniverse platform, given that posts are not assigned to a specific board? The mean response time of threads also raises questions about `Talk's` functionality as a question answering service, with boards such as `Science` in project `CC` taking over 15 days to respond, or `PH`'s `Help` board taking over 18 hours (on average) for a response, a finding which is significantly higher than other online communities (Zheng 2002; Mamykina et al. 2011), which typically see responses times of less than 10 minutes.

Does A Project's Subject Affect Participation?

The second significant observation made during the analysis relates to the participation of users within and across citizen science projects. Based upon the participant lifecycle and project cross-over analysis, we found that: (i) the lifecycle of a user typically out spans the life of a project, a finding which differs to the study of other online communities (Danescu-Niculescu-Mizil et al. 2013); and (ii) citizen scientists are more likely to participate in projects which have similarities in respects to the subject domain, as well as their age and size.

Traditionally in online communities, users are engaged with a particular topic and remain engaged for an extended period of time, however, the common user life cycle is shorter than the life cycle of the community as a whole. Users join a community, contribute and then stop (Danescu-Niculescu-Mizil et al. 2013). In the case of Zooniverse it appears that because users able to participate in multiple projects, individual projects underrun a life cycle which in some cases are shorter than the life cycle of the core users contributing to it. As Table 3 shows, the average `Talk` and `Task` lifespan of an active user across the ten different Zooniverse projects is longer than the duration of seven out of the ten projects. Given the different project launch dates we assume that housing multiple citizen science projects within one platform is beneficial for the participation and sustainability of Zooniverse users.

Furthermore, the participation of users across projects is dependent on a number of factors, including the subject domain of the project. Unlike other online communities (Whittaker et al. 1998a; Zhongbao and Changshui 2003; Adamic et al. 2008), our analysis shows that citizen scientists tend to participate in projects that are similar in domain (i.e., Astrophysics, Nature), both in terms of discussion and classification (see Table) We also observed that in addition to project domain, factors such as the age of the project and size of the user base affects user cross-over. Older projects such as `PH`, which contained the largest user base, had the highest proportion of cross-over participation, which suggest that their user base support the growth of other projects, potentially acting as a mechanism to overcome cold start problems. We consider these types of projects as 'home bases' which contain established users that exhibit a different 'code of conduct' in comparison to new users who may have joined on the rise of citizen science and Zooniverse.

Do Discussion Forums Facilitate Learning?

It has been often asked whether and to what degree novice volunteers of citizen science projects gain knowledge through participation (e.g. (Rotman et al. 2012)) or the consumption of forum contents (Mugar, Osterlund, and Hassman 2014), and whether this forms a motivation for continued participation. In order to answer this question, we looked for evidence of expertise transfer from experts to novice participants, or group learning, in which the scientific vocabulary and concepts transferred from some users to others, or transmuted over time as understanding of concepts changed. At the outset, we initially realised that the projects were not all similar in use of scientific language; some heavily used domain-specific vocabulary in their descriptions from launch, while others used more familiar terminology in their descriptions, leaving more space for expert vocabulary acquisition over their lifecycle.

In respect to the former group, `PH`, `SW`, `GZ`, and `PF` featured a core community with a significant proportion of users with a very stable vocabulary from from the initial launch, and remain constant throughout. We take the position that users within these projects were well-trained or already familiar in the subject domain given the users began with specific domain vocabulary when the projects were launched. We also consider the stability and use of expert vocabulary as a consequence of the natural connection of the Zooniverse platform to the astrophysics domain, where there was an already well-established community of individuals participating in projects such as Seti@Home (Anderson et al. 2002) and SpaceScience@Home (Méndez 2008). We also observed that SF and SG users were already using a domain specific vocabulary in the beginning of the project lifecycle (fish, scallop, lion, zebra), but one can still recognize an evolution towards an even higher density of names of very specific species (corallinealga, hermitcrab, cerianthid, bryozoan) suggesting that the users improved their knowledge about the domain.

In contrast to this, the results in Table 5 and Figure 3 indicate that users participating in CC, and NN did not feature a steady vocabulary during the project lifecycle. For NN and CC this is shown by the high proportion of help requests which address diverse problems and require specific answers provided by a core of supportive users. These findings support the initial board and thread analysis, which identified a higher proportion of threads in the help and chat boards.

We also question the complexity of the project's Task in relation to the shift in vocabulary. In projects such as SG, users are asked to identify 48 species from an image, where as in projects such as CC and NN, users are asked to complete complex Tasks, which we assume require a higher level of engagement. This may be the reason for the shift in vocabulary towards language related to general user interface terminology (field, click, record).

Conclusions

In this paper we explored the phenomenon of citizen science through the lens of the Zooniverse platform. We examined ten different projects and analysed over 50 million activities (both discussions and classifications) performed by over 250,000 users. The central contributions of this study are the insights gained in understanding (i) the relationships between Talk and Talk activities of citizen scientists; (ii) the factors that affect user engagement across multiple projects; and (iii) the adoption and transferability of expertise and specialized terminology though sustained participation.

By analysing the involvement of users in Task and Talk, we found that a less than 1% of the total user base represented the highly active users, responsible for over 70% of the discussion posts in Talk, and 29% of the Tasks performed. Whilst Task seems to exhibit 'wisdom of the crowd' properties similar to other crowdsourcing systems (Ortega, Gonzalez-Barahona, and Robles 2008; Wang et al. 2013), Talk could be characterized as 'wisdom of the few' - despite obtaining a large user-base, only a small minority of the citizen scientists consistently engage in discussions. We also noted how domain, size, and, the age/duration of a project has a positive impact on the extent to which the underlying community is willing to contribute to multiple projects, in both Talk and Task activities. While this type of behavior is common to other information sharing systems, including bulletin boards (Zhongbao and Changshui 2003), StackOverflow (Mamykina et al. 2011), and Quora (Wang et al. 2013), our findings suggest that Zooniverse users are more likely to participate in projects within a similar subject domain, rather than crossover into unfamiliar territory. In addition, the community created around some of the early (and arguably most successful) Zooniverse projects seem to be motivated to contribute to a wider range of projects in the Zooniverse family; by comparison, newer projects exhibit a higher variability of users. A final building block of our analysis looked into the learning effects created through participation in Talk and interaction with more expert users, as observed via the adoption of an increasingly specialized vocabulary of novice users. We observed two profiles in respect to the adoption of scientific terminology, projects that started with a online community that were already using highly specific terminology which remained steady throughout the project's lifecycle, and projects which acquired domain specific vocabulary over their lifecycle.

Our findings have implications not only for understanding and analysing the phenomenon of citizen science on the Web, but also for improving our knowledge of the participation of users of other online communities. We consider citizen science as a phenomenon which sits at the intersection of crowdsourcing, collective intelligence, and online community building, and argue that the study of platforms such as Zooniverse will offer rich insights into the participation and sustainability of users. Future work includes experiments to measure and predict the user engagement and to examine in more depth the activity profiles of users to help better understand the characteristics of successful citizen scientists. Ultimately, our work in online citizen science is also motivated by the furore around serendipitous citizen-led discoveries, resulting from the autonomous usage of socialisation and discussion facilities. This study revealed that the use of the microblogging feature in the Zooniverse can be interpreted as a kind of a user-driven coordination around hypothesis, because of the emergent use of tags. However, the question remains if this promotes or impedes the benefits of collaborative problem-solving and discovery in the discussion forums and leaves space for future research.

Acknowledgements

This work was supported by the EPSRC Theory and Practice of Social Machines Programme Grant, EP/J017728/1. We thank the anonymous reviewers for the outstandingly helpful comments on the paper.

References

Adamic, L. A.; Zhang, J.; Bakshy, E.; and Ackerman, M. S. 2008. Knowledge sharing and yahoo answers: Everyone knows something. In *17th International Conference on World Wide Web*, WWW '08, 665–674. ACM.

Agichtein, E.; Castillo, C.; Donato, D.; Gionis, A.; and Mishne, G. 2008. Finding high-quality content in social media. In *2008 International Conference on Web Search and Data Mining*, WSDM '08, 183–194. ACM.

Anderson, B. D. P.; Cobb, J.; Korpela, E.; and Lebofsky, M. 2002. SETI home which computing resources are. *Communications of the ACM* 45(11).

Anderson, A.; Huttenlocher, D.; Kleinberg, J.; and Leskovec, J. 2012. Discovering value from community activity on focused question answering sites: A case study of stack overflow. In *18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '12, 850–858. ACM.

Arguello, J.; Butler, B. S.; Joyce, E.; Kraut, R.; Ling, K. S.; Rosé, C.; and Wang, X. 2006. Talk to me: Foundations for successful individual-group interactions in online communities. In *SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, 959–968. New York, NY, USA: ACM.

Backstrom, L.; Huttenlocher, D.; Kleinberg, J.; and Lan, X. 2006. Group formation in large social networks: Membership, growth, and evolution. In *12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '06, 44–54. New York, NY, USA: ACM.

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

Butler, B.; Joyce, E.; and Pike, J. 2008. Don't look now, but we've created a bureaucracy: the nature and roles of policies and rules in wikipedia. In *SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, 1101–1110. ACM.

Campbell, C. S.; Maglio, P. P.; Cozzi, A.; and Dom, B. 2003. Expertise identification using email communications. In *12th International Conference on Information and Knowledge Management*, CIKM '03, 528–531. ACM.

Clauset, A.; Newman, M. E. J.; and Moore, C. 2004. Finding community structure in very large networks. *Physical Review E* 70:066111.

Cohn, J. P. 2008. Citizen science: Can volunteers do real research? *BioScience* 58(3):192–197.

Cooper, C. 2013. Zen in the art of citizen science: Apps for collective discovery and the 1 percent rule of the web. `http://blogs.scientificamerican.com/guest-blog/2013/09/11/zen-in-the-art-of-citizen-science-apps-for-collective-discovery-and-the-1-rule-of-the-web/`.

Danescu-Niculescu-Mizil, C.; West, R.; Jurafsky, D.; Leskovec, J.; and Potts, C. 2013. No country for old members: user lifecycle and linguistic change in online communities. In Schwabe, D., et al., eds., *WWW*, 307–318. ACM.

Fisher, D.; Smith, M.; and Welser, H. T. 2006. You are who you talk to: Detecting roles in usenet newsgroups. In *39th Annual Hawaii International Conference on System Sciences*, HICSS '06, 59.2–. IEEE Computer Society.

Fortson, L.; Masters, K.; Nichol, R.; Borne, K.; Edmondson, E.; Lintott, C.; Raddick, J.; Schawinski, K.; and Wallin, J. 2011. Galaxy zoo: Morphological classification and citizen science. *arXiv preprint arXiv:1104.5513*.

Girvan, M., and Newman, M. E. J. 2002. Community structure in social and biological networks. *National Academy of Sciences of the United States of America* 99(12):7821–6.

Gray, S. A.; Nicosia, K.; and Jordan, R. C. 2012. Lessons learned from citizen science in the classroom. a response to" the future of citizen science.". *Democracy and Education* 20(2):14.

Harper, F. M.; Raban, D.; Rafaeli, S.; and Konstan, J. A. 2008. Predictors of answer quality in online q&a sites. In *SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, 865–874. ACM.

Heinzelman, J., and Waters, C. 2010. Crowdsourcing crisis information in disaster-affected haiti. `http://www.humanitarianforum.org/data/files/crowdsourcinghaiti.pdf`.

Irwin, A. 1995. *Citizen science: A study of people, expertise and sustainable development*. Psychology Press.

Jordan, R. C.; Gray, S. A.; Howe, D. V.; Brooks, W. R.; and Enrenfeld, J. G. 2011. Knowledge gain and behavioral change in citizen-science programs. *Conservation Biology* 25(6):1148–1154.

Joyce, E., and Kraut, R. E. 2006. Predicting continued participation in newsgroups. *Journal of Computer-Mediated Communication* 11:2006.

Kairam, S. R.; Wang, D. J.; and Leskovec, J. 2012. The life and death of online groups: predicting group growth and longevity. In Adar, E.; Teevan, J.; Agichtein, E.; and Maarek, Y., eds., *WSDM*, 673–682. ACM.

Kawrykow, A.; Roumanis, G.; Kam, A.; Kwak, D.; Leung, C.; Wu, C.; Zarour, E.; Sarmenta, L.; Blanchette, M.; Waldispühl, J.; et al. 2012. Phylo: a citizen science approach for improving multiple sequence alignment. *PloS one* 7(3):e31362.

Khatib, F.; Cooper, S.; Tyka, M. D.; Xu, K.; Makedon, I.; Popović, Z.; Baker, D.; and Players, F. 2011. Algorithm discovery by protein folding game players. *Proceedings of the National Academy of Sciences* 108(47):18949–18953.

Kittur, A.; Chi, E.; Pendleton, B.; Suh, B.; and Mytkowicz, T. 2007a. Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. *World Wide Web* 1(2):19.

Kittur, A.; Suh, B.; Pendleton, B. A.; and Chi, E. H. 2007b. He says, she says: conflict and coordination in wikipedia. In *SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, 453–462. ACM.

Krieger, M.; Stark, E. M.; and Klemmer, S. R. 2009. Coordinating tasks on the commons: Designing for personal goals, expertise and serendipity. In *SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, 1485–1494. New York, NY, USA: ACM.

Kumar, R.; Novak, J.; and Tomkins, A. 2006. Structure and evolution of online social networks. In *12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '06, 611–617. New York, NY, USA: ACM.

Lampe, C., and Johnston, E. 2005. Follow the (slash) dot: Effects of feedback on new members in an online community. In *2005 International ACM SIGGROUP Conference on Supporting Group Work*, GROUP '05, 11–20. ACM.

Lintott, C. J., et al. 2008. Galaxy zoo: morphologies derived from visual inspection of galaxies from the sloan digital sky survey. *Monthly Notices of the Royal Astronomical Society* 389(3):1179–1189.

Mamykina, L.; Manoim, B.; Mittal, M.; Hripcsak, G.; and Hartmann, B. 2011. Design lessons from the fastest q&a site in the west. In *SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, 2857–2866. New York, NY, USA: ACM.

Méndez, B. 2008. SpaceScienceHome: Authentic Research Projects that Use Citizen Scientists. *EPO and a Changing World: Creating Linkages and Expanding Partnerships* 389:219.

Merkel, C. B.; Xiao, L.; Farooq, U.; Ganoe, C. H.; Lee, R.; Carroll, J. M.; and Rosson, M. B. 2004. Participatory design in community computing contexts: Tales from the field. In *8th Conference on Participatory Design: Artful Integration: Interweaving Media, Materials and Practices - Volume 1*, PDC 04, 1–10. New York, NY, USA: ACM.

Mugar, G.; Osterlund, C.; and Hassman, K. 2014. Planet Hunters and Seafloor Explorers: Legitimate Peripheral Participation Through Practice Proxies in Online Citizen Science. In *CSCW'14*.

Nov, O.; Arazy, O.; and Anderson, D. 2011. Dusting for science: motivation and participation of digital citizen science volunteers. In *iConference*, 68–74.

Ortega, F.; Gonzalez-Barahona, J. M.; and Robles, G. 2008. On the inequality of contributions to wikipedia. In *41st Annual Hawaii International Conference on System Sciences*, HICSS '08, 304–. Washington, DC, USA: IEEE Computer Society.

Quinn, A. J., and Bederson, B. B. 2011. Human computation: a survey and taxonomy of a growing field. In *SIGCHI Conference on Human Factors in Computing Systems*, 1403–1412.

Raddick, M. J.; Bracey, G.; Gay, P. L.; Lintott, C. J.; Murray, P.; Schawinski, K.; Szalay, A. S.; and Vandenberg, J. 2010. Galaxy zoo: Exploring the motivations of citizen science volunteers. *Astronomy Education Review* 9:010103.

Rafaeli, S.; Ravid, G.; and Soroka, V. 2004. De-lurking in virtual communities: A social communication network approach to measuring the effects of social and cultural capital. In *37th Annual Hawaii International Conference on System Sciences (HICSS'04)*, HICSS '04, 70203–. IEEE Computer Society.

Rotman, D.; Preece, J.; Hammock, J.; Procita, K.; Hansen, D.; Parr, C.; Lewis, D.; and Jacobs, D. 2012. Dynamic changes in motivation in collaborative citizen-science projects. In *ACM 2012 Conference on Computer Supported Cooperative Work*, 217–226. ACM.

Rowe, M.; Wagner, C.; Strohmaier, M.; and Alani, H. 2013. Measuring the topical specificity of online communities. In Cimiano, P.; Corcho, .; Presutti, V.; Hollink, L.; and Rudolph, S., eds., *ESWC*, volume 7882 of *Lecture Notes in Computer Science*, 472–486. Springer.

Silvertown, J. 2009. A new dawn for citizen science. *Trends in ecology & evolution* 24(9):467–471.

Wang, G.; Gill, K.; Mohanlal, M.; Zheng, H.; and Zhao, B. Y. 2013. Wisdom in the social crowd: An analysis of quora. In *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13, 1341–1352. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee.

Welser, H. T.; Gleave, E.; Fisher, D.; and Smith, M. 2007. Visualizing the signatures of social roles in online discussion groups. *The Journal of Social Structure* 8(2).

Westphal, A. J., et al. 2005. Stardust@ home: a massively distributed public search for interstellar dust in the stardust interstellar dust collector. In *36th Annual Lunar and Planetary Science Conference*, volume 36, 1908.

Whittaker, S.; Terveen, L.; Hill, W.; and Cherny, L. 1998a. The dynamics of mass interaction. In *1998 ACM Conference on Computer Supported Cooperative Work*, CSCW '98, 257–264. New York, NY, USA: ACM.

Whittaker, S.; Terveen, L.; Hill, W.; and Cherny, L. 1998b. The dynamics of mass interaction. In *1998 ACM conference on Computer supported cooperative work*, CSCW '98, 257–264. ACM.

Wiggins, A., and Crowston, K. 2010. Developing a conceptual model of virtual organisations for citizen science. *International Journal of Organisational Design and Engineering* 1(1):148–162.

Wiggins, A., and Crowston, K. 2011. From conservation to crowdsourcing: A typology of citizen science. In *System Sciences (HICSS), 2011 44th Hawaii International Conference on*, 1–10. IEEE.

Wilson, E. O. 2003. The encyclopedia of life. *Trends in Ecology & Evolution* 18(2):77–80.

Yang, J., and Leskovec, J. 2012. Defining and evaluating network communities based on ground-truth. *CoRR* abs/1205.6233.

Yu, J.; Wong, W.-K.; and Hutchinson, R. A. 2010. Modeling experts and novices in citizen science data for species distribution modeling. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, 1157–1162. IEEE.

Zhang, J.; Ackerman, M. S.; and Adamic, L. 2007. Expertise networks in online communities: structure and algorithms. In *16th international conference on World Wide Web*, 221–230.

Zheng, Z. 2002. Answerbus question answering system. In *2nd International Conference on Human Language Technology Research*, HLT '02, 399–404. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Zhongbao, K., and Changshui, Z. 2003. Reply networks on a bulletin board system. *Physical review. E, Statistical, nonlinear, and soft matter physics* 67(3 Pt 2):036117.

Zook, M.; Graham, M.; Shelton, T.; and Gorman, S. 2010. Volunteered geographic information and crowdsourcing disaster relief: a case study of the haitian earthquake. *World Medical & Health Policy* 2(2):7–33.