Collective Intelligence in Citizen Science – A Study of Performers and Talkers

Ramine Tinati, Max Van Kleek, Elena Simperl, Markus Luczak-Roesch, Nigel Shadbolt, University of Southampton

1. INTRODUCTION

Online citizen science [Silvertown 2009] can be seen as a form of collective intelligence Lévy [1997] and Woolley et al. [2010] in which the wisdom of the crowd is applied to the Web to advance scientific knowledge [Prestopnik and Crowston 2012] Thus far, online citizen science projects [Bonney et al. 2009; Gray et al. 2012] have applied millions of volunteers to solving problems in a wide array of scientific domains, ranging from the classification of galaxies [Fortson et al. 2011] to the completion of protein folding networks [Khatib et al. 2011].

Central to many of these projects are online messaging or discussion facilities designed to allow volunteers to ask one another questions and advice. Such facilities have in many cases yielded substantial, dedicated self-sustaining online communities. In this paper, we examine participation in such communities; specifically, whether participation in online discussion influences task completion within and across 10 distinct projects of a shared citizen science platform, the Zooniverse[1]. Our study was conducted on a dataset from December 2009 to July 2013, in which $250,000$ users contributed over 50 million tasks and $650,000$ discussion posts.

2. RELATED WORK

The hybrid nature of online citizen science, as part online community and part crowdsourcing platform, places it at a unique position at the intersection of two large and active research communities. Moreover, since both components work in tandem to yield outcomes that essentially would not have been possible without the other and the collective insights of the group, they can be considered as exhibiting collective intelligence, as described by [Skyring 2013].

Online communities have been studied from a variety of perspectives, from that of the individual to the entire system [Arguello et al. 2006], as peer-production information sharing repositories [Krieger and Stark 2009; Slattery 2009]. The time-evolution of such communities have also been studied [Kumar et al. 2006]. Specific focus on question-answering platforms [Harper et al. 2009], have considered questions such as; answer quality [Agichtein et al. 2008], topics coverage and language [Rowe et al. 2013], user profiles, roles, and detecting and measuring user expertise [Fisher et al. 2006; Pal et al. 2012].

Citizen science systems can be roughly characterised as being primarily concerned with citizen data collection [Zook et al. 2010], data analysis [Heinzelman and Waters 2010], and problem-solving [Kawrykow et al. 2012]. They generally fulfill three purposes: solving large-scale and complex tasks, creating and sustaining a community of amateur scientists and contributors for future projects, and educating users on science and scientific methods. Whilst studies have examined motivations for participating in these systems [Raddick et al. 2010; Rotman et al. 2012], there has yet been work documenting the process of scientific discovery as a collective intelligence phenomenon.

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

3. THE ZOONIVERSE PLATFORM

'Zooniverse' is a citizen science Web platform which hosts, as of January 2014, 30 separate citizen science projects spanning several scientific and humanities domains. Participants contribute to projects by performing data classification and analysis tasks on digital artifacts, comprising images, video and audio recordings. Each project is linked with a set of discussion forums and messaging facilities called *Talk*, which serve as the main tool for information sharing among participants.

4. RESULTS

First we summarise an analysis of the relationship between Talk participation and tasks performed across 10 Zooniverse projects. The analysis described here is a summary of early observations resulting from work[2] which provides a more thorough, multi-perspective analysis.

4.1 Talk versus Tasks: Do Those Who Discuss Contribute More?

We found that of the $250,071$ users in our data, only 40.5% of them had contributed both classifications and discussions. Since some of the users may have not been aware of the discussion facilities, we restricted our analysis to to those who posted at least once to avoid skewing our results. Figure 1 compares the number of classifications performed per user (x-axis) with the number of posts they made in the Talk system (y-axis). While the overall positive trend suggests that those who are generally more active perform more posts and tasks, participants completed vastly more tasks than made posts, performing a median 600 classifications versus 14 (

However, there exists a collection of users that not only have performed a much larger number of classifications, but also have participated extensively in the discussion forms. To focus on such highly active users, we extracted a subset that had performed at least 140 posts and 6000 classifications. This contained 2928 users across all projects, which in total represented 29.0% and 72.0% of the total number classification and discussions in our data set, respectively. Within this set, we found three different pattern of users, (1) users that sporadically perform classifications, (2) users that consistently perform classifications, (3) users that perform classifications very infrequently. Then, by clustering the users by their frequency of Talk entries and task completion, we identified five characteristic user types illustrated in Figure I.

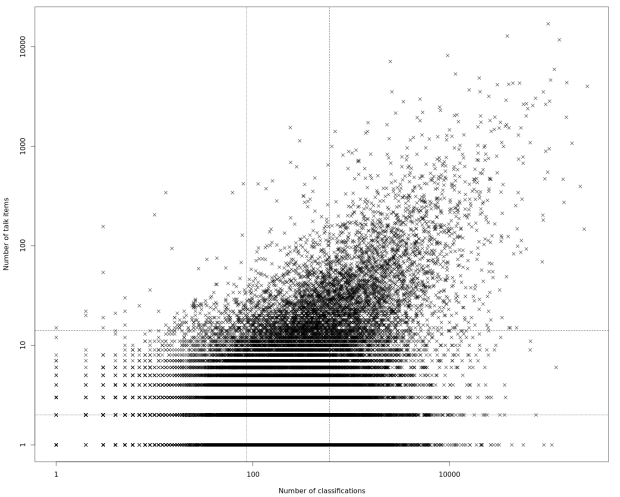

Fig. 1: Number of classifications (x-axis) and talk contributions (y-axis) of every single user.

4.2 Roles in the Discussion Forums

A second analysis pertained to roles played by participants in Talk discussions. While superficially, much of the role of participants in forum discussions consisted of fielding and answering questions, such questions and the conversations around them often served different purposes. Figure II illustrates an initial set of roles we identified, which illustrates a considerable breadth in the kinds of activities

[2]Zooniverse: Report on Talk and Task Analysis (technical report, unpublished) http://eprints.soton.ac.uk/id/eprint/361204

User type	Description
Casual Hobbyists	Moderate task completion making occasional Talk entries sustained over a long time, with slow participation decline
Short but Sweet	Highly active in both task and Talk, but short-lived participation
Birth of a Moderator	More classifications at the beginning with increasing involvement in Talk activities over time
Performer to Talker	High number of tasks performed to begin with, then starts to use Talk instead
Talker to Performer	Frequent use of Talk before engaging with tasks

Table I. : User behaviour stereotypes identified among the most active participants.

performed. While these roles overlap considerably with roles identified in online QA communities such as [Pal et al. 2012], the emergent roles of *Discoverer*, *Hypothesiser*, *Investigator*, were distinctive to citizen science that were not only crucial to the collaborative hypothesis process, but demonstrated that the citizens themselves were engaging in a scientific process at some level, rather than blindly performing tasks on behalf of a scientist.

Participant Role	Description
General Help Asker	Asked a general clarification or help question
Answerer	Answers a question with a definitive answer.
Informer	Posts unsolicited (but often authoritative) information to the community.
Moderator	Encourage/discourage certain posting behaviour, redirecting participants to a different thread or location where
Discoverer	Asked a question ultimately resulted in an unprecedented discovery
Hypothesiser	Poses a scientific question or proposes a possible answer.
Investigator/Validator	Took hypotheses posed and did further investigation/validation, presenting these results
Cheerleader	Encourages group to continue performing tasks, such as to reach milestones
Celebrator	Points to interesting objects for their aesthetic/fun value

Table II. : Roles in discussion forums

5. CONCLUSIONS

In this paper, we examined the phenomenon of citizen science through the lens of the Zooniverse platform. The central contributions are the insights gained in understanding the relationship between Talk and task completion in respects to online citizen science, and the discovery if emergent user roles. We discovered a relationship between those that completed tasks and participated in Talk discussions and found a set of 'active' users which were responsible for over 70% of the Talk content, thus assuming the role of the 'core community'. These results reflect other peer-production systems like Wikipedia [Kittur et al. 2007; Ortega 2008]; despite obtaining a large user-base, it is the activities of only a relatively small collection of users that produce content.

In addition to this study, our ongoing research of online citizen science has found noticeable cross-project phenomenon; we are discovering how language of communities evolves relative to their subject domain, and how serendipitous discoveries are achieved via collective behaviour. Ultimately, this has wider implications for understanding collective intelligence on the Web. Given that the current landscape of crowdsourcing is somewhat of a disjoint collection of communities, our findings suggest that it is beneficial to support and unite online communities within and between different crowdsourcing systems, independent of subject domain or topic.

REFERENCES

Eugene Agichtein, Carlos Castillo, Debora Donato, Aristides Gionis, and Gilad Mishne. 2008. Finding high-quality content in social media. In *Proceedings of the 2008 International Conference on Web Search and Data Mining (WSDM '08)*. ACM, 183–194.

Jaime Arguello, BS Butler, and Elisabeth Joyce. 2006. Talk to me: foundations for successful individual-group interactions in online communities. *Proceedings of the ...* (2006), 1–10. http://dl.acm.org/citation.cfm?id=1124916

Rick Bonney, Caren B Cooper, Janis Dickinson, Steve Kelling, Tina Phillips, Kenneth V Rosenberg, and Jennifer Shirk. 2009. Citizen science: a developing tool for expanding science knowledge and scientific literacy. *BioScience* 59, 11 (2009), 977–984.

Danyel Fisher, Marc Smith, and Howard T. Welser. 2006. You Are Who You Talk To: Detecting Roles in Usenet Newsgroups. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences - Volume 03 (HICSS '06)*. IEEE Computer Society, 59.2–.

Lucy Fortson, Karen Masters, Robert Nichol, Kirk Borne, Edd Edmondson, Chris Lintott, Jordan Raddick, Kevin Schawinski, and John Wallin. 2011. Galaxy Zoo: Morphological classification and citizen science. *arXiv preprint arXiv:1104.5513* (2011).

Steven A Gray, Kristina Nicosia, and Rebecca C Jordan. 2012. Lessons Learned from Citizen Science in the Classroom. A Response to" The Future of Citizen Science.". *Democracy and Education* 20, 2 (2012), 14.

F. Maxwell Harper, Daniel Moy, and Joseph A. Konstan. 2009. Facts or Friends?: Distinguishing Informational and Conversational Questions in Social Q&A Sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 759–768. DOI:http://dx.doi.org/10.1145/1518701.1518819

Jessica Heinzelman and Carol Waters. 2010. Crowdsourcing crisis information in disaster-affected Haiti. http://www.humanitarianforum.org/data/files/crowdsourcinghaiti.pdf. (2010).

Alexander Kawrykow, Gary Roumanis, Alfred Kam, Daniel Kwak, Clarence Leung, Chu Wu, Eleyine Zarour, Luis Sarmenta, Mathieu Blanchette, Jérôme Waldispühl, and others. 2012. Phylo: a citizen science approach for improving multiple sequence alignment. *PloS one* 7, 3 (2012), e31362.

Firas Khatib, Seth Cooper, Michael D Tyka, Kefan Xu, Ilya Makedon, Zoran Popović, David Baker, and Foldit Players. 2011. Algorithm discovery by protein folding game players. *National Academy of Sciences* 108, 47 (2011), 18949–18953.

Aniket Kittur, Ed Chi, Bryan A Pendleton, and Todd Mytkowicz. 2007. Power of the Few vs . Wisdom of the Crowd : Wikipedia and the Rise of the Bourgeoisie. *Algorithmica* (2007), 1–9.

Michel Krieger and Emily Stark. 2009. Coordinating tasks on the commons: designing for personal goals, expertise and serendipity. *Proceedings of the 27th International Conference on Human Factors in Computing Systems (CHI)* (2009), 1485–1494.

Ravi Kumar, Jasmine Novak, and Andrew Tomkins. 2006. Structure and evolution of online social networks. *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '06* (2006), 611.

Pierre Lévy. 1997. *Collective Intelligence: Mankind's Emerging World in Cyberspace*. Vol. 1. 277 pages. DOI:http://dx.doi.org/10.1109/MCI.2006.1672982

Felipe Ortega. 2008. On the inequality of contributions to Wikipedia. ... *Conference on System* ... (2008), 1–7. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4439009

Aditya Pal, Shuo Chang, and Joseph A Konstan. 2012. Evolution of Experts in Question Answering Communities.. In *ICWSM*.

Nathan Prestopnik and K Crowston. 2012. Exploring collective intelligence games with design science: A citizen science design case. In *ACM Group Conference*. http://crowston.syr.edu/system/files/designingcitizensciencegames.pdf

M Jordan Raddick, Georgia Bracey, Pamela L Gay, Chris J Lintott, Phil Murray, Kevin Schawinski, Alexander S Szalay, and Jan Vandenberg. 2010. Galaxy zoo: Exploring the motivations of citizen science volunteers. *Astronomy Education Review* 9 (2010), 010103.

Dana Rotman, Jenny Preece, Jen Hammock, Kezee Procita, Derek Hansen, Cynthia Parr, Darcy Lewis, and David Jacobs. 2012. Dynamic changes in motivation in collaborative citizen-science projects. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. ACM, 217–226.

Matthew Rowe, Claudia Wagner, Markus Strohmaier, and Harith Alani. 2013. Measuring the Topical Specificity of Online Communities.. In *ESWC (Lecture Notes in Computer Science)*, Philipp Cimiano, scar Corcho, Valentina Presutti, Laura Hollink, and Sebastian Rudolph (Eds.), Vol. 7882. Springer, 472–486. http://dblp.uni-trier.de/db/conf/esws/eswc2013.html#RoweWSA13

Jonathan Silvertown. 2009. A new dawn for citizen science. *Trends in ecology & evolution* 24, 9 (Sept. 2009), 467–71. DOI:http://dx.doi.org/10.1016/j.tree.2009.03.017

Kerry Skyring. 2013. Collective Intelligence: has the golden age of the 'citizen scientist' finally arrived? *dw.de* (Oct. 2013). http://dw.de/p/1A30R

SP Slattery. 2009. Edit this page: the socio-technological infrastructure of a wikipedia article. In *Proceedings of the 27th ACM international conference on Design of communication*. 289–295. http://dl.acm.org/citation.cfm?id=1622052

AW Woolley, CF Chabris, and A Pentland. 2010. Evidence for a collective intelligence factor in the performance of human groups. *science* October (2010), 686–688. http://www.sciencemag.org/content/330/6004/686.short

Matthew Zook, Mark Graham, Taylor Shelton, and Sean Gorman. 2010. Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake. *World Medical & Health Policy* 2, 2 (2010), 7–33.