

Event Based Analysis of a Citizen Science Community: Are New and Non-sustained Users Included?

Corey Jackson¹

¹ School of Information Studies, Syracuse University

Abstract

High turnover rates in online communities suggest the need for measures that move non-sustained community members towards sustained participation. Non-sustained members actively seek inclusion opportunities, but are often disappointed by the lack of access to existing members. One method of inclusion can be contact with existing members through dialogue on discussion boards. Understanding the structure of interactions between sustained and non-sustained members and can inform new strategies to address the high turnover rate and ensure community longevity. In this poster, we analyze the network structure of newcomer and existing member interaction through discussion posts. Through analysis of a citizen science community we ask: Are non-sustained and sustained participants engaged in conversations? The researcher analyzes the topological features of an affiliation network and centrality measures to determine the extent of interactions between these two groups. Finally, the researcher presents strategies to engage non-sustained participants in online.

Keywords: affiliation networks, network analysis, citizen science, newcomers, online communities

Citation: Jackson, C. (2014). Event Based Analysis of a Citizen Science Community: Are New and Non-sustained Users Included? In *iConference 2014 Proceedings* (p. 1139–1144). doi:10.9776/14405

Copyright: Copyright is held by the author.

Acknowledgements: This project was partially supported by the US National Science Foundation under grant number 1211071. I would also like to thank the Zooniverse Research team at Adler Planetarium and Syracuse University

Contact: cjacks04@syr.edu

1 Introduction

Sites relying on volunteer contribution must maintain an active pool of participants to ensure work is complete. Burke (2009) noted the high turnover rates, especially for new comers to online communities, which highlight the importance of early newcomer involvement in the community. Participant access and interaction with sustained users and more knowledgeable community members such as moderators and domain experts serves to orient non-sustained users to communities of practice and thus sustain their membership in them. These conversations and acknowledgment of work become essential for building relationships with community members (Griffin, Colella, & Goparaju, 2000) and enhancing participant learning of community norms.

Burke et al. also suggest newcomers actively look for ways to be included in a community and one way in which they can be involved is by posting a message and assessing their role in the community through replies (Burke, Kraut, & Joyce, 2009). The same research also showed community responsiveness would often lead to sustained participation. A similar study by (Arguello, 2006) found newcomers to Usenet groups are more likely to come back for subsequent visits if others reply to their comments. It is also known that the effects of newcomer sustained community involvement are stronger if more sustained users respond to newcomers (Kraut & Resnick, Under contract) Online communities vary in their responses to newcomers and some offer guides on how to build successful relationships with newcomers, for the purposes of fostering their confidence in contributing. In fact, Wikipedia has a page dedicated to how to handle new comers titled: *Wikipedia: Please do not bite the newcomers*, which gives instructions in newcomer interaction.

Given the importance of newcomer interaction, this research is concerned primarily with the responsiveness of a community of practice to non-sustained users in a citizen science project from the Citizen

Science Alliance’s Zooniverse suite of projects titled Planet Hunters. In this community, citizen scientist – amateurs and professionals alike are asked to annotate images to identify extrasolar planets. An example of the interface can be found in [Image 1]. Participants are asked to identify dips in light emissions from planets, which could highlight the presence of other transiting planets. After identifying the dips users are prompted to talk with other community members about the image they annotated [Image 2]. Participants can see what others have said about the image and more importantly ask questions about the image features which contributes to their understanding of the task and knowledge of the science about the project making this an important place for the exchange of information.

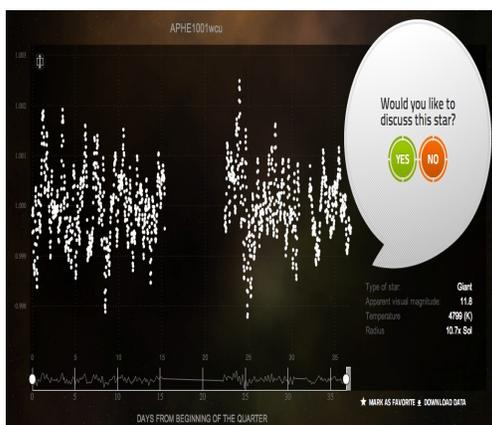


Figure 1: Annotation Interface

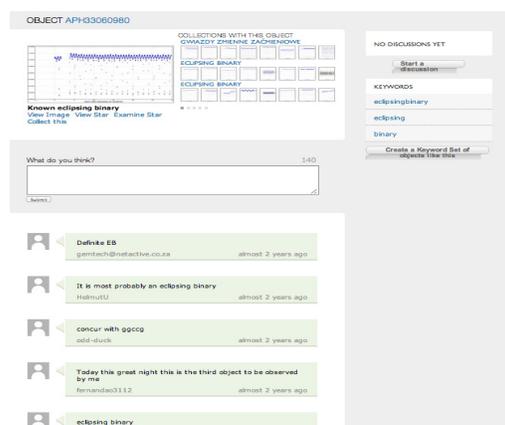


Figure 2: Discussion Interface

Traditional network analysis has focused on characteristics of nodes while removing context from the interaction. Combining people and the context through which they are connected could provide additional insight and allows one to make claims about the structure of the network and integrations with community participants (Easley and Kleinberg, 2010). Affiliation networks allow us to represent social interactions among collections of actors (Carrington, Scott, & Wasserman, 2004) through shared membership in a contextual manner. The research analyzes discussion posts to see if new comers and sustained users are conversing in a joint space. This method supports different perspectives on linkages between actors and events (Carrington, Scott, & Wasserman, 2004). In this case, membership in discussions serves to link actors to collectives in which they participate and link non-sustained participants to sustained community members. One of the earliest examples of affiliation networks was presented in a study by (Davis, Gardner, Gardner 1941) where networks of southern women attending parties were analyzed to determine class structure. In this research, using affiliation networks to characterize the network structure of discussion and participants is an appropriate method of analysis as we seek to answer our research questions: Are non-sustained and sustained users engaged in conversations? This work is purely descriptive and seeks to understand how one category of users is related to another.

2 Methods

The researcher used Cytoscape to construct an affiliation network of a dataset consisting of logs of discussion board posts from non-sustained participants and sustained users who appeared in the same posts during July 2013. Data was collected from members who are defined as non-sustained and the sustained participants who commented on their posts. In the case of Planet Hunters, sustained users are described as those who have made more than 1000 contributions in the form of image classifications. The choice to employ 1000 classifications as a distinctive feature which separates non-sustained and sustained participants emerged from interviews with sustained users who mentioned only after 1000 classifications did they become

competent in the task. Information about the amount of classifications a user contributed was also included in the dataset.

3 Results

The network can be visualized in image 3 and exhibits one large network and a number of isolated nodes. The thickness of edges represents comments from users who are sustained as established by our 1000 classifications or greater threshold. The network is comprised of distinct 1395 nodes and 1343 directed edges extending from nodes of participants. It is important to note that since the graph is directed no edges extend from discussion posts. The nodes in image 3 are of discussion posts and participants. The edges are weighted for participants with over 1000 classifications and the weights are further described by coloring the edges according the number of classifications within that group where the its scales from darkest to lightest as the number of classifications ascend.

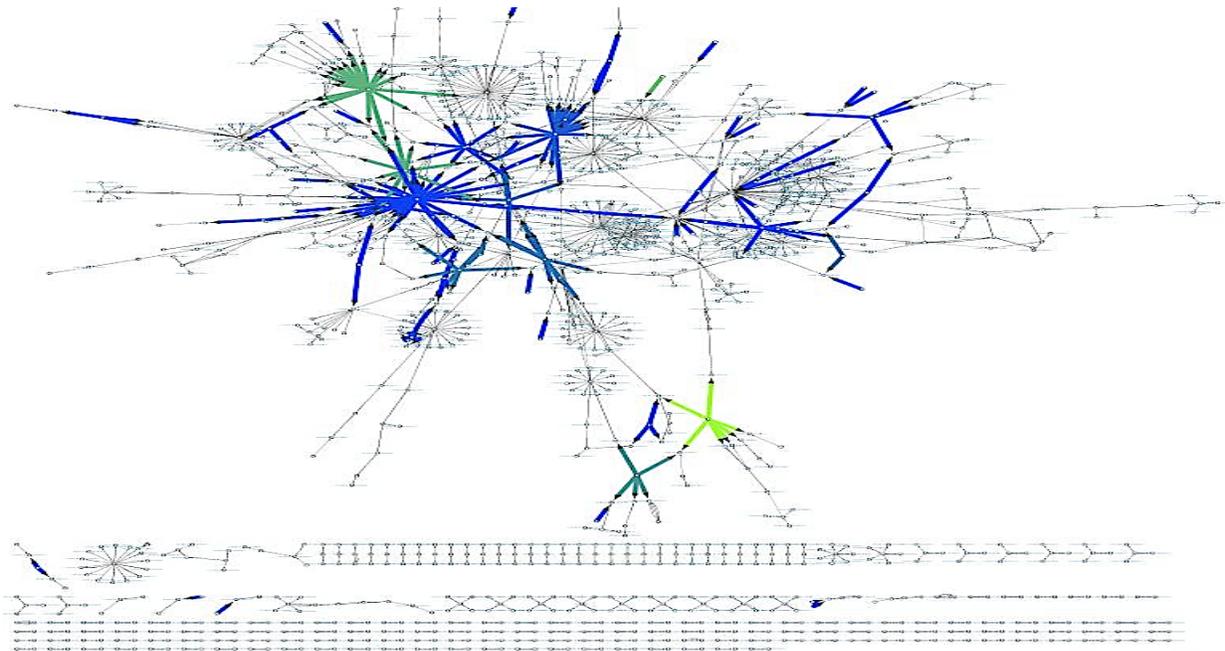


Figure 3: Affiliation Network of User Comments.

Interesting topological features have emerged in the network. Looking first at the primary network, a number of edges extending from the network, which are not bold, and many bolded edges in the central part of the network. Most edge weights are closely knit and few non-weighted edges exist in the central part of the network, which suggest a close knit between sustained users who comment. Table 1 also supports the lack of appearance of sustained users in conversations involving new comers. Given the production from sustained participants it is appropriate to expect small number sustained members to seek to include non-sustained participants. Another topographical feature of the network is the appearance of local isolates [Image 4]. Local isolates are nodes, which appear in the network, but are removed from the primary network structure -in this case conversation. This is an important finding and further analysis of the local isolates shows the appearance of 6 sustained users and 96 non-sustained users exhibiting this behavior.

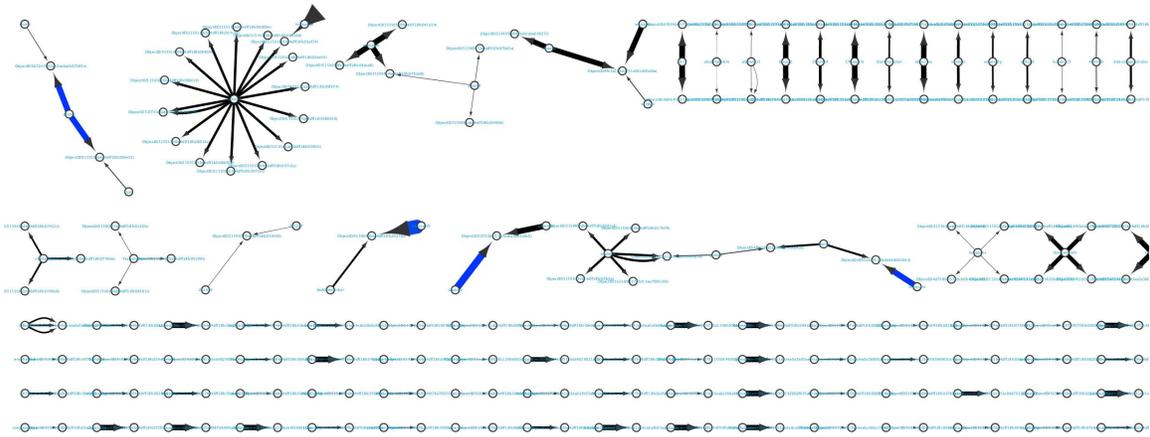


Figure 4: Local Isolates

There are a number of non-sustained users connected to the primary network and the structure of those interactions most often resembles node interaction in image 5. These participants most often have a greater number of posts, which increases the likelihood that sustained users will notice them. Many users in the large network are included only because of one interaction with a sustained user and another smaller group of participants are included because their path runs through another non-sustained user to a sustained user.

Classification Group	Users	Discussion Post	Classifications	Median Classification
Non-Sustained (0-999)	347	1039	52,457	41
Sustained (1000+)	57	211	1,008,028	4087

Table 1: Group Community Production

Analyzing this type of graph we can also begin to analyze the structure of the graph to look at the differences in centrality measures for the two groups (sustained and non-sustained participants). The researcher looks, in particular, at measures of centrality: betweenness, centrality, and closeness. These measures explain how nodes of participants are integrated in the network. Table 2 shows the three measures of centrality important to helping determine the level of inclusion for each group of nodes.

Degree centrality, which measures the number of direct connections to a node for sustained users, is less than that of non-sustained users, which suggest non-sustained participants are being engaged less frequently in conversations and are not included to the degree of sustained participants in the network. Closeness, which is a measure of how close a node is to other nodes in the network, is 0.19 for sustained users and 0.6 for non-sustained users. Again, sustained users are not seen as being close to other nodes in the community and are likely low for sustained users because of the lack of contact with isolates. Lastly, betweenness measures how a node exists as a bridge in the network. Again, non-sustained users show a betweenness value of 0.03, which is influenced by isolates which are disconnected from the primary network. This measure suggests sustained participants are essential in connected nodes in the graph.

	Non-Sustained	Sustained
Degree	2.99	3.70
Closeness	0.602	0.19
Betweenness	0.156	0.03

Table 2: Network Group Scores

4 Conclusion

The research focused on topological features of an affiliation network of discussion events and participants in an online community. A month of log data shows while non-sustained users often begin discussions, their comments often do not receive responses from sustained members suggesting they aren't engaged in conversations in the same space. New systems and designs should incorporate measures to identify non-sustained participants in communities where the exchange of information and knowledge is essential to performing work. Acknowledgement of non-sustained participant work is essential for building the community and building participant confidence in their contributions, which could move more non-sustained participants to a central and sustained role in the community. Future work will focus on analysis of interactions to determine if users whom have interacted with sustained users persist and eventually become sustained. One emergent area of research this work seeks to contribute to in the future is dynamic analysis of online communities using network analysis. Determining how communities form over time and how user roles change could lead to clues about important milestones and events in community formation.

5 References

- Arguello, Jaime and Butler, Brian S. and Joyce, Elisabeth and Kraut, Robert and Ling, Kimberly S. and Rose, Carolyn and Wang, Xiaoqing. 2006. Talk to me: foundations for successful individual-group interactions in online communities. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06), Rebecca Grinter, Thomas Rodden, Paul Aoki, Ed Cutrell, Robin Jeffries, and Gary Olson (Eds.). ACM, New York, NY, USA, 959-968. DOI=10.1145/1124772.1124916 <http://doi.acm.org/10.1145/1124772.1124916>
- Burke, M., Kraut, R., & Joyce, E. (2009). Membership Claims and Requests: Conversation-Level Newcomer Socialization Strategies in Online Groups. *Small Group Research*, 41(1), 4-40. doi:10.1177/1046496409351936
- David, Easley and Jon, Kleinberg (2010). Networks, Crowds, and Markets: Reasoning about a Highly Connected World. Cambridge University Press, New York, NY, USA.
- Davis, A.; Gardner, B. B.; Gardner, M. R. (1941). Deep South; a social anthropological study of caste and class. Chicago: University of Chicago Press.
- Griffin, A. E. , Colella, A., & Goparaju, S. (2000). Newcomer and Organizational Socialization Tactics: An Interactionist Perspective. *Human Resource Management Review*, 10(4), 453-474. doi:10.1016/S1053-4822(00)00036-X
- Kraut, R. E. & Resnick, P. (Under contract). Evidence-based social design: Mining the social sciences to build online communities. Cambridge, MA: MIT Press
- Marsden, P.V. (2005) Recent developments in network measurement. Chapter 2 (pp. 8-30) in P.J. Carrington, J. Scott, and S. Wasserman (eds.), Models and methods in social network analysis, (2005). Cambridge University Press.
- Mavrodiev, Pavlin, Tessone, Claudio J., and Schweitzer, Frank (2013). Quantifying the effects of social influence. *Sci. Rep.*, 3, 1360.

6 Table of Figures

Figure 1: Annotation Interface.....	1140
Figure 2: Discussion Interface	1140
Figure 3: Affiliation Network of User Comments.....	1141
Figure 4: Local Isolates	1142

7 Table of Tables

Table 1: Group Community Production	1142
Table 2: Network Group Scores	1142