Modeling Population Dynamics of Online Communities

Jacopo Baggio, Miles Manning, Lingfei Wu and Marco A. Janssen

Center for the Study of Institutional Diversity, School of Human Evolution and Social Change, PO Box 872402, Tempe 85287-2402 USA

Email: {jacopo.baggio@asu.edu; miles.manning@asu.edu; lingfei.wu@asu.edu; marco.janssen@asu.edu}

***Abstract*—Online communities are more dynamic in terms of composition compared to traditional physical communities. This can affect the conditions for cooperation. We discuss empirical results from online communities and extend the evolution of firms model of Axtell to discover the conditions of cooperation in groups.**

I. Introduction

IT was just a few decades ago that most people spend most of their time in a local community within in physical proximity of other community members. Nowadays we spend an increasing amount of time online. These online activities could relate to social networking, but also creating content for encyclopedia (e.g. Wikipedia), software (e.g. github), book review (e.g. Amazon), answering questions (e.g. stackoverflow). People can participate in diverse communities facilitated by online communication.

For communities to be productive members need to spend effort. For example, farmers in small-scale irrigation communities invest time and effort in maintenance of common infrastructure, meeting to coordinate the scared resource and sharing food in periods in need. Ostrom [1] has demonstrated that communities can be very effective in self-organizing their activities to be productive and avoiding overharvesting of the commons. A key factor thought to explain this are boundary rules which define entry and exit to the communities.

Online communities are different since boundary rules are less strict. People can participate in different communities and leave or reduce their effort if the direct benefits are not sufficient for the persons involved. This lead communities to compete for effort among potential members. The question we explore is to understand what the conditions are for productivity and long-levity of online communities.

The importance of this question can be seen by efforts to solve collective action problems by online communities such as creating encyclopedia (Wikipedia), and open source software (Sourceforge), and various sustainability problems [2]. But is groups are more fluid, cooperation within a group might be less stable. The study on the evolution of cooperation shows that repeated interaction (direct reciprocity) and indirected reciprocity (if having reliable reputation information) can explain high levels of cooperation in human societies. But when group composition changes frequently, what are the conditions in which cooperative arrangements can persist.

A relevant model in this context is the study of Axtell on firms [3]. Axtell formulate firms by having a population of agents where agents have preferences for income and leisure, and groups invest effort to the group, the firm, to produce output. There are increasing returns to cooperation, and agents can benefit from the group efforts since the output is divided equally among the participants. Agents define their effort level to maximize their utility. As a consequence agents may free ride on the investments of others. Agents are permitted to join other firms or start up new firms when it is welfare-improving to do so. As a firm becomes large, agents have little incentive to supply effort, since each agent's share is relatively insensitive to its effort level. This gives rise to free riders. As free riding becomes commonplace in a group, agents will leave. Hence firms change in size all the time, but the distribution of firm sizes is a stable Zip distribution similar to the empirical data [4].

We have replicated the Axtell model in Netlogo and will adjust it to explain observations in online communities. Let's first show some of the findings from online communities, and then we will discuss how we will change the Axtell model.

II. Patterns In Online communities

The data investigated in the current study are the log files of Baidu Tieba system, a collection of many topic-specific forums. Please refer to http://en.wikipedia.org/wiki/Baidu_Tieba for a more comprehensive introduction of this system. Among the millions of forums in the system, we select the top 1,000 forums, whose size (the average daily page views in two months) varies from hundreds to millions. For each forum, we recorded the hourly number of unique visitors and total clicks in 58 days from Feb. 28, 2013 to Apr. 27, 2013.

This work was supported by a grant of the National Science Foundation (ISS 1210856)

We find that the two rules suggested in the model of Malcai et al. [5] , i.e., the converged power-law distribution of organization size and the symmetric levy-stable distribution of the growth rate of average size in time, also shape the growth of online communities. Different from previous studies on organization evolution which usually focus on the dynamics of only one variable [4, 6, 7], we analyze both of the size (visitors) and productivity (clicks) of online communities. The richness of data allows us to explore the relationship between two variables in time, and we observe the super-linear scaling relationship of productivity vs. size, quantifying the economics of scales in collective action online. We also find that the scaling exponent is not universal, but decreases gradually to unity with the growth of community size, indicating the limits of the economics of scales. This pattern calls for a more comprehensive model than Malcai's on the growth dynamics of online communities.

We analyze the hourly-based distributions of the top 1,000 forums in terms of visitors and clicks, and find that the shape of the distributions do not change in time, as given in Figure 1. Both of the number of visitors and clicks satisfy power-law distribution, where the scaling exponents of the power-law distributions of visitors and clicks are 1:33 and 1:11, respectively.

We also looked at the distribution of effort among different communities, and find that 30% of the members participate in more than one forum during a day (Figure 2).

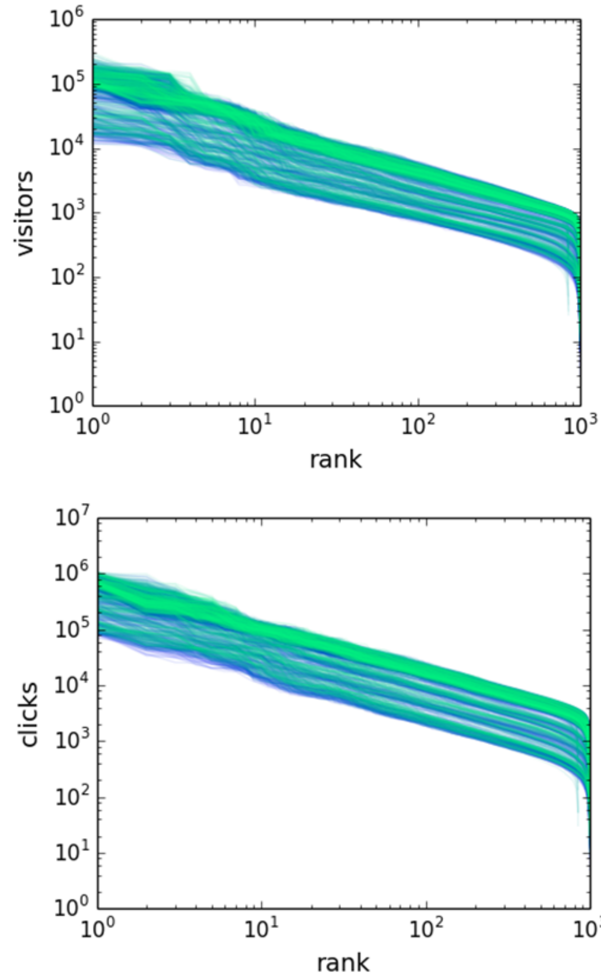

Figure 1. The hourly-based distributions of community size and productivity. (A) and (B) give the hourly distribution of visitors and clicks, respectively. The distributions were plotted in rank-order curves, in which y axis shows values and x axis shows the decreasing rank of the values. The color of data points change from blue to green gradually as time passes. We fitted the scaling exponents in every hour using ordinary least square regression in log-log axes and calculated the average value. We plotted the fitted line in red dotted line and denoted the value of the exponent in the lower left of (A) and (B).

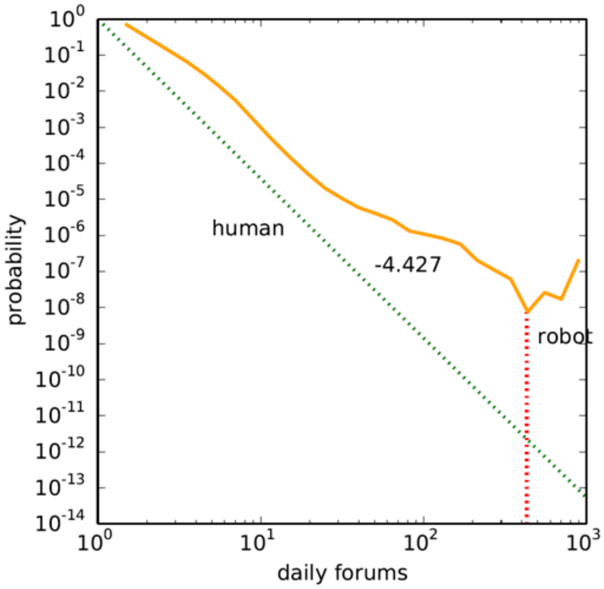

Figure 2. Distribution of daily clicks (left) and visited forums at individual level (right). The data contains the record of 16+ million active users. The results show that a user may visit many different forums in a day. The 430 forums as a separation between human and robot corresponds to the threshold of 1000 thread views/clicks per day, which is an arbitrary threshold to detect a robot.

III. EXTENSIONS

By September we expect to have done a more thorough empirical analysis of dynamics of online communities. For example, we like to explore the logs of stack overflow and identify population and effort levels of different topics and communities. In this extended abstract we use a simplistic number of clicks as effort. With stack overflow data we can measure effort based on feedback participants get on the quality of their work and the reputation they receive within the community.

We will extend the Axtell model by allowing agents to spend effort in more than one group, allow agents to have evolving social networks and allow agents to have preferences that affect which groups they move to. In the current Axtell model agents can move to groups that are known within their social network. There are no entree rules and we like to explore the effect of including entree rules (such having a minimum reputation level) on the population dynamics.

In the analysis we will explore the effect of increased mobility of the population on the ability to create cooperation. Janssen and Goldstone [8] showed that with traditional public good games this only happens with small levels of mobility. Since we observe cooperation in online communities with high levels of mobility, we will explore mobility rules, exit and entry rules, and reputation scores, to detect conditions for cooperation.

REFERENCES

[1] E. Ostrom, “Governing the Commons”, Cambridge University Press, 1990.

[2] J.L. Dickinson, R.L. Crain, H.K. Reeve, and J.P Schuldt,”Can evolutionary design of social networks make it easier to be ‘green’? Trends in Ecology & Evolution 28:9: 561-569, 2013

[3] R.L. Axtell, “The Emergence of Firms in a Population of Agents.” Working paper. The Brookings Institution. 1999. Available at www.brookings.edu/es/dynamics/papers

[4] R.L. Axtell, “Zipf distribution of us firm sizes.” Science 293: 1818-1820, 2000.

[5] O. Malcai, O. Biham, S. Solomon, “Power-law distributions and levy-stable intermittent fluctuations in stochastic systems of many autocatalytic elements.” Physical Review E 60: 1299, 1999.

[6] O. Biham, O. Malcai, M. Levy, and S. Solomon, “Generic emergence of power law distributions and lévy-stable intermittent fluctuations in discrete logistic systems.” Physical Review E 58: 1352, 1998.

[7] R.N. Mantegna and H.E. Stanley, “Scaling behaviour in the dynamics of an economic index.” Nature 376: 46-49, 1995.

[8] M.A. Janssen and R.L. Goldstone (2006), Dynamic-persistence of cooperation in public good games when group size is dynamic, Journal of Theoretical Biology 243(1): 134-142, 2006.