How Novelty is Created in a Web Service

Takashi Ikegami[1†], Mizuki Oka[2] and Yasuhiro Hashimoto[1]

[1]Department of General Systems Studies, The University of Tokyo, Tokyo, Japan
(E-mail: ikeg/hashi@sacral.c.u-tokyo.ac.jp)
[2]Department of Computer Science, Tsukuba University, Tokyo, Japan
(E-mail: mizuki@cs.tsukuba.ac.jp)

Abstract: A mechanism of generating novelty is studied in a web service. Analogous to living ecosystems in nature, web services form an artificial ecosystem consisting of many agents. Both systems develop new species (tags in case of a web service), new taxa (communities in case of a web service) and new niches (topics in case of a web service). In the previous works [1], using the network Hawkes process, we have found that a web service evolves to a point close to the edge of critical point. Here we show that a community of homogeneous users using the same tag patterns, is good at creating new tags. The formation of such community structure may be coincident with the critical state of the system. Together with the new analysis using Price equation, we analyze the evolution of the web services comparing with the biological ones.

Keywords: social tagging, novelty production, community structure, vocabulary similarity, edge of chaos, Price equation

1. INTRODUCTION

Open Ended Evolution (OEE) can be observed mainly in human technologies such as computers, airplanes, smart phones and all the other technologies that support our everyday life. OEE is defined as progressive improvement of quality and inventions of novelty in those technologies. Development of personal computers over decades is a good example. At the same time, such OEE is not solely caused by the singular development of technologies and ideas behind each product, but rather it is caused by a network effect of various technologies and ideas. Developing transistors will create better smart phones, which changes ways of human communication. Changes in human communication will determine the direction of smart phone development, and in the end requesting new design of transistors. This type of feedback loop has been in place between development of human technologies and human society for a long time.

The same sort of feedback is also found in between a web service and the user community. In case of social tagging systems, we see an evolution of new tags and their new combinations. The expansion of our cognitive space is facilitated through the creation of new words. New words provide us with new concepts and possibilities in everyday life, and the creation of new action patterns triggered by new ideas could lead to the creation of further new ideas [2]. In other words, social tagging dynamics is supposed to be a co-evolutionary system between human behaviors and vocabulary, wherein a new tag opens up new behavioral space and the behavior oriented by the cultural preference recurrently changes vocabulary space.

A mathematical treatment of word creation in social tagging analyses is often over-simplified such as Poisson process [3]. Yet some studies have developed sophisticated ideas, which assume a correlation between novelties [4] or a latent semantic structure behind word occurrences [5, 6]. In this study, we discuss a potential mechanisms of OEE by analyzing a new tag creation in a web service in relation to community formation and community size.

† Takashi Ikegami is the presenter of this paper.

2. ANALYSIS

2.1. Community formation

We demonstrate the empirical analysis on the photo-sharing social networking service "RoomClip" provided with tagging data by its operating company Tunnel, Inc. The data consists of a list of annotations, where each annotation has, the time stamp of the annotation created, ID of the photo, ID of the user who posted the photo, and a string of the tag. The data covers almost four years since the inception of the service, and the total numbers of distinct words and annotations are approximately 3.3×10^5 and 8.8×10^6, respectively. There are over 7×10^4 users who posted at least one photo.

First of all, we pay attention to the users community and characterize it with respect to tag usages [7]. The setup and procedure of the analysis is as follows:

1. Extract the users whose number of posts during the data period is greater than or equal to 100. This threshold is determined ad hoc in order to ensure their vocabulary size is sufficiently large.
2. Calculate a probability distribution of used words for all extracted users.
3. Calculate similarity between the probability distributions for every pair of the extracted users, and define a user similarity network in word usage.
4. Traversing from a loosely to densely connected network by changing a similarity threshold, observe the connectivity of the highly prductive users.

Here we define the novelty production rate of the user as the total number of words that were created by him/herself and used by more than 100 other users. The vocabulary similarity between a pair of users is evaluated

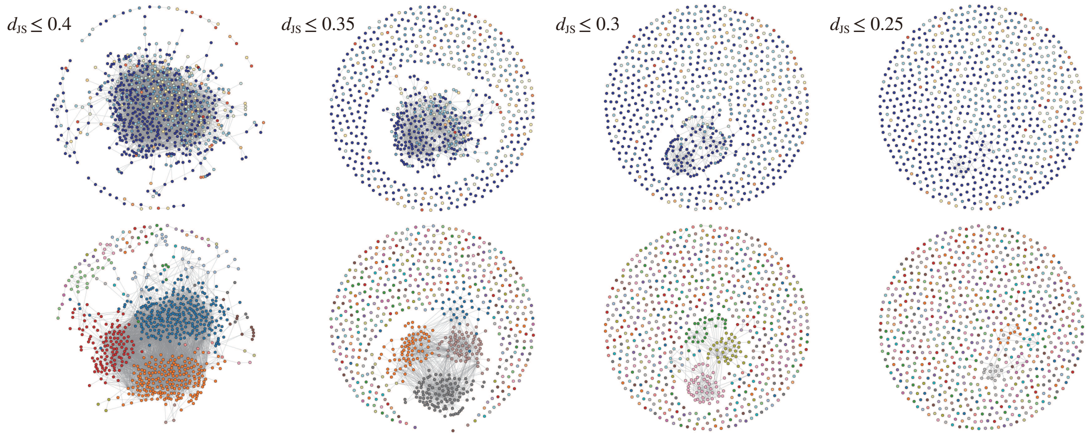

Fig. 1 The user similarity network. Each node is a user, and they are connected if d_{JS} is smaller than the threshold value 0.4, 0.35, 0.3, 0.25 from the left to the right, respectively. The top four figures show the number of word creations by individuals in color; shifting from blue, yellow to red means they created more words. The bottom four figures show the community structures detected in each network.

using the Jensen–Shannon divergence

$$d_{\mathrm{JS}}(p_i||p_j) = [d_{\mathrm{KL}}(p_i||q) + d_{\mathrm{KL}}(p_j||q)]/2,$$
$$q = (p_i + p_j)/2, \tag{1}$$

where p_i and p_j are the probability distribution of user i and j, respectively, and $d_{\mathrm{KL}}(p||q)$ is the Kullback–Leibler divergence from probability distribution q to p. The value of d_{JS} falls within a range between 0 and $\log_k 2$; here $k = e$ and therefore $0 \leq d_{\mathrm{JS}} \leq \ln 2$. Smaller d_{JS} means that the pair has more similar "vocabulary profiles".

The result is shown in Fig. 1. Laterally aligned top four figures show the networks with different threshold values of d_{JS} that gets smaller from left to right. The far left case ($d_{\mathrm{JS}} \leq 0.4$) exhibits a typical core-periphery structure [8], which has a densely connected part (core) surrounded by loosely connected parts (periphery). Using a smaller threshold value means that a more strongly connected part is focused, where we obtain a sub-network approximately associated with the core part of the core-periphery structure. At the same time, however, *community structure*, which means a loosely-connected set of densely-connected subgraphs, arises with such small threshold values. This is more clear if we apply a community detection method to the networks. The bottom four figures exhibit the detected community structures by the modularity optimization method [9], and the obtained values of modularity—0.35, 0.42, 0.49, 0.56 from left to right—tell that the community structure becomes relatively salient in a strongly connected part of the network.

On the other hand, the color of nodes in top four figures shows the novelty production rate of each user; red is high and blue is low. In the core part, a novelty production rate of users have a relatively lower value (see the case of $d_{\mathrm{JS}} \leq 0.3$ and 0.25). The user similarity network in word usage has a latent community structure in its strongly connected part. And on top of that, users who create new words which are used also by a certain amount of other users, are located out of such community structures or in peripheral parts of the network. Nevertheless, it should be noted that the novelty creating rate becomes larger in the higher order cliques in the core part [10].

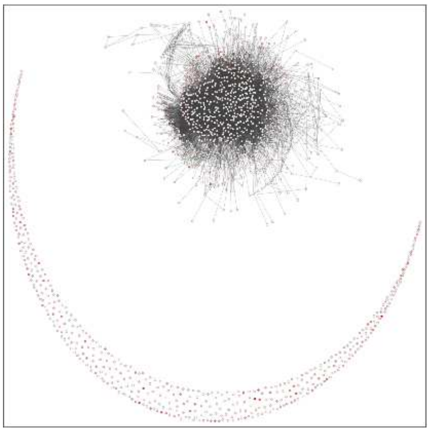

Fig. 2 A core-periphery structure in the user similarity network. Each node is a user, and those with higher tag creation rate is plotted with a red color [10]

Namely, the users in the peripheral parts and the higher-order cliques have the higher new tag production rate. This tendency becomes more clear in a different visualization of the same analysis depicted in Fig. 2. This suggests that users tend to create new tags when they have different profiles from others (i.e. users in a peripheral part) or if users strongly share the same interests (i.e. users in a higher order cliques of a core part). Those two rather opposite cases are observed in this analysis.

2.2. Community size

The above analysis lacks the analysis of dynamics how the tag creation rate changes over time. A single photo submission is associated with many tags. Actually the average number of tags is almost monotonically increasing from the beginning of the web service. A maximum number of tags will become over 100! In order to measure the growth of "complexity" in a quantitative manner, we measured the distinct number of tags as the number of

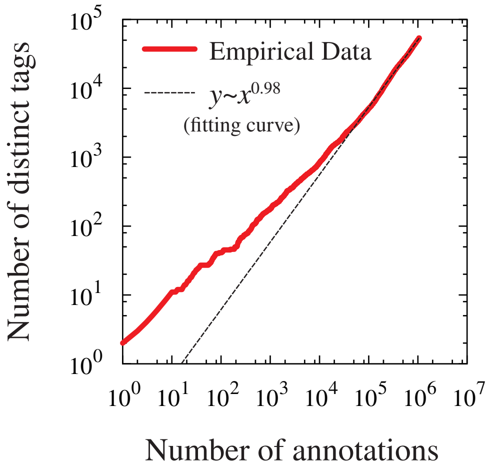

Fig. 3 The number of distinct tags observed in the service as a function of the number of annotations. The creation of new tags converges to the curve $x^{-0.98}$, so that the tag creation rate per annotation is almost constant ($\alpha = 0.05$ in this case).

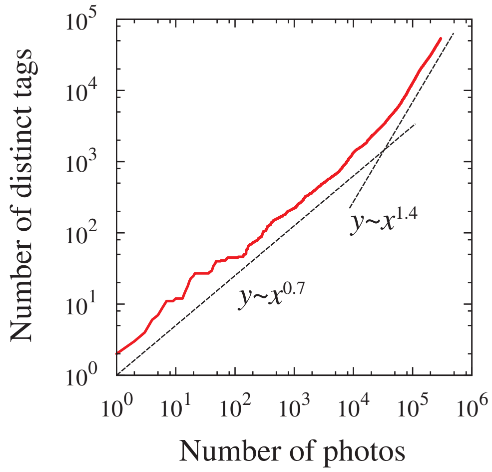

Fig. 4 The number of distinct tags observed in the service as a function of the number of photos. The creation of new tags changes from negative to positive at around 10^4 to 10^5.

annotations as well as the number of photos.

From our previous studies, we know that the volume of the entire dictionary (of tags) increases typically as t^β, by taking the number of annotation as t, which is known as Heaps' law. This exponent β takes the value between around 0.7 and 1 empirically, and in Fig. 3, we see it converges to almost one at around the 10^5 annotations. Beyond this number, the distinct number of tags is proportional to the total number of annotations.

The transition point in Fig. 3 is also computed as a function of the number of photo submissions. From Fig. 4, beyond the number of photo submissions around 10^4 to 10^5, we notice that the exponent of the fitted curve changes from 0.7 to 1.4 for the number of distinct tags. What causes the exponent change is not clear. It may have been caused by a kind of system changes in the web service or it may be simply caused when the system size (i.e. the total number of the users) becomes larger than a certain amount.

We assume that it may correspond to the critical point which is obtained from the network Hawkes analysis [**?**]. In the paper, we also discussed a simple mechanism of creating novelty when a system size goes beyond a certain size and showed that a simple boid system shows qualitatively different dynamics beyond a critical flock size (about 10k). It is said that there exists a critical community size (e.g. the Dunbar's number) due to the bounded cognitive capability of agents and the resulting stability of the community [11]. The current example provides yet another example of such critical size of community.

3. DISCUSSION

At the OEE workshop (published in [12]), we have discussed behavioral hallmarks and the hypothesized mechanism of OEE from various topics. In this paper, we have added several new observations to our previous findings. We propose a few hypothesis with the OEE phenomena in the web service.

- The novelty creation rate and a core strength of the network may be correlated. Our analysis shows that users of the higher order cliques produces more novel tags, as well as users of the peripheral part.
- The novelty creation rate doubles its value from 0.7 to 1.4, after a total amount of submissions excesses a certain number.
- A community develops several sub-communities of users with the similar profiles (i.e. possessing similar tag usages). Within these sub-communities, homogeneous community has a potential to create novelty.

These hypothesis should be refined by the further analysis, yet we expect that these correspond to Ackley's definition of indefinite scalability of OEE [13]. That is, " supporting open-ended computational growth without requiring substantial re-engineering." A growth of the system size leads to a potential "door-opening" innovation in each critical size. Also as it has been discussed in [12], the present work provides an another example of adaptive novelty, since the evolution of a web service can utilize finite combinations of old and new tags which will generate a qualitatively new niches for the users.

REFERENCES

[1] M. Oka, Y. Hashimoto, and T. Ikegami, "Open-ended evolution in a web system," in *Late breaking of the13th European Conference on Artificial Life*, ECAL2015, 2015.

[2] S. A. Kauffman, *Investigations*. Oxford University Press, 2000.

[3] C. Cattuto, V. Loreto, and V. D. P. Servedio, "A yule-simon process with memory," *Europhys. Lett.*, vol. 76, no. 2, pp. 208–214, 2006.

[4] F. Tria, V. Loreto, V. D. P. Servedio, and S. H.

Strogatz, “The dynamics of correlated novelties,” *Sci. Rep.*, vol. 4, p. 5890, 2014.

[5] C. Cattuto, A. Barrat, A. Baldassarri, G. Schehr, and V. Loreto, “Collective dynamics of social annotation,” *Proc. Natl. Acad. Sci. USA*, vol. 106, no. 26, pp. 10511–10515, 2009.

[6] A. Buchanan, N. H. Packard, and M. A. Bedau, “Measuring the evolution of the drivers of technological innovation in the patent record,” *Artificial Life*, vol. 17, no. 2, pp. 109–122, 2011.

[7] Y. Hashimoto and T. Ikegami, “Novelty production in tagging crowds,” in *Proceedings of the 2nd International Symposium on Swarm Behavior and Bio-Inspired Robotics*, SWARM’17, pp. 311–312, 2017.

[8] S. P. Borgatti and M. G. Everett, “Models of core/periphery structures,” *Social Networks*, vol. 21, no. 4, pp. 375–395, 1999.

[9] M. E. J. Newman, “Modularity and community structure in networks,” *Proc. Natl. Acad. Sci. USA*, vol. 103, no. 23, pp. 8577–8582, 2006.

[10] T. Ikegami, Y. Mototake, S. Kobori, M. Oka, and Y. Hashimoto, “Life as an emergent phenomenon: studies from a large-scale boid simulation and web data,” *Philosophical Transactions of the Royal Society A*, vol. 375, no. 2109, p. 20160351, 2017.

[11] R. I. M. Dunbar, “Neocortex size as a constraint on group size in primates,” *Journal of Human Evolution*, vol. 22, no. 6, pp. 469–493, 1992.

[12] T. Taylor, M. Bedau, A. Channon, D. Ackley, W. Banzhaf, G. Beslon, E. Dolson, T. Froese, S. Hickinbotham, T. Ikegami, B. McMullin, N. Packard, S. Rasmussen, N. Virgo, E. Agmon, E. Clark, S. McGregor, C. Ofria, G. Ropella, L. Spector, K. O. Stanley, A. Stanton, C. Timperley, A. Vostinar, and M. Wiser, “Open-ended evolution: Perspectives from the oee workshop in york,” *Artificial Life*, vol. 22, no. 3, pp. 408–423, 2016.

[13] D. H. Ackley and T. R. Small, “Indefinitely scalable computing = artificial life engineering,” in *Proceedings of the 14th International Conference on the Synthesis and Simulation of Living Systems*, ALIFE2014, pp. 606–613, 2014.