Understanding community evolution in Complex systems science

Qinna Wang, Eric Fleury

Understanding community evolution in Complex systems science

Qinna Wang[a] and Eric Fleury[b]
[a] qinna.wang@ens-lyon.fr
[b] eric.fleury@inria.fr

LIP ENS-LYON
D-NET INRIA
Université de lyon
46 Allée d'Italie Lyon 69364 France

Complex systems is a new approach in science that studying organized behaviours in computer science, biology, physics, chemistry, and many other fields. By collecting articles containing topic keywords relevant for the field of complex networks from ISI Web of knowledge during 1985-2009, we construct a science network, which connects ~ 215000 articles according to the proportion of shared references. Moreover, articles' publication time makes it dynamically evolve in time. We here use a two-step approach [3] to explore community evolution and study underlying information behind community changes. We firstly detect communities by applying Louvain algorithm [2] on each snapshot graph, and secondly construct relationships between partitions at successive snapshot graphs [1].

Communities may change, like fusion, split, disappearance and emergence. To construct relationships between communities, we use *community predecessor and successor*: given a community $C_i(t)$ found at time t, its predecessor is community $C_j(t-1)$, which has the maximum overlap size among all communities at time $t-1$, such as $C_j(t-1) = \arg\max_{C_k(t-1) \subseteq \mathcal{P}(t-1)} |C_k(t-1) \cap C_i(t)|$; its successor is community $\mathcal{C}_j(t+1)$, which has the maximum overlap size among all communities at time $t+1$, such as $C_j(t+1) = \arg\max_{C_k(t+1) \subseteq \mathcal{P}(t+1)} |C_k(t+1) \cap C_i(t)|$. Given a pair of clusters (X, Y), we use $X \to Y$ to denote that Y is X's successor while $X \leftarrow Y$ to denote that X is Y's predecessor. Besides, we define community $C_i(t)$'s *survival* is community $C_j(t+1)$ such as $C(i)(t) \leftrightarrows C_j(t+1)$, if and only if $C_i(t) \to C_j(t+1)$ and $C_i(t) \leftarrow C_j(t+1)$.

We use the survival to describe one community evolving stable. Furthermore, we also use community predecessor and successor to identify community dynamic events: given a community $\mathcal{C}(t)$, if it has more than one predecessors, then $\mathcal{C}(t)$ is a merged community; if it has more than one successors, then $\mathcal{C}(t)$ split in the next time step; if it has no predecessor, then $\mathcal{C}(t)$ is a new community; If it has no successor, then $\mathcal{C}(t)$ will vanish; otherwise, it evolves stable. A diagram (see Fig 1) shows several cases involving community dynamic events. We observe nearly all types of community changes: community $\mathcal{C}_2$ emerges at $t = 2$, community $\mathcal{C}_3$ disappears at $t = 3$, community $\mathcal{C}_2$ merges into $\mathcal{C}_1$ at $t = 4$ and community $\mathcal{C}_4$ is split from $\mathcal{C}_3$ at $t = 3$. For community $\mathcal{C}_1$, it evolves stable across the total four time steps although it is related to a merge event. It is like the change of $\mathcal{C}_2$ rather than $\mathcal{C}_1$.

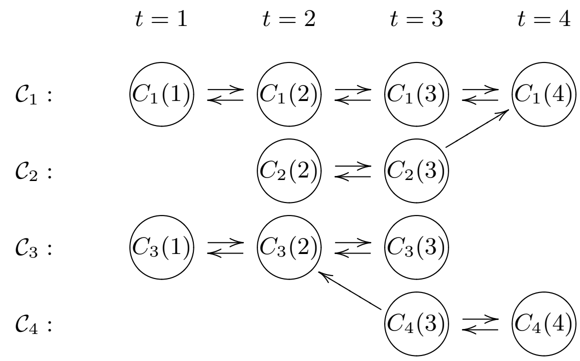

Fig. 1. Diagram of four communities over four time points, featuring continuation, emerge, disappearance, merge and split community events.

After applying our method on Complex Systems Science's snapshots: $G(1985-1995), G(1990-2000), G(1995-2005), G(2000-2010)$, we observe most of communities evolve stable over time, especially communities representing theoretical sciences like chaos theory(CHAOS), systems ecology(ECOLOGY), systems neuroscience(NEURAL NETs)[1,2]. These stable communities are also involving many change events, especially merges between function applications and science domains such as ISING-MODEL and PHOTOSYSTEM-II merged into chemistry during $1990-2000$. Results on community evolution suggest that theoretical sciences are the foundations of Complex System Science.

References

1. S. Asur, S. Parthasarathy, and D. Ucar. An event-based framework for characterizing the evolutionary behavior of interaction graphs. In *Proceedings of the 13th ACM SIGKDD on Knowledge discovery and data mining*, page 921. ACM, 2007.
2. V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *JSTAT*, 2008.
3. M. Spiliopoulou, I. Ntoutsi, Y. Theodoridis, and R. Schult. Monic: modeling and monitoring cluster transitions. In *Proceedings of the 12th ACM SIGKDD on Knowledge discovery and data mining*, pages 706–711. ACM New York, NY, USA, 2006.

[1] CHAOS, ECOLOGY and NEURAL NETs are the most popular key words shared by articles in the found communities while chaos theory, systems ecology, systems neuroscience are disciplinarians by considering popular keywords shared by articles in the found communities. For instance, the community nominated by CHAOS representing chaos theory have popular keywords like DYNAMIC, CHAOS, SYSTEM, NONLI-, COMPL-, MODEL, STABI- and so on.

[2] Through our method, we found many matching communities sharing the same popular keywords and the same disciplinarians, like CHAOS and ECOLOGY, and many matching communities sharing different popular keywords but representing the same disciplinarians, like biology:nervous system(RAT, BRAIN) whose articles focus on biology functions about humans' brain in place of rats' as time going.