Dynamic Network Motifs: Evolutionary Patterns of Substructures in Complex Networks

Yutaka Kabutoya, Kyosuke Nishida, and Ko Fujimura

NTT Cyber Solutions Laboratories, NTT Corporation,
1-1 Hikari-no-oka, Yokosuka-shi, Kanagawa, 239-0847 Japan

Abstract. We propose an entirely new approach to understanding complex networks; called "dynamic network motifs (DNMs)." We define DNMs as statistically significant local evolutionary patterns of a network. We find such DNMs in the networks of two web services, Yahoo Answers and Flickr, and discuss the social dynamics of these services as indicated by their DNMs.

Keywords: link analysis, subgraph mining.

1 Introduction

Many studies in the fields of engineering and science have tackled complex network analysis. In recent years, improvements in computational power have advanced the field of network mining, i.e. identifying all subgraphs in a given network [6,7].

Network motif (NM) analysis is one of the subgraph mining methods proposed by Milo et al. [4,5]. Motifs are small (usually from three to seven nodes in size) connected subgraphs that appear in the given network at higher frequency than in the equivalent (in terms of the number of nodes and edges) random networks. NM analysis is attracting a lot of attention because it enables a fuller understanding of complex networks in terms of their local structure.

A recent extension to data mining attempted to cover the evolution of subgraphs [2]. In [2], since the statistical significance of the evolution of subgraphs is not still defined, no discussion was given on what evolution of subgraphs is important for a given network.

To achieve this target, we propose a method, called *dynamic network motif (DNM)* analysis, that can analyze the evolution of subgraphs statistically. We define DNMs as statistically significant local evolutionary patterns of a network.

In this paper, we apply our method to networks based on datasets of two actual web services, Yahoo Answers[1] and Flickr[2]. The DNMs found in the network from Yahoo Answers are distinct from the DNMs found in the network from Flickr: This is due to the difference in communication properties, which influences the evolution of the network. We consider that the networks of Yahoo Answers and Flickr are evolving under the pressures of expertise and politeness, respectively.

[1] `http://answers.yahoo.com`

[2] `http://www.flickr.com`

X. Du et al. (Eds.): APWeb 2011, LNCS 6612, pp. 321–326, 2011.

2 Conventional Method: Network Motifs

Before introducing our method, we review network motifs (NMs), which are the basis of our method. NMs are defined to be statistically significant patterns of local structures in real networks. They are being studied by many researchers, including Milo et al. [5].

Milo et al. started with network $\boldsymbol{G} = \langle \boldsymbol{V}, \boldsymbol{E} \rangle$ where the interaction between nodes is represented by directed edges. The types of n-node subgraphs, which are defined to be subgraphs that have n nodes, are connected, and are not isomorphic to one another, were introduced. For example, there are 13 types of 3-node subgraphs as shown in Figure 1, and 199 types of 4-node subgraphs. The network is scanned for all possible n-node subgraphs (in many studies, $3 \leq n \leq 7$), and the number of occurrences of each type is recorded. To focus on those that are likely to be important, the real network is compared to K equivalent randomized networks. NMs are defined to be the type that recur in the real network with significantly higher frequency than in the equivalent networks.

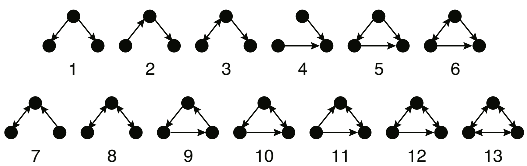

Fig. 1. All 13 types of triads

Furthermore, to compare the local structures of networks from different fields, Milo et al. [4] introduced the concept of the significance profile (SP) of a network. Here, m_i and $\hat{m}_{ik}$ denote the number of occurrences of type i in the real network and in the kth equivalent randomized network, respectively. First, the Z-score of the occurrences of type i is computed as:

$$z_i = \frac{m_i - \mu_i}{\sigma_i}, \tag{1}$$

where $\mu_i = \frac{1}{K}\sum_{k=1}^{K} \hat{m}_{ik}$ and $\sigma_i^2 = \frac{1}{K}\sum_{k=1}^{K}(\hat{m}_{ik} - \mu_i)^2$. The SPs are computed by normalizing the vector of Z-score $\boldsymbol{z} = \{z_i\}_i$ to length 1.

3 Proposed Method: Dynamic Network Motifs

We propose a method to analyze evolution of n node subgraphs (this paper mainly addresses the case of $n = 3$). We define dynamic network motifs (DNMs) as statistically significant evolutionary patterns of local structures in given networks.

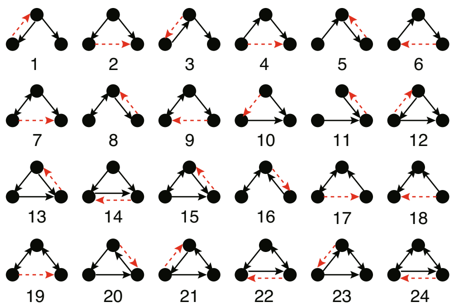

Fig. 2. All 24 evolutionary types of triads

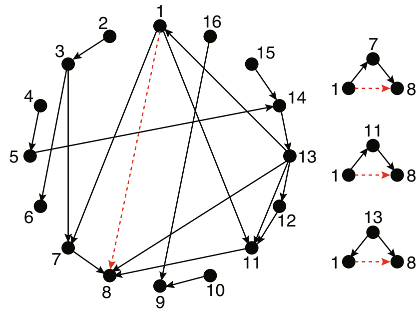

Fig. 3. Edge $\langle 1, 8\rangle$ yields three triad-evolution

Suppose we have network $\boldsymbol{G} = \langle \boldsymbol{V}, \boldsymbol{E}\rangle$ where each edge $e = \langle u, v\rangle \in \boldsymbol{E}$ represents an interaction from u to v that took place at a particular time, $t(e)$. Though potentially there are multiple interactions from u to v, we record only the oldest one. For two times $t_0 < t_1$, let $\boldsymbol{G}_0 = \langle \boldsymbol{V}, \boldsymbol{E}_0\rangle$ and $\boldsymbol{G}_1 = \langle \boldsymbol{V}, \boldsymbol{E}_1\rangle$ denote the subgraph of $\boldsymbol{G}$ consisting of all edges with a time-stamp before t_0 and t_1, respectively.

In this paper, we regard each single interaction that took place between t_0 and t_1, i.e. $e \in \boldsymbol{E}_1 - \boldsymbol{E}_0$, as evolution of network $\boldsymbol{G}_0$. The evolutionary type of a n node subgraph is determined by the type of the subgraph in $\boldsymbol{G}_0$ and the evolution $e \in \boldsymbol{E}_1 - \boldsymbol{E}_0$. The number of evolutionary types increases with subgraph size. For example, three node subgraphs offer 24 possible evolutionary types and four node subgraphs offer more than 200 evolutionary types. Let J denote the number of evolutionary types. Figure 2 shows all 24 evolutionary types of three node subgraphs, where black and red dashed lines indicate existing edges in $\boldsymbol{E}_0$ and new edges in $\boldsymbol{E}_1 - \boldsymbol{E}_0$, respectively.

We scan $\boldsymbol{G}_0$ and $\boldsymbol{E}_1 - \boldsymbol{E}_0$ for all possible evolution of n node subgraphs, and record the number of occurrence of each evolutionary type. In general, a single new edge is related to multiple subgraphs. In the case of Figure 3, for example, pair $\langle 1, 8\rangle$ yields three triad-evolution. Focusing on the three triads of $\langle 1, 7, 8\rangle$, $\langle 1, 11, 8\rangle$, and $\langle 1, 13, 8\rangle$, the pairs are related to evolutionary types 4, 4, and 2, respectively.

The number of occurrences of evolutionary type j in $\boldsymbol{E}_1 - \boldsymbol{E}_0$ of $\boldsymbol{G}_0$ is given by:

$$d_j = \sum_{\langle u,v\rangle \in \boldsymbol{E}_1 - \boldsymbol{E}_0} x_{uvj}, \tag{2}$$

where x_{uvj} represents the number of occurrences of evolutionary type j related to new edge $\langle u, v\rangle$ (e.g., $x_{uvj} = 2$ where $u = 1$, $v = 8$, and $j = 4$ in the case of Figure 3). Similar in NM analysis, to compute the statistical significance of evolutionary type j, we compare $\boldsymbol{E}_1 - \boldsymbol{E}_0$ to the equivalent randomized sets of edges $\hat{\boldsymbol{E}}_k$ in terms of the number of occurrences of each evolutionary type in

Table 1. Number of nodes, two times t_0 and t_1, and number of edges of social networks

	$\lvert\boldsymbol{V}\rvert$	t_0	$\lvert\boldsymbol{E}_0\rvert$	t_1	$\lvert\boldsymbol{E}_1\rvert$
Yahoo Answers	5,835	2010/07/21	18,796	2010/07/22	37,320
Flickr	7,776	2010/06/15	82,868	2010/06/18	173,344

$\boldsymbol{G}_0$. The randomized sets of edges are generated in a way similar to randomized networks in NM analysis.

The statistical significance of evolutionary type j is given by:

$$\zeta_j = \frac{d_j - \nu_j}{\tau_j}, \tag{3}$$

where $\hat{d}_{jk}$ represents the number of occurrences of evolutionary type j for $\hat{\boldsymbol{E}}_k$, $\nu_j = \frac{1}{K}\sum_{k=1}^{K}\hat{d}_{jk}$, and $\tau_j^2 = \frac{1}{K}\sum_{k=1}^{K}(\hat{d}_{jk} - \nu_j)^2$. The SPs of evolutionary types are computed by normalizing the vector of Z scores $\boldsymbol{\zeta} = \{\zeta_j\}_{j=1}^{J}$ as in NM analysis.

4 Experiments

4.1 Datasets

We exploited two social media sources, Yahoo Answers and Flickr, to gather the experimental data. Our Yahoo Answers data set is based on a snapshot crawled in the period of July 20th, 2010 to July 22nd, 2010, and our Flickr data set is based on a snapshot crawled using Flickr API[3] in the period of June 12th, 2010 to June 18th, 2010.

We extract a dense part from each graph as experimental data using parameter κ (set to three in each case). Dense is defined as occurring when all nodes have at least κ edges in each graph by Liben-Nowell et al. [3]. Table 1 lists details of social networks.

4.2 NM and DNM Analyses

In this paper, we focus on three node subgraphs. The analysis about subgraphs that have more than three nodes is future work.

NM analysis allows us to clarify the properties of interactions among nodes in terms of local structures in networks. Figure 4 displays network motif profiles of the Yahoo Answers network and the Flickr network.

From this figure, we can see that both media have strongly defined cliques (see triad 13 in the figure) compared to random networks. This triad, which represents the case that three people collaborate with each other, is a typical motif of social networks [4]. As well, we can see that only Yahoo Answers has

[3] http://www.flickr.com/services/api

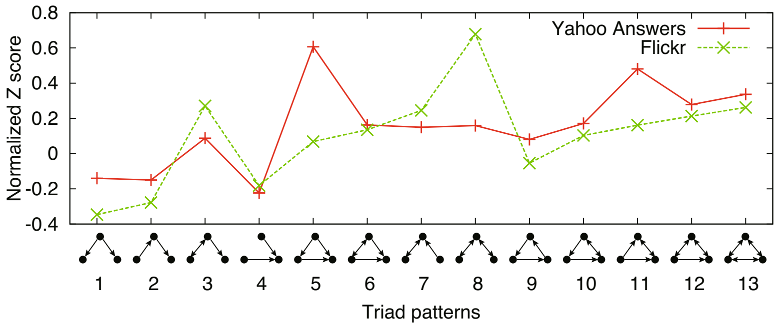

Fig. 4. Network motif profiles for social networks

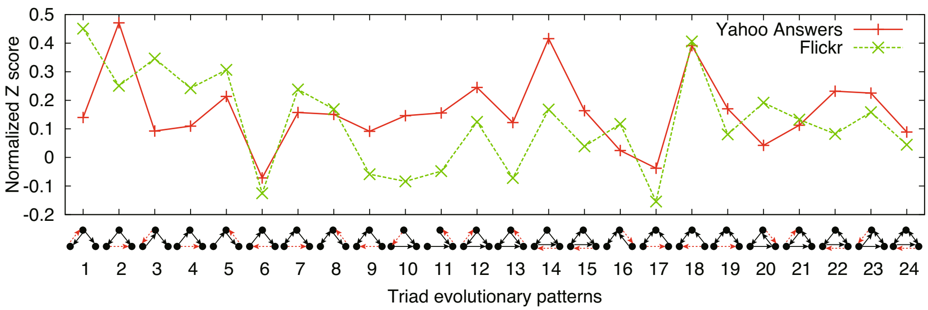

Fig. 5. Dynamic network motif profiles for social networks

a high number of feed forward loops (see triad 5 in the figure). This result matches that shown by Adamic et al. [1]. This motif indicates the expertise of Yahoo Answers users; people with high levels of expertise are willing to help people of all levels, whereas people of lower expertise help only those with even less expertise. On the other hand, only the Flickr network has the motif of two mutual dyads (see triad 8 in the figure). This motif indicates existence of core users who communicate mutually with the two other people in the triad.

DNM analysis enables us to discover the hidden properties of interaction among nodes in terms of the evolutionary patterns of substructures in networks. Figure 5 shows the dynamic network motif profiles of the Yahoo Answers network and the Flickr network.

First, we focus on common DNMs for both networks. From this figure, we can see that triad evolutionary type 18 is a common DNM for both networks, which is an intuitive result. This DNM is created by a user who wants to be invited into a community sending messages to one member after another. Note that neither triad evolutionary type 23 nor type 24 are DNM for both networks; therefore, interestingly, such users cannot necessarily become a member of the community.

Furthermore, we focus on the DNMs unique to each network. We can see that triad evolutionary type 2 is a DNM for the Yahoo Answers network whereas type 4 is not. This indicates that people with high levels of expertise help others faster than people of lower expertise. As well, we can see that type 1 is a DNM for the Flickr network whereas neither type 3 nor type 11 is. This DNM indicates that core users tend to send comments to others before they receive comments. This implies that core users are well-mannered in that they reply to others' comments.

5 Conclusion

We proposed an entirely novel notion to analyze complex networks more deeply, we call it *dynamic network motifs (DNMs)*. We defined DNMs as statistically significant local evolutionary patterns of a network. We applied our method to real networks from two web services, Yahoo Answers and Flickr, found DNMs in the networks, and analyzed the interaction properties of the two web services based on the DNMs.

In future work, we will conduct further experiments. First, we hope to confirm the efficiency of DNMs in the case of more than three node subgraphs. Second, we intend to use datasets gathered from social network services (SNS) such as Facebook[4].

References

1. Adamic, L., Zhang, J., Bakshy, E., Ackerman, M.: Knowledge sharing and Yahoo Answers: everyone knows something. In: Proceedings of the 17th International Conference on World Wide Web, pp. 665–674 (2008)
2. Inokuchi, A., Washio, T.: A fast method to mine frequent subsequences from graph sequence data. In: Proceedings of the 8th IEEE International Conference on Data Mining, pp. 303–312 (2008)
3. Liben-Nowell, D., Kleinberg, J.: The link-prediction problem for social networks. Journal-American Society for Information Science and Technology 58(7), 1019–1031 (2007)
4. Milo, R., Itzkovitz, S., Kashtan, N., Levitt, R., Shen-Orr, S., Ayzenshtat, I., Sheffer, M., Alon, U.: Superfamilies of evolved and designed networks. Science 303(5663), 1538–1542 (2004)
5. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U.: Network motifs: simple building blocks of complex networks. Science 298(5594), 824–827 (2002)
6. Nijssen, S., Kok, J.: A quickstart in frequent structure mining can make a difference. In: Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 647–652 (2004)
7. Yan, X., Han, J.: gSpan: Graph-based substructure pattern mining. In: Proceedings of the 2002 IEEE International Conference on Data Mining, pp. 721–724 (2002)

[4] `http://www.facebook.com`