Chapter 7.21
Motif Analysis and the Periodic Structural Changes in an Organizational Email-Based Social Network

Krzysztof Juszczyszyn
Wrocław University of Technology, Poland

Katarzyna Musiał
Wrocław University of Technology, Poland

ABSTRACT

Network motifs are small subgraphs that reflect local network topology and were shown to be useful for creating profiles that reveal several properties of the network. In this work the motif analysis of the e-mail network of the Wroclaw University of Technology, consisting of over 4000 nodes was conducted. Temporal changes in the network structure during the period of 20 months were analysed and the correlations between global structural parameters of the network and motif distribution were found. These results are to be used in the development of methods dedicated for fast estimating of the properties of complex internet-based social networks

INTRODUCTION

Communication technologies enabled the emergence of complex, evolving social networks built on various services like e–mail, P2P and community portals. In general they are similar to the traditional social networks based on relations between humans, but there are also some significant differences. First, the information about the users in virtual communities and their activities is stored in electronic form which allows precise inference of the network structure and parameters. On the other hand, the networks created by means of communication technologies show incomparable size and dynamics (for more differences between regular social networks and virtual ones see Sec.3).

When investigating the topological properties and structure of complex networks we must face a number of complexity–related problems. In large social networks, tasks like evaluating the centrality measurements, finding cliques, etc. require significant computing overhead. In this context the methods, which proved to be useful for small and medium sized networks fail when applied to the huge structures. Moreover, our knowledge about the actual network structure may be incomplete especially due to its size and dynamics.

In this work we present the results of the analysis of local topology structure of large e–mail based organizational social network. The investigated e–mail network of the Wroclaw University of Technology (WUT), consisting of over 4000 nodes (e–mail addresses of the users) was generated basing on server logs from the period of February 2006 – September 2007, analyzed with standard methods – clustering and centrality assessment. Periodic changes in the network structure, connected with the business profile of the organization (university) were discovered and analyzed. In the series of experiments the global features of the network (clustering coefficients, centrality as well as the number of edges and nodes) were checked for correlations with the distribution of small network subgraphs, called network motifs. The existence of relations between motif profile of the network and its global structural properties may allow their fast estimation.

Dependencies between global network characteristics and the distribution of local topology features may have numerous applications. They can help to estimate the measures like centrality and clustering without the complete knowledge of the network structure. This feature appears very appealing especially when we deal with evolving networks consisting of millions of nodes (like social networks of mobile phone users, Web communities and so on) and the distribution of motifs may be determined by various sampling techniques which do not assume the exhausting processing of the entire structure of the network.

In the following sections we briefly introduce the networks motifs concept as well as the basics of online social networks. Furthermore, the process of social network extraction from WUT e–mail logs is presented. Finally, the results of motif detection and temporal changes in the structure of this network are discussed.

NETWORK MOTIFS

Complex networks, both biological and engineered, were analyzed with respect to so–called network motifs (Milo, 2002). They are small (usually 3 up to 7 nodes in size) subgraphs which occur in the given network far more (or less) often then in the equivalent random (in terms of the number of nodes and edges) networks. Despite all known structural and statistical similarities, networks from different fields have very different local topological structure. It was recently shown that concentration of network motifs may help to distinguish and classify complex biological, technical and social networks (Milo et al.,2004). We can define so–called superfamilies of networks, which correspond to the specific significance profiles (SPs). To create SP for the motifs in a given network, the concentration of individual motifs is measured and compared to their concentration in a number of random networks. The statistical significance of motif M is defined by its Z–score Z_M:

$$Z_M = \frac{n_M - \left\langle n_M^{rand} \right\rangle}{\sigma_M^{rand}} \qquad (1)$$

where

n_M – the frequency of motif M in the given network,

$\left\langle n_M^{rand} \right\rangle$ and σ_M^{rand} – the mean and standard deviation of M's occurrences in the set of random networks,