Computer Communications 73 (2016) 167–175

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

Uncovering the fingerprint of online social networks using a network motif based approach

Alexandru Topirceanu*, Alexandra Duma, Mihai Udrescu

Department of Computers and Information Technology, Politehnica University Timisoara

A R T I C L E I N F O

Article history:
Available online 10 July 2015

Keywords:
Online social networks
Complex network topologies
Network motifs
Classification
Similarity

A B S T R A C T

Complex networks facilitate the understanding of natural and man-made processes and are classified based on the concepts they model: biological, technological, social or semantic. The relevant subgraphs in these networks, called network motifs, are demonstrated to show core aspects of network functionality and can be used to analyze complex networks based on their topological fingerprint. We propose a novel approach of classifying social networks based on their topological aspects using motifs. As such, we define the classifiers for regular, random, small-world and scale-free topologies, and then apply this classification on empirical networks. We then show how our study brings a new perspective on differentiating between online social networks like Facebook, Twitter and Google Plus based on the distribution of network motifs over the fundamental topology classes. Characteristic patterns of motifs are obtained for each of the analyzed online networks and are used to better explain the functional properties behind how people interact online and to define classifiers capable of mapping any online network to a set of topological-communicational properties.

1. Introduction

Complex networks cover an active area of scientific research inspired largely by the empirical study of real-world networks such as communication networks, economical networks and social networks. They are classified into four major types, based on the context which they model: biological networks (e.g., metabolic networks, transcription regulatory networks, protein–protein interaction networks, protein structure networks, neural networks, ecological networks, and natural food chains) [1,15,50], social networks (e.g. friendship networks, citation networks, voter networks, world markets, and political structures) [36,42,50], technological networks (e.g., computer networks, electrical circuits, and road networks) [1], and semantic networks (e.g. word-net [31] and recipe networks [43]). Without exception, all these networks can be represented as graphs, which include a wide variety of subgraphs. One fundamental property of networks are the so-called network motifs, which were introduced by Milo et al. [33]. They represent recurrent and statistically significant subgraphs or patterns in these complex networks. The fact that motifs repeat themselves in specific networks, or even among various networks, is highly correlated with the concepts of evolutionary theory. Each of these subgraphs, defined by a particular pattern of interactions between graph nodes, may reflect a framework in which particular functions are achieved efficiently. Motifs are considered to have a notable importance today because they may reflect underlying functional properties [30]. In light of their ability to uncover structural design principles of complex networks, motifs have been slowly adopted from Systems Biology into the broader perspective of Network Science. Although they foster a deep insight into the functional abilities of a network, their detection is computationally challenging even by current standards.

Particular research has been done in the areas of biology and genetics where motifs are associated with functional roles of transcription regulation networks which control the expression of genes [2]. Experimental studies show how motifs serve as basic building blocks of transcription networks. Another example is the understanding of how some cellular components are conserved across species but others evolve rapidly [54]. A notable study brings forward this new motif-inspired paradigm to uncover drug development strategies that help in the identification of drug target candidates [12]. A similar scientific track to our proposal is presented by Wang et al. in a study focused on detecting important nodes, not through the classic centrality metrics approach, but through specific motif patterns[49].

While conceptually (and functionally), complex networks can represent biological, technological, social or conceptual relationships between entities, we propose a motif-based analysis of networks from the topological perspective. As such, the fundamental topological families are: regular networks, random networks, small-world networks and scale-free networks [50]. Regular [8] and random

---

* Corresponding author. at Faculty of Automation and Computers, Politehnica University of Timisoara, Bd. Vasile Parvan 2, 300223, Timisoara, Romania. Tel.: +40 256 123456.

E-mail address: alext@cs.upt.ro (A. Topirceanu).

http://dx.doi.org/10.1016/j.comcom.2015.07.002

networks [16] represent the basics of complex networks. The effort to mathematically express accurate and realistic models of natural phenomena (e.g. social influence, collaboration, and internet communication) has been triggered by the observation of the three fundamental properties of complex networks: average path length, clustering coefficient and degree distribution [42,50]. The well-known models of small-world [51] and scale-free [4] networks both present these network properties. Since their introduction to the literature, a considerable amount of new networks have been added, yet all fall into one of the two categories: small-world or scale-free. To recreate natural processes with a higher fidelity, there are proposals which add the small-world property to scale-free models [18,20,28], or ones that add power-law degree distribution to the small-worlds [9,22,48,55].

Our work stems from an initiative to bring the concept of network motifs closer to the field of social networks analysis (SNA) and define a new way at looking at social graphs [14]. We bring substantial new insight in terms of the types of motifs analyzed, the size and number of real-world datasets and the results and conclusion based on this new research. Thus the motivation of this paper is to provide an analytical perspective over existing state of the art complex topologies using an novel approach – classification using the network structure, namely through network motifs.

In the second part of this paper, we apply this novel perspective to differentiate between online social networks. We use empirical data to demonstrate how real social networks can be classified with different levels of appurtenance to the four topological models. Even though similar in nature, it is shown in this paper that Facebook networks, Twitter networks and Google Plus networks have very distinct topological features, as revealed by the motif-based analysis. This points out to the different features the three social platforms have in the real world.

We set out to measure the motif distributions of sizes 3 and 4 on a comprehensive database of undirected online social networks. For this, we obtain encouraging results regarding the particular patterns each of the three mentioned online platforms reveals. Their fingerprint is highly visible in terms of distribution of triadic closures, which is correlated with the clustering of nodes and short paths in the graph. The mark of triads is important as it has been shown to drive the scaling and emergence of social networks in general [24]. Also, using our approach to reveal triadic closure formation is correlated with the predictability of evolving contacts in human proximity networks [40], an important aspect of modern communication frameworks. The classifiers we obtain for each of the three online social network classes are mapped onto the four topological families and also provide a new methodology of identifying key functional properties for new network data.

1.1. Motivation and outline

In light of the general concept-driven approach to complex and social networks analysis, we propose a new perspective of looking at networks from their topological point of view. This perspective is conceptualized in Fig. 1 using the four main complex network classes: regular, random, small-world and scale-free and is provided by in-depth network motif analysis. Thus, we bring forth the following main contributions:

- Large-scale computational generation and motif distribution analysis for the synthetic topology classes. We obtain a distinct motif pattern for each such class.
- Comprehensive motif analysis of online social networks (Facebook, Twitter, and Google Plus) from which we obtain three quantifiable characteristic motif fingerprints.
- Mapping and similarity assessment of empirical networks onto topology classes, and defining a general methodology for such an approach.

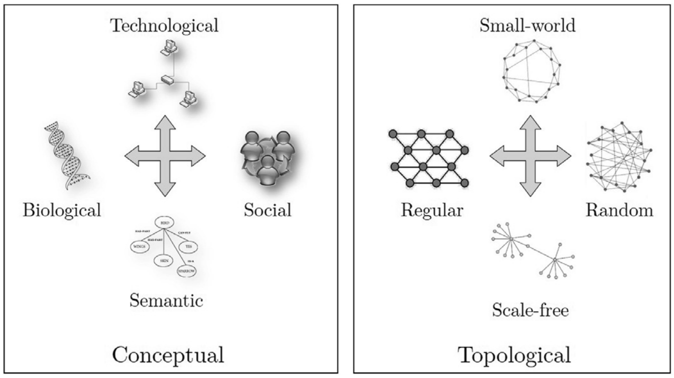

Fig. 1. The two classifications of complex networks: the conceptual perspective versus the topological perspective.

- Correlation and discussion of the individual motifs that occur in each fingerprint, and an outlining of the functional properties behind the three online social platforms.

2. A new perspective over the related work

Comparing complex networks is aimed at a deeper understanding of the interaction patterns between these systems [4,42,51], and extracting their common properties helps improve the models even further [3,23,51]. However, the predominant method of graph metric comparison suffers from limited information [27]. Some notable means of comparison are the distance ratio measure [7], used to compare individual mental models, a comparison from the data analysis perspective [27] and the study of the self-similarity of complex networks [41]. The network dimension is a key feature in understanding not only network topology, but also dynamical processes on networks, such as diffusion, percolation and other critical phenomena [13]. The fractal dimension d_B is proposed based on the belief that social networks are not invariant or self-similar under a length-scale transformation. Fractal dimension has been measured on multiple varied real world networks like the WWW, biological networks, and actor networks, and we will use it as an alternative to the standard metric comparison.

From a topological perspective there are studies done both in the direction of classifying social network models [23] and of structural pattern detection [37]. These methods however serve a higher level of meta-analysis rather than as measures of similarity.

The work done in the field of network motifs, since their introduction [33], has seen the definition of several super-families of evolved and designed networks by the same authors [32]. They present families of complex networks grouped together by the similar significance profiles (SP) of motifs in the networks compared to the normal occurrence in random networks. These families include:

- direct transcription interactions (in bacteria and yeast);
- signal-transduction interactions (cell signaling, neural networks);
- web hyperlinks and social networks;
- word-adjacency networks networks (in English, Spanish, Japanese).

Another study shows an alternative approach to the analysis of community structure by partitioning a network into a *core* of high degree nodes that are highly interconnected to each other, and a *periphery* of nodes that are not so well connected. The core has an important role in mediating most of the minimum path length motifs and has an integrative aspect over the topology [35].

With great preponderance, all studies revolve around the classification of networks – empirical or synthetic – from the conceptual point of view, into one of the mentioned four main categories.