Null Models For Social Networks

Jessica Su
Department of Computer Science
Stanford University
jtysu@stanford.edu

Sen Wu
Department of Computer Science
Stanford University
senwu@stanford.edu

ABSTRACT

Social motif has widely used for clustering large social network. However, graph generation algorithms often focus on degree distribution and lack of motif distribution. In this paper, we define the motif-driven graph generation problem and present a heuristic method which focus constructing the approximate motif distribution of large network. Our experiments on 67 real social network motif network distribution show that our approach obtains the generated graph with similar motif distribution within relevant error less then 0.02 in 10 hours.

Categories and Subject Descriptors

J.4 [**Social and Behavioral Sciences**]: Miscellaneous

General Terms

Algorithms, Experimentation

Keywords

Social networks

1. INTRODUCTION

2. BACKGROUND

In this section, we first give some basic concepts that we use throughout the paper. Then, we formulate the problem of motif-driven graph generation.

A **graph** is a representation of a set of objects and the connections between them. It is defined as a tuple (V, E), where V is a set of $|V| = N$ vertices ("nodes") and $E \subseteq V \times V$ is a set of edges. The nodes in a graph correspond to the objects, and the edges to the connections.

A graph is called **simple** if no node has an edge to itself and no two nodes have more than one edge between them. A graph is **undirected** if whenever (v, w) is an edge, (w, v) is also an edge. In our problem, we assume all graphs are simple and undirected.

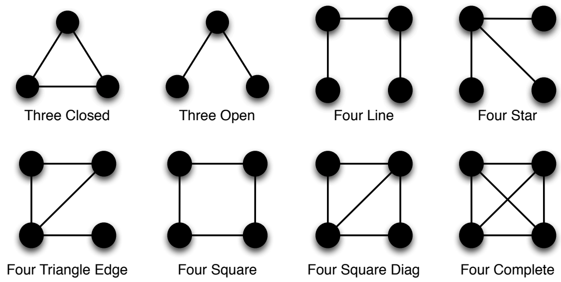

Figure 1: Graphical representation of the 3-motifs and 4-motifs.

A **subgraph** of G is a graph whose nodes and edges form subsets of the nodes and edges of the graph G. An **induced subgraph** of G on the vertices $V' \subseteq V$ is the graph consisting of the vertices V' and the edges between them.

A **motif** is a small, connected graph, often considered as a subgraph of a larger graph. In this paper, we only consider motifs with 3 or 4 vertices, which we term "3-motifs" and "4-motifs" respectively. A full list of motifs is shown in Figure 1.

The **motif distribution** of a graph G specifies how many of each motif type the graph contains. For example, a graph with motif distribution {mtThreeClosed: 12, mtThreeOpen: 16, mfFourLine: 22, mfFourSquare: 18, mfFourStar: 30, mfFourTriangleEdge: 10, mfFourSquareDiag: 17, mfFourComplete: 2} would contain twelve triangles, two 4-cliques, and so forth.

3. PROBLEM DEFINITION

Problem: To generate a network with desired specifications.

Input: The input of our problem consists of two components: the number of vertices and edges in the network, and the required motif distribution D.

Output: Our goal is to generate a graph G which has the required number of vertices and edges and a motif distribution that closely approximates D.

Ideally we would generate a graph with the exact motif distribution D. In the next section we show that problem is NP-hard, so the best we can hope for is an approximation.

The problem formulation is different from other graph generation problems [2, 5, 1, 4, 3], as in this paper we focus on generating graphs based on the motif distribution. Motif distributions are frequently used for analyzing large graphs in the social sciences.

4. EXACT MOTIF GENERATION IS NP-HARD

5. DATA AND OBSERVATIONS

6. OUR APPROACH

7. EXPERIMENTS

8. RELATED WORK

9. CONCLUSION

10. REFERENCES

[1] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.

[2] P. ERDdS and A. R&WI. On random graphs i. *Publ. Math. Debrecen*, 6:290–297, 1959.

[3] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random structures & algorithms*, 6(2-3):161–180, 1995.

[4] M. E. Newman. Random graphs with clustering. *Physical review letters*, 103(5):058701, 2009.

[5] D. J. Watts and S. H. Strogatz. Collective dynamics of Śsmall-worldŠnetworks. *nature*, 393(6684):440–442, 1998.