AI607: GRAPH MINING AND SOCIAL NETWORK ANALYSIS (FALL 2021)

Homework 2. Motifs and Graphlets

Release: September 24, 2021,
Due: October 8, 2021, 11:59pm

1 Introduction

Real-world graphs exhibit structural patterns that differentiate graphs in the same domain from random graphs or those in other domains. These structures are revealed by the occurrences of *network motifs*, which describe the connectivity pattern of a fixed number of nodes. In this assignment, you will understand the concept of network motifs and implement an algorithm for counting the instances of them. Moreover, you will find that network motifs reveal structural design principles of real-world graphs.

In this assignment, you will consider 3-node network motifs from M_1 to M_{13}, which are shown in Figure 1. Given a *directed* graph G, a set of three connected nodes is an *instance* of motif M_i if its connectivity pattern corresponds to M_i. You will count the number of instances of each motif M_i in G using the Exact Subgraph Enumeration (ESU) algorithm.

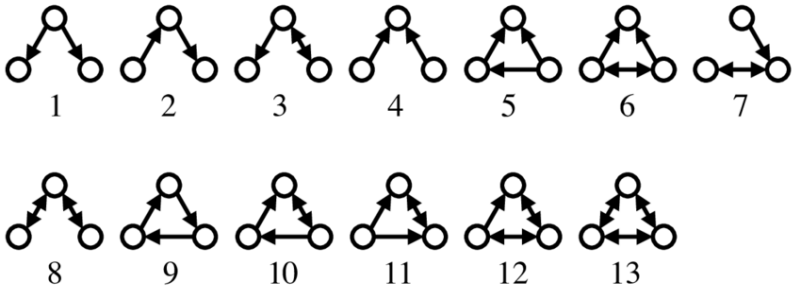

Figure 1: All 3-node network motifs denoted as M_1 to M_{13}.

2 Implementation [40 points]

In this section, you will count the number of instances of network motifs in the given graphs using the ESU algorithm. You are given **8 graphs**: 4 real-world graphs (`email`, `web`, `A`, and `B`) and 4 corresponding randomized graphs (`email_(A/B)_random` and `web_(A/B)_random`), which can be found in `data` folder. The `ESU` function returns the list of instances of three connected nodes.

1. (40 points) Implement the `countMotifs` function, which returns the absolute counts of the instances of each network motif, given a graph and its list of 3-node connected nodes.

You are allowed to use Numpy [1] and Scipy [2] for the implementation. You are **NOT** allowed to use any other external libraries (Snap.py, NetworkX, etc.).

3 Analysis [60 points]

In this section, you will analyze the structural properties of real-world graphs based on the number of instances of each network motif obtained in the previous section. Let C_i and $\tilde{C}_i$ be the number of instances of motif M_i in the real-world graph and the corresponding random graph, respectively. Then, the **significance** of M_i in a graph is defined as:

$$\triangle_i := \frac{C_i - \tilde{C}_i}{C_i + \tilde{C}_i + \epsilon},$$

where we simply set $\epsilon = 1$ in this assignment. Then, the **normalized significance** (NS) of M_i in the graph is defined as:

$$NS_i := \frac{\triangle_i}{\sqrt{\sum_{j=1}^{13} \triangle_j^2}}.$$

1. (40 points) Implement the `normalizedSignificance` function which computes the normalized significance of each network motif of a given graph. Plot a line graph where the x-axis is the network motif index (1 to 13), and the y-axis is the normalized significance of the corresponding network motif (NS_1 to NS_{13}).

2. (20 points) One between A and B is a web graph, and the other is an email. What do you think is a web graph? Provide reasons based on the normalized significance obtained in the previous step. Discuss your analysis in **HW2.ipynb**.

4 Notes

- Your implementation should run on TA's desktop within **10 minutes**.
- You may encounter some subtleties when it comes to implementation. Come up with your own design and/or contact Kyuhan Lee (kyuhan.lee@kaist.ac.kr) and Geon Lee (geonlee0325@kaist.ac.kr) for discussion. Any ideas can be taken into consideration when grading if they are written in the *readme* file.

[1] `https://numpy.org/install/`
[2] `https://scipy.org/install.html`

- In `HW2.ipynb`, do not modify the code outside of the "TASK" regions. For example, do not import additional python libraries outside "TASK" regions.

5 How to submit your assignment

1. Create hw2-[your student id].tar.gz, which should contain the following files:
   - **HW2.ipynb**: this should contain your implementation and answers in Section 3.2.
   - **readme.txt**: this file should contain the names of any individuals from whom you received help, and the nature of the help that you received. That includes help from friends, classmates, lab TAs, course staff members, etc. In this file, you are also welcome to write any comments that can help us grade your assignment better, your evaluation of this assignment, and your ideas.
2. All **2 files** should be included in a **single folder**.
3. Make sure that no other files are included in the tar.gz file.
4. Submit the tar.gz file at KLMS (`http://klms.kaist.ac.kr`).