ABBOUDI Mohammed Amine
mohammed-amine.abboudi@polytechnique.edu

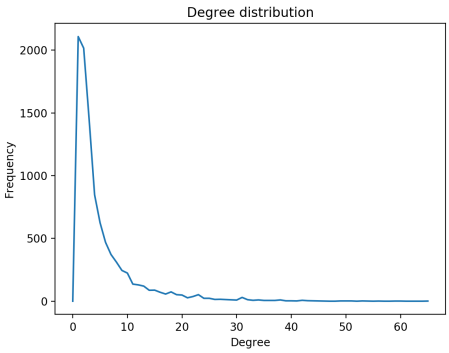
Lab session # 5
ALTEGRAD 2019

12/27/19

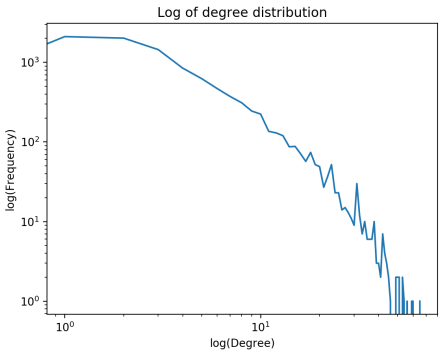
1 Question 1

The maximal degree attained by a vertex in an undirected graph is: $|V| - 1$; where $|V|$ is the number of vertices. It is referred to as a dominating vertex. Hence, the maximum number of edges in such a graph is: $\frac{|V| \times (|V|-1)}{2}$.
The maximum number of triangles than can be formed in an undirected graph is: $\binom{n}{3}$; $n = |V|$.

2 Question 2



(a) Degree histogram



(b) Log of degree histogram

We can see that a very small minority of authors are extremely collaborative while the majority almost always publishes with the same small amount of people. The degree distribution seems to be an exponential one or a Pareto distribution (based on the comparison of the original distribution and the log-log plot).

3 Question 3

Spectral clustering focuses on small eigenvalues because they give good approximations of zero eigenvalues. Since for a graph with d connected components, it will have 0 as eigenvalue with multiplicity d. So we would like to approximate their eigenvectors by selecting the d smallest eigenvalues and extracting clusters from them.
It seems that the problem to be optimized by spectral clustering is minimizing the probability that a random walk takes one from one cluster to another.

4 Question 4

$$l_{green} = 1, d_{green} = 2 \implies Q_{green} = \frac{1}{10} - (\frac{2}{20})^2 = \frac{9}{100}$$

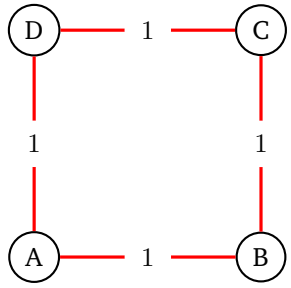
$$l_{blue} = 3, d_{blue} = 3 + 2 + 2 = 7 \implies Q_{blue} = \frac{3}{10} - (\frac{7}{20})^2 = \frac{3}{10} - \frac{49}{400} = \frac{71}{400}$$

$$l_{gray} = 5, d_{gray} = 3 + 2 + 3 + 3 = 11 \implies Q_{gray} = \frac{5}{10} - (\frac{11}{20})^2 = \frac{1}{2} - \frac{121}{400} = \frac{79}{400}$$

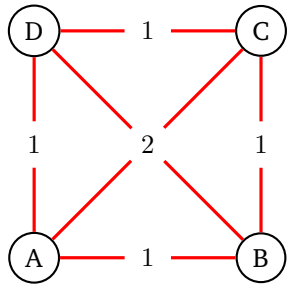
$$\implies \sum_{i}^{n_c} Q_i = \frac{9}{100} + \frac{71}{400} + \frac{79}{400} = 46.5\%$$

5 Question 5

A trivial way is to simply add an isolated vertex to the original graph. This can be circumvented by only considering the only considering non-zero labeled edges after the transformation.
A more straightforward example is the following. The first graph is:

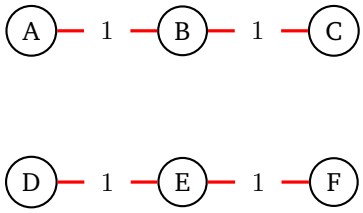


After Floyd-transformation, it becomes:

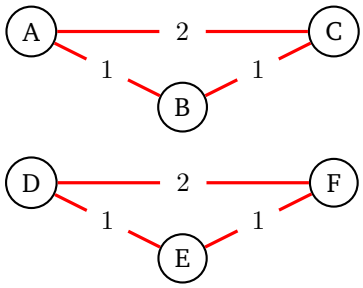


$$\Phi(G_1) = [2, 2, 1, 1, 1, 1, 0, ..., 0]$$

Another graph could be:



After Floyd-transformation, it becomes:



$$\Phi(G_2) = [2, 2, 1, 1, 1, 1, 0, ..., 0] = \Phi(G_1)$$

6 Question 6

A comparison of the achieved accuracy of the shortest path and graphlet kernels reveals that the shortest path kernel achieves an accuracy of 100% while the graphlet kernel lags behind with an error rate of 55%.
To explain the lack of performance we only consider in the following graphs that are comprised of a number of vertices strictly larger than 5. In this case, any sub-graph present in both instances of the dataset are in the form of a path. So a path graph and a cyclic graph of the same length have the same graphlet representation making the prediction as good as a random guess. Hence the poor score.

References