Virgile Rennard
virgile@rennard.org

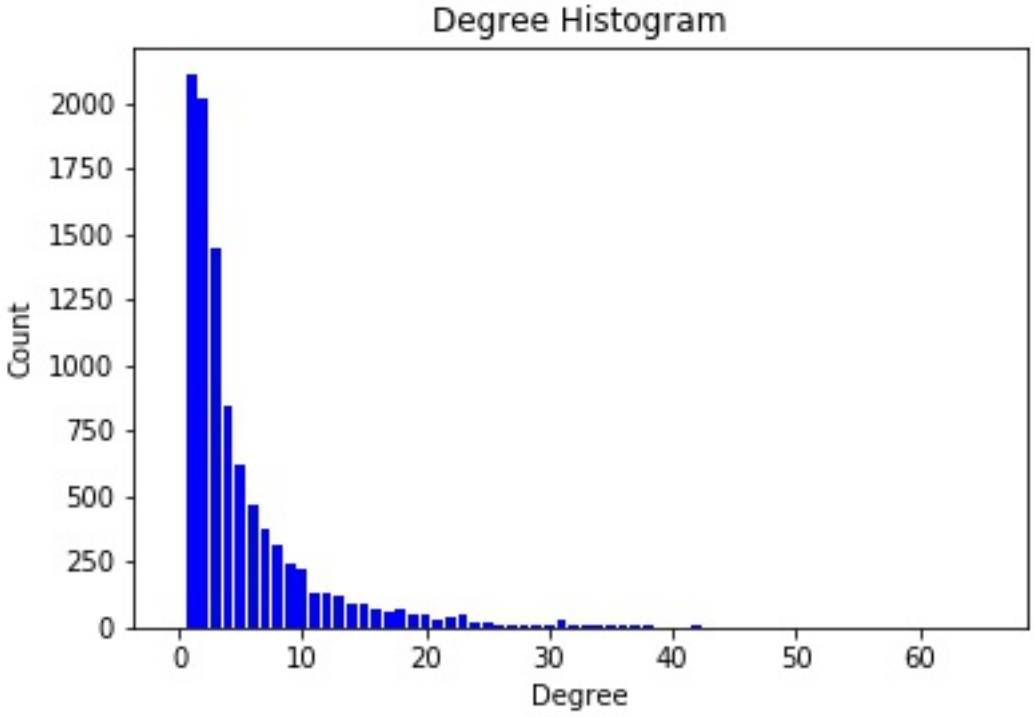
Lab session # 5
ALTEGRAD 2019

12/19/19

1 Question 1

In a graph of n nodes without self loops, the maximum number of edges is $\frac{n(n-1)}{2}$, the maximum number of triangles in such a graph is $C(3,n) = \frac{n(n-1)(n-2)}{6}$ and 0 if $n < 3$

2 Question 2



We can see that the number of nodes with a certain degree decreases as the degree increases. This is an exponential law

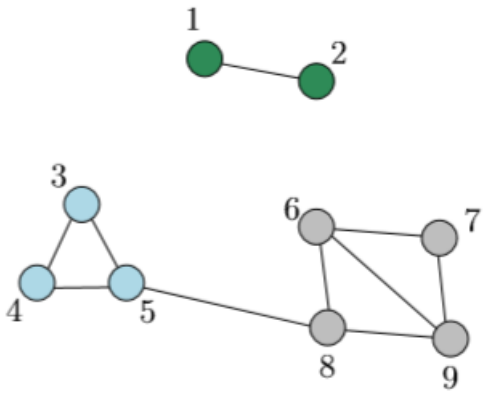
3 Question 3

The spectral clustering use a standard clustering method (such as the k-means, that we used here) on the relevent eigenvectors of the laplacian matrix. The relevant eigenvectors are the ones that correspond to the smallest eigenvalues of the Laplacian. This is because the eigenvalue 0 has a multiplicity that gives you the number of connected components, and that its eigenvector describe the connected components of the graph, as you get one eigenvector for each connected component where the non-zero entries correspond to members of a component. The clustering models the dataset as a graph and uses this decomposition to construct a better representation of the data. Once the custers are made, the k-means gets them easily.

4 Question 4

We want to compute the modularity for the graph
where the modularity is given by :

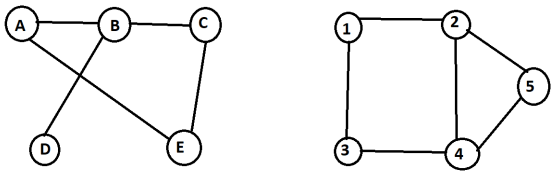
$$Q = \sum^{n_c} [\frac{l_c}{m} - (\frac{d_c}{2m})^2]$$



with $m = |E|$ the number of edges, n_c is the number of community, l_c is the number of edges within a community, and d_c is the sum of the degree of each nodes in the community. In our case we have : $n_c = 3, l_1 = 1, d_1 = 2, l_2 = 4, d_2 = 7, l_3 = 6, d_3 = 11, m = 10$ which gives us

$$Q = 0.09 + 0.1775 + 0.1975 = 0.465$$

5 Question 5



Both graphs map to [5,12,8,0,0] but are not isomorphic

6 Question 6

The graphlet kernel always outperforms the shortest path kernel. Whereas the shortest path kernel often gets us a perfect classification, the graphlet kernel oscillates between 0.4 and 0.55 accuracy. This stems form the fact that the kernel, defined as :

$$k(G, G') == f_G^T f_{G'}$$

is computed without taking account of the fact that the different graphs admit different sizes. One solution to it is to normalize the f_G by the number of all graphlets in G

References