Kyllian Asselin de Beauville
kyllian.asselin-de-beauville@polytechnique.edu

Lab session # 2
ALTEGRAD 2023

24/10/23

1 Question 1

Given an undirected graph with n nodes, we aim to determine the maximum number of edges and triangles the graph can have.

Maximum Number of Edges

Any node in the graph can connect to $n-1$ other nodes, leading to a total of $n(n-1)$ connections. However, due to the graph being undirected, each connection is counted twice (once for each node involved in the edge). Thus, to obtain the unique number of edges, we divide the total connections by 2.

The maximum number of edges an undirected graph of n nodes can have, E_{max}, is:

$$E_{\text{max}} = \frac{n(n-1)}{2} \tag{1}$$

Maximum Number of Triangles

A triangle in the graph is formed when three distinct nodes are all connected to each other. The maximum number of triangles is equivalent to the number of ways to select 3 nodes out of the n nodes. Using combinations, this is:

$$\binom{n}{3} = \frac{n!}{3!(n-3)!} = \frac{n(n-1)(n-2)}{6} \tag{2}$$

Therefore, the maximum number of triangles an undirected graph of n nodes can have, T_{max}, is:

$$T_{\text{max}} = \binom{n}{3} = \frac{n(n-1)(n-2)}{6} \tag{3}$$

2 Question 2

Having the same degree distribution does not imply that two graphs are isomorphic. While the degree sequence provides information about the number of edges connected to each node, it does not give a complete structural description of the graph as we can see in Figure 1.

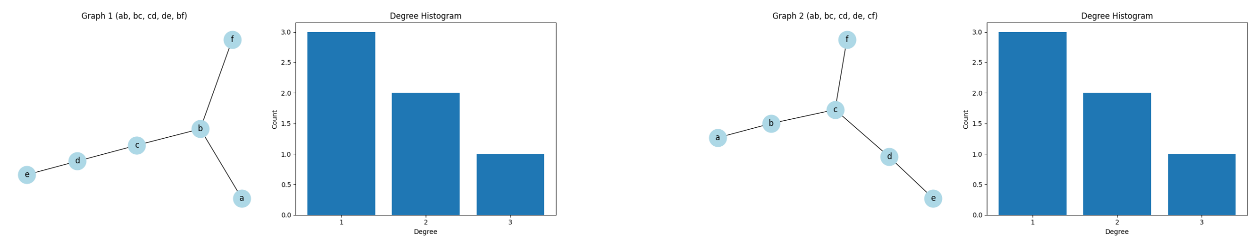

Figure 1: Counterexample

3 Question 3

The global clustering coefficient is defined as:

$$C = \frac{\text{number of closed triplets (triangles)}}{\text{number of closed triplets (triangles)} + \text{number of open triplets}} \tag{4}$$

Using this definition, we can compute the global clustering coefficient for different cycle graphs:

- C_3: There are 3 closed triplets and 0 open triplets. The global clustering coefficient for C_3 is given by:

$$C_{C_3} = \frac{3}{3} = 1 \tag{5}$$

- $C_4, C_5, \ldots$: There are 0 closed triplets and n open triplets. The global clustering coefficient for C_n, where $n \geq 4$, is given by:

$$C_{C_n} = \frac{0}{n} = 0 \tag{6}$$

So, the global clustering coefficient is 1 for C_3 (5), and 0 for any cycle graph C_n with $n \geq 4$ (6).

4 Question 4

Given the expression:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}([u_1]_i - [u_1]_j)^2 \tag{7}$$

This expression quantifies the smoothness of the eigenvector u_1 with respect to the graph structure encoded by the adjacency matrix A.

The term $([u_1]_i - [u_1]_j)^2$ represents the squared difference between the eigenvector values at nodes i and j, weighted by A_{ij}, which is 1 if nodes i and j are connected and 0 otherwise. The double summation accumulates these weighted squared differences for all node pairs, offering a global measure of how the eigenvector u_1 aligns with the graph's structure.

Since u_1 corresponds to the smallest eigenvalue of the Random Walk Laplacian L_{rw}, which is 0, its associated eigenvector is a vector of ones. Consequently, for any pair of nodes i and j that are connected (i.e., $A_{ij} = 1$), the term $([u_1]_i - [u_1]_j)^2$ will be 0, as all entries in the eigenvector u_1 are the same. Thus, the entire expression should output 0, indicating that u_1 perfectly aligns with the graph's structure.

5 Question 5

Modularity [2] is given by the following formula:

$$Q = \sum^{n_c} \left[\frac{l_c}{m} - (\frac{d_c}{2m})^2 \right] \tag{8}$$

where, $m = |E|$ is the total number of edges in the graph, n_c is the number of communities in the graph, l_c is the number of edges within the community c and d_c is the sum of the degrees of the nodes that belong to community c. Modularity takes values in the range $[-1, 1]$, with higher values indicating better community structure.

To compute the modularity for the given graphs (2), we identify the values for m, n_c, l_c and d_c for each community, and substitute them into the formula.

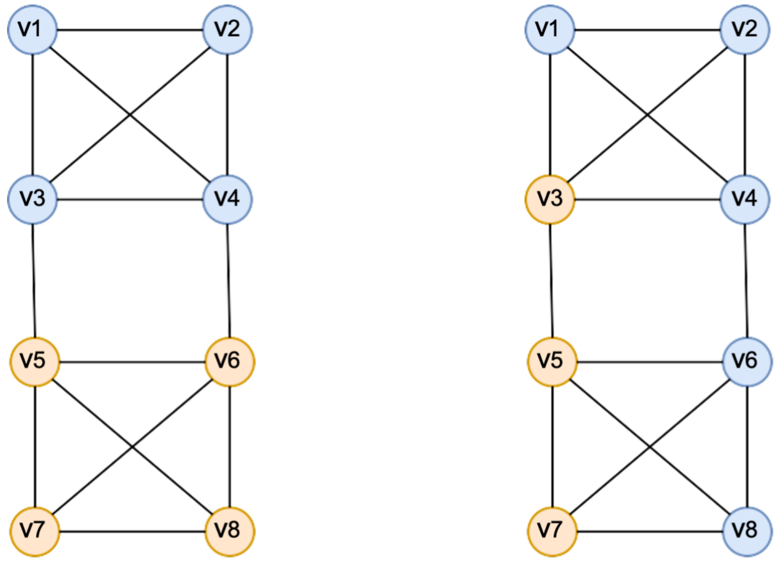

Figure 2: Graphs with 2 clusters. Cluster membership is indicated by node color.

Left Graph

I found the following parameters for the left graph:

$m = 14$, $n_c = 2$, $l_{c_{\text{blue}}} = 6$, $d_{c_{\text{blue}}} = 14$, $l_{c_{\text{orange}}} = 6$, $d_{c_{\text{orange}}} = 14$

Using the previously defined Formula 8, here is the computation:

$$\begin{aligned} Q_{\text{left}} &= \left[\frac{l_{c_{\text{blue}}}}{m} - \left(\frac{d_{c_{\text{blue}}}}{2m}\right)^2\right] + \left[\frac{l_{c_{\text{orange}}}}{m} - \left(\frac{d_{c_{\text{orange}}}}{2m}\right)^2\right] && (9) \\ &= \left[\frac{6}{14} - \left(\frac{14}{28}\right)^2\right] + \left[\frac{6}{14} - \left(\frac{14}{28}\right)^2\right] && (10) \\ &= \frac{5}{14} && (11) \\ &\simeq 0.36 && (12) \end{aligned}$$

Right Graph

I found the following parameters for the right graph:

$m = 14$, $n_c = 2$, $l_{c_{\text{blue}}} = 5$, $d_{c_{\text{blue}}} = 17$, $l_{c_{\text{orange}}} = 2$, $d_{c_{\text{orange}}} = 11$

Using the previously defined Formula 8, here is the computation:

$$\begin{aligned} Q_{\text{right}} &= \left[\frac{l_{c_{\text{blue}}}}{m} - \left(\frac{d_{c_{\text{blue}}}}{2m}\right)^2\right] + \left[\frac{l_{c_{\text{orange}}}}{m} - \left(\frac{d_{c_{\text{orange}}}}{2m}\right)^2\right] && (13) \\ &= \left[\frac{5}{14} - \left(\frac{17}{28}\right)^2\right] + \left[\frac{2}{14} - \left(\frac{11}{28}\right)^2\right] && (14) \\ &= -\frac{9}{392} && (15) \\ &\simeq -0.02 && (16) \end{aligned}$$

Based on these modularity scores (12 and 16), the community structure of the left graph appears stronger compared to the right graph, suggesting the left clustering is better than the right one.

6 Question 6

The shortest-path graph kernel is defined as follows in [1]:

Let G_1 and G_2 be two graphs that are Floyd-transformed into S_1 and S_2. We can then define our shortest-path graph kernel on $S_1 = (V_1, E_1)$ and $S_2 = (V_2, E_2)$ as

$$k_{\text{shortest paths}}(S1, S2) = \sum_{e_1 \in E_1} \sum_{e_2 \in E_2} k^{(1)}_{walk}(e_1, e_2) \tag{17}$$

where $k^{(1)}_{walk}$ is a positive definite kernel on edge walks of length 1.

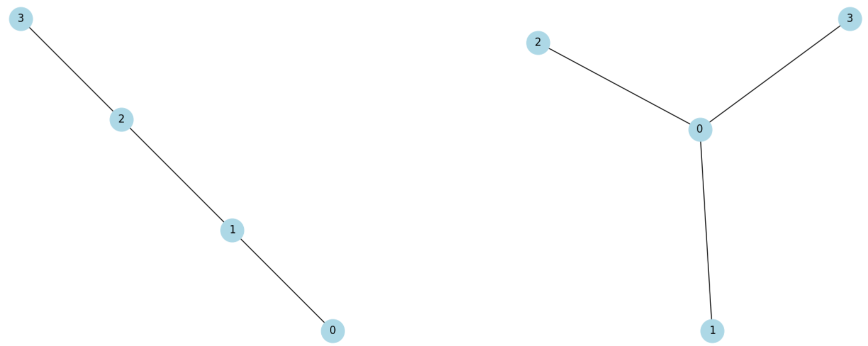

Figure 3: P_4 and S_4

Let's compute the Floyd-transformation of our graphs 3 (i.e. the shortest paths for all pair of nodes).

$$S_{P_4} = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 1 & 0 & 1 & 2 \\ 2 & 1 & 0 & 1 \\ 3 & 2 & 1 & 0 \end{bmatrix} \tag{18}$$

$$S_{S_4} = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 2 & 2 \\ 1 & 2 & 0 & 2 \\ 1 & 2 & 2 & 0 \end{bmatrix} \tag{19}$$

Figure 4: Floyd-transformed matrices

Given the matrix S_{P_4} (18) and considering only the upper triangle, the numbers 1, 2, and 3 occur 3, 2, and 1 times, respectively. This forms the feature map $\phi(P_4) = [3, 2, 1]$. For S_{S_4} (19), the numbers 1 and 2 occur 3 times each, forming the feature map $\phi(S_4) = [3, 3, 0]$.

From this, the shortest-path kernels are:

$$k_{\text{shortest paths}}(S_{P_4}, S_{P_4}) = \phi(P_4) \cdot \phi(P_4) = 3 \times 3 + 2 \times 2 + 1 \times 1 = 14 \tag{20}$$

$$k_{\text{shortest paths}}(S_{P_4}, S_{S_4}) = \phi(P_4) \cdot \phi(S_4) = 3 \times 3 + 2 \times 3 + 1 \times 0 = 15 \tag{21}$$

$$k_{\text{shortest paths}}(S_{S_4}, S_{S_4}) = \phi(S_4) \cdot \phi(S_4) = 3 \times 3 + 3 \times 3 + 0 \times 0 = 18 \tag{22}$$

7 Question 7

If $k(G, G') = f_G^\top f_{G'} = 0$, it implies that the two graphs, G and G', do not share any common graphlets [3] of size 3, i.e. the decomposition of these two graphs into graphlets of size 3 do not overlap. In other terms, they have no shared components of size 3.

The graphs from Figure 5 with graphlets of size 3 from Figure 6 respect this condition and therefore have no shared components. The complete graph K_4 has only 4 closed triangles, i.e. $4 \times G_1$, while the cyclic graph has only 4 open triplets, i.e. $4 \times G_2$.

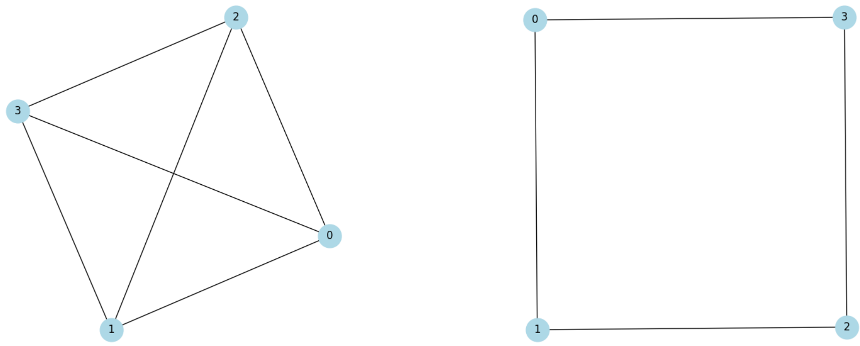

Figure 5: Complete graph K_4 and cyclic graph C_4

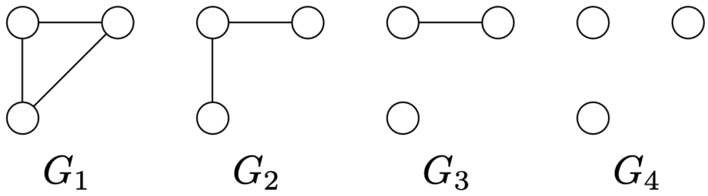

Figure 6: Graphlets of size 3

References

[1] K.M. Borgwardt and H.P. Kriegel. Shortest-path kernels on graphs. In *Fifth IEEE International Conference on Data Mining (ICDM'05)*, pages 8 pp.–, 2005.

[2] M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, jun 2006.

[3] N. Sherashidze, S.V.N. Vishwanathan, Tobias Petri, Kurt Mehlhorn, and Karsten Borgwardt. Efficient graphlet kernels for large graph comparison. *12th International Conference on Artificial Intelligence and Statistics (AISTATS), Society for Artificial Intelligence and Statistics, 488-495 (2009)*, 5, 01 2009.