Boubacar SOW
boubacar.sow@ens-paris-saclay.fr

Lab session # 2
ALTEGRAD 2023

10/24/23

1 Maximum Number of Edges and Triangles in an Undirected Graph

An undirected graph with n nodes without self-loops can have a maximum number of edges and triangles. The maximum number of edges E is given by:

$$E = \frac{n(n-1)}{2}$$

This is because each node can connect to $n-1$ other nodes, and we divide by 2 to avoid double-counting the edges (since it's an undirected graph).
The maximum number of triangles T is given by:

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

This is because we are choosing 3 nodes out of n to form each triangle.

2 Isomorphism of graphs.

Two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ having identical degree distributions does not necessarily imply that they are isomorphic. Isomorphism requires a bijective mapping $f : V_1 \to V_2$ such that $(v_i, v_j) \in E_1$ if and only if $(f(v_i), f(v_j)) \in E_2$.
A simple counterexample would be an hexagon graph (a cycle graph with 6 vertices) and another graph that consists of two disconnected triangles. Both graphs have the same degree distribution (each node has a degree of 2), but they are not isomorphic because there is no way to map the vertices of one onto the other while preserving the structure of the graphs.

3 Global clustering coefficient

The global clustering coefficient of any cycle graph that has more than 3 vertices is 0 because such graph do not have any triangle. If a cycle graph has 3 vertices, then it's global coefficient is 1.

4 Interpretation of the Expression Involving the Eigenvector of L_{rw}

The expression $\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}([u_1]_i - [u_1]_j)^2$ computes a weighted sum of the squared differences of the elements of the eigenvector u_1 associated with the smallest eigenvalue of the random walk Laplacian matrix L_{rw}. The weights are given by the adjacency matrix A, meaning that the differences are only considered for pairs of nodes that are connected (i.e., for which $A_{ij} \neq 0$).
This expression measures the total variation in the eigenvector values across all edges in the graph. In other words, it quantifies how much the eigenvector values change across edges in the graph. However, because u_1 is associated with the smallest eigenvalue of L_{rw}, and for a connected graph this smallest eigenvalue is 0 with a corresponding eigenvector of constant values (all entries are equal), this sum should be equal to 0. This is because for any pair of connected nodes i and j, we have $[u_1]_i = [u_1]_j$, so their difference is 0.
The expression outputs the weighted sum by the difference of the elements of the eigenvector u_1 of the adjacency matrix. This expression measures the total variation in the eigenvector values across all edges in the graph.

5 Compute (showing your calculations) the modularity of the clustering results shown in Figure 1.

The modularity Q for a given graph and its clustering can be computed using the formula:

$$Q = \sum_{c=1}^{nc} \left(\frac{lc}{m} - \left(\frac{dc}{2m} \right)^2 \right)$$

For Graph 1:

- Total number of edges, $m = 14$
- Number of communities, $nc = 2$
- For community 1: $lc1 = 6$, $dc1 = 14$
- For community 2: $lc2 = 6$, $dc2 = 14$

Plugging these values into the formula, we get:

$$Q_{\text{Graph 1}} = \left(\frac{6}{14} - \left(\frac{14}{2 * 14}\right)^2\right) + \left(\frac{6}{14} - \left(\frac{14}{2 * 14}\right)^2\right)$$
$$Q_{\text{Graph 1}} = 0.35$$

For Graph 2:

- Total number of edges, $m = 14$
- Number of communities, $nc = 2$
- For community 1: $lc1 = 2$, $dc1 = 11$
- For community 2: $lc2 = 5$, $dc2 = 17$

Plugging these values into the formula, we get:

$$Q_{\text{Graph 2}} = \left(\frac{2}{14} - \left(\frac{11}{2 * 14}\right)^2\right) + \left(\frac{5}{14} - \left(\frac{17}{2 * 14}\right)^2\right)$$
$$Q_{\text{Graph 2}} \ = -0.01$$

We can see here that the first clustering (the graph on the left) is better than the second clustering as it has greater modularity score.

6 Shortest Path Kernel Calculation for P_4 and S_4

Let's first compute the shortest paths of the individual graphs:

- **Path Graph** P_4: The shortest paths are of lengths 1, 2, and 3. The frequencies of these paths are 3, 2, and 1 respectively. We have $\phi(P4) = [3, 2, 1, 0, ..., 0]$
- **Star Graph** S_4: The shortest paths are of lengths 1 and 2. The frequencies of these paths are 3 each. We have $\phi(S4) = [3, 3, 0, 0, ..., 0]$

Now, let's calculate the shortest path kernel for the given pairs: $K(A, B) = \phi(A)@\phi(B)$ where @ is the scalar product.

- (P_4, P_4): All shortest paths of equal lengths will match. Therefore, $k(P_4, P_4) = (3 \times 3) + (2 \times 2) + (1 \times 1) = 14$.
- (P_4, S_4): Only the paths of lengths 1 and 2 will contribute to the kernel. Hence, $k(P_4, S_4) = (3 \times 3) + (2 \times 3) = 15$.
- (S_4, S_4): All shortest paths of equal lengths will match. Therefore, $k(S_4, S_4) = (3 \times 3) + (3 \times 3) = 18$.

7 Let k denote the graphlet kernel that decomposes graphs into graphlets of size 3. Let also G, G' denote two graphs and suppose that $k(G, G') = f_G^\top f_{G'} = 0$. What does a kernel value equal to 0 mean? Give an example of two graphs G, G' for which $k(G, G') = 0$ holds.

A kernel value of 0, i.e., $k(G, G') = f_G^\top f_{G'} = 0$, implies that there are no common graphlets of size 3 between the two graphs G and G' (or the two graphs are dissimilar with respect to the kernel function. This means that when the graphs are decomposed into graphlets of size 3, none of them are isomorphic between G and G'.

For instance, consider two simple graphs:

- Graph G: A path graph with 3 nodes and 2 edges (a straight line or a path graph).
- Graph G': A cycle graph with 3 nodes and 3 edges (a triangle).

In this case, the graphlet kernel $k(G, G')$ would be equal to 0 because the graphlets (subgraphs of size 3) in G and G' are not isomorphic. The graphlet in G is a path of length 2, while the graphlet in G' is a cycle of length 3. Since there are no common graphlets between G and G', the dot product $f_G^T f_{G'}$ equals 0.

References