Celine Hajjar
celinehajjar98@gmail.com

Lab session # 4
ALTEGRAD 2021

01/02/22

Note: In order to run the code provided with this report, please change the directory specified in the `os.chdir()` function at the beginning of every code file.

Question 1

G is a cycle graph with n nodes, so it contains n edges. Removing 2 of those edges will result in 2 connected components.

Question 2

Two graphs can have identical degree distributions while being non-isomorphic. A counterexample is shown in figure 1 (based on the course slides). In the second iteration of the Weisfeiler-Lehman test, the graphs have different label sets, and are thus non-isomorphic even though they have the same degree distribution.

(a) G_1 (b) G_2

Figure 1: Two non-isomorphic graphs with identical degree distributions

Question 3

I will base my answer on the following formula presented in the course slides:

$$C = \frac{\text{number of closed triplets}}{\text{total number of triplets}} = \frac{3 \times \text{number of triangles}}{\text{total number of triplets}} \tag{1}$$

where a triplet is defined as a set of two edges sharing a common node, and a triangle is defined as a set of three connected nodes.
In our case, if G was a complete graph, it would have $\frac{n(n-1)}{2}$ edges, but it only has $\frac{n(n-1)}{2} - 1$ edges so it can be seen like a complete graph missing one edge. An illustration of such a graph with $n = 5$ is presented in figure 2

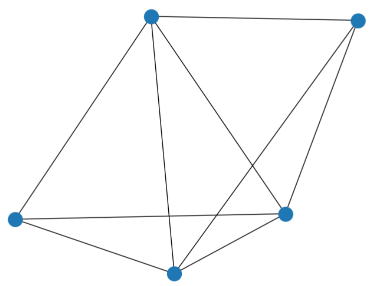

Figure 2: An example of a graph with 5 nodes and 9 edges

It is clear that in that case:

$$\text{number of closed triplets} = \text{number of all triplets - (n-2)} \quad (2)$$

because we have $n-2$ triplets that involve the two nodes that are not connected (and are thus open triplets). In a complete graph, we have $\times C_n^3$ triangles, and every triangle accounts for 3 triplets (each triplet is centered in one vertex of the triangle), so we have a total of $3 \times C_n^3$ triplets. However, in our case, the "triangles" that involve the two nodes that are not connected account for only one triplet (the one centered in the third node). So we have 2 less triplets in every such triangle. Knowing that we have $n-2$ such triangles, we have $2(n-2)$ less triplets than in the case of a complete graph. The number of all triplets becomes:

$$\text{number of all triplets} = 3C_n^3 - 2(n-2) \quad (3)$$

And the global clustering coefficient would be:

$$C = \frac{3C_n^3 - 3(n-2)}{3C_n^3 - 2(n-2)} \quad (4)$$

$$= 1 - \frac{n-2}{3C_n^3 - 2(n-2)} \quad (5)$$

This formula is verified for the graph in the example above ($n = 5$): `nx.transitivity(g)` returns the value of 0.875.

Question 4

In proposition 2 of [1] (section 3), it is shown that the multiplicity k of the smallest eigenvalue is equal to the number of connected components $A_1, A_2, ...A_k$.
Also, the Laplacian of the subgraph corresponding to the i^{th} component has the smallest eigenvalue with multiplicity 1, so the corresponding eigenvector is the constant one vector on the i^{th} connected component, and 0 elsewhere. Thus, the eigenvectors corresponding to the smallest eigenvalue are the indicator vectors of the connected components.

Question 5

The computation of the Laplacian matrix and its eigenvalues is a deterministic process. However, the initialisation of the k centers in the *k-means* algorithm is stochastic, so the outcome of the spectral clustering is stochastic as well.

Question 6

In both cases we have $m = 8$.

Graph a : Community 1 is blue and community 2 is green. The respective l_c and d_c are: $n_c = 2$, $l_1 = 4$, $d_1 = 9$, $l_2 = 3$, $d_2 = 7$. The modularity is:

$$Q = \frac{4}{8} - \left(\frac{9}{2 \times 8}\right)^2 + \frac{3}{8} - \left(\frac{7}{2 \times 8}\right)^2$$
$$= \frac{47}{128} \approx 0.37$$

Graph b : Community 1 is blue, community 2 is orange, and community 3 is green. The respective l_c and d_c are: $n_c = 3$, $l_1 = 1$, $d_1 = 4$, $l_2 = 2$, $d_2 = 8$, $l_3 = 1$, $d_3 = 4$.
The modularity is:

$$Q = \frac{1}{8} - \left(\frac{4}{16}\right)^2 + \frac{2}{8} - \left(\frac{2}{16}\right)^2 + \frac{1}{8} - \left(\frac{4}{16}\right)^2$$
$$= \frac{5}{16} \approx 0.31$$

We find the community structure detected in the first case is better.

Question 7

We have:
$\phi(P_4) = [3, 2, 1, 0, \ldots, 0]$, and $\phi(K_4) = [6, 0, \ldots, 0]$.
Which yields the following kernels:

- $k(P_4, P_4) = \langle \phi(P_4), \phi(P_4) \rangle = 14$
- $k(P_4, K_4) = \langle \phi(P_4), \phi(K_4) \rangle = 18$
- $k(K_4, K_4) = \langle \phi(K_4), \phi(K_4) \rangle = 36$

References

[1] Ulrike von Luxburg. A tutorial on spectral clustering. *CoRR*, abs/0711.0189, 2007.