Basile Terver
basile.terver@polytechnique.edu

Lab session # 2
ALTEGRAD 2023

10/23/23

1 Task 1

```
the graph has  9877 nodes.
the graph has  25998 edges.
is the graph directed False.
```

2 Question 1

We can have at most $\frac{n(n-1)}{2}$ edges into a simple undirected graph, which corresponds to a complete graph. In a simple undirected graph, we can have at most $\binom{n}{3}$ triangles. Indeed, to have the maximum number of triangles, we must consider a complete graph, where choosing a triangle is equivalent to choosing 3 vertices.

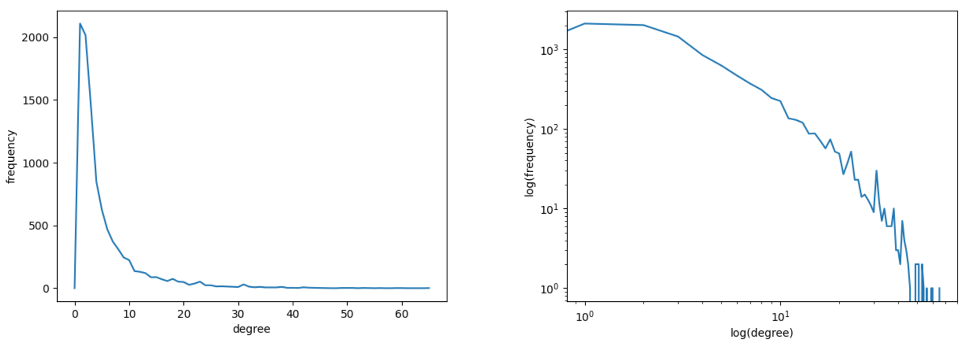
3 Task 2

```
The graph has 429 connected components.
The largest connected component in the graph has 8638 nodes.
the largest connected component in the graph has 24827  edges.
This corresponds to  0.9549580736979768  of the original graph's edges.
This corresponds to  0.8745570517363572  of the original graph's nodes.
```

4 Task 3

```
min degree of graph is  1
max degree of graph is  65
median degree of graph is  3.0
mean degree of graph is  5.264351523742027
```

5 Task 4



(a) Plot of the degree histogram, which gives the frequency of each (integer) degree value.

(b) Plot of the degree histogram using log-log axis.

6 Question 2

A counterexample is the one of a hexagon and two triangles. In both cases, all vertices have same degree 2, so the degree distribution is all in degree 2. Those graphs are obviously not isomorphic.

7 Task 5

```
the global clustering coefficient of the HepTh graph is  0.2839996525875546.
```

8 Question 3

The global clustering coefficient of C_3, which is a triangle, is one.
The global clustering coefficient of C_4, which has no triangle, is zero. Same for all $C_n, n \geq 4$.

9 Question 4

We assume that there are no isolated vertices in the graph so that the degree matrix D has no zero coefficients on the diagonal and is therefore invertible.
$\boldsymbol{L}_{rw}$ is called the left (random-walk) normalized Laplacian matrix of the graph G. Let the Laplacian matrix of G be defined by $L := D - A$, where D is the diagonal degree matrix and A is the adjacency matrix.
We know that the left (random-walk) normalized Laplacian matrix is similar to the normalized symmetric Laplacian matrix, more precisely we have

$$\boldsymbol{L}_{rw} = D^{-1/2}\boldsymbol{L}^{sym}D^{1/2}, \tag{1}$$

where the normalized symmetric Laplacian is given by

$$\boldsymbol{L}^{sym} = D^{-1/2}LD^{-1/2} \tag{2}$$

On the other hand, for any vector $u \in \mathbb{R}^n$, we have

$$\begin{aligned}
\sum_{i,j=1}^{n} A_{ij}(u_i - u_j)^2 &= \sum_{i,j=1}^{n} A_{ij}u_i^2 - 2\sum_{i,j=1}^{n} A_{ij}u_iu_j + \sum_{i,j=1}^{n} A_{ij}u_j^2 \\
&= \sum_{i=1}^{n} D_{ii}u_i^2 - 2\sum_{i,j=1}^{n} A_{ij}u_iu_j + \sum_{j=1}^{n} D_{jj}u_j^2 = 2\left(\sum_{i=1}^{n} D_{ii}u_i^2 - \sum_{i,j=1}^{n} A_{ij}u_iu_j\right) \\
&= 2(u^\top Du - u^\top Au) = 2u^\top Lu
\end{aligned}$$

Using the above this yields

$$\sum_{i,j=1}^{n} A_{ij}(u_i - u_j)^2 = 2u^\top Lu = 2u^\top D^{1/2}\boldsymbol{L}^{sym}D^{1/2}u = 2u^\top D\boldsymbol{L}_{rw}u \tag{3}$$

In our case, u is the eigenvector associated with the smallest eigenvalue of $\boldsymbol{L}_{rw}$. This eigenvalue is zero because $\boldsymbol{L}_{rw}$ and $\boldsymbol{L}^{sym}$ are similar so they have same eigenvalues and $\boldsymbol{L}^{sym}$ has n non-negative, real valued eigenvalues with the smallest one equal to zero. Thus, we have $\boldsymbol{L}_{rw}\boldsymbol{u}_1 = 0$ and

$$\sum_{i,j=1}^{n} A_{ij}([\boldsymbol{u}_1]_i - [\boldsymbol{u}_1]_j)^2 = 0. \tag{4}$$

10 Question 5

10.1 Left graph

We have $m = 6 + 2 + 6 = 14$, 2 communities with $l_1 = 6 = l_2$. In both communities, 2 vertices have degree 3 and the two others have degree 4, so $d_1 = d_2 = 14$. The modularity of this partition is

$$Q_l = 2\left(\frac{6}{14} - \left(\frac{14}{2\cdot 14}\right)^2\right) = 2\left(\frac{3}{7} - \frac{1}{4}\right) = 2\frac{12-7}{28} = \frac{5}{14}.$$

10.2 Right graph

We have $m = 6 + 2 + 6 = 14$, 2 communities with $l_1 = 2, l_2 = 5$. In both communities, 2 vertices have degree 3 and the two others have degree 4, so $d_1 = 4 + 4 + 3 = 11, d_2 = 3 + 4 + 4 + 3 + 3 = 17$. The modularity of this partition is

$$Q_r = \left(\frac{1}{7} - \left(\frac{11}{2\cdot 14}\right)^2\right) + \left(\frac{5}{14} - \left(\frac{17}{2\cdot 14}\right)^2\right) \simeq -0,023.$$

We have a higher modularity for the left graph, which has better community structure so this result is consistent.

11 Task 9

```
Modularity of the spectral clustering with 50 clusters: 0.16207588664036468
Modularity of the random clustering with 50 clusters: 0.0014939040949855456
```

12 Question 6

The feature maps of the two graphs are

$$\phi(P_4) = [3, 2, 1, 0, ...],$$
$$\phi(S_4) = [3, 3, 0, 0, ...]$$

Therefore, using the usual inner product on $\mathbb{R}^{\mathbb{N}}$, we have

$$k(P_4, P_4) = 3^2 + 2^2 + 1 = 14,$$
$$k(P_4, S_4) = 3^2 + 3 \cdot 2 = 15,$$
$$k(S_4, S_4) = 2 \cdot 3^2 = 18.$$

13 Question 7

A kernel value $k(G, G')$ equal to zero means that f_G do not have any graphlets of size 3 in common.
An exemple of such pair of graphs is given by two graphs of 4 nodes.

- A complete graph G_a, such that each pair of nodes is linked by an edge. Its feature vector is $[4, 0, 0, 0]$.
- A graph G_b made of one path of 3 nodes (2 edges) and one isolated vertex. Its feature vector is $[0, 1, 2, 1]$.

14 Task 13

```
Accuracy of shortest path kernel classification is  1.0
Accuracy of graphlet kernel classification is  0.3
```

References