Simon Queric
simon.queric@telecom-paris.fr

Lab session # 2
ALTEGRAD 2023

10/23/23

Analysing a Real-World Graph

Question 1

Let $G = (V, E)$ be an undirected graph without self-loops. The maximum number of edges of this graph is $\binom{|V|}{2} = \frac{|V|(|V|-1)}{2}$. The maximum number of triangles is $\binom{|V|}{3} = \frac{|V|(|V|-1)(|V|-2)}{2}$.

Question 2

Let us consider the two graphs below.

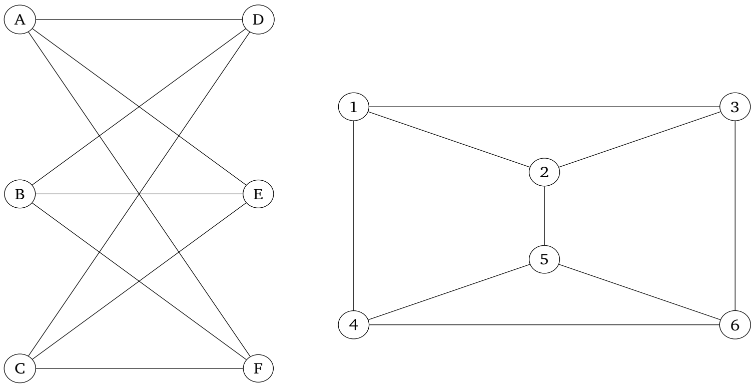

Figure 1: Example of two graphs with same degree distribution that aren't isomorphic.

These two graphs have same degree distribution. Indeed, each node has three neighbors. However the left one is bipartite (with partition $\{\{A, B, C\}, \{D, E, F\}\}$) and not planar, while the right one isn't bipartite (it contains triangles, e.g $\{1, 2, 3\}$) and planar. The planarity and the bipartite property are two graph invariants. Hence these two graphs are not isomorphic.
Here is an other example of two graphs with identical degree distribution $(1, 2, 2, 2, 3)$ but that are not isomorphic (the left one is bipartite, the right one is not) :

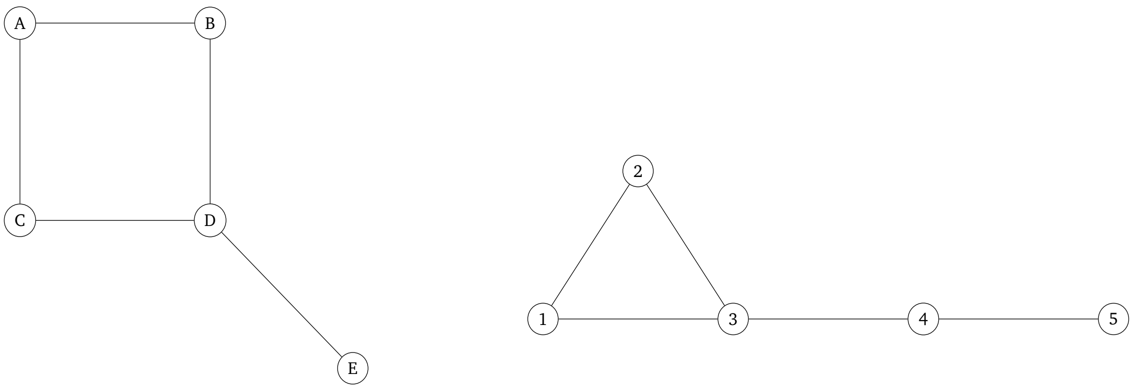

Figure 2: An other example of two graphs with same degree distribution that aren't isomorphic.

Question 3

For $n \geqslant 3$, we write c_n the *global clustering coefficient* of the graph C_n.

The graph C_3 has no open triplet and one closed triplet, hence $c_3 = 1$.
For $n \geqslant 4$, C_n has no closed triplet hence $c_n = 0$.

Community Detection

Question 4

Notice that, the adjacency matrix and the degrees are related by the formula $d_i = \sum_{j=1}^n A_{i,j}$ and $d_j = \sum_{i=1}^n A_{i,j}$ where d_i is the degree of the node i. We write λ an eigenvalue of the Laplacian matrix. Let u be an eigenvector of L_{rw} associated to λ, we have $(I - D^{-1}A)u = \lambda u$ which leads to $Au = (1-\lambda)Du$.

$$\begin{aligned}
\sum_{i=1}^n \sum_{j=1}^n A_{i,j}(u_i - u_j)^2 &= \sum_{i=1}^n \sum_{j=1}^n A_{i,j}(u_i^2 - 2u_iu_j + u_j^2) \\
&= \sum_{i=1}^n u_i^2 \sum_{j=1}^n A_{i,j} - 2\sum_{i=1}^n \sum_{j=1}^n A_{i,j}u_iu_j + \sum_{j=1}^n u_j^2 \sum_{i=1}^n A_{i,j} \\
&= 2\sum_{i=1}^n u_i^2 d_i - 2u^T Au \\
&= 2u^T Du - 2(1-\lambda)u^T Du \\
&= 2\lambda u^T Du
\end{aligned}$$

Now, we can show that $\lambda_1 = 0$.
What we showed above is that $2\lambda_1 u_1^T Du_1 \geqslant 0$ because $A \geqslant 0$. Hence $\lambda_1 \geqslant 0$.
Then, let $\mathbf{1}$ be the vector of size n full of 1.
Then

$$L_{rw}\mathbf{1} = (I - D^{-1}A)\mathbf{1} = \mathbf{1} - D^{-1}d = \mathbf{1} - \mathbf{1} = 0$$

where d is the vector of degrees.
Hence, 0 is an eigenvalue of L_{rw} and it's the smallest. We can conclude that

$$\sum_{i=1}^n \sum_{j=1}^n A_{i,j}(u_{1,i} - u_{1,j})^2 = 0$$

Question 5

In the left graph, we have $m = 14$ edges. For $c \in \{1,2\}$ we have $l_c = 6, d_c = 14$. Finally the modularity associated to this clustering is $Q_l = 2\left(\frac{6}{14} - \left(\frac{14}{2\times14}\right)^2\right) = 0.35$.
In the right graph, we have $m = 14$ edges. Let's say that the orange cluster is the cluster 1 and that the blue one is the cluster 2. Then, $l_1 = 2, l_2 = 5, d_1 = 11, d_2 = 17$ and $Q_r = \left(\frac{2}{14} - \left(\frac{11}{2\times14}\right)^2\right) + \left(\frac{5}{14} - \left(\frac{17}{2\times14}\right)^2\right) = -0.02$.
The two graphs are the same and we can see visually that the left clustering is better than the right one. We also have $Q_r < Q_l$: it's an example of the modularity metric which measures the quality of clustering. The left clustering is good while the right one is poor.

Graph Classification using Graph Kernels

Question 6

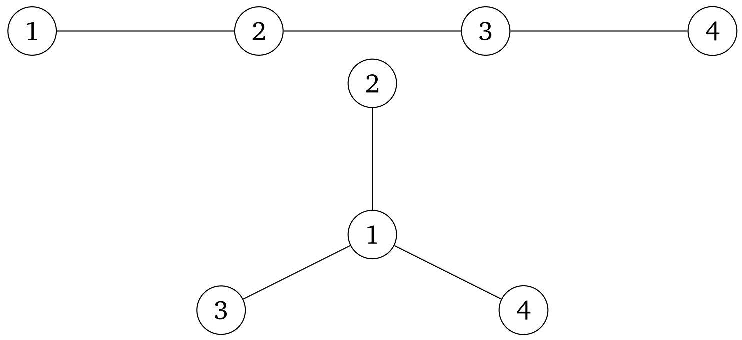

Figure 3: Drawing of P_4 (above) and S_4 (below).

In the path graph P_4, we count 3 shortest paths of length 1, 2 of length 2 and 1 of length 3. We have $\phi(P_4) = (3,2,1)^T$.
In the star graph S_4, we count 3 shortest paths of length 1 and 3 of length 3. We have $\phi(S_4) = (3,0,3)^T$.
Hence :

$$\begin{aligned} k_{SP}(P_4, P_4) &= \langle \phi(P_4), \phi(P_4) \rangle = 14 \\ k_{SP}(P_4, S_4) &= \langle \phi(P_4), \phi(S_4) \rangle = 12 \\ k_{SP}(S_4, S_4) &= \langle \phi(S_4), \phi(S_4) \rangle = 18 \end{aligned}$$

Question 7

A graphlet kernel value equal to zero means that the two graphs G and G' don't have any graphlet in common. For instance, if we take i, j in $\{1,2,3,4\}$ then $f_{G_i}^T f_{G_j} = \delta_{i,j}$. So two different graphlets have a kernel value equal to zero.
Here is another example of two graphs whose kernel is equal to zero :

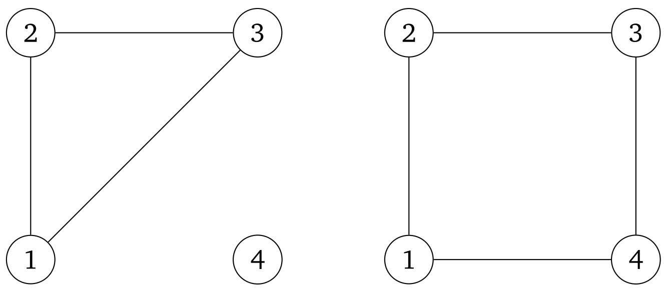

Figure 4: Example of two graphs, G on the left, G' on the right such that $k(G, G') = 0$.

Indeed, we have $f_G = (1,0,3,0)^T$ and $f_{G'} = (0,4,0,0)^T$, then $k(G, G') = 0$.