1 Question 1

The number of edges is the number of 2-combinations you can form with n edges i.e $\binom{n}{2} = \frac{n(n-1)}{2}$.

The number of triangles is the number of 3-combinations you can form with n edges i.e : $\binom{n}{3} = \frac{n(n-1)(n-2)}{6}$

2 Question 2

The degrees follow an exponential distribution.

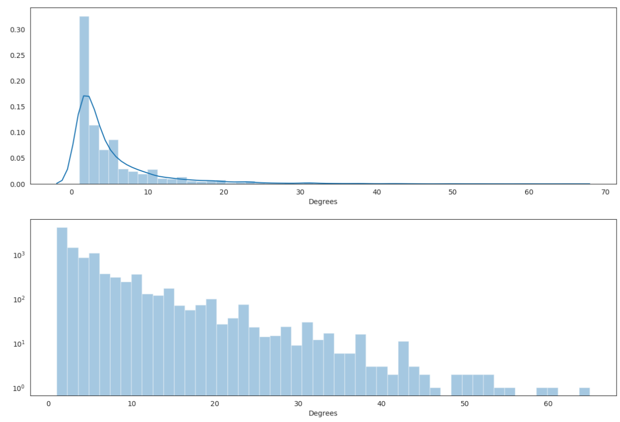

Figure 1: Distribution of the degrees

3 Question 3

Each of the eigenpair of the laplacian is associated to a "component" of the graph. The more connected it is, the smaller the eigenvalue will be. For instance, if the graph has two connected components, the laplacian will have two eigenvectors associated to the eigenvalue 0. This generalises to other graphs where low eigenvalues will highlight clusters in the data.

The eigenvalue decomposition solves a min cut problem (or a variant such as ratio cut or ncut).

$$\min cut(A, B) = \sum_{i \in A, j \in B} w_{ij}$$

4 Question 4

$m = 10$, $n_c = 3$
Green community : $\frac{1}{10} - \left(\frac{2}{20}\right)^2$
Blue community : $\frac{3}{10} - \left(\frac{7}{20}\right)^2$
Grey community : $\frac{5}{10} - \left(\frac{11}{20}\right)^2$

$Q = \frac{9}{10} - \frac{141}{400} = 0.465$

5 Question 5

For any graph G, any graph G' which is constructed by adding new nodes of degree 0 will have the same representation.

6 Question 6

Shortest path gives a perfect accuracy whereas graphlet struggles to beat the random baseline. In a cycle of size n, the longest shortest path will be of size $\frac{n}{2}$ whereas it will be of size n for a path, therefore these features are adapted to these graphs. However, there is almost no way for the graphlet kernel to differentiate a path from a cycle, which explains the poor performance. Here, the graphlet kernel samples subgraphs of size 3, and loses the global information of the graph (and especially the difference between the cycle and the path). Therefore, it does not work here.