Exercise 5 – Triangles and Clustering Coefficient

①	②	∑
6	4	10

Preliminaries

In this exercise[1], we will implement triangle counting and clustering coefficient on some real networks. In addition to the normal tasks (Task 1 and 2), you may get 3 points from accomplishing the extra task (Task 3). However you cannot get more than 10 points in total. Please submit your solution before July 10, 2017 – 23:59.

1 Triangle Counting (6 Points)

Please write a function that computes the number of triangles in an undirected network. Apply this function to the following three networks from KONECT.

- Jazz musicians network[2].
- Reactome network[3].
- Amazon (MDS) network[4].

Note your algorithm should remove self-loops first, and should be efficient enough to handle large networks. We give 2 points for each network, if your algorithm gives the *exact* result within 3 minutes. Compare your result with the statistics given on the webpages. Please submit both results and related code.

2 Clustering Coefficient (4 Points)

In an undirected network, the *local clustering coefficient* $c(u)$ of a node u is defined as the probability that two randomly picked neighbours of u are also connected.

Using a subset of the Facebook social network[5], compute the local clustering coefficient for all users.

The *global clustering coefficient* for the whole network can be defined in two ways:

- As the average over all nodes' local clustering coefficient.
- As the probability that two incident edges are completed by a third edge to form a triangle.

Compute both types of global clustering coefficient. Are they equal? Why?

[1] This exercise is based on the exercise by Jérôme Kunegis.
[2] `http://konect.uni-koblenz.de/networks/arenas-jazz`
[3] `http://konect.uni-koblenz.de/networks/reactome`
[4] `http://konect.uni-koblenz.de/networks/com-amazon`
[5] `http://konect.uni-koblenz.de/networks/facebook-wosn-links`

3 Random Graphs (Extra)

In the Erdős–Rényi random graph model, an undirected simple graph $G(n, m)$ is constructed by connecting n nodes with m edges randomly. The probability that an edge exists between a pair of nodes is equal, independent from every other edge.

- Consider only the global clustering coefficient defined in the second way in Task 2. What is the expected value of the global clustering coefficient C for $G(n, m)$? Please give the expression of C in terms of n and m.
- Please give the degree distribution $P(d)$ for $G(n, m)$.
  Hint: what is the probability that a randomly picked node has degree d?