

Heidelberg University
Institute of Computer Science
Database Systems Research Group

Lecture: Complex Network Analysis

Prof. Dr. Michael Gertz

Assignment 1
Graph Theory and Networks in Python

https://github.com/nilskre/CNA_assignments

Team Member: Patrick Günther, 3660886,
Applied Computer Science
rh269@stud.uni-heidelberg.de

Team Member: Felix Hausberger, 3661293,
Applied Computer Science
eb260@stud.uni-heidelberg.de

Team Member: Nils Krehl, 3664130,
Applied Computer Science
pu268@stud.uni-heidelberg.de

1 Problem 3-1 Clustering Coefficients

We now look at random networks in the $G(N, p)$ ensemble. Assume the link probability $p = \frac{a}{N^z}$, with $a \geq 0$ and $z \geq 0$, and a and z are independent of N .

1. Given an undirected complete graph with N nodes, calculate the total number of triangles. Note that the order of nodes matters, e.g., $A \rightarrow B \rightarrow C$ is not the same as $C \rightarrow B \rightarrow A$.

First one needs to find the number of unique triangles in the complete graph, which is calculated by $\binom{N}{3}$. For the calculation of the global clustering coefficient, this is the amount of triangles to be used and matches the post on Moodle from Shideh Almasian:

*I delete the post and the comments because my answer was wrong. So the correct answer is that "the order of nodes does **not** matter", ABC is a single triangle and that counts towards the global clustering coefficient.*

In case one additionally wants to define the total number of triangles by the order of nodes, one has multiply with $3!$. So the total number of triangles in a complete graph asked by the task description is $3! \cdot \binom{N}{3}$.

2. Can the number of triangles in a graph be larger than the number of edges? Are you able to find a graph with more triangles than edges? If so, draw such a graph.

We use the stricter definition of the number of triangles in a graph that does not consider the order of nodes. We again use the example of a complete graph. The number of edges in such a graph is defined by $L = \binom{N}{2} = \frac{N(N-1)}{2}$. $\binom{N}{3} > \binom{N}{2}$ holds for $N = 6$. See the following graph in Figure 1:

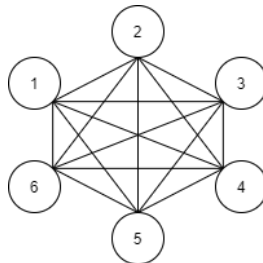


Figure 1: A complete graph with $N = 6$.

This means the number of triangles in a graph can be larger than the number of edges.

3. Prove that the number of connected triples in an undirected graph is equal to half the **sum** of non-zero off-diagonal entries of the adjacency matrix A^2 , where A denotes the adjacency matrix.

Let (n_i, n_j, n_k) be a connected triple in a graph with $N \geq 3$. This means that there is a path from n_i to n_k of length 2. We know that for such a path must hold $A_{ij}A_{jk} = 1$. Therefore, A^2 holds paths of length 2, i.e. connected triples in its non-zero off-diagonal entries. As an illustration, consider the following simple graph in Figure 2 and the following Equation 3:

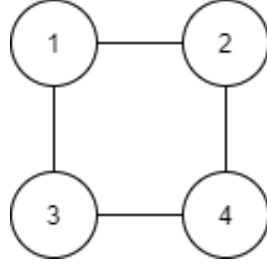


Figure 2: A simple graph with 4 connected triples.

$$\begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 & 2 \\ 0 & 2 & 2 & 0 \\ 0 & 2 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{pmatrix}$$

Half the sum of all non-zero off-diagonal entries of the adjacency matrix A^2 equals exactly the number of connected triples, namely 4.

4. Given a graph with its adjacency matrix A , find an expression for the global clustering coefficient in terms of the elements of the adjacency matrix. Notice that the number of triangles in a graph is **related** to the trace of the matrix A^3 .

We know

$$C = \frac{3 \cdot \#triangles}{\#connected\ triples}.$$

The also know that the relation between the number of triangles $n_{triangles}$ and the trace of A^3 is exactly $6 \cdot n_{triangles} = tr(A^3)$. Therefore it follows:

$$C = \frac{\text{tr}(A^3)}{\sum_{i \neq j} A_{ij}^2}$$

2 Problem 3-2 Parametrized Random Networks

We now look at random networks in the $G(N, p)$ ensemble. Assume the link probability $p = \frac{a}{N^z}$, with $a > 0$ and $z \geq 0$, and a and z are independent of N .

1. Determine the *average degree* $\langle k \rangle$ in the limit $N \rightarrow \infty$ for the following values of the parameters a and z :

We use the equation $\langle k \rangle = p(N - 1)$.

- $a = 0.5, z = 1$

$$\langle k \rangle = \lim_{N \rightarrow \infty} \frac{0.5}{N} (N - 1) = \lim_{N \rightarrow \infty} 0.5 - \frac{0.5}{N} = 0.5 \quad (1)$$

$\langle k \rangle = 0.5$ and $\langle k \rangle < 1$, which means there **is no GC**.

- $a = 2, z = 1$

$$\langle k \rangle = \lim_{N \rightarrow \infty} \frac{2}{N} (N - 1) = \lim_{N \rightarrow \infty} 2 - \frac{2}{N} = 2 \quad (2)$$

$\langle k \rangle = 2$ and $\langle k \rangle > 1$, which means there **is a GC**.

- $a > 0, z = 2$

$$\langle k \rangle = \lim_{N \rightarrow \infty} \frac{a}{N^2} (N - 1) = \lim_{N \rightarrow \infty} \frac{a}{N} - \frac{a}{N^2} = 0 \quad (3)$$

$\langle k \rangle = 0$ and $\langle k \rangle < 1$, which means there **is no GC**.

- $a > 0, z = 0.5$

$$\langle k \rangle = \lim_{N \rightarrow \infty} \frac{a}{\sqrt{N}} (N - 1) = \lim_{N \rightarrow \infty} a\sqrt{N} - \frac{a}{\sqrt{N}} = \infty \quad (4)$$

$\langle k \rangle \rightarrow \infty$ and $\langle k \rangle > 1$, which means there **is a GC**.

In which of the above cases does the random network contain a giant component in the limit $N \rightarrow \infty$?

See above.

2. Given $G(N, p)$ with the link probability $p = \frac{a}{N^z}$ and the values of $a > 0$ and $z \geq 0$. Determine the average degree $\langle k \rangle$ in the limit $N \rightarrow \infty$.

Again with $\langle k \rangle = p(N - 1)$:

$$\langle k \rangle = \lim_{N \rightarrow \infty} \frac{a}{N^z} (N - 1) = \lim_{N \rightarrow \infty} aN^{1-z} - \frac{a}{N^z} \quad (5)$$

This means:

$$\langle k \rangle = \begin{cases} \infty, & \text{if } z < 1 \\ 0, & \text{if } z \geq 1 \end{cases}$$

3. Determine the conditions on a and z for which these random networks are critical, again in the limit $N \rightarrow \infty$.

Networks are critical if $\langle k \rangle = 1$.

We have this equation from the previous subtask:

$$\langle k \rangle = \lim_{N \rightarrow \infty} aN^{1-z} - \frac{a}{N^z} \quad (6)$$

Since here $1 - z$ must be 0 and aN^0 must be 1, the conditions are: z must be 1 and a must be 1.

3 Problem 3-3 Differences between real and random networks

See *problem_3.ipynb*.