Heidelberg University
Institute of Computer Science
Database Systems Research Group

Lecture: Complex Network Analysis

Prof. Dr. Michael Gertz

Assignment 3
Random Graph Models and Statistical Characterizations

`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

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

   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:

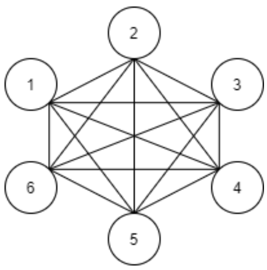

Figure 1: A complete graph with $N = 6$.

   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 is 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:

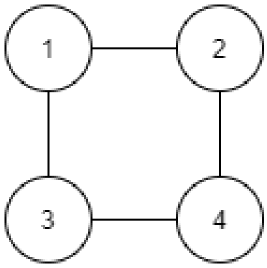

Figure 2: A simple graph with 4 connected triples.

$$A * A = A^2 :$$

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

   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.

$$\#connected\ triples = \frac{\sum_{i \neq j} A^2_{ij}}{2} = \frac{2+2+2+2}{2} = 4 \tag{1}$$

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 (trace = sum of the matrix diagonal elements) of the matrix A^3.

   We know

$$C = \frac{3 \cdot \#triangles}{\#connected\ triples}. \tag{2}$$

The $\#triangles$ is related to the trace of A^3.

- $A =$ adjacency matrix: Which nodes can be reached within distance 1?
- $A^2 =$ Which nodes can be reached within distance 2?
- $A^3 =$ Which nodes can be reached within distance 3?

If one node can reach itself within distance 3 it forms a triangle. Therefore we look at the diagonal of the A^3 matrix (trace). Each triangle is counted three times (each node of the triangle = one count). That is why the result is divided by 3. Furthermore the graph is undirected and every triangle is counted twice, resulting in the division by 2. A triangle begins and ends at the same node and consists of three edges. Finally we receive the following equation:

$$\#triangles = \frac{tr(A^3)}{2 * 3} = \frac{tr(A^3)}{6} \tag{3}$$

The $\#connected\ triples$ is known from equation 1. Now we insert $\#connected\ triples$ and $\#triangles$ into the given equation 2:

$$C = \frac{3 \cdot \#triangles}{\#connected\ triples} = \frac{3 \cdot \frac{tr(A^3)}{6}}{\frac{\sum_{i \neq j} A_{ij}^2}{2}} = \frac{1 \cdot \frac{tr(A^3)}{2}}{\frac{\sum_{i \neq j} A_{ij}^2}{2}} = \frac{tr(A^3)}{2 \cdot \frac{\sum_{i \neq j} A_{ij}^2}{2}} = \frac{tr(A^3)}{\sum_{i \neq j} A_{ij}^2}$$

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 $N \to \infty$ for the following values of the parameters a and z:

   We use the equation $\langle k \rangle = p(N-1)$. Then we insert the given equation for p $\langle k \rangle = \frac{a}{N^z}(N-1)$. Now the different values for a and z can be inserted and their behavior for $N \to \infty$ can be observed:

- $a = 0.5,\ z = 1$

  $$\langle k\rangle = \lim_{N\to\infty} \frac{0.5}{N}(N-1) = \lim_{N\to\infty} \frac{0.5N - 0,5}{N} = \lim_{N\to\infty} 0.5 - \frac{0.5}{N} = 0.5 \tag{4}$$

  $\langle k\rangle = 0.5$ and $\langle k\rangle < 1$, which means there **is no GC**.
- $a = 2,\ z = 1$

  $$\langle k\rangle = \lim_{N\to\infty} \frac{2}{N}(N-1) = \lim_{N\to\infty} \frac{2N-2}{N} = \lim_{N\to\infty} 2 - \frac{2}{N} = 2 \tag{5}$$

  $\langle k\rangle = 2$ and $\langle k\rangle > 1$, which means there **is a GC**.
- $a > 0,\ z = 2$

  $$\langle k\rangle = \lim_{N\to\infty} \frac{a}{N^2}(N-1) = \lim_{N\to\infty} \frac{aN-a}{N^2} = \lim_{N\to\infty} \frac{a}{N} - \frac{a}{N^2} = 0 \tag{6}$$

  $\langle k\rangle = 0$ and $\langle k\rangle < 1$, which means there **is no GC**.
- $a > 0,\ z = 0.5$

  $$\begin{aligned}\langle k\rangle = \lim_{N\to\infty} \frac{a}{N^{\frac{1}{2}}}(N-1) = \lim_{N\to\infty} \frac{aN-a}{N^{\frac{1}{2}}} &= \lim_{N\to\infty} aN^{1-\frac{1}{2}} - \frac{a}{\sqrt{N}} \\ &= \lim_{N\to\infty} a\sqrt{N} - \frac{a}{\sqrt{N}} = \infty\end{aligned} \tag{7}$$

  $\langle k\rangle \to \infty$ 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 $N \to \infty$?

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 $N \to \infty$.

   Again with $\langle k\rangle = p(N-1) = \frac{a}{N^z}(N-1)$:

   $$\langle k\rangle = \lim_{N\to\infty} \frac{a}{N^z}(N-1) = \lim_{N\to\infty} \frac{aN-a}{N^z} = \lim_{N\to\infty} aN^{1-z} - \frac{a}{N^z} \tag{8}$$

   This means:

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

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

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

   We have this equation from the previous subtask:

$$\langle k \rangle = \lim_{N \to \infty} aN^{1-z} - \frac{a}{N^z} \tag{9}$$

   For $N \to \infty$ the right term $\frac{a}{N^z}$ becomes irrelevant. To get a critical random network the first term needs to be 1. 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*.