Heidelberg University
Institute of Computer Science
Database Systems Research Group

Lecture: Complex Network Analysis

Prof. Dr. Michael Gertz

Assignment 2

Graph Properties and Random Graphs

`https://github.com/nilskre/CNA_assignments`

Team Member: Patrick Guenther, 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 2-1 Erdos-Renyi Network

Consider an Erdos-Renyi network with N = 80 nodes, connected to each other with probability p = 0.05.

1. What is (i) the expected number of links in the graph and (ii) the expected degree of a node?

   (i) The expected number of links in the graph:

   $$\langle L \rangle = p\frac{N(N-1)}{2} = 0.05\frac{80 \cdot (80-1)}{2} = 158 \tag{1}$$

   The expected number of links in the graph is 158.

   (ii) The expected degree of a node:

   $$\langle k \rangle = p(N-1) = 0.05 \cdot (80-1) = 3.95 \tag{2}$$

   The expected degree of a node in the network is 3.95.

2. In which regime is the network?

   $\langle k \rangle$ is 3.95, thus greater than 1 and not in the subcritical regime.

   $\langle k \rangle < ln(N)$ since $3.95 < 4.38$, thus it is not in the connected regime.

   This means that the network is in the **supercritical regime**.

3. What is the probability to find exactly $L = 200$ links in the graph?

   $$p_L = \binom{\frac{N(N-1)}{2}}{L} p^L (1-p)^{(N(N-1)/2)-L} \tag{3}$$

   $$p_{200} = \binom{\frac{80(80-1)}{2}}{200} 0.05^{200} (1-0.05)^{(80(80-1)/2)-200} \approx 1.26e^{-4} \tag{4}$$

   The probability to find exactly 200 links in the graph is arround $1.26e^{-4}$.

4. What is the probability that a node i in the graph has degree $k_i = 5$ (using the binomial distribution)?

   $$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k} \tag{5}$$

Node	Degree
1	3
2	2
3	2
4	1
5	5
6	1
7	1
8	3
9	1
10	1
11	1
12	1

Table 1

$$p_5 = \binom{80-1}{5} 0.05^5 (1-0.05)^{80-1-5} \approx 0.158 \tag{6}$$

The probability of a node i having a degree of 5 is 0.158.

5. Use maximum likelihood estimation to estimate the model parameters (N, p) for the shown graph.

   From Table 1 follows that the average degree of the network is $\langle k \rangle \approx 0.54$.

$$\langle k \rangle = p(N-1) \tag{7}$$

$$p = \frac{\langle k \rangle}{N-1} = \frac{0.54}{12-1} \approx 0.049 \tag{8}$$

   The model parameters are $N = 12$ and $p = 0.049$.

   Alternative:

   $p(G)$ is maximized for

$$p = \frac{m}{n(n-1)} \tag{9}$$

where m is the number of edges and n is the number of vertices in the graph.

$$p = \frac{11}{12(12-1)} \approx 0.083 \tag{10}$$

Following this method, p would be 0.083.

2 Problem 2-2 Three-Dimensional Lattice Network

Consider an undirected network $G = (V, E)$ with $N = l^3$ nodes corresponding to points on a regular three-dimensional lattice $\{1, ..., l\}x\{1, ..., l\}x\{1, ..., l\}$. Two nodes x and y are connected if and only if $d(x, y) = 1$, where $d(x, y)$ denotes the Euclidean distance. A visualization of such a graph is shown in the following image:

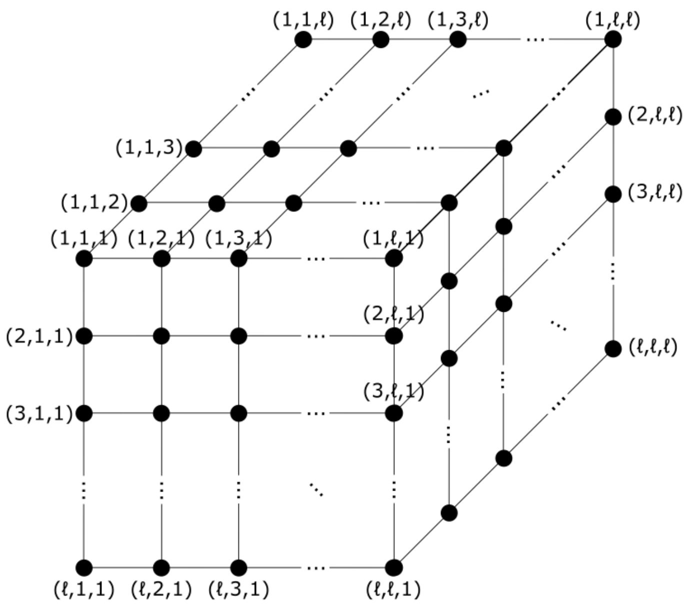

Figure 1: Three-Dimensional Lattice Network

1. What is the diameter d_{max} of the graph?

   The diameter is the longest shortest path in the graph. One example for a longest shortest path is the distance from the left upper edge in the front (1,1,1) to the right lower edge in the back (l,l,l). The diameter is therefore $d_{max} = l^3$.

2. Provide an expression (in terms of N and/or l) for the probability p_i that a randomly chosen node has degree i. You may assume that $l >= 3$. What are the consequences for $N- > infinity$?

   Due to the fact that there are three different types of nodes (inner nodes, corner nodes and border nodes) with differing degree three different expressions are provided in the following.

   First the distribution of the different node types is clarified:

   - inner node: 6 edges; $(l-2)^2$ nodes
   - corner node: 3 edges; 8 nodes
   - border node: 4 edges; $6*l^2-8-6*l = 6*l-8$ nodes (nodes on all six sides: $6*l^2$; substract all edge nodes: -8; count each outer edge only once: $-6*l$)

   Secondly the expressions for the probability distribution are derived based on the following formula with $N = l^3$:

$$p_k = \binom{N-1}{k} * p^k * (1-p)^{N-1-k} \tag{11}$$

   For inner nodes with $i = 6$:

$$p_6 = \binom{l^3-1}{6} * (\frac{(l-2)^2}{l^3})^6 * (1 - \frac{(l-2)^2}{l^3})^{l^3-1-6} \tag{12}$$

   For corner nodes with $i = 3$:

$$p_3 = \binom{l^3-1}{3} * (\frac{8}{l^3})^3 * (1 - \frac{8}{l^3})^{l^3-1-3} \tag{13}$$

   For border nodes with $i = 4$:

$$p_4 = \binom{l^3-1}{4} * (\frac{6l-8}{l^3})^4 * (1 - \frac{6l-8}{l^3})^{l^3-1-4} \tag{14}$$

   For $N- > infinity$ also $l- > infinity$, that is why the probability for an inner node increases and the probability for a corner or a border node decreases.

3. What is (i) the clustering coefficient of a node i and (ii) the average clustering coefficient of this network?

(i) the clustering coefficient of a node i

The clustering coefficient is defined as:

$$C_i = \frac{2 * L_i}{k_i * (k_i - 1)} \tag{15}$$

with L_i being the number of edges between the neighbors of the node and k_i being the number of edges from the chosen node.

The clustering coefficient is 0 for all types of nodes, because there are no edges (diagonal edges between the nodes) between the neighbors of the node ($L_i = 0$).

(ii) the average clustering coefficient of this network

The average clustering coefficient is defined as:

$$< C >= \frac{1}{N} \sum_{i=1}^{N} C_i \tag{16}$$

As described in (i) the clustering coefficient is zero for all nodes. Consequently the average clustering coefficient is also zero.

4. Now assume we have a different three-dimensional lattice graph with nodes corresponding to points $\{1, ..., l\}x\{1, ..., l\}x\{1, ..., l\}$, where two nodes x and y are connected if and only if $d(x, y) <= 3$. How does this change the average clustering coefficient in the limit $N- > infinity$?

   Now diagonal edges between the nodes are present. That is why the clustering coefficient is not zero any more.

   $For N- > infinity < C > - > 0,5$