CS-E5740 - Complex Networks
Exercise set 1

Hugues Verlin (584788)
hugues.verlin@aalto.fi

November 11, 2016

1 Basic network properties

a) The adjacency matrix A It is a $n - by - n$ boolean (or integer) matrix, (with n, the number of nodes) where a *true* value (or a 1) at $A[i][j]$ indicates an edge from node i to node j.
For the example, we have :

$$A = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

b) The edge density ρ of the graph The edge density of a network is the fraction of edges out of all possible edges. This is defined as (with n, number of nodes in the graph, and m, number of edges in the graph):

$$\rho = \frac{m}{\binom{n}{2}} = \frac{2m}{n\,(n-1)}$$

For the example, we have:

$$\rho = \frac{2 \times 9}{8 \times (8-1)} \approx 0.321$$

c) The degree k_i of each node $i \in V$ and the degree distribution P(k) The degree k_i of each node is the number of links that each node has. In the example, we have:

$$\begin{array}{ccc} k_1 = 1 & k_2 = 1 & k_3 = 2 \\ k_4 = 5 & k_5 = 3 & k_6 = 3 \\ & k_7 = 2 & k_8 = 1 \end{array}$$

The degree distribution $P(k)$ is the probability that the degree of a node picked at random is k. It is defined as:

$$P(k) = {}^{n_k}/_{n} \quad \text{with } n_k \text{, number of nodes of degree k}$$

In the example, we have:

$$\forall j \in \{0; 4\} \cup [6, +\infty[, \; P(j) = 0 \quad P(1) = {}^{3}/_{8} \quad P(2) = {}^{1}/_{4}$$
$$P(3) = {}^{1}/_{4} \quad P(5) = {}^{1}/_{8}$$

d) The mean degree $\langle k \rangle$ of the graph The mean degree is the average degree of the graph. Hence, it is define as:

$$\langle k \rangle = \frac{1}{N} \sum_i d_i$$

where N is the number of vertices, and d_i is the degree of the vertice i.
In the example, we have:

$$\langle k \rangle = \frac{1}{8} (1 + 1 + 2 + 5 + 3 + 3 + 2 + 1) = \frac{9}{4}$$

e) The diameter d of the graph The diameter d is the largest distance in the graph:

$$d = \max_{i,j \in V} d_{i,j}$$

With the example, we have:

$$d = 4$$

f) Clustering coefficient and average clustering coefficient The clustering coefficient of a node v_i is the quotient between the number of edge between its neighbours and the number of possible edges between its neighbours.

$$C_i = \frac{E}{\binom{k_i}{2}} = \frac{2E}{k_i (k_i - 1)}$$

where E is the number of edges between v_i's k neighbours.
The average clustering coefficient is then:

$$\langle C \rangle = \frac{1}{n} \sum_i C_i$$

In the example, we have:

- For the clustering coefficients :
  - $C_1 = 0$
  - $C_2 = 0$
  - $C_3 = 1$
  - $C_4 = \frac{2 \times 2}{5(5-1)} = \frac{1}{5}$
  - $C_5 = \frac{2 \times 2}{3(3-1)} = \frac{2}{3}$
  - $C_6 = \frac{2 \times 1}{3(3-1)} = \frac{1}{3}$
  - $C_7 = 0$
  - $C_8 = 0$
- For the average: $\langle C \rangle = \frac{1}{n} \sum_i C_i = \frac{1}{8} \left(1 + \frac{1}{5} + \frac{2}{3} + \frac{1}{3}\right) = \frac{11}{40}$

2 Computing network properties programmatically

Please, see the code in the attached file.

2. a) Visualize the network

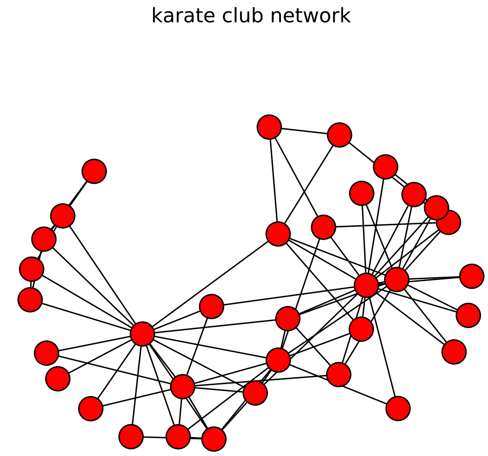

Figure 2.1: Karate club network

2. b) Calculate the edge density

The results are the same:

```
D from self-written algorithm: 0.13903743315508021
D from NetworkX function: 0.13903743315508021
```

2. c) Calculate the average clustering coefficient

The results are also the same:

```
C from self-written algorithm: 0.5706384782076824
C from NetworkX function: 0.5706384782076822
```

2. d) Calculate the degree distribution $P(k)$ and complementary cumulative degree distribution 1-CDF(k) of the network

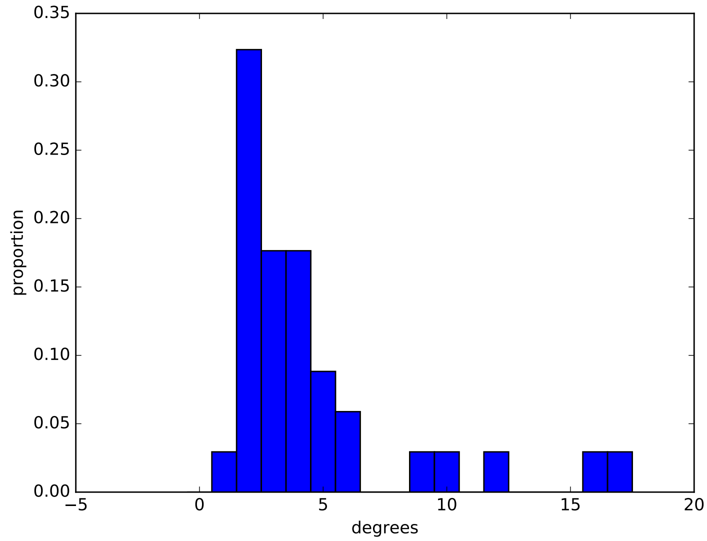

Figure 2.2: degree distribution $P(k)$

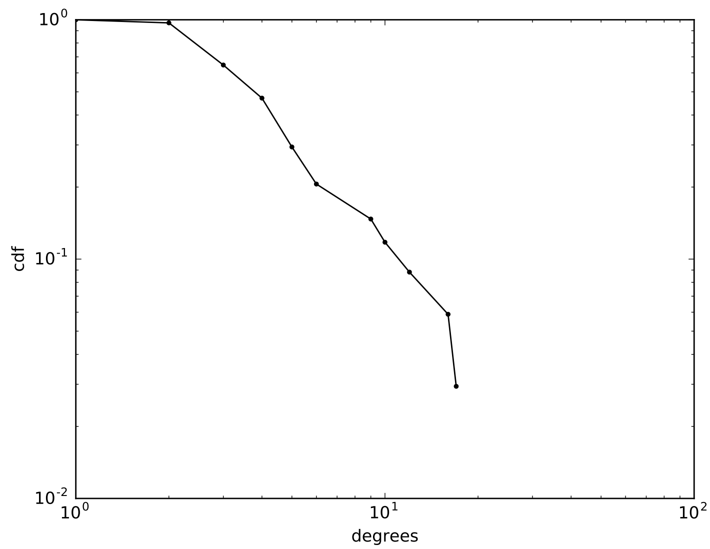

Figure 2.3: complementary cumulative degree distribution 1-CDF(k)

2. e) Calculate the average shortest path length $\langle l \rangle$

Here is the output of the program:

```
<l> from NetworkX function: 2.408199643493761
```

2. f) Create a scatter plot of C_i as a function of k_i

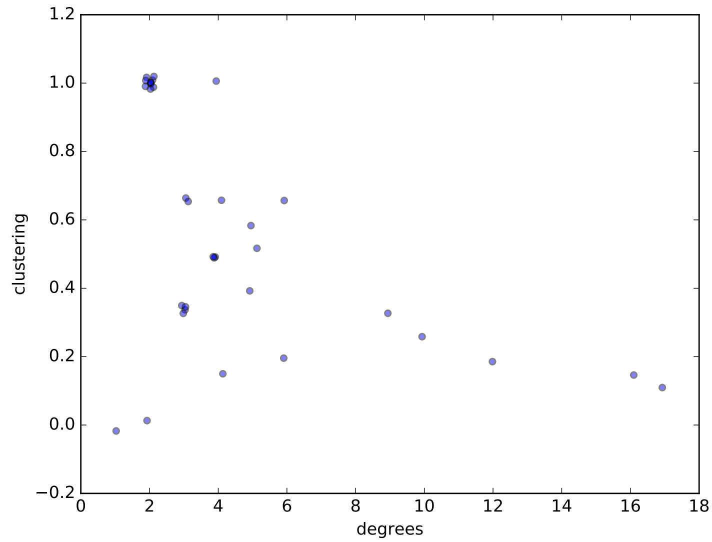

Figure 2.4: Scatter plot of C_i as a function of k_i

3 Path lengths in simple model networks

3. a) Ring lattice

For a ring lattice, the diameter d is equal to the number of node divided by 2 as the largest distance is the one from any node to diametrically opposite node.

$$d_{\text{ring lattice}} = \frac{N}{2}$$

3. b) Two-dimensional square lattice

For a two dimensional square lattice, the largest distance start from a corner and go to the complete opposite corner (up-left $\rightarrow$ bottom-right for example).

Then, the shortest from one corner to the other is for example along the side of the square. Then, as $N = L^2$, we have to travel $2 \times L$. Therefore,

$$d_{\text{square lattice}} = 2 \times L = 2\sqrt{N}$$

3. c) Cayley tree

- Number of nodes

For each added layer, it splits k times in $(k-1)$ nodes (so $k \times (k-1)^l$), where l is the number of layers. Thus, the number of nodes is:

$$N = (k-1)^0 + k\,(k-1)^1 + k\,(k-1)^2 + \cdots + k\,(k-1)^l = 1 + k\sum_{i=0}^{l-1}(k-1)^l = 1 + k\left(\frac{(k-1)^l - 1}{k-2}\right)$$

As we only consider *Cayley trees*, with $k = 3$, we have:

$$N = 1 + 3\left(\frac{2^l - 1}{2-1}\right) = 1 + 3\left(2^l - 1\right)$$

- **Diameter**

If l is strictly greater than 1, it is very to see that the largest distance is increased by 2 each time we add a new layer to the tree. It follows that:

$$d = 2 \times l$$

We can then express l in term of N:

$$l = \log_2 \left(\frac{N-1}{3} + 1 \right)$$

Therefore,

$$d_{\text{cayley tree}} = 2 \times \left(\log_2 \left(\frac{N-1}{3} + 1 \right) \right)$$

3. d) Analysis

- **If N is increased, which network's diameter grows fastest?**

We have

$$\lim_{N \to +\infty} \frac{N/2}{2\sqrt{N}} = \lim_{N \to +\infty} \sqrt{N} = +\infty$$

$$\lim_{N \to +\infty} \frac{N/2}{2\left(\log_2\left(\frac{N-1}{3}+1\right)\right)} = \lim_{N \to +\infty} \frac{N}{\log(N)} = +\infty$$

Therefore, the network with the fastest diameter grow is the *ring lattice*.

- **And slowest?**

We also have:

$$\lim_{N \to +\infty} \frac{N/2}{2\sqrt{N}} = \lim_{n \to +\infty} \frac{\log N}{\sqrt{N}} = 0$$

Therefore, the slowest diameter grow belongs to the *cayley tree*.

- **Which of these networks fulfill the 'small-world' property?**

For the *cayley tree,* we have

$$\lim_{N \to +\infty} \frac{2\left(\log_2\left(\frac{N-1}{3}+1\right)\right)}{\log N} = \lim_{N \to +\infty} \frac{2\left(\log_2(N) - \log_2 3\right)}{\log N} = \lim_{N \to +\infty} \frac{2 \times \log(N)}{\log 2 \times \log N} = \frac{2}{\log 2}$$

As $\frac{2}{\log 2}$ is a constant, we can conclude that :

$$d_{\text{cayley tree}}(N) = \Theta\left(d_{\text{small world}}(N)\right)$$

This shows that the *cayley tree* fullfill the "small world" property.

4 Counting number of walks using the adjacency matrix

4. a) Draw the induced subgraph G^*

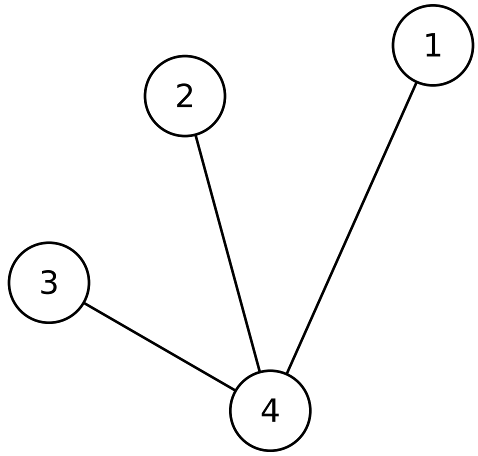

Figure 4.1: Induced graph G^*

4. b) Compute the number walks of length 2

- 1 to 1: 1
- 1 to 2: 1
- 1 to 3: 1
- 1 to 4: 0
- 2 to 2: 1
- 2 to 3: 1
- 2 to 4: 0
- 3 to 3: 1
- 3 to 4: 0
- 4 to 4: 3

4. c) Compute the matrix A^2, what can you notice?

$$A = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \qquad A^2 = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

We can notice that the numbers of walks are similar to the numbers in the matrice. For example, we have $1 \to 2 = 1$ and $A^2_{1,2} = 1$ or $1 \to 4 = 0$ and $A^2_{1,4} = 0$.

4. d) Compute the number of walks of length three from node 3 to node 4 in G^*

From node 3 to node 4, there is 3 walks.

$$A^3_{3,4} = A_{3,1} \times A^2_{1,4} + A_{3,2} \times A^2_{4,2} + A_{3,3} \times A^2_{4,3} + A_{3,4} \times A^2_{4,4}$$
$$A^3_{3,4} = 0 \times 1 + 0 \times 1 + 0 \times 1 + 1 \times 3$$

We get the same number using both ways.

4. e) Show that the element $A^m_{i,j}$, $m \in N$ corresponds to the number of walks of length m between nodes i and j

Proof by induction:

- Let's consider the case $m = 1$.
  $A_{i,j}$ indicates if there is a path between node i and j. Therefore, it corresponds to the number of walk of length $m = 1$.

- Let's now suppose that the property is true for all $m = n$, $n \in N$.

Using the hypothesis, we know that $a_{ij}^{(m)}$ — the $ij : th$ entry of A^m — is the number of walks of length m from any node v_i to v_j .

By definition, we have $a_{ij}^{(m+1)} = a_{i1}a_{1j}^{(m)} + a_{i2}a_{2j}^{(m)} + \cdots + a_{in}a_{2n}^{(m)} = \sum_{m=1}^{n} a_{im}b_{mj}$.

Then:
$a_{i1}a_{1j}^{(m)}$ is equal to the number of walks of length m from v_1 to v_j times the number of walks of length 1 from v_i to v_1. It is also the number of walks of length $m+1$ from v_i to v_j , where v_1 is the second vertex.

This argument holds for each $k \in \mathbb{N}$. Indeed, $a_{it}a_{tj}^{(m)}$ is the number of walks from v_i to v_j in which v_k is the second vertex. Therefore, the sum is the number of all possible walks from v_i to v_j. □

5 Bipartite networks

5. a) Construct the two unipartite projections of the network

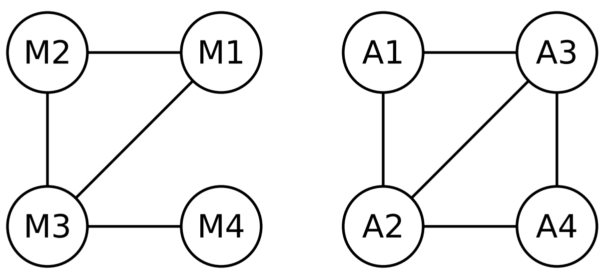

Figure 5.1: Movies and actors unipartite graphs

5. b) Show that, in general, it is not possible to uniquely reconstruct a bipartite network from its two unipartite projections

Here is a counter example constructed from the two unipartite graphs:

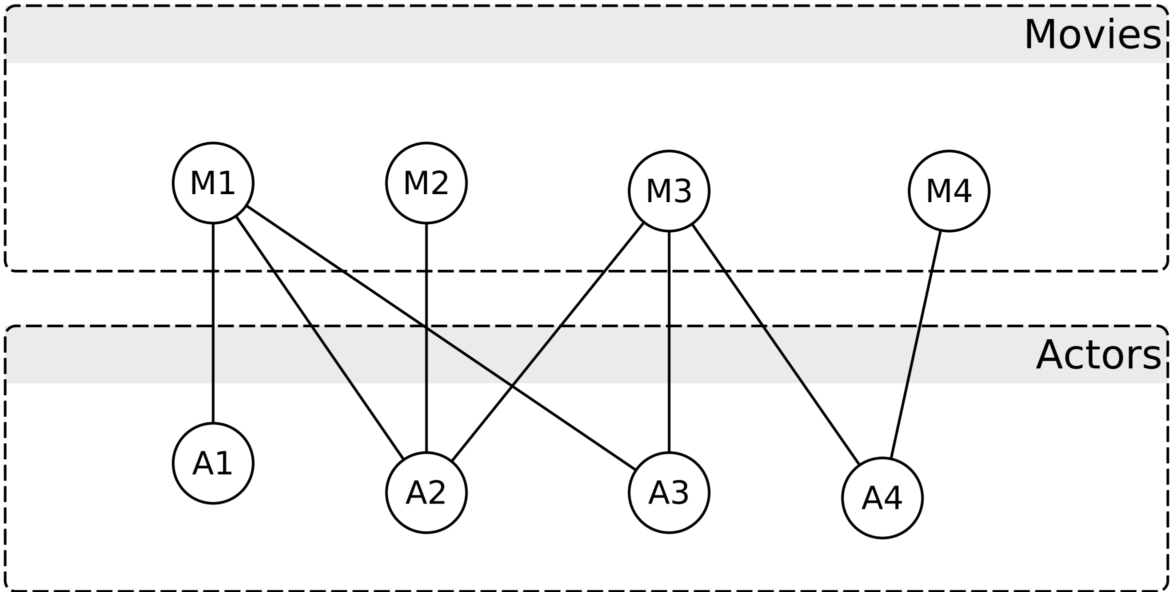

Figure 5.2: Counter example — other bipartite network built from the two unipartite graphs

6 Ensemble averages by enumeration

6. a) Calculate, using pen and paper, $\langle k\rangle$, $\langle C\rangle$, and $\langle d^*\rangle$ for $G(N=3, p=1/3)$

For $N=3$, we have 8 possible graphs:

- – 1 graph with no links
  – 3 graphs with 1 link
  – 3 graphs with 2 links
  – 1 graph with 3 links

$$\begin{aligned}\langle k\rangle &= \sum_{i=0}^{3} \pi_i k\left(G_i\right)\\ &= \left(1-\frac{1}{3}\right)^3 k\left(G_0\right) + 3\times\frac{1}{3}\times\left(1-\frac{1}{3}\right)^2 k\left(G_1\right) + 3\times\left(\frac{1}{3}\right)^2\times\left(1-\frac{1}{3}\right)\times k\left(G_2\right) + \left(\frac{1}{3}\right)^3 k\left(G_3\right)\\ &= \frac{8}{27}\times 0 + 3\times\frac{4}{27}\times\frac{2}{3} + 3\times\frac{2}{27}\times\frac{4}{3} + \frac{1}{27}\times\frac{6}{3}\\ &= \frac{2}{3}\end{aligned}$$

$$\begin{aligned}\langle C\rangle &= \sum_{i=0}^{3} \pi_i C\left(G_i\right)\\ &= \frac{8}{27}\times C\left(G_0\right) + \frac{12}{27}\times C\left(G_1\right) + \frac{6}{27}\times C\left(G_2\right) + \frac{1}{27}\times C\left(G_3\right)\\ &= 0+0+0+1\times\frac{1}{27}\\ &= \frac{1}{27}\end{aligned}$$

$$
\begin{aligned}
\langle d^* \rangle &= \sum_{i=0}^{3} \pi_i d^* (G_i) \\
&= \frac{8}{27} \times d^* (G_0) + \frac{12}{27} \times d^* (G_1) + \frac{6}{27} \times d^* (G_2) + \frac{1}{27} \times d^* (G_3) \\
&= 0 + \frac{12}{27} \times 1 + \frac{6}{27} \times 2 + \frac{1}{27} \times 2 \\
&= \frac{25}{27} \approx 0,9259
\end{aligned}
$$

6. b) Calculate, using pen and paper, $\langle k \rangle$, $\langle C \rangle$, and $\langle d^* \rangle$ for $G(N = 3, p)$

$$
\begin{aligned}
\langle k \rangle &= \sum_{i=0}^{3} \pi_i k (G_i) \\
&= (1-p)^3 k (G_0) + 3p (1-p)^2 k (G_1) + 3p^2 (1-p) k (G_2) + p^3 k (G_3) \\
&= 2p
\end{aligned}
$$

$$
\begin{aligned}
\langle C \rangle &= \sum_{i=0}^{3} \pi_i C (G_i) \\
&= (1-p)^3 C (G_0) + 3p (1-p)^2 C (G_1) + 3p^2 (1-p) C (G_2) + p^3 C (G_3) \\
&= p^3
\end{aligned}
$$

$$
\begin{aligned}
\langle d^* \rangle &= \sum_{i=0}^{3} \pi_i d^* (G_i) \\
&= (1-p)^3 d^* (G_0) + 3p (1-p)^2 d^* (G_1) + 3p^2 (1-p) d^* (G_2) + p^3 d^* (G_3) \\
&= -2p^3 + 3p
\end{aligned}
$$