Web Structure Mining – Analysing Networks

Exercise sheet

Keti Korini & Alexander Brinkmann

Networks with Undirected Edges

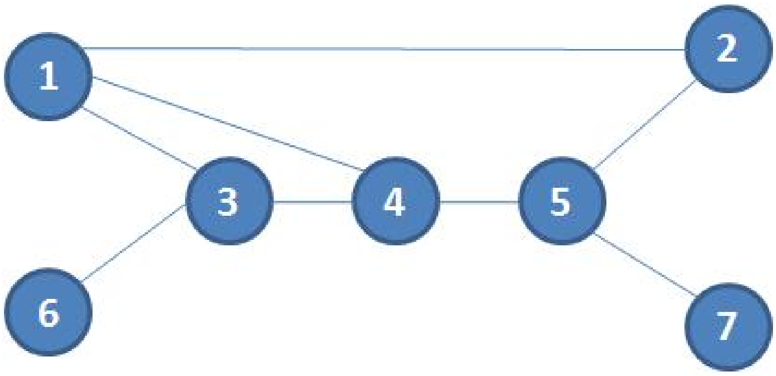

Figure 1: Undirected Graph.

Task 1: Adjacency Matrix

Represent the undirected graph displayed in Figure 1 using an adjacency matrix.

Solution:

	1	2	3	4	5	6	7
1	0	1	1	1	0	0	0
2	1	0	0	0	1	0	0
3	1	0	0	1	0	1	0
4	1	0	1	0	1	0	0
5	0	1	0	1	0	0	1
6	0	0	1	0	0	0	0
7	0	0	0	0	1	0	0

Task 2: Density

Compute the graph's density.

Solution:

$$density(G) = \frac{\sum_{i=1}^{|V|}\sum_{j=1,i\neq j}^{|V|} a_{ij}}{|V|(|V|-1)} = \frac{16}{7\cdot 6} = \frac{16}{42} = 0.38$$

Task 3: Centrality

We are interested in the actor (vertex) of the undirected graph with the highest **centrality** value. Use the following three measures to calculate the centrality values of the vertices. Don't forget to **normalize the results** whenever possible/needed.

Degree

Closeness The distance *dist* is defined as the number of edges on the shortest path between two vertices (e.g. $dist(vertice_1, vertice_7) = 3$)

Betweenness Start with listing all possible shortest paths, then compute the betweenness.

Solution:

Degree

	1	2	3	4	5	6	7	Degree
1	0	1	1	1	0	0	0	$\frac{1}{2}$
2	1	0	0	0	1	0	0	$\frac{1}{3}$
3	1	0	0	1	0	1	0	$\frac{1}{2}$
4	1	0	1	0	1	0	0	$\frac{1}{2}$
5	0	1	0	1	0	0	1	$\frac{1}{2}$
6	0	0	1	0	0	0	0	$\frac{1}{6}$
7	0	0	0	0	1	0	0	$\frac{1}{6}$

Closeness

a. Formulas:

$$\text{closeness: } C_c(v) = \left(\sum_{i=1,i\neq v}^{|V|} d(i,v)\right)^{-1}$$
$$\text{normalization: } C_c'(v) = \frac{|V|-1}{\sum_{i=1,i\neq v}^{|V|} d(i,v)}$$

with d = length of shortest path

b. Computing closeness for a single node (v_4):

$$C'_c(v_4) = \frac{7-1}{\sum_{i=1,i\neq v_4}^{7} dist(i,v_4)}$$
$$= \frac{6}{d(v_1,v_4)+d(v_2,v_4)+d(v_3,v_4)+d(v_5,v_4)+d(v_6,v_4)+d(v_7,v_4)}$$
$$= \frac{6}{1+2+1+1+2+2} = \frac{2}{3}$$

c. Closeness of all nodes:

- $C_c(v_1) = 6\ /\ 10 = 0{,}6$
- $C_c(v_2) = 6\ /\ 11 = 0{,}54$
- $C_c(v_3) = 6\ /\ 10 = 0{,}6$
- $C_c(v_4) = 6\ /\ 9 = 0{,}67$
- $C_c(v_5) = 6\ /\ 10 = 0{,}6$
- $C_c(v_6) = 6\ /\ 15 = 0{,}4$
- $C_c(v_7) = 6\ /\ 15 = 0{,}4$

Betweenness

a. Formulas:

$$\text{betweenness: } C_B(v) = \sum_{i,j\neq v, i\neq j} \frac{\sigma_{ij}(v)}{\sigma_{ij}}$$

where σ_{ij} is the number of shortest paths between nodes i and j, and $\sigma_{ij}(v)$ is the number of shortest paths between i and j that go through node v.

b. Shortest paths (and unnormalized betweenness scores for v_4):

From v_x	Shortest Path to v_y	$\frac{\sigma_{ij}(v_4)}{\sigma_{ij}}$
1	1**2**	0
1	1**3**	0
1	~~1**4**~~	$\times$
1	12**5**, 14**5**	$\frac{1}{2}$
1	13**6**	0
1	125**7**, 145**7**	$\frac{1}{2}$
2	21**3**	0
2	~~21**4**, 25**4**~~	$\times$
2	2**5**	0
2	213**6**	0
2	25**7**	0
3	~~3**4**~~	0
3	34**5**	1
3	3**6**	0
3	345**7**	1
4	~~4**5**~~	$\times$
4	~~43**6**~~	$\times$
4	~~45**7**~~	$\times$
5	543**6**	1
5	5**7**	0
6	6345**7**	1

b. Computing betweenness for a single node (v_4):

$$C_B(v_4) = \sum_{i,j \neq v, i \neq j} \frac{\sigma_{ij}(v_4)}{\sigma_{ij}} = \frac{1}{2} + \frac{1}{2} + 1 + 1 + 1 + 1 = 5$$

Normalization for undirected graphs:

$$C'_B(v) = \frac{C_B(v)}{(|V|-1)(|V|-2)/2} = \frac{C_B(v)}{6 \cdot 5/2} = \frac{C_B(v)}{15}$$

c. Betweenness of all nodes:

- $C_B(v_1) = \frac{2.5}{15} = 0.1667$
- $C_B(v_2) = \frac{1}{15} = 0.0667$
- $C_B(v_3) = \frac{5}{15} = 0.3333$
- $C_B(v_4) = \frac{5}{15} = 0.3333$
- $C_B(v_5) = \frac{5.5}{15} = 0.3667$
- $C_B(v_6) = C_B(v_7) = \frac{0}{15} = 0$

Networks with Directed Edges

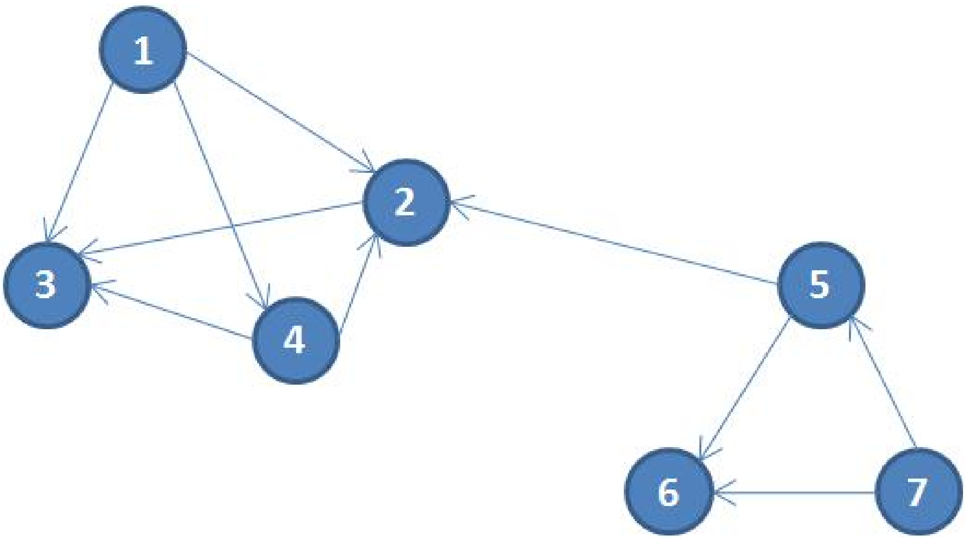

Figure 2: Directed Graph.

Task 4: Prestige

Have a look at the **directed** graph shown in Figure 2. This time we are interested in the actor of the directed graph with the highest **prestige** value. Apply the following measures. Don't forget to **normalize the results** whenever possible/needed.

Degree

Rank: Set the initial values by using the in-degree values. Compute two *rounds*.

Proximity In a directed graph the distance is not commutative (for example $dist(vertice_7, vertice_2) = 2$, but $vertice_2$ cannot reach $vertice_7$). If a vertice cannot be reached from the current vertice do not include it in the set I_v.

Solution:

Degree

Corresponds in directed graphs to a node's in-degree:

$$degree(v_1) = 0$$
$$degree(v_2) = 3$$
$$degree(v_3) = 3$$
$$degree(v_4) = 1$$
$$degree(v_5) = 1$$

$$degree(v_6) = 2$$
$$degree(v_7) = 0$$

(divide by $N - 1$, where N is the number of nodes, so as to normalize.)

Rank

a. Formula:

$$P_R(v) = \sum_{i=1}^{|V|} a_{iv} \cdot P_R(i)$$

Where a_{iv} represents an entry in the adjacency matrix.

b. Computing rank iteratively:

Initial values:
$P_R(v_1) = 0$
$P_R(v_2) = 3$
$P_R(v_3) = 3$
$P_R(v_4) = 1$
$P_R(v_5) = 1$
$P_R(v_6) = 2$
$P_R(v_7) = 0$

Round 1:
$P_R(v_2) = 2$
$P_R(v_3) = 4$
$P_R(v_4) = 0$
$P_R(v_5) = 0$
$P_R(v_6) = 1$

Round 2:
$P_R(v_2) = 0$
$P_R(v_3) = 2$
$P_R(v_6) = 0$

Proximity

a. Formulas:

$$P_P(v) = \frac{|I_v|/(|V| - 1)}{(\sum_{i \in I_v} dist(i, v))/|I_v|}$$

Where

- I_v is the set of actors who can reach v
- $dist(i, v)$ is the geodesic distance from i to v
- $|I_v|/(|V| - 1)$ is the proportion of factors who can reach v
- $(\sum_{i \in I_v} dist(i, v))/|I_v|$ is the average geodesic distance from a node to v

b. Computing proximity for a single node (v_2):

$I_v = 4$ (nodes 1, 4, 5 and 7)

$$P_P(v_2) = \frac{\frac{4}{6}}{(1 + 1 + 1 + 2)/4} = \frac{0.66}{1.25} = 0.53$$

c. Proximity of all nodes:

- $P_p(v_1) = 0$
- $P_p(v_2) = (4 / 6) / ((1+1+1+2) / 4) = 0{,}533$

- $P_p(v_3) = (5\ /\ 6)\ /\ ((1{+}1{+}1{+}2{+}3)\ /\ 5) = 0{,}521$
- $P_p(v_4) = (1\ /\ 6)\ /\ ((1)\ /\ 1) = 0{,}167$
- $P_p(v_5) = (1\ /\ 6)\ /\ ((1)\ /\ 1) = 0{,}167$
- $P_p(v_6) = (2\ /\ 6)\ /\ ((1{+}1)\ /\ 2) = 0{,}333$
- $P_p(v_7) = 0$