ARS: Analisi Reti Sociali
First Half

06/09/2017

Name ______________________________
Student ID ______________________________

Note: Whenever an exercise requires the application of a known formula both the formula and its solution must be **reported** and **discussed**.

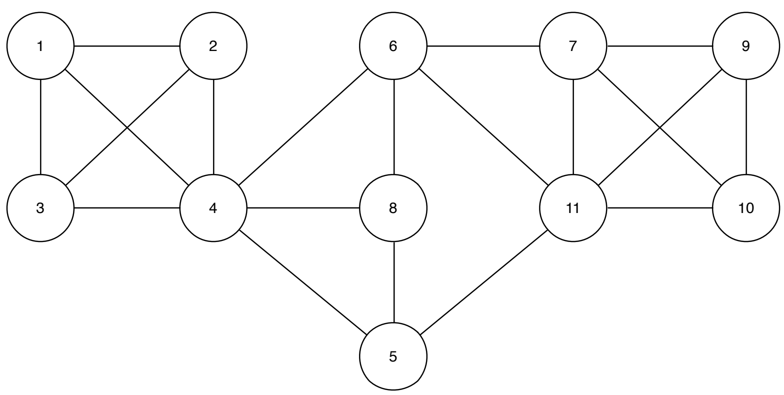

Figure 1:

Exercise 1: Paths & Centrality [5 points]
Given the graph $\mathcal{G}$ shown in Figure 1:

- Compute the diameter of $\mathcal{G}$;
- List all the shortest paths among the pairs [1, 10] and [4, 9];
- Degree Centrality of all nodes;
- Closeness Centrality of 1, 5, 6;

- Local Clustering Coefficient of 1, 5, 6.

Solution.

- $diameter(G) = 4$ (e.g. path among nodes 1 and 9);
- Shortest Paths
  - $(1, 10) \rightarrow 1-4-5-11-10;\ 1-4-6-11-10;\ 1-4-6-7-10$
  - $(4, 9) \rightarrow 4-6-7-9;\ 4-6-11-9;\ 4-5-11-9$
- Degree Centrality: $d_1 = 3$, $d_2 = 3$, $d_3 = 3$, $d_4 = 6$, $d_5 = 3$, $d_6 = 4$, $d_7 = 4$, $d_8 = 3$, $d_9 = 3$, $d_{10} = 3, d_{11} = 5$;
- Closeness Centrality:
  - $C(1) = \frac{1}{10} * ((1 * 3) + (2 * 3) + (3 * 2) + (4 * 2)) = 2, 3$
  - $C(5) = \frac{1}{10} * ((1 * 3) + (2 * 7)) = 1, 7$
  - $C(6) = \frac{1}{10} * ((1 * 4) + (2 * 6)) = 1, 6$
- Clustering Coefficient
  - $CC(1) = 1$
  - $CC(5) = \frac{1*2}{3*2} \simeq 0, 3$
  - $CC(6) = \frac{2*2}{4*3} \simeq 0, 3$

Exercise 2: Graph Construction [6 points]
Given 13 nodes - identified with letters - and, at most, 25 edges build a graph such that all the following conditions hold:

- The graph is composed by two separated components;
- The *shortest path* that connects A and B has length 4;
- The clustering coefficient of node C and D are equal to, respectively, $\frac{1}{2}$ and $\frac{1}{4}$;
- Node D has the highest Degree Centrality;
- Node E has the highest Closeness Centrality (in its component);
- Edge (A,F) has the lowest betweenness centrality.

Solution.
Several solutions are possible.

Exercise 3: Synthetic models[5 points]
Let $G = (V, E)$ be a graph with $|V| = 2000$ nodes and $|E| = 40000$ edges.

- Which value of m allows to generate G with the BA model?
- Which value of p allows to generate G with the ER model?
- What is the expected degree of the largest hub in BA? and the average degree in ER?
- In which regime is the ER graph?

Solution.

BA graph:

- $|E| = m * |V| \rightarrow m = \frac{|E|}{|V|} = 20$
- $k_{max} = k_{min} * |V|^{\frac{1}{\gamma-1}} \simeq 894$

ER graph:

- $|E| = p * \frac{|V|(|V|-1)}{2} \rightarrow p = \frac{2|E|}{|V|(|V|-1)} \simeq 0,02$
- $< k >= \frac{2|E|}{|V|} = 40$
- Supercritical regime: $p = 2 * 10^{-2} > \frac{1}{|V|} = 5 * 10^{-4}$,

ARS: Analisi Reti Sociali
Second Half

06/09/2017

Name ______________________________
Student ID ______________________________

Note: Whenever an exercise requires the application of a known formula both the formula and its solution must be **reported** and **discussed**.

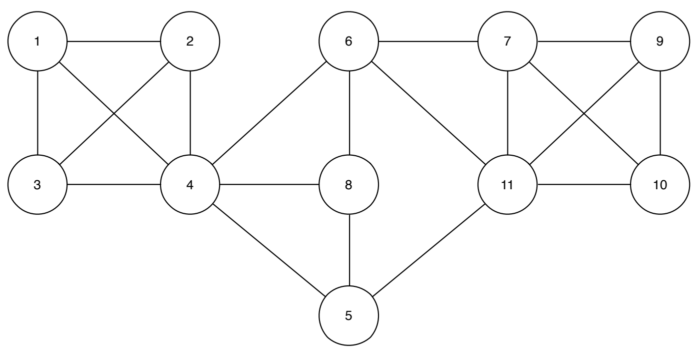

Figure 2:

Exercise 1: Community Evaluation [5 points]
Given the Graph $\mathcal{G}$ shown in Figure 2 compare the following partitions:

P1 = [(1,2,3,4,5,8), (6,7,9,10,11)]

P2 = [(1,2,3,4), (5,6,7,8,9,10,11)]

Which partition is the best with respect to the *Modularity* score? and for *Conductance*?

Solution.
Modularity:

P1 $wP1 = min(mod(C1), mod(C2)) = mod(C2) = mod(C1)$

- $mod(C1) = (\frac{6}{20} - \frac{21}{40})^2 \simeq 0,05$
- $mod(C2) = (\frac{5}{20} - \frac{19}{40})^2 \simeq 0,05$

P2 $wP2 = min(mod(C1), mod(C2)) = mod(C1)$

- $mod(C1) = (\frac{4}{20} - \frac{15}{40})^2 \simeq 0,03$
- $mod(C2) = (\frac{7}{20} - \frac{25}{40})^2 \simeq 0,07$

Best partition w.r.t. modularity:
$max(wP1, wP2) = wP1$

Conductance:

P1 $wP1 = max(Cond(C1), Cond(C2)) = Cond(C2)$

- $Cond(C1) = \frac{2*3}{2*9+3} \simeq 0,28$
- $Cond(C2) = \frac{2*3}{2*8+3} \simeq 0,31$

P2 $wP2 = max(Cond(C1), Cond(C2)) = Cond(C1)$

- $Cond(C1) = \frac{2*3}{2*6+3} \simeq 0,4$
- $Cond(C2) = \frac{2*3}{2*11+3} \simeq 0,24$

Best partition w.r.t. AND:
$min(wP1, wP2) = wP1$

Exercise 2: Threshold Model [5 points]
Given the graph $\mathcal{G}$ shown in Figure 2 apply the Threshold model considering the following two scenarios:

S1
- Set of initial infected nodes: I = $\{5\}$;
- Node threshold τ: 1/4.
- Set of blocked nodes (i.e. nodes that are not allowed to change their status): B = $\{11, 3\}$

S2
- Set of initial infected nodes: I = $\{4\}$;
- Node threshold τ: 1/3 if its id is *even*, 1/4 otherwise.

Consider a node infected at time t iff *at least* τ% of its neighbors were already infected at time $t-1$.

Solution.

S1 0: I_0:$\{5\}$, B_0: $\{3, 11\}$

1: I_1:$\{8\}$, B_1: $\{3, 11\}$

2: I_2:$\{4, 6\}$, B_2: $\{3, 11\}$

3: I_3:$\{1, 2, 7\}$, B_3: $\{3, 11\}$

4: I_4:$\{9, 10\}$, B_4: $\{3, 11\}$

5: I_5:$\{\}$, B_5:$\{3, 11\}$

S2 0: I_0:$\{4\}$

1: I_1:$\{1, 2, 3, 5, 8\}$

2: I_2:$\{6\}$

3: I_3:$\{11\}$

4: I_4:$\{7, 9, 10\}$

5: I_5:$\{\}$

Exercise 3: Community Discovery [4 points]
Given the Graph $\mathcal{G}$ shown in Figure 1 compute its communities applying k-clique for k=3,4,5. Compute AND (average node degree) score for the partition identified with k=3 and k=4.

Solution.

k=3 : $[(1, 2, 3, 4), (6, 7, 9, 10, 11), (4, 5, 6, 8)]$

k=4 : $[(1, 2, 3, 4), (7, 9, 10, 11)]$

k=5 : $[]$

For $k = 3$:
$AND = min(3.75, 3.16, 4.25) = 4.25$
For $k = 4$:
$AND = min(3.75, 3.75) = 3.75$

Exercise 4: Open Question [2 points]
How does the DEMON algorithm work? And the Girvan-Newman one? Describe the peculiarities of the two approaches and of the communities they are able to identify.

Solution.
Grivan-Newman is a top-down iterative approach that removes edges by decreasing betweenness centrality score. After every edge removal the betweenness of the remaining ones is recomputed. The identified communities do not overalp.