ARS: Analisi Reti Sociali
First Half

05/07/2017

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**.

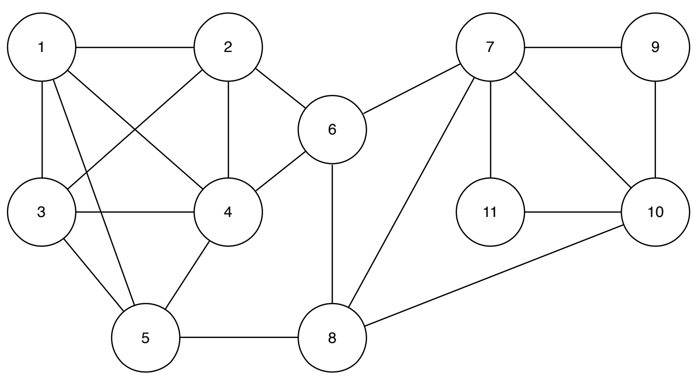

Figure 1:

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 [1,9], [5,9];
- Degree Centrality of all nodes;
- Closeness Centrality of 1, 6, 11;
- Local Clustering Coefficient of 4, 5, 7.

Solution.

- $diameter(G) = 4$ (e.g. path among nodes 1 and 9);
- Shortest Paths
  - $(1,9) \rightarrow 1-2-6-7-9;\ 1-4-6-7-9;\ 1-5-8-10-9;\ 5-8-7-9$
  - $(5,9) \rightarrow 5-8-7-9;\ 5-8-10-9$
- Degree Centrality: $d_1 = 4$, $d_2 = 4$, $d_3 = 4$, $d_4 = 5$, $d_5 = 4$, $d_6 = 4$, $d_7 = 5$, $d_8 = 4$, $d_9 = 2$, $d_{10} = 4, d_{11} = 2$;
- Closeness Centrality:
  - $C(1) = \frac{1}{10} * ((1*4) + (2*2) + (3*2) + (4*2)) = 2$
  - $C(6) = \frac{1}{10} * ((1*4) + (2*6)) = 1,6$
  - $C(11) = \frac{1}{10} * ((1*2) + (2*3) + (3*5)) = 2,3$
- Clustering Coefficient
  - $CC(4) = \frac{2*6}{5*4} = 0,6$
  - $CC(5) = \frac{2*3}{4*3} = 0,5$
  - $CC(7) = \frac{2*4}{5*4} = 0,4$

Exercise 2: Graph Construction [6 points]
Given 12 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 E and F has length 3;
- The clustering coefficient of node F and E are equal to, respectively, $\frac{1}{2}$ and $\frac{1}{3}$;
- Node B has the highest Degree Centrality;
- Node C has the lowest Closeness Centrality (in its component);
- Edge (F,G) has the lowest betweenness centrality.

Solution.
Several solutions are possible.

Exercise 3: Synthetic models[5 points]
Let $G = (V, E)$ be a graph with $|V| = 1400$ nodes and $|E| = 35000$ 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:

- $|E| = m * |V| \rightarrow m = \frac{|E|}{|V|} = 25$
- $k_{max} = k_{min} * |V|^{\frac{1}{\gamma-1}} \simeq 935$

ER graph:

- $|E| = p * \frac{|V|(|V|-1)}{2} \rightarrow p = \frac{2|E|}{|V|(|V|-1)} \simeq 0,036$
- $< k >= \frac{2|E|}{|V|} = 50$
- Supercritical regime: $p = 3,6 * 10^{-2} > \frac{1}{|V|} = 7 * 10^{-4}$,

ARS: Analisi Reti Sociali
Second Half

05/07/2017

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**.

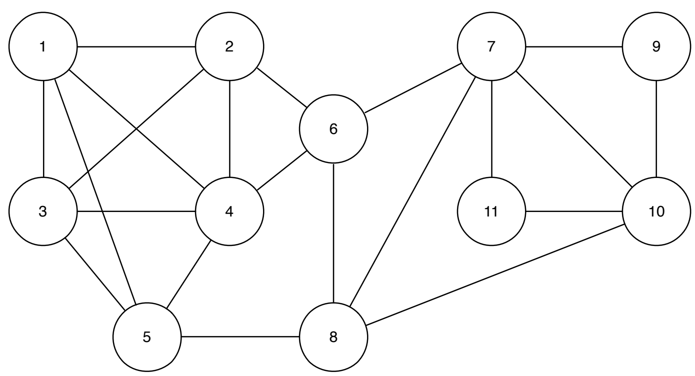

Figure 2:

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), (6,7,8), (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 $wP1 = min(mod(C1), mod(C2), mod(C3)) = mod(C3)$

- $mod(C1) = (\frac{5}{21} - \frac{21}{42})^2 \simeq 0,068$
- $mod(C2) = (\frac{3}{21} - \frac{13}{42})^2 \simeq 0,028$
- $mod(C3) = (\frac{3}{21} - \frac{8}{42})^2 \simeq 0,002$

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

- $mod(C1) = (\frac{6}{21} - \frac{25}{42})^2 \simeq 0,096$
- $mod(C2) = (\frac{5}{21} - \frac{17}{42})^2 \simeq 0,028$

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

Conductance:

P1 $wP1 = max(Cond(C1), Cond(C2), Cond(C3)) = Cond(C2)$

- $Cond(C1) = \frac{2*3}{2*9+3} \simeq 0,29$
- $Cond(C2) = \frac{2*7}{2*3+7} \simeq 1$
- $Cond(C3) = \frac{2*4}{2*2+4} \simeq 1$

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

- $Cond(C1) = \frac{2*3}{2*11+3} \simeq 0,24$
- $Cond(C2) = \frac{2*3}{2*9+3} \simeq 0,29$

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

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 = $\{1, 10\}$;
- Node threshold τ: 1/2.
- Set of blocked nodes (i.e. nodes that are not allowed to change their status): B = $\{2, 4\}$
- Nodes in B are blocked only during *odd* iterations.

S2
- Set of initial infected nodes: I = $\{4\}$;
- Node threshold τ: 1/3 during *even* iterations, 1/4 during *odd* iterations.

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

NB: in both S1 and S2 the first diffusion step is an *odd* one.

Solution.

S1
0: I_0:$\{1, 10\}$, B_0: $\{2, 4\}$
1: I_1:$\{9, 11\}$, B_1: $\{2, 4\}$
2: I_2:$\{7\}$, B_2: $\{\}$
3: I_3:$\{8\}$, B_3: $\{2, 4\}$
4: I_4:$\{5, 6\}$, B_4: $\{\}$
5: I_5:$\{\}$, B_5:$\{2, 4\}$
6: I_6:$\{2, 3, 4\}$, B_6: $\{\}$

S2
0: I_0:$\{4\}$
1: I_1:$\{1, 2, 3, 5, 6\}$ (odd: $\tau = 1/4$)
2: I_2:$\{8\}$ (even: $\tau = 1/3$)
3: I_3:$\{7, 10\}$ (odd: $\tau = 1/4$)
4: I_4:$\{9, 11\}$ (even: $\tau = 1/3$)

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 both the the IED (internal edge density) and AND (average node degree) scores for the partition identified with k=3.

Solution.

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

k=4 : $[(1, 2, 3, 4, 5)]$

k=5 : $[]$

For $k = 3$:
$IED = min(0.73, 0.6) = 0.6$
$AND = min(4.16, 4.2) = 4.16$

Exercise 4: Open Question [2 points]
How does the k-clique 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.

k-clique is a fixed-topology based approach. It searches for cliques (complete subgraphs) of k nodes, then it extends them with all those neighboring nodes that are connected at k-1 of their vertexes. The identified communities can overlap.