ARS: Analisi Reti Sociali
First Half

15/06/2017

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**.

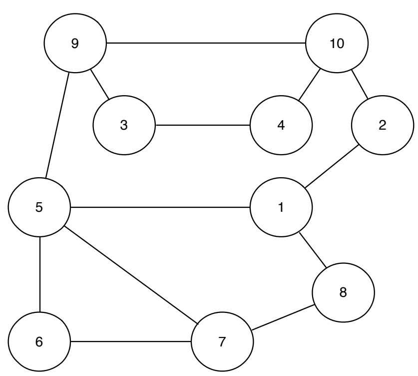

Figure 1:

Exercise 1: Paths & Centrality [6 points]
Given the graph $\mathcal{G}$ shown in Figure 1:

- Compute the diameter of $\mathcal{G}$;
- List all the shortest paths among the pairs [4,7], [2,3];
- Add the minimum number of edges such that an Eulerian path can be identified on $\mathcal{G}$. Define what an Eulerian path is.
- Degree Centrality of all nodes;

- Closeness Centrality of 7, 2, 9;
- Local Clustering Coefficient of 1, 6, 7.

Solution.

- $diameter(G) = 4$ (e.g. path among nodes 3 and 8);
- Shortest Paths
  - $(4, 7) \rightarrow 4 - 3 - 9 - 5 - 7;\ 4 - 10 - 9 - 5 - 7$
  - $(2, 3) \rightarrow 2 - 10 - 9 - 3;\ 2 - 10 - 4 - 3$
- An Eulerian path is a path traversing *all* the graph edges *once*.
  To build it on $\mathcal{G}$ a single edge must be added (1,7).
  The Eulerian path is thus:
  (9,3)(3,4)(4,10)(10,9)(9,5)(5,6)(6,7)(7,5)(5,1)**(1,7)**(7,8)(8,1)(1,2)(2,10);
- Degree Centrality: $d_1 = 3$, $d_2 = 2$, $d_3 = 2$, $d_4 = 2$, $d_5 = 4$, $d_6 = 2$, $d_7 = 3$, $d_8 = 2$, $d_9 = 3$, $d_{10} = 3$;
- Closeness Centrality:
  - $C(7) = \frac{1}{9} * ((1 * 3) + (2 * 2) + (3 * 3) + (4 * 1)) \simeq 2,2$
  - $C(2) = \frac{1}{9} * ((1 * 2) + (2 * 4) + (3 * 3)) \simeq 2,1$
  - $C(9) = \frac{1}{9} * ((1 * 3) + (2 * 5) + (3 * 1)) \simeq 1,7$
- Clustering Coefficient
  - $CC(1) = \frac{2*0}{2*1} = 0$
  - $CC(6) = \frac{2*1}{2*1} = 1$
  - $CC(7) = \frac{2*1}{3*2} \simeq 0,33$

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 three separated components;
- There exists a path of length 4 between nodes A and C;
- The clustering coefficient of node F and A are equal to, respectively, $\frac{1}{3}$ and 1;
- Node F has the highest Degree Centrality;
- Node H has the lowest Closeness Centrality (in its component);

- Edge (B,C) has the highest betweenness centrality (in its component).

Solution.
Several solutions are possible.

Exercise 3: Preferential Attachment[4 points]
Let $\mathcal{G}$ be a BA graph with $N = 1250$ and $m = 13$:

- How many edges in $\mathcal{G}$?
- What is the expected degree of the largest hub?
- What fraction of edges is incident on the largest hub?

Solution.

- $|E| = m * N = 1250 * 13 = 16250$
- $k_{max} = k_{min} * N^{\frac{1}{\gamma-1}} \simeq 460$ ($k_{min} = m$, and $\gamma = 3$)
- $ratio = \frac{k_{max}}{|E|} \simeq 0,028$

ARS: Analisi Reti Sociali
Second Half

15/06/2017

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**.

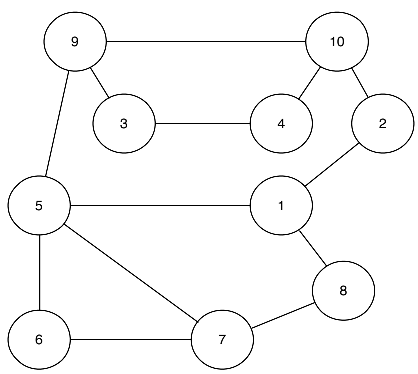

Figure 2:

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

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

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

Which partition is the best with respect to the *modularity* score? which is the best for *average node degree (AND)*? and for *internal edge density (IED)*?

Solution.
Modularity:

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

- $mod(C1) = (\frac{5}{13} - \frac{12}{26})^2 \simeq 0.006$
- $mod(C2) = (\frac{5}{13} - \frac{14}{26})^2 \simeq 0.02$

P2 $wP2 = min(mod(C1), mod(C2), mod(C3)) = mod(C2)$

- $mod(C1) = (\frac{3}{13} - \frac{9}{26})^2 \simeq 0.013$
- $mod(C2) = (\frac{3}{13} - \frac{7}{26})^2 \simeq 0.001$
- $mod(C3) = (\frac{4}{13} - \frac{10}{26})^2 \simeq 0.006$

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

AND:

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

- $AND(C1) = \frac{1}{5} * (3 + 2 + 2 + 3 + 2) = 2,4$
- $AND(C2) = \frac{1}{5} * (3 + 2 + 3 + 4 + 2) = 2,8$

P2 $wP2 = min(AND(C1), AND(C2), AND(C3)) = AND(C2)$

- $AND(C1) = \frac{1}{3} * (4 + 2 + 3) = 3$
- $AND(C2) = \frac{1}{3} * (2 + 3 + 2) = 2,3$
- $AND(C3) = \frac{1}{4} * (3 + 2 + 2 + 3) = 2,5$

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

IED:

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

- $IED(C1) = \frac{2*5}{5*4} = 0.5$
- $IED(C2) = \frac{2*6}{5*4} = 0.6$

P2 $wP2 = min(IED(C1), IED(C2), IED(C3)) = IED(C2) = IED(C3)$

- $IED(C1) = \frac{2*3}{3*2} = 1$
- $IED(C2) = \frac{2*2}{3*2} \simeq 0.7$
- $IED(C3) = \frac{2*4}{4*3} \simeq 0.7$

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

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

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* τ% of its neighbors were already infected at time $t-1$.

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

Solution.

S1 I_0: $\{1, 10\}$
I_1: $\{4, 8\}$
I_2: $\{3\}$
I_3: $\{9\}$
I_4: $\{5\}$
I_5: $\{6, 7\}$

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

Exercise 3: Resilience [4 points]
Given the graph $\mathcal{G}$ shown in Figure 2 remove 5 nodes, one at a time, applying the following selection criteria:

- random selection;
- decreasing degree centrality.

Compare how each removal impacts on the graph structure in terms of *number of connected components* and *average component size*. Specify the node removed - and obtained partition - at each iteration: if multiple nodes generate the same score solve the tie removing the one with lowest identifier.

Solution.

Decreasing Degree Centrality

it_1 : node 5, degree 4, components 1, avg. size 9

it_2 : node 10, degree 3, components 2, avg. size 4

it_3 : node 1, degree 2, components 3, avg. size 2,6

it_4 : node 3, degree 2, components 4, avg. size 1,75

it_5 : node 7, degree 2, components 5, avg. size 1

Random

it_1 : node 2, degree 2, components 1, avg. size 9

it_2 : node 5, degree 4, components 2, avg. size 8

it_3 : node 1, degree 1, components 2, avg. size 3,5

it_4 : node 7, degree 2, components 3, avg. size 2

it_5 : node 10, degree 2, components 3, avg. size 1,7

NB: node degrees *must* be recomputed after each removal.

Exercise 4: Open Question [2 points]
How does the Label Propagation algorithm work? Describe its steps and the rationale behind them.

Solution.
The label propagation algorithm has three steps:

- Step 0: initializzation. Each node is coloured with a different colour (i.e. its label);
- Step 1: randomization. Each node, with probability p, change its colour with the color of one of its neighbors;
- Step 2: iteration. During an iteration t each node change its colour with the most frequent one among its neighbors at time $t-1$. Step 2 is repeated until convergence is reached. In case of ties (i.e. overlapping colouring generating a ping-pong effect) the disputed node is assigned to multiple communities.

The idea underneath LP is to identify internally densely connected network modules by simulation a percolation phenomenon. The number of communities is not defined beforehand, overlapping is allowed.