Social Network Analysis
First Half

18/06/2018

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**.

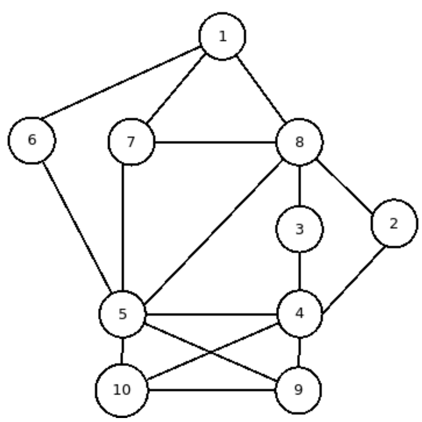

Figure 1:

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 [2,6], [1,9];
- Add the minimum number of edges such that an *Hamiltonian path* can be identified on $\mathcal{G}$. Define what an Hamiltonian path is.

- Degree Centrality of all nodes;
- Closeness Centrality of 1, 2, 10;
- Local Clustering Coefficient of 10, 5, 4.

Solution.

- $diameter(G) = 3$;
- Shortest Paths
  - $(1, 9) \rightarrow [1, 6, 5, 9], [1, 7, 5, 9], [1, 8, 5, 9]$
  - $(2, 6) \rightarrow [2, 4, 5, 6], [2, 8, 5, 6], [2, 8, 1, 6]$
- An Hamiltonian path is a path traversing *all* the graph nodes *once*. An Hamiltonian path is already present in the graph: [3,4,9,10,5,6,1,7,8,2];
- Degree Centrality: $d_1 = 3$, $d_2 = 2$, $d_3 = 2$, $d_4 = 5$, $d_5 = 6$, $d_6 = 2$, $d_7 = 3$, $d_8 = 5$, $d_9 = 3$, $d_{10} = 3$;
- Closeness Centrality:
  - $C(1) = 2$
  - $C(2) \simeq 1.88$
  - $C(10) \simeq 1.77$
- Clustering Coefficient
  - $CC(4) = 0.3$
  - $CC(5) \simeq 0.27$
  - $CC(10) = 1$

Exercise 2: Graph Construction [6 points]
Given 12 nodes - identified with letters - and, at most, 20 edges build a graph such that all the following conditions hold:

- The graph is composed by two separated components;
- There exists a path of length 5 between nodes E and C;
- The clustering coefficient of node C and A are equal to, respectively, $\frac{1}{2}$ and 1;
- Node E 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).

Exercise 3: Preferential Attachment[4 points]
Let $\mathcal{G}$ be a BA graph with $N = 43210$ and $m = 11$:

- 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.

- $|E| = m * N = 43210 * 11 = 475310$
- $k_{max} = k_{min} * N^{\frac{1}{\gamma-1}} \simeq 2287$ ($k_{min} = m$, and $\gamma = 3$)
- $ratio = \frac{k_{max}}{|E|} \simeq 0,0048$

Social Network Analysis
Second Half

18/06/2018

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**.

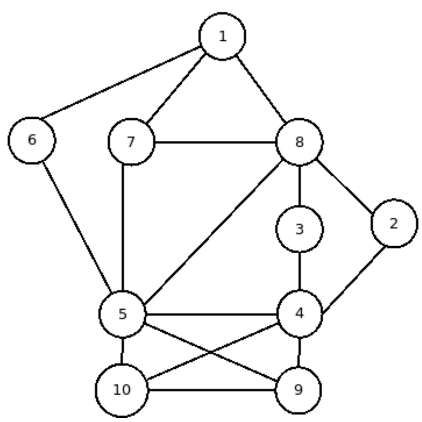

Figure 2:

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

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

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

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

Solution.
AND:

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

- $AND(C1) = \frac{1}{5} * (5 + 6 + 2 + 3 + 3) = 3.8$
- $AND(C2) = \frac{1}{5} * (3 + 2 + 2 + 3 + 5) = 3$

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

- $AND(C1) = \frac{1}{4} * (3 + 6 + 2 + 3) = 3.5$
- $AND(C2) = \frac{1}{4} * (2 + 2 + 5 + 5) = 3.5$
- $AND(C3) = \frac{1}{2} * (3 + 3) = 3$

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

IED:

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

- $IED(C1) = \frac{2*7}{5*4} = 0.7$
- $IED(C2) = \frac{2*5}{5*4} = 0.5$

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

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

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 = $\{1, 9\}$;
- Set of blocked nodes (i.e. nodes that are not allowed to change their status): B = $\{3\}$
- Node threshold τ: 1/3.

S2
- Set of initial infected nodes: I = $\{7\}$;
- Node threshold τ: 1/3 during *even* iterations, 1/2 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 S2 the first diffusion step is an *odd* one.

Solution.

S1 I_0: $\{1, 9\}$

I_1: $\{6, 7, 10\}$

I_2: $\{4, 5, 8\}$

I_3: $\{2\}$

S2 I_0: $\{7\}$

I_1: $\{\}$ (odd: $\tau = 1/2$)

I_2: $\{1\}$ (even: $\tau = 1/3$)

I_3: $\{6\}$ (odd: $\tau = 1/2$)

I_4: $\{5, 8\}$ (even: $\tau = 1/3$)

I_5: $\{2, 3\}$ (odd: $\tau = 1/2$)

I_6: $\{4, 9, 10\}$ (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:

- increasing degree centrality;
- decreasing degree centrality.

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

Solution.

Decreasing Degree Centrality

it_1 : node 5, degree 6, components 1, avg. size 9

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

it_3 : node 4, degree 4, components 4, avg. size 1,75

it_4 : node 1, degree 2, components 5, avg. size 1,2

it_5 : node 10, degree 1, components 5, avg. size 1

Increasing Degree Centrality

it_1 : node 6, degree 2, components 1, avg. size 9

it_2 : node 3, degree 2, components 1, avg. size 8

it_3 : node 2, degree 2, components 1, avg. size 7

it_4 : node 1, degree 2, components 1, avg. size 6

it_5 : node 8, degree 2, components 1, avg. size 5

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

Exercise 4: Open Question [2 points]
Explain the Granovetter's "strength of weak ties" hypothesis. How does it relate to community discovery and network diffusion phenomena?