Social Network Analysis

04/06/2018

Name ______________________________
Student ID ______________________________

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

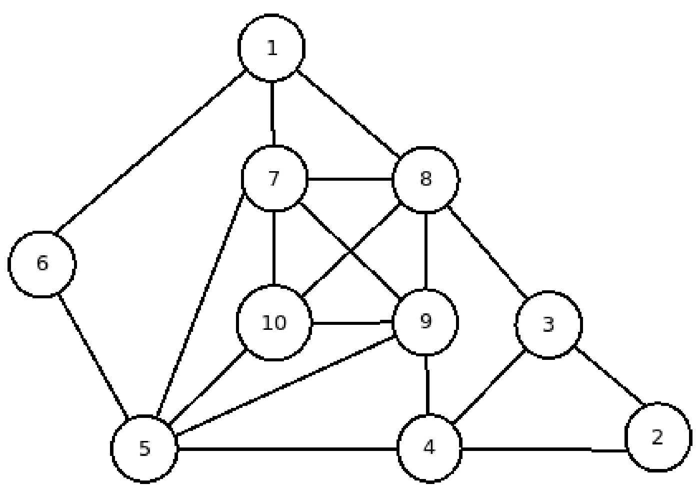

Figure 1:

Exercise 1: Community Evaluation [6 points]
Given the Graph $\mathcal{G}$ shown in Figure 1 compare the following partitions:

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

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

P3 = [(7, 8, 9, 10, 5), (6, 1), (2, 3, 4)]

Which partition is the best with respect to the *modularity* score? and for *conductance*?

Solution.

Modularity:

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

- mod(C1) = $(\frac{4}{19} - \frac{15}{2*19})^2 \simeq$ **0.0339**
- mod(C2) = $(\frac{6}{19} - \frac{23}{2*19})^2 \simeq 0.0837$

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

- mod(C1) = $(\frac{3}{19} - \frac{10}{2*19})^2 \simeq$ **0.0110**
- mod(C2) = $(\frac{4}{19} - \frac{17}{2*19})^2 \simeq 0.0560$
- mod(C3) = $(\frac{3}{19} - \frac{11}{2*19})^2 \simeq 0.0173$

P3: $wP3 = min(mod(C1), mod(C2), mod(C3)) = mod(C2)$

- mod(C1) = $(\frac{5}{19} - \frac{24}{2*19})^2 \simeq 0.1357$
- mod(C2) = $(\frac{2}{19} - \frac{5}{2*19})^2 \simeq$ **0.0006**
- mod(C3) = $(\frac{3}{24} - \frac{9}{2*24})^2 \simeq 0.0062$

Best partitions w.r.t. modularity: P1
$max(wP1, wP2, wP3) = wP1 = wP1$

Conductance: NB: conductance ranges in [0,1]. Due to the relaxation proposed by of the provided formula values greater than 1 may appear: in such scenario, for coherence, we set them to 1.

P1: $wP1 = max(cond(C1), cond(C2)) = cond(C1)$

- cond(C1) = $\frac{2*7}{2*4+7} \simeq$ **0.9333**
- cond(C2) = $\frac{2*7}{2*11+7} \simeq 0.4827$

P2: $wP2 = max(cond(C1), cond(C2), cond(C3)) = cond(C1) = cond(C2) = cond(C3)$

- cond(C1) = $\frac{2*6}{2*3+6} \simeq$ **1**
- cond(C2) = $\frac{2*9}{2*4+9} \simeq$ **1**
- cond(C3) = $\frac{2*7}{2*2+7} \simeq$ **1**

P3: $wP3 = max(cond(C1), cond(C2), cond(C3)) = cond(C2)$

- cond(C1) = $\frac{2*6}{2*9+6} \simeq 0.5$
- cond(C2) = $\frac{2*3}{2*1+3} \simeq$ **1**
- cond(C3) = $\frac{2*3}{2*3+3} \simeq 0.6666$

Best partition w.r.t. conductance: P1
$min(wP1, wP2, wP3) = wP1$

Exercise 2: Community Discovery [6 points]
Given the Graph $\mathcal{G}$ shown in Figure 1 compute its communities applying k-clique for k=3,4,5.

Solution. The communities identified by k-clique are:

k=3 : $\{1, 4, 5, 7, 8, 9, 10\}$, $\{2, 3, 4\}$

k=4 : $\{5, 7, 8, 9, 10\}$

k=5 : $\{\}$

Exercise 3: Threshold Model [8 points]
Given the graph $\mathcal{G}$ shown in Figure 1 apply the Threshold model considering the following two scenarios:

S1
- Set of initial infected nodes: I = $\{1\}$;
- Node threshold τ: 1/4 for even nodes, 1/5 for odd nodes.

S2
- Set of initial infected nodes: I = $\{6, 2\}$;
- Node threshold τ: 1/4
- Set of blocked nodes (i.e. nodes that are not allowed to change their status): B = $\{9, 10\}$

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

Solution. The nodes infected in the two scenarios during each iteration are:

S1

I_0 : {1}
I_1 : {6, 7}
I_2 : {5, 8, 9, 10}
I_3 : {3, 4}
I_4 : {2}

S2

I_0 : {6, 2}
I_1 : {1, 3, 4}
I_2 : {5, 8}
I_3 : {7}

Exercise 4: Resilience [6 points]
Given the graph $\mathcal{G}$ shown in Figure 1 remove 6 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. Multiple solutions are possible (especilly due to the random node selection). One of the possible is the following:

Decreasing Degree Centrality

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

it_2 : node 8, degree 5, components 1, avg. size 8

it_3 : node 4, degree 3, components 2, avg. size 3.5

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

it_5 : node 1, degree 1, components 3, avg. size 1.7

it_6 : node 2, degree 1, components 3, avg. size 1.3

Random

it_1 : node 8, degree 5, components 1, avg. size 9

it_2 : node 10, degree 3, components 1, avg. size 8

it_3 : node 5, degree 4, components 1, avg. size 7

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

it_5 : node 3, degree 1, components 1, avg. size 5

it_6 : node 7, degree 2, components 2, avg. size 2

Exercise 5: Multiple-choice Questions [4 points]
Identify the correct answer(s), among the ones proposed, for to the following questions:

Q1: Which of the following models produce a deterministic diffusion pattern (once fixed the initial infection seeds and its parameters)?

- (A) SI
- (B) SIS
- (C) SIR
- **(D)** Threshold model

Q2: Does the presence of a community structure impact the unfolding of a diffusive phenomena?

- (A) No
- (B) Yes, it slows down the diffusion
- (C) Yes, it speeds up the diffusion
- **(D)** Other – specify: communities speed up intra-community diffusion while slowing down inter-community diffusion.

Q3: Which of the following community quality measures involves a comparison with a null model? which null model?

- **(A)** Modularity, Erdos-Renjy graph
- (B) Internal Edge density, Small-Wolrd model
- (C) Conductance, Barabasi-Albert graph
- (D) Conductance, Erdos-Renjy graph

Q4: Which of the following algorithms produce complete coverage node partitions?

- **(A)** Louvain
- (B) Demon
- (C) k-cliques
- **(D)** Girvan-Newman

Exercise 6: Open Question [2 points]
What does it mean for a problem to be *ill posed*? Discuss the reasons for which Community Discovery is considered an ill posed problem and the implications this have in practice.

Solution. An ill posed problem cannot be described using a single formal definition, thus allowing for different interpretations. The main implications for CD lie in the possible formulation of etherogeneous community definitions (and related algorithmic solutions) that cannot be easily compared within each other.