Name	Solution
ASU ID Number	

CSE 472: Social Media Mining

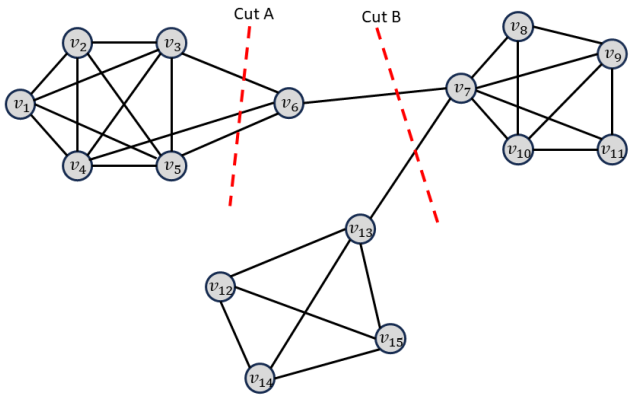
Homework III - Community Analysis, Information Diffusion

Prof. Huan Liu
Due at 2023, October 27^{th}, 11:59 PM

This is an ***individual*** homework assignment. Please submit a digital copy of this homework to **Gradescope**. This is a fillable PDF and you are able to type into answer boxes provided for each question.

1. **[Community Analysis]**

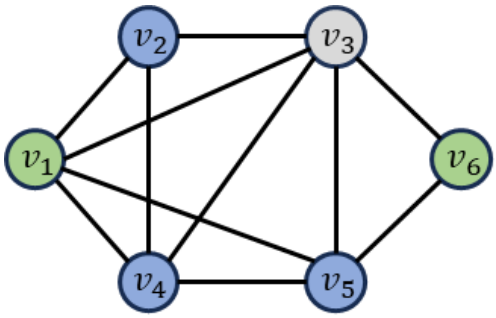
   (a) Consider the following graph. Which cut is more balanced based on **Ratio Cut**? What about **Normalized Cut**? Provide the justifications for both types of cut.



For Cut A:
Ratio Cut ({v1, v2, v3, v4, v5},{v6, v7, v8, v9, v10, v11, v12, v13, v14, v15}) =
(1/2) * (3/5 + 3/10) = 0.45
Normalized Cut ({v1, v2, v3, v4, v5},{v6, v7, v8, v9, v10, v11, v12, v13, v14, v15}) =
(1/2) * (3/23 + 3/37) = 0.1057

For Cut B:
Ratio Cut ({v1, v2, v3, v4, v5, v6},{v7, v8, v9, v10, v11}, {v12, v13, v14, v15}) =
(1/3) * (1/6 + 2/5 + 1/4) = 0.2722
Normalized Cut ({v1, v2, v3, v4, v5, v6},{v7, v8, v9, v10, v11}, {v12, v13, v14, v15}) =
(1/3) * (1/27 + 2/20 + 1/13) = 0.0713
Based on the Ratio cut, Cut B is more balanced. Based on Normalized Cut, Cut B is more balanced.

(b) One partitioning of three group is shown in the following graph, indicated by different colors, $P1 : (v_1, v_6)$, $P_2 : (v_2, v_4, v_5)$, and $P_3 : (v_3)$. Compute **Modularity** for each of the three groups, and then compute **Normalized Modularity**.



Q1 = 2*(0 - (4*2)/22) + (0 - (4*4)/22) + (0 - (2*2)/22) = -18/11 = -1.6363

Q2 = 2*(1 - (3*4)/22) + 2*(1 - (4*4)/22) + 2*(0 - (3*4)/22) + (0 - (3*3)/22) + (0 - (4*4)/22)
+ (0 - (4*4)/22) = -3/2 = -1.5

Q3 = (0 - (5*5)/22) = -25/22 = -1.1363

Q = (1/22) * (Q1 + Q2 + Q3) = (1/22) * ((-18/11) + (-3/2) + (-25/22)) = -47/242 = -0.1942

(c) Compute the following metric for the given figure:

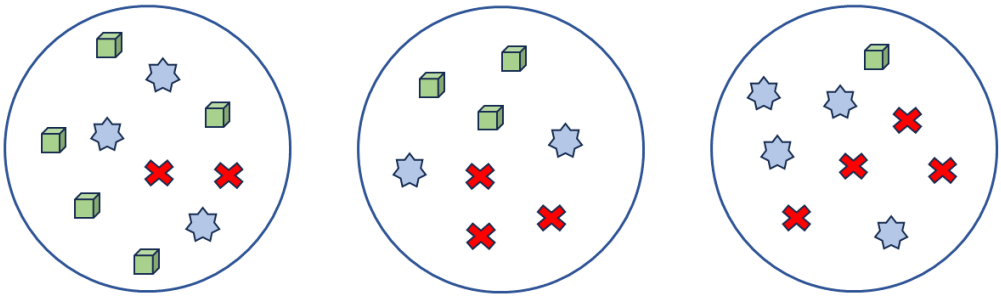


Figure 1: The communities.

Normalized Mutual Information (NMI): [It is necessary to set the default base value for all logarithms to 2.]

H: [1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,3,3,3,3,3,3,3,3,3] , n = 27
L: [1,1,1,1,1,1,2,2,2,3,3,1,1,1,2,2,3,3,3,1,2,2,2,2,3,3,3,3]

MI = 5/27*log2((27*5)/(10*9)) + 3/27*log2((27*3)/(10*9)) + 2/27*log2((27*2)/(10*9))
+ 3/27*log2((27*3)/(8*9)) + 2/27*log2((27*2)/(8*9)) + 3/27*log2((27*3)/(8*9))
+ 1/27*log2((27*1)/(9*9)) + 4/27*log2((27*4)/(9*9)) + 4/27*log2((27*4)/(9*9))
≈ 0.108136

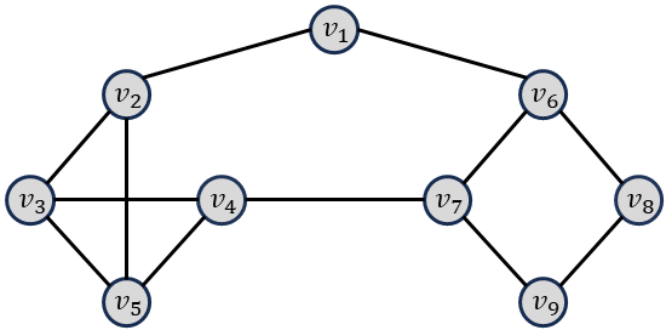
H(H) = -10/27*log2(10/27) -8/27*log2(8/27)-9/27*log2(9/27)
≈ 1.579013

H(L) = -9/27*log2(9/27) -9/27*log2(9/27)-9/27*log2(9/27)
≈ 1.584962

NMI = MI / sqrt(H(L) * H(H))
≈ 0.108136 / (sqrt(1.584962*1.579013)) ≈ 0.068354

2. [**Information Diffusion**]

(a) For the following graph, assume that node i activates node j when $|i - j| \equiv 1 \pmod 3$ and nodes 5 and 9 are activated at time 0. Follow **the Independent Cascade Model (ICM)** and detail all steps until it converges (Traverse the nodes in numerical order).



(t = 1) node v5 ==> v4: |5-4| ≡ 1 (mod 3) ==> v4 is activated.
v2: |5-2| ≡ 0 (mod 3), v3: |5-3| ≡ 2 (mod 3) ==> v2,v3 is not activated.
node v9 ==> v8: |9-8| ≡ 1 (mod 3) ==> v8 is activated.
v7: |9-7| ≡ 2 (mod 3) ==> v7 is not activated.
Activated nodes = {v4,v5,v8,v9}
(t = 2) node v4 ==> v3: |4-3| ≡ 1 (mod 3) ==> v3 is activated.
v7: |4-7| ≡ 0 (mod 3) ==> v7 is not activated.
node v8 ==> v6: |8-6| ≡ 2 (mod 3) ==> v6 is not activated.
Activated nodes = {v3,v4,v5,v8,v9}
(t = 3) node v3 ==> v2: |3-2| ≡ 1 (mod 3) ==> v2 is activated.
Activated nodes = {v2,v3,v4,v5,v8,v9}
(t = 4) node v2 ==> v1: |2-1| ≡ 1 (mod 3) ==> v1 is activated.
Activated nodes = {v1,v2,v3,v4,v5,v8,v9}
(t = 5) node v1 ==> v6: |1-6| ≡ 2 (mod 3) ==> v6 is not activated.
Activated nodes = {v1,v2,v3,v4,v5,v8,v9}

Activated nodes = {v1,v2,v3,v4,v5,v8,v9} and nodes v6 and v7 did not get activated!

(b) If you were given permission to introduce only one additional edge within the existing network, connecting either node 5 or node 9 with a node that is currently not connected to it, would it be possible to activate all nodes starting from nodes 5 and 9? Please provide an explanation for your response.

Since, nodes v6 and v7 are not activated, by connecting node v6 to either v2 or v5, we can ensure the activation of all nodes in the network. Similarly, connecting v7 to either v3 or v8 guarantees the activation of all nodes.

New Edge: v2 - v6
node v2 ==> v6: |2-6| ≡ 1 (mod 3) ==> v6 is activated.
node v6 ==> v7: |6-7| ≡ 1 (mod 3) ==> v7 is activated.
New Edge: v5 - v6
node v5 ==> v6: |5-6| ≡ 1 (mod 3) ==> v6 is activated.
node v6 ==> v7: |6-7| ≡ 1 (mod 3) ==> v7 is activated.
New Edge: v3 - v7
node v3 ==> v7: |3-7| ≡ 1 (mod 3) ==> v7 is activated.
node v7 ==> v6: |7-6| ≡ 1 (mod 3) ==> v6 is activated.
New Edge: v8 - v7
node v8 ==> v7: |8-7| ≡ 1 (mod 3) ==> v7 is activated.
node v7 ==> v6: |7-6| ≡ 1 (mod 3) ==> v6 is activated.

Good Luck