

# Homework 3

CSIT6000K- Social Networks and Social Computing: A Data Science Perspective

Due date: May 2, 11:59 PM

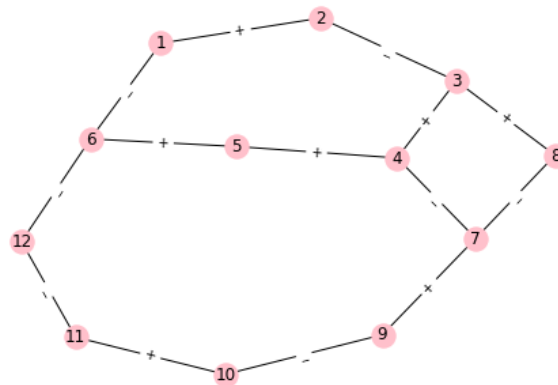
Name: \_\_\_\_\_  
Student ID: \_\_\_\_\_  
E-mail: \_\_\_\_\_

## Notes:

- In this HW, we have 5 theoretical questions. Both hand-written (scanned) and typed answers are accepted.
- You can discuss your ideas with your friends; however, the final answers should all be your own work. ZERO-Tolerance on Plagiarism: All involved parties will get zero mark.
- Send your solutions as a PDF file and name it “hw3\_studentID\_studentName”
- Submit to Canvas only. Any submission by email will be ignored.
- For each hour delay, you will lose 2% of your homework’s grade. Two days after the deadline’s pass, no points will be given anymore. So, please consider submitting your answers on time.

## **Problem 1: Signed Networks (20%)**

The following shows a signed graph that is not complete:



(a) Explain whether the graph shown above is structurally balanced or not?

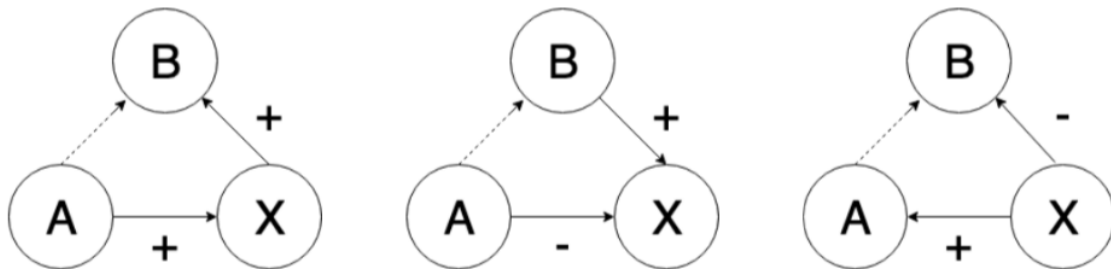
(b) Using the balance theory, predict all the missing signs by filling in the following table below. You only need to fill in the upper-right half, as this table is symmetrical. Explain briefly how you obtain the sign predictions. If the graph is not balanced, change at most 2 of the signs to make the whole graph balanced.

	1	2	3	4	5	6	7	8	9	10	11	12
1	X	+				-						
2		X	-									
3			X	+				+				
4				X	+		-					
5					X	+						
6						X						-
7							X	-	+			
8								X				
9									X	-		
10										X	+	
11											X	-
12												X

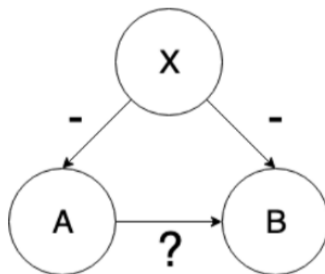
**Problem 2: User Evaluation (30%)**

(a) Show an example triad and the predictions from the balance theory and status theory that differ from each other.

(b) Using the Status theory, predict the sign of the dotted edge in each of the following graphs. (NOTE: Please provide a short explanation)



(c) In the figure below, how should the fraction of positives for the link  $A \rightarrow B$  deviate from the generative baseline of A and the receptive baseline of B according to the status theory? Explain your answer clearly.



### Problem 3: Graph Representation Learning (15%)

Perform random walk in the undirected graph in Figure 1.

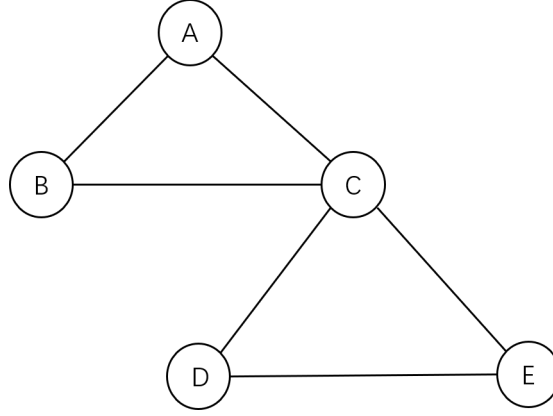


Figure 1: Undirected Graph for Problem 3

Suppose we start from node *A* at Step 0. At each step, we move to one of the neighboring nodes with **equal probability**. Then for each node we have possibly moved to, we continue to move from it to the next node using the same probability rule.

In the first two steps, the probability of walking to each node is shown in Table 1. Please run and obtain the probabilities for Step 2 to 4. Besides filling in the table, you should also write down the computation processes.

Note: Theoretically, the probabilities of walking to certain nodes will converge to a stationary distribution, if the number of steps is large enough. This question aims to provide you with a sense of such convergence.

Step	Node A	Node B	Node C	Node D	Node E
0	1	0	0	0	0
1	0	0.5	0.5	0	0
2					
3					
4					
...					

Table 1: Probabilities in Random Walk

**Problem 4: Contagion and Viral Market (15%)**

- $S(t)$  is the number of susceptible individuals in the population at time  $t$ ;
- $I(t)$  is the number of infected individuals in the population at time  $t$ ;
- $R(t)$  is the number of recovered individuals in the population at time  $t$ ;
- Read Lecture Note 9

Recall that in the SIR model,  $dS/dt = -\beta SI$ , which can be approximated as

$$S(t+1) - S(t) = -\beta S(t)I(t)$$

The other two equations for  $dI/dt$  and  $dR/dt$  can also be similarly interpreted.

(a) Assume that  $S(0) = 10$ ,  $I(0) = 5$ ,  $R(0) = 0$ ,  $\beta = 0.05$ , and  $\delta = 0.25$ . Obtain the values (to two decimal places) of  $S(t)$ ,  $I(t)$  and  $R(t)$  for  $t = 1, 2, \dots, 10$ .

$t$	0	1	2	3	4	5	6	7	8	9	10
$S(t)$	10										
$I(t)$	5										
$R(t)$	0										

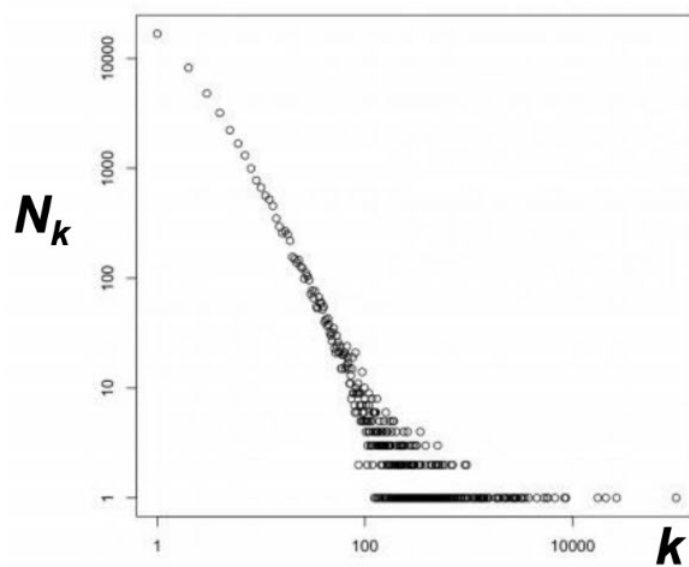
(b) Now we change  $\beta$  to 0.02, and keep all the other numbers the same. Obtain the values (to two decimal places) of  $S(t)$ ,  $I(t)$  and  $R(t)$  for  $t = 1, 2, \dots, 10$ .

$t$	0	1	2	3	4	5	6	7	8	9	10
$S(t)$	10										
$I(t)$	5										
$R(t)$	0										

(c) Interpret the results in parts (a) and (b) based on the virus strength.

### Problem 5: Degree Distribution (20%)

The following plot is the degree distribution of a certain network.  $N_k$  is the number of nodes with degree  $k$ .



- (a) Why are there horizontal lines on right side of the figure?
- (b) Explain why fitting the power law coefficient with linear regression on the above figure is a good or bad idea.
- (c) For the coefficient in the power law distribution, explain why we usually have  $2 < \alpha < 3$  in real networks.