Social Computing **CS60017** **Autumn Semester, 2015**
Mid-Sem **Maximum Marks: 50** **Time Limit: 2 Hours**

This exam contains 4 pages (including this cover page) and 7 problems.

You may *not* use your books or notes for this exam. Be *precise* in your answers. All the *sub-parts* of a problem should be answered at *one place* only. On multiple attempts, *cross* any attempt that you do not want to be graded for.

There are no clarifications. In case of doubt, you can take a valid assumption, state that properly and continue.

1. (7 points) For the network shown in Figure 1, answer the following

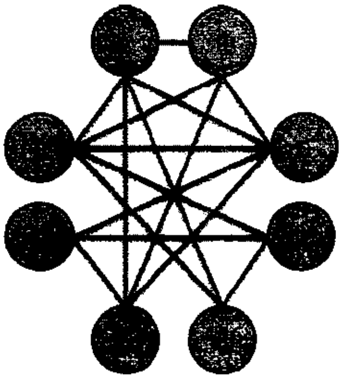


Figure 1: Network to be used for Q1

   (a) Calculate the density of this network. (2 points)

   (b) Draw the 1.5-degree egocentric network of node A. (1 point)

   (c) Calculate the closeness centrality for nodes A and G in the network. Which of these nodes is more central? (4 points)

2. (10 points) Consider the network shown in Figure 2.

   Assume that these 5 nodes correspond to 5 different Twitter users. The connections are shown but are not directed. Mark the directions in these edges such that users which come earlier in lexicographical ordering follow the users which come later. Thus, A follows E and so on.

   (a) Suppose that users B and C are known to be spammers. Use the CollusionRank algorithm to obtain the CollusionRank scores for all the nodes. It is sufficient to show the first iteration. You should update the scores in lexicographical ordering. Assume the value of α to be 0.8. (5 points)

   (b) Suppose that you have to run random walk with restarts on this network with C being the restart node. Write down the set of equations you will use to obtain the scores of each node. [You are not required to show the computations] (5 points)

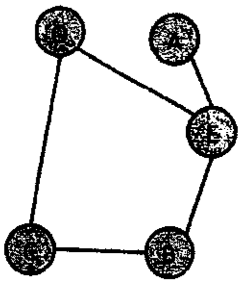


Figure 2: Network to be used for Q2

3. (7 points) Answer the following questions:

   (a) Arrange the following social media applications/websites in ascending order (from least formal to most formal) of the typical level of formality of conversation/language usage on them. (2 points)

   - LinkedIn
   - WhatsApp
   - Facebook
   - Twitter

   (b) In the network shown in Figure 2, let Q and E denote the first 2 users of a given hashtag. Label all the nodes in the graph with their internal degree and entering degree. You must specify clearly as to which nodes are being labeled for internal degree. (2 points)

   (c) Suppose that you are given the stickiness and persistence value for a hashtag. Can you use these values to approximate the exposure curve for this hashtag? Explain. (3 points)

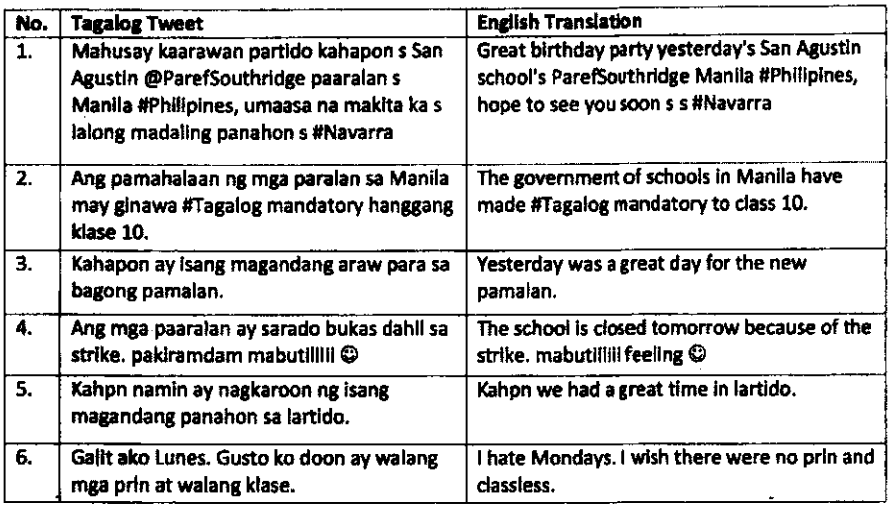
4. (10 points) The following table shows 6 Tagalog Tweets from a teenager from Manila, Philippines. In order to understand these tweets, you thought you will use an online Tagalog-to-English machine translation (MT) service. The translations generated by the MT system are shown as well.

   (a) Identify the Tagalog words from the tweets, which you believe are not typed correctly. Provide the correct spellings and the English meanings of those words. If you cannot decide on a meaning, you can mention a few guesses). [Hint: there are 9 such words] (5 points).

   (b) Which of the spelling mistakes you identified were unintentional? (1 point)

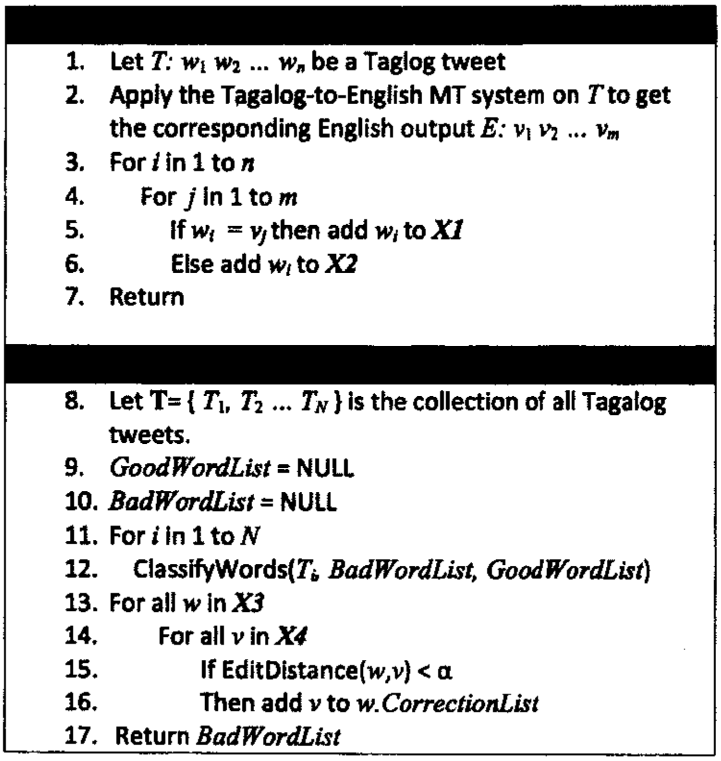
   (c) Given below is a sketch of a procedure that intends to automatically spot informal spellings in a Tagalog tweet corpus and then generate a list of potential corrections (i.e., the correct spelling) for each of these words. Four variable names were replaced by **X1, X2, X3** and **X4**. Identify the original variable names for **X1** to **X4** (Hint: they appear in other parts of the pseudocode). (2 points)

   (d) After running the above procedure, it was seen that while the informal spellings were indeed captured well, a lot of the correct spellings were also identified as informal spelling. It was discovered that this flaw could be rectified to some extent by changing the condition of the if-statement in Step-15 to:



No.	Tagalog Tweet	English Translation
1.	Mahusay kaarawan partido kahapon s San Agustin @ParefSouthridge paaralan s Manila #Philipines, umaasa na makita ka s lalong madaling panahon s #Navarra	Great birthday party yesterday's San Agustin school's ParefSouthridge Manila #Philipines, hope to see you soon s s #Navarra
2.	Ang pamahalaan ng mga paralan sa Manila may ginawa #Tagalog mandatory hanggang klase 10.	The government of schools in Manila have made #Tagalog mandatory to class 10.
3.	Kahapon ay isang magandang araw para sa bagong pamalan.	Yesterday was a great day for the new pamalan.
4.	Ang mga paaralan ay sarado bukas dahil sa strike. pakiramdam mabutilllll ☺	The school is closed tomorrow because of the strike. mabutilllll feeling ☺
5.	Kahpn namin ay nagkaroon ng isang magandang panahon sa lartido.	Kahpn we had a great time in lartido.
6.	Galit ako Lunes. Gusto ko doon ay walang mga prin at walang klase.	I hate Mondays. I wish there were no prin and classless.

Figure 3: To be used for Q4



```
1.  Let $T$: $w_1$ $w_2$ ... $w_n$ be a Taglog tweet
2.  Apply the Tagalog-to-English MT system on $T$ to get
    the corresponding English output $E$: $v_1$ $v_2$ ... $v_m$
3.  For $i$ in 1 to $n$
4.      For $j$ in 1 to $m$
5.          If $w_i$ = $v_j$ then add $w_i$ to $X1$
6.          Else add $w_i$ to $X2$
7.  Return

8.  Let T= { $T_1$, $T_2$ ... $T_N$ } is the collection of all Tagalog
    tweets.
9.  $GoodWordList$ = NULL
10. $BadWordList$ = NULL
11. For $i$ in 1 to $N$
12.     ClassifyWords($T_i$, $BadWordList$, $GoodWordList$)
13. For all $w$ in $X3$
14.     For all $v$ in $X4$
15.         If EditDistance($w$,$v$) < $\alpha$
16.         Then add $v$ to $w.CorrectionList$
17. Return $BadWordList$
```

Figure 4: To be used for Q4

If $(EditDistance(w, v) < \alpha$ AND $w.frequency/v.frequency < \beta)$
And then removing the w from the BadWordList if the procedure was not able to find any potential corrections for the word w. What do you think was the cause behind the

false positives (correct spellings identified as informal), and how it got rectified with the above correction? Can you guess the value or the bound on the threshold β? (2 points)

5. (4 points) For the Friedkin-Johnsen model of opinion formation based on social influence, obtain an expression for the opinion profile at time t in terms of the initial opinion profile and the known matrices.

6. (8 points) Consider the following variant of the Naming Game model. Keeping everything else in the game exactly same, we introduce the concept of *overhearers* who are a subset of the population and update their inventories as follows – in every round of the game, each overhearer overhears the word transmitted by the speaker to the listener; if the word is in her inventory, she removes all the words from her inventory except this word (i.e., treats the event as a success) else she adds this word in her inventory (i.e., treats the event as a failure). Assuming that at each step of the game N^{δ} agents are randomly selected as overhearers re-estimate: (i) the number of unique words in the system close to maxima, (ii) the scaling of ${N_w}^{max}$ and (iii) the scaling of $t_{m\hat{a}x}$.

7. (4 points) In the context of the re-tweeting convention study discussed in class, *criticality* of a node x is defined as the percentage of nodes that adopted a variation after they were exposed to the variation exclusively from that node x or its descendants in the diffusion graph. Experimentally, obtained values of criticality are noted in the table below. Explain in at most 2-3 lines what do these values tell about the nodes in the diffusion graph? Assume that the diffusion graph is small and has only 500 nodes.

Variation	Criticality (%)
via	2
HT	1.8
Retweet	1.1
Retweeting	2.4
RT	0.5
R/T	1.6
Recycle	4.9