This exam contains 2 pages (including this cover page) and 8 problems.

You may *not* use your books, notes, or any calculator on this exam.

1. (5 points) Consider the classical model of opinion dynamics with fixed stochastic matrix A of the form

$$A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix}$$

Determine the stationary probability vector π.

2. (7 points) Consider a variant of the naming game model where the failure rule remains intact while the success rule is modified as follows: both the agents delete a particular word with some probability p. Formulate the evolution equation for the scaling of the maximum memory (N_w^{max}). Find the expression for N_w^{max} if (i) $p = 1/N$ and (ii) $p = 1/\sqrt{N}$.

3. (5 points) Show that for the faster version of the reservoir sampling algorithm, the expected value of a skip $S(n, t)$ is given by $\frac{t-n+1}{n-1}$.

4. (3 points) Consider the scientific collaboration network presented in Figure 1. How would the network structure change after the "named" entities have been correctly resolved.
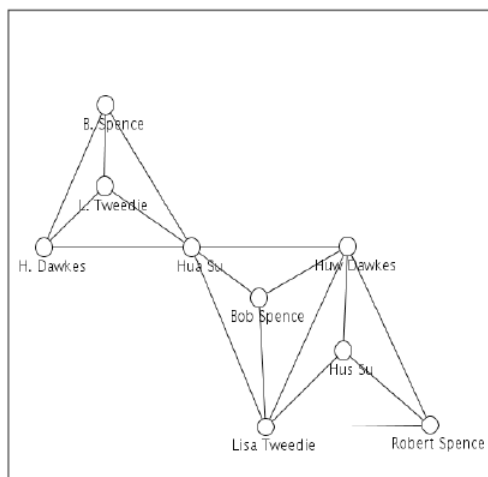

Figure 1: Scientific collaboration network.

5. (7 points) We discussed the paper, "Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter" during one of the classes. The authors studied the mechanism of information diffusion across various hashtag categories. The categories and and a few examples are given below.

Category	Examples
Celebrity	mj, brazilwantsjb, regis, iwantpeterfacinelli
Music	thisiswar, mj, musicmonday, pandora
Games	mafiawars, spymaster, mw2, zyngapirates
Political	tcot, glennbeck, obama, hcr
Idiom	cantlivewithout, dontyouhate, musicmonday
Sports	golf, yankees, nhl, cricket
Movies/TV	lost, glennbeck, bones, newmoon
Technology	digg, iphone, jquery, photoshop

The authors manually annotated these hashtags. Instead, can you propose an approach to label the hashtags among one of these categories automatically? If you are going to use features, each feature should be accompanied with an appropriate justification, as to why do you think that feature would be important.

6. (7 points) Consider the following network, consisting of 6 nodes. Write down the equations that you will use to compute the hub as well as authority scores of various nodes in the (a) HITS and (b) SALSA frameworks.
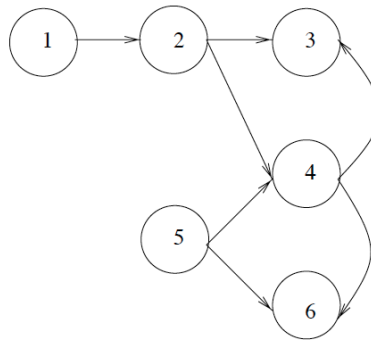
7. (6 points) Cosider a ring lattice with k nodes and $2k$ edges, such that each node in the lattice is connected to 4 nearest nodes, 2 immediately previous nodes and 2 immediately next nodes. Compute the local clustering coefficient of any arbitrary node in the lattice.

8. (10 points) In the case of *supervised random walks*, you learn a function $f_w(\psi_{uv})$ that outputs an edge strength a_{uv} between any two edges u and v.

(a) (3 marks) Explain how you use these edge strengths for the task of link prediction.

(b) (7 marks) During the task of link prediction, you use an initial distribution. What initial distribution should you use? If the choice of initial distribution does not matter, you should prove that.