

**Course:** CSC591/791, Graph Data Mining: Theory, Algorithms, and Applications

**Final Exam:** Comprehensive, home-take exam.

**STUDENT's ID:** \_\_\_\_\_

---

This exam is intended to be an **individual** effort. You are allowed to use and reference the materials posted on Moodle, cited in the exam, your homework assignments, your notes, the class slides, and the lecture assignments. You are **NOT** allowed to use the other sources, such as your classmates and the internet at large.

1. Good luck!
  2. You get BONUS 5 points if your solution is in LaTeX (both source and PDF)
  3. Otherwise, a PDF (even a scanned hand-written solution) must be uploaded into Moodle
  4. The Total Exam Score by the TA: \_\_\_\_\_ out of **(100 points)**
- 

*From the NCSU Code of Student Conduct:*

## DEFINITIONS OF ACADEMIC DISHONESTY

1. Academic dishonesty is the giving, taking, or presenting of information or material by a student that unethically or fraudulently aids oneself or another on any work which is to be considered in the determination of a grade or the completion of academic requirements or the enhancement of that student's record or academic career.
2. A student shall be guilty of a violation of academic integrity if he or she:
  - represents the work of others as his or her own;
  - obtains assistance in any academic work from another individual in a situation in which the student is expected to perform independently;
  - gives assistance to another individual in a situation in which that individual is expected to perform independently;
  - offers false data in support of laboratory or field work.

3. The act of submitting work for evaluation or to meet a requirement is regarded as assurance that the work is the result of the student's own thought and study, produced without assistance, and stated in that student's own words, except as quotation marks, references, or footnotes acknowledge the use of other sources. Submission of work used previously must first be approved by the instructor.
4. Regulations regarding academic dishonesty are set forth in writing in order to give students general notice of prohibited conduct. They should be read broadly and are not designed to define academic dishonesty in exhaustive terms.
5. If a student is in doubt regarding any matter relating to the standards of academic integrity in a given course or on a given assignment, that student shall consult with the faculty member responsible for the course before presenting the work.

By signing this exam, the student acknowledges the above terms and agrees to abide by NCSU policies on academic integrity.

Agreed \_\_\_\_\_ (specify your ID)

Date \_\_\_\_\_

1. \_\_\_\_\_ out of **(15 points)**: Given an infectious disease following the SIS virus propagation model with transmission probability  $\beta = 0.3$  and healing probability  $\delta = 0.5$ , in a contact network with 10 nodes and the following edges:  $\{(0, 1), (0, 3), (1, 4), (2, 3), (2, 6), (5, 6), (5, 7), (6, 7), (7, 9), (8, 9)\}$ , answer:
  - (a) \_\_\_\_\_ out of **(1 point)**: What is the spectral radius of the network?
  - (b) \_\_\_\_\_ out of **(1 point)**: What is the effective strength of the virus?
  - (c) \_\_\_\_\_ out of **(1 point)**: According to the theorem studied in class [1], can the infection result in an epidemic?
  - (d) \_\_\_\_\_ out of **(9 points)**: Using the NetShield algorithm [2], which 3 nodes should be immunized to minimize the spread of the infection?
  - (e) \_\_\_\_\_ out of **(3 points)**: After immunizing the 3 nodes selected in (4), can the infection result in an epidemic?

**References:**

- 1 B. A. Prakash, D. Chakrabarti, M. Faloutsos, N. Valler, C. Faloutsos. Got the Flu (or Mumps)? Check the Eigenvalue! arXiv:1004.0060 [physics.soc-ph], 2010.
- 2 H. Tong, B. A. Prakash, C. Tsourakakis, T. Eliassi-Rad, C. Faloutsos, D. H. Chau. On the Vulnerability of Large Graphs. In ICDM, 2010.

2. \_\_\_\_\_ out of **(15 points)**: Given random variables  $x, y, w, z$ , **prove or disprove** the following statements about their conditional independence ( $\perp$ ) relationships (see the lecture on  $D$ -separation and example below; if you decide to disprove the stmt, then a counter-example of a DAG will suffice):

- (a) \_\_\_\_\_ out of **(0 points)**: Example:  $(x \perp y, w | z) \implies (x \perp y | z)$

**Proof:**

$$p(x, y | z) = \sum_w p(x, y, w | z) = \sum_w p(x | z) p(y, w | z) = p(x | z) \sum_w p(y, w | z) = p(x | z) p(y | z)$$

- (b) \_\_\_\_\_ out of **(6 points)**:  $(x \perp y, z) \& (x, y \perp w | z) \implies (x \perp w | z)$

- (c) \_\_\_\_\_ out of **(9 points)**:  $(x \perp y, z) \& (x \perp y | w) \implies (x \perp y | z, w)$

3. \_\_\_\_\_ out of **(15 points)**: Suppose you need to design an efficient algorithm for analyzing random walks of an undirected simple graph  $G$ . Let  $A$  be its adjacency matrix with Singular Value Decomposition (SVD)  $A = U \Lambda U^t$ .

- (a) \_\_\_\_\_ out of **(5 points)**: Prove that the power matrix  $A^k = U\Lambda^k U^t$ .
- (b) \_\_\_\_\_ out of **(10 points)**: Devise an efficient algorithm to compute  $A^k$  for different values of  $k$ . What is its Big-O complexity?
4. \_\_\_\_\_ out of **(10 points)**: Let  $G$  be an undirected complete bi-partite graph. Prove that if  $\lambda$  is an eigenvalue of its adjacency matrix then  $-\lambda$  is also its eigenvalue. Hint: Assume that after the permutation of rows and columns, the adjacency matrix is of the form:

$$A = \begin{pmatrix} 0 & B \\ B^t & 0 \end{pmatrix}$$

5. \_\_\_\_\_ out of **(20 points)**: Consider the following training examples:

Instance	$x_1$	$x_2$	$x_3$	class, $y$
1	T	T	T	+
2	T	T	T	+
3	F	F	T	+
4	F	T	F	-
5	T	F	T	-
6	T	F	F	-
7	F	F	F	-
8	F	T	F	-

- (a) \_\_\_\_\_ out of **(10 points)**: You will be using a naïve Bayes classifier for this question. Given a test example with attributes  $x_1 = T$ ,  $x_2 = T$ , and  $x_3 = F$ , which class will be assigned to this test example? Show your work clearly.
- (b) \_\_\_\_\_ out of **(10 points)**: Repeat the question in part (a) using the following Bayesian Belief network as the classifier:

$$x_3 \rightarrow x_2 \rightarrow y \rightarrow x_1$$

6. \_\_\_\_\_ out of **(10 points)**: Show the key steps of the Junction Tree algorithm for the following DAG (Note: No need to show the details of the last step, i.e., the Belief Propagation inference step):

$$A \rightarrow T \rightarrow E \rightarrow D; S \rightarrow L \rightarrow E \rightarrow X; S \rightarrow B \rightarrow D$$

7. \_\_\_\_\_ out of **(15 points)**: Using the Maximum Likelihood estimation, find the parameters for the following Bernoulli Trials problem (show all the derivations):  
INPUT:

- $M$  iid coin flips:  $D = H, H, T, H, \dots$

- Model:  $p(H) = \theta$  and  $p(T) = 1 - \theta$

OUTPUT:  $\theta_{ML}^*$

- (a) \_\_\_\_\_ out of **(10 points)**: What is the likelihood function  $l(\theta; D) = \log p(D|\theta)$ ? (Hint: Introduce a random variable  $x$  that is equal to 1 if the trial is a Head and it is equal to zero if the trial is a Tail.)
- (b) \_\_\_\_\_ out of **(5 points)**: What is  $\theta_{ML}^*$ ?