CS 685 The Structure of Information Networks

Course Information

January 24, 2005

1 Instructor Information

Instructor:	Prof. David Williamson
Office:	236 Rhodes Hall
Office hours:	Mondays 2:30-3:30, Thursdays 11-12, and by appt.
Email:	`dpw@cs.cornell.edu`
Phone:	(607) 255-4883
Course web site:	`www.cs.cornell.edu/courses/cs685/2005sp`

2 Lectures

Lectures will be MWF 1:25-2:15 in Hollister 368.

3 Prerequisites

The formal prerequisite is CS 482. In practice, I will be assuming some previous exposure to basic graph algorithms (e.g. breadth-first search, strongly connected components), linear algebra (e.g. eigenvalues and eigenvectors), and discrete probability. Please talk to me if you have questions about whether you have the necessary background.

That being said, the course is not a Theory course. The prerequisites are needed for understanding some of the lectures and papers we will read, but not necessarily for doing the assigned work.

4 Textbooks and readings

There is no required textbook. We will mostly be looking at papers from the past 5-7 years. Links to these papers will be posted on the course web site.

5 Assigned work

There will be two types of assigned work: reaction papers and a class project. In addition, you are expected to contribute actively to class discussions.

There will be 4-5 reaction papers, assigned on a biweekly basis. The assignment will include a list of papers that have been covered in a segment of lectures. The reaction paper will cover two related papers associated with that segment of lectures; at least one of these must not be on the given list. The reaction paper will then be about three pages of discussion related to the papers. The discussion should go beyond a mere summary

of the two papers. For instance, you might discuss some weaknesses in the papers, and suggest alternative approaches; formulate different hypotheses than the papers and suggest experiments that could be used to validate the hypotheses; discuss related work from a different field that the authors were not aware of, or a potential application of the work to a different field.

Reaction papers will be assigned grades of $\surd+$, $\surd$, $\surd-$, and 0 (the last for incomplete assignments or negligible effort).

The class project is a 10-15 page paper similar to those listed on the course web site. Preliminary proposals for the project are due by spring break in order for me to give you early feedback. The final project will be due at the end of the semester, with an exact deadline to be given later. Projects can be done in groups of up to three people; I will consider larger groups if the scope of the project is correspondingly larger. Standard types of projects include surveying a topic of the course in much greater depth than presented in class, experimental evaluation of algorithms, empirical studies of various datasets (pointers will be given on the course web site to some examples), mathematical modelling of some phenomenon of information networks, and algorithmic development to address some issue of information networks.

The final grade will be based 45% on the reaction papers, 45% on the project, and 10% on in-class participation.

6 Collaboration

You are invited to discuss the course papers and materials, and to discuss your work on your reaction papers with others. However, the choice of papers to discuss in the reaction papers must be your own, and the reaction paper must be your own work (see section below). As mentioned above, the class project can be conducted in groups of up to three people.

7 Academic integrity

You are expected to maintain standards of academic integrity. You must cite sources of ideas that are not your own, quote statements that are not your own, etc. In particular, paraphrasing others' work without attribution is considered plagiarism, and can be expected to result in failing the class. For a detailed discussion of what is and is not appropriate, see pages 16–22 of `web.cornell.edu/UniversityFaculty/docs/AI.Acknow.pdf`.

8 Topics

Here is a very rough outline of the course content, which is subject to change without notice.

- Introduction. The web graph & structure.
- Finding authoritative sites (HITS). Variants of HITS.
- Random walks on the web, PageRank. PageRank variants. Web page sampling.
- Community finding. Eigenvector analysis. Combinatorial methods.

- Problems in web search. Focused crawling and sampling.
- Web graph modelling. Power-law distributions.
- Small-world phenomena.
- Meta-search. Rank aggregation.
- Designing information networks: peer-to-peer search.
- The future of information networks. The semantic web.

9 Your information

Please fill out the information below and return it by the end of the lecture.

Name _____

Email address _____

Major _____ Year _____

Are you taking this class for credit? _____