

CME 305: Discrete Mathematics and Algorithms

Instructor: Professor Amin Saberi (saberi@stanford.edu)

Course Project Handout

Here are some suggested project topics. However you are free to suggest a topic of your own provided you discuss it with Amin.

Cake Cutting Algorithms and Fair Division.

The problem of “fair division”, that is, fairly dividing resources or costs among a collection of people, is one of the most basic problems that any society has to address. A Google search on the phrase “fair allocation” returns over 100K links, referring to the division of sports tickets, health resources, computer networking resources, voting power, intellectual property licenses, costs for environmental improvements, etc. Many of these situations demand formal protocols for division. For example, the 1982 Convention of the Law of the Sea, which was signed by 159 countries, specifies a simple cut-and-choose protocol for dividing seabed mining tracts. As more of the business and society interactions migrate to the Internet, it will become even more critical to have formal, well-studied, protocols for fair division,

There is wide literature on fair division within the fields of economics, political science, mathematics, operations research, and computer science. In the recent years there have been several academic books published on the subject: Young 1994, Robertson and Webb 1998, Moulin 2003, Barbanel, 2004 and two books by Brams and Taylor, 1996 & 1999 providing overviews. Many variations of the basic problem exist, for example, the situation with divisible resources (goods or sometimes “bads,” like chores or other burdens) is quite different from the situation with indivisible resources, and problems involving the division of money can be quite different from those involving the division of resources. But most of the variation comes from the fact that there are many reasonable way to formalize “fairness”, including max-min fairness, proportional fairness, envy-free fairness, etc. which may or may not lead to stable allocations in the sense of say Nash Equilibrium, or strong Nash Equilibrium.

Sponsored Search

A great reference for this topic is Chapter 28 in Algorithmic Game Theory (AGT) (available free on the web). The topic deals with algorithms for the purposes of maximizing the effectiveness of advertising allocation based on user queries in search engines. The following is an excerpt from (AGT) describing some of the background.

Web search engines like Google and Yahoo! monetize their service by auctioning off advertising space next to their standard algorithmic search results. For example, Apple or Best Buy may bid to appear among the advertisements usually located above or to the right of the algorithmic results whenever users search for ipod. These sponsored results are displayed in a format similar to algorithmic results: as a list of items each containing a title, a text description, and a hyperlink to the advertisers Web page ... Advertisers bid for placement

on the page in an auction-style format where the larger their bid the more likely their listing will appear above other advertisements on the page. By convention, sponsored search advertisers generally pay per click, meaning that they pay only when a user clicks on their advertisement, and do not pay if their advertisement is displayed but not clicked ... Typically, in sponsored search mechanisms, the advertisers specify a list of pairs of keywords and bids as well as a total maximum daily or weekly budget. Then, every time a user searches for a keyword, an auction takes place among the set of interested advertisers who have not exhausted their budgets.

Many papers are available on the subject. A good starting point is Amin's paper "AdWords and Generalized On-line Matching". Also see a SIAM news article discussing this work: www.stanford.edu/~saberis/sara.pdf.

Netflix Challenge

This project regards the Netflix competition, the aim of which is to better design movie recommendations based on available information. Complete details for challenge which was first held in October 2006 can be found at <http://www.netflixprize.com/index>. The algorithms for this purpose are known as collaborative filtering algorithms which power most commercial recommendation systems.

One possibility for this project is to incorporate information from IMDB regarding similar movie types, lists of movies that actors have appeared in, etc.

Diffusion Processes on Social Networks

A good introductory book in the topic is "The Tipping Point: How Little Things Can Make a Big Difference" by Malcolm Gladwell. For an academic introduction see "Influential Nodes in a Diffusion Model for Social Networks" by Kempe, Kleingberg and Tardos.

Spread of Epidemics in Networks

This topic deals with modelling and understanding the spread of viruses in human societies and in computer networks. One possible applied project would be to simulate different models of the spread of virii and methods for controlling them. For a theoretical background see Amin's paper "On the Spread of Viruses on the Internet" and the "Epidemic Thresholds on Scale-Free Graphs" by Blanchard et. al..

Parallel Implementations of Graph Algorithms

How can we design graph algorithms that scale efficiently in large scale parallel computations? This is an active area of research with many applications. For a good starting point regarding high performance computing applications see David Bader's research page: <http://www.cc.gatech.edu/~bader/research.html>

Search & Browsing Library

The Library of Congress wants to create an interface using modern search engine techniques. This is a massive project involving several ICME students. Talk to Amin if you are interested.

Sudoku, Latin Squares and Graph Coloring

While Sudoku has recently caught fire around the globe, the underlying mathematical structure have been studied for a long time. Sudoku is a variant of Latin Squares the name of which comes from Leonhard Euler, which gives an idea how long the concept has been around.

One open problem is the size of the "defining set of a latin square", i.e. the minimum number of symbols that are necessary to uniquely define a Latin square. It is conjectured to be lower bounded by $n^2/4$, which is far from the best proven bound.

Random Graph Generation

Random graphs are an important and growing area of research in graph and probability theories, with applications in many fields such as computer science, biology, chemistry, physics, etc. The notion of random graphs was introduced by Erdos and Renyi in their seminal 1959 paper. This is known as the $G(n,p)$ model, where edges between n vertices are assigned i.i.d. with probability p . This model results in many interesting properties, but it fails to capture many of the properties associated with real-life graphs and networks such as heavy tailed degree distribution and high clustering coefficients. To address these concerns there are many other models of random graphs which have been introduced.

Generating LDPC Codes

The aim of this project is looking at various implementations of low density parity checking (LDPC) codes. Within the context of information theory this method ensures error free

message passing over unreliable networks. A good starting reference is “LDPC Codes: An Introduction” A. Shokrollahi.

Network Formation Games

Here game theory concepts are to model the formation of networks as a result of player interactions. Here are two excellent recent papers looking at different models: E. Anshelevich et. al. “Near-optimal Network Design with Selfish Agents” and Fabrikant et. al. “On a Network Creation Game”.

The following two books also offer a good introduction: Tim Roughgarden’s “Selfish Routing And The Price Of Anarchy” and Chapter 19 of Algorithmic Game Theory (available free on the web).

Graphs arising from Social Networks

With the advent of social networking sites such as myspace and facebook, the social sciences are suddenly awash in data. Analysis of this data is the subject of much recent research. The following seminar webpage has a good description of the field:

<http://www.dagstuhl.de/en/program/calendar/semhp/?semid=30825>

In addition Jon Kleinberg’s home page is an excellent starting point.

<http://www.cs.cornell.edu/home/kleinberg/>

Expander Graphs

This project will focus on a class of graphs called expanders. These graphs have important applications in areas of theoretical computer science such as complexity theory, error correcting codes, de-randomization, communication networks etc. For a nice introduction to applications see the following talk.

www.cs.huji.ac.il/~nati/PAPERS/nips04_talk.pdf

There are many classes of graphs which make good expanders (one such example is the family of d -regular graphs). The analysis of expanders involves several branches of mathematics including graph theory, linear algebra and probability theory and others which makes them interesting mathematical objects to study. For theoretical background take a look at the following course webpage.

<http://www.math.ias.edu/~boaz/ExpanderCourse/>

Bioinformatics and Graph Theory

Study of genetic, metabolic and signaling networks. This is a broad area of research which is basically in its infancy, as large data sets are just beginning to emerge. In class we have seen examples with DNA sequencing and reconstruction of phylogenetic trees. These among other topics are covered in more detail in “An Introduction To Bioinformatics Algorithms” by Jones and Pevzner. The following course webpage covers a lot of literature on these topics.

<http://www.biostat.wisc.edu/~craven/cs838-s03.html>