COT4810 - Topics in Computer Science

Dr. Pawel M. Wocjan

Email: wocjan@cs.ucf.edu
Web: http://www.eecs.ucf.edu/~wocjan

Suggested Topics and Articles:

http://www.eecs.ucf.edu/~wocjan/Teaching/COT4810-Spring2007/

Suggested Topics

- Real-life Applications of Graph Theory and Linear Algebra
  - Ranking the Importance of Web Pages
  - Reputation Management in Networks
  - Detecting Communities in Complex Networks
- Quantum Speed-Ups of Graph-Theoretical Problems

Ranking the Importance of Web Pages

- Imagine a library containing billions of documents but with no centralized organization and no librarians.

- In addition, anyone may add a document at any time without telling anyone.

- You may feel sure that one of the documents contained in the collection has a piece of information that is vitally important to you, and, being impatient like most of us, you'd like to find it in a matter of seconds.

- How would you go about doing it?

- Stated in this way, the problem seems impossible. Yet this description is not too different from the World Wide Web, a huge, highly-disorganized collection of documents in many different formats.

- Of course, we're all familiar with search engines so we know that there are efficient solutions; but how do they work?

Ranking the Importance of Web Pages

Many of today's search engines use a two-step process to retrieve pages related to a user's query:

1. Traditional text processing is done to find all documents using the query terms, or related to the query terms of semantic meaning.

   With the massive size of the web, this first step can results in thousands of retrieved pages related to the query.

2. To make this list manageable for a user, many search engines sort this list by some ranking criterion.

   One popular way to create this ranking is to exploit the additional information inherent in the web due to its hyperlinking structure.

   Link analysis has become the means to ranking.

Ranking the Importance of Web Pages

- You will examine two of the most important link-based ranking systems:
  - **PageRank** which is “the heart Google's software and continues to provide the basis for all of their web search tools” and
  - **HITS** which forms the basis for the new search engine Teoma
- You will explore (in detail) the mathematical and algorithmic principles underlying theses algorithms:
  - Theory of Markov chains (random walks)
  - Linear algebra
- You will have the opportunity to implement these algorithms and test them.

Reputation Management in Networks

- Reputation management is the process of tracking an entity's actions and other entities' opinions about those actions; reporting on those actions and opinions; and reacting to that report creating a feedback loop.

- Reputation management has come into wide use with the advent of widespread computing. Reputation management systems use various predefined criteria for processing complex data to report reputation.

Examples of Reputation Management

- **eBay** is an on-line marketplace, a forum for the exchange of goods. The feedback system on eBay asks each user to post his opinion (positive or negative) on the person with whom he or she transacted.

  Since having primarily positive feedback will improve a user's reputation and therefore make other users more comfortable in dealing with him, users are encouraged to behave in acceptable ways – that is, by dealing honestly with other users, both as buyers and as sellers.

- **Wikipedia** is a multilingual, Web-based, free-content encyclopedia project written collaboratively by volunteers, allowing most of its articles to be edited by almost anyone with access to the Web site.

  Wikipedia's size, stature, and growing prominence in the larger world have attracted many users – some of them troublesome. The small-town method, where reputation is managed implicitly, has begun to break down; it is no longer possible for any one user to know the majority of other users, to have any sort of personal opinion on all of them.

- **Peer-to-peer file sharing networks** (such as Gnutella) have many benefits over standard client-server approaches to data distribution, including improved robustness, scalability, and diversity of available data.

  However, the open and anonymous, open nature of these networks leads to a complete lack of accountability for the content a peer put on the network, opening the door to abuses of these networks by malicious peer.

You will examine the EigenTrust algorithm for reputation management for peer-to-peer networks:

- it provides each peer in the network a unique global trust value based on the peer's history of uploads; it aims to reduce the number of “bad” files

- the mathematical and algorithmic ideas underlying EigenTrust are similar to those employed in the PageRank algorithm for ranking the importance of webpages

Detecting Communities in Complex Networks

- Many systems of current interest to the scientific community can usefully be represented as networks. Examples include the Internet and the World Wide Web, social networks, citation networks, and biochemical networks.

- Each of these networks consists of a set of nodes representing, for instance, computers or routers on the Internet or people in a social network, connected together by links or edges, representing data connections between computers, friendships between people, and so forth.

- One network feature that has been emphasized in recent work is community structure, the gathering of nodes into groups such that there is a higher density of edges within groups than between them.

Detecting Communities in Complex Networks

You will examine several different proposal for detecting communities using

- flow-based approach
- resistor network approach
- random walks
- ...

You will have the opportunity to implement some of the algorithms and test them.

Quantum Speed-Ups of Graph Theoretical Problems

- The cross-disciplinary research in quantum computing is motivated by the ambitious aim to build a quantum computer
- A quantum computer is a (hypothetical) computational device that harness physical phenomena unique to quantum mechanics to realize a fundamentally new mode of information processing.
- The goal is to solve certain computational problems more efficiently than any classical computer can do.

You will explore quantum speed-ups for (basic) graph-theoretical problems such as matching and network flows.

The algorithms are classical but use a quantum subroutine.

The subroutine implements Grover's search algorithm which enables us to find elements with certain properties faster than any classical search algorithm.