Complex Network Theory – Spring 2012

Term Projects

Project 1:
Title: Traffic Dynamics in the Indian Railway Network
Brief description: The Indian Railway Network (IRN) is one of the largest transportation networks in the world, carrying millions of passengers and tons of freight every day. Hence the dynamics of traffic flow in this network is a socially important and interesting issue. While there have been several studies on traffic flow in transportation networks (including a few by us), almost all of them have been on static snapshots of the network. In this project, we propose to adopt a different approach, as follows. Recently, there has been a lot of research interest over analysis of dynamic or temporal networks, in which an edge remains in existence for a finite interval of time, as long as two entities or nodes are interacting with each other (e.g. mobile call graphs). A transportation network (e.g. railway network) is actually a dynamic network where stations are nodes, and stations *A* and *B* are connected by an edge only when there is a train travelling from *A* to *B* (or vice versa). This captures the real scenario, since resources need to be allocated on the *A* to *B* route only when there is a train travelling along this route. The network will be a weighted network, where the weight of an edge at a certain time instant is the number of trains travelling along that route at the given time. We have collected a detailed schedule of express trains in IRN, whereby we know the exact position of each express train over the entire day. In this project, we shall use this data to:
(i) Create a Web application to visualize the flow of railway traffic in India over a day
(ii) Study the basic properties of the temporal network, e.g. in periods of congestion
Knowledge of Python (or some similar language), and creating Web applications is desirable for this project.
Mentor: Saptarshi Ghosh (saptarshi.ghosh@gmail.com)

Project 2:
Title: Extracting topical news headlines from Twitter
Brief description: The Twitter online social network has become a great source of real-time news. Recent studies have shown that information regarding events happening 'now' is often more readily obtained from Twitter than from regular news sources. In this project, we desire to build an application that automatically extracts news headlines on certain topics from Twitter. Given a topic (e.g. football or photography), we have ways to identify a set of Twitter accounts that are `experts' on the topic (i.e. sources of authoritative information). On a given day, we can crawl the recent tweets of these experts and then attempt to identify important news from among the tweets. The challenges include identifying topical keywords, differentiating among different news stories (e.g. news stories from different parts of the world), judging the relative importance of the multiple news stories, and so on. For instance, if multiple experts tweet about the same news, then this news is definitely important; but for this, we need to identify which tweets relate to the same news story.
Ultimately, we shall build a Web application to present a Webpage of news headlines based on the choice of topics of a user (similar to the Yahoo or CNN Webpage). Knowledge of Python (or some similar language), and creating Web applications is desirable for this project.
Mentor: Saptarshi Ghosh (saptarshi.ghosh@gmail.com)

Common information for Term Projects 3 – 6:
Lots of research have been done on maximizing the utilization of the available network resource while disseminating a given piece of information from a single node to all other network nodes in a distributed fashion (i.e. algorithms can only use local information). Unfortunately, no algorithm could prove itself to be the best algorithm and there is also no hard theoretical bound (i.e. what maximum utilization can be achieved is also not known). Therefore, there is still a requirement of a better algorithm. That is why, even at the end of 2011, we cannot make conference calls in Skype with more than 25 users (this is officially reported, i.e., applicable in an ideal situation; the practical limit is much lower). In our lab we have got some very basic idea of how the maximum coverage can be achieved under given constraints of time and bandwidth. But it is not yet proved that this is the maximum. An algorithm has been developed that can do much better than the state-of-the-art. With this background, the following problems come out directly which are required to be done very soon.

Project 3:
Title: Mathematical proof of new coverage algorithm
Brief description: Mathematically prove that the derived algorithm is the best possible one.
Mentor: Sudipta Saha (sudipta.saha.22@gmail.com)

Project 4:
Title: Performance measurement in very large wireless ad-hoc networks
Brief description: This problem can be divided into the following parts (use of NS3 may be applicable):
a. Making a brief survey and finding out the best possible algorithm for information dissemination in these networks
b. Implementing and testing the derived algorithm in very large random geometric graphs with different dimensions
c. Comparing its performance with the best known algorithm (which is already surveyed)
d. From the insights available through the results, adding some necessary tweaks to make it suitable for the above networks
Mentor: Sudipta Saha (sudipta.saha.22@gmail.com)

Project 5:
Title: Performance measurement in very large Power-law networks with different clustering coefficients
Brief description: This problem can be divided into the following parts (use of NS3 may be applicable):
a. Making a brief survey and finding out the best possible algorithm for information dissemination in these networks
b. Implementing and testing the derived algorithm
i. In very large computer generated Power-Law networks with different clustering coefficients
ii. In some real world data (possibly Skype)
c. Comparing its performance with the best known algorithm (which is already surveyed)
d. From the insights available through the results, adding some necessary tweaks to make it suitable for the current environment

Mentor: Sudipta Saha (sudipta.saha.22@gmail.com)

Project 6:
Title: Derivation of algorithm with better coverage
Brief description: In a very large mobile network, the direction to be followed to reach a certain mobile node at a certain time is very significant information. However, maintaining this information on the fly is a heavy-weight task. Therefore, under some simplifying assumptions, we plan to exploit the availability of this information at each node for at least their neighbors. Deriving a special variation of the already developed algorithm that can give much better coverage is the final objective. The flow of the work is similar to those mentioned above.
Mentor: Sudipta Saha (sudipta.saha.22@gmail.com)

Project 7:
Title: A study of unit co-occurrence networks in Web search query logs from different geographical regions
Brief description: Complex networks can be found everywhere, even among the units (words or multiword expressions) in Web search query logs. In this assignment, we will study such co-occurrence networks, which can reveal interesting properties about the structure of queries. Specifically, we are interested in a comparative study of the networks from logs of different geographical regions, to see if there exist significant search pattern/content differences. If any, these differences can be used for search personalization and synthetic query generation for research purposes. We will use data from USA (AOL, 2006) and Australia (Bing, 2010).
Mentor: Rishiraj Saha Roy (rishiraj.saharoy@live.in)

Project 8:
Title: An analysis of the interaction of kernel and periphery units within Web search queries
Brief description: Past research on co-occurrence network analysis from Web search query logs has revealed the existence of two kinds of units (words or multiword expressions) in queries – those from the kernel and the periphery. Such a distinction exists in natural language text as well, where we find most day-to-day sentences formed by units from the kernel. When peripheral units occur, they are mostly surrounded by ones from the kernel, helping humans infer the context. But preliminary results from query analysis show that it is not the case with search queries. In this project, we try to characterize the interaction of kernel and periphery units in queries. This can help the search engine in query interpretation, and us to formulate concrete hypotheses for understanding the language of Web queries.
Mentor: Rishiraj Saha Roy (rishiraj.saharoy@live.in)
Project 9:
Title: Degree-preserving link rewiring scheme based on local knowledge of the network
Brief description: It has been found that some networks are robust to some attacks (node removal and edge removal) whereas other networks are highly fragile to the same attacks. It is observed that resilience of a network is an intrinsic property which largely depends upon the degree-degree correlation existing in it. Supervising all the link formations in large dynamic real-world networks in order to maintain a resilient degree-degree correlation is absurd. In this project we will try to develop a degree-preserving link rewiring scheme based on local knowledge of the network and improve the resilience of the overall network.
Mentor: Animesh Srivastava (animesh.srivastava@yahoo.com)

Project 10:
Title: The effect of perturbing the spectra of complex networks
Brief description: “Robustness” or “resilience” properties of complex networks are commonly analyzed by perturbing the network. The simplest perturbations, which we call elementary changes, are the omission or addition of a link and/or a node or the rewiring of links. Any topological perturbation can be constructed as a sequence of elementary changes. The degree of degradation of the network performance, measured in terms of some sets of graph metrics (such as giant component size), under a certain topological perturbation is commonly regarded as a measure of the robustness of that network. In this work, we will try to explore the effect of perturbing the spectra of complex networks. A preliminary work has been done "Spectral perturbation and *reconstructability* of complex networks" in PRE 81, 016101 2010.
Mentor: Animesh Srivastava (animesh.srivastava@yahoo.com)

Project 11:
Title: Modular characterization and analysis of technological networks
Brief description: The case study of digital logic circuits in a microprocessor (important scientific question: do the identified structural modules correspond to different functional blocks?) [Methodology: Similar to R K Pan et al, PLOS One 5(2): e9240]
Mentor: Sitabhra Sinha (sitabhra@imsc.res.in)

Project 12:
Title: Network analysis of natural signals
Brief description: The case study of musical sounds (important scientific question: can network representation of the time-frequency decomposition of musical sounds allow machine characterization of different musical genres or even, composers?) [Methodology: Wavelet analysis: B T Grenfell et al, Nature 414, 716 (2001); network reconstruction: R K Pan and S Sinha, PRE 76, 046116 (2007)]
Mentor: Sitabhra Sinha (sitabhra@imsc.res.in)

Project 13:
Title: Can hierarchical and/or modular organization help in network functional performance?
Brief description: The case study of neural network models for solving decision problems (a la Multi-layer perceptron) (Important scientific question: Does the hierarchical and modular structure of the nervous system make it robust w.r.t. dynamical and structural perturbations in taking decisions based on incomplete and noisy data?) [Methodology: Neural network models: Hertz, Krogh and Palmer, Introduction to the theory of neural computation, Addison-Wesley; modular neural network: N Pradhan et al, EPL 94, 38004 (2011)]
Mentor: Sitabhra Sinha (sitabhra@imsc.res.in)

Project 14:
Title: Stability versus complexity in networks with mesoscopic organization
Brief description: The case study of multiple species population equilibria in ecological networks (important scientific question: How do the Jacobian eigenvalues of large networks of coupled dynamical systems arrange in the complex plane when the network is modular and/or

hierarchical and/or have core-periphery organization?) [Methodology: S Sinha, Physica A, 346, 147 (2005)]
Mentor: Sitabhra Sinha (sitabhra@imsc.res.in)

Project 15:
Title: Evolution of modularity in a network of agents playing cooperation-defection games
Brief description: Important scientific question: Can cooperation strategies and network communities co-evolve when agents are allowed to form and break links based on information about past behavior of other agents? [Methodology: Game theory: C. Hauert and G. Szabo, Am. J. Phys. 73, 405 (2005); Games on networks: J. Gomez-Gardenes et al., PRL 98, 108103 (2007)]
Mentor: Sitabhra Sinha (sitabhra@imsc.res.in)

Project 16:
Title: Interaction dynamics of competing contagia on a network
Brief description: Important scientific question: When several contagia are simultaneously diffusing on a network, how will they interact and compete? What is the character of the resulting epidemic? [Methodology: M. Kuperman and G. Abramson, PRL 86, 2909 (2001)]
Mentor: Sitabhra Sinha (sitabhra@imsc.res.in)

Project 17:
Title: Diffusion dynamics in networks switching between different mesoscopic structures
Brief description: Case study of epidemic spreading in an age-structured population of agents (Important scientific question: What are the conditions under which an epidemic will spread in a population whose members interact with different sets of neighbors at different times?) [Methodology: M. Kuperman and G. Abramson, PRL 86, 2909 (2001)]
Mentor: Sitabhra Sinha (sitabhra@imsc.res.in)

Project 18:
Title: Opinion formation on time-varying networks
Brief description: –
Mentor: Animesh Mukherjee (animeshm@gmail.com)

Term Papers

Paper 1:
Title: Community analysis of time-varying networks
Brief description: –
Mentor: Niloy Ganguly (niloy@cse.iitkgp.ernet.in)

Paper 2:
Title: Network of networks
Brief description: –
Mentor: Animesh Mukherjee (animeshm@gmail.com)

Paper 2:
Title: Overlapping Communities
Brief description: –
Mentor: Animesh Mukherjee (animeshm@gmail.com)

Reading Assignments

Assignment 1:
Title: Literature survey on content distribution in wireless networks
Brief description: A content delivery network or content distribution network (CDN) is a system of computers containing copies of data placed at various nodes of a network. When properly designed and implemented, a CDN can improve access to the data it caches by increasing access bandwidth and redundancy and reducing access latency. Data content types often cached in CDNs include Web objects (text, graphics, URLs and scripts), downloadable objects (media files, software and documents), applications, live streaming media, and database queries (ref: Wikipedia). CDN nodes are usually deployed in multiple locations, often over multiple backbones. These nodes cooperate with each other to satisfy requests for content by end users, transparently moving content to optimize the delivery process. Optimization can take the form of reducing bandwidth costs, improving end-user performance (reducing page load times and improving user experience) or increasing global availability of content. Many CDNs are available – some of the popular CDNs are PeerCast, PPLive, PPStream (Akamai Technologies), Amazon CloudFront and YouTube. In most cases, CDN is developed for wired networks with either centralized servers (YouTube and Akamai) or with P2P technology. Due to the availability of cheap handheld devices and ubiquitous wireless connectivity, a huge demand for content has been noticed from wireless users. In 2010, the total wireless Internet traffic is 37% of overall Internet traffic and it has been predicted that by 2015, it will cross 50%. This fact motivates us to rethink about some new content delivery technique which can reduce network traffic. The goal of this project is to study the already existing literature in this area.
Mentor: Sourav Dandapat (sourav.dandapat@gmail.com)