Complex Network Theory
Project Abstracts

1. Bollywood Actors' Network

Definition: Each actor is a node in the network. There is an edge between two nodes (read actors) if they have co-acted in a film. If they have co-acted in "w" films then "w" defines the weight of the edge.

Works to be done

a) Build the Network
b) Analyze the network (should include the studies of weighted degree distribution, clustering index, actor communities etc.)
c) The data is available for movies of 1970s, 1980s, 1990s and 2000 onwards (till 2005). The change in the structural properties of the networks in each decade is to be studied separately.

The overall objective of this work would be to answer certain questions like,

a) Do popular actors always sign films comprising other popular actors (assortativity)
b) What is the trend of Indian cinema over the decades (were there drastic changes)
c) How popular an actor is at the current time and many others that can be done as the work proceeds.
d) Do actors' son always become actors.

- Number of Students – 2
- Data Source – Visit the site http://www.bollywood.de/ , you are supposed to search such other similar sites.
- References - Self-organization of collaboration networks José J. Ramasco,[1,]* S. N. Dorogovtsev,[2,3,4] and Romualdo Pastor-Satorras[5,] Weighted competition scale-free network Shijun Wang and Changshui Zhang. You are supposed to vide other references and come up with new possible studies.
- Contact Person - Mainly Animesh Mukherjee

2. IIT Collaboration Network

Definition: Each author is a node in the network. There is an edge between two nodes (read authors) if they have co-authored a paper. If they have co-authored "w" papers then "w" defines the weight of the edge.

Works to be done

a) Build the Network
b) Analyze the network (should include the studies of weighted degree distribution, clustering index, actor communities etc.)
c) Try to find specialization communities. Find out the necessity of separate specializations.
d) Validate law of scientific productivity.

- Number of Students – 3
- Data Source – This is a bit difficult task.
- References – Vide standard references on collaboration networks.
- Contact Person - Mainly Prof. Niloy Ganguly

3. Graphemic and Phonemic Networks

Definition: Each word is a node in the network. There is an edge between every two nodes (read words) of the network with edge-weight being the edit-distance between the words (both graphemic and phonemic).
Each node also has a weight equal to its usage frequency.
The network properties are to be studied at various edit-distance based thresholds.

Works to be done

a) Build the Network for languages Hindi, Bengali and English.
b) Analyze the network (should include the studies of weighted degree distribution, clustering index, word-level communities etc.)
c) Use the various definitions of edit-distance (transpose allowed/not allowed etc.) and study if there is any change in the nature of the network properties.

The work has direct implications on natural language evolution. For instance ideally words with very close edit-distance that are very frequent should not form the part of the same community.
The work also has applications in building spell-checkers.

- Number of Students – 2
- Data Source – CIIL corpora for Hindi and Bengali. Emille/BNC for English
- References - You are supposed to vide references on edit-distance based spell checkers.
- Contact Person - Mainly Animesh Mukherjee

4. WebSocial Network

You can for example study the orkut network (may be the blog network also)

- Number of Students – 2
- Data Source –
- References – Vide standard references on social networks.
- Contact Person - Mainly Prof. Niloy Ganguly and Prof. Pabitra Mitra

5. Information Percolation

The work is primarily theoretical.

- Number of Students – 2
- Data Source –
- References –
- Contact Person - Mainly Prof. Niloy Ganguly and Bivas Mitra

6. Word Network

Definition: Each word is a node in the network. There is an edge between two nodes (read words) if they co-occur in a sentence. The edge weight is defined as the weighted sum of the number of sentences in which they have co-occurred. Each node also has a weight equal to its usage frequency.

Works to be done

a) Build the Network for languages Hindi, Bengali and English.
b) Analyze the network (should include the studies of weighted degree distribution, clustering index, word-level communities etc.)

The work has direct implications in the study of cross-linguistic similarities and dissimilarities. Network analysis will give insight about the dependency grammar for each of the languages. The work also includes frame-net/verb-net induction and MWE (multi-word expression) analysis.

- Number of Students – 2
- Data Source – CIIL corpora for Hindi and Bengali. Emille/BNC for English
- References - You are supposed to vide references from the works of Ferrer i Cancho and Ricard Solé.
- Contact Person - Mainly Monojit Choudhury and Animesh Mukherjee

7**. Entropy Optimization**

This is a theoretical work in which you are supposed to recreate the work of Ferrar i Cancho and Ricard Solé on ***"Optimization of Complex Networks"***. Once this is done the optimization function may be changed and new studies can be carried out. There is also enough scope to change the optimization algorithm which will be carried out towards the end of the semester.

- Number of Students – 2
- Contact Person - Mainly Monojit Choudhury

8. Network of Musical Strings

Definition: Each node is a musical note or a sequence of notes (note string of a given length). The edge weight is defined as the weighted average of the distance between the note strings in a musical composition.

Works to be done

a) Data source is partially available for 10 ragas, and scripts can be written to download data from the net as well. But some data is expected to be created by the students from Bhatkhande.
b) Build the Network.
c) Analyze the network (should include the studies of weighted degree distribution, clustering index, word-level communities etc.)

This work has an objective to find out the similarities/dissimilarities between various Raags/ th´ats. The various musical genres also need to be studied.

- Number of Students – 2
- Contact Person - Mainly Monojit Choudhury

9. Airlines/Postal/Railways… Networks

Read out the standard references and suggest plans

- Number of Students – 2
- Data Source -
- Contact Person - Mainly Prof. Niloy Ganguly

10. Social Network of Mahabharata (dropped)

The work is yet to be finalized.

- Number of Students – 2
- Contact Person – Prof. Niloy Ganguly, Monojit Choudhury and Animesh Mukherjee.

***** You are free to come up with any other project proposal that you find interesting and relevant**

10. Network of NGOs *

Each NGO is a node and there is an edge between two NGOs if they work in a same area. The number of areas they work together in defines the edge weight.

Work to be done

a) Build the Network
b) Preliminary analysis
c) Try to draw inferences on the necessity of so many NGOs based on the community structures

Number of Students – 2
Contact Person – Prof. Niloy Ganguly and Animesh Mukherjee

* Suggested by Namita Jarika