# Complex Networks (CNT 2014) Term Projects

**Topic 1: Mining tweet streams posted during disasters to improve situational awareness**
**Mentor: Saptarshi Ghosh**

In the present day, online social networks (OSNs) like Twitter and Facebook are important platforms to share and find information. Especially, at the time of natural disasters (e.g., bomb blasts, riots, cyclones, earthquakes), these OSNs are increasingly being used to share real-time updates and news. In this project, we shall attempt to improve the usability of OSNs during such events. During such disasters, Twitter sees a rapid stream of tweets related to the event. Some of these tweets contain valuable situational information, while many other tweets reflect the opinion / sentiment of the people. This tweet stream is almost always too fast and too large for being effectively processed by human beings. We will attempt to make this task easier, by summarizing the tweet stream in real-time, and identifying the most important tweets which provide new information, and so on.

**Pre-requisites:** ability to analyze large amounts of text data. knowledge of scripting languages like Python desirable

**Topic 2: Analysis of Ego-structure in Collaboration Network**
**Mentor: Tanmoy Chakraborty**

In this study, we wish to propose and validate social networks based theoretical model for exploring scholars' collaboration (co-authorship) network properties associated with their citation-based research performance (i.e., h-index). Using structural holes theory, we want to focus on how a scholar's egocentric network properties of density, efficiency and constraint within the network associate with their scholarly performance. For our analysis, we shall mostly use publication data in the Computer Science domain. In particular, we shall design our experiment in order to answer following questions: (i) correlation of the performance of scholars with their ego-network measures, (ii) evolution of a scholar's ego structure over the years and possible driving factors behind it, (iii) how the ego structure of scholars in one research area differs from the others in a particular domain.

**Topic 3: Predicting Future Popularity of publication Forums**
**Mentors: Tanmoy Chakraborty, Saptarshi Ghosh**

New scholarly venues (e.g., conferences and journals) are emerging as research fields expand. Predicting the popularity of these new venues is imperative to assist researchers, librarians, and research institutions. However, prediction based only on traditional citation-based metrics have limitations and are no longer the only or the best choice to determine the impact of scholarly venues. In this project, we will concentrate particularly on the conferences related to the computer science domain and try to predict the popularity of new conferences. Estimation of possible popularity is of great significance and is quite challenging. We plan to utilize several features of fundamental characteristics for these conferences, namely features related to its program committee members, features related to the quality of the papers published previously in these conferences (both paper centric and author centric) etc. We will implement a system which will take a series of features of a particular conference

as input and try to produce as output the estimated popularity after a given time period.

**Topic 4: Edge based attack on time varying network**
**Mentor: Souvik Sur**

**Abstract:** Attack on a network known as the process of deactivation of nodes/ edges in the network. It helps us to identify important portion of a network in order to maintain the overall connectivity. In the domain of time varying network, it has been seen that, state-of-the-art attack strategies are basically some **variant of node based attack** strategies. Still there is **no pure edge based attack** strategy in time varying network domain. In this project, we will explore the following aspects of edge based attack,

1. Effectiveness of random edge based attack.
2. Effectiveness of edge-clustering coefficient based attack.
3. Effectiveness of edge emergence factor based attack.
4. Comparison among above mentioned schemes as well as with the node based schemes.

**Topic 5: Understanding protein stickiness using the theory of random threshold graphs**
**Mentor: Sudipta Saha**

Various proteins stick together in order to perform various important functions in the body. Prediction of such protein associations is a very important step in understanding various biological phenomena and their pathways. Researchers experimentally discovered various such protein pairs along with the details of their stickiness properties. In this work, we try to visualize this protein stickiness phenomena through a different formalization. There are various physical attributes associated with these proteins. We observe a link between the value of the attributes of these proteins and the event that a pair of proteins stick together. A protein-interaction network can be constructed from these protein stickiness information where the nodes will be the individual proteins and an edge between two proteins would indicate the existence of the stickiness between them. The edges can be assigned appropriate weights calculated from the available stickiness properties. In order to understand the existence of stickiness among various proteins under different situations, like enhanced temperature or interaction with other agents, we plan to construct a secondary structure from the main network by pruning edges whose weights cross a certain threshold value. We believe that this whole phenomena of protein stickiness can be modeled by a special kind of threshold graph constructed in this way. Thus, in this work we primarily attempt to reveal the generalized rule behind the association of different proteins.

**Topic 6: Analysis of the HIV infection and the scope of AIDS**
**Mentor: Sudipta Saha**

It is well known that HIV affected people become more prone to various other diseases because of massive fall in their immunity. However, it has been clinically seen that in certain cases an HIV affected person stays healthy for many years and does not grow the AIDS (Aqua Immune Deficiency Syndromes) at all for the entire life. On the other hand, it has been also seen that in certain cases the person grows the AIDS very soon and dies. Even, various other cases have been clinically observed

where the patients survive for a comparatively longer or lesser time than expected. Thus, despite the existence of HIV in the body, various other factors are also associated with the processes of a person growing AIDS. Some of these attributes help a person to survive a long, while, due to some other of these, the chance of acquiring AIDS from HIV increases. Many such attributes have been clinically identified - such as - the context, existence of many other bacteria or virus - like tuberculosis, pneumonia etc., family status of the patient, environment of the patient, role of various antibiotics etc. In this study we aim to understand the significance of the role of each such attribute. We plan to employ network-analysis methodologies to appropriately distinguish between helpful and detrimental attributes which would definitely allow HIV affected persons to survive longer.

**Topic 7: Development of a network analysis package**
**Mentor: Sudipta Saha**
Design and development of a GUI using java2D to display dendrogram from the outputs of well known clustering algorithms. The GUI will also incorporate few common network analysis tools.

**Topic 8: Devising strategies to obtain efficient broadcasting in delay tolerant networks:**
**Mentor: Sandipan Sikdar**
A delay tolerant network(DTN) is characterized by a sparse node population, and by lack of full network connectivity at virtually every time. Eventual packet delivery to the destination can be achieved only through node mobility. It can be noted that a contact opportunity occurs only when two nodes come in close proximity with one another and each such contact opportunity is characterized by a time of contact. Major routing protocols are epidemic, two-hops and spray and wait. The packets are delivered to the system using a push or a pull technique. In pull technique, after establishing a contact with a node v, node u requests a set of packets and if it is present in v then v transfers a subset of it to u. In case of push technique, after establishing a contact with node v, node u checks whether it has a set of packets which v does not have. If there is such a set then u delivers a subset of it to v. In this project we plan to check whether we can come up with a combined strategy to obtain a more efficient broadcast algorithm.

**Topic 9: Context-aware Wi-Fi association for mobile clients**
**Mentors: Swadhin Pradhan & Sourav Dandapat**
When we switch on the Wi-Fi in our smartphones, it gets associated with the nearby Access Point (AP) with strongest RSSI signal. However, if there is no AP in its range, it continuously searches for AP and thereby drains significant amount of energy. Moreover, when the client is moving, it suffers from frequent dissociation and hence reduces battery level due to unnecessary scanning.
Human mobility model study shows that people follow specific set of routes in daily life. So, people generally encounter same set of Wi-Fi APs during their daily transportation. Hence, it can be easily learned about the need of scanning for APs at a particular location by mining the AP association history stored in a smartphone. In this project, we aim to develop a kernel-level Wi-Fi AP scanning and association scheme which will replace the native scanning system in current android smartphones. Moreover, we want to test the system in IIT-Kharagpur campus-wide Wi-Fi network.

**Topic 10: Opinion Mining in social networks**
**Mentor: Abir De**
Opinion propagation is a crucial phenomenon in any social network and has been studied extensively in the context of influence propagation, opinion maximization etc. Existing literature mostly consider synchronous opinion propagation models that reach consensus after a finite number of time steps. In contrast to that, in this work, we attempt to model opinion propagation in a natural setting, where: (a) opinions are continuous and bounded, (b) opinions propagate asynchronously, and (c) opinions need not stabilize over time. We consider a linear model for opinion propagation as has been considered in though they do not consider asynchronous propagation and potential temporal instability.

**Topic 11: Analyzing group performances in cricket**
**Mentors: Suman Kalyan Maity and Swadhin Pradhin**
In this project, we shall try to study the impact of group performance of players in the team on the outcome of the game of cricket. The basic intuition is that average group performance is better than individual brilliance. We shall be using complex network approach to come up with an appropriate ranking function which ranks teams not only based on the number of wins but on the quality of wins. Similarly, the same approach would be used to rank individual players of the teams based on their performance. We then analyze the impact of group performance and individual performance of players on the various aspects of the game that includes batting, bowling, fielding and all the fields taken together which captures the all-round performance.

**Tpoic 12: Evolution of linguistic Styles in Twitter**
**Mentor : Suman Kalyan Maity**
Given their ubiquity, immediacy and accessibility, social media channels such as Twitter have emerged as the de facto medium for information sharing, and communication about various topics from breaking news to personal stories. Twitter houses many features that make its language distinct. On the one hand, unlike on traditional media like blogs, magazines and newspapers, posts on Twitter (tweets) are inherently much shorter and constrained by a hard 140 character limit. Thus, the language of twitter is of great interest to us. In this project, we are interested to analyze the linguistic style in twitter and how it evolves with its growth as social media. The linguistic change corresponds to its word usage, is it becoming more informal with the intrusion of the chattish language in it?

**Topic 13: KnowYourPoliticalLeader: Analyzing the Indian politics unfolded in Social media**
**Mentor : Suman Kalyan Maity**
The landscape of public sentiments is not flat in a democratic governance. In fact, it is remarkably complex in a multi-party democracy such as India, where more than 50 major (few hundred in total) political parties participate in the election process and the Indian constitution permits alliance of various political parties. The aim of this project is to analyze the public sentiments as reflected in social media (twitter) and to know the underlying political relationship by means of complex network tools and techniques.

**Topic 14: Studying Deleted Tweets in Twitter**
**Mentor: Parantapa Bhattacharya**

This project involves collecting and studying information about Tweets that people delete on Twitter. There has been a recent work [1] studying Tweet deletions. However it leaves several questions unanswered. In this work we would look into the questions as to why people delete tweets, and are there significant characteristic specialties in the tweets that are deleted and the users who delete them.

[1] http://www.cs.cmu.edu/afs/.cs.cmu.edu/Web/People/shomir/cscw2013_tweets_are_forever.pdf

**Topic 15: Implementing Community Detection Algorithms for igraph**
**Mentor: Parantapa Bhattacharya and Tanmoy Chakraborty**

igraph [1] is a popular graph library written in C with frontends in Python and R. A number of community detection algorithms have been implemented in igraph. However there are many other popular ones that are not. Most authors when developing graph algorithms use their custom code instead of using standard libraries (for various reasons). Many use their own input and output formats. Which makes using these quite inconvenient to use.

This project involves implementing the following algorithms using igraph (C language). It also involves understanding existing implementations written in various languages (mostly C++ and Java). We intend to submit
the versions you write to the igraph project so other researchers can
also benefit from your work.
Algorithms to be implemented are BigClam [2], OSLOM [3], COPRA [4], SLPA [5], and MaxPerm [6].

[1] http://igraph.sourceforge.net/
[2] http://infolab.stanford.edu/~crucis/pubs/paper-nmfagm.pdf
[3] http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0018961
[4] http://arxiv.org/pdf/0910.5516.pdf
[5] https://sites.google.com/site/communitydetectionslpa/
[6] Unpublished - Ask Tanmoy for Paper.