

# **TELCOM 2125 – Network Science and Analysis**

## **Projects' Descriptions**

Projects will be strictly personal. You are highly encouraged to come up with your own project related to network science but following are some ideas for projects. Projects can be either analytical or data analysis-based using available datasets (see data sources for ideas). You will have to give me a one-page description of your project by the end of January. You will need to make a case on why this problem is important. This will require some preliminary literature review and clear description of the project contribution.

### 1) **Assortative Mixing**

Homophily refers to the phenomenon where connections in a network are biased from node characteristics. For example, it has been shown that friendship networks exhibit homophily with regards to the race of each vertex, that is, edges connect nodes of the same race with much higher probability compared to nodes of dissimilar race. There exist metrics that quantify this effect for networks where nodes are associated with enumerative or scalar characteristics. Little attention has been given to formal treatments of assortative mixing with vector attributes of nodes.

- Literature review on existing metrics for assortative mixing via vector attributes (if any)
  - Benefits and drawbacks of these metrics
- Define new metric(s) for assortative mixing that consider vector attributed
  - Apply these metrics on appropriate datasets and analyze the results

### 2) **Time evolving random graphs**

Random graph models, even though not realistic, are often used as a first approach to obtain intuition behind various phenomena observed in real world. The Erdos-Renyi random graph model  $G(n, p)$ , is static in the sense that the edges are stable in time. However, in many real life scenarios (e.g., wireless ad-hoc/mesh networks) edges can be time varying, i.e.,  $p$  is a function of time  $t$ .

- Literature review on existing time varying random graph models
- Analyze the properties of a random graph model  $G(n, p(t))$ 
  - Degree distribution
  - Local clustering coefficient
  - Connectivity and average path length

### 3) **Joint in- and out- degree distribution in directed networks**

Typically directed networks are characterized by two separate degree distributions, that is, the in- and out- degree distributions. The latter do not

say the whole story about the network. For instance, we are not able to study possible correlations between the in- and out- degree of nodes. In order to study similar problems we need to define the joint degree distribution  $p_{jk}$ , which is the probability that a randomly selected vertex of the directed network will have in-degree equal to  $j$  and out-degree equal to  $k$ .

- Literature review on existing studies on the joint-degree distribution of directed networks
- Using appropriate data from directed networks:
  - Compute the joint degree distributions
  - Identify correlations between the in- and out- degrees of nodes
  - Study the (directed) clustering coefficient as a function of the joint degree distribution
  - Model the joint degree distribution with appropriate bivariate probability distribution

#### 4) Effect of initially infected point on epidemic spread

There is a large volume of research on epidemic spread and the network structure. The main result can be summarized from the fact that in random graphs there exists a positive epidemic threshold, while for scale free networks this threshold is zero (that is, the epidemic always spreads). However, these results can be viewed as *average* results. Even in a scale free network, there can be instances that depending on the network characteristics (e.g., centrality, degree etc.) the initially infected set of nodes the epidemic dies out before spreading in the whole network.

- Literature review on existing studies related with the effect of the initially infected nodes on an epidemic spread
- Either analytically or through simulations (considering SI and SIR models), identify the critical network characteristics of the initially infected nodes at the spread of the epidemic
  - We are mainly interested in scale free topologies and in random graphs at the regime above the epidemic threshold

#### 5) Study of collaboration networks

Many times implicit network structures can reveal a number of interesting things. For instance, by analyzing co-authorship data we can create collaboration networks, where two authors are connected if they have co-authored at least one paper.

- Literature review related with collaboration networks and summarization of main findings
- Analysis of the social properties of the collaboration network
- Study of correlation between scientific productivity and network characteristic of a scientist
- Study of assortativity matching in the collaboration network

#### 6) Navigability in real networks

One of the major properties of real-world social networks is that they exhibit the so-called small-world effect. The latter essentially describes the fact that two individuals are usually within a small number of hops in the social graph. However, how can one find these paths? The famous and ingenious experiment of Stanley Milgram has shown that people are indeed able to find these short paths. Much later computer scientist try to model the way people find these paths and came up with some models/algorithms.

- Literature review on navigability in networks
- Study the applicability of Kleinberg's model on real world social networks (appropriate data for this project will be provided upon request)

#### 7) **Location-based social networks**

In location-based social networks people are connected not only through their declared friendships, but (implicitly) also with regards to the locations they visit. Places can also be associated with each other if we consider the flow of people from one place to another.

- Literature review on related studies with your project
- Analyze the location trails of users to construct and study co-location graphs of users
- Compare co-location graphs with social graph and identify possible ways to rate friendships
- Study assortativity mixing with regards to locations
- Analyze graphs where vertices are venues and connections between venues exist depending on the flow of people among these places

#### 8) **Signed networks**

Networks do not always need to capture positive relationships (e.g., friendships, trust). They can also represent negative relations (e.g., enemies, distrust).

- Literature review on related studies on signed networks
- Analysis of "signed" network characteristics
  - E.g., signed degree distribution, average positive path length etc.
- Study of the structural balance theorem on the datasets considered
- Prediction models for the sign of a social link

#### **Data Sources:**

- a) <http://snap.stanford.edu/data/index.html>
- b) <http://www.network-science.org/>

#### **Structure of report:**

The report should be written in a research paper format, preferably using LATEX. It needs to include an appropriate title, abstract, introduction and related work section along with your work. Language needs to be as clear and formal as possible. Given that the project consists 60% of your final grade it needs to be of high quality.