SOCIAL NETWORK ANALYSIS

Mentors (Jadavpur University) –

- Professor Sarbani Roy
- Subhayan Bhattacharya

Students (Indian Institute of Engineering Science and Technology, Shibpur) –

- Anupam Sanidhya
- Shivang Srivastava

TOPICS –

- Information Propagation

Prerequisites –

Knowledge of python programming language and its various libraries such as Numpy, Matplotlib and NetworkX, which were further used for the study and implementation of various types of graphs, such as simple, bipartite, directed, undirected, multi, disconnected, etc.

Properties of Graphs –

The following properties of graphs were studied –

1.)**Density :** The proportion of direct ties in a network relative to the total number possible.

2.)**Average Path Length :** Average distance between pairs of nodes.

4.)**Eccentricity :** The eccentricity of a node 'v' is the maximum distance from v to all other nodes in the graph.

3.)**Diameter :** The maximum eccentricity of any node in the graph, which is, the greatest distance between any pair of nodes.

4.)**Radius :** The minimum eccentricity of a graph.

5.)**Center :** The set of nodes with eccentricity equal to the radius.

6.)**Centrality :** It refers to a group of metrics that aim to quantify the "importance" or "influence" (in a variety of senses) of a particular node (or group) within a network. The types of centrality studied are listed below –

a.)Degree Centrality : The degree centrality for a node is the fraction of nodes it is connected to.

b.)Betweenness Centrality : Betweenness centrality of a node v is the sum of the fraction of all-pairs shortest paths that pass through v.

c.)Closeness Centrality : Closeness centrality of a node v is the reciprocal of the sum of the shortest path distances from v to all other nodes except v. Higher values of closeness indicate higher centrality.

d.)Eigen Vector Centrality : Uses the power method to find the eigenvector for the largest eigenvalue of the adjacency matrix of a graph.

7.)**Clustering Coefficient :** It is the measure of the degree to which nodes in a graph tend to cluster together. The types of clustering coefficients are described below –

a.)Local Clustering Coefficient : The local clustering of each node in a graph is the fraction of triangles that actually exist over all possible triangles in its neighbourhood.

b.)Global Clustering Coefficient : It is the number of closed triplets of nodes (or 3 x triangles) over the total number of triplets (both open and closed).

c.)Average Clustering Coefficient : It is the mean of local clusterings in a graph.

8.)**Assortativity :** It measures the similarity of connections in the graph with respect to the node degree. It was mostly found to be negative.

- Properties of a few different Graphs (Jupyter Notebook)
- Comparison of properties of graphs

References –

a.) Characterizing the structural diversity of complex networks across domains

Kansuke Ikehara (i) and Aaron Clauset (i,ii,iii)
(i.)Department of Computer Science, University of Colorado, Boulder, CO, USA
(ii.)BioFrontiers Institute, University of Colorado, Boulder, CO, USA
(iii.)Santa Fe Institute, Santa Fe, NM, USA

b.) Classification of Social Network Sites based on Network Indexes and Communication Patterns

F.Toriumi (i.), I.Okada (ii.), H.Yamamoto (iii.), H.Suwa (iv.), K.Izumi(v.) and Y.Hashimoto (v.)
(i.)Nagoya University, Aichi, Japan1
(ii.)Soka University, Tokyo, Japan2
(iii.)Rissho University, Tokyo, Japan4
(iv.)University of Electro-Communications, Tokyo, Japan4
(v.)The University of Tokyo, Tokyo, Japan5

c.) Relationship Classification in Large Scale Online Social Networks and Its Impact on Information Propagation

Shaojie Tang (i.), Jing Yuan (ii.), Xufei Mao (iii.), Xiang-Yang Li (i.), Wei Chen (iv.), Guojun Dai(v.)
(i.)Department of Computer Science, Illinois Institute of Technology
(ii.)Department of Computer Science, NanJing University
(iii.)Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia Beijing University of Posts and Telecommunications
(iv.)Microsoft Research Asia
(v.)Institute of Computer Application Technology, Hangzhou Dianzi University

d.) Analysis of Topological Characteristics of Huge Online Social Networking Services

Yong-Yeol Ahn
Department of Physics
KAIST, Deajeon, Korea
yongyeol@gmail.com

Seungyeop Han
Division of Computer Science
KAIST, Daejeon, Korea
syhan@an.kaist.ac.kr

Haewoon Kwak
Division of Computer Science
KAIST, Daejeon, Korea
syhan@an.kaist.ac.kr

Sue Moon
Division of Computer Science
KAIST, Daejeon, Korea
sbmoon@cs.kaist.ac.kr

Hawoong Jeong
Department of Physics
KAIST, Deajeon, Korea
hjeong@kaist.ac.kr

e.) Classes of small-world networks

L. A. N. Amaral, A. Scala, M. Barthe´le´my, and H. E. Stanley
Center for Polymer Studies and Department of Physics, Boston University, Boston, MA02215

f.) Classification of scale free networks

K.-I. Goh (i.), E.S. Oh (i.), H. Jeong (ii.), B. Kahng (i.), and D. Kim (i.)
(i.)School of Physics and Center for Theoretical Physics, Seoul National University, Seoul 151-747, Korea
(ii.)Department of Physics, Korea Advanced Institute of Science and Technology, Daejon 305-701, Korea

g.) Maximizing the Spread of Influence through a Social Network

David Kempe
Dept. of Computer Science
Cornell University, Ithaca NY
kempe@cs.cornell.edu

Jon Kleinberg
Dept. of Computer Science
Cornell University, Ithaca NY
kleinber@cs.cornell.edu

Ev´a Tardos
Dept. of Computer Science
Cornell University, Ithaca NY
eva@cs.cornell.edu

Information Propagation

Information propagation is the process by which any data is spread over the networks by the means of communication among the social entities. There are two widely used models of information propagation –
(i) Linear Threshold Model
(ii) Independent Cascade Model

Linear Threshold Model –

- A node *v* has random threshold ~ θ_v *[0,1]*
- A node *v* is influenced by each neighbor *w* according to a *weight* $b_{w,v}$ such that $\sum_{w \text{ neighbor of } v} b_{w,v} \leq 1$
- A node *v* becomes active when at least (weighted) θ_v fraction of its neighbors are active $\sum_{w \text{ active neighbor of } v} b_{w,v} \geq \theta_v$

Independent Cascade Model –

- When node *v* becomes active, it has a single chance of activating each currently inactive neighbor *w.*
- In spite of its success, the same node will never get another chance to activate the same inactive neighbor and the activation attempt succeeds with probability p_{vw}.
- The deterministic model is a special case of IC model. In this case, p_{vw}=1 for all (v,w).

A few real life network data were assessed such as that of Facebook NIPS, Wiki-Vote, Epinions, Avogato, Slash Dot, etc.

- Code for Independent Cascade Model
- Code for Linear Threshold Model
- Implementations of the models on real life networks

The weights, thresholds and in the models were generated randomly with values ranging between 0 and 1 and probabilities were generated according to the weight distribution between infected and target nodes.

In general, Independent Cascade Models seemed to perform better than Linear Threshold Models as the former would affect a higher percentage of nodes.
This was because in LTM, even with the influence of more than one neighbour, a threshold value had to be overcome which might not have been possible in some cases. While in ICM, more the number of neighbours meant more probabilities of getting infected for a particular node.