CSI 445/660 – Network Science – Fall 2015

Suggestions for Course Projects

Rules for Course Projects:

1. Each project team may consist of *up to two* students.
2. The same project may be carried out by different teams.
3. Each team should do a demo of the completed project and submit a short report discussing the project and the results obtained.
4. All the project demos must be completed by 5 PM on Monday, Dec. 14, 2015. Project reports (along with source code) must also be turned in by that time.

Project Suggestions:

The following are brief outlines for some course projects. Students should meet with Ravi to finalize the details regarding these projects.

In addition to the projects mentioned below, students may propose other projects. However, they should discuss their project proposals with Ravi to make sure that the projects are appropriate for the class and that they can be completed by the deadline mentioned above.

Suggestion 1: Generation and properties of large **directed** graphs under the Erdös-Renyi (ER) random graph model.

The ER model of directed random graphs is similar to the undirected model except that for any pair of nodes x and y, both the directed edges (x, y) and (y, x) are considered independently and added with the chosen probability p. A goal of this project is to generate random directed graphs for various combinations of values of n and p and experimentally study the impact of various values for the product np on the properties of the resulting directed graphs. The experiments should try all three ranges for the product np, that is, $np < 1$, $np = 1$ and $np > 1$. For these combinations, measures that will be of interest include (i) the number of nodes with indegree 0, (ii) the number of nodes with outdegree 0, (iii) the size of a largest weakly connected component, (iv) the size of a largest strongly connected component and (v) the number of directed cycles of length k (for small vales of k).

Students may choose any reasonable subset of measures and study how these measures are affected by the various ranges of values for np.

Additionally, students may also compute and report how for a given n, the value of p affects other parameters (e.g. maximum number of nodes reachable from any node, maximum depth of breadth-first-search tree, the size of a largest strongly connected component).

Suggestion 2: Generation and properties of large **directed** graphs whose indegrees have a power-law distribution.

An algorithm (from Chapter 18 of the text by Easley & Kleinberg) based on the concept of preferential-attachment for generating large directed graphs (where the indegrees of nodes have a power-law distribution) will be discussed in class. The goal of the project is to implement the algorithm for various combinations of the values of the parameters n (the number of nodes) and p (the probability value used in deciding whether the copy step is performed) used by the algorithm and study how well the generated graphs satisfy the predicted power law behavior.

Additionally, students may also compute and report how the value of p used by the algorithm affects other appropriate directed graph parameters (e.g. maximum depth of a BFS tree, the maximum number of nodes reachable from any node, the number of weakly connected components).

Suggestion 3: Generation and properties of large **undirected** graphs whose degrees have a power-law distribution.

An algorithm (due to Albert and Barabasi) based on the concept of preferential-attachment for generating large undirected graphs (where the degrees of nodes have a power-law distribution with exponent 3) will be discussed in class. The goal of the project is to implement the algorithm for various combinations of the values of the parameters n (the total number of nodes), m_0 (the initial number of nodes) and m (the number of edges added at each step) used by the algorithm and study how well the generated graphs satisfy the predicted power law behavior.

Additionally, students may also compute and report how the value of m used by the algorithm affects other graph parameters (e.g. size of giant component, average clustering coefficient, diameter of a giant component).

Suggestion 4: Generation and properties of large **undirected** graphs generated using the Watts-Strogatz model.

An algorithm (due to Watts and Strogatz) based on the concept of rewiring for generating large undirected small-world graphs will be discussed in class. The goal of the project is to implement the algorithm for various combinations of the values of the parameters n (the total number of nodes), K (the average node degree) and β (the rewiring probability) used by the algorithm and study how well the generated graphs satisfy the predicted small-world property.

Additionally, students may also compute and report how the values of K and β used by the algorithm affect other graph parameters (e.g. average clustering coefficient, average distance between a pair of nodes and diameter of the graph).

Students may also implement and study the Newman-Watts modification of the above method by starting with some other initial network (e.g. a grid).

Suggestion 5: Study of **hitting times** for undirected graphs under the ER model.

Recall that the hitting time for a node pair (u, v) is the expected number of edges used under the uniform random walk model to go from u to v. The goal of this project is to study for a given n, how the parameter p used in the ER model affects the maximum and average hitting times of a giant component generated under the ER model.

One can also carry out hitting time studies for other undirected graph models such the power-law graphs generated by the Albert-Barabasi algorithm or the small-world networks generated by the Watts-Strogatz model.

Suggestion 6: Study of diffusion in networks.

Several models for diffusion of contagions in networks (e.g. threshold models, SIR model, Voter model) will be discussed in class. For any one of these models, students should carry out the following simulation-based study.

(a) Choose a giant component of an undirected or directed ER network and another undirected or directed network with a power-law degree distribution. (The two networks must have the same number of nodes.)

(b) For a given number s of randomly chosen **seed** nodes (i.e., nodes which are initially infected), determine through simulation the number of infected nodes at the end. (The number s of seed nodes should vary over a range.)

(c) For a given number s of seed nodes which are chosen to be the s nodes of maximum degree in the network, determine through simulation the number of infected nodes at the end. (The range for s should be the same as in (c).)

Students may also experiment with other methods of choosing seed nodes.

For models involving transmission probabilities, the simulation experiments should also vary this probability value over a suitable range.

Suggestion 7: Construct a social network and determine some of its centrality values.

The goal of this project is to construct a reasonably large social network (with a few thousand nodes) using social media data and determine several of its centrality parameters. If the nodes correspond to people, it will be interesting to identify the top-k most central people in the network according to the centrality measures used. It will also be of interest to determine whether the degree distribution for the network has the power-law property.

The methodology used in this project may be similar to that used in the following paper (available through the class home page):

> E. Yan and Y. Ding, "Applying Centrality Measures to Impact Analysis: A Co-authorship Network Analysis", *J. American Institute of Information Science and Technology*, Vol. 60, No. 10, 2009, pp. 2107–2118.
