Final Project Ideas

Note: online social networks data that can be used in final projects are available at: http://www.cse.usf.edu/~jhblackb/datasets.html

1. For a set of real networks (some social, some biological or technological), **compute the correlation between edge betweenness for every pair of nodes and the overlap between their neighborhoods**. In *"Structure and tie strength in mobile communication networks", PNAS, May 1, 2007, vol 104, no. 18* it is suggested that these metrics are inversely correlated. If this is true for a representative set of networks, then the overlap (which is a local metric) can be a good estimation of edge betweenness (which is a global metric and extremely computationally intensive to compute).

   Notes:
   - This project seems to be appropriate for the MapReduce/Hadoop infrastructure.
   - Datasets:
     - Legislature data (weighted)
     - Facebook interactions (weighted)
     - Flickr data (weighted)
     - IMDB (from imdb.org)

2. **Path structure and strength of tie in the legislature data** (variant of forbidden triad). The basic idea is that the greater the number of two paths connecting a pair of nodes and the stronger the ties are in the two paths, the more likely it is that the pair is connected by a tie and as number of two paths and their strength increases so too does the strength of the tie connecting the pair. Data from legislators and the bills they sponsor and cosponsor can provide weights, w_{ij}, the number of ties legislator i cosponsors bills introduced (sponsored) by legislator j. Interesting complications arise because the ties are directed – the two paths from i to j through a set of k's could impact the ij tie but also the ji tie.

3. Compare the centrality measures of different nodes in a Facebook social graph to understand the topology and the position of a node in the graph.

   Notes:

   - More insightful discussions can come from comparing results on various graphs. Interpret the differences.

4. Propose a hypothesis to test, collect data from a relevant online source, and evaluate the hypothesis.

   Examples of hypothesis:
   - Diffusion hypothesis: is there some kind of adoption? Identifying trends, etc
   - Homophily-driven behaviors: similar behavior because their attributes are similar. Selection vs. influence.

- Horizon of observability evaluation through joining causes in Facebook: collect data for understanding if joining causes is influenced by existing relationships. Note that you will need time information in your dataset.

Components:

- Understand/decide what is the social network information to collect (identify nodes, edges, direction if directed network, weights if weighted network, and other information you need for your study).
- Decide how to collect the information (if data not readily available). For example, write a crawler, a Facebook application, etc. Identifying the correct data is a team task: the SNA team members are expected to contribute significantly on this component; the CS team members will assess if the collection of data is possible.
- If the data is really huge (e.g., Facebook) trivial sampling should not be statistically meaningful. You may want to use Burt and Ronchi's 1994 "Measuring a Large Network Quickly" or other studies.
- Evaluate hypothesis using a significantly large dataset (or/and different datasets) on a parallel or distributed environment studied in the CS class.

5. Confirm or infirm a result from social network analysis (such as, for example, Friendkin's Horizon of observability or the forbidden triad hypothesis) using much larger and *mediated* social networks (e.g., online social networks or other content producing/sharing systems). An overarching question for your investigation could be: do mediated networks have different characteristics than what is accepted in traditional SNA?

   Some examples:

   - Verify the forbidden triad result on some large online social network.
   - Verify the horizon of observability claim that people are not aware of people who are further than 2 hops in their social graph:
     - Intuition is that the facility of browsing (given enough time to waste) would allow people to learn more about distant nodes.
     - Would writing a Facebook application be useful to see if users look at the profiles or activity of friends of friends of friends of …?

6. Identifying ring voters in a community such as `reddit.com` or `dig.com`. The original problem was posed as a job interview question (http://www.thesixtyone.com/#/info/settings/jobs/). The problem, however, is relevant for other contexts, as well, such as eBay or rating posts in any product review listing.

   > Excerpt from the original job interview question: A social news website called 'Reddigg' has hired you as a consultant to help them with a potentially serious problem. Reddiggers submit news articles in hopes that their submission will make it to the front page -- an article has a chance of getting posted once it receives enough upvotes.

Reddigg suspects spammers have found a way to manipulate the system by commanding fake users to regularly vote for their own articles, hence forming 'voting rings'. You've been hired to identify suspected voting rings based on recent user data.

Your task may not be as straightforward as it seems however. The caveat is that the spammers may have fashioned their sock puppets to act like real users. To create some misdirection, fake users may sometimes pass on voting for targeted articles or even vote on articles that they have no association with.

Examples of other projects did in the past (the first was published):

7. Political Change in North Africa and the Middle East? (proposed by Ginger Johnson)

Tunisia – December 17th 2010: The trigger was December 17's Jasmine Revolution, when a young man set himself on fire in Sidi Bouzidi, south of Tunis. For the next four weeks, demonstrations intensified — morphing from economic grievances to governance and corruption, and helped by damaging disclosures by WikiLeaks. It ended Zine El-Abidine Ben Ali's 23-year reign.

Southern Sudan – January 9th-15th 2011: A referendum took place in Southern Sudan early in 2011 on whether the region should remain a part of Sudan or become independent. The referendum was one of the consequences of the 2005 Naivasha Agreement between the Khartoum central government and the Sudan People's Liberation Army/Movement (SPLA/M) negotiated to stop two decades of civil war. On January 30th, 2011 the referendum commission published preliminary final results with 98.83% voting in favor of independence. The predetermined date for the creation of an independent state is July 9th, 2011 – the newest country on the African continent if Northern President Omar al-Bashir abides by his promise to recognize the vote.

Egypt – January 25th 2011: Egypt has been on edge for almost a year. It was the scene of food riots in 2010 and started out 2011 with Islamist attacks on the country's Coptic Christians. Egypt is also due for elections later this year, and for the first time, Mohammed ElBaradei, Nobel Peace laureate and former head of the IAEA, has emerged as an alternative to the 30-year rule of Hosni Mubarak. Current protestors in the region are calling for the immediate resignation of Mubarak.

Yemen – February 3rd 2011: What seemed like thousands of anti-government protesters gathered near Sanaa University in Yemen's capital early on the morning of February 3rd, a clear indication that many in the country were not placated by President Ali Abdullah Saleh's recent announcement that he would not seek re-election. Many are saying that across the Red Sea, the heady influence of the Jasmine Revolution has taken root in Yemen. Since late January 2011, Yemenis have been repeating a popular Yemeni chant in the streets "Yemen is not another Somalia, but another Tunisia".

In many ways, the change did not actually start from Tunisia. Some analysts point to the referendum in Sudan as the point when North Africans and Middle Easterners

realized that people's power actually works. Communication technology has played a major role in these unprecedented political events. Facebook, Twitter, WikiLeaks and other social media continue to play a crucial role in the protest movements, just as they did in Iran in 2009. Several online news sources have already declared this to be "the first Wiki-revolution."

This project would seek to analyze the role of social media networks in these contemporary political events.

8. The distribution of federal relief after disasters in the U.S. and the role of political influence in the presidential disaster declaration process. There is an extensive online database for disaster declaration requests from 1953-2004 including awards and denials. There is also an extensive online database related to voting patterns of congressional representatives, committee appointments, tenure, and party affiliation. A Social Network Analysis of these relationships may reveal **potential tendencies for disaster awards versus denials based on network ties**. (proposed by Richard Salkowe)