Huynh, Peter

April 22, 2014

Professor Elizabeth White

CSCI 2270: Data Structures

**Colloquium Write-up: Computational Challenges within Large-Scale Social Networks**

Social networks have evolved drastically within the past decade. With the rise of micro media blogs, such as Twitter and Tumblr, and socializing websites, like Facebook, YouTube, or Reddit, it has become obvious why the world is so readily available to communicate with one another through the tap of a screen or the push of a button. However, with the expansion of the these intricate networks, and the progression of self-investment every user places into these media websites, the computation analysis and organization of the user information has rose into a challenge that may be more than placing information into a simple spread sheet or graph. These social media networks have reached a state in computation where information has become intertwined with real life, and it has become a system where one aspect of collected knowledge can be associated towards other pieces of intellectual properties. The world wide web, and all of its resources has evolved into just what its title implies, a web of information spanning across the world, all adherent towards other aspects of its own web. I was curious about how a modern computational analyst would tackle the issue of finding and calculating patterns within a complex social dataset of an individual user’s digital identity, hence I decided to attend the Johan Ugander’s seminar explaining his research about the social network computation.

Within Johan’s presentation of social network analysis, Johan discussed the foundation of a social network’s information, in terms of a mapping structure. Johan stated that an individual’s mapped social network is similar to mankind’s early interpretations of DNA structures, in the sense that they were organic connecting webs, seemingly unsystematic, yet computable, and formed sectioned off loops based on certain criteria. He then gave an example from his research, displaying the graph of an unnamed volunteer’s Facebook social web. The user’s data web represented his digital friendships on Facebook, and how each of his friend correlated with one another. This inorganic web showed cluster zones with all points connecting towards a central point, each point representing a person’s Facebook identity. The central point was the user’s Facebook page, meaning that all points on that web had to connect towards his page and those people were associated with him in some way. What I found to be extremely interesting was how the pattern of the web helped depict social circles and important friends on Facebook. For example, if you carefully study the cluster zones within the web, one cluster represents a friendship circle of the user’s coworkers, and another represented the user’s family and college friends. There was also a sub-central point, which represented the user’s spouse. This visually computed information allowed the analyst to determine who was the volunteer’s coworker, wife, or family friend. It’s just amazing that a simple constructed web allowed for such an interpretation.

Upon viewing this visually interpreted data, I began to wonder how these individual users became aware of other mutual friends on Facebook, and how they even joined social network websites in the first place. According to the Economist’s article, Social Contagion: Conflicting Ideas, social media connections act like a spreading virus, becoming more viral as people are exposed to mutual users that are part of a social network website. The article states that the spread of online ideas, or in this case friendships, start based on a given individual’s exposure to other individual people who have already been exposed to such ideas and friendships. This viral behavior is primarily spread towards the individual through invitations to join the idea, the social network concept. Hence, the reason why those cluster of mutual friendships occur are primarily due to actual social interactions instances and information. Users on Facebook are able to detect possible mutual friends throughout the social site and establish friend requests with actual friends based on a separate individual’s friends in comparison to the friends they have on the site. This system is like comparing two social webs, and seeing which web points, or nodes, are the same person. Then, when there is a match-up of a certain set of individuals, or a certain amount of related individual nodes, Facebook establishes a list of potential mutual friends. The connection of mutual friends and new social network users mimic a similar pattern to a contagion, in the sense that information and friend connections establish based on spreading knowledge and data.

Another aspect of Johan’s research was the correlation between mutual friends of the volunteer, through data only gathered directly from the volunteer’s Facebook information. Johan stated that almost everything is computationally tractable, and then displayed a three-dimensional sparse graph depicting thousands of small, dense points representing the frequency of people’s relationships with one another in comparison to the type of relationship they had with only another, digitally. For example, one axis of the graph displayed a frequency of three people who were directly friends with each other, forming a triangular relationship, while the other axis showed a frequency of three mutual friends who had no relation to one another, forming a three point node system, with no connecting nodes. Upon carefully examining the graph, it can be easily seen that the triangular relationships formed farm more frequently as the social web was closer towards the center of a social cluster on the web, while the three unconnected node system often appeared more frequently between two social clusters or entirely away from a cluster. I find this frequency graph rather fascinating. Johan was able to gather data from data to create this graph. This means he determined the nature of social connections relating to the data of social connections towards one person. This meta-analysis allowed for him to find a pattern in Facebook relationships of separate individuals from a single individual.

One more notable example node system, given a specific name by Johan, is called the Forbidden Triad. This node system type is unique for its complex occurrence rate and appearance frequency. The Forbidden Triad is a social relationship case where three points within the social system are connected to a single, separate point, but do not form a full triangle relation. This means that there is a single individual that has a mutual friend with two other individuals, but the individuals are not mutually connected. Johan then goes on to explain that typically the majority of these Forbidden Triads then to form triangular relations, if given time, but these are rare occurrences within his research, granted because he only gather the data at a set instance. He states that the pattern of this relation type appears in the entirety of the social network web diagram and mostly does not propagate into the central part of the web.

This tidbit of thought had me pondering about the closeness of each relationship between Facebook users. Though Johan was able to derive the frequencies of mutual relationship with his data computing, I was unsure if he was able to confirm the solidity of these node system types or the individual’s true relationships between Facebook friends, just through social network computing. As an article on the popular technology media website Wired, titled Facebook Study: It’s a Small(er) World After All, stated: “Facebook’s study found that instead of the average of six degrees of separation between each of us [You and your friends], Facebook users are separated by an average of four degrees.” This research also states that the degree that separates people across states, countries, and continents lesser, with local separation showing the most impact. According to this study stated on Wired, people forge stronger relationships on social media than they do in reality. This leaves me to wonder, does the mutual connection between two individual friends on a social network website correlate to actual situations in real life? Although the degree of social separate is stated to be smaller, this indicates a difference from reality and the digital world, which means the computation of social networks are not directly corresponding to real life instances. Or, does the fact that social relations become stronger online solidify the computation research of social networks towards real life? I suppose the real question I have about this research is its application outside of the digital pattern recognition and algorithmic interpretation. Does this research apply further than finding patterns in a social network market, is it able to confirm real life relations from digital connections?

One other notable aspect of this seminar I found particularly interesting was the friendship paradox introduced by Johan’s research. As cleverly stated within the Opinionator New York Time’s article based on Johan’s research, Friends You Can Count On, “The average number of friends of friends is always greater than the average number of friends of individuals. This phenomenon has been called the friendship paradox.” This is a prime example of the massively interconnected viral web that social media and networks have established with their collective information. With this paradox, it is possible to trace the interconnected social web an endless amount of degrees away from the central point, the targeted user, and find nearly every possible active user within that social site. This implements a possible issue within the computation of social networks: To what degree point in the analyzed information of a user become arbitrary and irrelevant to the targeted user? This imposed problem is difficult to answer, since the branching network data has to continue expanding outward in degree in order to establish a full perspective of a single user’s mutual relations. However, computing when the stop the data branching was not explained within this colloquium, and I do not understand how it should exactly works nor when to implement such a process.

There are also other aspects of the lecture that I couldn’t quite grasp my mind around. For example, there was a moment where Johan explained the physical process and efficiency of his research. He explained that the graphs used for Facebook’s friend suggestion system ran on seventy eight different machines in order to establish a list of possible mutual friend for Facebook users. Each machine processed a different multitude of queries, and he explained that this process reduced fifty percent of time usage, bandwidth, and sixty three percent of incoming traffic. He also explained the algorithm of his research, and the logic behind this algorithm, but it was hard for me to understand it at that current moment. I attempting to do a bit of research behind his sorting and computing algorithm, the Label Propagation algorithm, but found it hard to grasp with my current understanding of computational sciences.

Overall, this computation analysis on social networking datasets has been rather interesting, and it is clear how applicable it can be within the media industry. The pressing demand of structural computation on social data can also be reapplied in other form of data as well. For example, one analyst can use these methods of evaluating targeted information and expanding relations to determine frequencies of patterns and cluster patterns in conjunction to other abstract concepts. If someone were to use this method of analysis to research the patterns of actual contagions, the results can be yielded in the same manner, except within a different context and subject matter.

**Bibliography**

Isaac, Mike. "Facebook Study: It’s a Small(er) World After All" *Wired.com*. Conde Nast Digital,

20 Nov. 0011. Web. 13 Feb. 2014. <http://www.wired.com/2011/11/facebook-social-

graph-study/>.

Strogatz, Steven. "Friends You Can Count On." *Opinionator New York Times*. New York Times,

17 Sept. 2012. Web. 8 Apr. 2014. <http://opinionator.blogs.nytimes.com/2012/09/17/

friends-you-can-count-on/?\_php=true&\_type=blogs&\_r=0>.

T.C. "Conflicting Ideas." *The Economist*. The Economist Newspaper, 03 Apr. 2012. Web. 20

Feb. 2014. <http://www.economist.com/blogs/babbage/2012/04/social-contagion>.

Ugander, Johan. *Vimeo*. Proc. of Johan Ugander on “Computational Challenges in Large-Scale

Social Networks”, Colorado, Boulder. CU CS Colloquia, 11 Mar. 2014. Web. 11 Mar. 2014. <http://vimeo.com/88824384>.

Ugander, Johan. *Johan Ugander's Homepage*. Cornell University. Web. 12 Feb. 2014.

<https://people.cam.cornell.edu/~jugander/>.