CAP 6315: Social Networks and Big Data Analytics

Term project proposal guideline [Proposal due by March 17]

The goal of the term project is to allow students to practice knowledge learnt from the class and work on specific research projects during the second period of the class.

The final outcomes of the project will be turned into a formal technical report. Students can choose to use a term project report to substitute the final exam, or they can choose to take final exam and submit a simple term project (details will follow).

Students using term project to substitute final exam: Students can choose to substitute the final exam using term project. If this is the case, the term project will contribute 15+20=35 point to your final. The report shall have 4000 words minimum including motivation and description of the research problems, technical solutions, validations, conclusion, and reference etc. (I will have detailed requirement posted online soon). I will assist each student (or each team) to polish the report and try to find a suitable venue to publish the report, if possible.

Students participating final exam: For students who still want to take the final exam, the term project will contribute 15 points to your final. Your report shall have 1000 words minimum, including statement of the research problem, designs, and validations. (I will have detailed requirement posted online soon).

Type of project: Two types of term projects are available. The first type of project is suitable for students who have rather limited programming knowledge, and the second type of project is more suitable for students with sufficient programming skills and also want to challenge themselves in specific research topics.

Team: Students are free to choose any type of project suitable for their own background. They can form a team to work on the term project, with each team consisting of maximum 3 students.

Term Project Proposal: [Proposal Deadline: March 17] To help each student better prepare for the term project, I require that each student to submit a brief description of your term project (with title, 2-3 sentences project aim, and team members (if any)). [1 point]

Term Project Examples

Please note that the following topics are only for your reference. You can suggest any other topic in your mind.

Category I: Literature Review Tasks

- **Social Network Marketing Strategy:** Social networks are now becoming a major source of online advertising. Many companies and individuals are now selling in social networks, and at the very least level, they have built a profile in the social network in order to provide effective and efficient customer interactions. You can write a review paper to summarize existing marketing strategies in social networks. Your survey can provide answers to many important questions, such as what type of marketing strategies is commonly used in different social networks, what are the key aspects of marketing in social networks?
- **Social Network Applications and Case Studies:** The rising of the social networks has resulted in many exciting applications, where people use information collected from social network sites, such as Twitter, to analyze and predict social events. For example, there are applications which use Twitter data to predict earthquakes, stock market price etc. You can write a review paper to summarize some most recent exciting applications which use social network data.
- **Review of Social Network Sites and Resources:** Online social network sites evolve very fast, and there are also many tools/algorithms/data sources available for research or education purposes. You can write a review paper to categorize and summarize main stream social network sites, and then outline major sources available for public usages. (*this type of project can focus on specific domains such as social network site and resources for health, animal sciences etc*.).
- **Privacy in Social networks and the treatment**: When users are sharing their information through social networking sites, they may be at risk of leaking privacy information and be vulnerable to criminal or other attacks (especially for teenagers or minors). You can write a review paper to summarize the type of privacy issues in social networks, common mistakes and risks about privacy in online social networks, and explain data analytics tools/algorithms/treatments which can be used for improve the privacy for social networks.
- **Cyberbulling in Social Networks:** Cyberbulling is bulling that use electronic devices/technologies. This type of actions has becoming a major issue, due to the increasing popularity of the social networking tools and devices. You can write a survey paper to define types of cyberbulling, the characteristics of cyberbulling, and analytics method for discovery and treatment of the cyberbulling methods.
- **Literature Review for Finding Influential Nodes in Social Networks:** In social networks, each individual has respective role in terms of his/her social influence. In the class, we mainly discussed social propagation models and method to find a target set with maximum social influence. You can write a literature review to summarize existing methods in characterizing the social influence of the nodes in the network.
- **Literature Review for Node Similarity Measures in Social Networks:** We have studied quite a few node similarity measures in the class, and there are many other measures which can be

used to assess node similarities (for example, a group at CMU has been focusing on studying network node similarity for many years: http://www.cs.cmu.edu/~dkoutra/tut/sdm14.html). You can do a literature review to survey methods/algorithms/measures which focus on network node similarity.

Category II: Programming Tasks

- **Consistency Validation between Structure-based Node Similarity and Content-based Node Similarity:** In citation networks and many other types of networks, each node may contain detailed node content (such as the keywords of a paper). Therefore, we can use node features to calculate the similarity between two nodes. This project will intend to study to which extend the structure-based node similarity is consistent (or inconsistent) with the node content similarity. You can study node characteristics (such as node degrees, node centrality scores etc), and further conclude what are the impact factors which will make two similarities (structure-based vs. node-based) to be more consistent (or more inconsistent).
- **Outlier Detection in Social networks:** An outlier is an instance which has a large offset with respect to the majority population. In social network setting, an outlier may be a node whose connections are inconsistent to the node content. You can design a network outlier detection method to find special nodes in the network, and may further rank all nodes in the network to specify their outlier scores.
- **Gender/age Prediction in Social Network:** In social networks, the content of the node and the topology structures of the nodes (e.g., the node degree and the centrality scores) may help identify the gender (or age) information of the node. You can design a new algorithm to predict the biometric information of the node in the social network.
- **The robustness/sensitivity of the network measures with respect to the changes of the network:** We have introduced a number of node based criteria to assess the centrality scores, betweeness of the nodes, etc. The main objective is to study whether these measures are sensitive to the random changes of the network. You can randomly add or remove some network edges, and then evaluate the robustness of the node centrality scores with respect to the changes introduced into the network.
- **The robustness/sensitivity of the community detection algorithm with respect to the random noise:** The main objective is to study whether one or multiple community detection algorithms are sensitive to the random changes of the network connections. You can randomly add or remove some network edges, and then evaluate the robustness of the community detection results with respect to the randomness introduced into the network.
- **Supervise Learning for Link Prediction:** In the class, we have learned methods useing supervised learning methods for link/edge prediction. You can implement a similar framework for social network link prediction. You will need to build a training set with each instance representing one edge, and feature values include some characteristics of the edges (such as the common

neighbors of the two nodes between the edge). You can build a dataset to test the algorithm performance, by using different features and different learning algorithm, for link prediction.

Some Useful Public Social Network Datasets:

1. Stanford Social Network Repository

http://snap.stanford.edu/data/

2. UCI Network data repository

https://networkdata.ics.uci.edu/index.php

3. University of Arizona social network repository

http://socialcomputing.asu.edu/pages/datasets

4. University of Maryland Networked Data repository (containing node labels)

http://linqs.cs.umd.edu/projects//projects/lbc/index.html

Submission:

Please submit your proposal via BB by Thursday March 17. The proposal should include following information.

1. The title of the project.
2. A brief description about what you intend to do in the project. [2-3 sentences]
3. Name of the team members, if you decide to team up with other students.