[Project #1] WallStreetBets

Reddit forum WallStreetBet had rallied the day-traders in the past year for holding and investing in certain stocks on WallStreet. This resulted in a global phenomenon and also in a chaos in the stock markets for some bigger companies. In this project, you can take an explorative approach to find out the role played by this community in relation to the stock prediction, along with how the users campaigned on the forum e.g. is there specific users or groups of users (sub-communities) coordinating together who are more influential on other people's opinion? Any other well-motivated and interesting ideas are also welcome.

Dataset: Available (Reddit), students are encouraged to contribute in data collection

Required skills to successfully complete the project during the course

- You will need to be self-motivated and proactive, which means you should be able to set your own milestones and finish them on time.
- Strong interest in analyzing data and finding common or new patterns from it
- Programming knowledge (Python, R) is a plus
- Data, network analytics tools (SNAP, Gephi, NetworkX) is a plus
- Web development skills (HTML, CSS, JS) is a plus.

Benefits

- You will be working with one or more PhD students in the lab, who will guide you during the project.
- Opportunity to gain hands-on experience in Social Network Analysis and Machine Learning research
- Opportunity to tackle a challenging and novel problem that no one has solved before
- Opportunity for having a top-tier academic publication, depending on the quality and the amount of novel contribution of the project
- Opportunity to gain experience in data visualisation tools.

Contact: Students are welcome to contact Ehsan (euhaq@connect.ust.hk) to further discuss the project or general idea brainstorming discussion related to other proposals.

[Project #2] Rumors/Trolls Detection during COVID-19

Rumors and trolls are one of the most used mediums to sway public opinions. In this project, you can use the data from public discussions related to COVID-19 to detect such rumors shared by users. This work can be effective, in general, in handling the fake news and has a real life impact. We have already labelled some of the rumors based on keywords specific heuristics.
You can improve on the detection model, and for example, study the propagation of such rumors on social media.

Dataset: Available (Twitter)

Required skills to successfully complete the project during the course

- You will need to be self-motivated and proactive, which means you should be able to set your own milestones and finish them on time.
- Strong interest in analyzing data and finding common or new patterns from it
- Programming knowledge (Python, R) is a plus
- Data, network analytics tools (SNAP, Gephi, NetworkX) is a plus
- Web development skills (HTML, CSS, JS) is a plus.

Benefits

- You will be working with one or more PhD students in the lab, who will guide you during the project.
- Opportunity to gain hands-on experience in Social Network Analysis and Machine Learning research
- Opportunity to tackle a challenging and novel problem that no one has solved before
- Opportunity for having a top-tier academic publication, depending on the quality and the amount of novel contribution of the project

Opportunity to gain experience in data visualisation tools.

Contact: Students are welcome to contact Ehsan (euhaq@connect.ust.hk) to further discuss the project or general idea brainstorming discussion related to other proposals.

[Project #3] Political Polarisation

Based on the social events happening around us, users carry certain offline opinions that are depicted in their online interactions e.g. retweeting, or mentioning other users in their messages. Such interactions can be modelled as networks and hence used to study political discourse of certain population set. Political Polarisation is one of the major research directions in such discourse analysis. Such polarisation has several real life implications and can be extended to areas such as election predictions or policy making.

Dataset: Available (Twitter), students are encouraged to contribute in data collection

Apart from the data from other media, We also have developed testbed that can be used to study the polarisation of users on news sharing platforms rather than social media.

Required skills to successfully complete the project during the course

- You will need to be self-motivated and proactive, which means you should be able to set your own milestones and finish them on time.
- Strong interest in analyzing data and finding common or new patterns from it
- Programming knowledge (Python, R) is a plus
- Data, network analytics tools (SNAP, Gephi, NetworkX) is a plus
- Web development skills (HTML, CSS, JS) is a plus.

Benefits

- You will be working with one or more PhD students in the lab, who will guide you during the project.
- Opportunity to gain hands-on experience in Social Network Analysis and Machine Learning research
- Opportunity to tackle a challenging and novel problem that no one has solved before
- Opportunity for having a top-tier academic publication, depending on the quality and the amount of novel contribution of the project
- Opportunity to gain experience in data visualisation tools.

Contact: Students are welcome to contact Ehsan (euhaq@connect.ust.hk) to further discuss the project or general idea brainstorming discussion related to other proposals.

[Project #4] De-polarisation system

Polarisation of opinions can lead to the building of echo chambers where users are less recipients to opposing opinions. Currently the recommendation algorithms are designed in a way that enforce personalisation and hence there is a tendency for the echo chambers to be built.

The focus of the project is to build a tool that can let users reach to the opinions that are of opposing viewpoints. The project would focus on using network based approaches to identify the echo chamber and lead users to the information otherwise not reachable.

Our initial testbed is live on the chrome webstore, students will have the opportunity to take part in the project and develop a real life system.

Required skills to successfully complete the project during the course

- We would expect students to know any front end development such as Web, Mobile apps, or Browser extensions etc. Knowledge of python will also be useful.

Benefits

- You will be working with one or more PhD students in the lab, who will guide you during the project.
- Opportunity to gain hands-on experience in Social Network Analysis and Machine Learning research
- Opportunity to tackle a challenging and novel problem that no one has solved before
- Opportunity for having a top-tier academic publication, depending on the quality and the amount of novel contribution of the project

Contact: Students are welcome to contact Ehsan (euhaq@connect.ust.hk) to further discuss the project or general idea brainstorming discussion related to other proposals.

[Project #5]Social Media and Epidemic Modeling (Corona)

We have collected large scale datasets from different social media related to recent CoronaVirus. We would like to study the evolution of users' perception about the epidemic. We would like to see how the perception of disease changes over the time, we will use the social media data, global media coverage, along with the statistics of epidemic in this project.

Dataset: Available (Twitter, Weibo), students are encouraged to contribute in data collection

Required skills to successfully complete the project during the course

- You will need to be self-motivated and proactive, which means you should be able to set your own milestones and finish them on time.
- Strong interest in analyzing data and finding common or new patterns from it
- Programming knowledge (Python, R) is a plus
- Data, network analytics tools (SNAP, Gephi, NetworkX) is a plus
- Web development skills (HTML, CSS, JS) is a plus.

Benefits

- You will be working with one or more PhD students in the lab, who will guide you during the project.
- Opportunity to gain hands-on experience in Social Network Analysis and Machine Learning research
- Opportunity to tackle a challenging and novel problem that no one has solved before
- Opportunity for having a top-tier academic publication, depending on the quality and the amount of novel contribution of the project
- Opportunity to gain experience in data visualisation tools.

Contact: Students are welcome to contact Ehsan (euhaq@connect.ust.hk) to further discuss the project or general idea brainstorming discussion related to other proposals.

[Project #6] Hashtags Evolution and Adaptation

Hashtags have become an inclusive part of online interactions and have the topical representation of the content users are posting. Some hashtags have become globally standard such as ‘selfie’, ‘tbt’. There are some other hashtags that emerge as part of some social movement or some particular event such as ‘metoo’, ‘blacklivesmatter’. Network based approaches can be utilised to predict the virality and contextual usage of hashtags.

Dataset: Available (Twitter, Instagram), students are encouraged to contribute in data collection

Required skills to successfully complete the project during the course

- You will need to be self-motivated and proactive, which means you should be able to set your own milestones and finish them on time.
- Strong interest in analyzing data and finding common or new patterns from it
- Programming knowledge (Python, R) is a plus
- Data, network analytics tools (SNAP, Gephi, NetworkX) is a plus
- Web development skills (HTML, CSS, JS) is a plus.

Benefits

- You will be working with one or more PhD students in the lab, who will guide you during the project.
- Opportunity to gain hands-on experience in Social Network Analysis and Machine Learning research
- Opportunity to tackle a challenging and novel problem that no one has solved before
- Opportunity for having a top-tier academic publication, depending on the quality and the amount of novel contribution of the project
- Opportunity to gain experience in data visualisation tools.

Contact: Students are welcome to contact Ehsan (euhaq@connect.ust.hk) to further discuss the project or general idea brainstorming discussion related to other proposals.

[Project #7] Nationalist behavior in social media

Events related to political turmoil such as protests and wars or disasters at national scale make the online users diverge from their usual behavior. Such behavior can carry latent information that can be used to better understand the people's sentiments and get real time responses from the larger population.

Dataset: Available (Twitter), students are encouraged to contribute in data collection

Required skills to successfully complete the project during the course

- You will need to be self-motivated and proactive, which means you should be able to set your own milestones and finish them on time.
- Strong interest in analyzing data and finding common or new patterns from it
- Programming knowledge (Python, R) is a plus
- Data, network analytics tools (SNAP, Gephi, NetworkX) is a plus
- Web development skills (HTML, CSS, JS) is a plus.

Benefits

- You will be working with one or more PhD students in the lab, who will guide you during the project.
- Opportunity to gain hands-on experience in Social Network Analysis and Machine Learning research
- Opportunity to tackle a challenging and novel problem that no one has solved before
- Opportunity for having a top-tier academic publication, depending on the quality and the amount of novel contribution of the project
- Opportunity to gain experience in data visualisation tools.

Contact: Students are welcome to contact Ehsan (euhaq@connect.ust.hk) to further discuss the project or general idea brainstorming discussion related to other proposals.

[Project #8] Astroturfing campaigns on Social Media

Dataset: Available

Social media has been an effective tool in campaigns for various purposes e.g. political, marketing etc. It also presents an opportunity for people to communicate on it in an indirect manner to sway other people's opinion on certain issues. Such usage of social media is referred to as astroturfing and in recent years it has gotten a very high interest from researchers. We will study and predict any such campaigns during the recent years' Black live matter protests and US-Election datasets.

Dataset: Available (Twitter)

Required skills to successfully complete the project during the course

- You will need to be self-motivated and proactive, which means you should be able to set your own milestones and finish them on time.
- Strong interest in analyzing data and finding common or new patterns from it
- Programming knowledge (Python, R) is a plus
- Data, network analytics tools (SNAP, Gephi, NetworkX) is a plus
- Web development skills (HTML, CSS, JS) is a plus.

Benefits

- You will be working with one or more PhD students in the lab, who will guide you during the project.
- Opportunity to gain hands-on experience in Social Network Analysis and Machine Learning research
- Opportunity to tackle a challenging and novel problem that no one has solved before
- Opportunity for having a top-tier academic publication, depending on the quality and the amount of novel contribution of the project

Opportunity to gain experience in data visualisation tools.

Contact: Students are welcome to contact Ehsan (euhaq@connect.ust.hk) to further discuss the project or general idea brainstorming discussion related to other proposals.

[Project #9] Business Popularity in LBSNs

Location-Based Social Networks (LBSNs), with the most popular being Yelp and Foursquare, have become a vital part of our society due to their assistance on users' needs. For example, foreign tourists in San Fransisco can easily find highly-reputed restaurants even if it is their first time in the city. As LBSNs have accumulated various information about users and businesses within it, this information allows us to solve practical problems of predicting the survival rate of the businesses in LBSNs.

As an extension of the survival rate prediction, can we accurately predict the popularity of the business within LBSNs based on a few initial information on the business? One can use various techniques such as traditional ML algorithms (e.g., Logistic Regression or Random Forest) or advanced DL methods (e.g., Transformers or Graph Neural Networks) or develop a novel DL method. Finally, this work can help new business owners in LBSNs to make their business more prosperous.

Contact: Interested students can actively discuss and brainstorm new ideas as well as perform the analysis while consulting with Reza (rhadimogavi@connect.ust.hk).

Available Datasets: (1) Yelp (2) Foursquare

Required skills to successfully complete the project during the course

- You will need to be self-motivated and proactive, which means you should be able to set your own milestones and finish them on time.
- Strong interest in analyzing data and finding common or new patterns from it
- Programming knowledge (Python, R) is a plus
- Data, network analytics tools (SNAP, Gephi, NetworkX) is a plus
- Data processing frameworks (Pandas, PySpark) is a plus

Benefits of participating in this project

- Opportunity to gain hands-on experience in Social Network Analysis and Machine Learning research
- Opportunity to tackle a challenging and novel problem that no one has solved before

Reference papers (can be a good starting point)

[1] 2018. IMWUT. The Role of Urban Mobility in Retail Business Survival. K D'Silva et al. [2] 2017. WWW. Restaurant Survival Analysis with Heterogeneous Information. J Lian et al.

[Project #10] Interpretable Churn

Location-Based Social Networks (LBSNs), with the most popular being Yelp and Foursquare, have become a vital part of our society due to their assistance on users' needs. For example, foreign tourists in San Fransisco can easily find highly-reputed restaurants even if it is their first time in the city. As LBSNs have become widely used by users, understanding user engagement and predicting user churn are essential to the maintainability of the services. Many researchers largely focus on churn prediction problem to predict users who are about to leave the services. Although previous studies achieve high performance by adopting various ML techniques and DL approaches, we are still far from explaining why those users churn from the services. In other words, the interpretability of the models is limited in prior works.

Can we develop a new method or algorithm in order to enable "Interpretable Churn"? "Interpretable churn" would allow us to provide human-readable explanations to the service providers, maintainers, or governmental decision-makers for their various services of interest.

Contact: Interested students can actively discuss and brainstorm new ideas as well as perform the analysis while consulting with Reza (rhadimogavi@connect.ust.hk).

Available Datasets: (1) Yelp

Required skills to successfully complete the project during the course

- You will need to be self-motivated and proactive, which means you should be able to set your own milestones and finish them on time.
- Strong interest in analyzing data and finding common or new patterns from it
- Programming knowledge (Python, R) is a plus
- Data, network analytics tools (SNAP, Gephi, NetworkX) is a plus
- Data processing frameworks (Pandas, PySpark) is a plus

Benefits of participating in this project

- Opportunity to gain hands-on experience in Social Network Analysis and Machine Learning research
- Opportunity to tackle a challenging and novel problem that no one has solved before

Reference papers (can be a good starting point)

[1] 2019. IMWUT. GeoLifecycle: User Engagement of Geographical Exploration and Churn Prediction in LBSNs. YD Kwon et al.

[Project #11] Active User Contributions in Stack Overflow

Community Question Answering (CQA) platforms have been an important source of knowledge sharing. Users can raise or answer questions and rate other's questions/answers on the websites. Alas, some questions are unlikely to get answered soon. In this project, we want to find out what makes an effective question that attracts crowd attention better and get answered quicker, with higher scores and more contributions from the CQA community memebers. By carrying out quantitative data analysis, we hope to summarize traits of effective questions and users' posting attributes and propose possible guidelines for asking questions more efficiently.

Contact: Interested students can actively discuss and brainstorm new ideas as well as perform the analysis while consulting with Reza (rhadimogavi@connect.ust.hk).
Dataset: Available, students are encouraged to contribute in data collection

Required skills: to successfully complete the project during the course

- You will need to be self-motivated and proactive, which means you should be able to set your own milestones and finish them on time.
- Strong interest in analyzing data and finding common or new patterns from it
- Programming knowledge (Python, R) is a plus
- Data, network analytics tools (SNAP, Gephi, NetworkX) is a plus
- Data processing frameworks (Pandas, PySpark) is a plus

Benefits of participating in this project

- Working on some interesting research in the area of community question answering websites
- Opportunity to gain hands-on experience in Natural Language Processing and Machine Learning research