Final Project Description

NETWORK SCIENCE FOR THE WEB - DS 608 - AUG - DEC 2019

Analysis of Social Media

Data Elements in general are:

- Users
  - Ids, locations…
- Social Network
  - Friends, Followers, Leaders
- Posts
  - Text
    - Likes, mentions
    - Favorite, Tweets, Retweets etc
- Time or Event (Epoch) Marker dimension
  - Data captured at regular intervals, or whenever an external (pre-defined) event occurs

Twitter Data Anaytics

- **#TAG = Topic**
  - **Users can tweet across topics – you can analyze by topic (Ex: #AyodhyaHearing)**
- **Users**
  - Can follow certain other users (follwers)
  - Certain Others can follow this user (friends)
- **Network**
  - A Tweet network is a directed network, showing who is following whom (acquaintance network)
    - A 'Leader' is one who has many 'followers'
  - A Retweet network: nodes, *i* and *j* have an edge placed between them if *j* retweets *i* - is also called a ***diffusion*** network
    - Not all followers of a leader may retweet a leader
    - Many users may retweet a user they may not follow

Follower network Vs Retweet Network
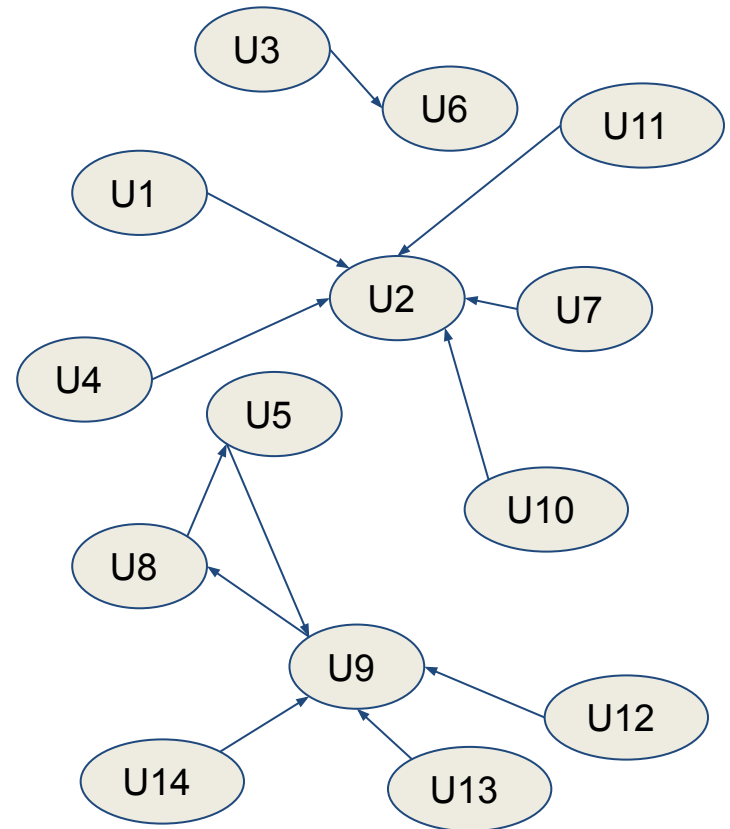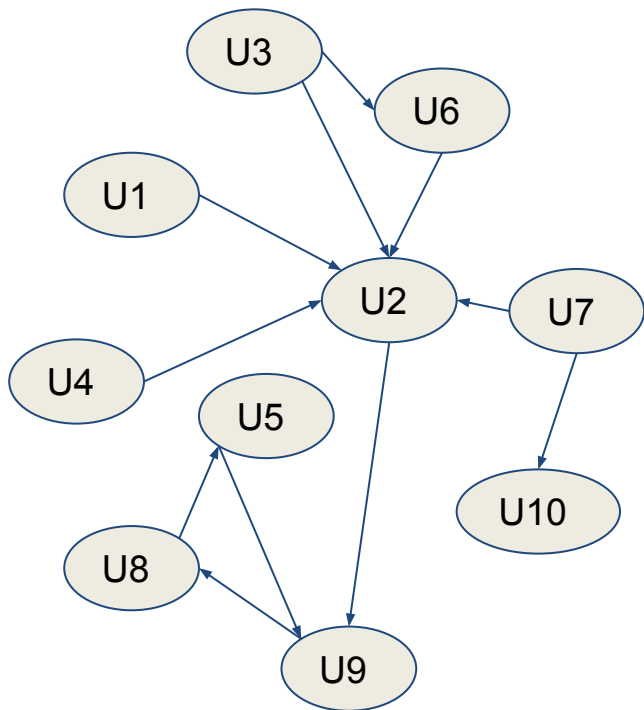
Influencers by network position

- **Follower-Followee Network**
  - Follow/followee represents the acquaintance, user with high followers has higher reach and influence on the network
  - Detect communities
- **Retweet Network**
  - Retweeting can be considered as the retweeter *being influenced* by the orginal tweeter, in effect , amplifying the position of the original tweeter
  - Given a Retweet graph, you can calculate how many people a tweet reached through retweets averaged over the path length (hops) – the higher the number – higher the Reach

Dynamics

- Time ticker or event marker
  - User Dimension
    - New users may appear, old users may turn private, or vanish
  - Network Dimension
    - New links may form, old links may vanish
    - Edges may acquire new weights
      - E.g. Weight = freq of tweets/hour
  - Post Dimension
    - Stance and/or Sentiment may vary over time

Project Details

- Each Team is given 2 datasets
- Each dataset will have 2 files
  - Tweet information
  - Follower network information
- Students have to extract retweet network from tweets
- For each dataset, students should (Time-series) analyse and present a story with 3 dimensions:
  - Network analysis
  - Social Learning over Sentiment
  - Detect Social Bots in the network

Disclaimer

Students should not publish the datasets on any public forums

(Twitter prohibits the above action and if violated, there can be implications on students and IIITB)

Expectations from the project-1 (Network Analysis)

- Tell a story about a set of users(for each dataset)
  - A network view of a community
    - Use community detection tools, centrality measures
    - Extract data and build a ReTweet graph
      - Analyse for structure, centrality
    - Visualize and Analyze dynamics over time
      - Are the communities growing ?, Shrinking ?
      - Are Influencers growing or shrinking in stature/following ?
    - What is the rationale for your analysis ?
    - Are the influencers and their propagators saying the same thing ?
      - What is the level of alignment or lack of it ?

Expectations from the project-2 (Social Learning)

- Is there is shift in the sentiment over time?
    - As a whole group ?
    - As a community ?

    Develop a rationale for your observations

  - Is there any observable social learning over sentiments expressed ?
    - Build a sentiment driven revision model based on the DeGroot social Learning model
    - Examine the trajectory of learning and the possible fit using initial sentiments against predicted vs. actual final sentiments over time

Expectations from the project-3 (Bot detection)

- Is there a social bot in the twitter space that is attempting to sway public opinion ?
  - Identify the bot using variables such as
    - Time of day, Frequency of posts
    - Rapidity (over time) of followers following a leader
    - Rate of growth of a central character who is retweeted with high frequency
    - Frequency of Tweets vs. ReTweets
    - Alignments of Tweets with Retweets in terms of 'Bag of Words' and sentiment
    - Rate of change of sentiment

Stances & Sentiment of Tweets
(Optional)

- Stance: Users stand towards the topic
- Sentiment: Users feeling about the topic
- Text Mining of Tweet text can reveal
  - If tweet == junk ?
  - If tweet <aligned> with #TAG
  - If tweet exposes a sentiment
    - -1 = Strongly negative
    - 0 = neutral
    - 1 = Strongly positive
- Analysis of a content of Tweets can reveal
  - Does all users belonging to one community have similar stance?
  - How close the sentiments in the community are
    - E.g. A 'camp' of 'negative' posts

Submission details (as Team)

- Present the analysis of both datasets
  - Can use 'rtweet' R package to perform analysis
- Submit a report on the analysis of both datasets

Thank you