COMP30850 Assignment 2

Deadline: Friday 10th March 2023

Overview:

The goal of this assignment is to create and characterise a range of different network representations, created from pre-collected Twitter data for a specific *Twitter List* of user accounts that relates to a specific topic (e.g. business, technology etc).

The assignment should be implemented as a single Jupyter Notebook (not a script). Your notebooks should be clearly documented, using comments and Markdown cells to explain the code and interpret the results of your analysis.

Data:

Select one of the Twitter lists from the page below and download the associated ZIP dataset file. No further Twitter data collection is required.

http://mlg.ucd.ie/modules/COMP30850/lists/

Each dataset consists of three different files:

1. *list_members.jsonl*: A JSON Lines (JSONL) file, where each line is a separate JSON document. Each JSON document contains the Twitter user profile of a list member.
2. *followers.csv*: A tab-separated text file, indicating the follower relations between pairs of list members. The first two columns contain a pair of ordered user names X and Y, indicating that user X follows user Y. The third column indicates the date on which X started following Y.
3. *tweets.jsonl*: A JSON Lines file, where each line is a separate JSON document. Each JSON document stores a tweet posted by one of the list members.

Tasks:

From your data, you should (i) *construct* and (ii) *characterise* in detail the six Twitter representations listed below.

When constructing each network, ensure that you use an appropriate type of network which includes the relevant nodes, edges, and attributes. Explain the construction process using Markdown cells.

When characterising each network, you should include a range of analysis approaches which have been covered during the module, and discuss the results using Markdown cells. Plots and tables should be included where appropriate.

1. **Follower network:** Here each node in the network should represent a user from your list and edges encode follower relations.
2. **Reply network:** Here nodes will either be users from your list or users from the wider Twitter platform - i.e. users who are not necessarily members of the list, but received replies from users on the list. Edges encode these replies.

3. **Mention network:** Here nodes will be either users from your list or users from the wider Twitter platform - i.e. users who are not necessarily members of the list, but were mentioned in tweets by users on the list. Edges encode these mentions in tweets.
4. **User-hashtag network:** A bipartite network, where one set of nodes represents the users from your list, and the other set of nodes will represents the hashtags which appear in their tweets. Edges encode the user-hashtag relations.
5. **Hashtag co-occurrence network:** Here each node will correspond to a hashtag appearing in one or more of the tweets of users from your list. Edges represent the co-occurrence of hashtags in the same tweets.
6. **Dynamic follower network**: Using the timestamp information which indicates when each user X started following user Y, construct and characterise three time window networks representing the evolution of follower relations.

Guidelines:

- The assignment should be completed individually. All submissions will be subject to plagiarism checking. Any evidence of plagiarism will result in a 0 grade.
- The grade awarded will depend on the complexity of the analysis and level of detail, i.e. network construction, characterisation, interpretation etc.
- Submit your assignment via the COMP30850 Brightspace page.
- Your submission should be in the form of a single ZIP file containing your Jupyter notebook (i.e. IPYNB file).
- Hard deadline: Submit by the end of 10th March 2023.
  Penalties will apply for late submissions:
  - 1-5 days late: 1 grade point deduction, e.g. B to B-
  - 6-10 days late: 2 grade point deduction, e.g. B to C+
  - Assignments will not be accepted later than 10 days without Extenuating Circumstances formally approved by UCD.

Related Links:

- Details on the JSON Lines (JSONL) file format:
  http://jsonlines.org
- Description of the Twitter JSON format for representing user profiles
  https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/object-model/user
- Description of the Twitter JSON format for representing tweets
  https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/object-model/tweet