Week 09

Network growth

Tuesday, October 19

INFO 5613: Network Science

Brian Keegan
brian.keegan@colorado.edu

University of Colorado Boulder

Agenda

- Networks are not static! Nodes, edges, and their attributes change in time
- Final assignment

Network dynamics

- Dynamics ***of*** networks – how the nodes and links in a network change in time
  - This week!
- Dynamics ***on*** networks – how the attributes of nodes and links in a network change in time
  - Next two weeks!

Temporal sampling

- Defining the temporal boundaries of when our network starts and stops is crucial (Week 2!)

- **Instantaneous networks** – what nodes and edges are present in an instant of time?
- **Snapshot network** – what nodes and edges are present only between a start and stop time?
- **Cumulative network** – what nodes and edges have been present from a start through a stop time?

Case study: Boulder politics on Twitter

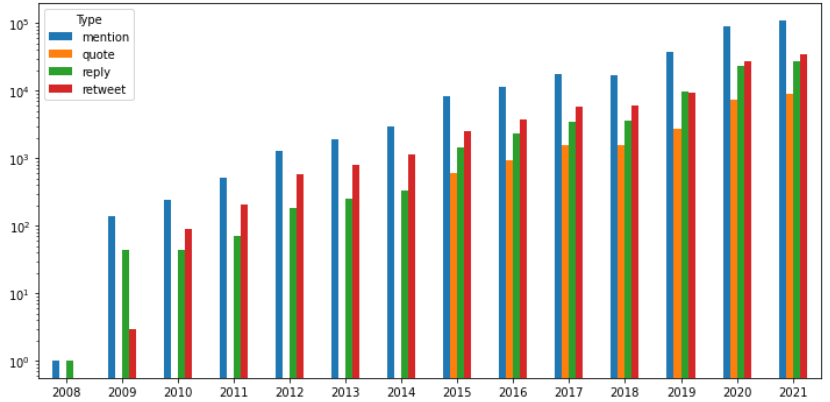
- Previously: friend relationships among 692 accounts around City Council members and candidates
- Analyze four types of content-level interactions in tweets
  - **Mention**: account X references account Y in the text of a tweet
    - https://twitter.com/AaronBrockett12/status/1447755284005326854
  - **Reply**: account X tweets in reply to a tweet from account Y
    - https://twitter.com/AaronBrockett12/status/1449767415139635201
  - **Retweet**: account X retweets a tweet from account Y
    - https://twitter.com/mitchellbyars/status/1450474271013212170
  - **Quote**: account X embeds a tweet from account Y
    - https://twitter.com/AaronBrockett12/status/1450472507060916230

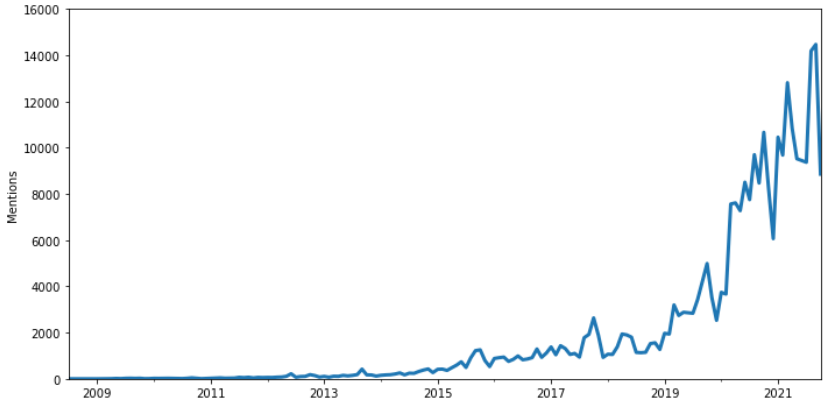
Data collection

- Use the GET statuses/user_timeline API endpoint to retrieve each of 692 user's tweet histories
  - Limited to only ~3,200 most recent tweets!
  - Super-active tweeters will have a smaller window biased towards more recent tweets
  - Less-active tweeters will have a larger window, potentially going back to account (or platform!) creation
  - Get all tweets from a user, not just most recent 3,200? …It's a lot of trouble!
- Include replies and retweets, though these can be filtered
- Just over 6.4 gigabytes of data and 6 hours to retrieve in all!
- For each user's timeline
  - Load file from disk, parse out the four types of content interactions, add to an edgelist
  - Filter edgelist to only include 692 friends, ignore self-interactions, save as "interaction_edgelist.csv"
  - See Appendix in Week 09 notebook for details

Analysis

- Load “interaction_edgelist.csv” and parse Timestamp column
- Filter to the mentions (blue) since it’s most active
- Group mentions by month
  - Plot count over time → strong bias towards recent events!
  - Create an edgelist of monthly mentions
- Make snapshot and cumulative temporal graphs
  - Snapshot: create a network from only edges in that month
  - Cumulative: create a network from all edges through that month
- Let’s jump into the notebook!





Final assignment

What to do?

- Some interest in data collection and analysis
- Each notebook will include a section on collecting and analyzing data
  - Mostly from Wikipedia, but open to other suggestions

Next class

Readings

- Readings
  - Andriani, P. and McKelvey, B. (2009). From gaussian to paretian thinking: Causes and implications of power laws in organizations. *Organization Science*, 20(6): 1053–1071
    - I probably cite this paper in 25% of reviews I write… power laws aren't just caused by preferential attachment!
  - Kossinets, G. (2006). Empirical analysis of an evolving social network. *Science*, 311(5757): 88–90
    - An intuitive and straight-forward analysis of an evolving communication network at a university