



# CS 8803 Social Computing: Event and News Analytics

*Munmun De Choudhury*

[munmund@gatech.edu](mailto:munmund@gatech.edu)

Week 13 | November 12, 2014

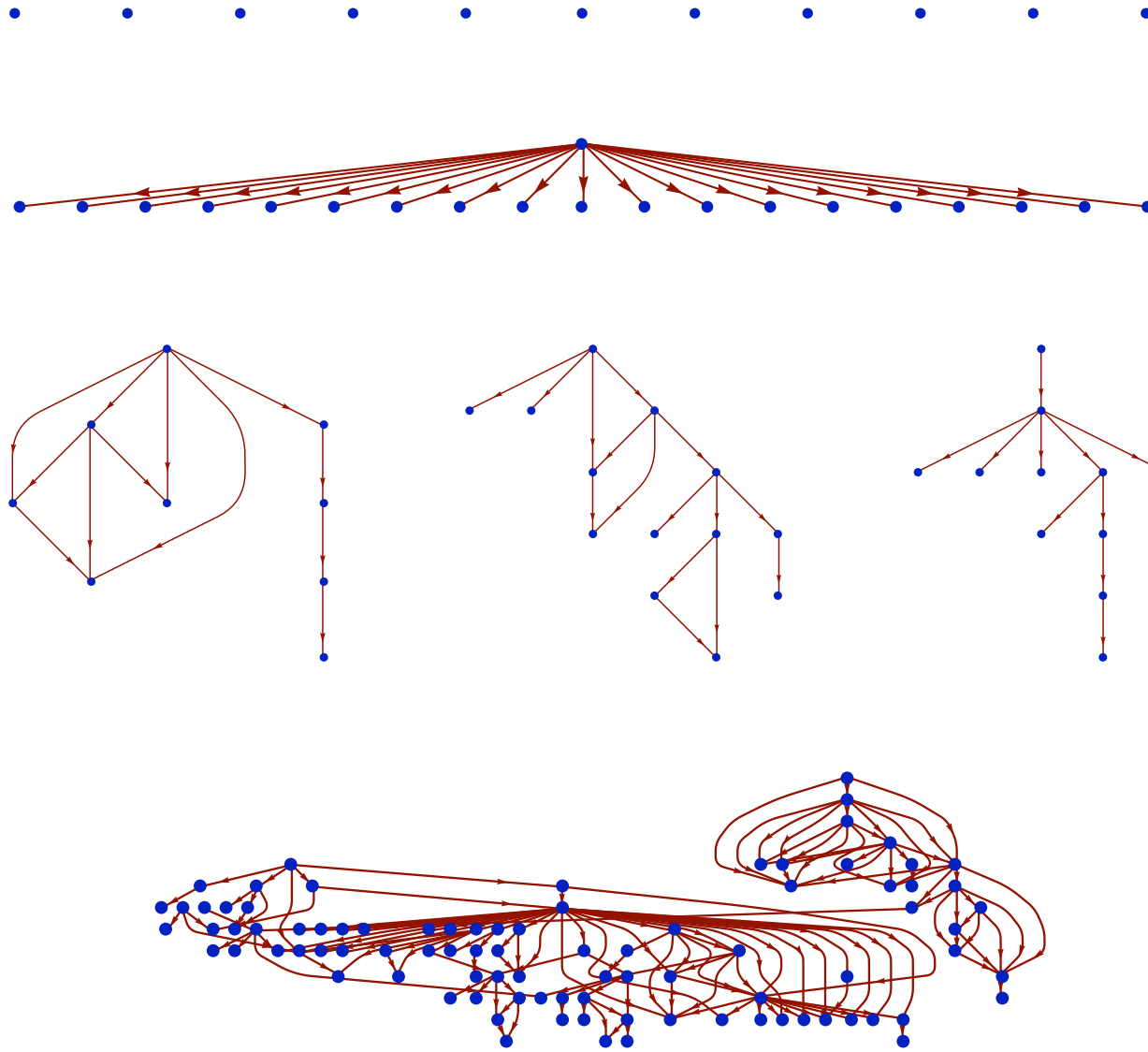
Why is trend identification an important problem? Explain with some case examples.

What is Twitter, a Social  
Network or a News Media?

Do you really agree that Twitter is a news media?

# Summary

- Examine the potential of Twitter as a medium of information sharing
- Data was all of Twitter – crawl was made in 2009 – 41.7 million user profiles, 1.47 billion social relations, 4, 262 trending topics, and 106 million tweets
- Findings:
  - In its follower-following topology analysis authors found a non-power-law follower distribution, a short effective diameter, and low reciprocity
  - Retweet popularity based influence is different from follower count based influence
  - 85% trending topics are headline news
  - Quick diffusion of retweets occur, independent of the follower count of the originator



Information cascades on Twitter (based on retweets) – Bakshy et al, 2010

# Characterizing Debate Performance via Aggregated Twitter Sentiment

# Summary

- The article develops an analytical methodology and visual representations to help understand temporal dynamics of event sentiment as they unfold over time on social media
  - Particularly analyze the first Presidential debates of the 2008 US Presidential elections
- Help a journalist or public affairs people
- Data – authors queried the Twitter Search API for #current, #debate08 and #tweetdebate
  - 1,820 tweets from 664 people during the 97-minute debate and 1,418 tweets from 762 people in the 53 minutes following the debate
- Identify controversial topics and how sentiment changed for both candidate on a moment to moment basis



# Hip and Trendy: Characterizing Emerging Trends on Twitter

# Summary

- Interpret and identify emerging trends in social media
- Contributions:
  - Develop a taxonomy of the trends present in the data (NYC)
  - Identify important dimensions according to which trends can be categorized
  - Key distinguishing features of trends
  - Quantitative examination of the computed features for different categories of trends, and how they can be used to distinguish between categories
- Data – 48,000,000 Twitter messages posted by New York City users of Twitter between September 2009 and March 2010
- Two types of trends detected – exogenous events (offline world news) and endogenous events (Twitter only)

Kwak et al. found that retweet popularity based influence is different from follower count based influence. What could be the reason?

Kwak et al. found that 85% trending topics are news. To what extent the design of Twitter (and its trending topic) algorithm is responsible for it?

Twitter now allows some basic algorithmic curation of feeds. Do you think this may cause Twitter to no longer be a news medium?

Is Facebook a news medium? If not, why not?  
Would you consider Reddit to be one?

What makes a social media a news media? Take Twitter's example. Conversely, is the New York Times a "social media" now that you can comment on articles?

Diakopoulos et al.'s paper and the paper we read earlier from the same author (SRSR system for journalists) indicate the potential of using social media as a reliable way to tap on citizen journalism. Why aren't such systems still not quite available for the mainstream?



Diakopoulos et al. used Mechanical Turk to rate sentiment of tweets. What is the strength and weakness of this approach?

Diakopoulos et al. found that tweets were mostly negative. What could be the reason? (Early use of Twitter for these scenarios; skewed population who only complains on Twitter)

The method used in Naaman et al. for trend identification uses burst detection. Could it be used to detect unforeseen events, such as a fire or an earthquake?

Diakopoulos et al. found that tweets favored Obama more than McCain during the first debate which eventually aligned with 2008 election outcomes. To what extent could this be a valid predictor?

# Wednesday Nov 19's class

- Atlanta Computational Social Science Workshop
- Schedule: <http://css-workshop.gatech.edu/schedule.html>
- Attend Prof. Noah Smith's (CMU) talk – 10am to 11am, Friday Nov 21
  - Talk title: "Machine Learning About People From Their Language"
- *OR*
- Attend Prof. Arthur Spirling's (Harvard) talk – 1:30-2:30pm, Friday Nov 21
- **Location: TSRB Ballroom**
- *Either do Nov 19's reading reflections, or attend one of the above talks. If later, attendance will count for the grade toward Nov 19's reading reflections*

# Next class

- Monday 11/17
- Topic: “Location and Mobility”
- Assigned readings due by 11:59 pm Sunday
- **No class next Wednesday 11/19**