A quick introduction to...
#FluxFlow

A technical paper from Jian Zhao, Nan Cao, Zhen Wen, Yale Song, Yu-Ru Lin, and Christopher Collins

Presented for DS8006 @ Ryerson University - Masters of Data Science

By Bernie Najlis on Feb 1st, 2017

11 AM 12 PM 01 PM 02 PM 03 PM 04 PM 05 PM 06 PM 07 PM 08 PM 09 PM 10 PM 11 PM

Agenda

- What is **#FluxFlow** ?
- What problem does it try to solve ?
- How does it solve it ?
- Case Study – Hurricane Sandy
- Questions
- Summary

What is **#FluxFlow**?

- Visual tool to analyze "anomalous information" spreading on Social Media

- What is "anomalous information" in the context of Social Media?
  - Rumor, misinformation
  - I.e.: During 2011 London riots, misinformation (police response to a fatal shooting) spread in social media fueled the riots even more …

- Combines an interactive visual analysis system with innovative Machine Learning models applied to Social Media

What problems does **#FluxFlow** solve?

- Some hard to detect conversations in social media are the ones that have significant impact

- Conventional anomaly detection models can not be easily adapted to social media

- Analysis based solely on scores is difficult (high dimensionality, why models assign certain scores)

- **Note**: #FluxFlow is not for live stream analysis, it works on batch mode

How does **#FluxFlow** work?
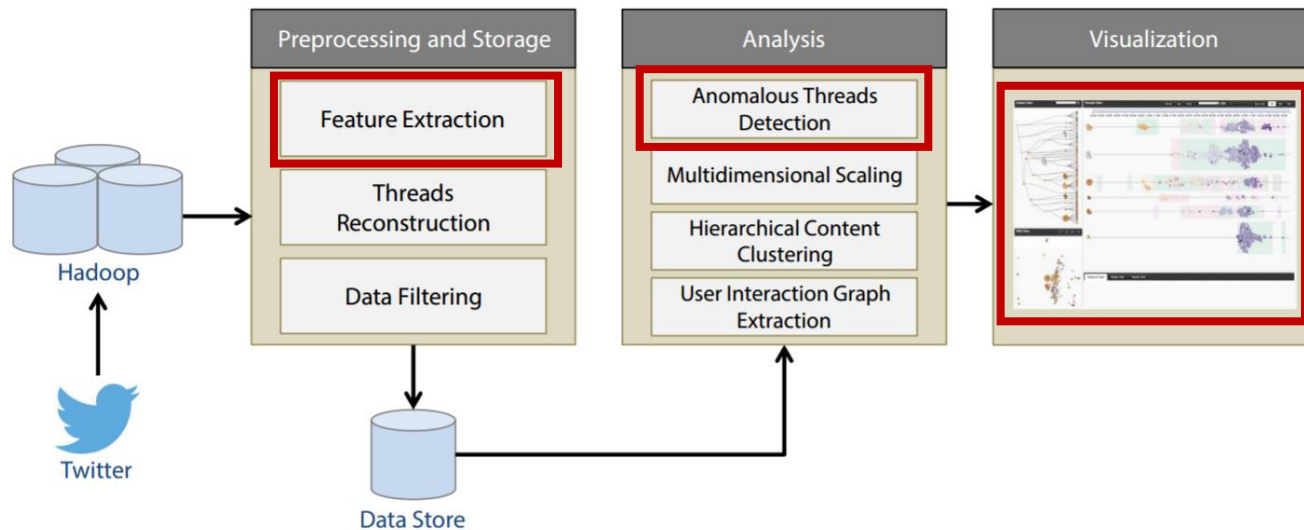
- Data is collected on a Hadoop Cluster

- Map Reduce pre process for:
  - Feature Extraction
  - Thread Reconstruction
  - Data Filtering
  - Store pre processed data into a (relational?) DB

- Analysis assigns an anomaly score for each retweeting thread and ranks them

- Interactive visualization layer allows analyst to drill down and track information spread patterns

How does **#FluxFlow** work – Under the hood…

- Feature Extraction
  - **Twitter user interaction graph:** weigh link from user a to user b based on number of retweets and mentions from a to b
  - **User profile**: # friends, #followers, #tweets, acct age, …
  - **User network**: EgoNet similarity, rel among thread users,…
  - **Temporal**: retweet count in time, number of devices tweeted from, time interval between retweets, …
  - **Content**: Has question mark?, Sentiment detection score

- Thread anomaly detection
  - OCCRF (One Class Conditional Random Fields) Machine Learning Model detects sequential anomalies (unsupervised) based on spread patterns
  - Model with 3,000 threads trained in 2 hours (Xeon 2.13 GHz – 40 Cores – 32GB)
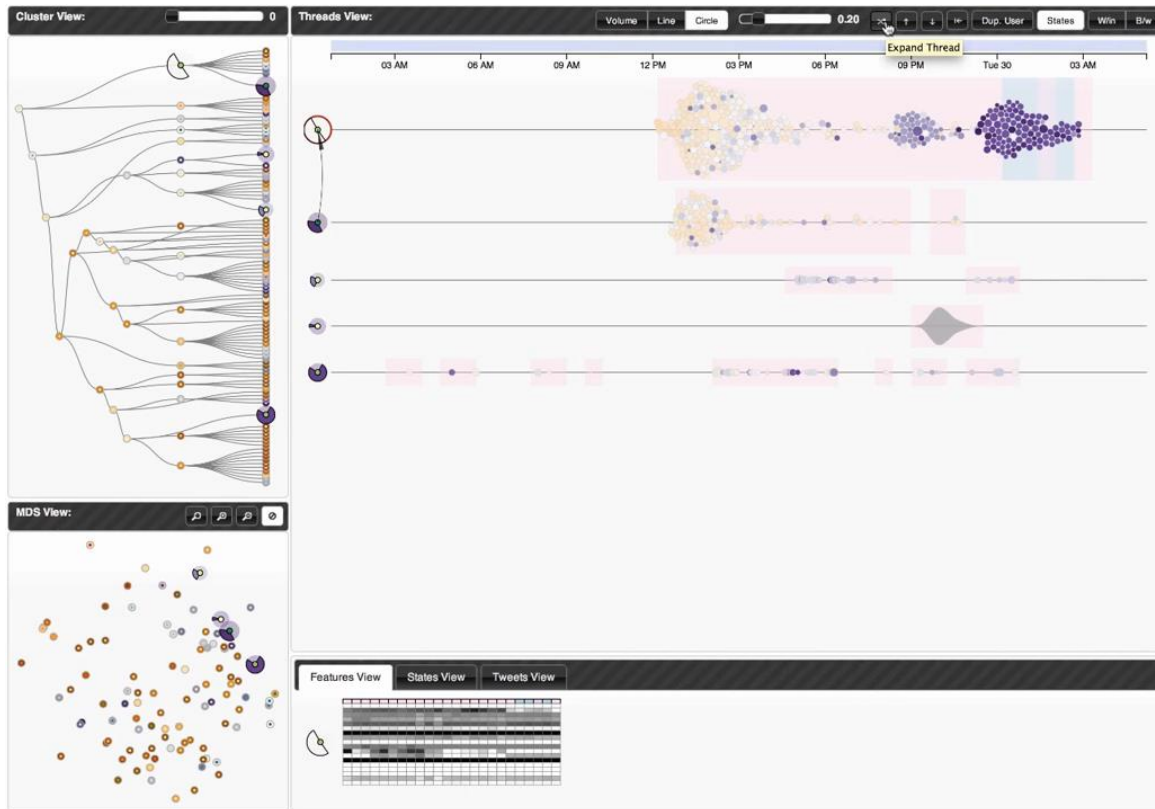
- Visualization
  - Cluster view
  - MDS to represent thread distances in 2D
  - Threads view
  - Information Panel
    - Features View
    - States View
    - Tweets View

How does **#FluxFlow** visualize data?

a) Thread glyph design to aggregate the main thread information

b) Volume Chart

c) Linear circle view

d) Volume circle view
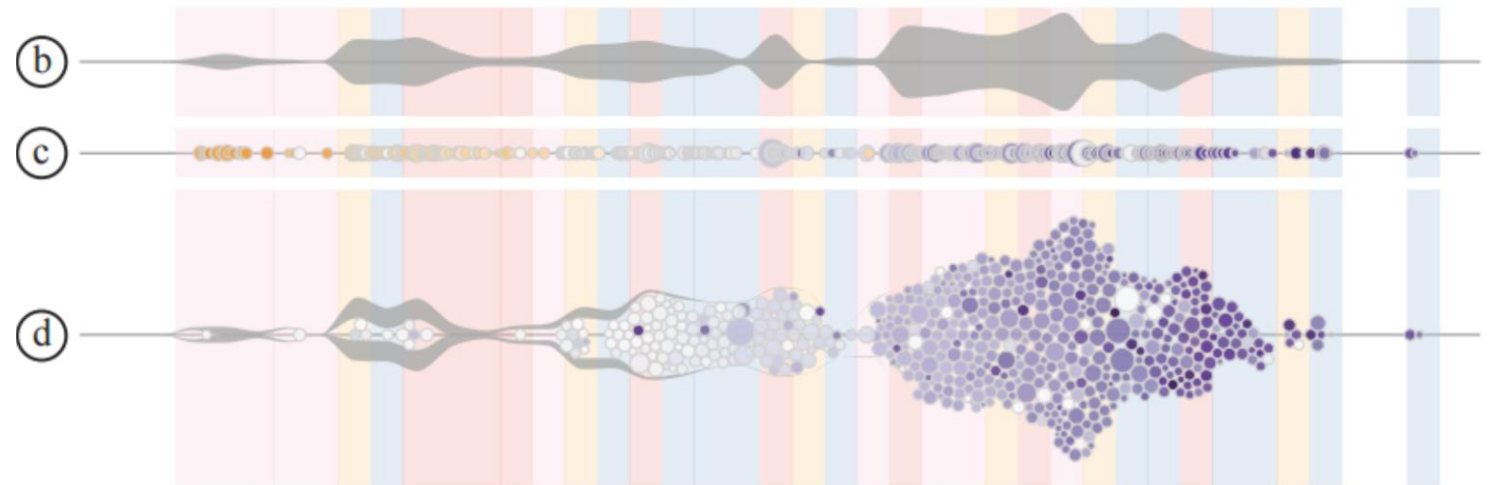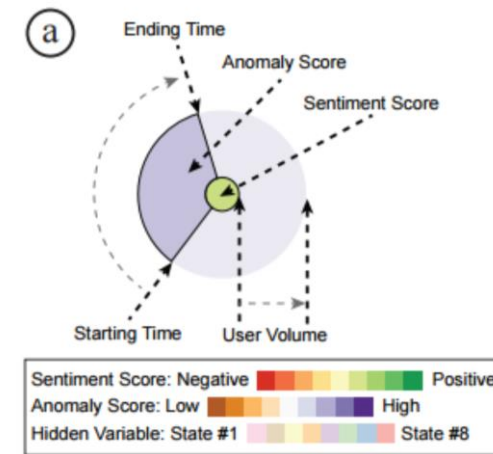
#FluxFlow Case Study – Hurricane Sandy

- Hurricane Sandy dataset contains 52 million tweets during October 29, 2012
- Bottom-up approach
  - *"Obama tells marine _____ ave to guard the Tomb of the Unk_____ hey Refuse"* **RUMOUR**
- Top-down approach
  - *"Power o_____ merica"* **CHEAT**
  - *"A shark s_____ down the street"*
- Connection and comparison
- Deeper insights into the model

http://www.cs.toronto.edu/~jianzhao/webapp/FluxFlow/fluxflow.html

More Questions??? Oh, yeah…

- What other social media rumors you know of?

- What kind of issues do you think social media rumors can cause?

- What do you like (or do not like) about #FluxFlow?

Summary

- Helps detecting 'anomalous information' (i.e. rumours)
- Innovative approach combining machine learning and dynamic visuals to social media
- Makes social media analytics much more interactive
- Domain experts consulted during research praised #FluxFlow
- Research team to optimize model training to reduce time and potentially add real time analysis