An Agent-based Model of Posting Behavior During Times of Societal Unrest

Krishna Bathina*, Aruna Jammalamadaka**, Jiejun Xu**, Tsai-Ching Lu**
*Indiana University, bathina@Indiana.edu
**HRL Laboratories {ajammalamadaka, jxu, tlu}@hrl.com

Hashtag Activism

- First mentioned regarding the #OccupyWallStreet movement
- Has since become a global phenomenon due to increased social media usage and awareness of news events
- Shown to be an effective publicity technique
  - #IceBucketChallenge, #BlackLivesMatter, #ShoutYourAbortion
- Although efficacy for political protests may be questionable, there is a positive correlation between social media usage and political participation

#OccupyWallStreet   #UmbrellaRevolution   #SOSVenezuela

Research Goals

- **Case Study: Protests in Ferguson, Missouri**
  - On August 9, 2014 Michael Brown was shot dead by then police officer Darren Wilson. Protests began later that day and were met by officers in riot gear
  - Tensions escalated and included a declared state of emergency, a state mandated curfew, and calling in the National Guard. State of emergency was lifted on Sept. 3, 2014.
  - Large international following on Twitter, Facebook, and Tumblr with many popular hashtags like #ferguson, #mikebrown, #handsupdontshoot
- **Objective:** Accurately model and simulate changes in the behavior of Tumblr users due to the Ferguson protests
- **Approach:** Design an agent-based model (ABM) that uses the real historical event to change the behavior of agents accordingly.
- **Talk Outline:**

Data Collection → **Preliminary Analysis** → **Model Description** → **Validation of Results**
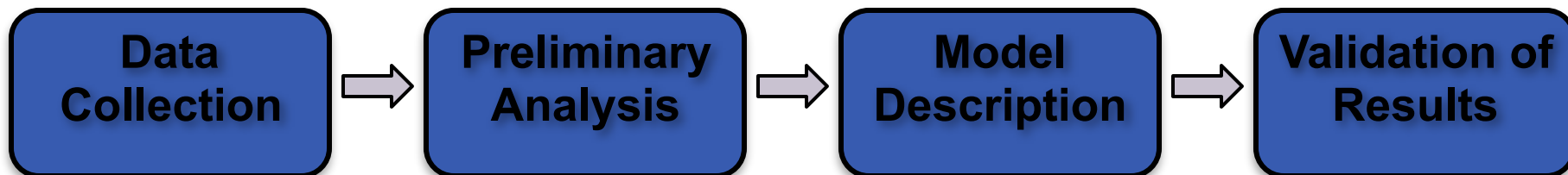
Data Collection

- **Full Tumblr network dataset (413,867 nodes, 23M edges)** from 2012 to end of 2014 allows us to accurately validate our model

- **Time periods:**
  - **"Before"**: 5/1 – 8/8, 3 months prior to the protest
  - **"During"**: 8/9 – 9/3 beginning of the protest until the day the state of emergency was lifted.

ferguson, mikebrown, michaelbrown, tcot, justiceformikebrown, handsupdontshoot, mediablackout, darrenwilson, dontshoot, fergusonshooting, opferguson, policestate, missouri, ripmikebrown, handsup, police, iftheygunnedmedown, fergusonriot, prayforferguson, stlouis, standwithferguson, blacklivesmatter, nojusticenopeace, peaceinferguson, fergusonpd, whereisjustice, occupyferguson, ccot, arrestdarrenwilson, officergofuckyourself, curfew, crimebutnotime, fergusonpolice, fergusononfireusa, michael, brown, humanrightsferguson, michealbrown, mikebrownfuneral, blackyouthmatter, fergusonsolidarity, militarizationofpolice, justiceformichael, justiceformichaelbrown, mikebrownrally, militarizedpolice, dcferguson, stoppolicebrutality, policemilitarization, pleasedontshoot, fergusonqs, sosferguson, copwatch, resistwithferguson, mikebrownnola, furgeson, iammikebrown, fergusonscanner, protests, endpoliceterror, badgecam, fergusonriots, feedferguson, direnferguson, violenceincites, fergusonlive, freeferguson, endpolicebrutality, fergusontapes, fergusoncoverup, nojustice, fergson, standup, justiceformike

*Jules, B.: Hashtags of Ferguson (2014), https://medium.com/on-archivy/ hashtags-of-ferguson-8f52a0aced87.vbia14pwu

- Find all users who mentioned a **protest hashtag*** at least once in the "During" period and gather all their posts and re-posts from "Before" and "During"

- Label all tags that are not in the above list as non-protest tags for this study

- Extract all posts and re-blogs for 10k randomly sampled non-protest tag users for our preliminary study.

- This results in a total of **220M posts, 764M tags**

- During the protest about **1.7% of posts** and **2.1% of tags** were about the Ferguson protest

Preliminary Data Analysis

- Using 4 established macro-level metrics*, is there a statistical difference in behavior before and during the protest?

- Due to size of dataset, all p-values were significant (-log(p) > 30), therefore the effect size Δx and KS-test statistic are reported

Metric	Definition
Hashtag Lifetime	Longest consecutive number of days a tag was posted (normalized by total # days)
Hashtag Popularity	Number of posts of a tag per day averaged over days where it was posted at least once
User Entropy	"Breadth of attention"; Shannon entropy of the tags posted by a given user per day averaged over days in which they posted
User Activity	Number of posts per user per day averaged over days in which they posted

*Weng, L., Flammini, A., Vespignani, A., Menczer, F.: Competition among hashtags in a world with limited attention. Scientific reports 2 (2012)

- Analyses A and B show that **User Activity of protestors during the protest was lower than for both groups before the protest**

- Analyses C and D show that **protest tags have much higher Popularity**

| Comparison	Popularity $\Delta\tilde{x}$	Popularity Z	Time $\Delta\tilde{x}$	Time Z	Activity $\Delta\tilde{x}$	Activity Z	Entropy $\Delta\tilde{x}$	Entropy Z
A. Non-protesters during vs protesters during	-0.19	0.02	0.07	0.10	-12.79	0.28	-2.14	0.31
B. Protesters before vs protesters during	-0.05	0.03	-0.03	0.89	-10.65	0.01	-1.63	0.02
C. Non-protest memes during vs protest memes during	2.35	0.78	0.52	0.89				
D. Non-protest memes before vs protest memes during	12.95	0.82	0.47	0.89				

Model Description

- Inspired by previous models of the limited attention of social media users*, our ABM proceeds as follows:
  - An agent (Tumblr user) is chosen randomly to post a hashtag
  - Central agent receives hashtag #4 from their neighbor and updates their memory
  - Attention/Memory is limited, therefore the new hashtag replaces the oldest hashtag
  - Central agent is chosen to post and chooses either a novel hashtag (#5) or existing hashtags from memory (#2, #4).
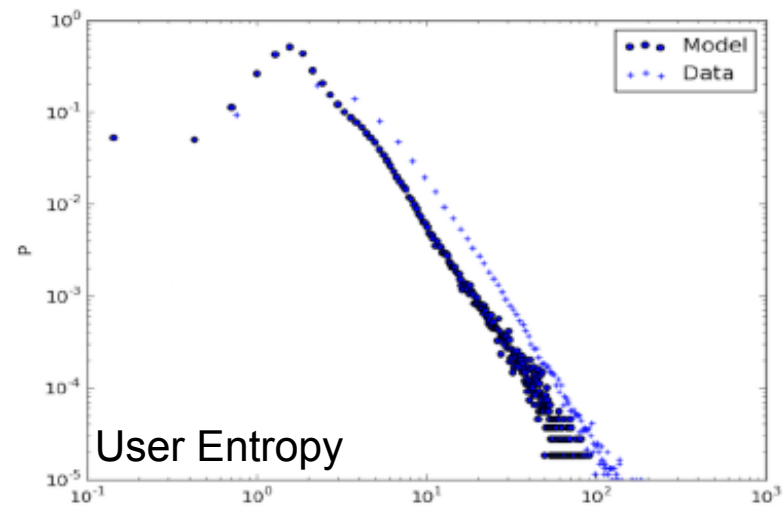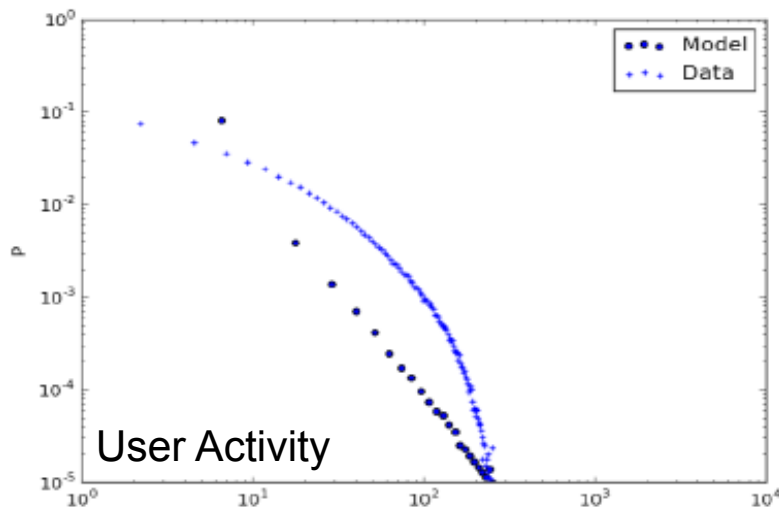 - Their memory is then updated according to what they posted.
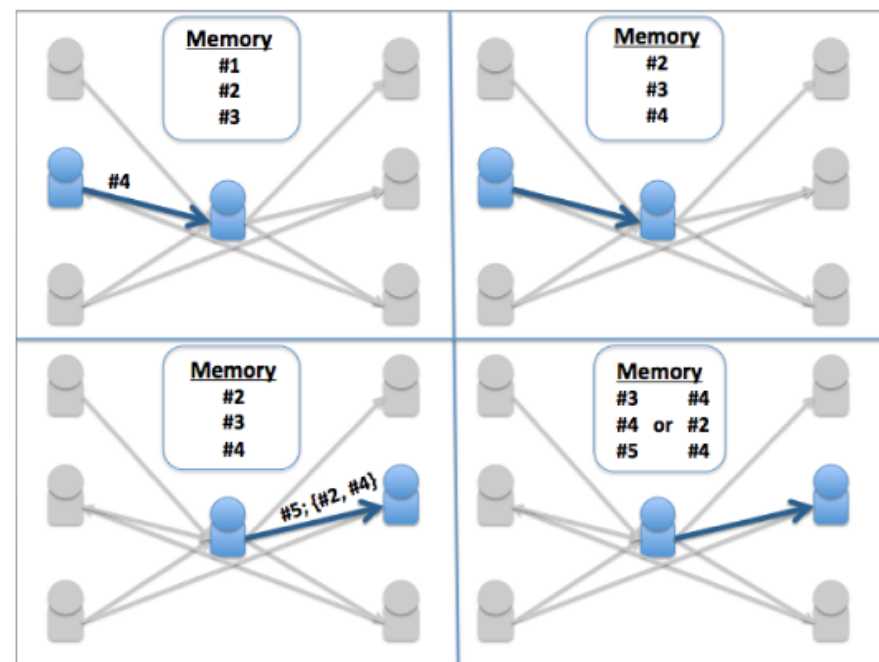
*Weng, L., Flammini, A., Vespignani, A., Menczer, F.: Competition among hashtags in a world with limited attention. Scientific reports 2 (2012)
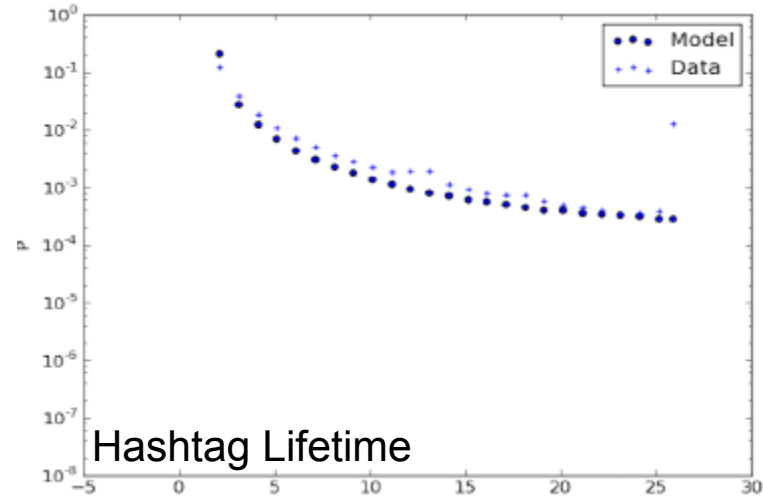
Model Description

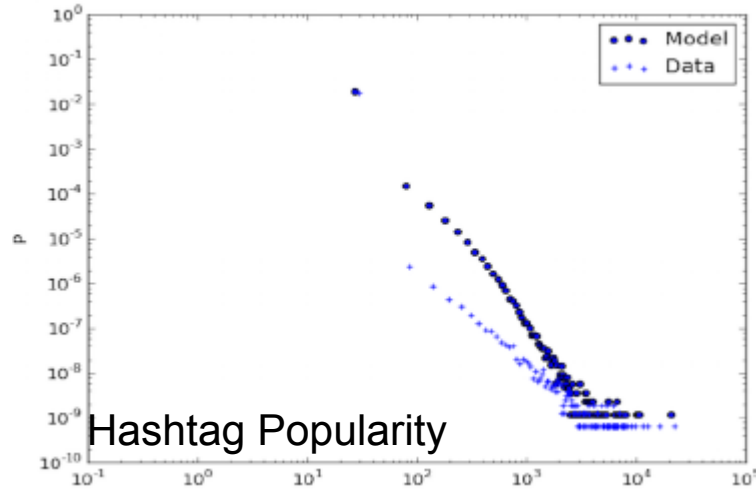- Our model differs from (Weng 2012) in that *posting behavior changes once the protest begins.*

- The agent can still post a novel hashtag or multiple from their recent memory depending on parameters P_n and P_r

- An initial number of agents, equal to the number of actual protesters on the first day are randomly chosen as protestors, and protest tags are added to their memory

- If the proportion of protest tags in an agents memory is $> P_m$, the agent becomes an *"activist"* and consequently only posts protest tags

- The "Before" model is initialized with the Tumblr network environment and parameters derived from the data

- The "During" model is initialized with the output of the "Before" model and same network environment
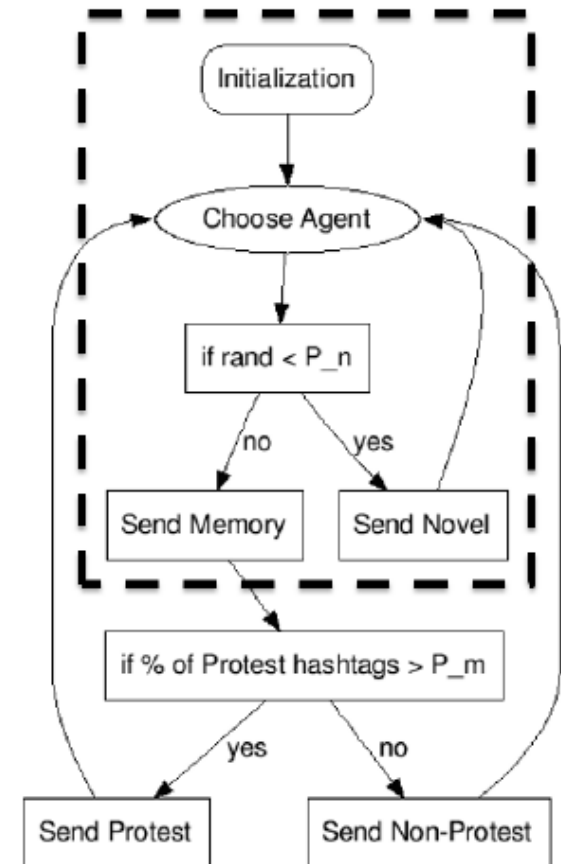
Results

- Model Validation is performed by visually comparing the same **4 macro-scale metrics** resulting from the model simulation to those of the observed data *during* the protest
- **User Activity:** Our model shows a linear decrease in attention due to our assumption that agents post proportional to their out degree*. However, observed results indicate that agents with a moderate number of posts per day are more common than we expected.
- **User Entropy:** matches fairly well with a slight increase to a peak around 1.0 followed by a rapid decrease. Suggests that most users tend to post with little variety per day. Modulated by the novelty parameter P_n.

User Activity

User Entropy

*Chung, F., Lu, L.: The average distances in random graphs with given expected degrees. Proceedings of the National Academy of Sciences 99(25), 15879{15882 (2002)

Results

Hashtag Popularity

Hashtag Lifetime

- **Hashtag Popularity:** although plots have similar shapes, tends to be overestimated. Possibly due to protest tags being posted more frequently than actually observed.
- **Hashtag Lifetime:** similar distributions, flattening out as time increases indicates that the majority of Tumblr hashtags are not re-blogged. Modulated by re-blog parameter P_r
- **Discussion:**
    - Overall difference in macro-level metric distributions other than User Activity are small, therefore we believe that our model was successfully validated by the dataset
    - However we acknowledge that ultimate validation of the model should be performed on an entirely new protest dataset

Questions?

Thank you!