The digital evolution of Occupy Wall Street

Emilio Ferrara – SOIC – CNetS – IU Bloomington
ferrarae@indiana.edu

Joint work with:
MD Conover, C Davis, K McKelvey, F Menczer and A Flammini

Information diffusion research at Indiana University

Truthy is a research project that helps you understand how discussion spreads on Twitter. We currently focus on tweets about political, social movements and events from the past 90 days.

For Researchers
- Data, methods, and visualization
  Study online communication networks with interactive interfaces that visualize data and allow you to identify interesting users.
- Download recent tweets, trend volume by time, network layers, and visualize social alerts (such as predicted political parties, sentiment, use, engagement, etc).

For Citizens
- Gallery
  Read descriptions of interesting memes and take a look at static images to learn what is possible with this research.
- Mentions
  Leave some of your conversation online to see how a hashtag on Twitter evolved during a given time period.

For Journalists
- Political Coverage
  Identify important events and pay attention to key players with our interactive visualizations of political conversations.

For Developers
- Programming
  Use tools to work with our data, statistics, and images using the API.

The truthy project
Meme diffusion on Social Media

Motivations of the protest
- Social and wealth inequalities, taxation disparities
- Non sustainable capitalistic market models
- Political corruption, corporate influence of government

Mottos and Motives
- “We are the 99%”
- Pacifist protest

Key dates and events
- #occupywallstreet Adbusters
- Sep 17th, 2011 – Zuccotti park encampment
- May 1st, 2012 – Attempt to revive the movement
- Massive on-the-ground protests in NY, DC and California

Research questions:

- How did protesters use Online Social Media (OSM)?
- How much localized the diffusion of Occupy-related information on OSM is?
- What type of users Occupy involved in OSM discourse?
- What’s the demographics of Occupy users on OSM?
- Did Occupy change online users’ behaviors, connectivity and interests?

What’s the impact of Occupy on Online Social Media?
“Geography” dataset [1]:
- **Occupy corpus**
  - Any tweet containing #ows or #occupy[*]
  - Total of 1.5M tweets [676K RT] produced by 257K distinct users
- **Domestic politics corpus (baseline)**
  - Any tweet containing #tcot (top conservatives on Twitter) or #p2 (Progressives 2.0)
  - Total of 825K tweets [259K RT] produced by 68K distinct users

“Temporal” dataset [1]:
- **Occupy corpus**
  - From Sep 1st, 2011 to Aug 31st, 2012
  - Any tweet containing #ows or #occupy[*]
  - Total of 1.82M tweets produced by 447K distinct users
- **Random sample**
  - From Jun 1st, 2011 to Aug 31st, 2012 [3 months before ows]
  - Any tweet produced by a random set of 25K ows users
  - Total of 7.74M tweets

[1] Data collected from Twitter garden-hose (10% sample rate)

Only ~1% of total tweets contains geo-data (e.g., GPS coordinates)

People in encampments and during marches were committed to spread news on OSM

Goal: geocoding w/ high accuracy level high-volume Twitter traffic

Methods: Bing API + blacklist + whitelist + fuzzy string matching

Result: 55.7% and 29.3% of Occupy and Domestic politics users geocoded from users’ self-reported locations
The geography of Occupy

Volume of traffic per state

NY, California and DC are the main actors of the Occupy discourse. Some states very active in political discourse, such as Kentucky or Alabama, show little to no interest in Occupy-related topics.
Deviation from the baseline

The intensity in the color represents how much the amount of Occupy-related traffic deviates from that of domestic politics per state.

Content production and consumption

Ratio = # RT originating from users in the state / # RT retweeted by users in the state
Content diffusion

Interstate communication

Occupy-related discourse show a prominent hub-and-spoke structure differently from domestic politics (on the left). Multiscale backbone extraction – confidence level $\alpha = 0.15$. 
Collective framing: the social processes whereby movement participants negotiate the shared language and narrative frames that help define the movement's identity and goals.

Resource mobilization: the work to marshal the physical and technological infrastructure, human resources, and financial capital necessary to sustain ongoing activity.

Local vs. global communication

Occupy communication patterns exhibit heightened diagonal activity (more than 3 times) than domestic politics ones.
Occupy discourse on Twitter has highly localized geospatial structure: a large amount of traffic is produced and consumed locally per state. This might be explained by Resource Mobilization.

Interstate communication is driven by high-profile locations acting as information broadcasters. This might represent the Collective Framing process.

Proximity to on-the-ground events plays a big role: users from NY, DC and California are the main actors of the discourse. They produce much more Occupy-related information than that they consume, unlike other states.

Summary of geo-analysis

The evolution of Occupy

This sign was put in front of a media tent at the Occupy Toronto camp in St. James Park (Toronto). Image by Hillary Burridge.
Activity volume

On-the-ground events (circles) and Twitter data-stream outages (blue bands) are highlighted. Bins have 12-hour length.

Politics and protest keywords

The top hashtags adopted by the 25K random users related to domestic politics and foreign social movements.
Attention allocation

Engaged User Ratio describes the proportion of active users in each time-step who produced at least one topicically-related tweet. Engaged User Attention Ratio describes, among these users, the share of average attention allocated to each topic.

In-group connectivity

Reported values represent means and 95% confidence intervals for each time-step.
Occupy discourse on Twitter is highly correlated in time with on-the-ground events. Spikes and high volume of traffic coincide with relevant protests and police actions, marches and strikes.

The phase of explosive growth has exhausted few weeks after the beginning of the protest.

The volume of Occupy-related traffic has diminished by orders of magnitude in the latest observation period.
Occupy captured the attention of users with pre-existing interests in politics and social protests.

Occupying users’ interests and amount of attention dedicated to politics and social discourse did not exhibit any remarkable variation over time.

Occupy users were already highly interconnected before the movement’s start.

The extent to which Occupy users interacted before the movement’s start and at the end of the observation period is substantially unchanged.

Summary of temporal analysis (II)

How groups of users create, interact and evolve in Online Social Media?

Can we generalize some features observed for Occupy (e.g., explosive growth, fast decay, etc.) to other classes of topics discussed on OSM?

How the social and the topical space interact each other to determine the success of certain topics of discussion and the failure of the vast majority of the others?

Current and future work