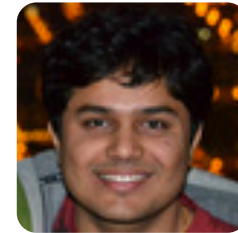
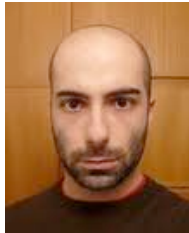
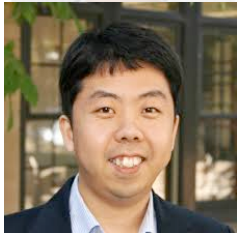


The study of social phenomena via social media

Alessandro Flammini
Indiana University



Thanks to:

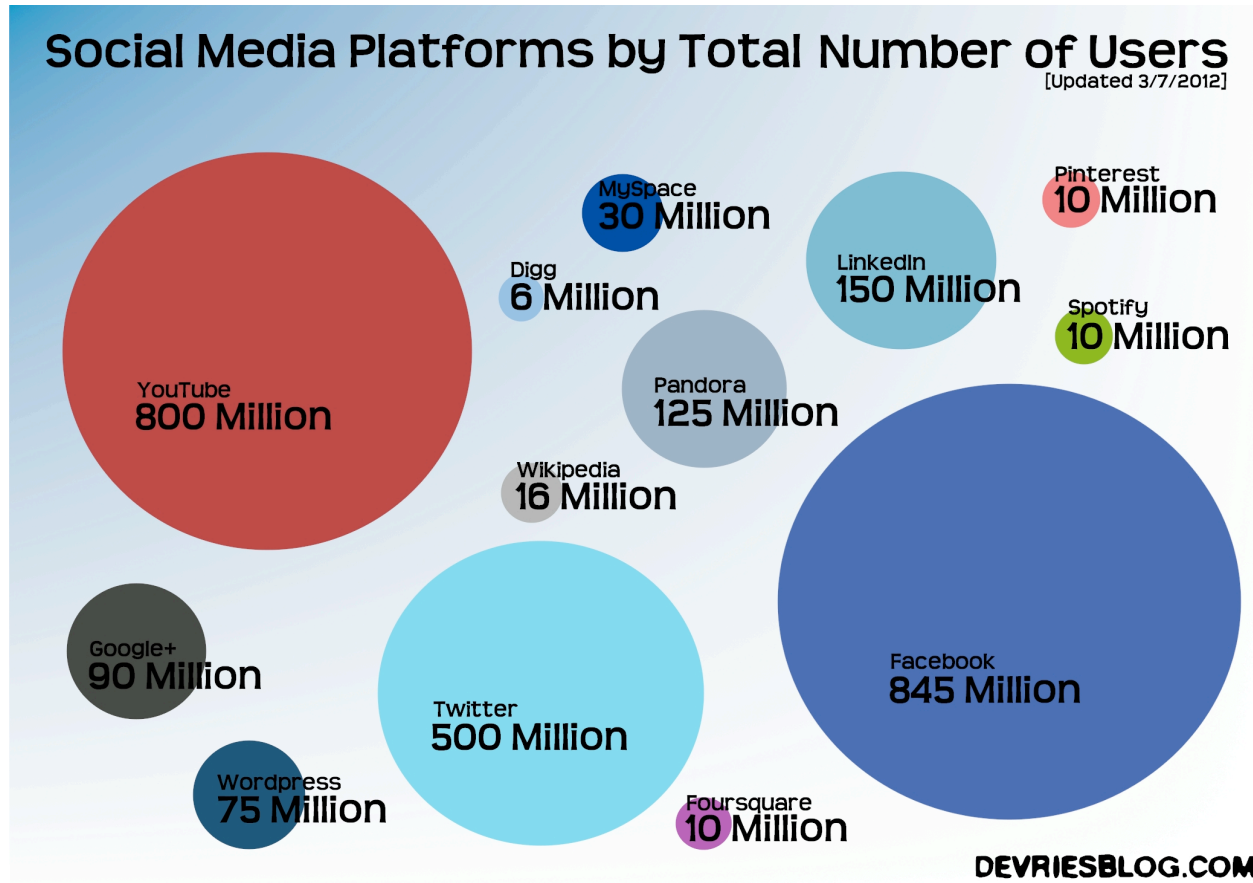


And to:

James S. McDonnell Foundation



Why do we care about Online Social Media?



Information shared on SM matters

People increasingly makes choices based on info share on SM:

purchasing product
political information
health
lifestyle

eternal information consumed (TV shows, blogs, news, books, music)

SM have been leveraged to:

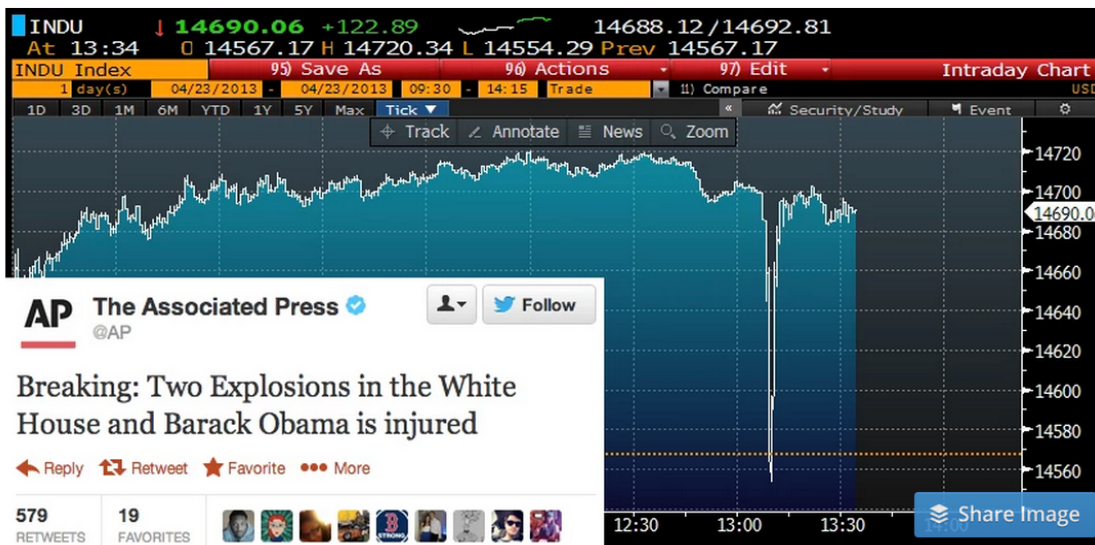
anticipate stock market
forecast movie revenue
probe consumer preferences
coordinate action in crisis
predict outcome of political elections



Syrian hackers claim AP hack that tipped stock market by \$136 billion. Is it terrorism?



By Max Fisher April 23, 2013 [Follow @Max_Fisher](#)



This chart shows the Dow Jones Industrial Average during Tuesday afternoon's drop, caused by a fake A.P. tweet, inset at left.

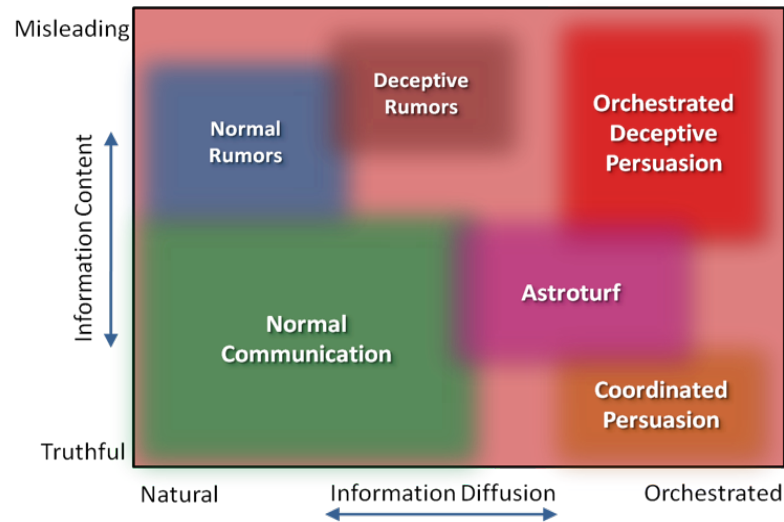
Advertisement

Thanks for the feedback! [Undo](#)

We'll use your feedback to review ads on site.

Social Media: many actors, many agendas

- Detecting persuasion campaigns on Twitter



- Automatic detection of artificial users



Visit us @truthy.indiana.edu

truthy.indiana.edu

Memes

Gallery

Movies

About

Press

FAQ

Meme Search

Search

Truthy

Information diffusion research at Indiana University

Truthy is a research project that helps you understand how communication spreads on Twitter. We currently focus on tweets about [politics](#), [social movements](#), and [news](#) from the past 90 days.

For Researchers

DATA, STATISTICS, AND VISUALIZATIONS

Study online communication networks with interactive interfaces that visualize data and allow you to identify interesting users.

Download recent tweets, tweet volume over time, network layout, and statistics about users (such as predicted political partisanship, sentiment score, language, and activity).

For Citizens

GALLERY

Read descriptions of interesting memes and take a look at static images to learn what is possible with this research.

MOVIES

Create movies of communication over time to see how a hashtag on Twitter evolved during a given time period.

For Journalists

ELECTION COVERAGE (COMING SOON)

Cover the elections with Truthy on your side. Identify prominent ideas and pay attention to key players with our interactive election coverage tool.

For Developers

API (COMING SOON)

Write scripts to work with our data, statistics, and images using the API.

Recent Updates



Truthy @ Indiana
[truthyatindiana](#)

[truthyatindiana](#) Recent Truthy movies about: [#uw](#), [#uwo](#), [#p2](#). Check them out at [youtube.com/truthyatindiana](#)
11 days ago · reply · retweet · favorite

[truthyatindiana](#) Recent Truthy movies about: [#25s](#), [#anonymous](#), [#uw](#). Check them out at [youtube.com/truthyatindiana](#)
6 days ago · reply · retweet · favorite

[truthyatindiana](#) Truthy is tracking at least 5607145 different memes; which ones are the most interesting? [truthy.indiana.edu](#)
5 days ago · reply · retweet · favorite



[Join the conversation](#)

System Info



From the Gallery



Copyright 2011 Indiana University
Truthy, Indiana.edu
The #tcot hashtag represents the "Top Conservatives On Twitter"

Social Networks as a macroscopic system

How does the network structure affects information diffusion?

How does the network evolves in response to the spread of information it supports?

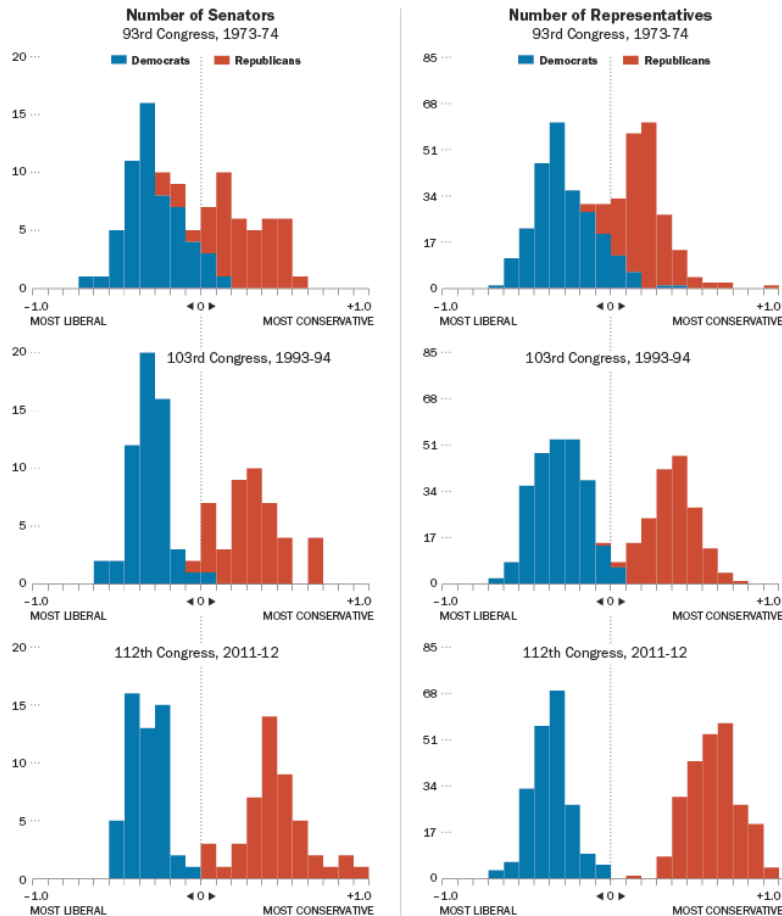
How do memes compete for our attention?

How does our finite attention affects the lifetime, popularity, and diversity of the information we consume?

Political discourse on Twitter & the issue of polarization

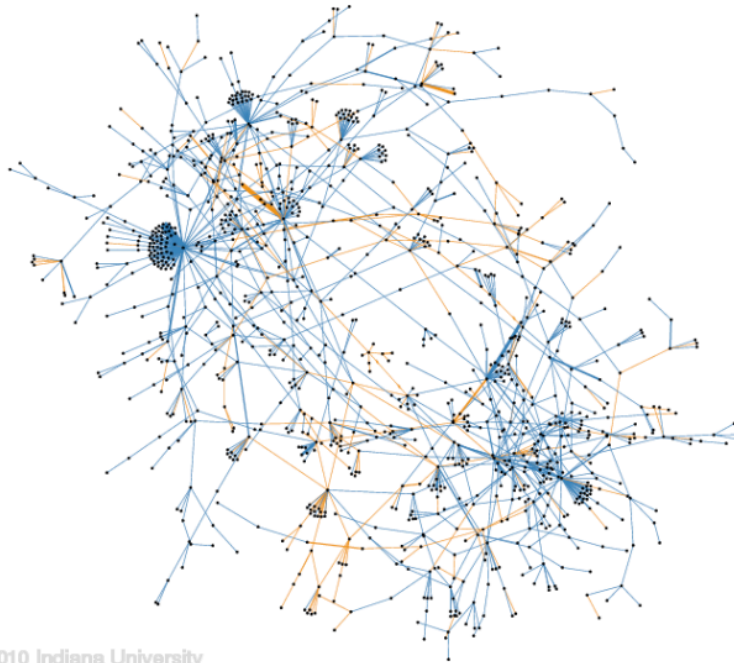
In Congress as Well as Public, the Center Increasingly Cannot Hold

Ideological scores of senators and representatives based on roll-call votes. Negative numbers represent liberal views and positive numbers conservative views



Is people polarized as their congressmen?

Data and Networks



Copyright 2010 Indiana University



thefloatingfrog RT @SebastienPage 5 Ways to Share Your Social Media Identity <http://bit.ly/Mepx2>

less than 5 seconds ago from web



SebastienPage 5 Ways to Share Your Social Media Identity <http://bit.ly/Mepx2>

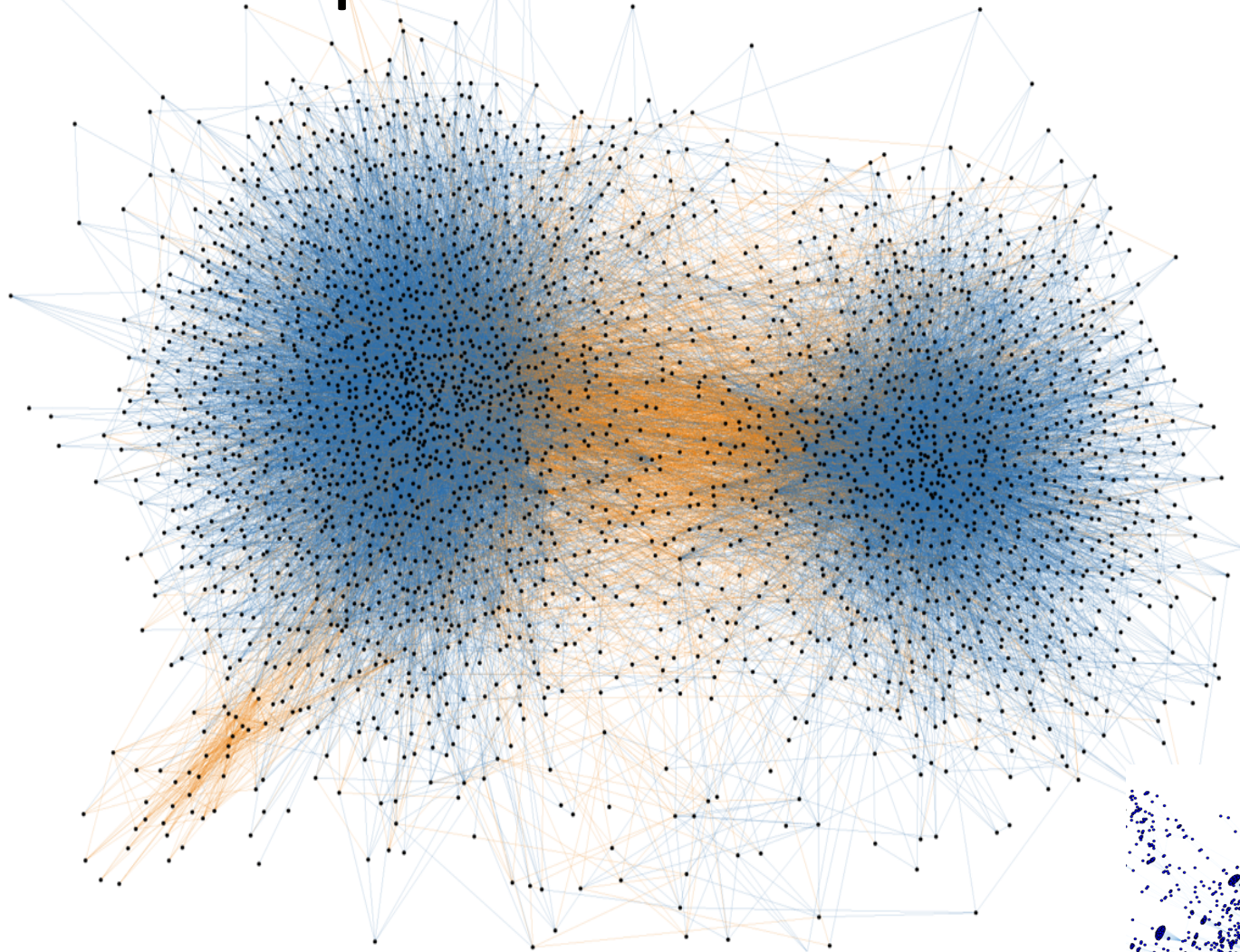
less than 20 seconds ago from HootSuite



right 2010 Indiana University

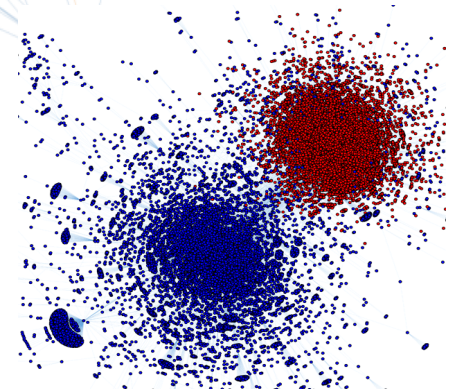
truthy.indiana.edu

How the political network looks like

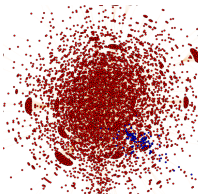
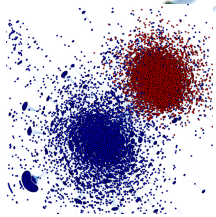
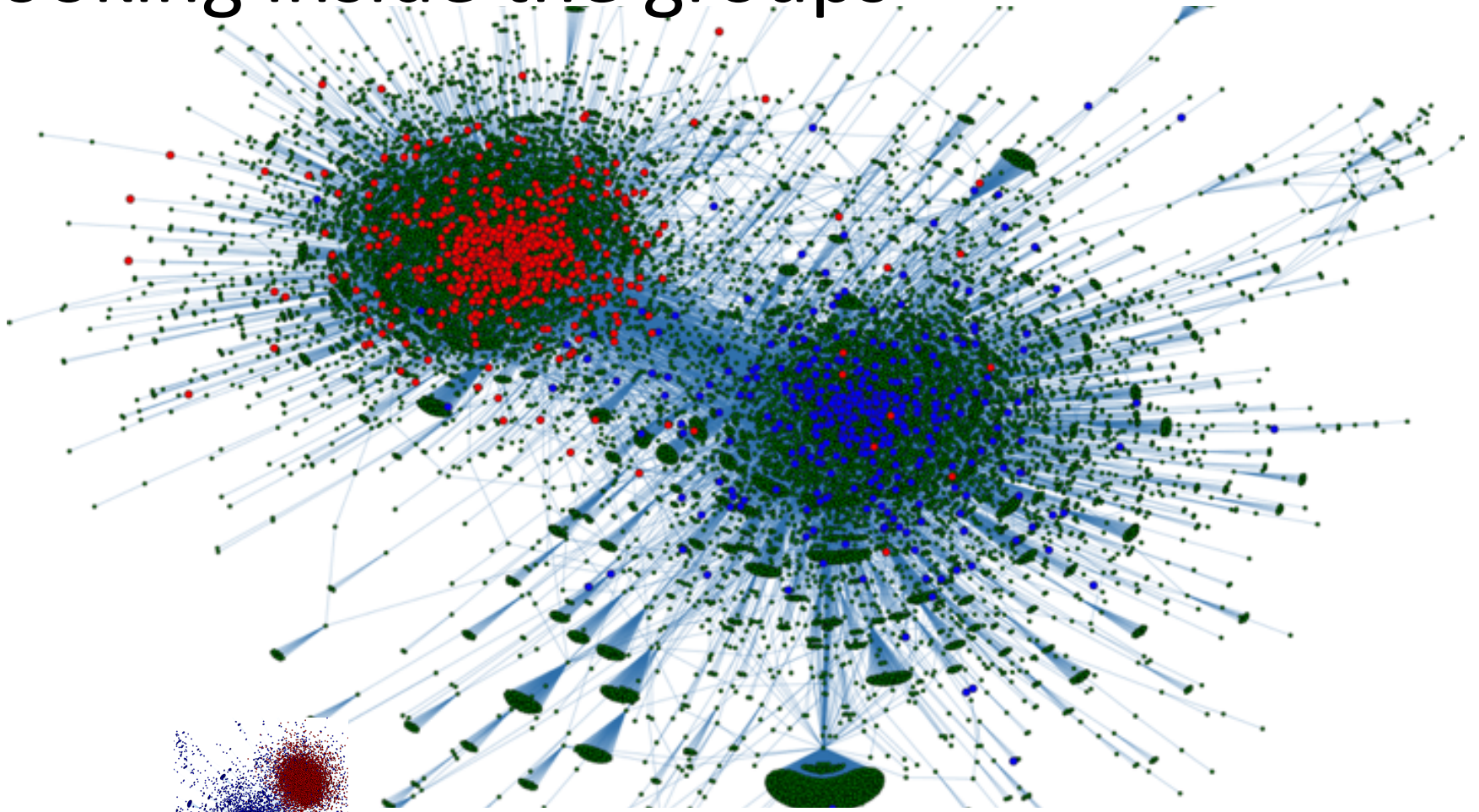


3 months of data. ~300k tweets. ~20k users.

On the right the result of label propagation clustering algorithm

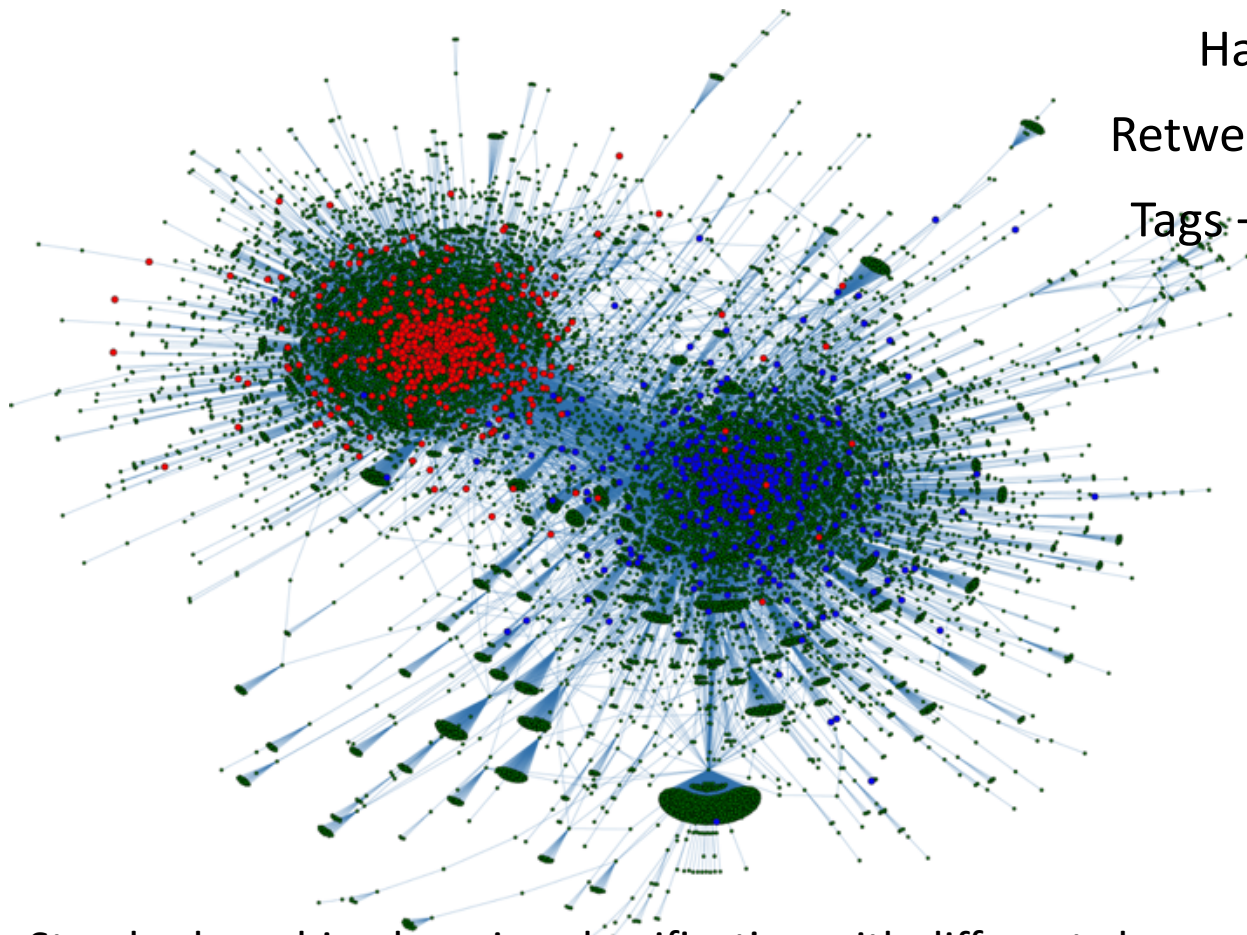


Looking inside the groups



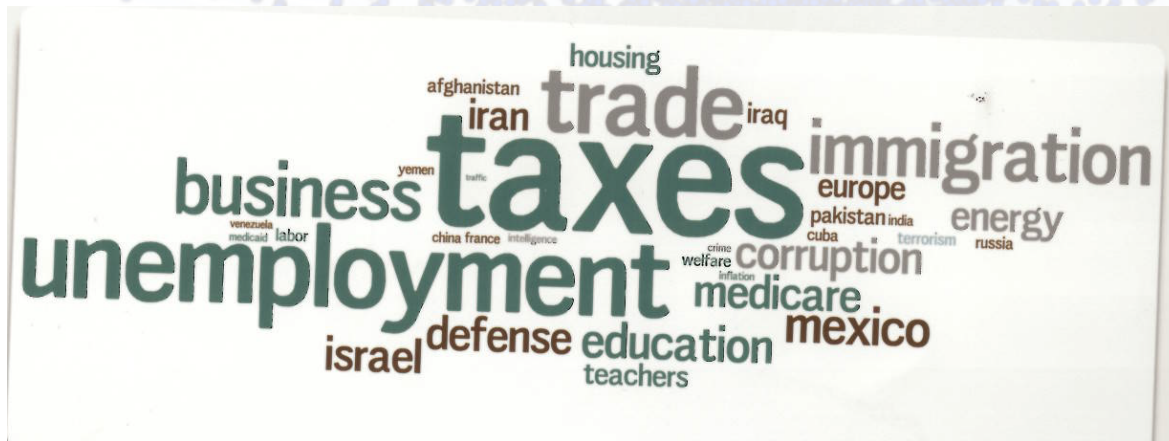
	Left	Right	Und.	Nodes
Retweet	1.1%	93.4%	5.3%	7,115
	80.1%	8.7%	11.1%	11,355
Mention	39.5%	52.2%	8.1%	7,021

Classification of users political leaning



Features	Accuracy
Text (TF-IDF)	79%
Hashtags	91%
Retweet network	95%
Tags + Network	95%

Standard machine learning classification with different classes of features.
Text is in bag of words + TF-IDF representation



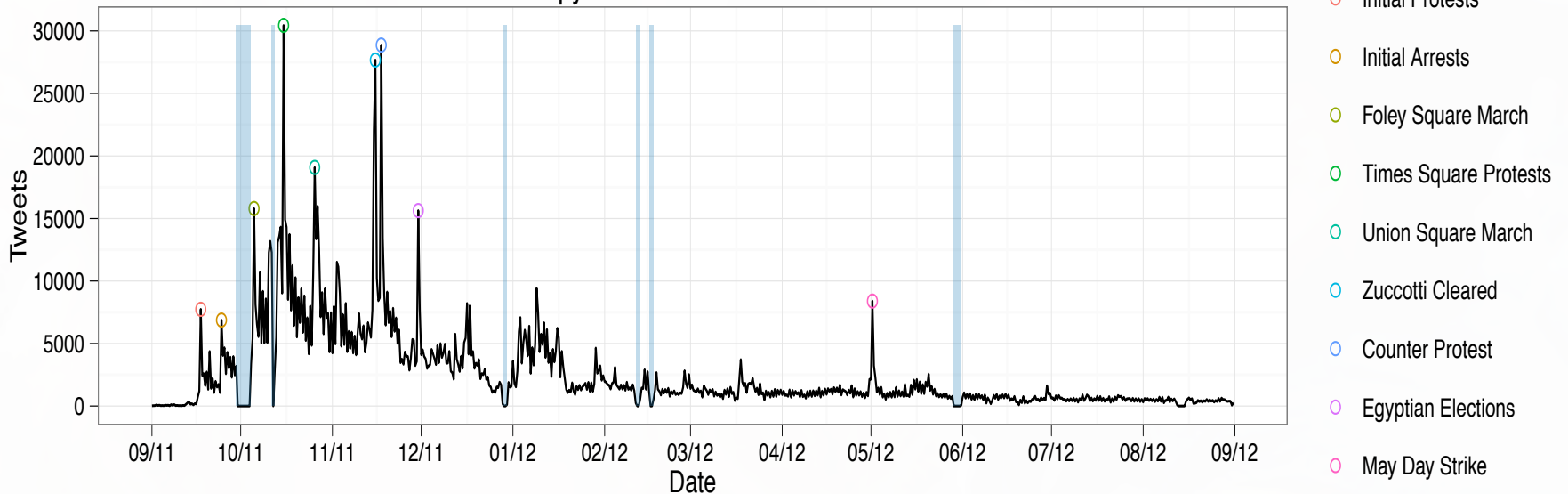
Words were ranked according to their Term Frequency – Inverse Document Frequency, Represented here in Log-scale

Occupy Wall Street



Twitter traffic & timeline

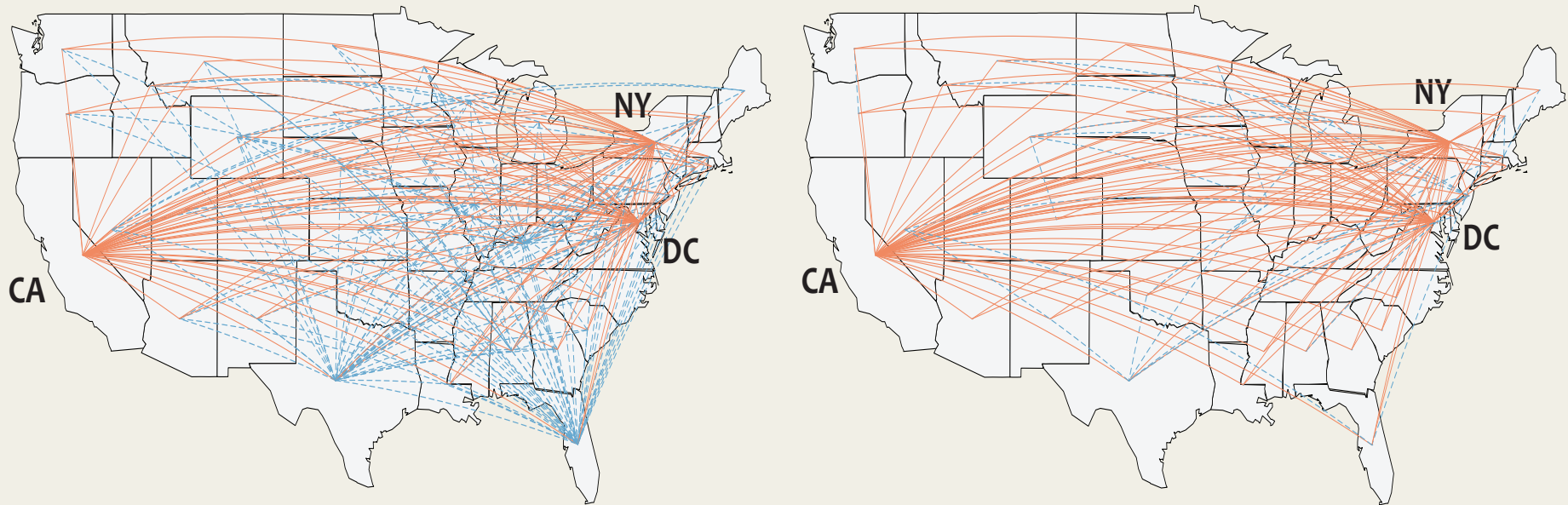
Occupy Wall Street Traffic Volume



On-the-ground events (circles).

Twitter data-stream outages (blue bands) are highlighted. Bins have 12-hours length.

The geography of online OWS



Multi-scale backbone extraction – confidence level $\alpha = 0.15$

Occupy-related discourse (right) shows a prominent hub-and-spoke structure with features distinct from conversations on domestic politics (left).

Collective Framing vs Resource mobilization

Collective framing

Interstate		Intrastate	
Token	Ratio	Token	Ratio
wall	.590	city	2.254
nyc	.600	tonight	1.737
street	.699	march	1.669
news	.718	join	1.494
99%	.756	solidarity	1.387
bank	.763	day	1.354
don't	.782	square	1.333
media	.837	please	1.243
peaceful	.845	park	1.220
nypd	.847	now	1.179

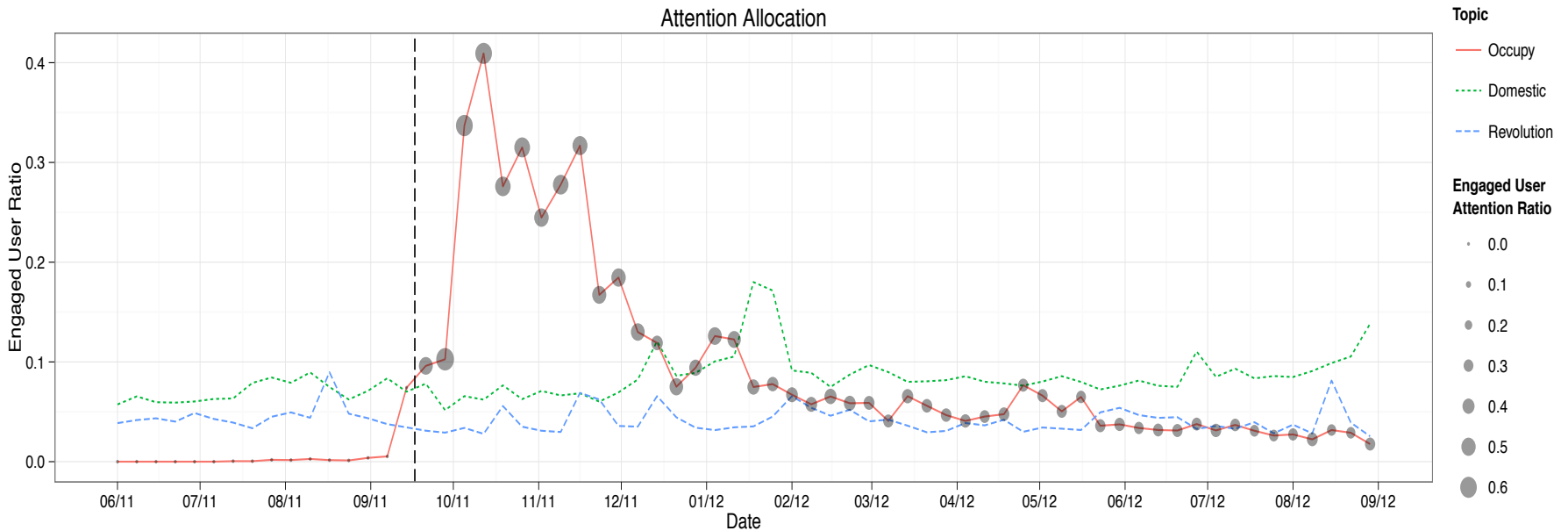
Resource mobilization

$$\text{Ratio} = \frac{P(\text{Token}|\text{Intrastate})}{P(\text{Token}|\text{Interstate})}$$

Collective framing: the social processes whereby participants negotiate the shared language and narrative frames to define the movement's identity and goals.

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

People engagement

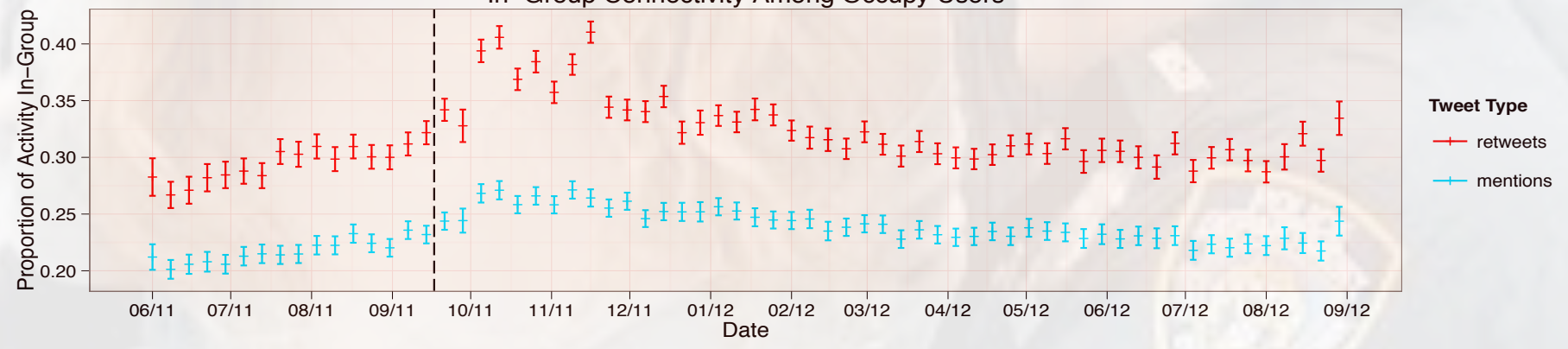


Point size is proportional to user engagement, measured as average (across users) ratio of OWS-related tweets over entire production

At the core of the protest

Domestic Politics	Social Movements
#tcot	#syria
#p2	#bahrain
#teaparty	#egypt
#gop	#yemen
#anonymous	#libya
#obama	#tahrir
#tlot	#wiunion
#jobs	#iranelection
#ronpaul	#assange
#romney	#wikileaks
#sopa	#jan25
#ndaa	#14feb
#obama2012	#assad
#ocra	#greece
#twisters	#damascus
#sgp	#gaddafi
#politics	#feb14
#solidarity	#scaf
#gop2012	#antisec
#p21	#arabspring
#topprog	#tunisia
#obamacare	#noscaf
#mapoli	#syrian

In-Group Connectivity Among Occupy Users



Longitudinal connectivity level of 25k users – reported as mean and 95% confidence intervals

Conclusions

The background features a network of stylized human figures. Most are light blue, but one figure in the center foreground is a darker orange. They are interconnected by thin white lines, suggesting a social network or data flow. The figures are arranged in a somewhat circular pattern, with some overlapping.

SM data can be leveraged to answer questions relevant to social science

Technical issues I carefully hide: data collection, filtering, archiving, retrieval, analysis

May possible biases

Opportunity to observe in the “wild”,
without relying o what people says about the self or others

Complement traditional approaches