Social Media &
The Networked Public Sphere

Michael Conover

@vagabondjack
truthy.indiana.edu

SCHOOL OF INFORMATICS AND COMPUTING
INDIANA UNIVERSITY
Bloomington

CNetS | Center for Complex Networks and Systems Research
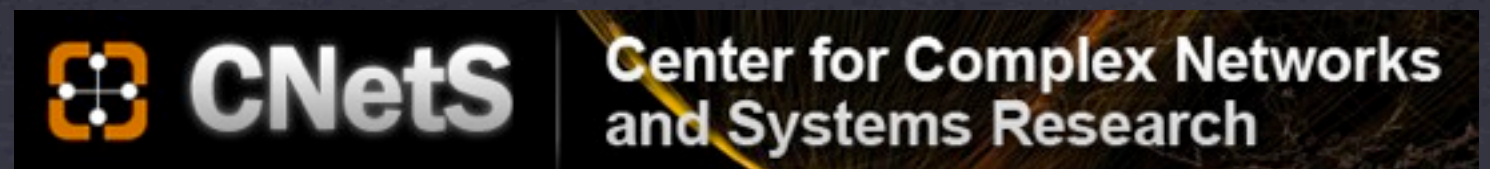
Vignettes

Setting the Stage

Polarization

Prediction

Asymmetry

Looking Forward

Information & Democracy

Deliberative Democracy

The Networked Public Sphere

A Phase Transition

[1] Benkler, Y. The Wealth of Networks (2006)
[2] Habermas, J. The Structural Transformation of the Public Sphere (1962)

A Disruptive Technology

Accelerants

Key Components

Backlash from the State

[1] Aday, S. Blogs & Bullets: New Media in Contentious Politics (2010)
[2] Gladstone, B. The Influencing Machine (2011)

Filter Bubbles

General Interest Intermediaries

Birds of a Feather

Social Pressures

[1] Sunstein, C. Republic.com 2.0 (2007)

[2] McPherson, M. Birds of a Feather: Homophily in Social Networks (2001)

[3] Conover, M.D., et al. Political Polarization on Twitter (2011)

ASTROTURF DETECTION

#P2   #TEAPARTY   #TLOT

#NVSEN   #GOP   #TCOT

NETWORKS OF POLITICAL DISCOURSE

ATOMIC STRUCTURE

#P2   #TEAPARTY   #TLOT

#NVSEN   #GOP   #TCOT

NETWORKS OF POLITICAL DISCOURSE
ATOMIC STRUCTURE

Monday, September 26, 2011

NETWORKS OF POLITICAL DISCOURSE
AGGREGATE STRUCTURE

Monday, September 26, 2011
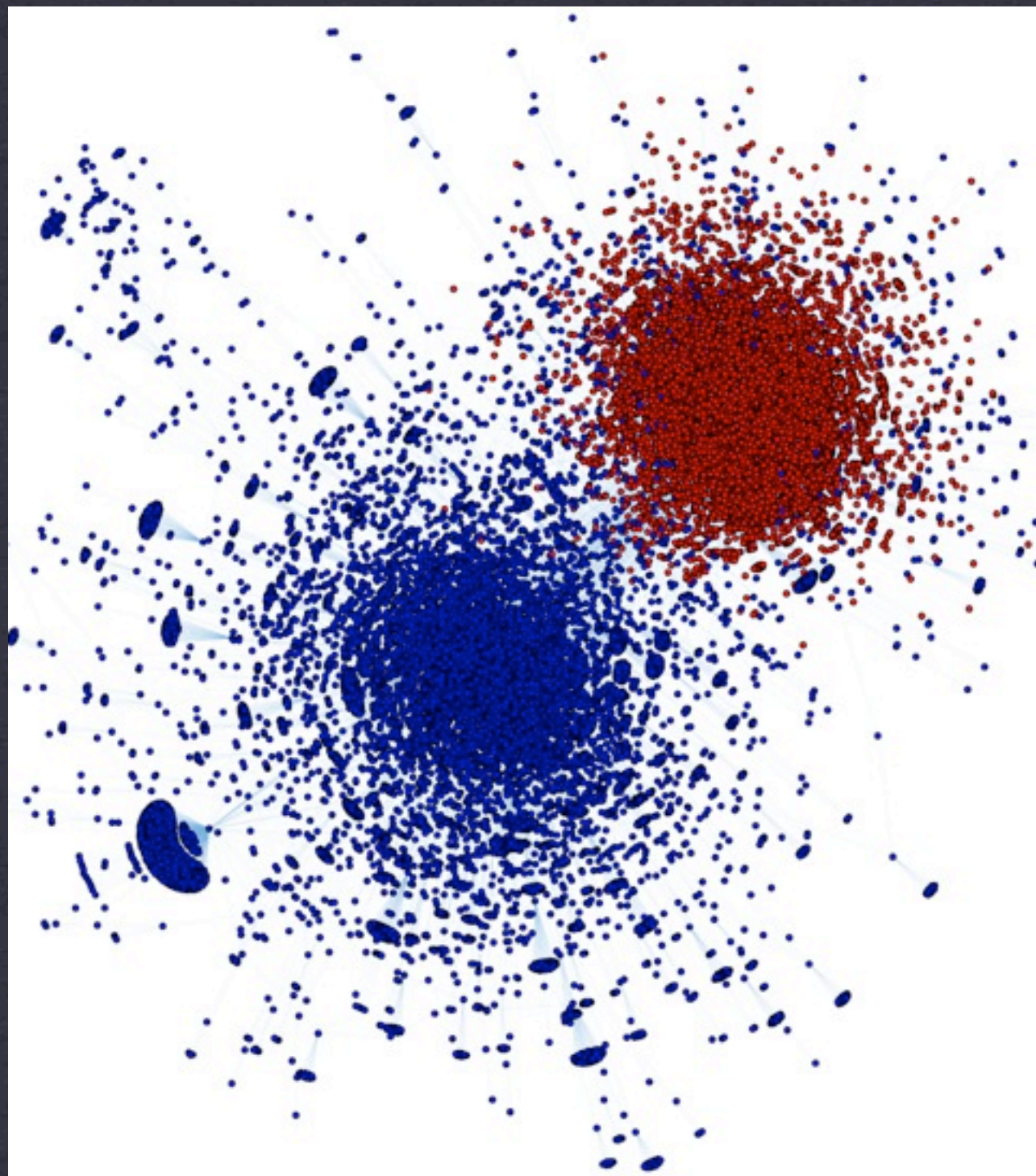
RETWEET
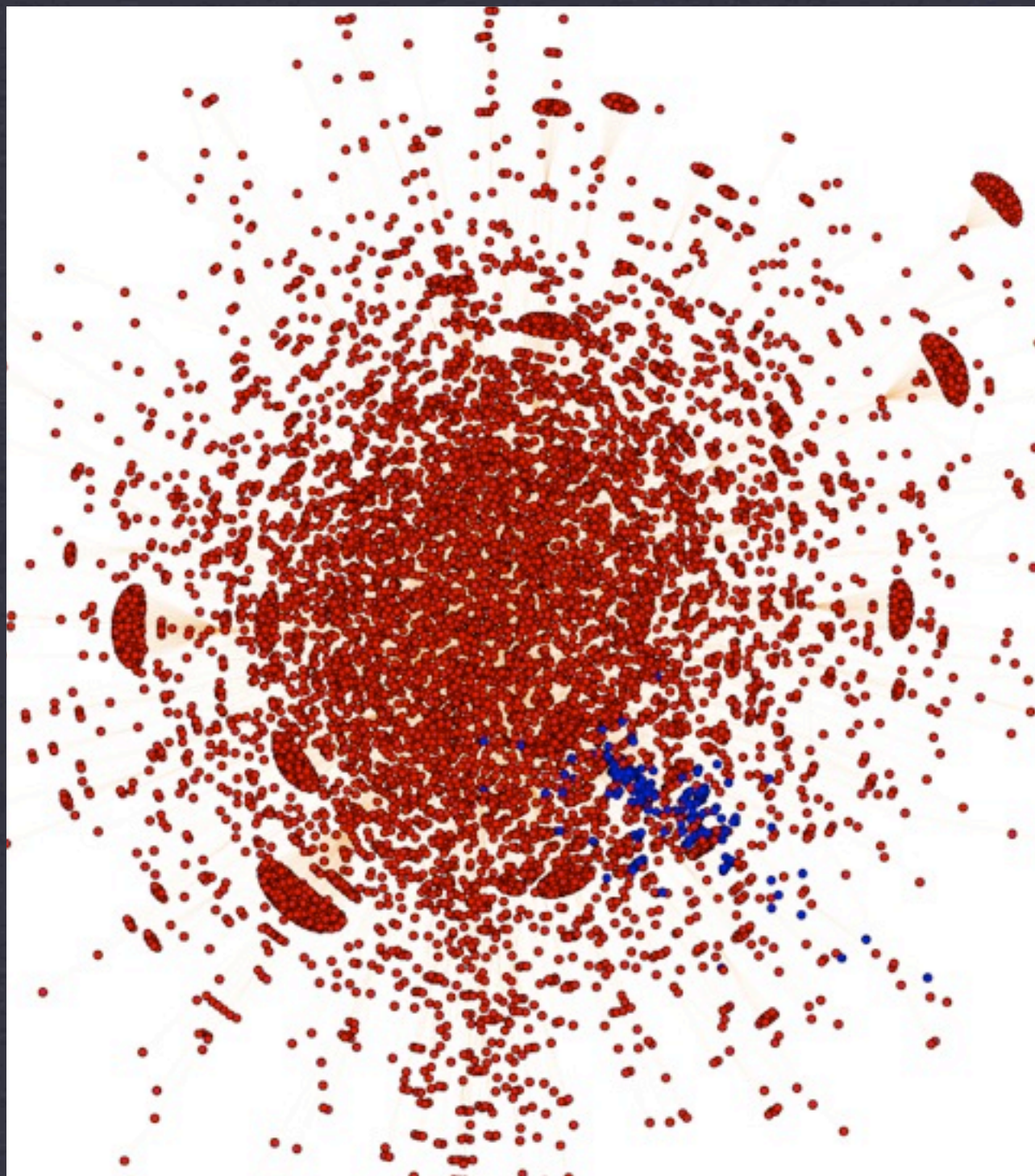
MENTION

NETWORKS OF POLITICAL DISCOURSE

MULTI-MODE COMMUNICATION

Monday, September 26, 2011

Cluster Analysis

Seeded Label Propagation

Hill-Climbing Algorithm

Inter-run Agreement
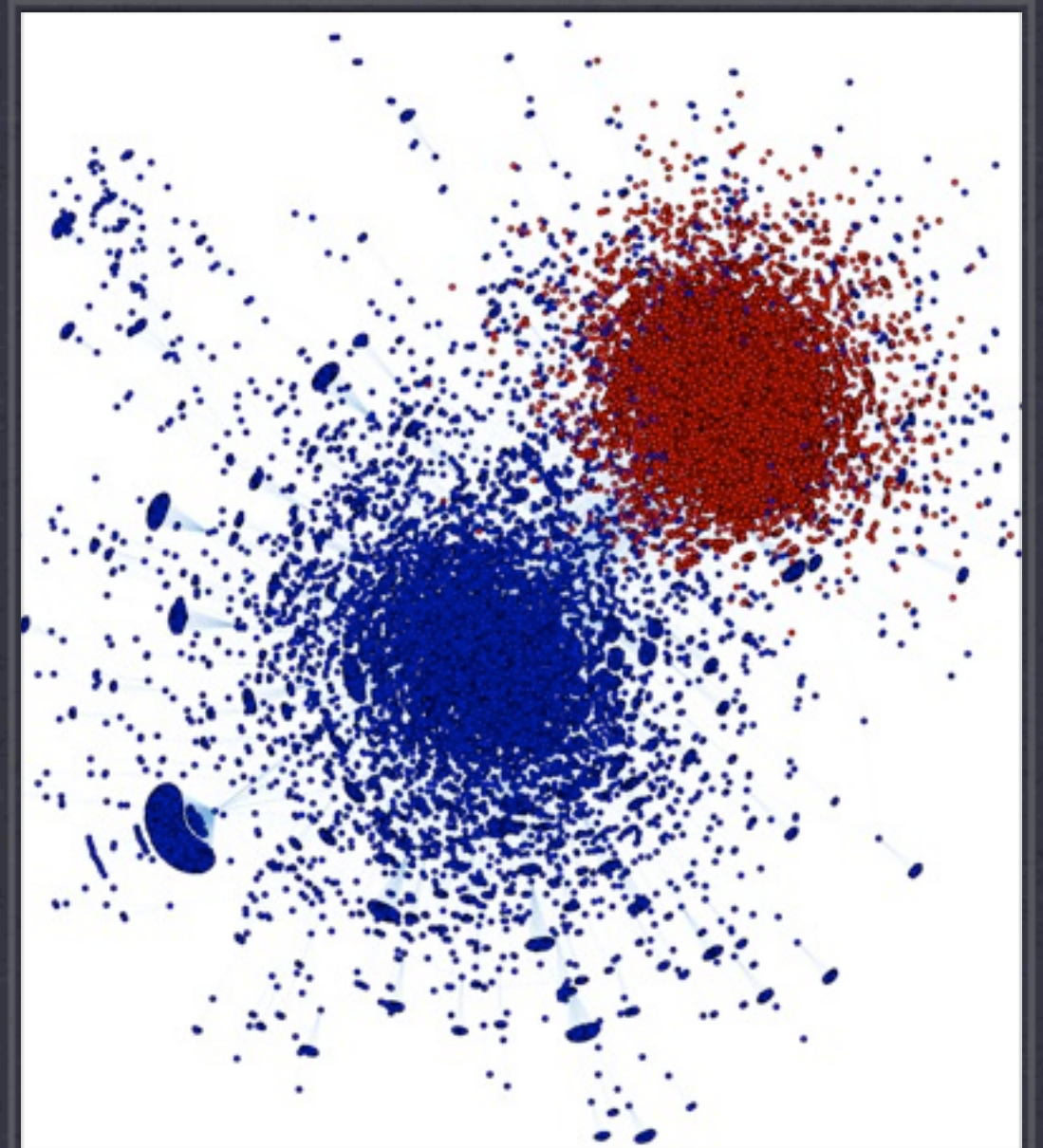
QUALITATIVE CONTENT ANALYSIS
LEFT UNDECIDABLE RIGHT

1,000 RANDOM USERS

2 AUTHORS, 1 NON-AUTHOR JUDGE

ANALYTICAL RESULT
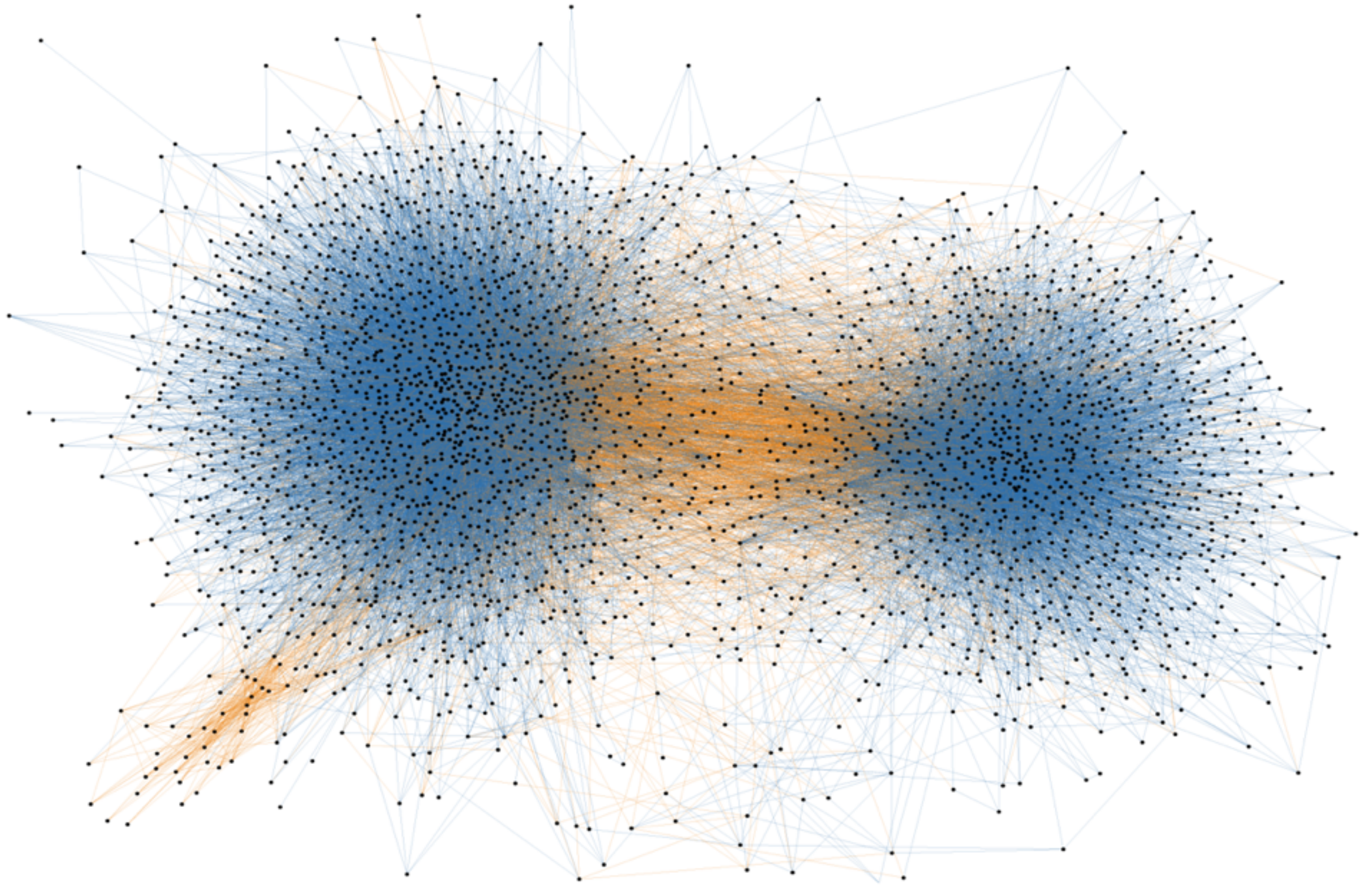
$$E[R \rightarrow L] = k_R \cdot \frac{U_L}{U_L + U_R}.$$

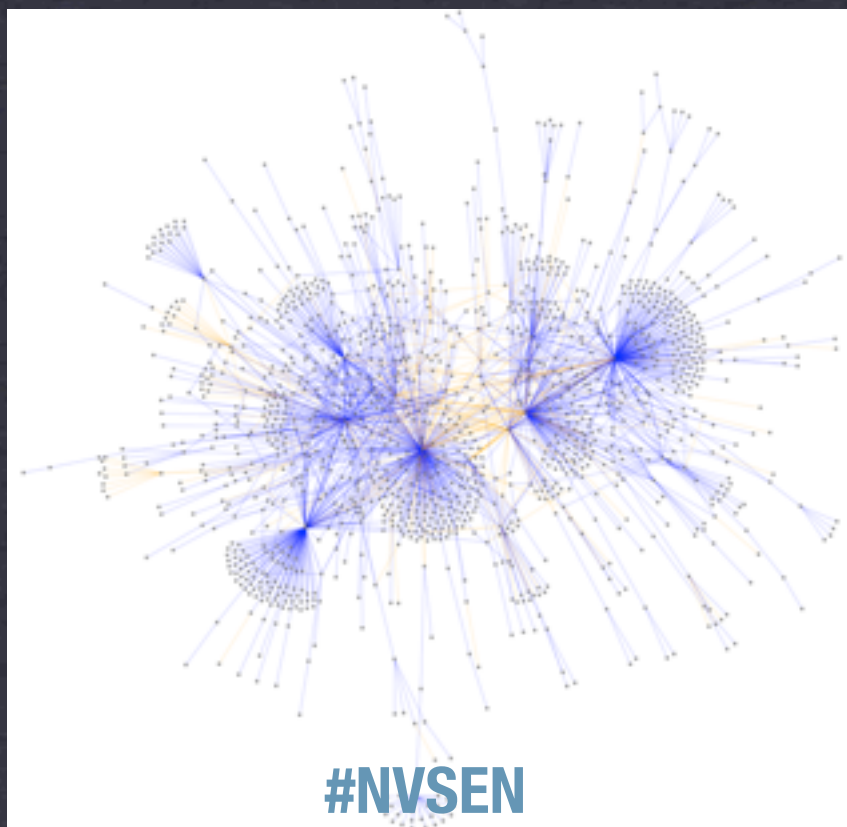|	Cluster	Left	Right	Und.	Nodes
Retweet	A	1.1	**93.4**	5.3	7,115
	B	**80.1**	8.7	11.1	11,355
Mention	A	**39.5**	**52.2**	8.1	7,021
	B	9.5	85.7	4.7	154
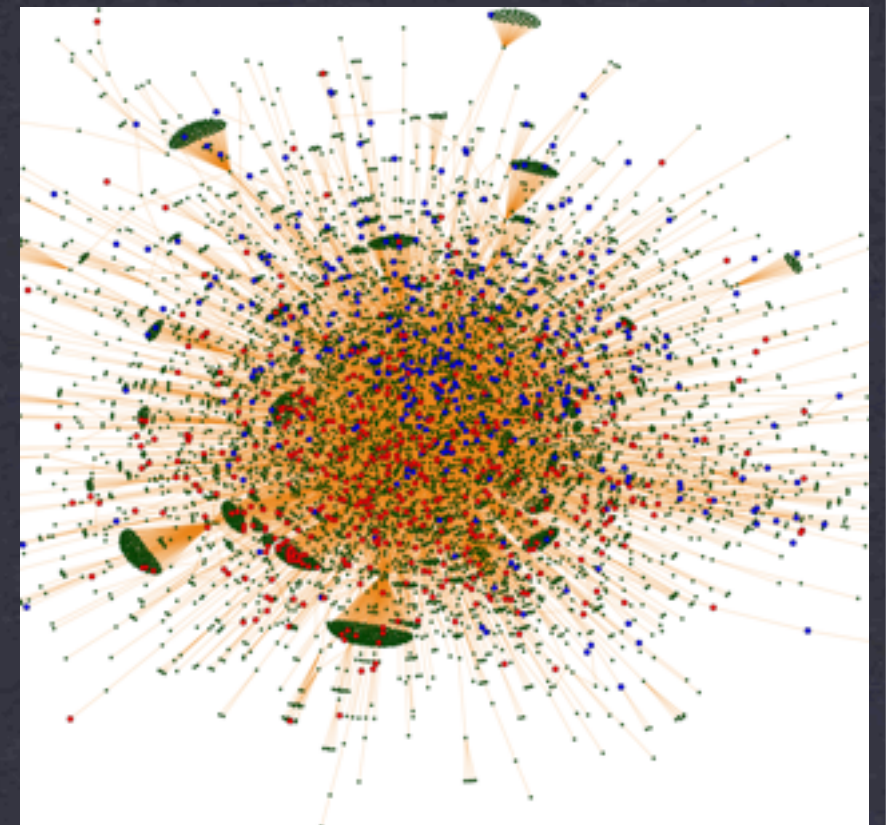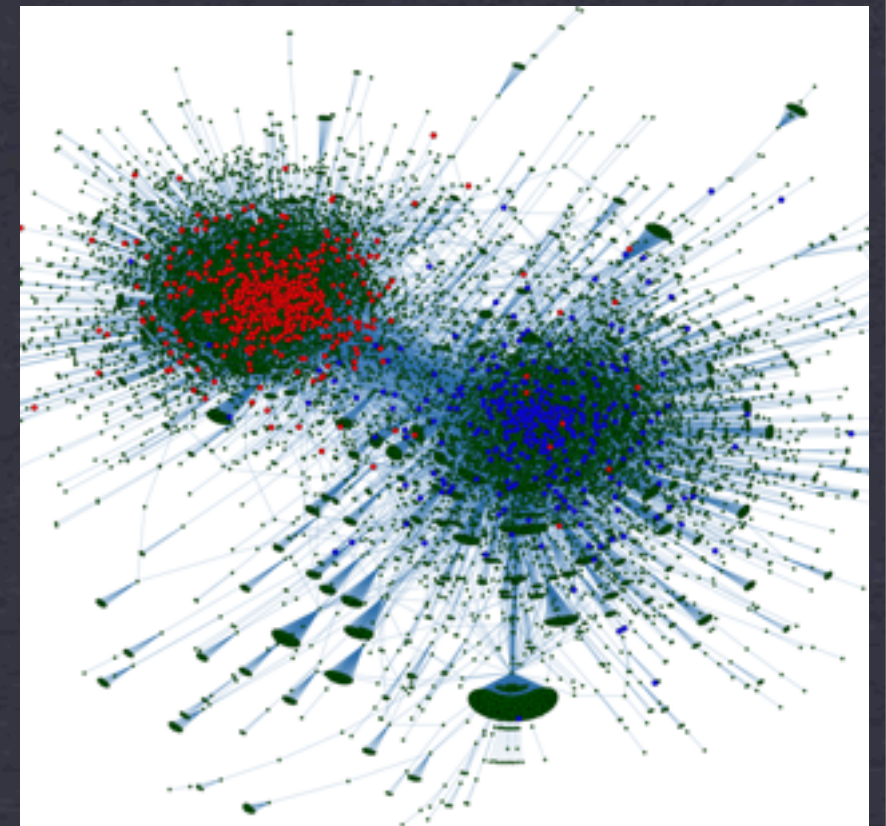
POLITICAL IDENTITY
PARTISAN DIVIDE

CROSS-IDEOLOGICAL INTERACTION
AGGREGATE STRUCTURE

Monday, September 26, 2011

Please follow @username for an outstanding progressive voice! **#p2 #dems #prog #tcot**

Couple Aborts Twin Boys for Being Wrong Gender .. http://bit.ly/xyz **#tcot #christian #teaparty #p2 #prolife**

Content Injection

Content Injection

IMPORTANCE OF A CONTENT STREAM

$$I(t, A) = \frac{N(t, A)}{\sum_t N(t, A)}$$

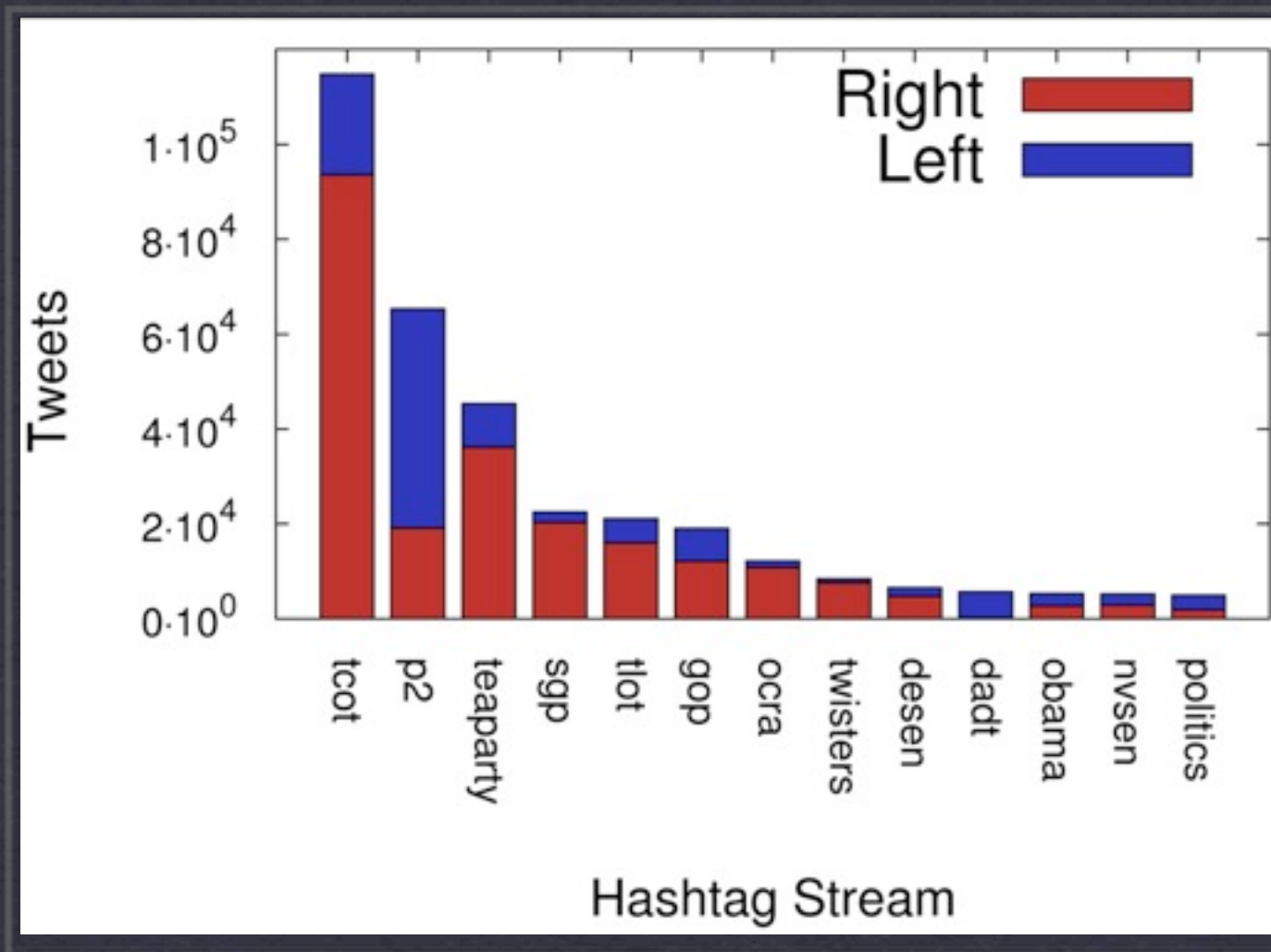| Far Left	Moderate Left	Centrist	Moderate Right	Far Right
#healthcare	#aarp	#democrats	#rangel	#912project
#judaism	#citizensunited	#social	#waste	#twisters
#hollywood	#democratic	#seniors	#saveamerica	#gop2112
#2010elections	#banksters	#dnc	#american	#israel
#capitalism	#energy	#budgets	#gold	#foxnews
#security	#sarahpalin	#political	#repeal	#mediabias
#publicoption	#progressives	#gopproud	#mexico	#constitution
#recession	#stopbeck	#christian	#terrorism	#patriots
#dreamact	#iraq	#media	#gopleader	#rednov
#topprogs	#women	#nobel	#palin12	#abortion

$$V(t) = 2 \frac{I(t, R)}{I(t, R) + I(t, L)} - 1$$

PARTISAN LANGUAGE

MEASURE **VALENCE**   TABLE **POLITICAL HASHTAGS**

Cross-Ideological Interaction

CROSS-IDEOLOGICAL INTERACTION
AGGREGATE STRUCTURE

Predicting Political Affiliations

4 Billion Dollar Industry

Peddling Influence

Supplement Traditional Polling

Learning Scenario

DAISY (1964)

[1] Conover, M.D., et al. Predicting the Political Alignment of Twitter Users (2011)

Full Text Baseline

Linear SVM Trained on Unigrams

Stopwords, Mentions, URLs, Hashtags

$$TFIDF_{ij} = \frac{n_{ij}}{\sum_k n_{k,j}} \cdot \log \frac{|U|}{|U_i|}$$

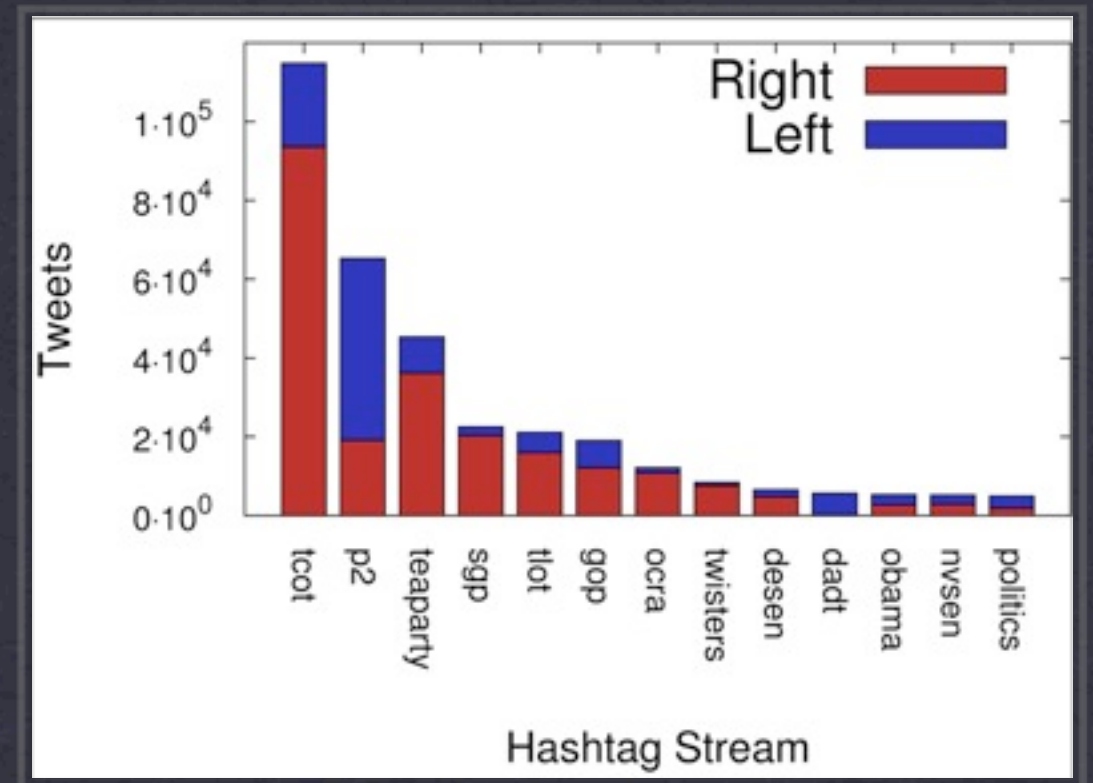| Confusion Matrix			Accuracy
	R	L	79.2%
R	266	107	
L	75	431	Lower Bound: 72.6%

Stream Identifiers

Linear SVM Trained on Hashtags

Content Injection

Generalizability

Condense Topic Information

Confusion Matrix			Accuracy
	R	L	
R	**331**	**42**	90.8%
L	**41**	**465**	Lower Bound: 83.5%

Topic Detection

Latent Semantic Analysis of Hashtags

TABLE VI

MOST EXTREME HASHTAG COEFFICIENTS FOR SECOND LEFT SINGULAR VECTOR. THIS LINEAR COMBINATION OF HASHTAGS APPEARS TO CAPTURE VARIANCE ASSOCIATED WITH POLITICAL ALIGNMENT.

Hashtag	Coeff.	Hashtag	Coeff.
#tcot	0.380	#p2	-0.914
#sgp	0.030	#dadt	-0.071
#ocra	0.020	#p21	-0.042
#hhrs	0.013	#votedem	-0.039
#twisters	0.012	#lgbt	-0.038
#tlot	0.011	#p2b	-0.032
#whyimvotingdemocrat	0.009	#topprog	-0.027
#rs	0.005	#onenation	-0.025
#ftrs	0.004	#dems	-0.023
#ma04	0.004	#gop	-0.021
#tpp	0.003	#hcr	-0.017
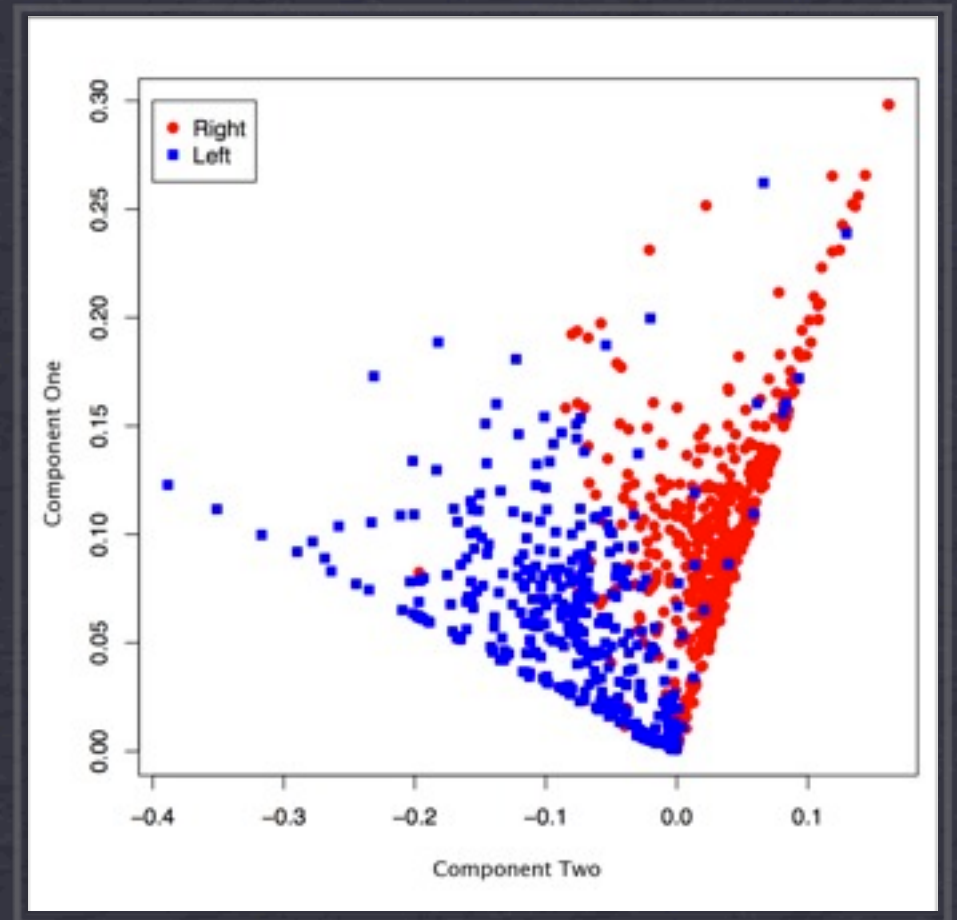
Cluster Membership

Boolean Classification

Generalizability

Composite Performance
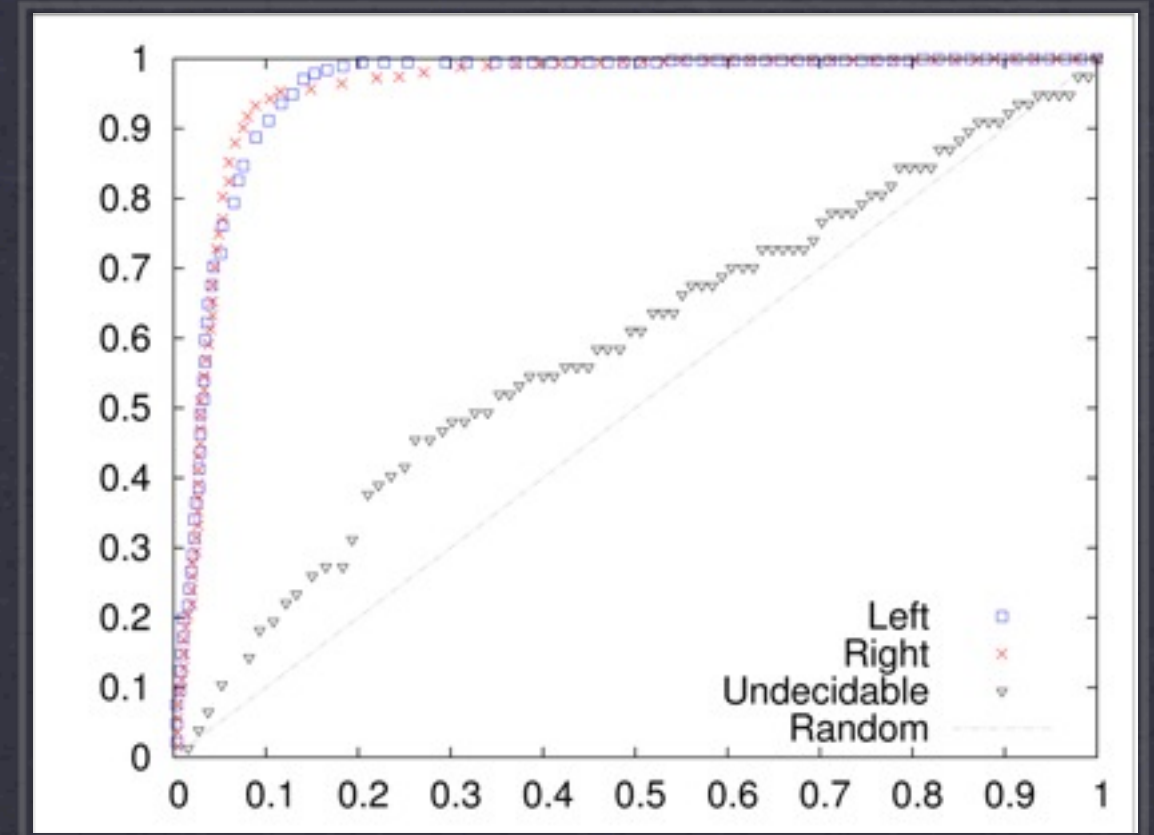
COMPOSITE ROC CURVE

Confusion Matrix			Accuracy
	R	L	**94.9%**
R	366	6	
L	38	468	Lower Bound: 87.3%

Knowledge Discovery

$4 Billion Political Advertising Industry
Non-Obvious Behavioral Data
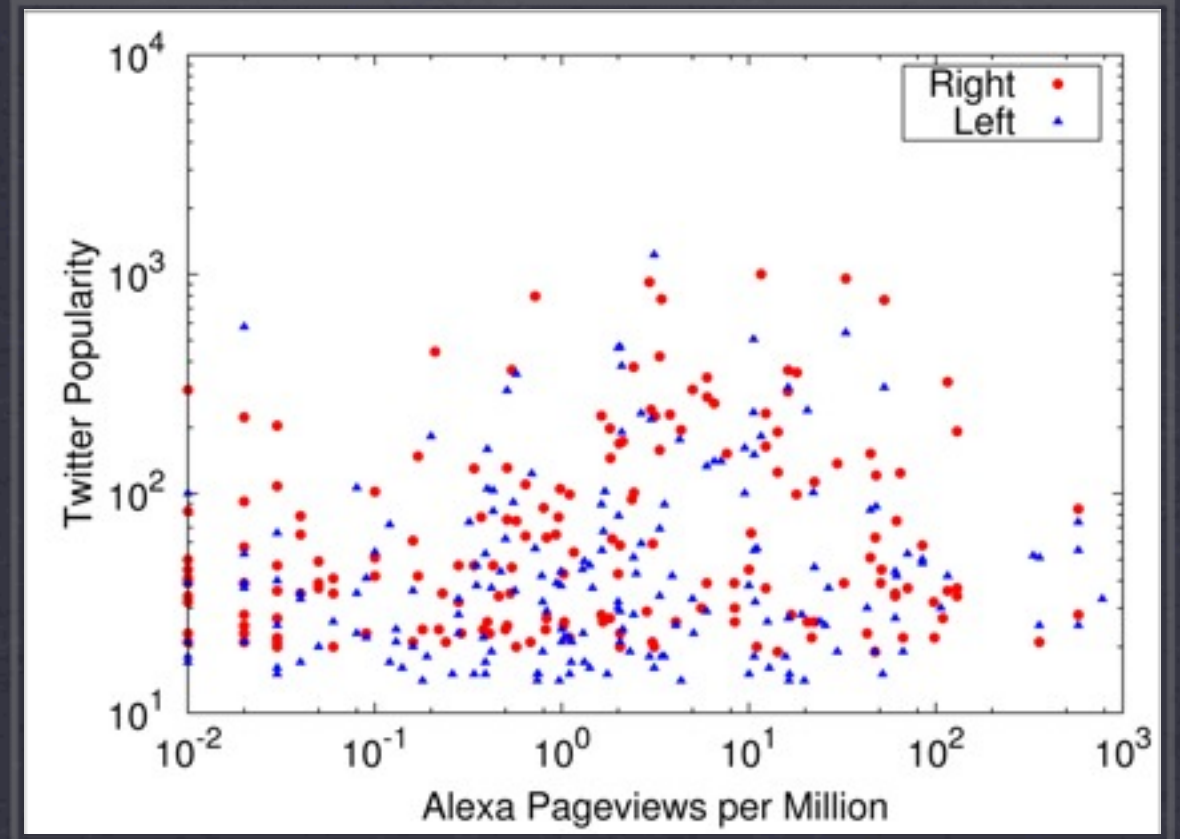
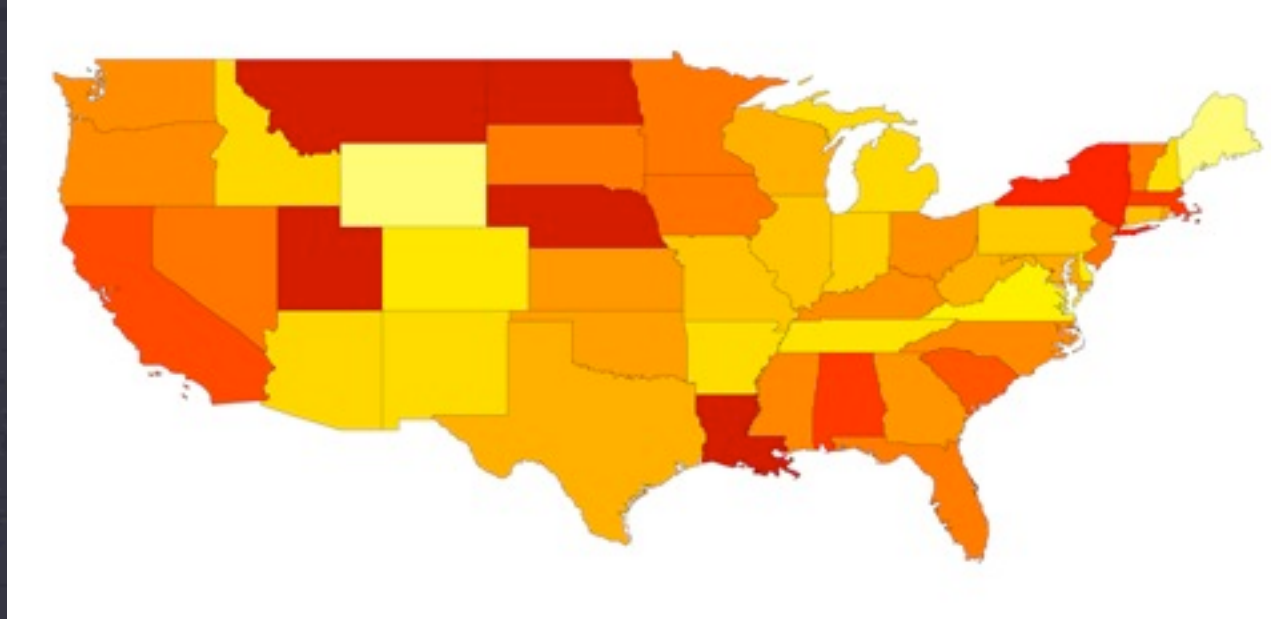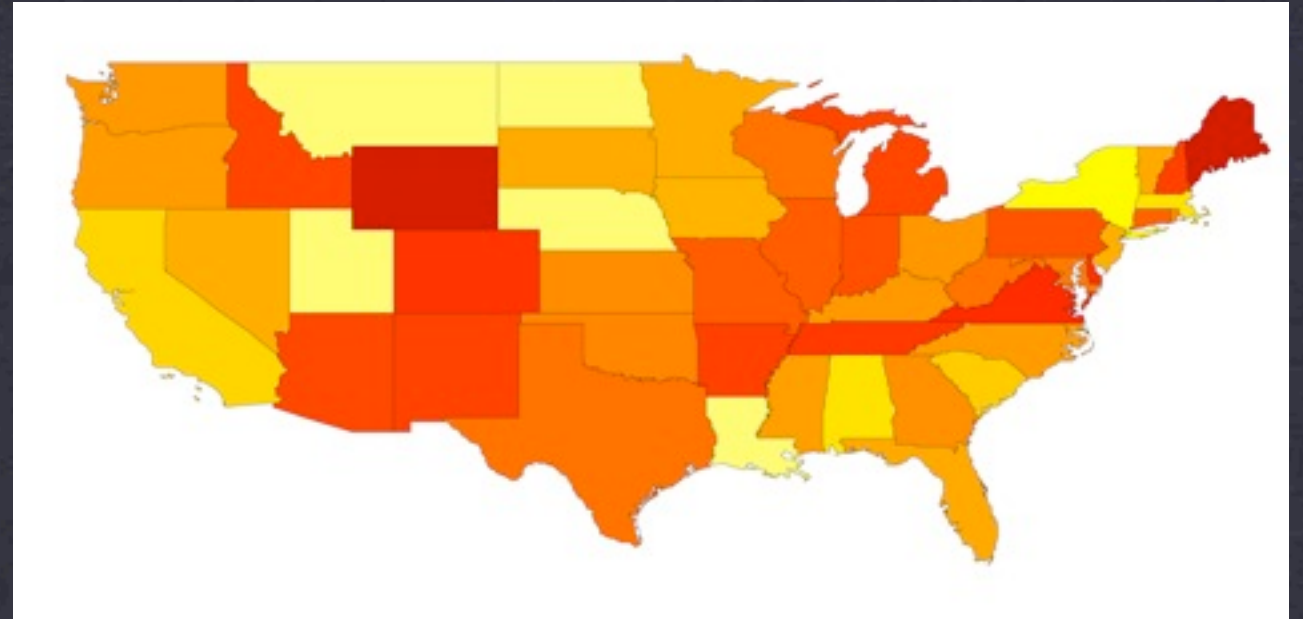| Popular Left	Popular Right
feedproxy.google.com	feedproxy.google.com
mediamatters.org	hotair.com
politicalwind.com	gop2112.com
youtube.com	youtube.com
dailykos.com	redstate.com
truthy-out.org	firstthings.com
msnbc.msn.com	americanthinker.com
harryreid.com	google.com
www.google.com	survivalstation.org
realclearpolitics.com	newsbusters.org
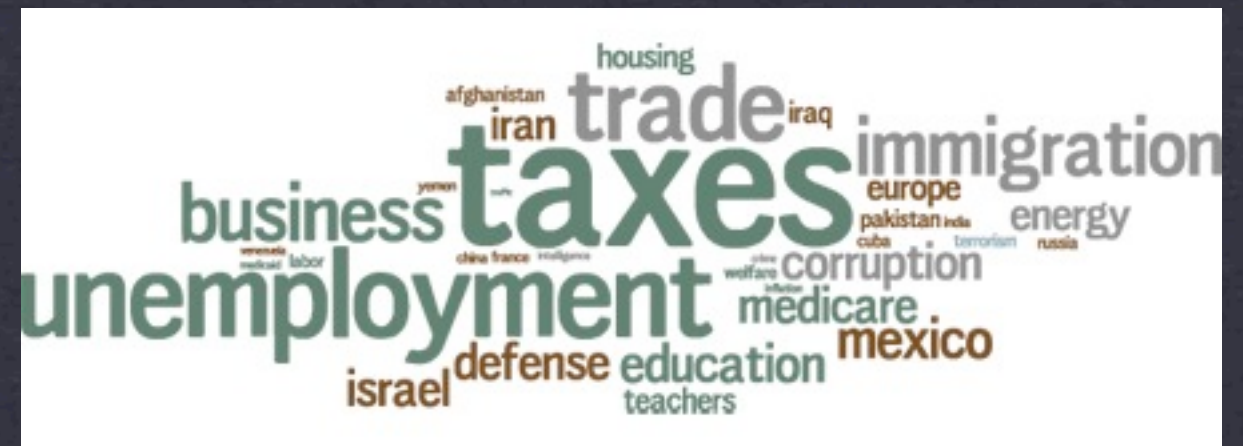
LEFT-LEANING   RIGHT-LEANING

POLICY ISSUES

LEFT-LEANING

RIGHT-LEANING

POLITICAL GEOGRAPHY

Monday, September 26, 2011

Information Consumption

Online News Consumption is a Core Political Behavior

Social Media is a Major Component

Tangible Changes

Distribute Political Information

Focus on Ideological Issues

Formation of Advocacy Groups

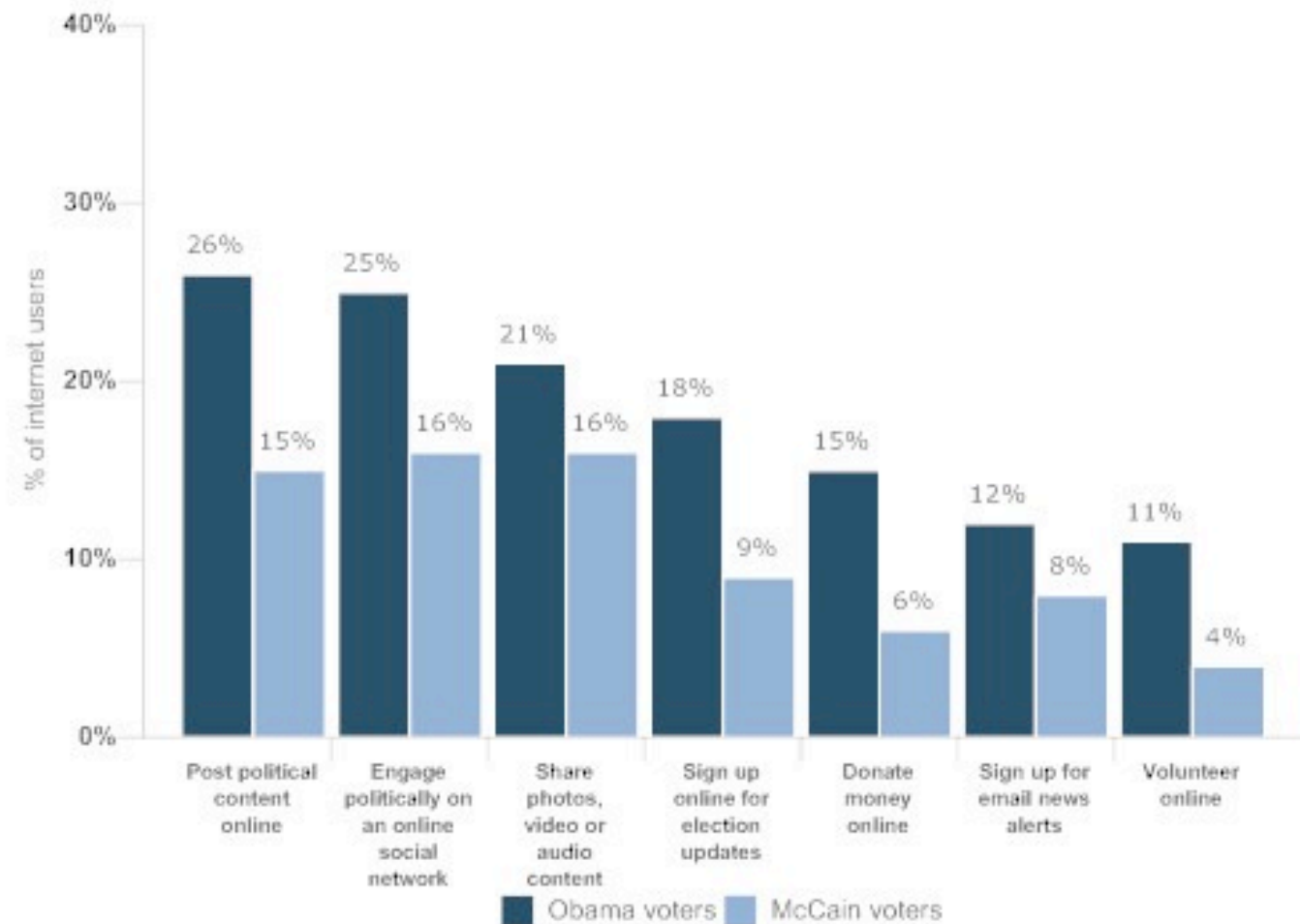**Aggregation of Small-Scale Acts**

Political Dividends

[1] Conover, M.D., et al. Partisan Asymmetries in Online Political Activity (invited)

Historical Topology

Obama voters lead the way in online political activism — Key differences between online McCain and Obama supporters. Pew Internet & American Life Project Post-Election Survey, November-December 2008.

[1] Pew Internet & American Life Project. The Internet's Role in Campaign 2008 (2009)
[2] Habermas, J. The Structural Transformation of the Public Sphere (1962)
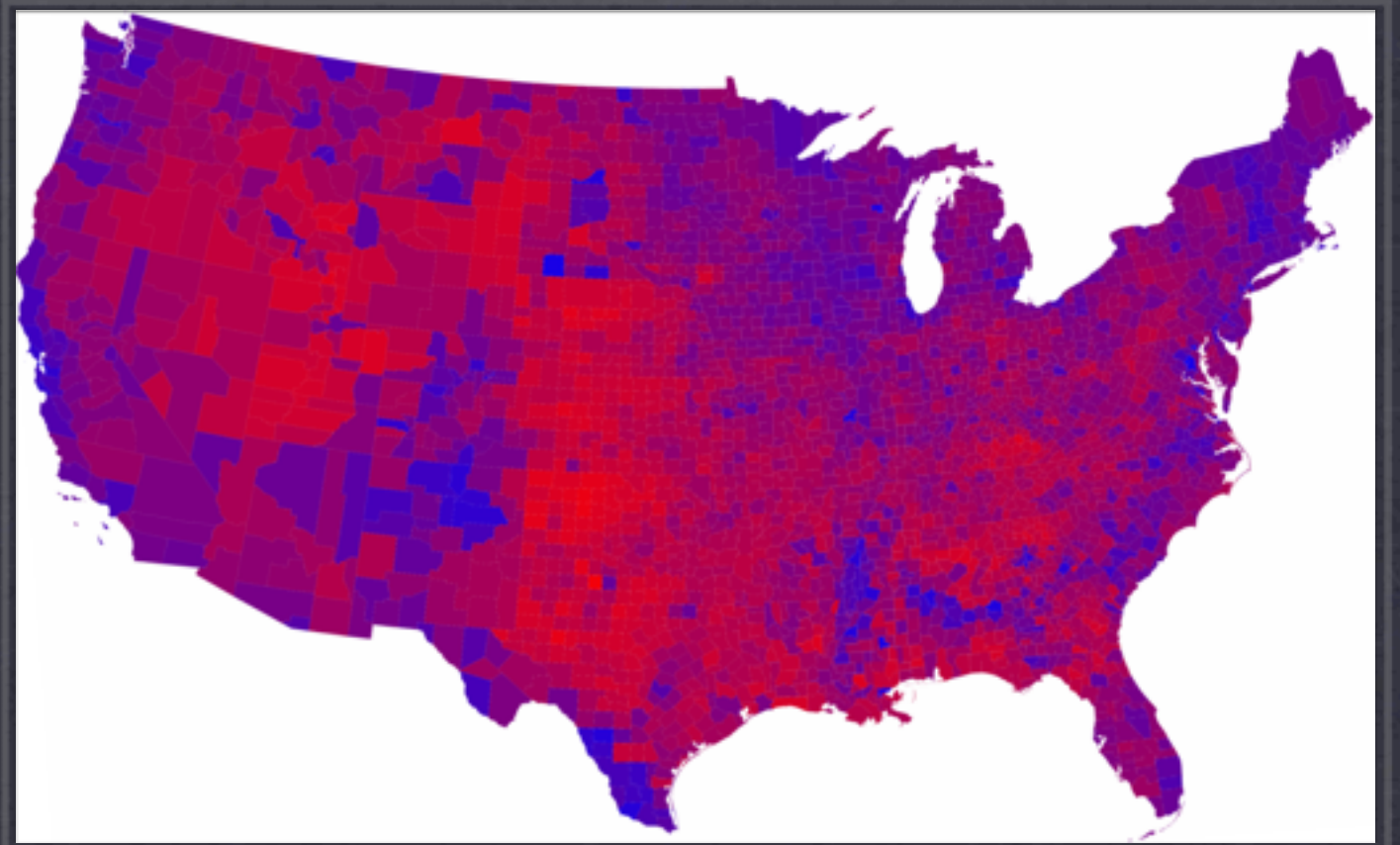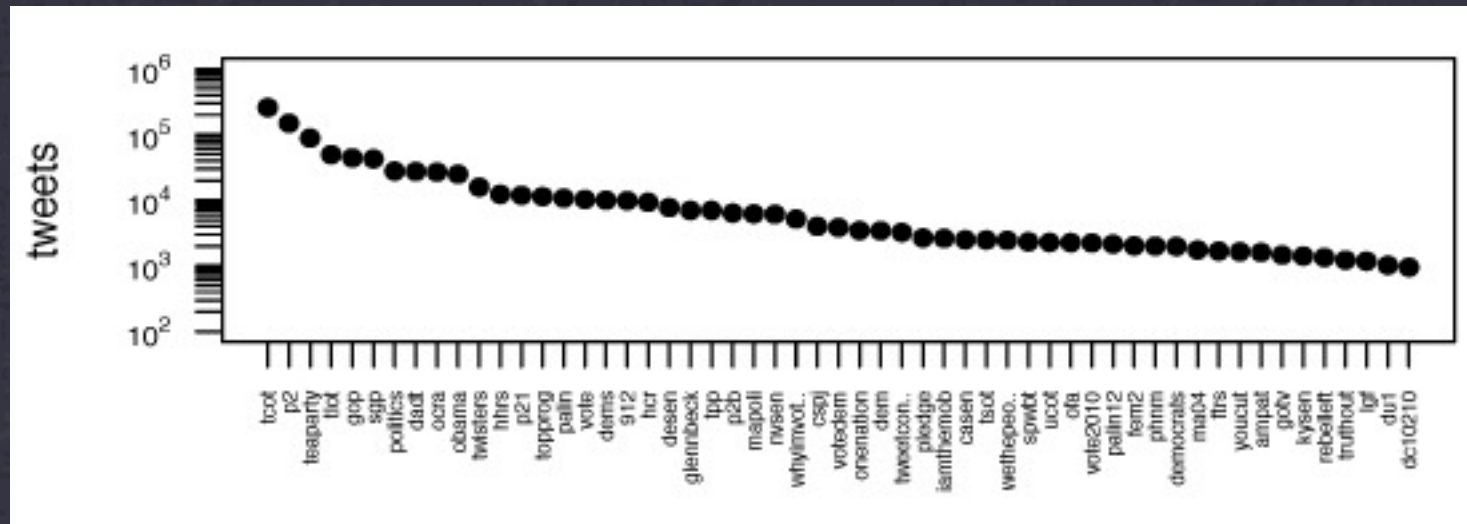
Framework

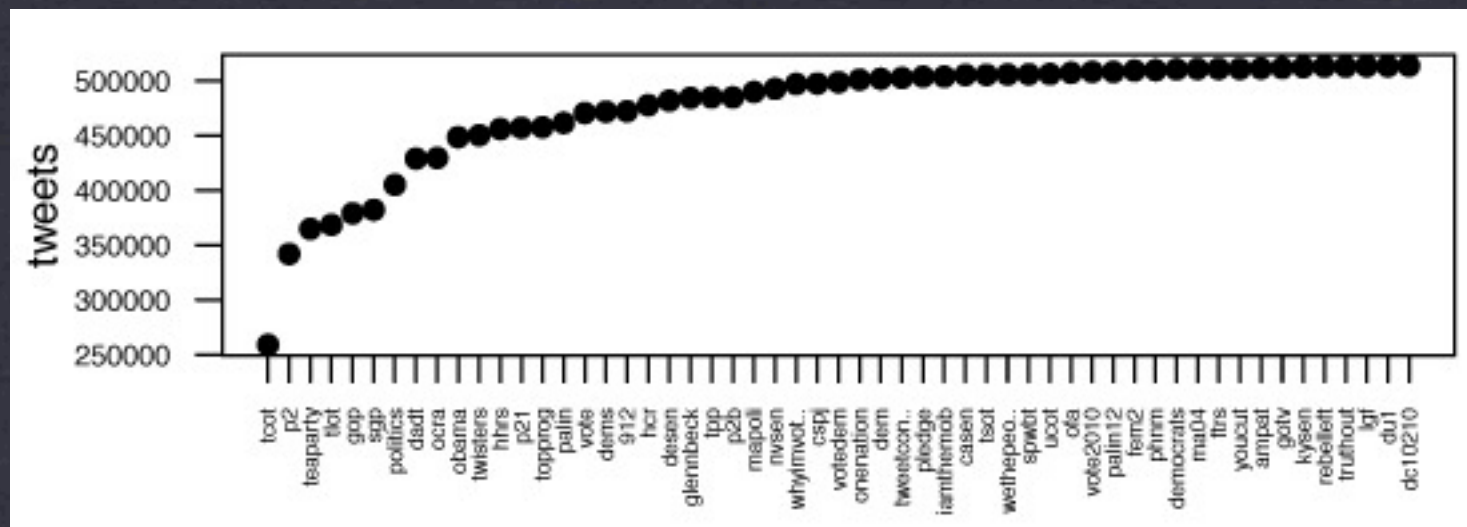## Expanded Dataset

- Timespan
- Social Network
- Singletons
- Non-Political

Behavior & Connectivity

Tweets Associated with Each Hashtag
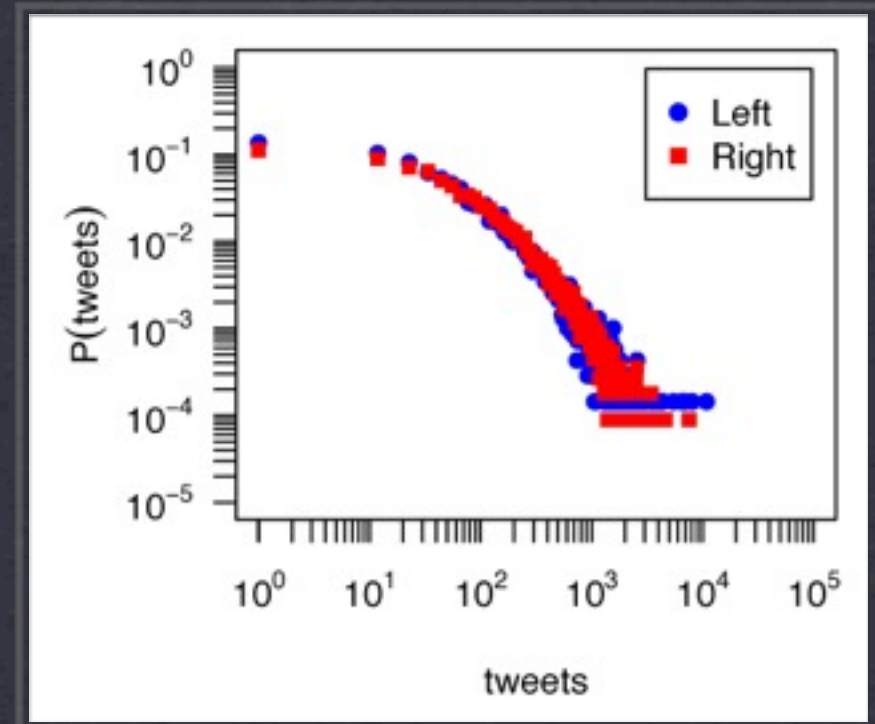

Cumulative Unique Tweets
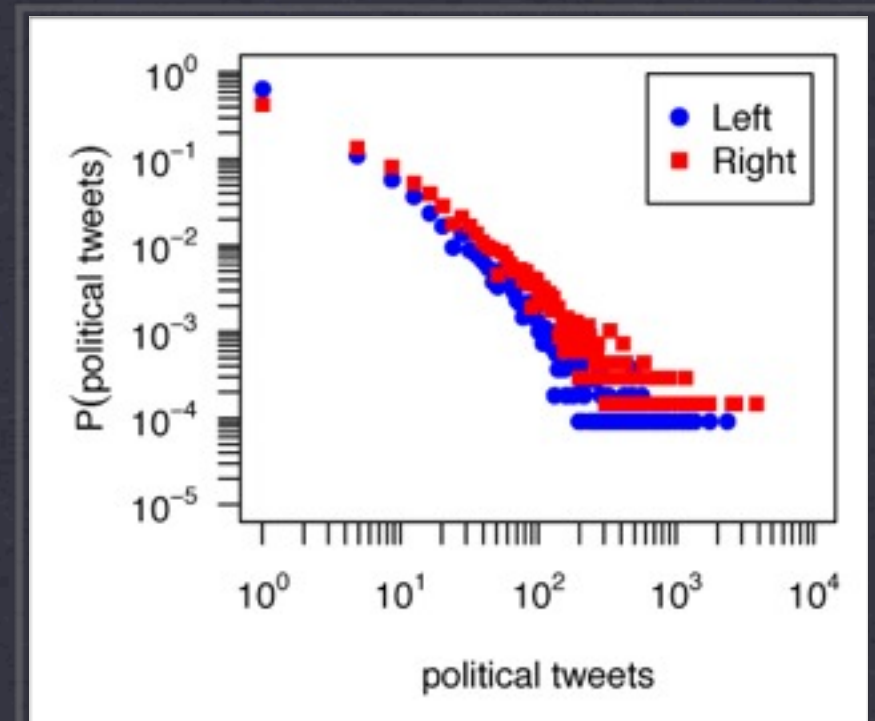
Representativeness

Communication Behavior

Equiprobable to Tweet

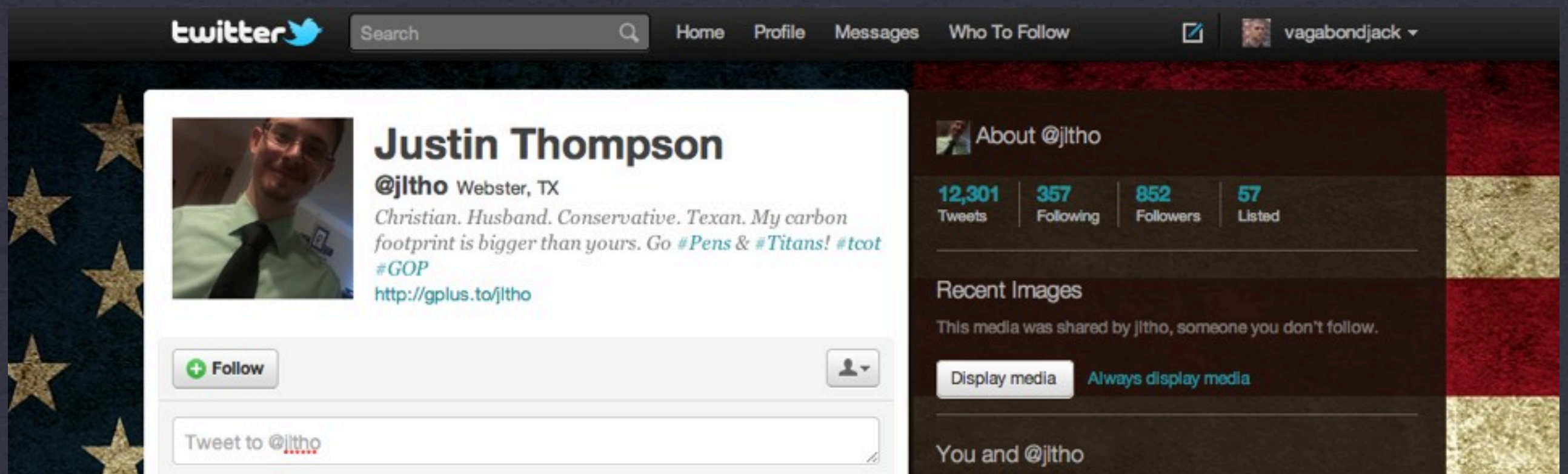85% more attention to political communication (22% vs. 12%)

54% more total political content

ALL TWEETS

POLITICAL TWEETS
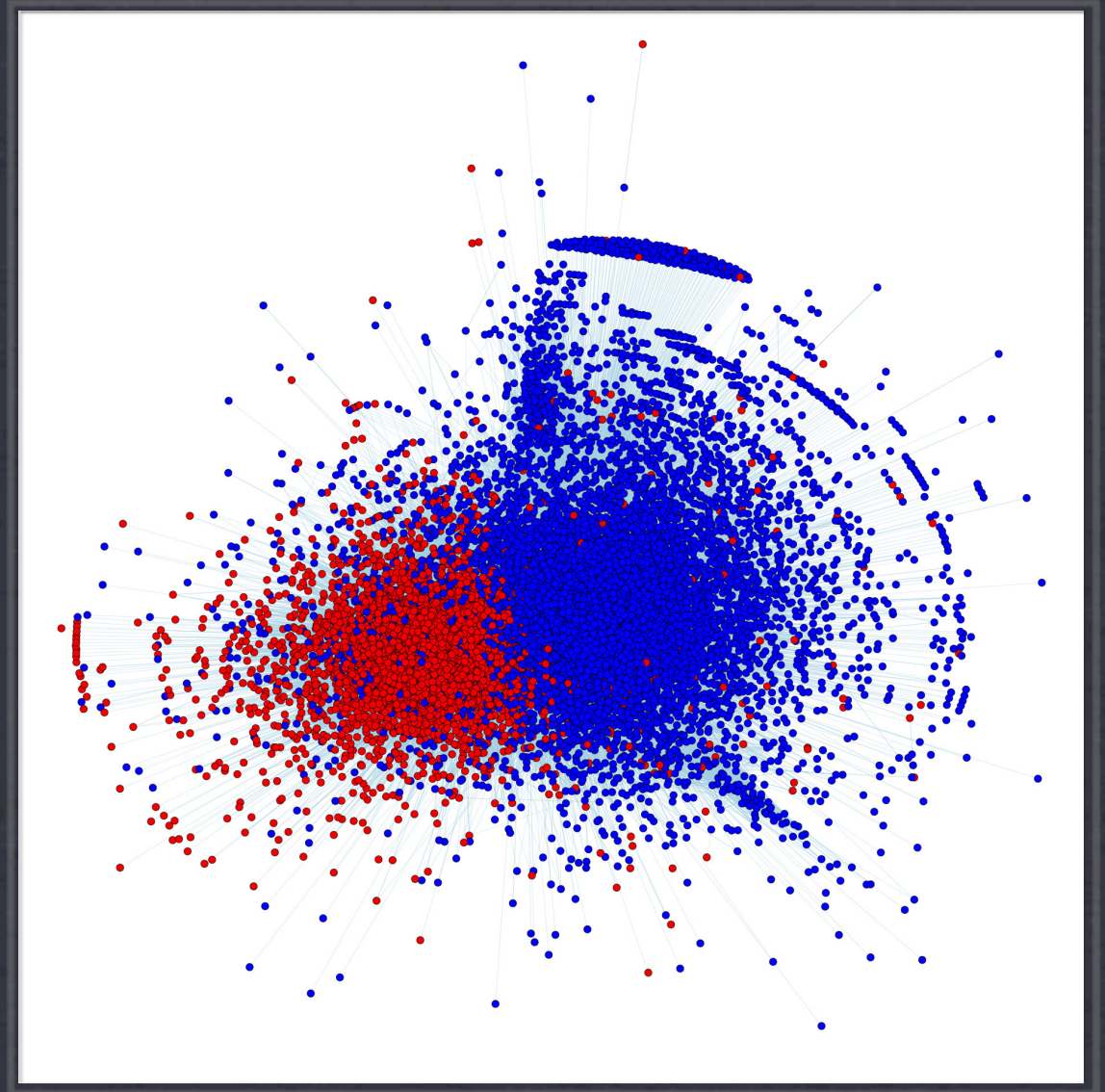
Communication Behavior

Partisan Self-Identification (38.7% vs. 24.6%)
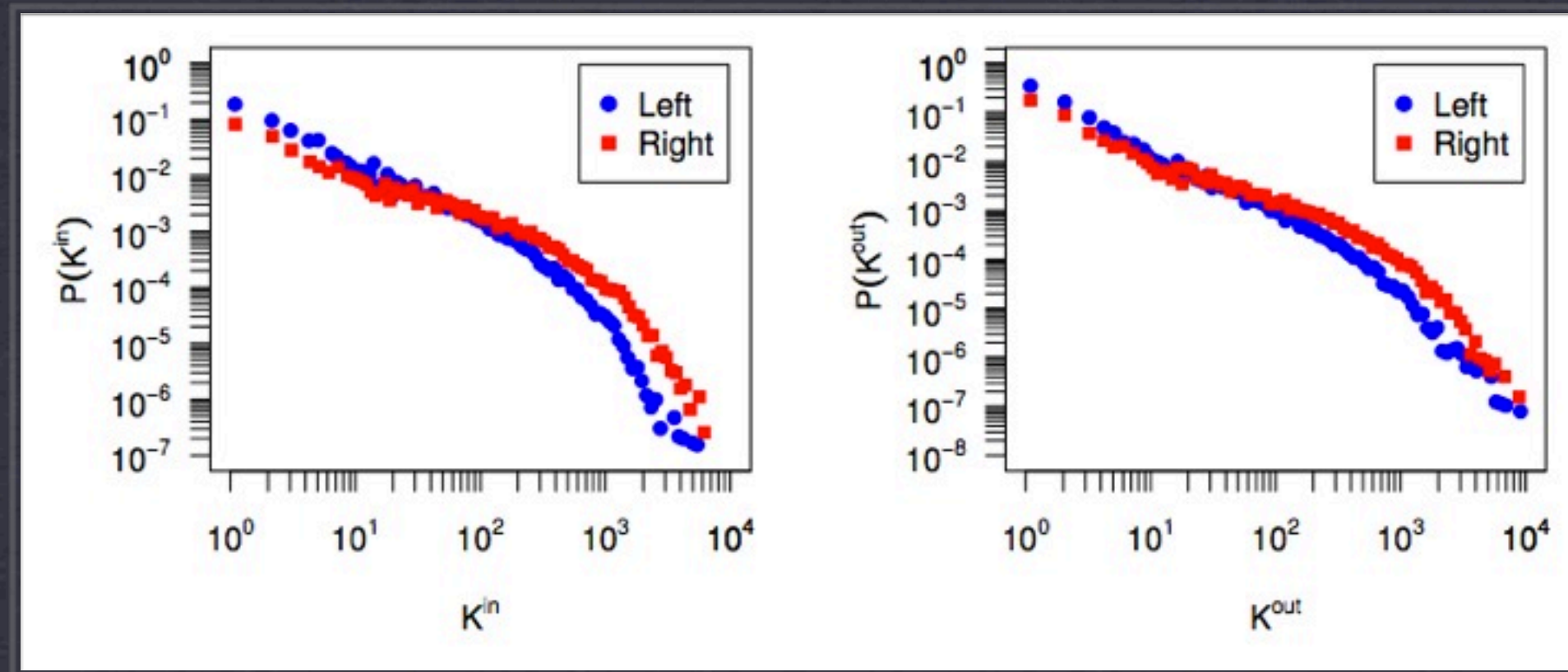
All Tweets: 43.4% vs. 36.5%

Political Tweets: 62.5% vs. 50.8%

Connectivity

Follower Network as Social Substrate
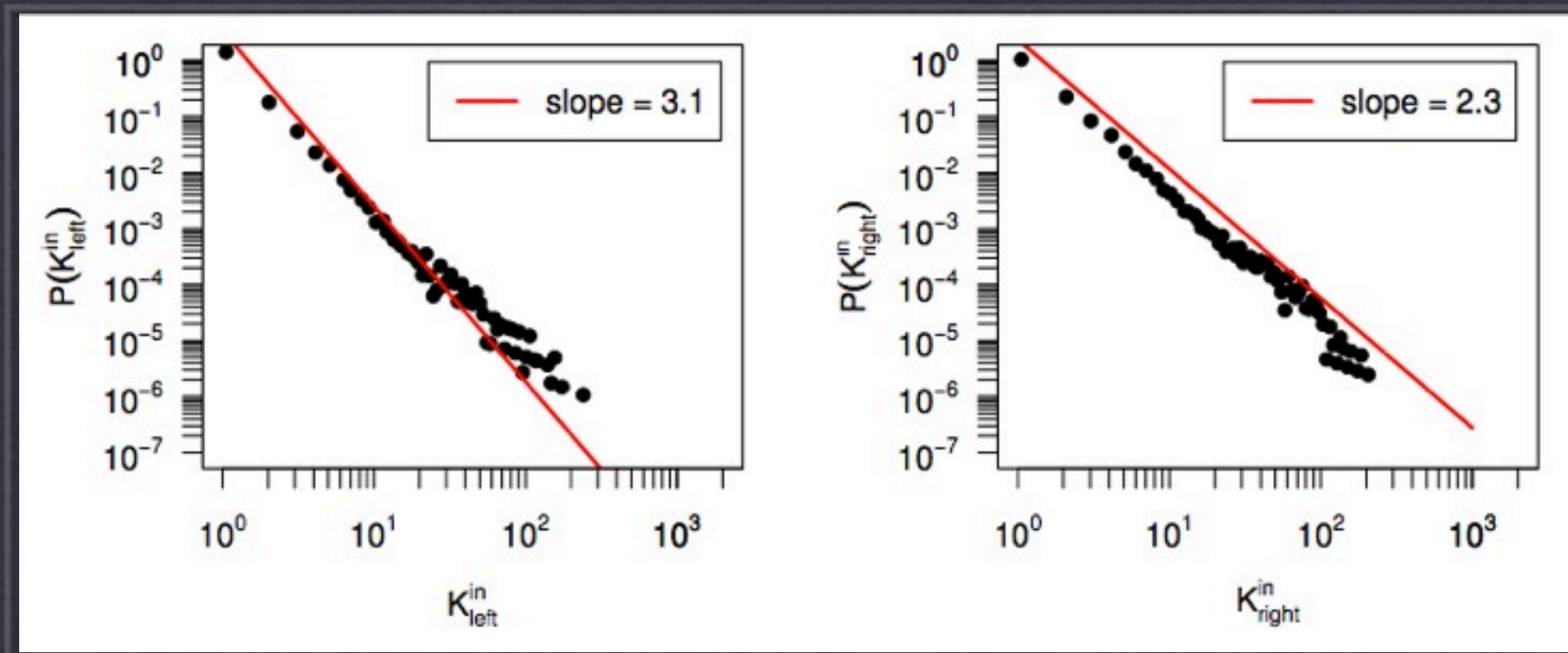
Information Diffusion

Follower Network

In-Degree　　　　　　　　　　**Out Degree**

Community	Nodes	Edges	Avg. Degree	Clust. Coeff.	Reciprocity
Left	9,941	803,329	80.80	.134	42.8%
Right	6,426	1,503,417	233.95	.221	64.8%

Follower Network

Community	Nodes	Edges	Avg. Degree	Clust. Coeff.	Reciprocity
Left	11,353	32,772	2.88	.032	13.5%
Right	7,115	39,713	5.58	.045	12.1%

Retweet Network
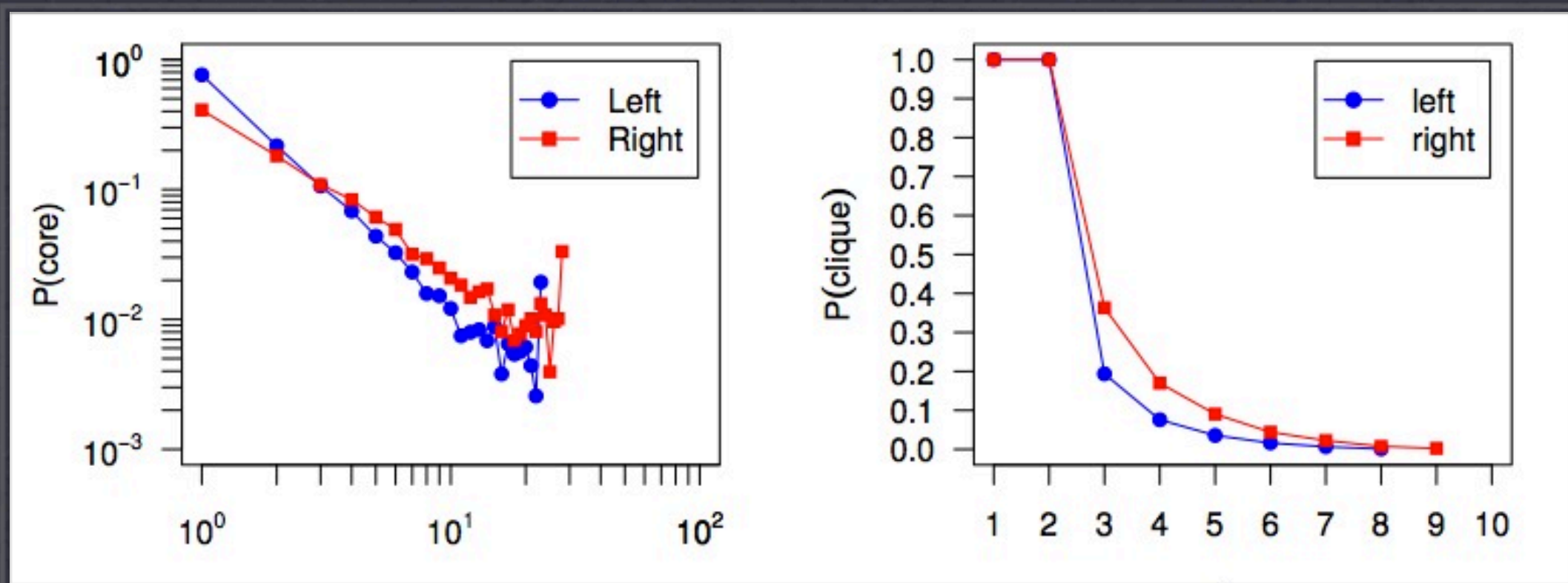
K-Core

K-Clique

Epidemic Spreading

Complex Contagion

Retweet Network

Future Directions

Modeling & Temporal Dynamics

Resource Repository
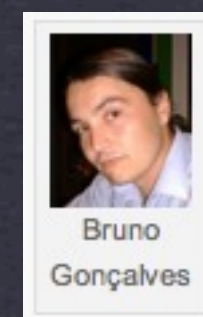
Forecasting & Prediction

Acknowledgements

Monday, September 26, 2011
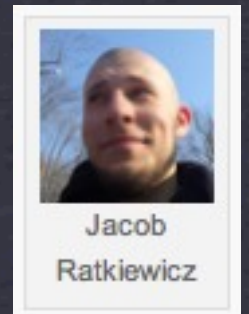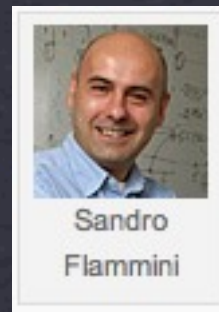
Social Media & The Networked Public Sphere

cnets.indiana.edu/groups/nan/truthy

Sandro Flammini

Fil Menczer

Jacob Ratkiewicz

Michael Conover

Bruno Gonçalves

Matt Francisco

midconov@indiana.edu

@vagabondjack

CNetS

CENTER FOR COMPLEX NETWORKS AND SYSTEMS RESEARCH
INDIANA UNIVERSITY SCHOOL OF INFORMATICS AND COMPUTING

Monday, September 26, 2011