The Party is Over Here: Structure and Content in the 2010 Election

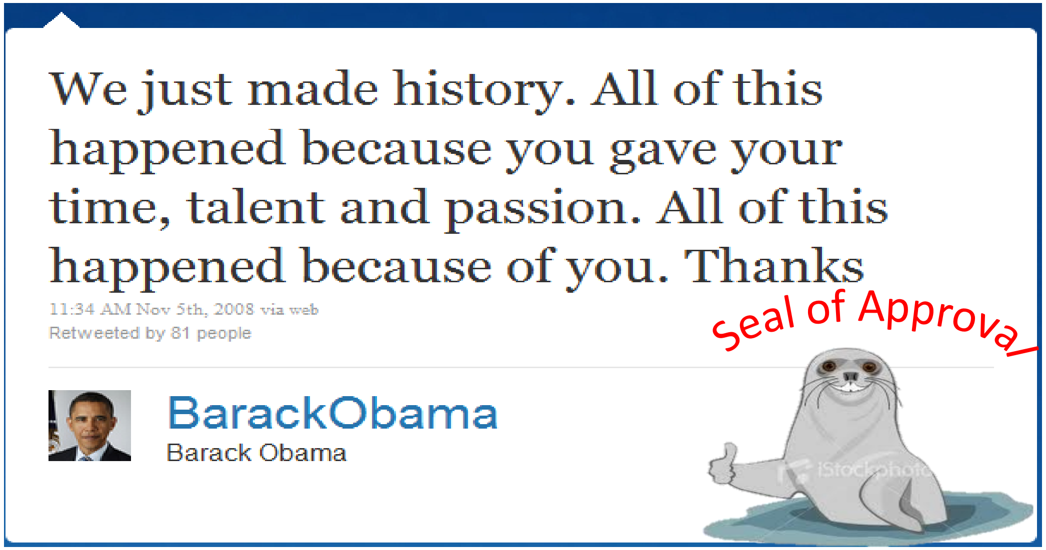
Avishay Livne, Matt Simmons, Eytan Adar, Lada Adamic

University of Michigan

ICWSM 2011

Twitter and Politics

- 14% of Internet users engaged in political activity via social media in 2008 => 22% in 2010 [Smith 2008, 2010].



Political Polarization on Twitter – [Conover et al. 2011]

Characterizing Debate Performance via Aggregated Twitter Sentiment – [Diakopoulos & Shamma 2010]

Twitter in Politics

Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment – [Tumasjan et al. 2010]

Hypothesis: Portion of tweets mentioning a party => portion of votes this party will get.

Achieved MAE of 1.65%, traditional polls achieved 0.8%-1.48%.

German federal elections 2009.

In the same time a German octopus was used to predict the World Cup ‘10 results...

… correctly predicting 8/8 matches

Paul the Octopus

<u>Our goal: Can we extract more information from Twitter?</u>

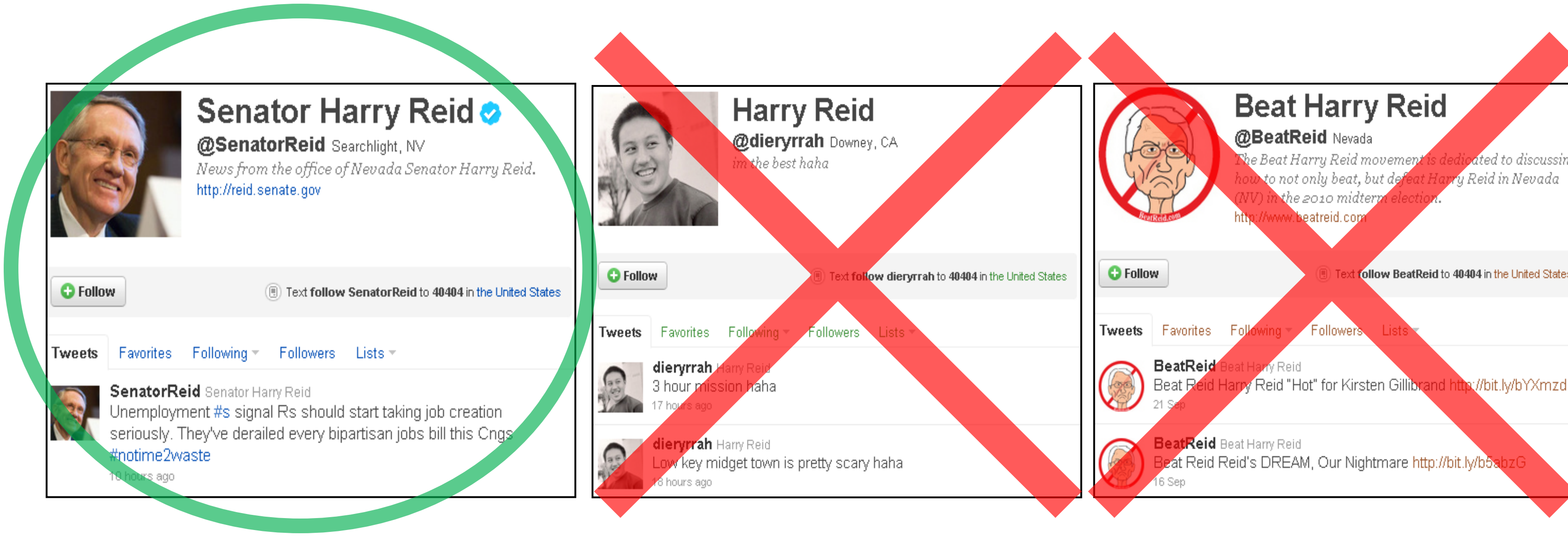
Agenda

- Background
- Data
- Network analysis
- Content analysis
- Predicting election results

Getting The Data

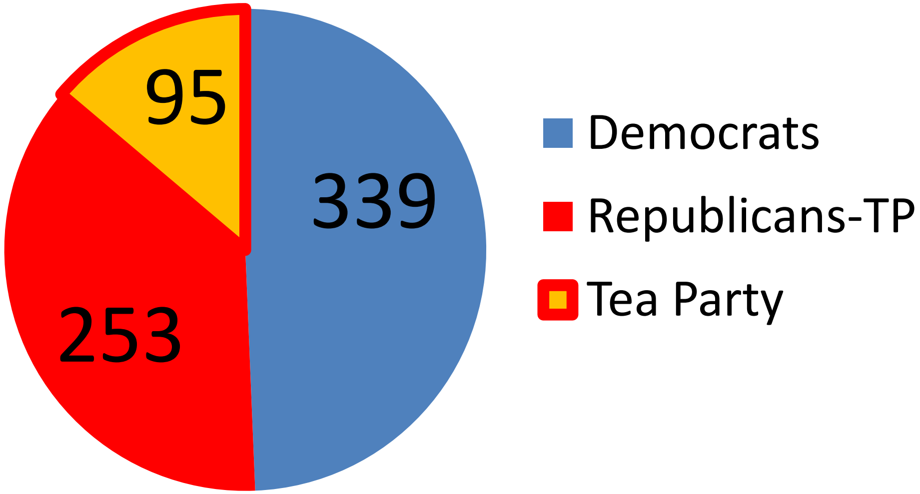
Unlike other studies we looked at **candidates'** accounts.

Search: "Harry Reid site:twitter.com"



Data

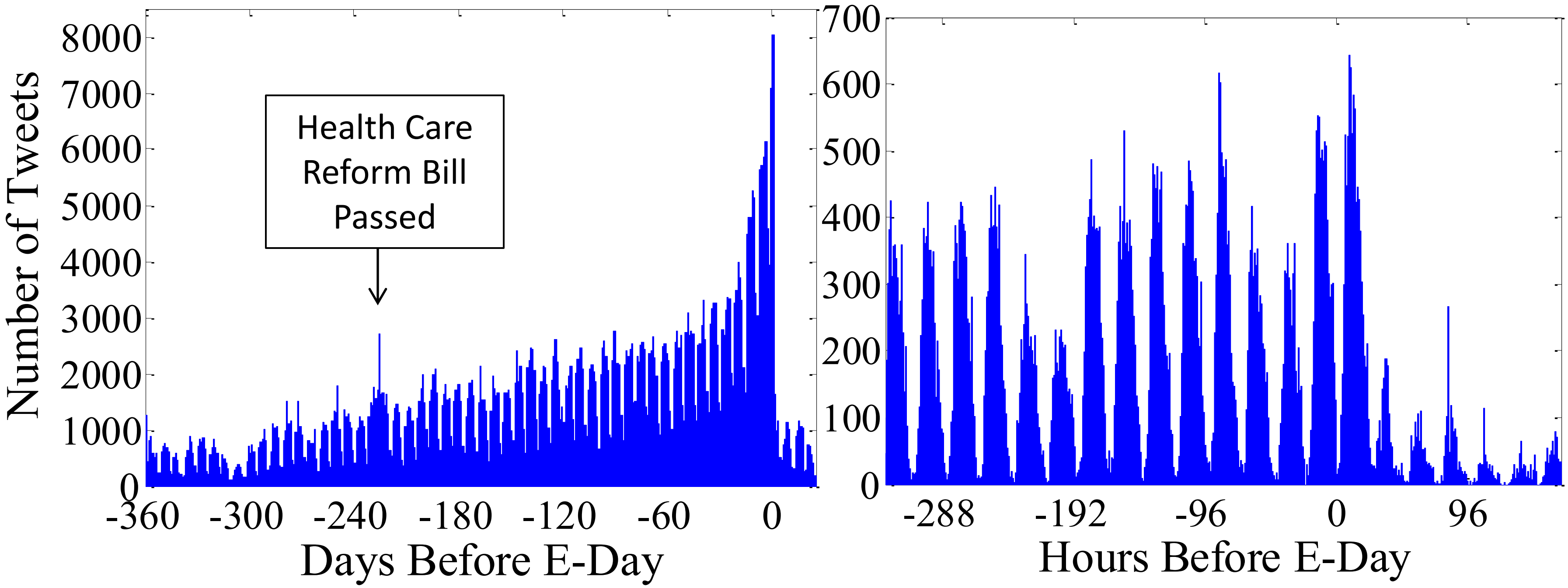
687 manually filtered users, 50% of the candidates.



*Tea Party – political movement, supported Republicans. Gained wide attention in 2010.

Data

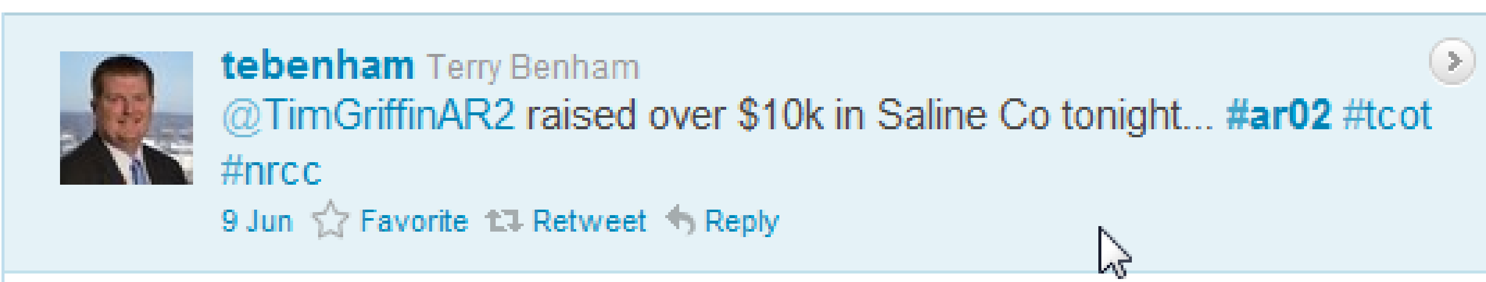
- 460K tweets over 2 years + 233K URLs.
- 4429 directed edges: A→B == user A follows user B



(a) (b)

Usage Analysis

Party (number of users)	Democrats (339)	Republicans-TP (253)	Tea Party (95)
tweets	551	723	901
tweets /day	2.66	2.97	5.21
retweets	40	52.3	82.6
@mentions	172.6	260.5	472.7
#hashtags	196	404	753
#/tweet	0.37	0.54	0.68

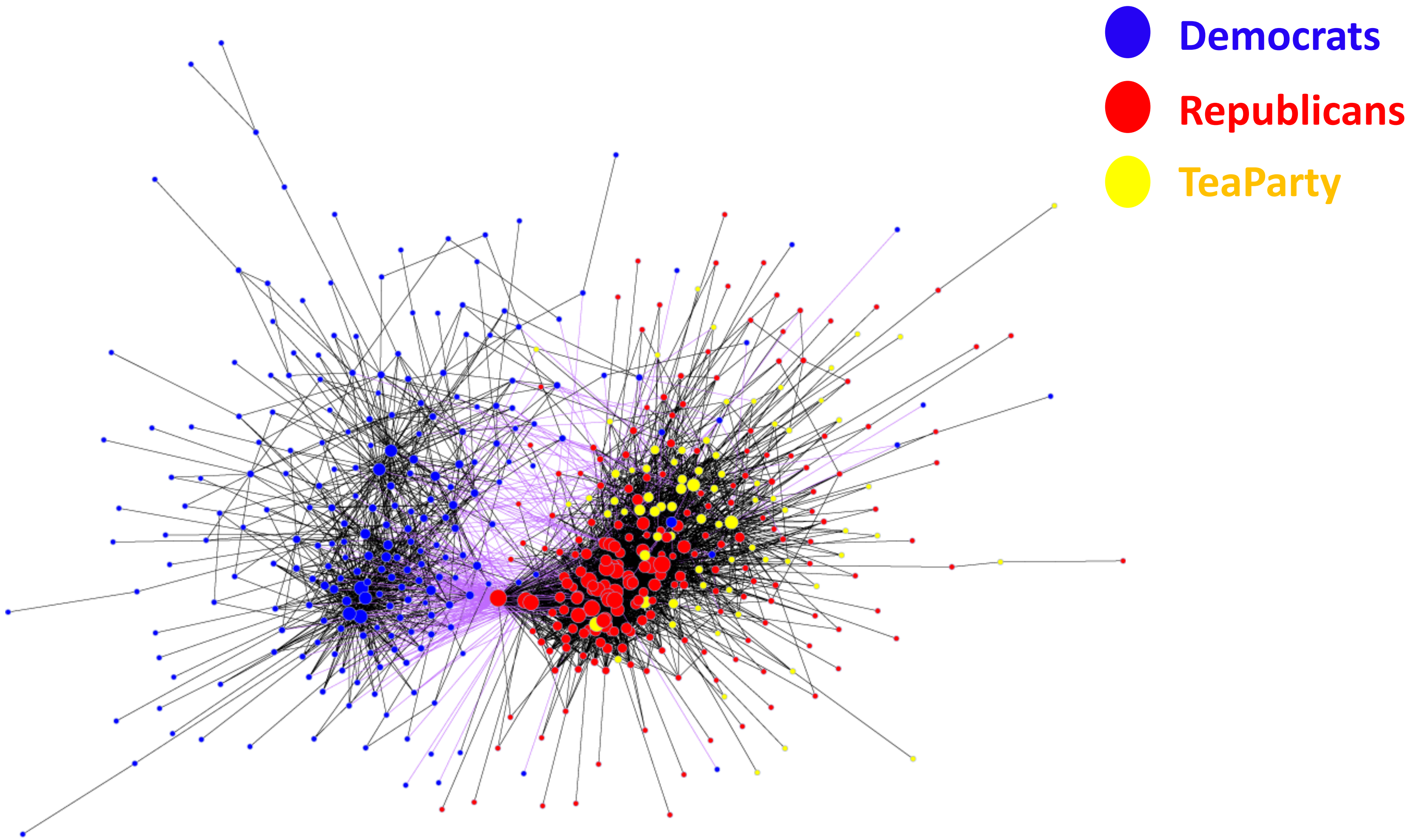


Agenda

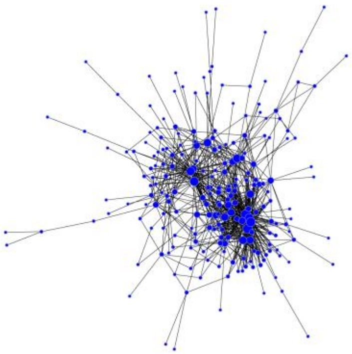
- Background
- Data
- Network analysis
- Content analysis
- Predicting election results

Graph Analysis

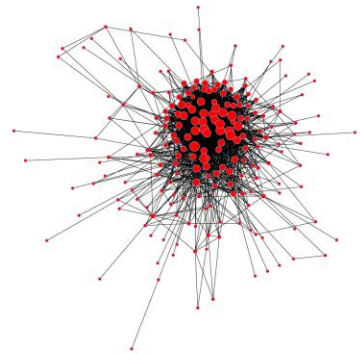
Edge(A,B) = A follows B



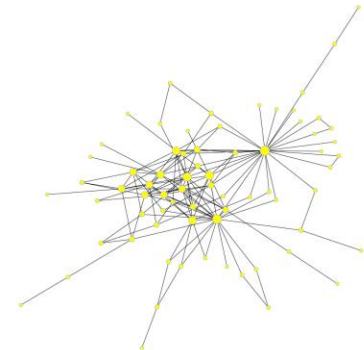
Graph Analysis



Democratic



Republican



TeaParty

$$Density = \frac{Edges}{Possible\ Edges} = \frac{E}{N(N-1)}$$

	Democrats	Republicans-TP	Tea Party	Republicans+TP
Density	0.007	0.032	0.02	0.025
Edges/Nodes	2.55	8.37	1.82	8.97

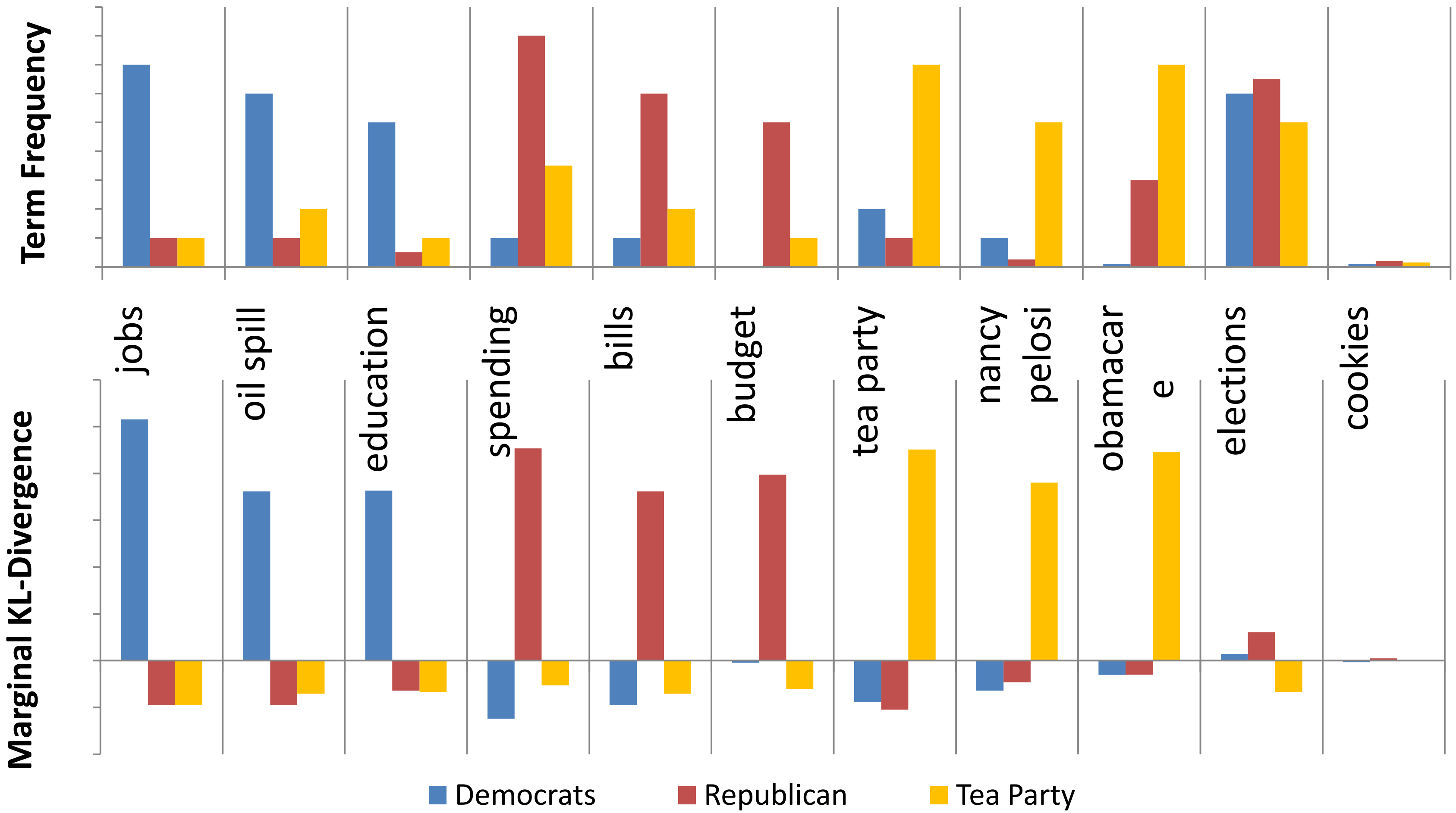
Supports previous studies [Adamic & Glance 2005]

Agenda

- Background & data
- Network analysis
- Content analysis
- Predicting election results

Language Modeling

How frequent each candidate/party used each term?



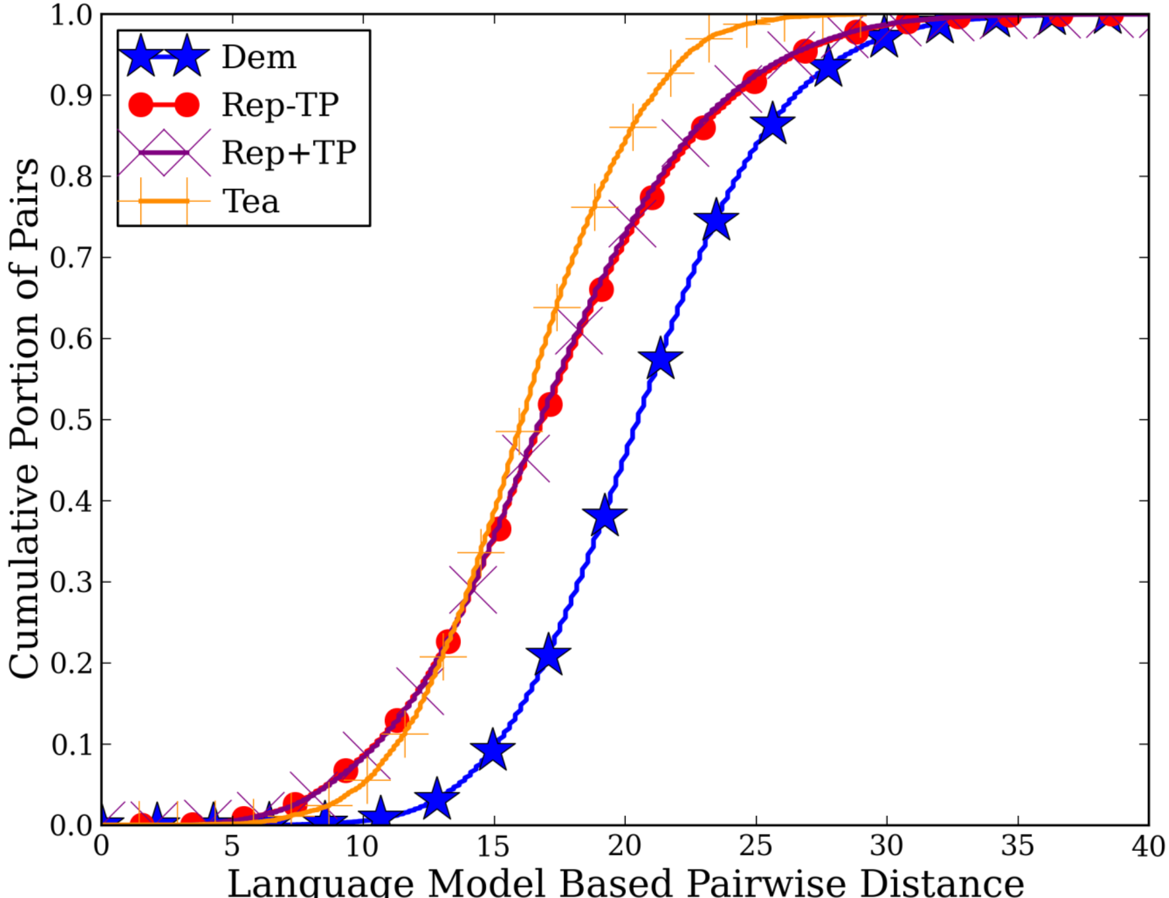
Extract distinguishing terms with KL-divergence

Divergent terms

Democrats	Republicans-TP	Tea Party
education	spending	barney_frank
jobs	bills	conservative
oil_spill	budget	tea_party
clean_energy	wsj (wall street journal)	clinton
afghanistan	bush	nancy_pelosi
reform	deficit	obamacare

Party Cohesiveness

How similar are pairs within each party



Similar ⟵⟶ Not similar

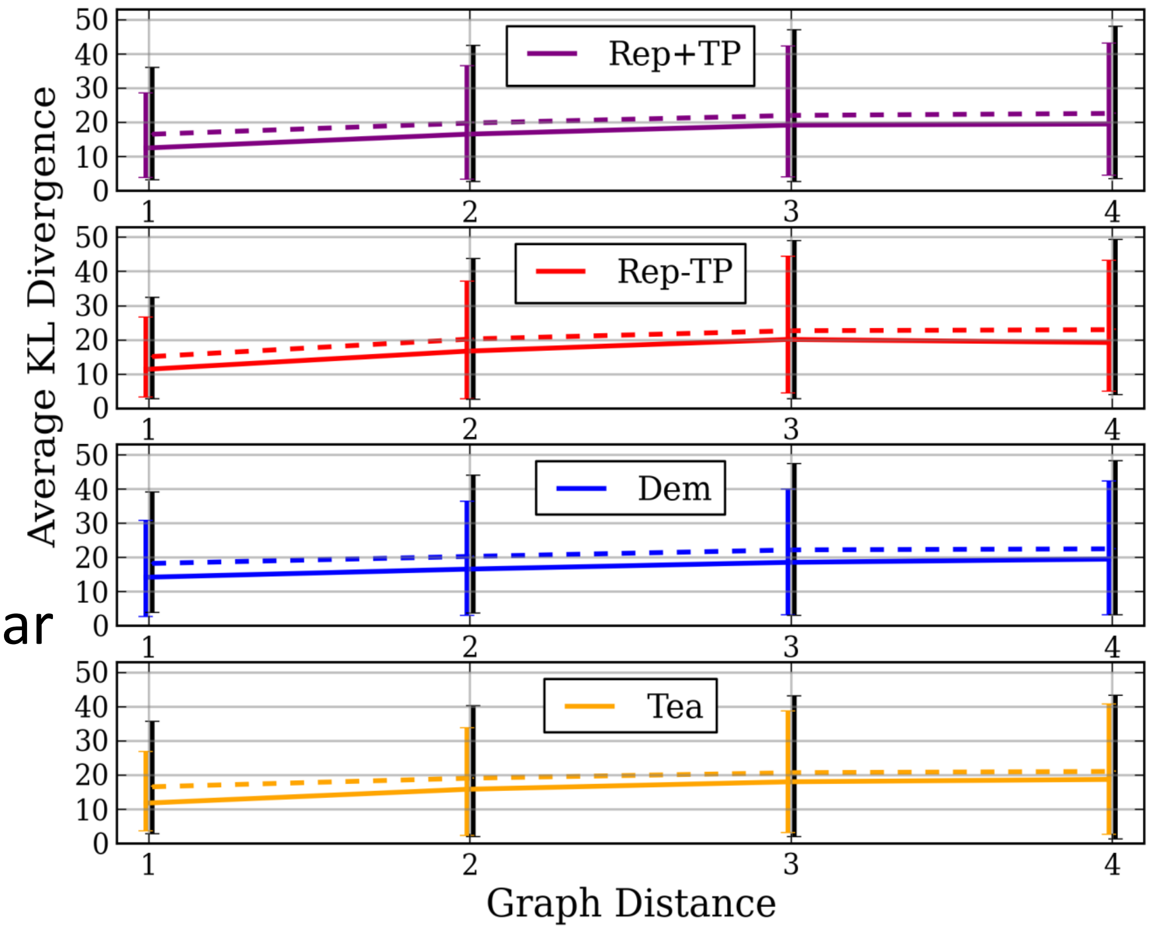
Language Similarity vs. Graph Distance

With retweets

Without retweets

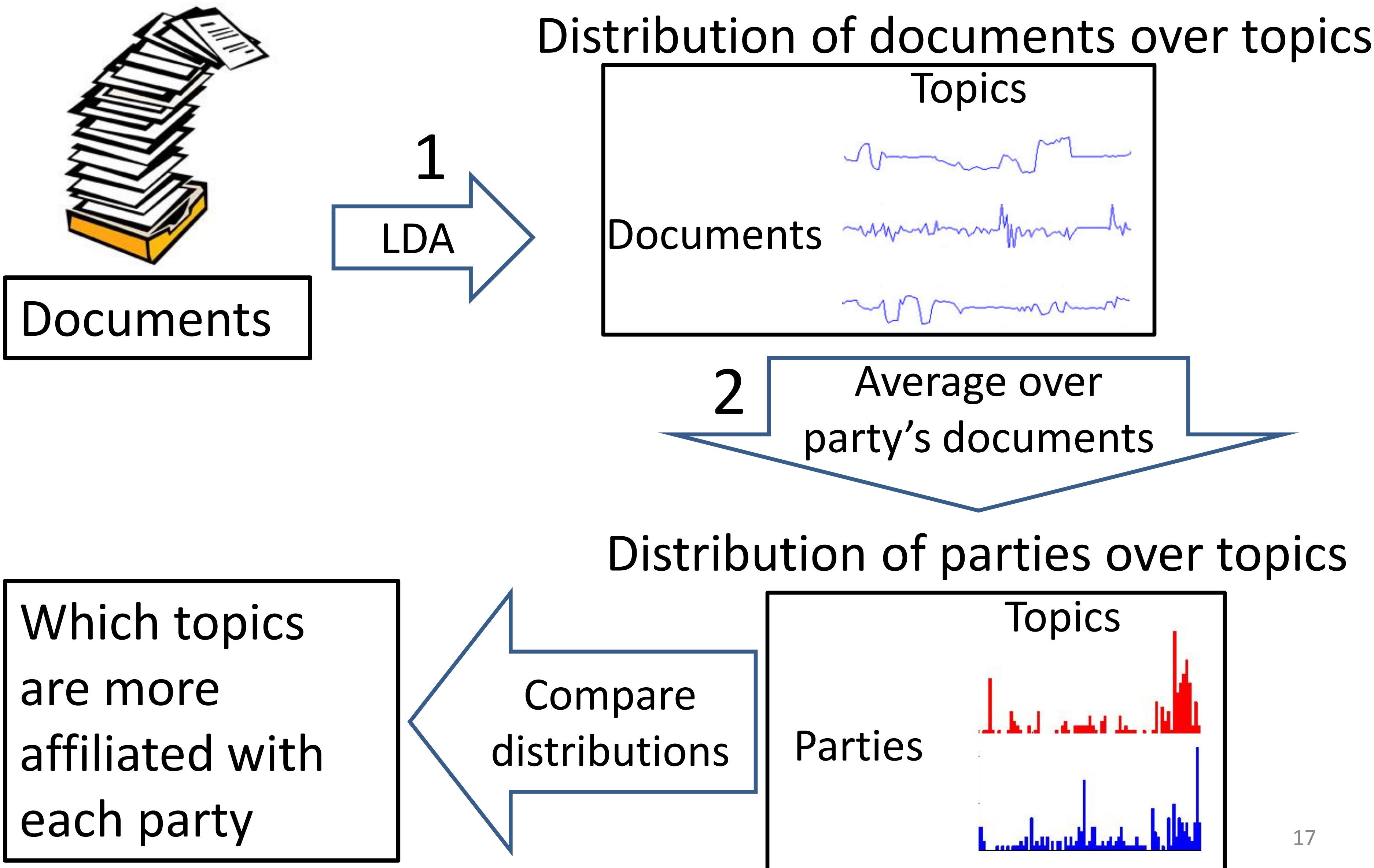
Not similar

Similar



- The **closer** two candidates are the more **similar** their content is.
- This trend diminishes after 3 steps.

Latent Topics



Topics

Topic terms	Affinity Difference
tax, jobs, spending, Obama, stimulus	-0.047618
health, care, bill, house, reform	-0.032136
tcot, barney, teaparty, Sean Bielat, twisters	-0.020878
live, show, interview, radio, fox	-0.018375
ff, great, followfriday, twitter, followers	-0.012113
obama, people, dont, good, government	-0.010277
great, county, meeting, day, tonight	-0.007769
campaign, tcot, twitter, facebook, support	-0.007624
john, david, ad, Pelosi, Sharron Angle	-0.002737
vote, endorsement, Harmer, ca10, candidate	-0.001998
change, view, changed, committee, energy	0.002625
great, day, parade, good, time	0.002746
ar02, ar2, Tim Griffin, vote, join	0.003104
Obama, oil, president, hearing, BP	0.007417
day, happy, great, women, honor	0.018366
bill, house, voted, senate, reform	0.028481
jobs, small, energy, great, business	0.074132

Agenda

- Background & data
- Network analysis
- Content analysis
- Predicting election results

Predicting Election Results - Individual Features

	Feature	Coefficient	P-value	Accuracy
External	Same party	2.67	<0.0001	78.9%
	Incumbent	3.16	<0.0001	76.9%
Usage	Retweets	-0.001	0.15	58.4%
	Hashtags	-0.0002	0.11	58.1%
	Tweets	-0.0002	0.08	57.8%
	Replies	-0.0003	0.08	57.5%
Network	Indegree	0.25	<0.0001	74.6%
	Closeness (all)	486.7	<0.0001	73.5%
	PageRank	486.7	<0.0001	66.4%
	Closeness (in)	1017.2	<0.0001	64.7%
	HITS Authority	0.442	<0.001	63.8%
Content	Corpus dissimilarity	-0.28	<0.0001	66.7%
	Party dissimilarity	-0.047	<0.05	55.9%

Logistic regression, 10-fold cross validation.

Predicting Election Results

Models Comparison

Name	Features	Accuracy
Baseline	Incumbent, Party, Same party	81.5%
Baseline + content	Tweets, Corpus dissimilarity, Incumbent, Party, Same party	83.8%
Baseline + network	Incumbent, Party, Same party, Closeness (all), Closeness (out)	84.0%
All but kl-corpus	Tweets, Incumbent, Party, Same party, Closeness (all), Closeness (out)	85.5%
All	Tweets, Corpus dissimilarity, Incumbent, Party, Same party, Closeness (all), Closeness (out)	88.0%

Next, top model...

Predicting Election Results

Top model's coefficients

Feature	Coefficient	Normalized	P-value
Intercept	-5.931	-3.641	<0.001
Tweets	-.000827	-3.328	<0.001
KL divergence from corpus	-.252	-3.232	<0.01
Incumbent	1.597	2.717	<0.01
Republican	1.374	2.899	<0.01
Tea Party	.605	1.085	0.28
Closeness (all)	23.820	5.443	<0.001
Closeness (out)	-76.750	-3.395	<0.001
Same party	1.931	3.888	<0.001

Summary

- Twitter is prevalent in campaigns
- Tea party more aggressive and Twitter-savvy
- Republicans more aligned (network)
- Democrats less cohesive (content)
- Content similarity vs. network distance
- Network and language improve prediction

Future Work

- More extensive topic & sentiment analysis
- Temporal analysis
- Integrating funding
- Integrating constituents (the crowd)

Thanks

Thanks to Abe Gong for helpful insights.

More information in the paper.

Contact info: avishay@umich.edu

Thank you!