Are Social Media more Social than Media? Measuring Ideological Homophily and Segregation on Twitter

Presented by Jiahao Lu (Kyle)

Introduction

- Social media represent a rapidly growing source of information for citizens around the world.
- Three key features of social media distinguish it from other forms of media and social interactions:

1. Social media allow users to not only consume information but also to produce information.
2. The information to which users are exposed depends upon self-chosen links among users.
3. Information on social media travels more rapidly and broadly than in other forms of social interactions.

Introduction

- On the one hand, social media may expose individuals to a more diverse set of viewpoints by allowing users to reach beyond their traditional geographic domains

- On the other hand, political content on social media may be highly polarized due to homophily, a tendency of users to develop links with like-minded users.

 › Motivated to compare the degree of homophily to that in offline social networks.

Introduction

- It is natural to conjecture that social media has displaced, at least partially, other sources of information for voters, most prominently traditional media

- face-to-face social interactions are characterized by higher ideological segregation when compared to traditional media settings

  - Motivated to compare the degree of ideological segregation on social networks to ideological segregation on both traditional media outlets and in face-to-face social interactions.

2 key findings

- They find that the network they constructed shares important features with face-to-face interactions.
- They find that Twitter is much more segregated than traditional media, and is more in-line with ideological segregation in face-to-face interactions
  - Taken together, the results suggest that social media may be a force for increasing isolation and ideological segregation in society.

Background
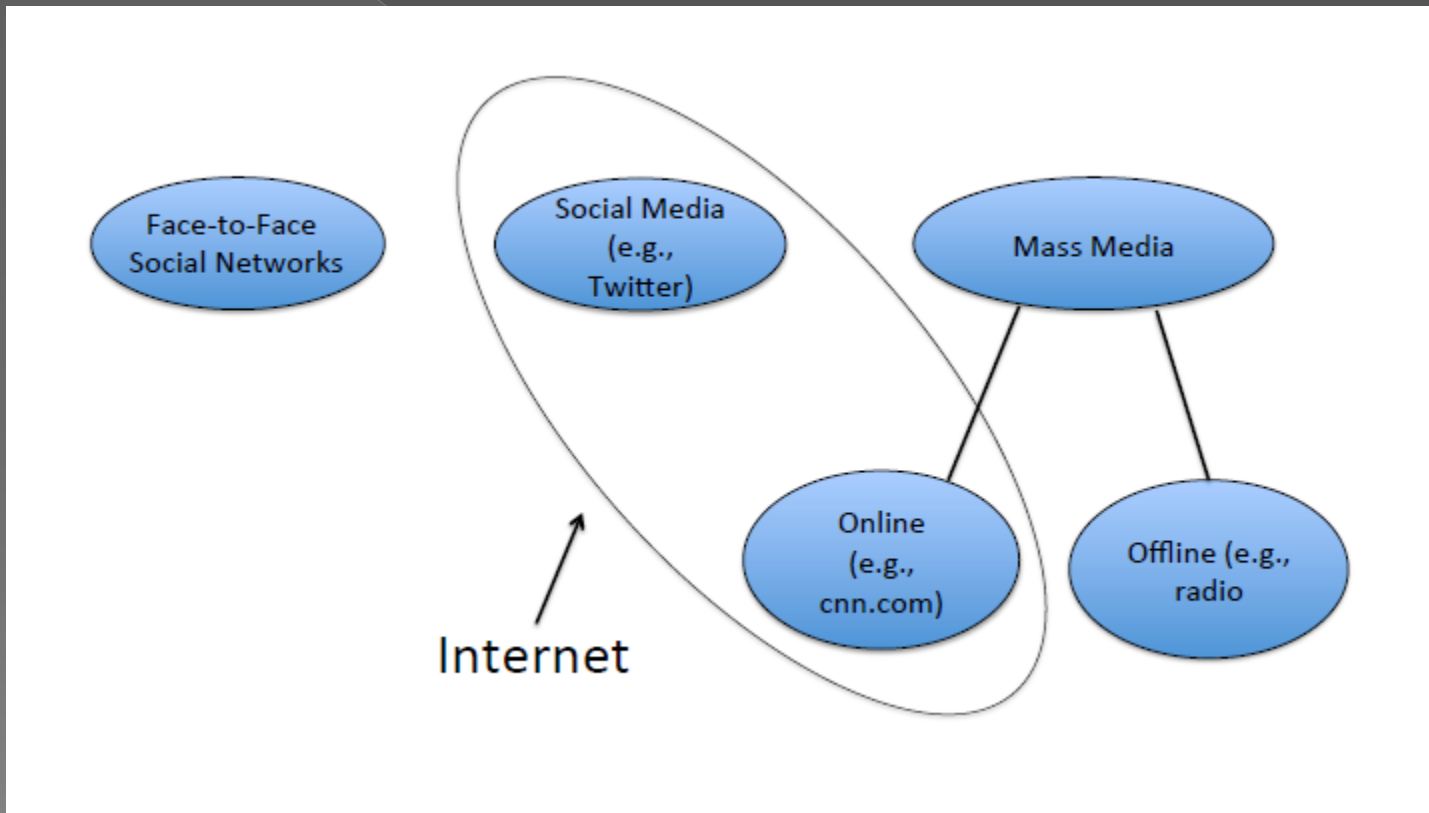
Social Media, politics and Ideology

- During the 2012 election campaign, 20 percent of American adults received at least some campaign information through Facebook and 5 percent did so through Twitter (Kohut et al., 012a). This represents a significant increase compared to the 2010 campaign when about 6 percent received information about the campaign through any social media and less than 1 percent reported doing so through Twitter (Smith, 2011).

- In 2008, 3 percent of American adults reported having used social media to post political news, a figure which reached 17 percent by 2012 (Smith, 2013). Also in 2012, 12 percent of American adults said they had used social media to "friend" a political candidate's account and 12 percent said they had started or joined a group on a social media website dedicated to a particular political or social cause, having both risen from an earlier level of 3 percent in 2008.

Twitter Interface and Experience

- Twitter is a unique social platform through which users connect with each other and gather information from a variety of sources.
  - Following: the connection is not necessarily mutual
  - Tweet: a message or update
  - Retweet: a republishing of another user's tweet
  - Home: a interface that displays the user's timeline or feed, where users are exposed to tweets from their entire set of followees.

Data: Construction of Network

Table 1: House Candidate Statistics

Twitter Usage and Party Affiliation of Candidate

	Total	Republicans		Democrats	
		N	Percent of total	N	Percent of total
Candidates with account	751	383	51%	368	49%
Candidates without account	74	31	42%	43	58%
Candidate accounts	976	537	55%	439	45%

Table 2: Candidate Account Statistics

Candidates with Twitter Accounts

	Median	Mean	Std. Dev.	Min	Max	N
Number of accounts	1	1.299601	0.5132769	1	4	751
Unique number of followers	1231	4982.264	24089.89	1	399223	707
Total number of tweets	462	850.6511	1328.663	0	12757	751
Year account was created	2010	2010.223	1.268873	2007	2012	686

Data: Voters
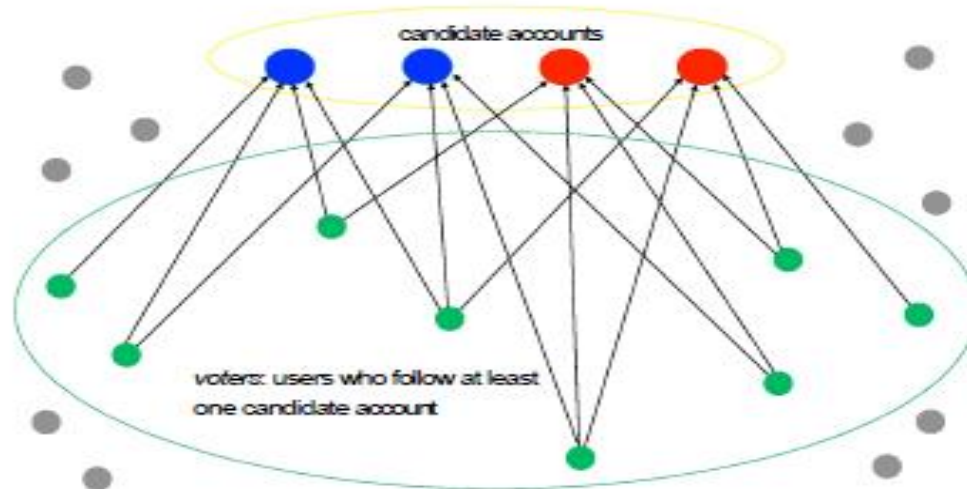
(a) Selecting sample of users (*voters*)
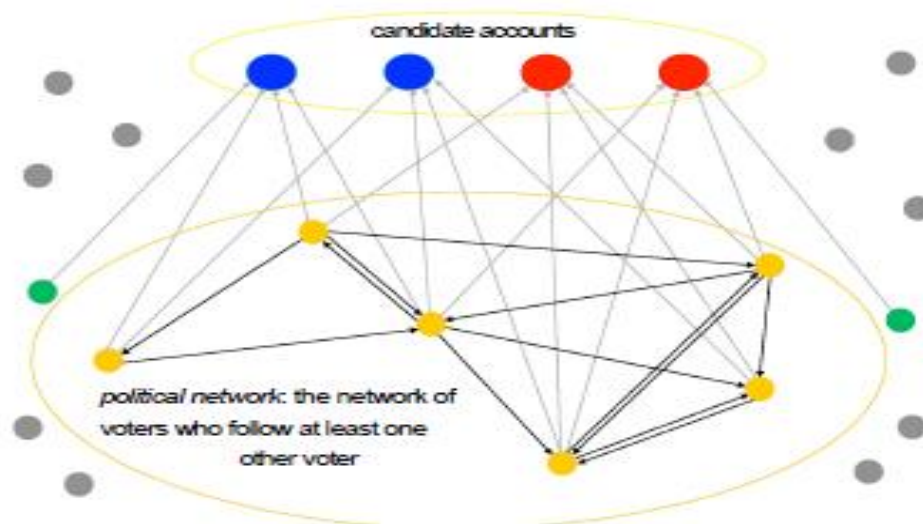
(b) Connecting selected users (*political network*)

Figure 2: Constructing the Network of Politically-Engaged Twitter Users

Data: Voters

Table 3: Voter Statistics

	Median	Mean	Std. Dev.	Min	Max	N
Total followers	48	330.5213	967.6255	0	92607	1547585
Voter followers	9	79.05342	269.9794	1	7683	1103832
Total followed	189	512.7784	862.027	0	74283	1547585
Voters followed	14	74.07017	211.1889	1	10029	1178095
Tweets	90	1710.496	7793.589	0	1300966	1547585
Year account was created	2010	2010.34	1.28955	2006	2012	1547585
Year voter was last mentioned	2013	2012.473	0.9154292	2007	2013	1363228

Voter Geography

- One aspect of the network that we are particularity interested in examining is the degree of spatial homophily among voters.
  - Segregation in social media may be fully determined by geographic segregation
  - If ideological segregation is driven by geography, then segregation should fall when examining links within a geographic area.
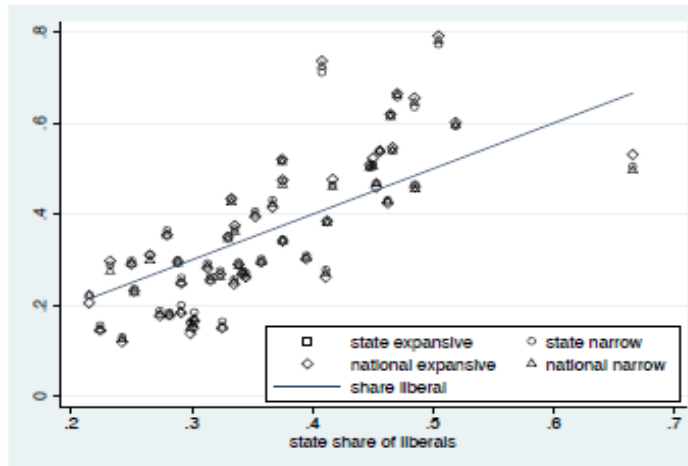
Voter Ideology

- characterize voters as liberal, conservative, or moderate based upon the party affiliation of the candidates that they chose to follow.
  - › Voters that exclusively followed Democratic candidates are coded as liberals
  - › Voters that exclusively followed Republican candidates are coded as conservatives
  - › For voters who follow a mix of Democratic and Republican candidates, we have 2 separate methods for categorizing these voters:
    - Broad measure of moderate: any voter who follow at least one Republican and one Democratic candidate is coded as moderate
    - Narrow measure of moderate: only those who follow an equal number of Democratic and Republican candidates are regarded as moderate
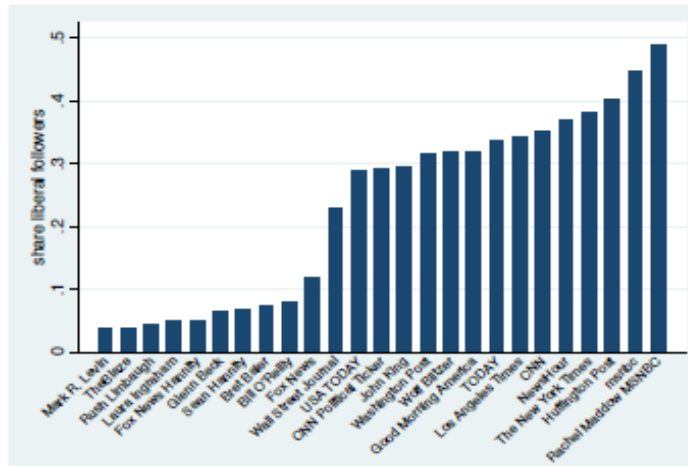
Voter Ideology

Table 6: Distribution of Voter Ideology

Ideology measure	User Supplied State		
	Liberal	Conservative	Moderate
State expansive	36%	59%	5%
State narrow	37%	60%	2%
National expansive	32%	53%	14%
National narrow	36%	59%	5%

Voter Ideology

(a) Share of State Liberal Voters and Liberal Twitter Users

(b) Share of Media Outlet Liberal Followers

Figure 5: Validation of Ideology Measure for Twitter Users

Empirical Framework

- Measures of Homophily in Social Networks

For measures of homophily, we follow Currarini et al. (2009). In our setting, there are three different voter types, liberals, moderates and conservatives, denoted by $t \in \{l,m,c\}$. Let I be the total number of voters who follow at least one voter and I_t be the total number of type t voter followers, such that $I = I_l + I_m + I_c$. Then $w_t = \frac{I_t}{I}$ is the fraction of type t in the voter population. Similarly, let J be the total number of voters who are followed by at least one voter and J_t be the total number of type t voters being followed, such that $J = J_l + J_m + J_c$. Let v_{it} denote the number of type t followees of voter i. Then $s_t = \frac{1}{I_t}\sum_{i \in I_t} v_{it}$ denotes the average number of type t voters followed by type t voters (same) and $d_t = \frac{1}{I_t}\sum_{i \in I_t} v_{i-t}$ denote the average number of non-type t voters followed by type t voters (different). With these in hand, we define three related measures of voter type homophily.

Measures of Homophily in Social Network

DEFINITION 1: The homophily index for type t voters is as follows:

$$H_t = \frac{s_t}{s_t + d_t}.$$

DEFINITION 2: The profile $(s_l, d_l, s_m, d_m, s_c, d_c)$ satisfies *relative homophily* if $w_t > w_{t'}$ implies $H_t > H_{t'}$.

DEFINITION 3: The profile $(s_l, d_l, s_m, d_m, s_c, d_c)$ satisfies:

1. *baseline homophily* if for all type t,

$$H_t = w_t.$$

2. *inbreeding homphily* for type t if

$$H_t > w_t.$$

3. *heterophily* for type t if

$$H_t < w_t.$$

DEFINITION 4: The *inbreeding homophily* of type t is

$$IH_t = \frac{H_t - w_t}{1 - w_t}.$$

Measuring Ideological segregation

For comparison with other measures of ideological isolation, we can compute the isolation index following Gentzkow and Shapiro (2011). In particular, for each voter j, let v_{jc} denote the number of conservative followers and v_{jl} the number of conservative followers. We can then define the *share conservative* of voter j as the fraction of his or her partisan (i.e., not moderate) followers who are conservative:

$$share\ conservative_j = \frac{v_{jc}}{v_{jl}+v_{jc}}.$$

Finally, group t's conservative exposure is defined as follows:

$$conservative\ exposure_t = \frac{1}{\sum_{j\in J}v_{jt}}\sum_{j\in J}v_{jt}\cdot share\ conservative_j,$$

where $\sum_{j\in J}v_{jt}$ is the sum over the total number of type t followers for each voter followee j.[10] The term above is then the followers-weighted (v_{jt}) average exposure of type t voters.[11] With this in hand, the isolation index is then:
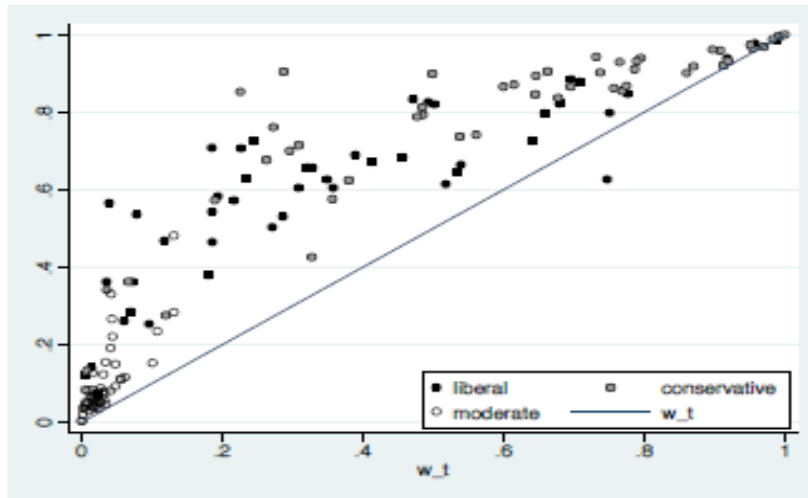
$$isolation = conservative\ exposure_c - conservative\ exposure_l.$$

Ideological Homophily in Social Media

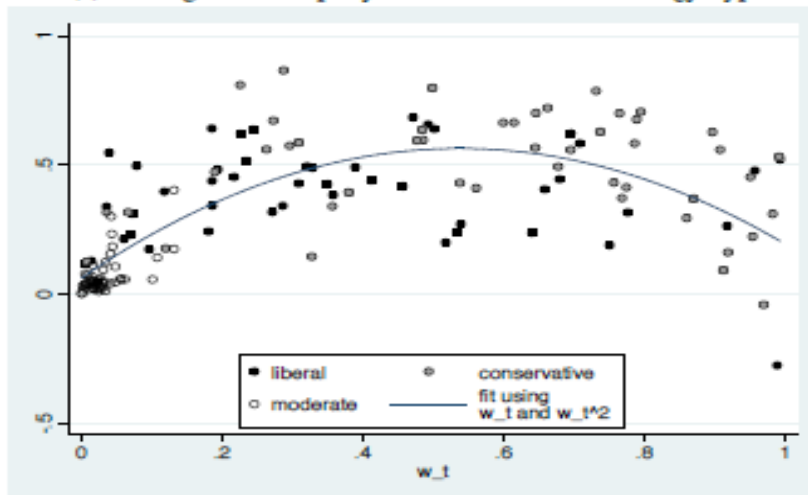## Table 8: National Ideological Homophily

	Population share	Same-type voters followed	Per capita voters followed	H index	IH index
Liberal	0.412	40.885	60.983	0.670	0.440
Conservative	0.543	67.857	83.776	0.810	0.584
Moderate	0.045	6.746	76.667	0.088	0.045

Ideological Homophily in Social Media

(a) Ideological Homophily and State Share of Ideology Type

(b) Ideological Inbreeding Homophily and State Share of Ideology Type

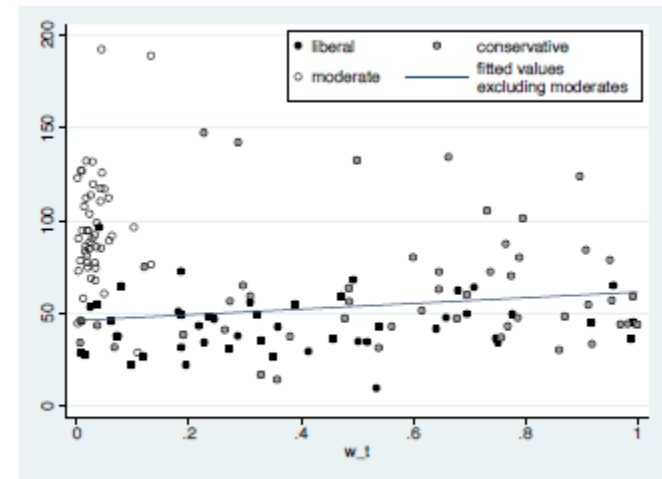Figure 6: Group Size and within Group Ideological Homophily

Figure 7: Group Size and Per Capita Twitter Voters Followed

Segregation in Social Media

Table 9: Ideological Segregation

	Conservative exposure of		
	Conservatives	Liberals	Isolation Index
Social Media			
National	0.795	0.306	0.489
State	0.792	0.376	0.416
Face-to-face interactions			
Political discussants	0.796	0.402	0.394
People you trust	0.675	0.372	0.303
Family	0.69	0.447	0.243
Neighborhood	0.627	0.439	0.187
Work	0.596	0.428	0.168
Voluntary associations	0.625	0.48	0.145
County	0.682	0.622	0.059
ZIP code	0.637	0.543	0.094
Media			
National newspapers	0.612	0.508	0.104
Internet	0.606	0.531	0.075
Local newspapers	0.695	0.647	0.048
Magazines	0.587	0.54	0.047
Cable	0.712	0.679	0.033
Broadcast news	0.677	0.66	0.018

Note: Source for data on face-to-face interactions and media is Gentzkow and Shapiro (2011)

Conclusion

- Our results suggest that social media are highly segregated along ideological lines and thus emphasize these potential problems associated with the flow of information in segregated networks.

- Using information on links between followers of accounts associated with candidates in the 2012 election, we find that Twitter is segregated along ideological lines. In particular, followers of Republican candidates are much more likely to be linked to other followers of Republican candidates and likewise for followers of Democratic candidates. This Twitter political network is at least as segregated along ideological lines as face-to-face interactions and is much more segregated than traditional media outlets, such as newspapers and television. Taken together, these results suggest that social media may be a force for increasing ideological segregation.