A Measure of Polarization on Social Media Networks Based on Community Boundaries

Pedro Calais Guerra, **Wagner Meira Jr.**

Federal University of Minas Gerais (UFMG), Brazil

Claire Cardie, Robert Kleinberg

Cornell University

July 10, 2013

Outline

Polarization: a Social Science Concept

Social process whereby a social group is divided into two opposing sub-groups having **conflicting** positions and viewpoints

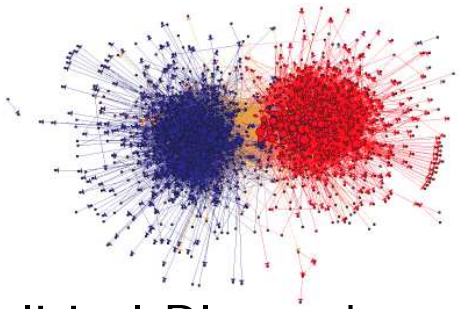


Figure: 2004 U.S. Political Blogosphere [Adamic et al. 2005]

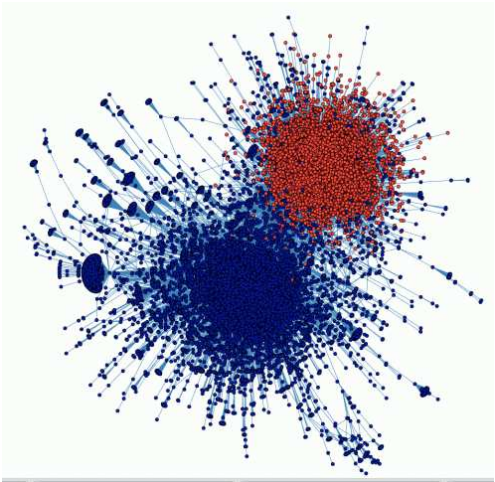
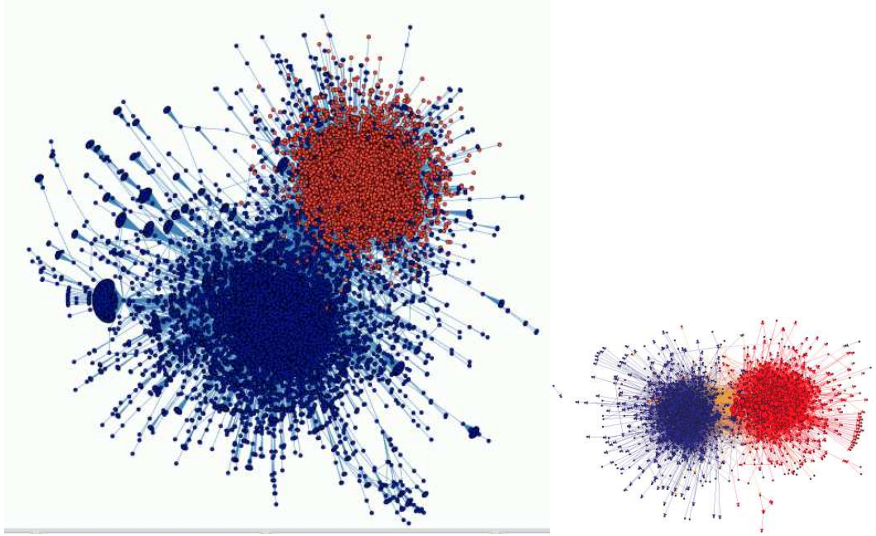


Figure: 2010 U.S. Political Twitter Network [Conover et al. 2010]

Polarization: a Social Science Concept

- Politics – parties
- Public Policies – same-sex marriage, abortion, gun control
- Sports – supporters of rival clubs



Key aspects:

- **antagonism**
- **extreme, biased opinions**

Why study polarization in social media?

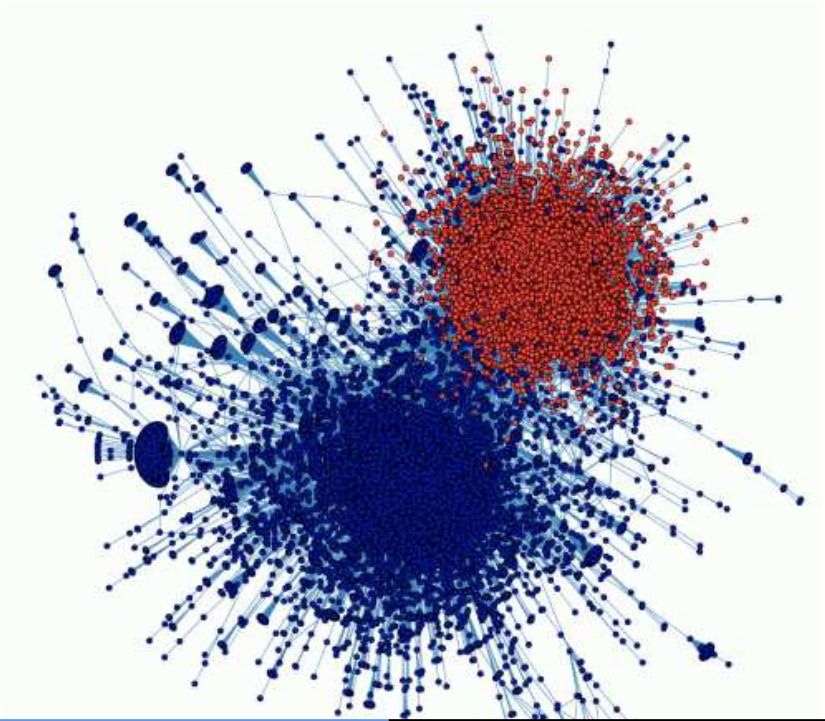
- **1.** Polarization causes **segregation** and **conflict** in the society [Paul DiMaggio 1996; Mouw and Sobel 2001].
- **2.** Polarization can be a key information for **opinion analysis** [Calais et al 2011, Tan et al 2011]
- **3.** Polarization $\rightarrow$ strong **bias** on opinions [Walton 1991]

Outline

How is polarization measured on OSNs?

Polarization $\rightarrow$ Cohesive and dense subgroups [Conover et al. 2011, Adamic et. al. 2005, Scott et. al, 2009, Zhang et. al 2010]

- clusters having many internal connections among nodes
- few connections to the other group



How is polarization measured on OSNs?

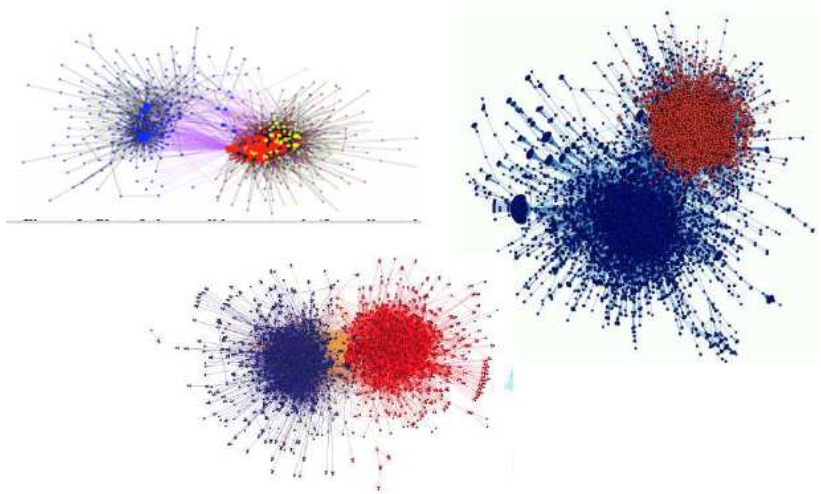
- **Modularity** [Newman 2006]
  - Measures the **strength** of division of a network into modules
  - Compares the number of edges inside a cluster with the expected on a **random graph**

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \frac{s_i s_j + 1}{2} \quad (1)$$

- range: $[-1/2, 1)$

How is polarization measured?

- high modularity $\rightarrow$ polarization [Conover et al. 2011, Adamic et. al. 2005, Scott et. al, 2009, Zhang et. al 2010]



Modularity seems to make sense...

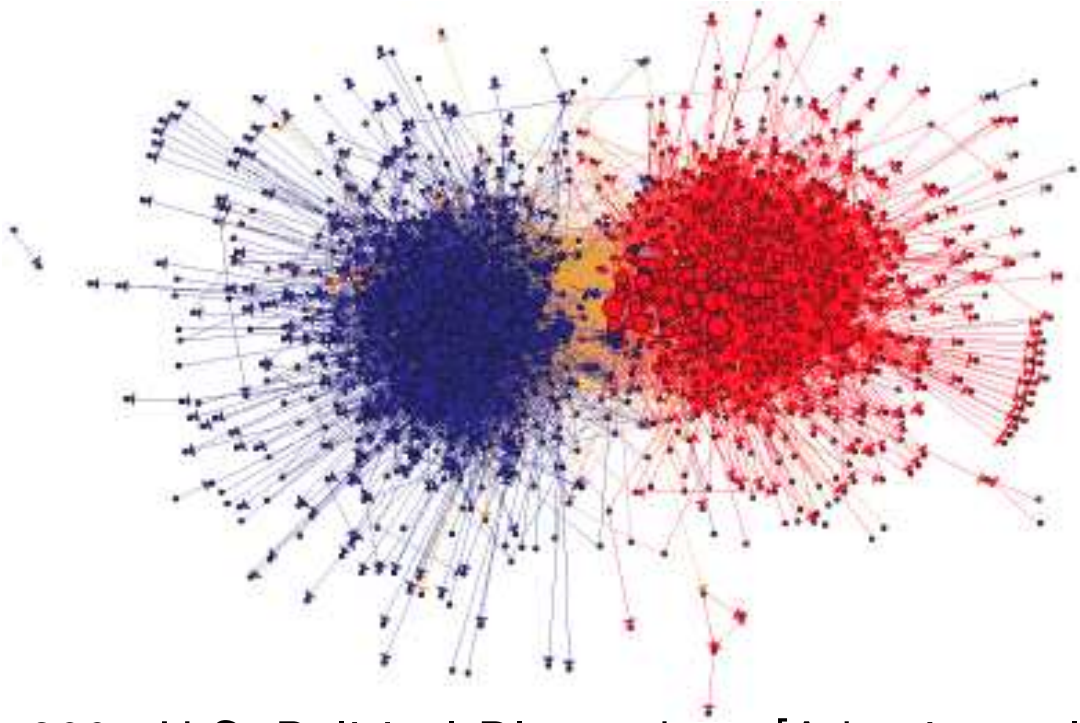


Figure: 2004 U.S. Political Blogosphere [Adamic et al. 2005]

Q = 0.42 (high!)

... but what about non-polarized social networks?

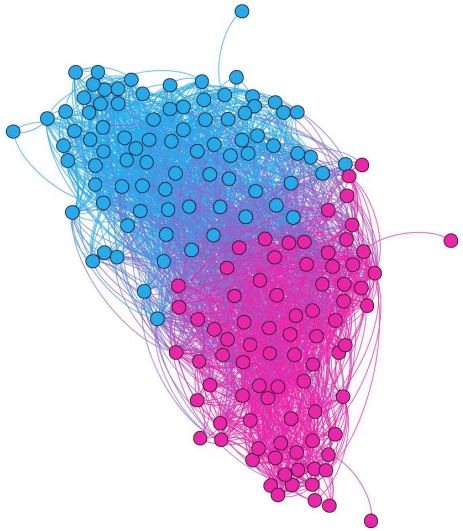


Figure: Facebook Friends (grad and undergrad communities). No polarization at all!

$$\mathbf{Q = 0.24}$$

Modularity and Polarization

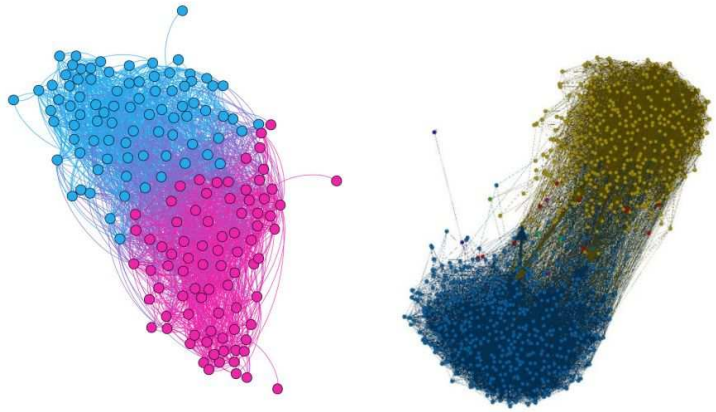


Figure: Q=0.24 and Q=0.42

- How much is "high" modularity?

Modularity and Polarization

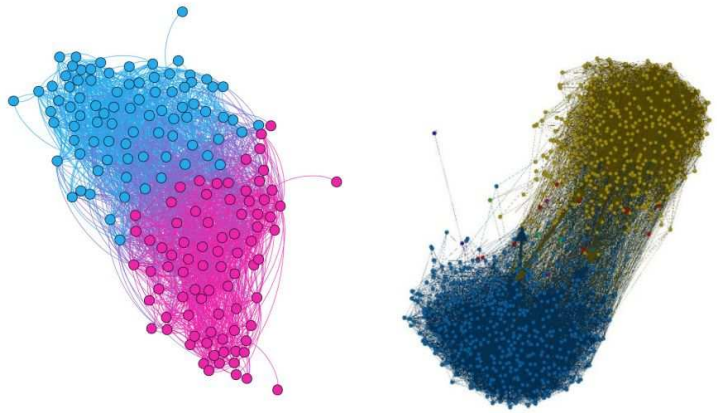


Figure: Q=0.24 and Q=0.42

- How much is "high" modularity?
- Modularity is not a *direct* measure of Polarization.

Modularity and Polarization

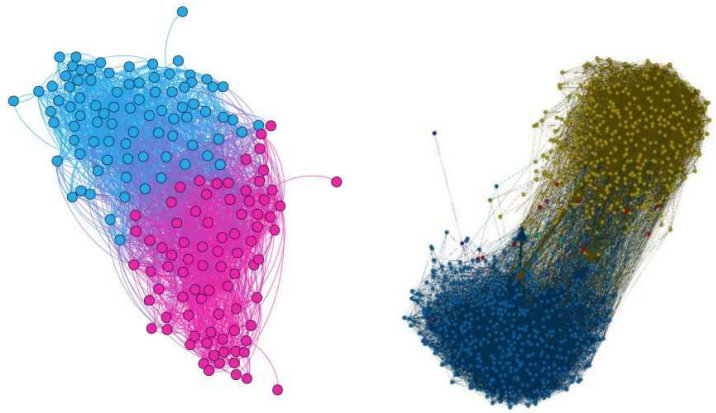


Figure: Q=0.24 and Q=0.42

- How much is "high" modularity?
- Modularity is not a *direct* measure of Polarization.
- We want a **negative** measurement when no polarization is present.

Modularity and Polarization

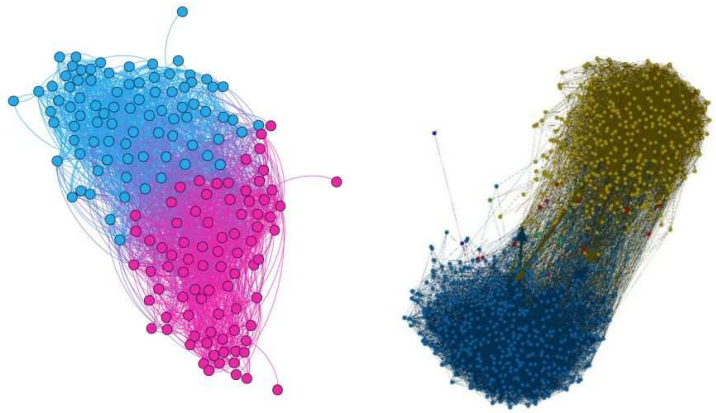
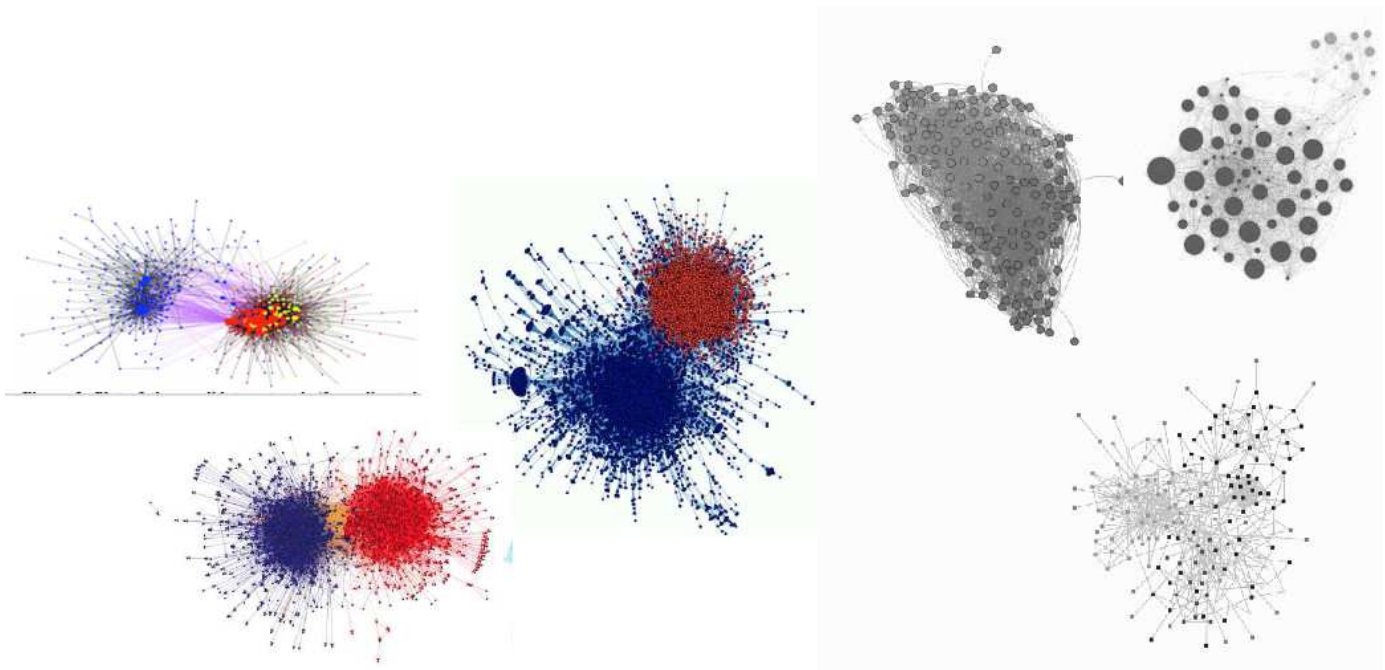


Figure: Q=0.24 and Q=0.42

- How much is "high" modularity?
- Modularity is not a *direct* measure of Polarization.
- We want a **negative** measurement when no polarization is present.
- We seek a structural pattern that highlights *antagonism*.

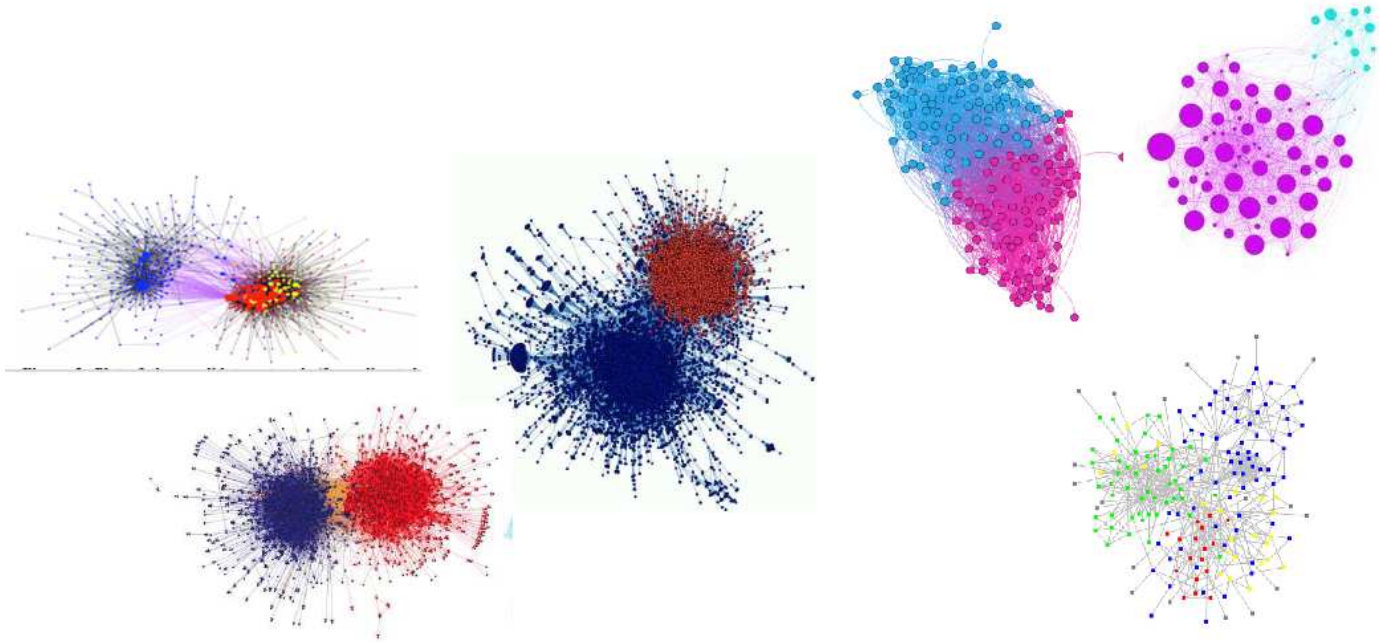
The bias in the literature

Polarized networks vs. Non-polarized networks



The bias in the literature

Polarized networks vs. Non-polarized networks



Outline

Community Boundary

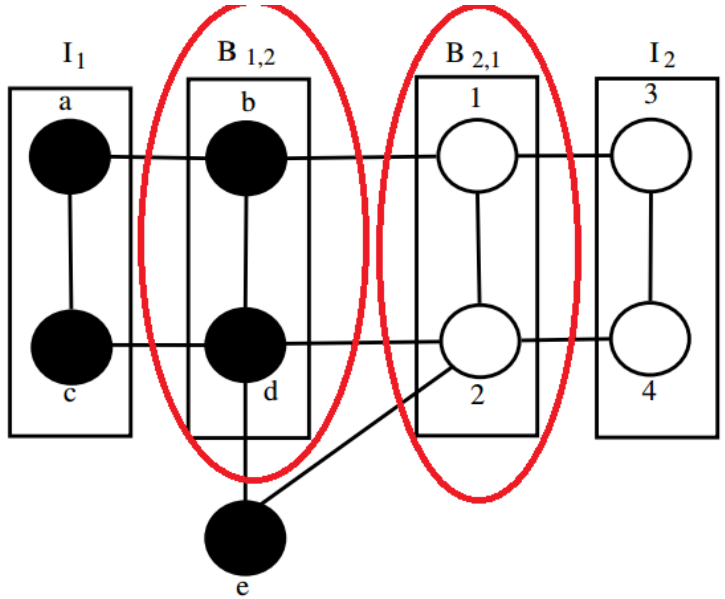
community **boundary**: set of all nodes v that

1. have at least one edge that connecting to the other community **($d_b(v)$)**;

Community Boundary

community **boundary**: set of all nodes v that

1. have at least one edge that connecting to the other community **($d_b(v)$)**;
2. have at least one edge connecting to a member of its community which does not link to the other community **($d_i(v)$)**.



$$P_v = \frac{d_i(v)}{d_b(v) + d_i(v)} - 0.5 \tag{2}$$

$$P_v = \frac{d_i(v)}{d_b(v) + d_i(v)} - 0.5 \tag{2}$$

- $P_v > 0 \rightarrow v$ prefers internal connections **(antagonism?)**

$$P_v = \frac{d_i(v)}{d_b(v) + d_i(v)} - 0.5 \quad (2)$$

- $P_v > 0 \rightarrow v$ prefers internal connections **(antagonism?)**
- $P_v < 0 \rightarrow v$ prefers connections with members of the other group **(increased homophily!)**

$$P_v = \frac{d_i(v)}{d_b(v) + d_i(v)} - 0.5 \tag{2}$$

- $P_v > 0 \rightarrow v$ prefers internal connections **(antagonism?)**
- $P_v < 0 \rightarrow v$ prefers connections with members of the other group **(increased homophily!)**

$$P = \frac{1}{|B|} \sum_{v \in B} P_v \tag{3}$$

Comparing Q and P

- **Non-polarized networks:**
- Facebook friends: Q = +0.24, P= -0.24
- NY Giants and NY Knicks (Twitter): Q = +0.15, P= -0.002

Comparing Q and P

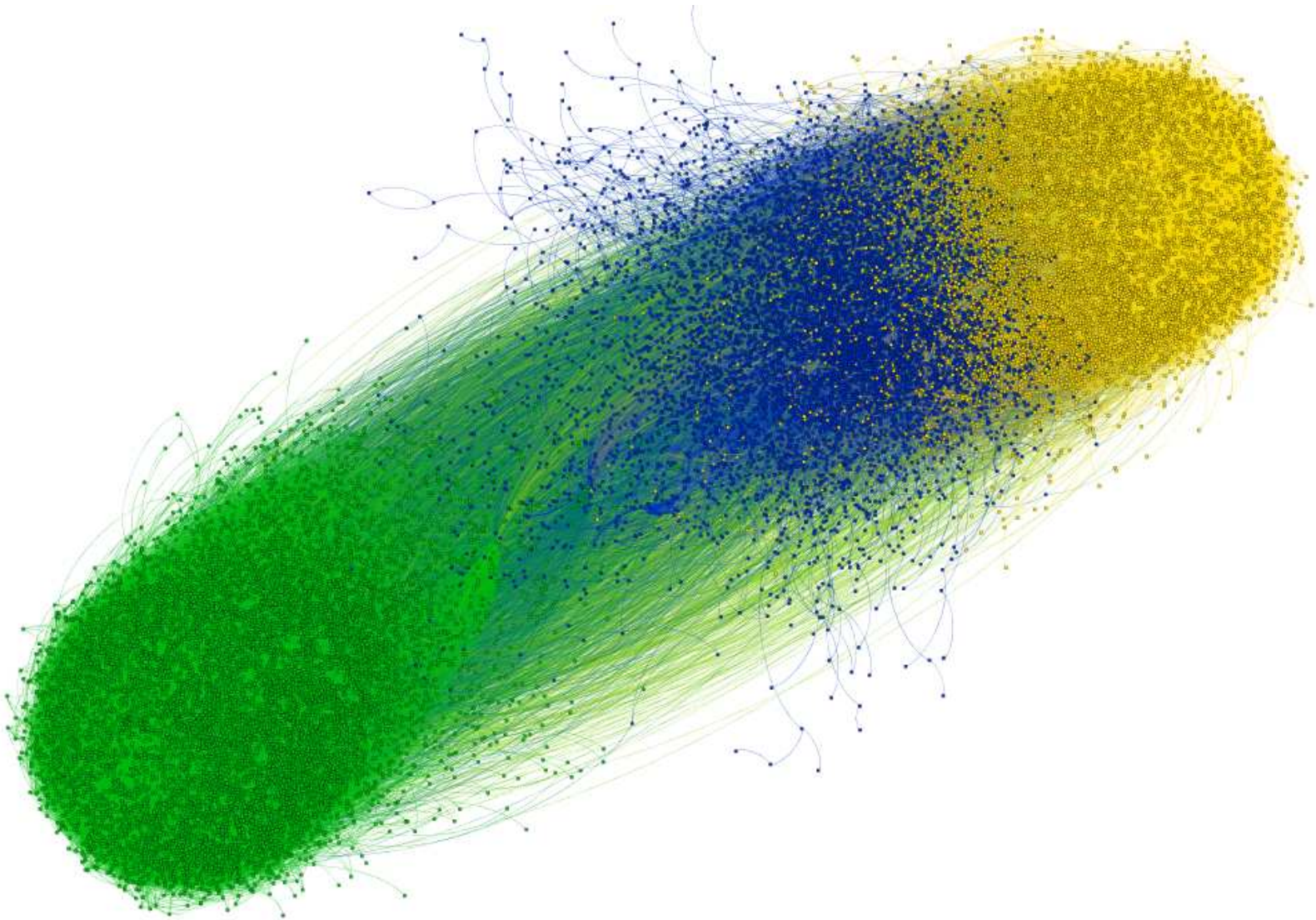
- **Non-polarized networks:**
- Facebook friends: Q = +0.24, P= -0.24
- NY Giants and NY Knicks (Twitter): Q = +0.15, P= -0.002
- **Polarized networks:**
- Brazilian Soccer (Twitter): Q = +0.39, P= +0.20
- 2004 U.S. Political Blogosphere: Q =+0.42, P= +0.18

Outline

Gun Control Debate on Twitter

- keywords: `gun control`, `guns`, `mass shootings` and `NRA`
- from December 14, 2012 to February 10, 2013
- ~3.8 million tweets

Gun Control Debate on Twitter – retweet graph



Gun Control Debate on Twitter

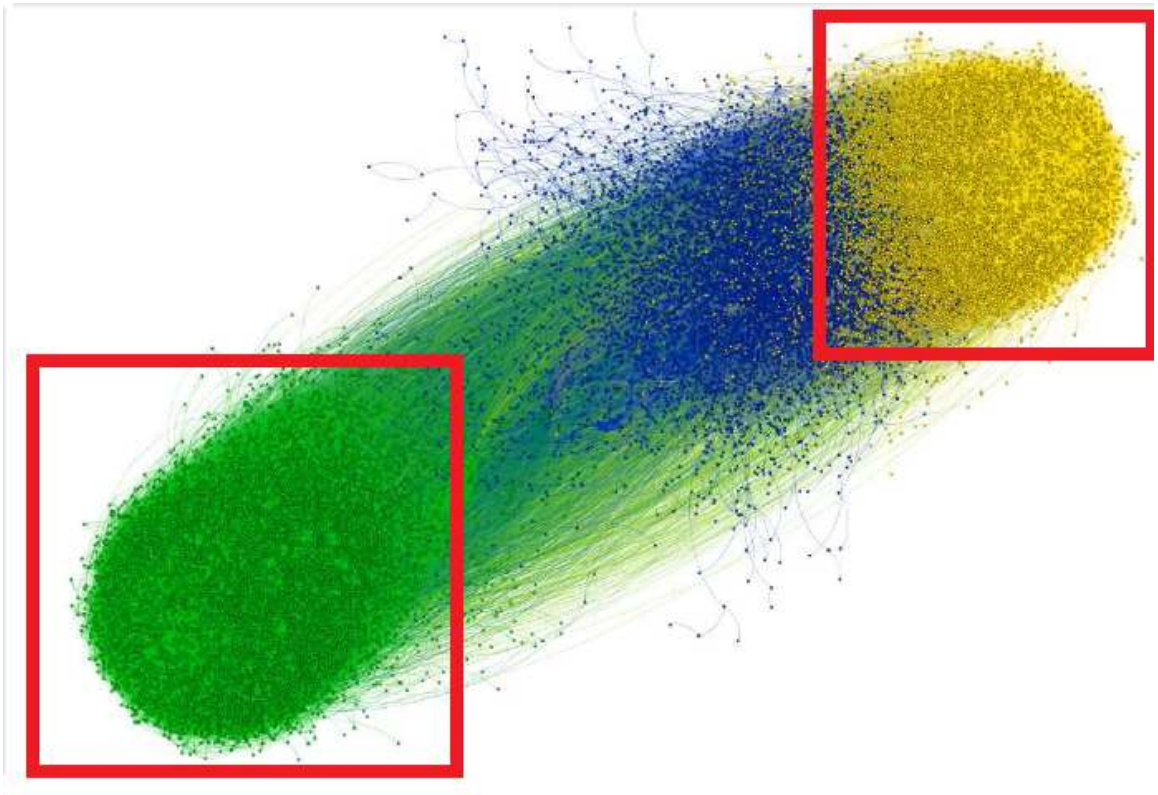


Figure: Q(green,yellow) = **+0.47**, P= **+0.32**

Gun Control Debate on Twitter

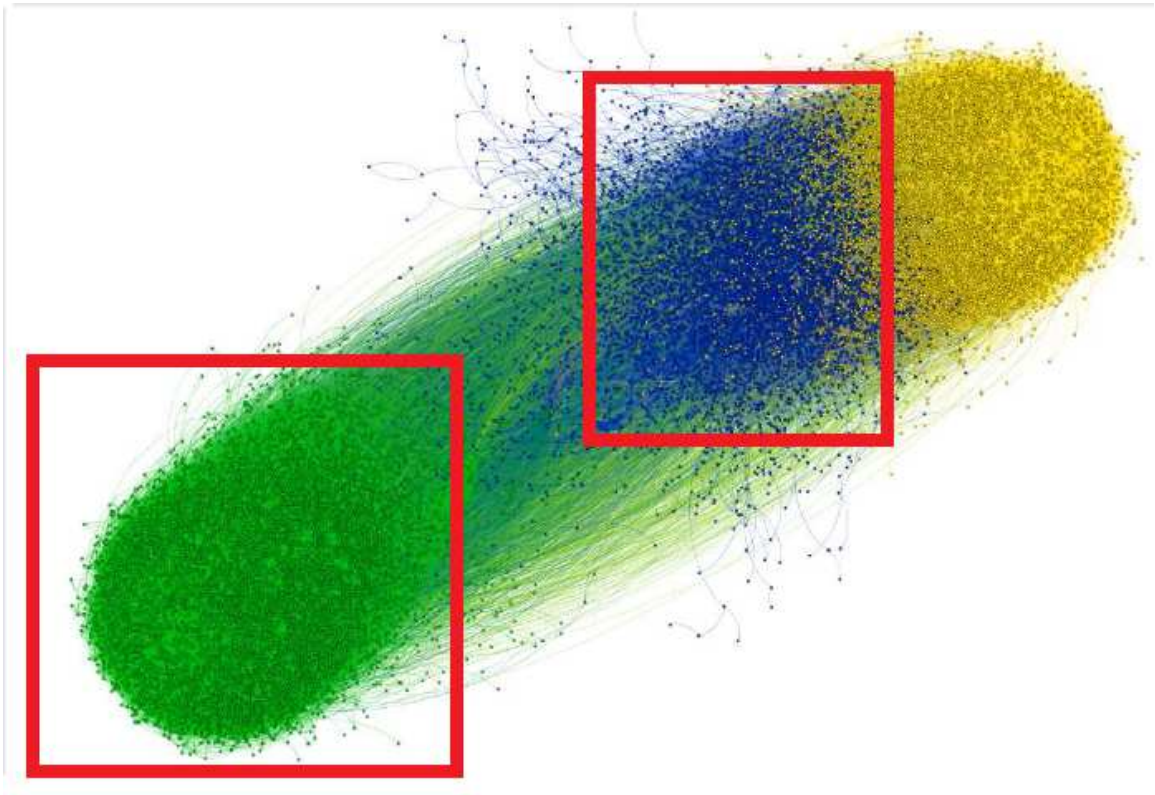


Figure: Q(green,blue) = **+0.31**, P = **+0.23**

Gun Control Debate on Twitter

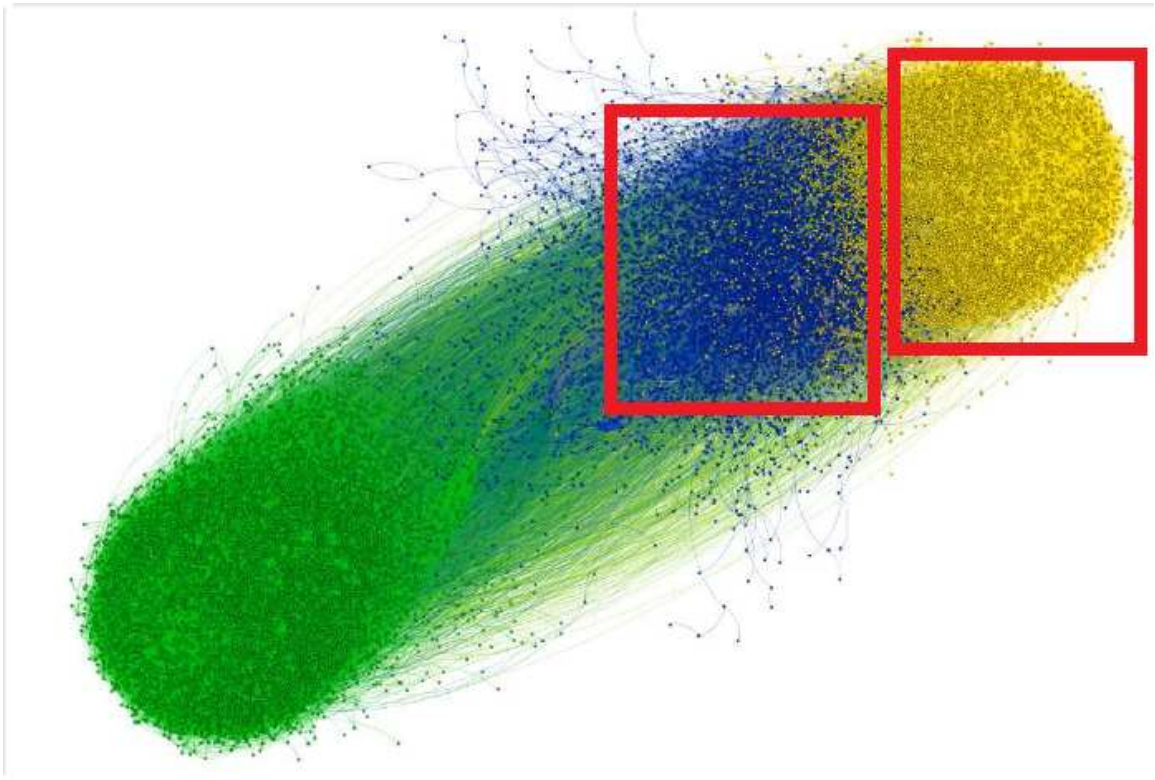


Figure: Q(blue,yellow) = **+0.26**, P = **-0.14**

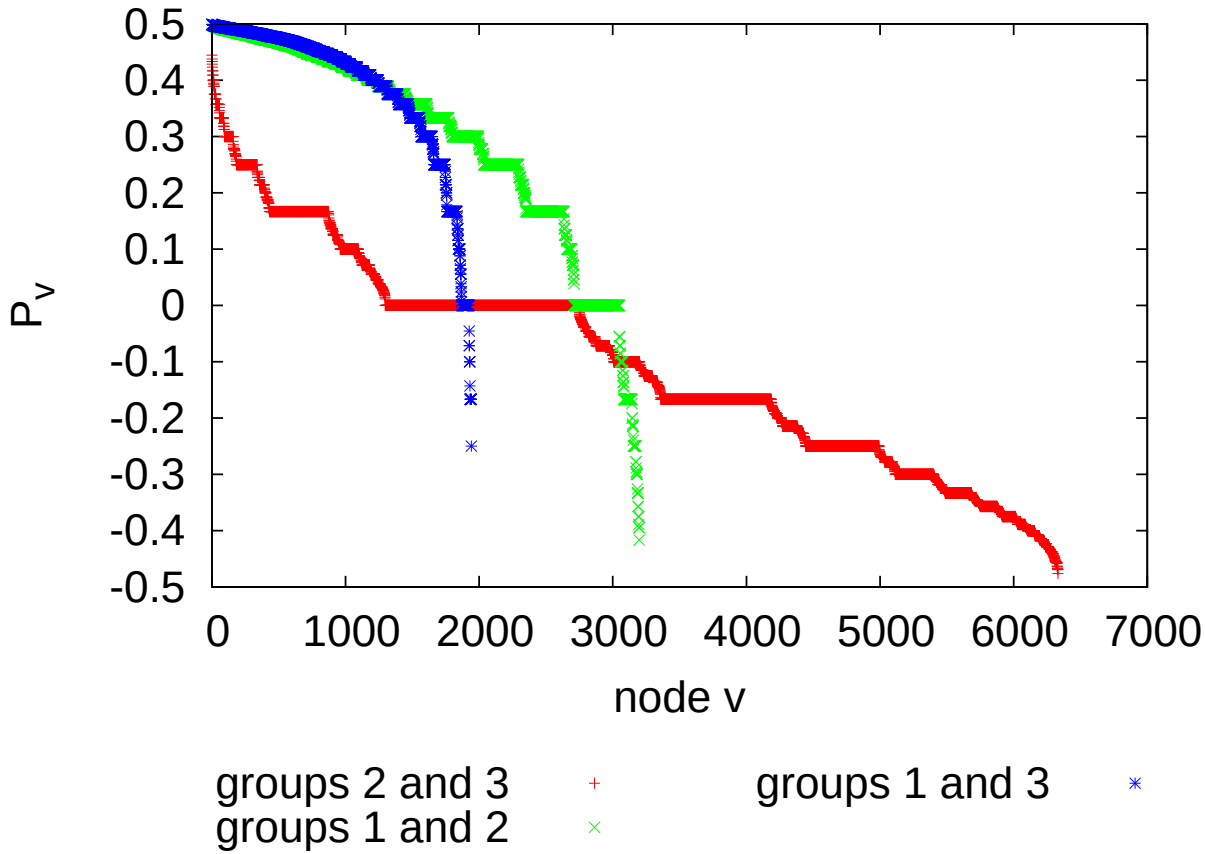
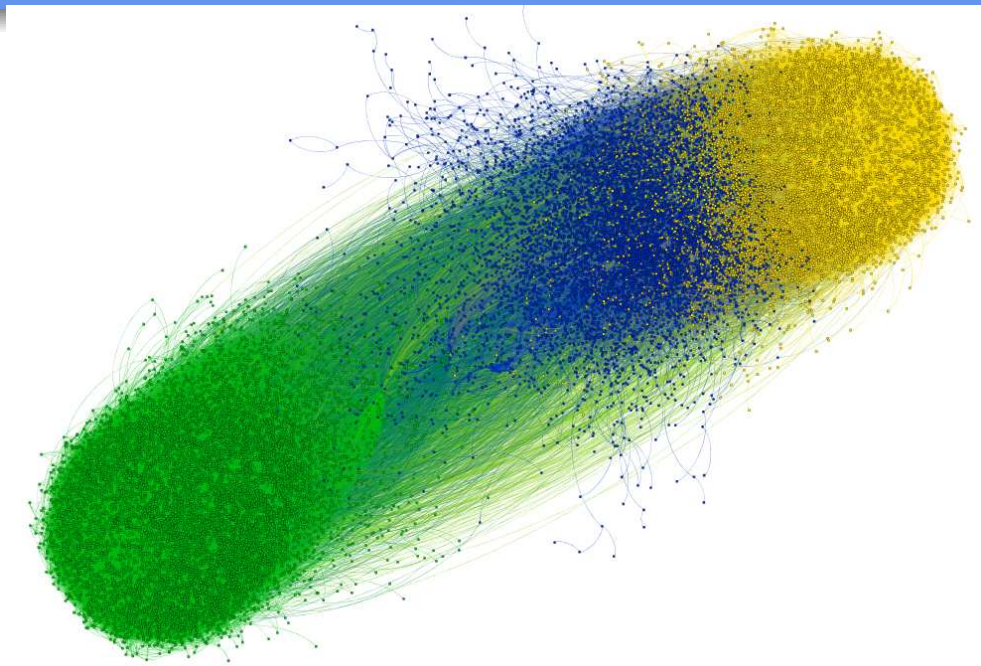


Figure: Distribution of P_v for communities debating Gun Control.



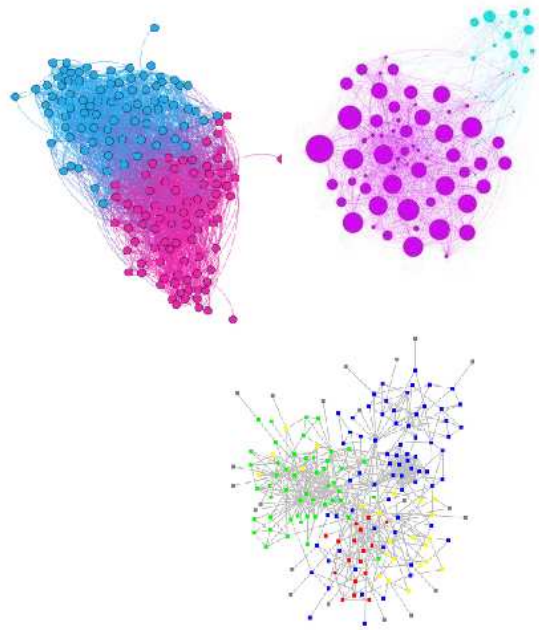
Figure

- conservatives (no gun control)
- independent (pro-gun control)
- liberals (pro-gun control)

Outline

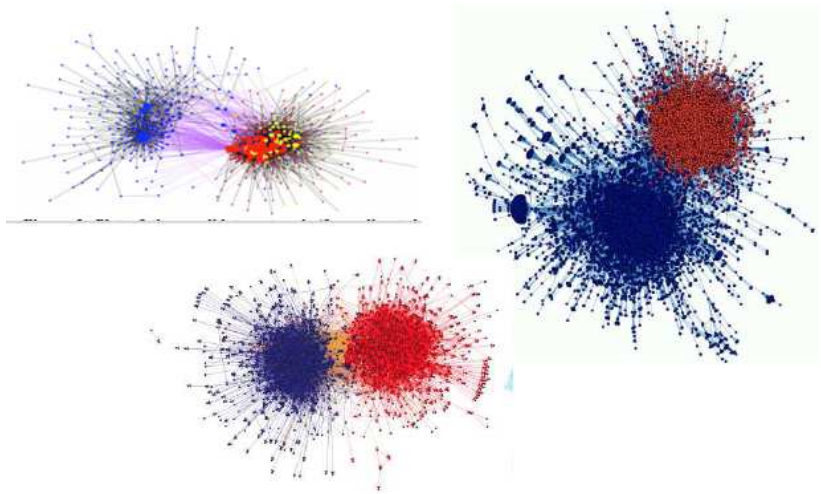
Concentration of popular nodes in the boundary

- **non-polarized communities**: promotes the existence of high-degree nodes in the boundaries
  1. fans of football and fans of basketball
  2. college friends and family



Concentration of popular nodes in the boundary

- **polarized communities**: popular nodes far from the boundary, as strong representatives of their group viewpoints



two ranks:

- r: a rank of all nodes in the graph sorted by degree

two ranks:

- r: a rank of all nodes in the graph sorted by degree
- r_b : a rank of all nodes in the graph according to the number of cross-boundary connections

two ranks:

- r: a rank of all nodes in the graph sorted by degree
- r_b : a rank of all nodes in the graph according to the number of cross-boundary connections

- Are high-ranked nodes in r also high-ranked in r_b?
- Spearman's correlation $\rightarrow$ [-1,1]

Concentration of popular nodes on the boundary

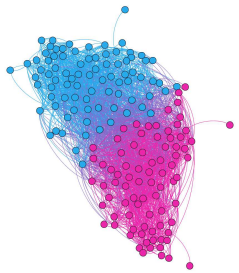


Figure: Facebook Friends (grad and undergrad communities)

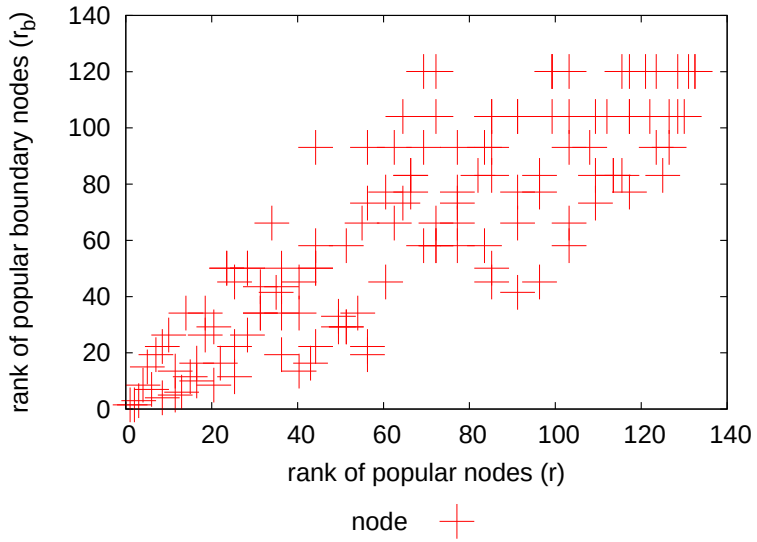


Figure: $\rho = 0.84 \rightarrow$ high correlation, many popular nodes in the boundary, no polarization

Gun Control Debate

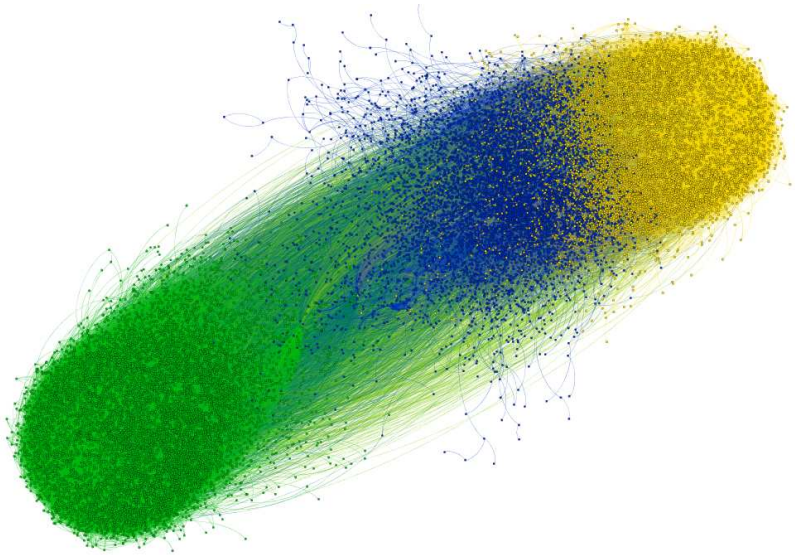


Figure: Communities on Gun Control Debate

- comm. green and yellow: $\rho = \mathbf{0.21}$
- comm. green and blue: $\rho = \mathbf{0.23}$
- comm. blue and yellow: $\rho = \mathbf{0.70}$ **lack of polarization**

Outline

Summary

1. we examine **both** polarized and non-polarized social networks
2. we propose a new metric of polarization (P)
   1. focus on **antagonism**
   2. tends to generate **negative** measurements for non-polarized social networks
3. we show that non-polarized networks tend to show a higher concentration of popular nodes in the intersection

Summary

1. we examine **both** polarized and non-polarized social networks
2. we propose a new metric of polarization (P)
   1. focus on **antagonism**
   2. tends to generate **negative** measurements for non-polarized social networks
3. we show that non-polarized networks tend to show a higher concentration of popular nodes in the intersection
4. Q, P and ρ can be used together

Ongoing Work

1. **multipolarization:** multiple sides!
2. support, antagonism, **indifference**
3. mathematical relationship between Q, P and ρ

More details on `www.dcc.ufmg.br/~pcalais`

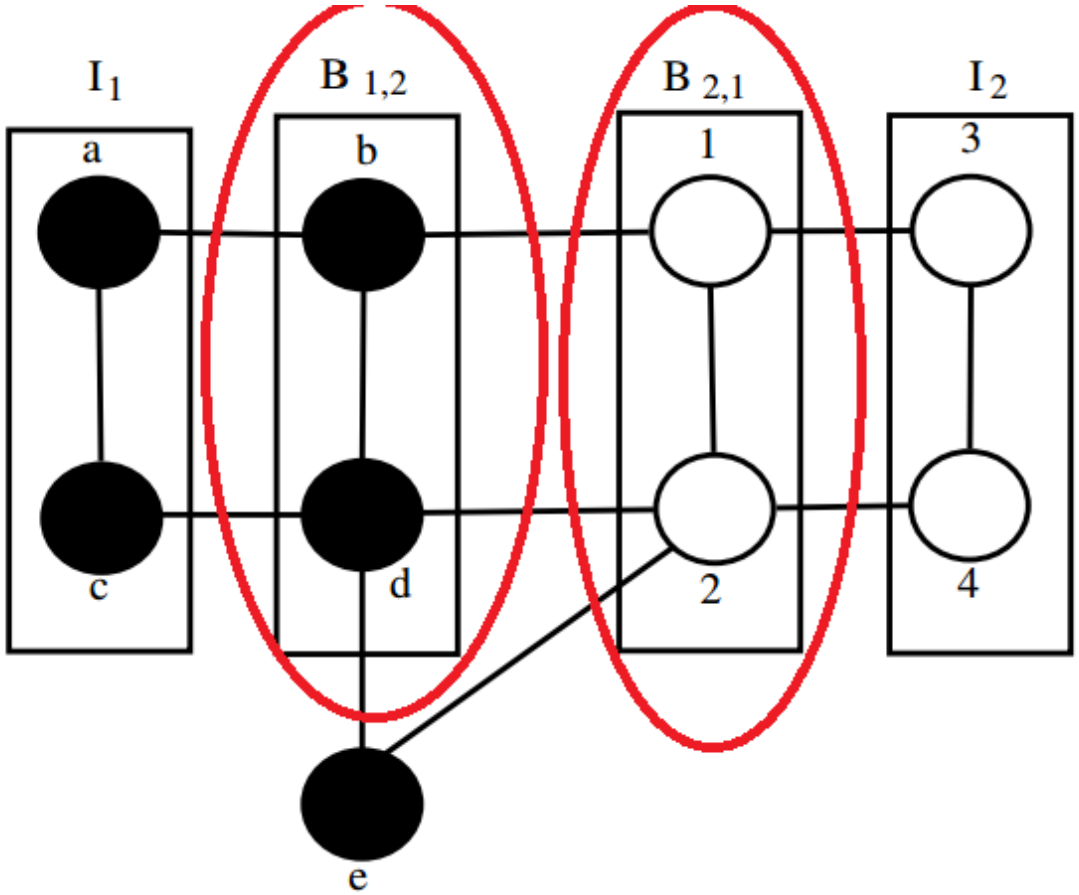
A Measure of Polarization on Social Media Networks Based on Community Boundaries

Pedro Calais Guerra, **Wagner Meira Jr.**

Federal University of Minas Gerais (UFMG), Brazil

Claire Cardie, Robert Kleinberg

Cornell University



Figure

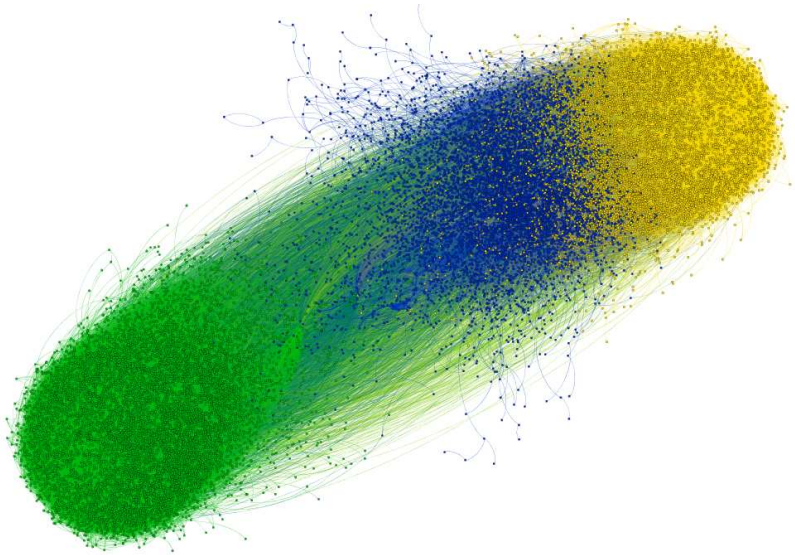


Figure: Communities on Gun Control Debate

- comm. green and yellow: $\rho = \mathbf{0.21}$ Q $= 0.47$
- comm. green and blue: $\rho = \mathbf{0.23}$ Q $= 0.31$
- comm. blue and yellow: $\rho = \mathbf{0.70}$ Q $= 0.26$