Follow Your Ideology: Measuring Media Ideology on Social Networks

January 2017

Or How We Determined That Fox News is Conservative

January 2017

Fox News' Special Report, while right of center, was closer to the center than any of the three major networks' evening news broadcasts.

Groseclose and Milyo, 2005, abstract

Fox News' Special Report, while right of center, was closer to the center than any of the three major networks' evening news broadcasts.

Groseclose and Milyo, 2005, abstract

The fourth and fifth most centrist outlets are the Drudge Report and Fox News' Special Report with Brit Hume

Groseclose and Milyo, 2005, p.1221

Fox News' Special Report, while right of center, was closer to the center than any of the three major networks' evening news broadcasts.

Groseclose and Milyo, 2005, abstract

The fourth and fifth most centrist outlets are the Drudge Report and Fox News' Special Report with Brit Hume

Groseclose and Milyo, 2005, p.1221

Media outlet	Estimated ADA score
Fox News' Special Report with Brit Hume	39.7
Drudge Report	60.4

Fox News' Special Report, while right of center, was closer to the center than any of the three major networks' evening news broadcasts.

Groseclose and Milyo, 2005, abstract

The fourth and fifth most centrist outlets are the Drudge Report and Fox News' Special Report with Brit Hume

Groseclose and Milyo, 2005, p.1221

Media outlet	Estimated ADA score
Fox News' Special Report with Brit Hume	39.7
Drudge Report	60.4
Average American voter	50.1
Average Republican legislator	16.1
Average Democratic legislator	84.4

Motivation

Why measure ideology of the news media?

Motivation

Why measure ideology of the news media?

- Mediated consumption of policy relevant information

Motivation

Why measure ideology of the news media?

- Mediated consumption of policy relevant information
    - Politicians

Motivation

Why measure ideology of the news media?

- Mediated consumption of policy relevant information
  - Politicians $\longrightarrow$ Media

Motivation

Why measure ideology of the news media?

- Mediated consumption of policy relevant information
  - Politicians $\longrightarrow$ Media $\longrightarrow$ Public

Motivation

Why measure ideology of the news media?

- Mediated consumption of policy relevant information
  - Politicians $\longrightarrow$ Media $\longrightarrow$ Public
  - Events in the world

Motivation

Why measure ideology of the news media?

- Mediated consumption of policy relevant information
  - Politicians $\longrightarrow$ Media $\longrightarrow$ Public
  - Events in the world $\longrightarrow$ Media

Motivation

Why measure ideology of the news media?

- Mediated consumption of policy relevant information
  - Politicians $\longrightarrow$ Media $\longrightarrow$ Public
  - Events in the world $\longrightarrow$ Media $\longrightarrow$ Public

Motivation

Why measure ideology of the news media?

- Mediated consumption of policy relevant information
  - Politicians $\longrightarrow$ Media $\longrightarrow$ Public
  - Events in the world $\longrightarrow$ Media $\longrightarrow$ Public
- Worry:

Motivation

Why measure ideology of the news media?

- Mediated consumption of policy relevant information
  - Politicians $\longrightarrow$ Media $\longrightarrow$ Public
  - Events in the world $\longrightarrow$ Media $\longrightarrow$ Public
- Worry:
  - Biased media coverage

Motivation

Why measure ideology of the news media?

- Mediated consumption of policy relevant information
  - Politicians $\longrightarrow$ Media $\longrightarrow$ Public
  - Events in the world $\longrightarrow$ Media $\longrightarrow$ Public
- Worry:
  - Biased media coverage $\longrightarrow$ Biased voters' opinions

How to measure ideology?

– Direct methods:

How to measure ideology?

- Direct methods:
  - Expressed positions on issues (Puglisi and Snyder, 2011)

How to measure ideology?

- Direct methods:
  - Expressed positions on issues (Puglisi and Snyder, 2011)
    - $\rightarrow$ Rare and self-selected

How to measure ideology?

- Direct methods:
  - Expressed positions on issues (Puglisi and Snyder, 2011)
    - $\rightarrow$ Rare and self-selected
- Indirect methods:

How to measure ideology?

- Direct methods:
  - Expressed positions on issues (Puglisi and Snyder, 2011)
    - $\rightarrow$ Rare and self-selected
- Indirect methods:
  - Citations of interest groups (Groseclose and Milyo, 2005)

How to measure ideology?

- Direct methods:
  - Expressed positions on issues (Puglisi and Snyder, 2011)
    - $\rightarrow$ Rare and self-selected
- Indirect methods:
  - Citations of interest groups (Groseclose and Milyo, 2005)
  - Language similarity wrt legislators (Gentzkow and Shapiro, 2010)

How to measure ideology?

- Direct methods:
  - Expressed positions on issues (Puglisi and Snyder, 2011)
    - $\rightarrow$ Rare and self-selected
- Indirect methods:
  - Citations of interest groups (Groseclose and Milyo, 2005)
  - Language similarity wrt legislators (Gentzkow and Shapiro, 2010)
    - $\rightarrow$ Complexity of automated text analysis

How to measure ideology?

- Direct methods:
  - Expressed positions on issues (Puglisi and Snyder, 2011)
    - $\rightarrow$ Rare and self-selected
- Indirect methods:
  - Citations of interest groups (Groseclose and Milyo, 2005)
  - Language similarity wrt legislators (Gentzkow and Shapiro, 2010)
    - $\rightarrow$ Complexity of automated text analysis
    - $\rightarrow$ Agenda coverage vs slant.

How to measure ideology?

- Direct methods:
  - Expressed positions on issues (Puglisi and Snyder, 2011)
    - $\rightarrow$ Rare and self-selected
- Indirect methods:
  - Citations of interest groups (Groseclose and Milyo, 2005)
  - Language similarity wrt legislators (Gentzkow and Shapiro, 2010)
    - $\rightarrow$ Complexity of automated text analysis
    - $\rightarrow$ Agenda coverage vs slant.
    - $\rightarrow$ Good discrimination across parties; not so good within.

How to measure ideology?

- Direct methods:
  - Expressed positions on issues (Puglisi and Snyder, 2011)
    - $\rightarrow$ Rare and self-selected
- Indirect methods:
  - Citations of interest groups (Groseclose and Milyo, 2005)
  - Language similarity wrt legislators (Gentzkow and Shapiro, 2010)
    - $\rightarrow$ Complexity of automated text analysis
    - $\rightarrow$ Agenda coverage vs slant.
    - $\rightarrow$ Good discrimination across parties; not so good within.
- Our contribution: new measure to quantify ideology based on structure of social media networks (Twitter)

Learning about ideology using social networks

3 sets of informational cues about ideology (on social media)

Learning about ideology using social networks

3 sets of informational cues about ideology (on social media)

- Content (tweets' text)

Learning about ideology using social networks

3 sets of informational cues about ideology (on social media)

- Content (tweets' text)
- Shared content ("retweets")

Learning about ideology using social networks

3 sets of informational cues about ideology (on social media)

- Content (tweets' text)
- Shared content ("retweets")
- Audience ("followers")

Learning about ideology using social networks

3 sets of informational cues about ideology (on social media)

- Content (tweets' text)
- Shared content ("retweets")
- Audience ("followers")

Users' decision to follow are informative because they are costly:

Learning about ideology using social networks

3 sets of informational cues about ideology (on social media)

- Content (tweets' text)
- Shared content ("retweets")
- Audience ("followers")

Users' decision to follow are informative because they are costly:

- Opportunity costs: exposure to one source reduces exposure to other sources

Learning about ideology using social networks

3 sets of informational cues about ideology (on social media)

- Content (tweets' text)
- Shared content ("retweets")
- Audience ("followers")

Users' decision to follow are informative because they are costly:

- Opportunity costs: exposure to one source reduces exposure to other sources
- Psychological costs: messages that are uncongenial to the person's existing political beliefs cause discomfort.

Learning about ideology using social networks

3 sets of informational cues about ideology (on social media)

- Content (tweets' text)
- Shared content ("retweets")
- Audience ("followers")

Users' decision to follow are informative because they are costly:

- Opportunity costs: exposure to one source reduces exposure to other sources
- Psychological costs: messages that are uncongenial to the person's existing political beliefs cause discomfort.

Focus on politically interested users: decisions are more motivated by ideology, and thus more informative.

Learning about ideology using social networks

3 sets of informational cues about ideology (on social media)

- Content (tweets' text)
- Shared content ("retweets")
- Audience ("followers")

Users' decision to follow are informative because they are costly:

- Opportunity costs: exposure to one source reduces exposure to other sources
- Psychological costs: messages that are uncongenial to the person's existing political beliefs cause discomfort.

Focus on politically interested users: decisions are more motivated by ideology, and thus more informative.

Ideology: primary dimension of political conflict, *common* to elites and voters (Jessee 2009, Tausanovitch and Warshaw 2013)

A spatial model of following behavior

– Users' and media outlets' ideology (θ_i and ϕ_j) are defined as latent variables to be estimated.

A spatial model of following behavior

- Users' and media outlets' ideology (θ_i and ϕ_j) are defined as latent variables to be estimated.
- Data: "following" decisions as a series of binary choices ($\mathbf{Y_{ij}}$).

A spatial model of following behavior

- Users' and media outlets' ideology (θ_i and ϕ_j) are defined as latent variables to be estimated.
- Data: "following" decisions as a series of binary choices ($\mathbf{Y_{ij}}$).
- Spatial following assumption: users choose to follow the set of media outlets that maximize the objective function:

$$ {}_{y_{1,\ldots,J}} \left[\sum_{j=1}^{J} \alpha_j(y_j) - \beta_i(y_j) - y_j(\gamma||\theta_i - \phi_j||^2) \right] $$

A spatial model of following behavior

- Users' and media outlets' ideology (θ_i and ϕ_j) are defined as latent variables to be estimated.
- Data: "following" decisions as a series of binary choices ($\mathbf{Y_{ij}}$).
- Spatial following assumption: users choose to follow the set of media outlets that maximize the objective function:

$$_{y_{1,\ldots,J}} \left[\sum_{j=1}^{J} \alpha_j(y_j) - \beta_i(y_j) - y_j(\gamma||\theta_i - \phi_j||^2) \right]$$

where:

A spatial model of following behavior

- Users' and media outlets' ideology (θ_i and ϕ_j) are defined as latent variables to be estimated.
- Data: "following" decisions as a series of binary choices ($\mathbf{Y_{ij}}$).
- Spatial following assumption: users choose to follow the set of media outlets that maximize the objective function:

$$\underset{y_{1,\ldots,J}}{} \left[\sum_{j=1}^{J} \alpha_j(y_j) - \beta_i(y_j) - y_j(\gamma||\theta_i - \phi_j||^2) \right]$$

where:

α_j measures *popularity* of outlet j

A spatial model of following behavior

- Users' and media outlets' ideology (θ_i and ϕ_j) are defined as latent variables to be estimated.
- Data: "following" decisions as a series of binary choices ($\mathbf{Y_{ij}}$).
- Spatial following assumption: users choose to follow the set of media outlets that maximize the objective function:

$$\underset{y_{1,\ldots,J}}{} \left[\sum_{j=1}^{J} \alpha_j(y_j) - \beta_i(y_j) - y_j(\gamma||\theta_i - \phi_j||^2) \right]$$

where:

α_j measures *popularity* of outlet j

β_i measures *following cost* of user i

A spatial model of following behavior

– Users' and media outlets' ideology (θ_i and ϕ_j) are defined as latent variables to be estimated.
– Data: "following" decisions as a series of binary choices ($\mathbf{Y_{ij}}$).
– Spatial following assumption: users choose to follow the set of media outlets that maximize the objective function:

$$\underset{y_{1,\ldots,J}}{} \left[\sum_{j=1}^{J} \alpha_j(y_j) - \beta_i(y_j) - y_j(\gamma||\theta_i - \phi_j||^2) \right]$$

where:

α_j measures *popularity* of outlet j

β_i measures *following cost* of user i

– Following Bonica (2013), we estimate θ_i and ϕ_j using correspondence analysis. Mathematically close to log-linear ideal point model (Lowe, 2008).

Data

- List of 2,363 Twitter accounts of journalists, news programs and national news outlets in the U.S.

Data

- List of 2,363 Twitter accounts of journalists, news programs and national news outlets in the U.S.
  - Source: websites, Twitter lists, etc.

Data

- List of 2,363 Twitter accounts of journalists, news programs and national news outlets in the U.S.
  - Source: websites, Twitter lists, etc.
  - Collected lists of followers from REST API

Data

- List of 2,363 Twitter accounts of journalists, news programs and national news outlets in the U.S.
  - Source: websites, Twitter lists, etc.
  - Collected lists of followers from REST API
- Followers:

Data

- List of 2,363 Twitter accounts of journalists, news programs and national news outlets in the U.S.
  - Source: websites, Twitter lists, etc.
  - Collected lists of followers from REST API
- Followers:
  - From total number of unique followers (72,259,123)...

Data

- List of 2,363 Twitter accounts of journalists, news programs and national news outlets in the U.S.
  - Source: websites, Twitter lists, etc.
  - Collected lists of followers from REST API
- Followers:
  - From total number of unique followers (72,259,123)...
  - $\rightarrow$ ...filter those who follow <5 journalists; or located outside US.

Data

- List of 2,363 Twitter accounts of journalists, news programs and national news outlets in the U.S.
  - Source: websites, Twitter lists, etc.
  - Collected lists of followers from REST API
- Followers:
  - From total number of unique followers (72,259,123)...
  - $\rightarrow$ ...filter those who follow <5 journalists; or located outside US.
  - $\rightarrow$ ...keep only politically interested: follow >2 political actors

Data

- List of 2,363 Twitter accounts of journalists, news programs and national news outlets in the U.S.
  - Source: websites, Twitter lists, etc.
  - Collected lists of followers from REST API
- Followers:
  - From total number of unique followers (72,259,123)...
  - $\rightarrow$ ...filter those who follow <5 journalists; or located outside US.
  - $\rightarrow$ ...keep only politically interested: follow >2 political actors

Total: n = 4,140,572 *informants*

Data

- List of 2,363 Twitter accounts of journalists, news programs and national news outlets in the U.S.
  - Source: websites, Twitter lists, etc.
  - Collected lists of followers from REST API
- Followers:
  - From total number of unique followers (72,259,123)...
  - $\rightarrow$ ...filter those who follow <5 journalists; or located outside US.
  - $\rightarrow$ ...keep only politically interested: follow >2 political actors

  Total: n = 4,140,572 *informants*
- Members of Congress

Data

- List of 2,363 Twitter accounts of journalists, news programs and national news outlets in the U.S.
  - Source: websites, Twitter lists, etc.
  - Collected lists of followers from REST API
- Followers:
  - From total number of unique followers (72,259,123)...
  - $\rightarrow$ ...filter those who follow <5 journalists; or located outside US.
  - $\rightarrow$ ...keep only politically interested: follow >2 political actors

Total: n = 4,140,572 *informants*

- Members of Congress
  - Source: NYTimes Congress API

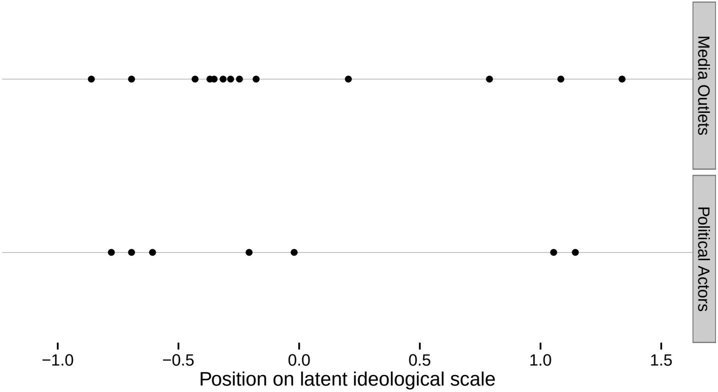
Data

- List of 2,363 Twitter accounts of journalists, news programs and national news outlets in the U.S.
  - Source: websites, Twitter lists, etc.
  - Collected lists of followers from REST API
- Followers:
  - From total number of unique followers (72,259,123)...
  - $\rightarrow$ ...filter those who follow <5 journalists; or located outside US.
  - $\rightarrow$ ...keep only politically interested: follow >2 political actors

  Total: n = 4,140,572 *informants*
- Members of Congress
  - Source: NYTimes Congress API
  - Collected lists of followers from REST API

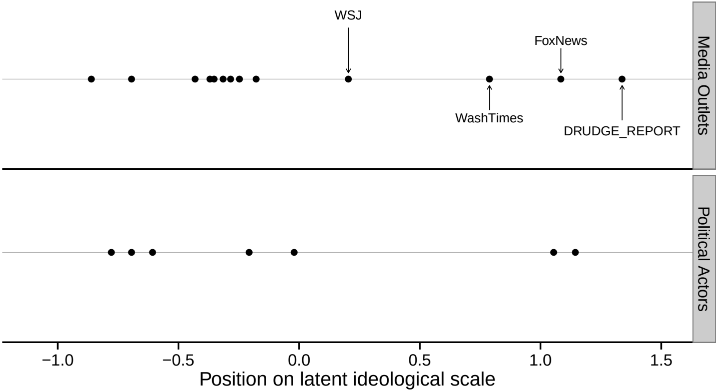
Results

Ideology estimates for main outlets and political actors



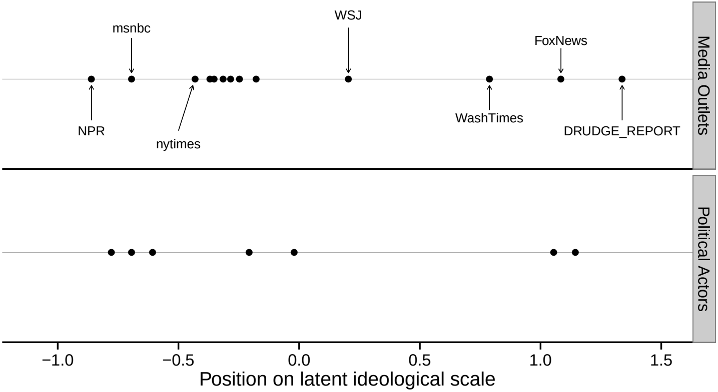
Results

Ideology estimates for main outlets and political actors



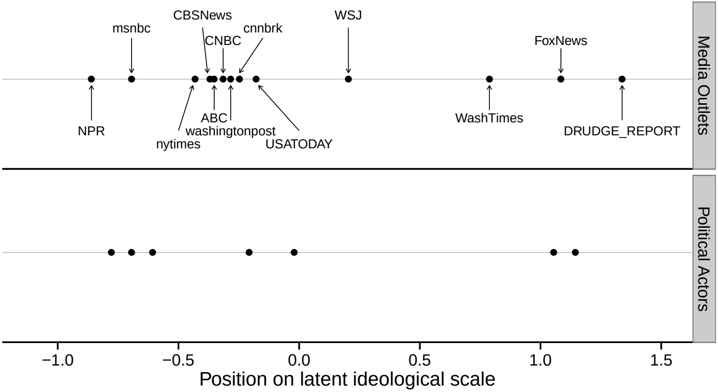
Results

Ideology estimates for main outlets and political actors



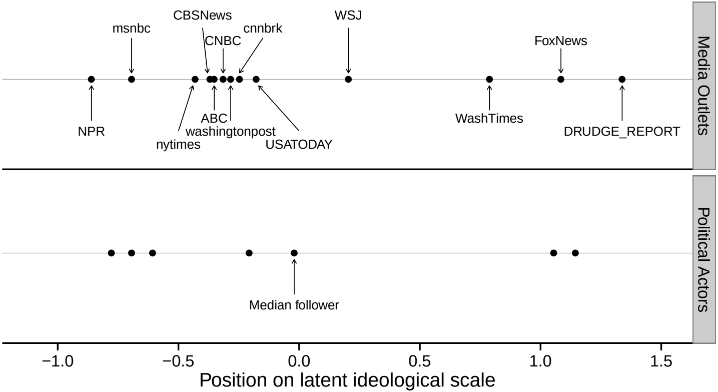
Results

Ideology estimates for main outlets and political actors



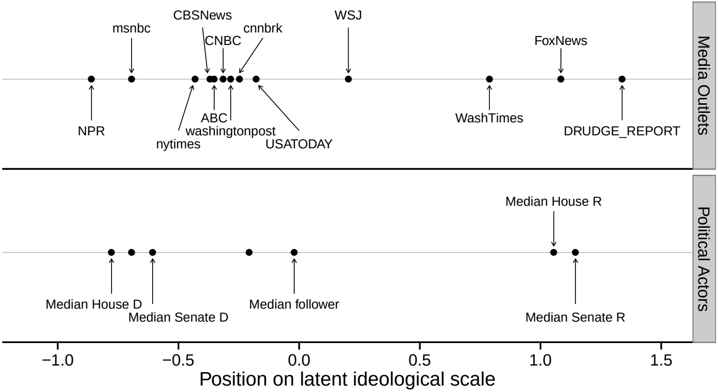
Results

Ideology estimates for main outlets and political actors



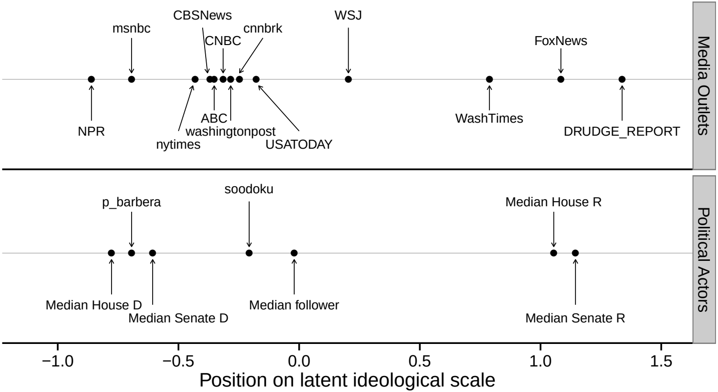
Results

Ideology estimates for main outlets and political actors



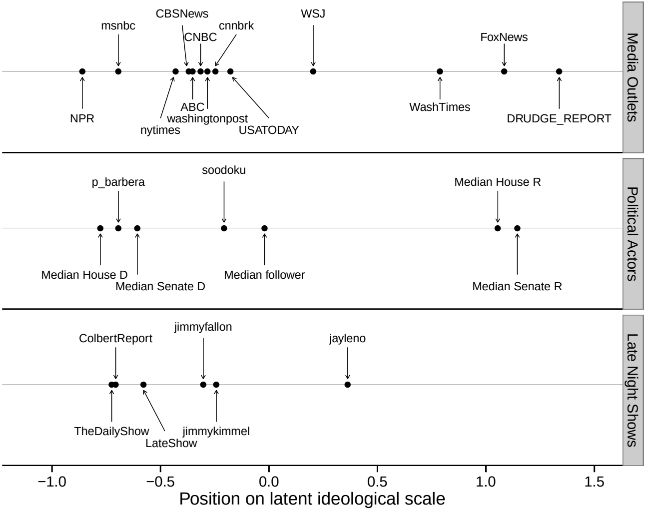
Results

Ideology estimates for main outlets and political actors



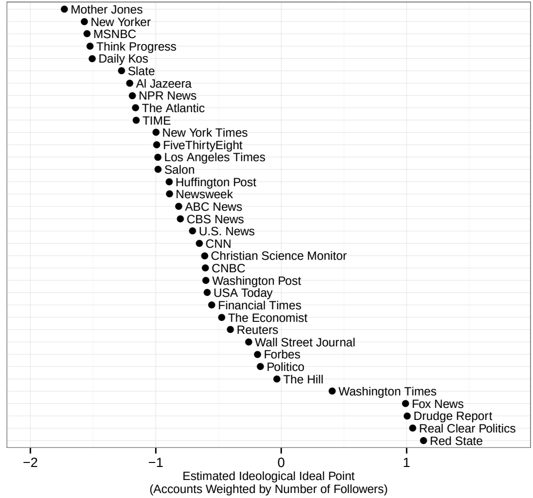
Results

Ideology estimates for main outlets and political actors



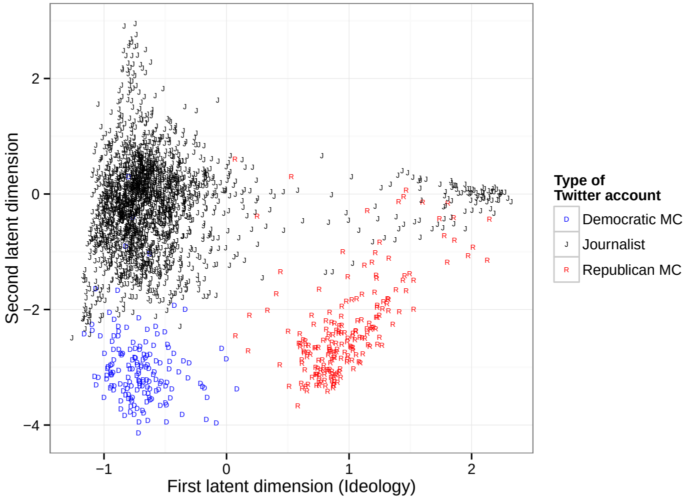
Results

Weighted estimates of media ideology (all outlets)



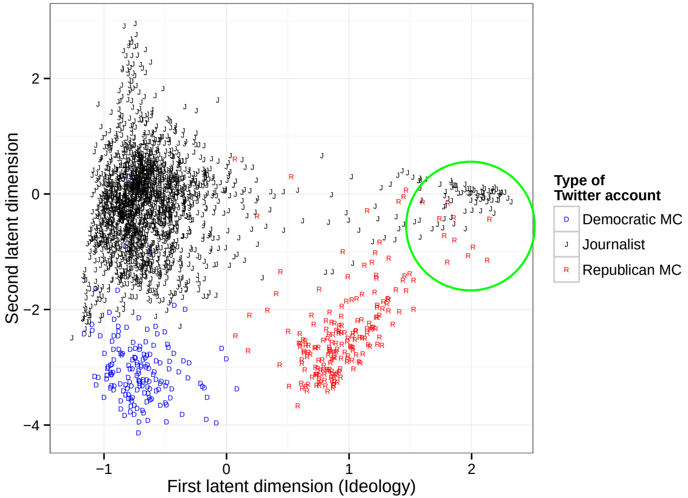
Results

Two-dimensional ideology estimates for outlets and politicians



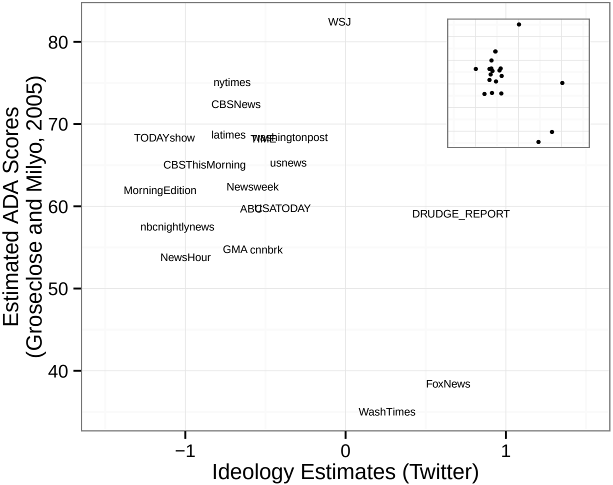
Results

Two-dimensional ideology estimates for outlets and politicians



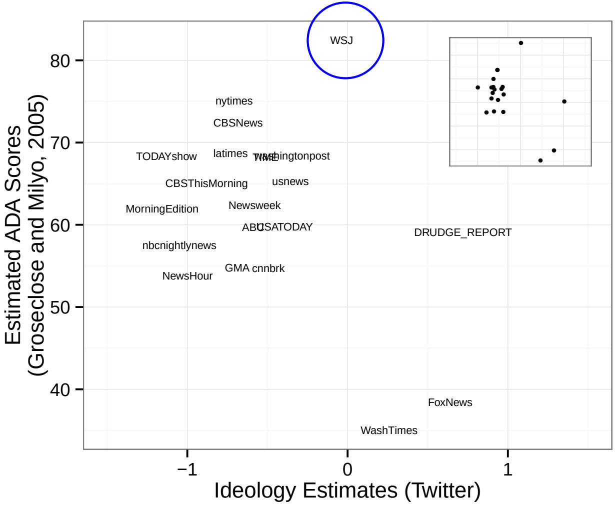
Construct validity

Comparison with estimates in Groseclose and Milyo (2005)



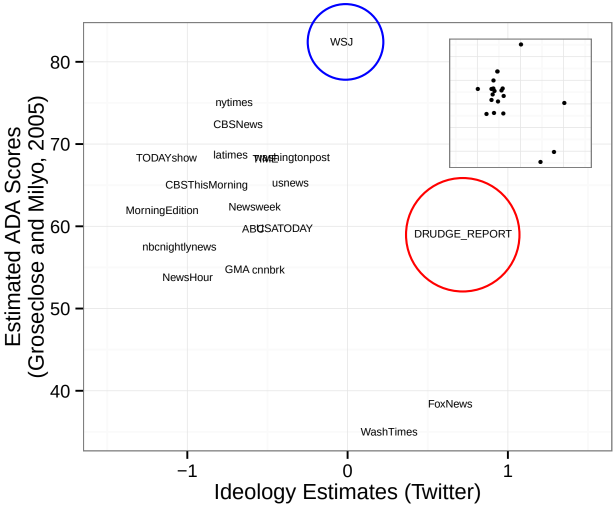
Construct validity

Comparison with estimates in Groseclose and Milyo (2005)



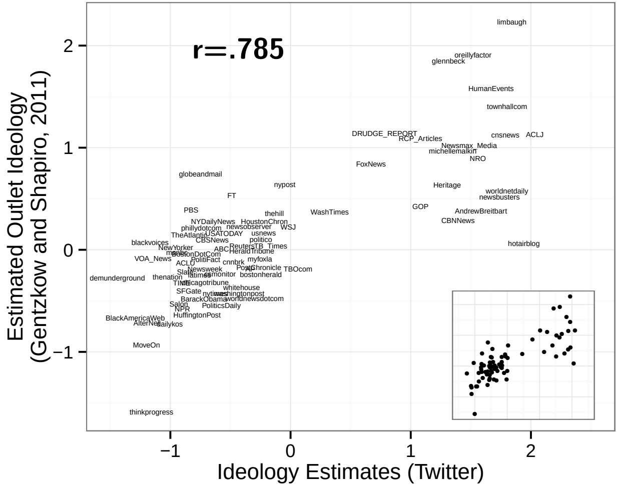
Construct validity

Comparison with estimates in Groseclose and Milyo (2005)



Construct validity

Comparison with estimates in Gentzkow and Shapiro (2011)



Convergent validity

- How consistent are our estimates consistent with those resulting from scaling tweets' text (e.g. Toff and Kim, 2013)

Convergent validity

- How consistent are our estimates consistent with those resulting from scaling tweets' text (e.g. Toff and Kim, 2013)
- Data: 3,200 most recent tweets from all media outlets and journalists in our sample, and Members of Congress.

Convergent validity

– How consistent are our estimates consistent with those resulting from scaling tweets' text (e.g. Toff and Kim, 2013)

– Data: 3,200 most recent tweets from all media outlets and journalists in our sample, and Members of Congress.

$\rightarrow$ Source: Twitter REST API

Convergent validity

- How consistent are our estimates consistent with those resulting from scaling tweets' text (e.g. Toff and Kim, 2013)
- Data: 3,200 most recent tweets from all media outlets and journalists in our sample, and Members of Congress.
  - $\rightarrow$ Source: Twitter REST API
- Method: *Wordscores* (Laver, Benoit, and Garry 2003), using legislators' tweets as "reference texts"

Convergent validity

- How consistent are our estimates consistent with those resulting from scaling tweets' text (e.g. Toff and Kim, 2013)
- Data: 3,200 most recent tweets from all media outlets and journalists in our sample, and Members of Congress.
  - → Source: Twitter REST API
- Method: *Wordscores* (Laver, Benoit, and Garry 2003), using legislators' tweets as "reference texts"
- Results: our method has better out-of-sample performance.

Convergent validity

- How consistent are our estimates consistent with those resulting from scaling tweets' text (e.g. Toff and Kim, 2013)
- Data: 3,200 most recent tweets from all media outlets and journalists in our sample, and Members of Congress.
  - → Source: Twitter REST API
- Method: *Wordscores* (Laver, Benoit, and Garry 2003), using legislators' tweets as "reference texts"
- Results: our method has better out-of-sample performance.

Convergent validity

- How consistent are our estimates consistent with those resulting from scaling tweets' text (e.g. Toff and Kim, 2013)
- Data: 3,200 most recent tweets from all media outlets and journalists in our sample, and Members of Congress.
  - $\rightarrow$ Source: Twitter REST API
- Method: *Wordscores* (Laver, Benoit, and Garry 2003), using legislators' tweets as "reference texts"
- Results: our method has better out-of-sample performance.
  - Correlation with legislators' ideal points estimated using their roll-call votes (Jackman, 2014):

Convergent validity

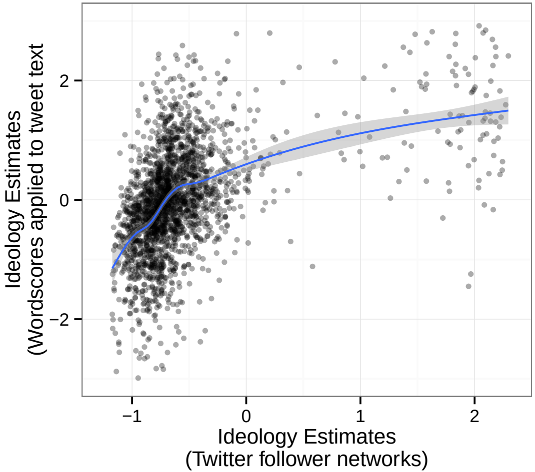
- How consistent are our estimates consistent with those resulting from scaling tweets' text (e.g. Toff and Kim, 2013)
- Data: 3,200 most recent tweets from all media outlets and journalists in our sample, and Members of Congress.
  - → Source: Twitter REST API
- Method: *Wordscores* (Laver, Benoit, and Garry 2003), using legislators' tweets as "reference texts"
- Results: our method has better out-of-sample performance.
  - Correlation with legislators' ideal points estimated using their roll-call votes (Jackman, 2014):
    - → Network-based method: ρ=0.94

Convergent validity

- How consistent are our estimates consistent with those resulting from scaling tweets' text (e.g. Toff and Kim, 2013)
- Data: 3,200 most recent tweets from all media outlets and journalists in our sample, and Members of Congress.
  - → Source: Twitter REST API
- Method: *Wordscores* (Laver, Benoit, and Garry 2003), using legislators' tweets as "reference texts"
- Results: our method has better out-of-sample performance.
  - Correlation with legislators' ideal points estimated using their roll-call votes (Jackman, 2014):
    - → Network-based method: ρ=0.94
    - → Text-based method: ρ=0.89

Convergent validity

Comparing text-based and network-based methods: ρ=0.47



Applications

- Do journalists' private political beliefs affect the content they produce?

Applications

- Do journalists' private political beliefs affect the content they produce?
- Are 'moderate' outlets really centrist or just ideologically diverse?

Applications

- Do journalists' private political beliefs affect the content they produce?
- Are 'moderate' outlets really centrist or just ideologically diverse?
- Ideological Self-selection by Outlet Type

Private political beliefs and ideological slant of content

Private political beliefs and ideological slant of content

– Survey evidence that only a few journalists identify as conservatives

Private political beliefs and ideological slant of content

- Survey evidence that only a few journalists identify as conservatives
- Journalists reject such inference: “political beliefs do not translate into how we cover the news”

Private political beliefs and ideological slant of content

- Survey evidence that only a few journalists identify as conservatives
- Journalists reject such inference: “political beliefs do not translate into how we cover the news”
- How correlated are journalists’ private beliefs and the content they produce?

Private political beliefs and ideological slant of content

- Survey evidence that only a few journalists identify as conservatives
- Journalists reject such inference: “political beliefs do not translate into how we cover the news”
- How correlated are journalists’ private beliefs and the content they produce?
- Analysis:

Private political beliefs and ideological slant of content

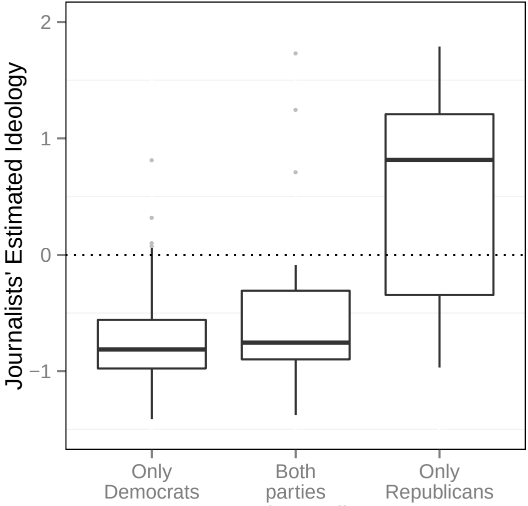
- Survey evidence that only a few journalists identify as conservatives
- Journalists reject such inference: "political beliefs do not translate into how we cover the news"
- How correlated are journalists' private beliefs and the content they produce?
- Analysis:
  - Matched campaign contribution records from DIME database (Bonica, 2013) with Twitter accounts

Private political beliefs and ideological slant of content

- Survey evidence that only a few journalists identify as conservatives
- Journalists reject such inference: “political beliefs do not translate into how we cover the news”
- How correlated are journalists’ private beliefs and the content they produce?
- Analysis:
  - Matched campaign contribution records from DIME database (Bonica, 2013) with Twitter accounts
  - Estimated ideology for n=306 journalists

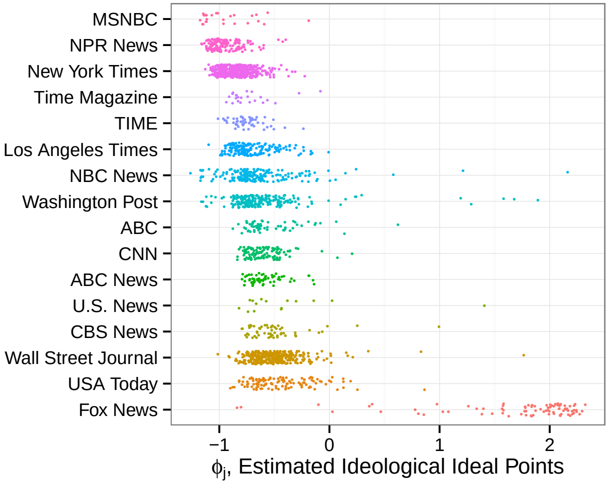
Private political beliefs and ideological slant of content

Journalists' ideology estimates and their campaign contributions

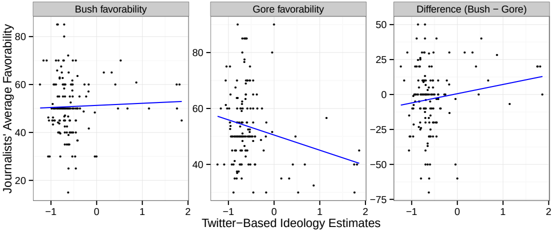


Intra-media ideological heterogeneity

Journalists are clustered within outlets, but high heterogeneity



Ideological Self-Selection



Discussion

– Implications:

Discussion

- Implications:
  - Re-analysis of selective exposure at content unit level

Discussion

- Implications:
  - Re-analysis of selective exposure at content unit level
  - Perceived ideological slant better measure of ideology than those based on content analysis?

Discussion

- Implications:
  - Re-analysis of selective exposure at content unit level
  - Perceived ideological slant better measure of ideology than those based on content analysis?
  - Insights on text classification: can we combine text and Twitter networks to improve scaling method?

Discussion

- Implications:
  - Re-analysis of selective exposure at content unit level
  - Perceived ideological slant better measure of ideology than those based on content analysis?
  - Insights on text classification: can we combine text and Twitter networks to improve scaling method?
  - Potential for comparative study of media system polarization

Discussion

- Implications:
  - Re-analysis of selective exposure at content unit level
  - Perceived ideological slant better measure of ideology than those based on content analysis?
  - Insights on text classification: can we combine text and Twitter networks to improve scaling method?
  - Potential for comparative study of media system polarization
- Wrapping up...

Discussion

- Implications:
  - Re-analysis of selective exposure at content unit level
  - Perceived ideological slant better measure of ideology than those based on content analysis?
  - Insights on text classification: can we combine text and Twitter networks to improve scaling method?
  - Potential for comparative study of media system polarization
- Wrapping up...
  - New method to recover reliable and valid estimates of ideology of a large set of journalists and media outlets.

Discussion

- Implications:
  - Re-analysis of selective exposure at content unit level
  - Perceived ideological slant better measure of ideology than those based on content analysis?
  - Insights on text classification: can we combine text and Twitter networks to improve scaling method?
  - Potential for comparative study of media system polarization
- Wrapping up...
  - New method to recover reliable and valid estimates of ideology of a large set of journalists and media outlets.
  - Journalists' private beliefs and content slant are correlated

Discussion

- Implications:
  - Re-analysis of selective exposure at content unit level
  - Perceived ideological slant better measure of ideology than those based on content analysis?
  - Insights on text classification: can we combine text and Twitter networks to improve scaling method?
  - Potential for comparative study of media system polarization
- Wrapping up...
  - New method to recover reliable and valid estimates of ideology of a large set of journalists and media outlets.
  - Journalists' private beliefs and content slant are correlated
  - Significant heterogeneity in journalists' ideology within outlets

Correspondence analysis (Greenacre, 1993)

- $\mathbf{Y}$ matrix of nonnegative numbers ($n \times m$)
- Correspondence matrix: $\mathbf{P} = \mathbf{Y} / \sum_{ij} y_{ij}$
- Row and column masses, $\mathbf{r}$ and $\mathbf{c}$, where $r_i = \sum_j p_{ij}$ and $c_j = \sum_j p_{ij}$ (row and column sums)
- Diagonal matrices $\mathbf{D}_r = \mathrm{diag}(\mathbf{r})$ and $\mathbf{D}_c = \mathrm{diag}(\mathbf{c})$
- Basic computational algorithm:
  1. Standardized residuals: $\mathbf{S} = \mathbf{D}_r^{1/2}(\mathbf{P} - \mathbf{r}\mathbf{c}^T)\mathbf{D}_c^{1/2}$
  2. SVD: $\mathbf{S} = \mathbf{U}\mathbf{D}_\alpha\mathbf{V}^\mathbf{T}$ where $\mathbf{U}^\mathbf{T}\mathbf{U} = \mathbf{V}^\mathbf{T}\mathbf{V} = \mathbf{I}$
  3. Compute coordinates:

     Rows: ¯ $= \mathbf{D}_r^{1/2}\mathbf{U}$
     Columns: ` $= \mathbf{D}_c^{1/2}\mathbf{V}$
  4. Supplementary points (e.g. column $\mathbf{h}$ of length n):
     - 4.1 Standardize: $\tilde{\mathbf{h}} = \mathbf{h} / \sum_i h_i$
     - 4.2 Project: $\mathbf{g} = \tilde{\mathbf{h}}^T$¯