# Joint Non-negative Matrix Factorization for Learning Ideological Leaning on Twitter

Preethi Lahoti[1], Kiran Garimella[2], Aristides Gionis[2]

[1] Max Planck Institute for Informatics, [2] Aalto University

# Access to Diverse Information Around World …

# Filtered and Cherry Picked Content …





# Twitter

(Retweet or Follow other users)

User – User
social graph

# Twitter

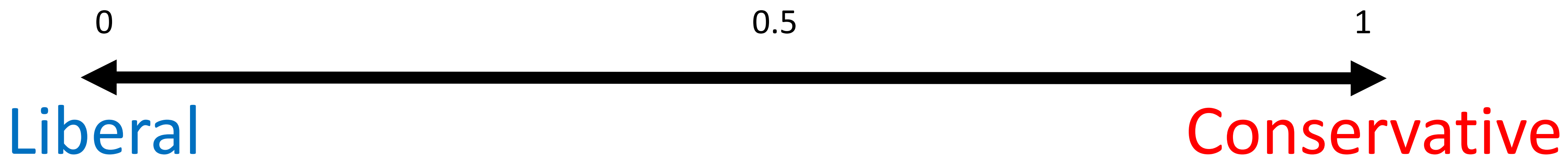
(Retweet or Follow other users)

User – User
social graph

(Retweet a news article)

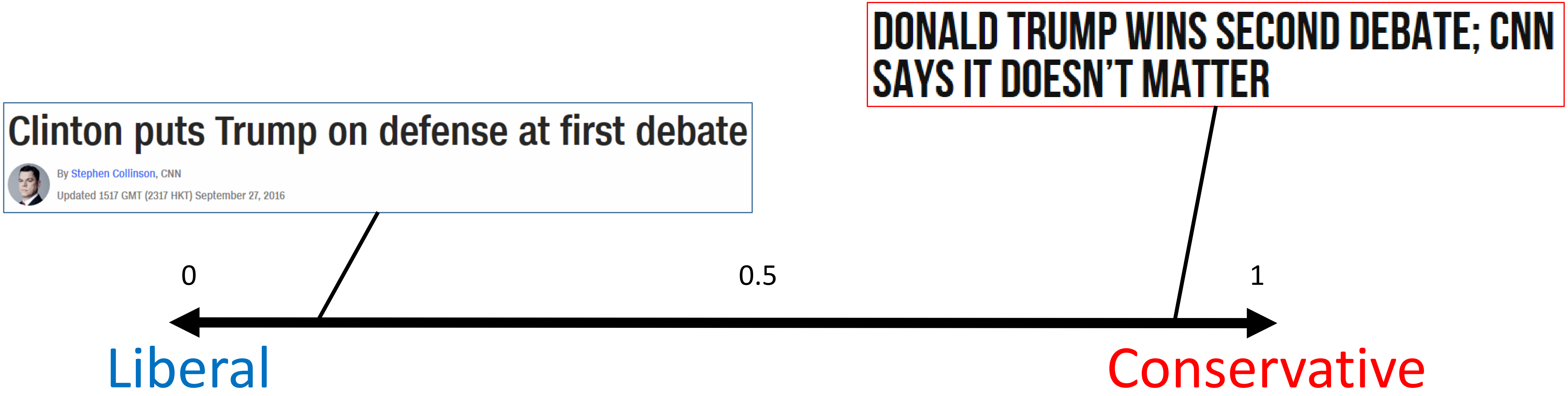
User - Content
content graph

# User and Content Ideology



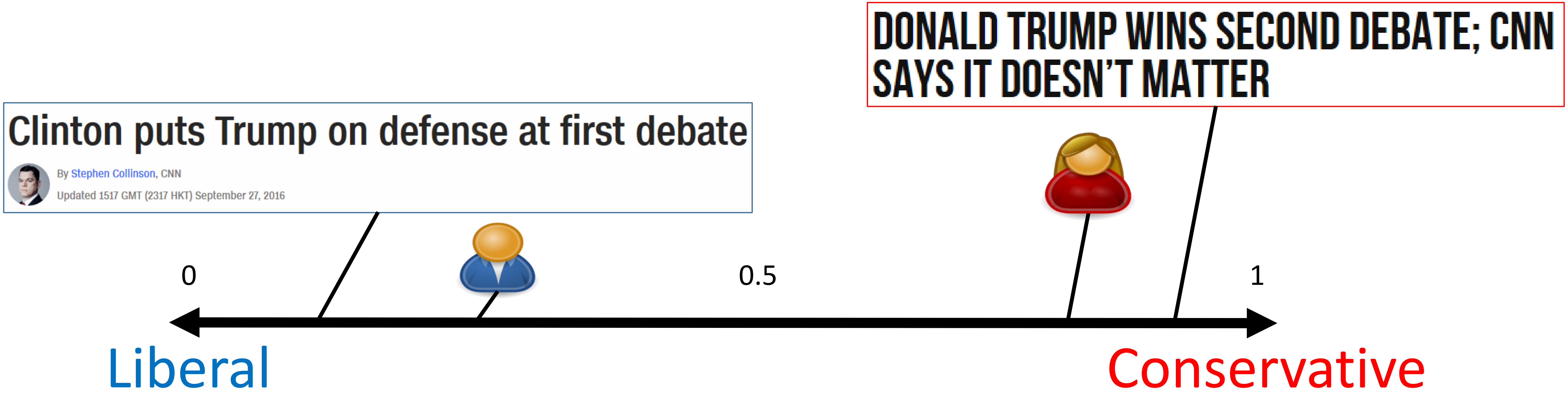

Ideology Scale

# User and Content Ideology




Ideology Scale

# User and Content Ideology




Ideology Scale

# Filter bubble…

User A

User B

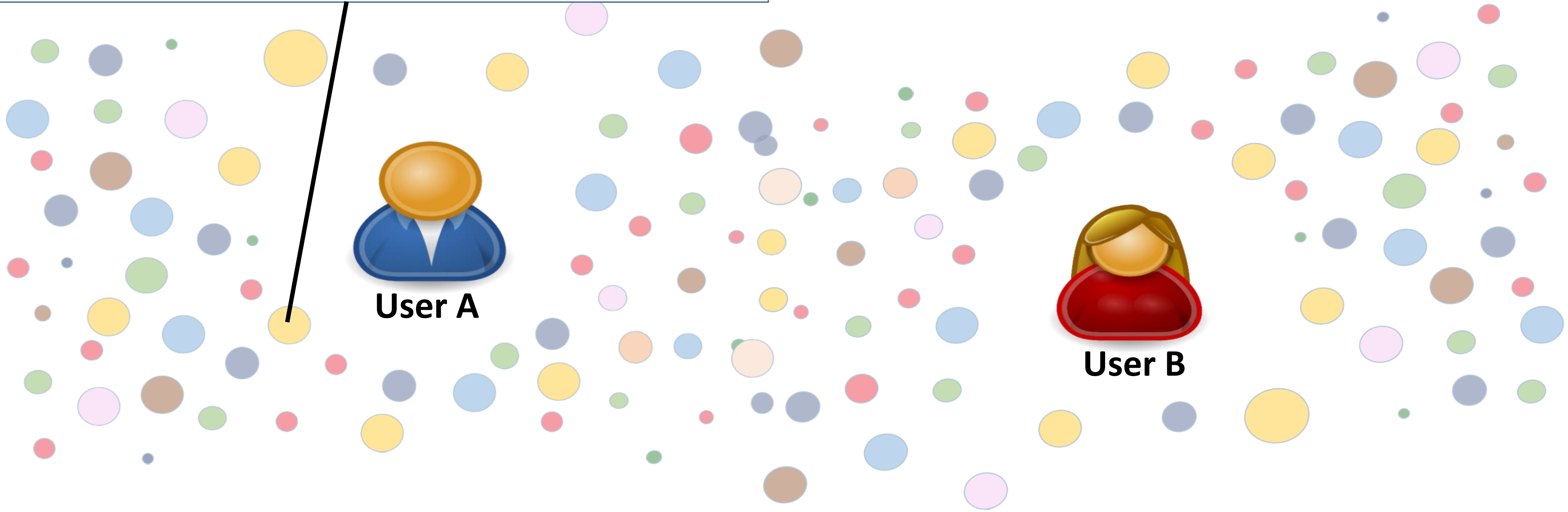
Topic: Presidential Debate

# Filter bubble …


Clinton puts Trump on defense at first debate

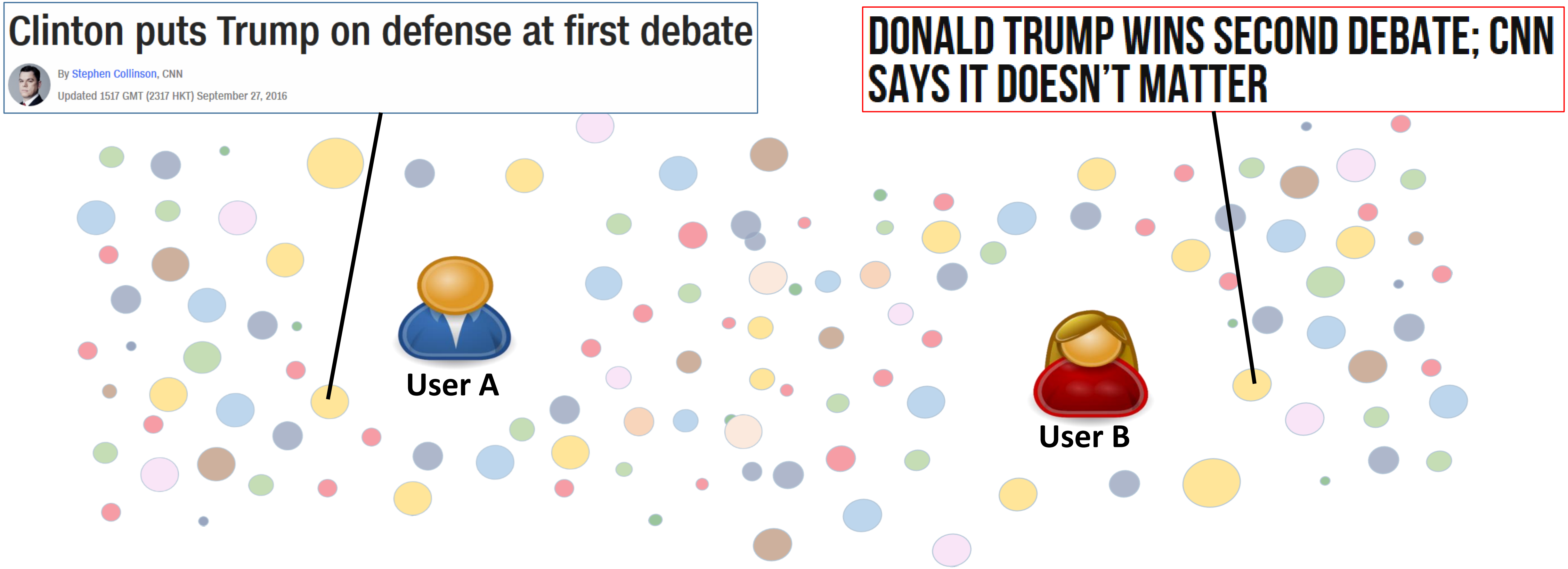
By Stephen Collinson, CNN

Updated 1517 GMT (2317 HKT) September 27, 2016




Topic: Presidential Debate

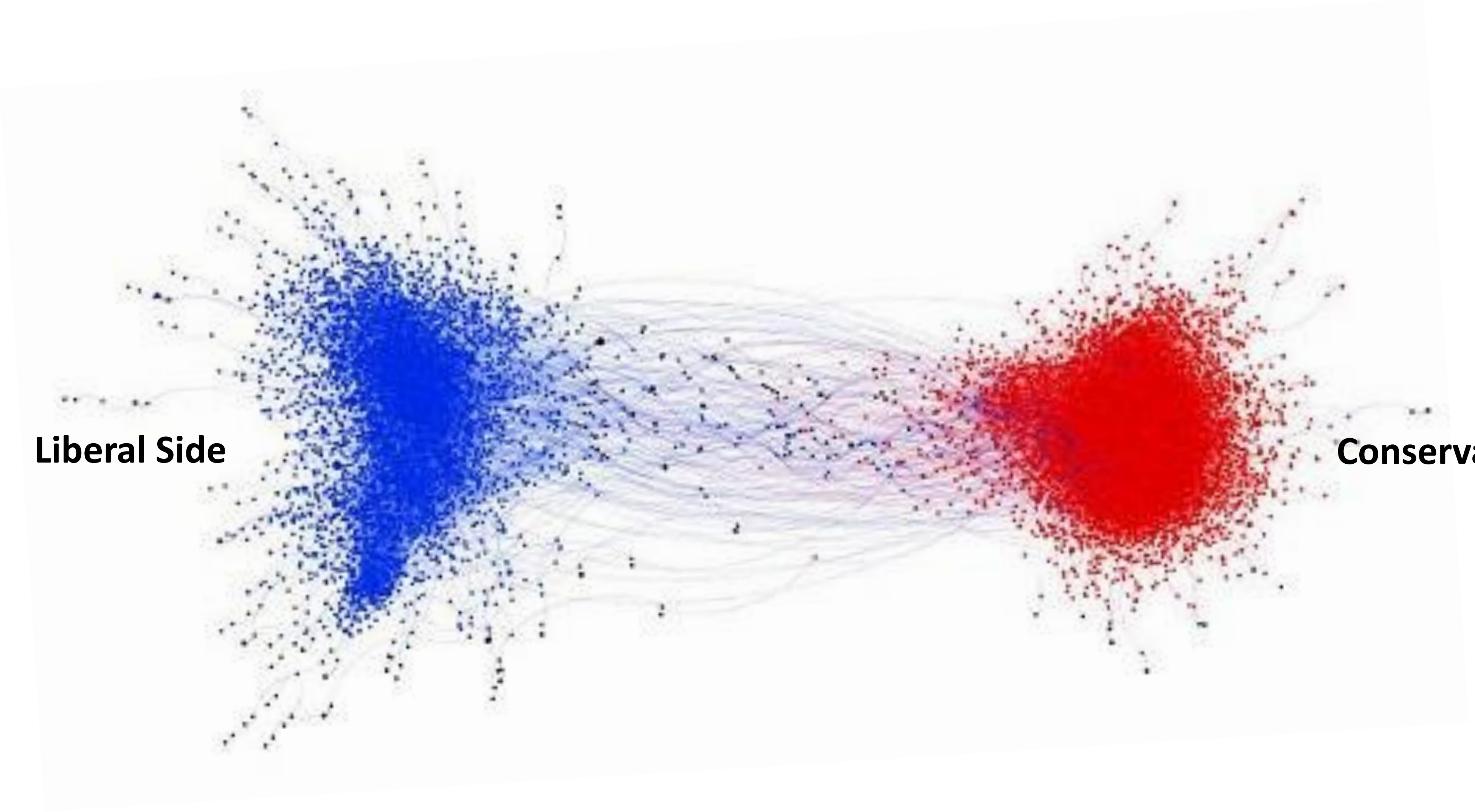
# Filter bubble…



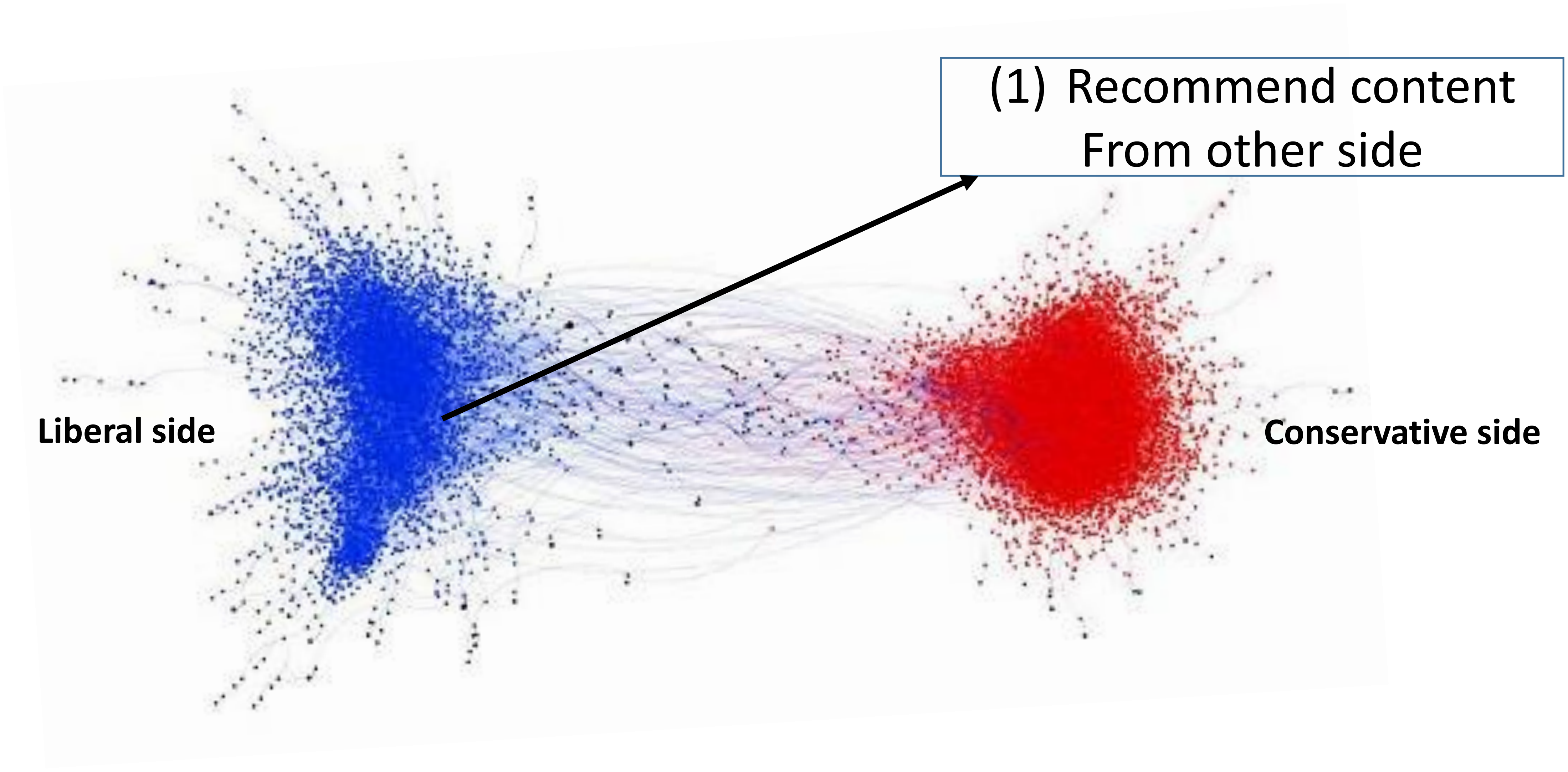

Topic: Presidential Debate

# Motivation

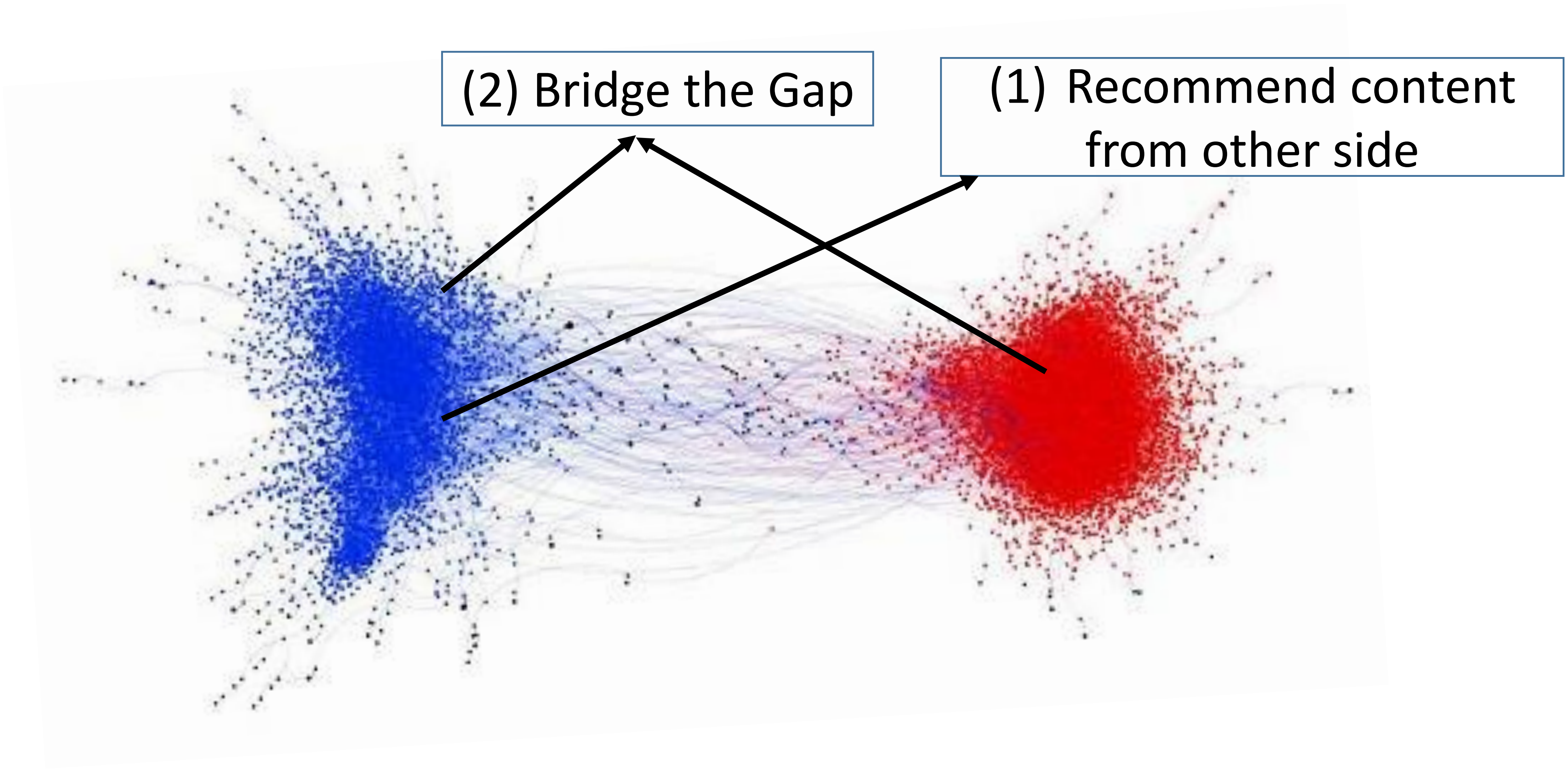
# Ideological segregation, polarization, biased views



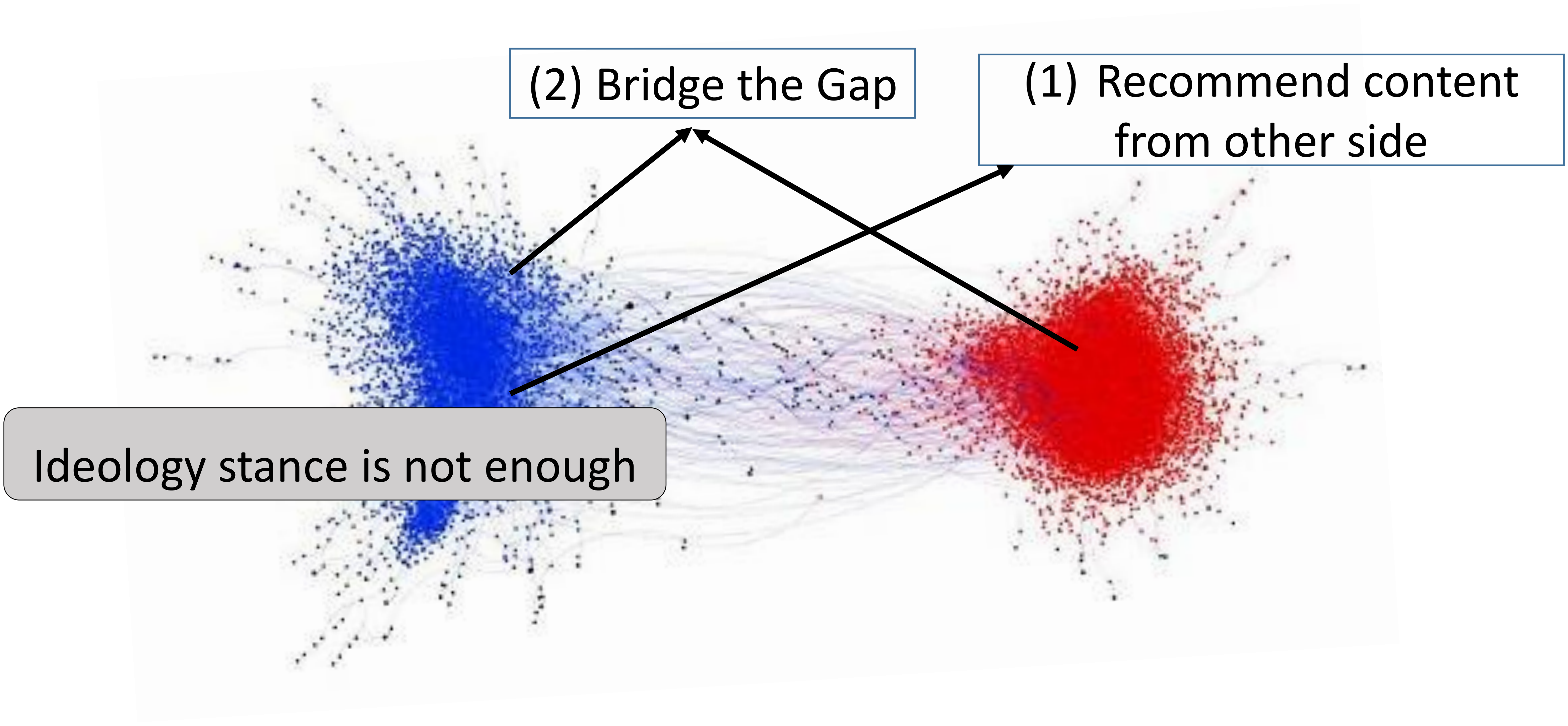
# Bursting the filter bubble…



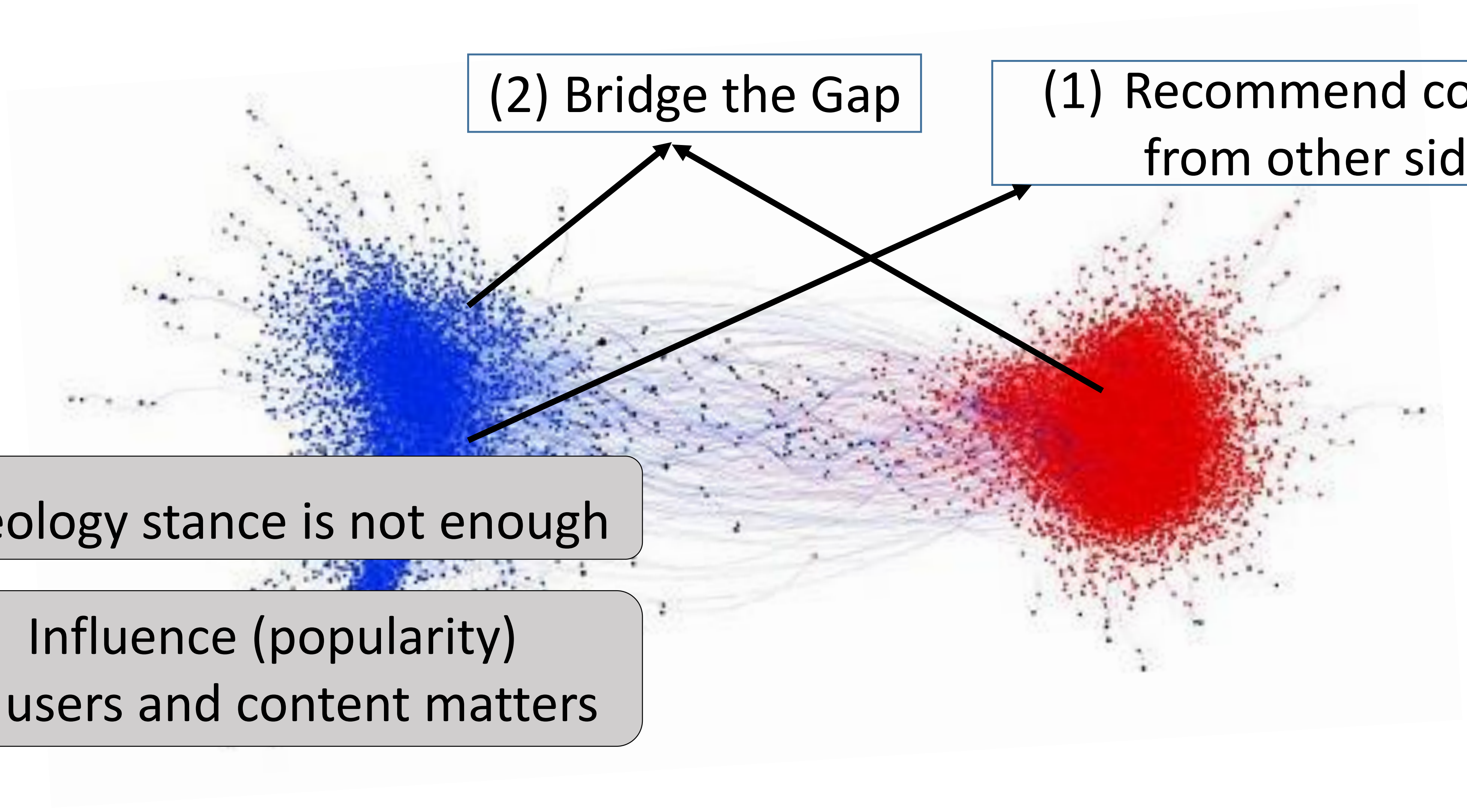
# Bursting the filter bubble…



# Bursting the filter bubble…



# Bursting the filter bubble…




Ideology stance is not enough

Influence (popularity) of users and content matters

# Problem Statement

- Input:

social graph
(A)

content graph
(C)

# Problem Statement

- Input:

social graph
(A)

content graph
(C)

- Learn the shared latent space between A and C

# Problem Statement

- Input:

social graph
(A)

content graph
(C)

- Learn the shared latent space between A and C
- Discover ideology-popularity latent dimensions

# Problem Statement

- Input:

social graph
(A)

content graph
(C)

- Learn the shared latent space between A and C
- Discover ideology-popularity latent dimensions
- Estimate ideology and popularity scores for users and content

# Proposed Methodology

# Orthogonal Non-negative Matrix Factorization as Co-Clustering Model [Ding et al]

$$X \approx [U][H][V^T]$$

clustering of columns of X

Input matrix (e.g., user – item)

clustering of rows of X

association between row and column clusters

# Combining Link and Content

$$J = \|A - UH_uU^T\|_F^2 + \|C - UH_sV^T\|_F^2$$

user-user matrix

user-content matrix

**Shared Latent Space:**

- A and C are related via users
- row datatype of matrix A is the same as rows of C

# Combining Link and Content

$$J = \|A - UH_uU^T\|_F^2 + \|C - UH_sV^T\|_F^2$$

Joint Matrix factorization

user-user matrix

user-content matrix

**Shared Latent Space:**

- A and C are related via users
- row datatype of matrix A is the same as rows of C

# Learning Shared Latent Space

$$J = \|A - UH_uU^T\|_F^2 + \|C - UH_sV^T\|_F^2$$

Joint Matrix factorization

user-user matrix

user-content matrix

Shared latent factors (U and V)

## Shared Latent Space:

- A and C are related via users
- row datatype of matrix A is the same as rows of C

# Learning Hidden Manifolds in The Data

$$J = \|A - UH_uU^T\|_F^2 + \|C - UH_sV^T\|_F^2 + \boxed{\lambda\, Tr(U^T L_u U) + \lambda\, Tr(V^T L_s V)}$$

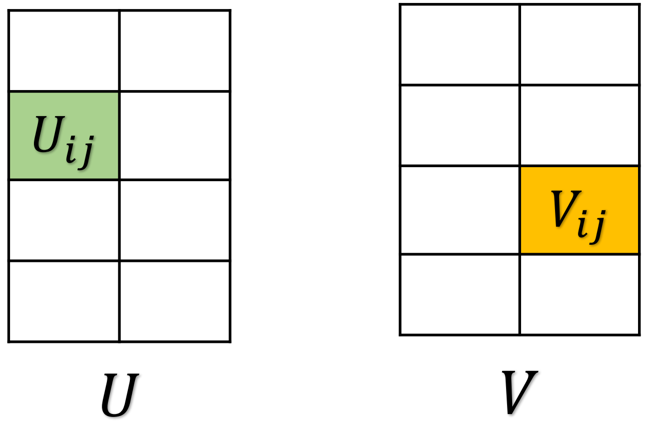
User Manifold and content Manifold are tied together

- Users connected in social graph tend to be ideologically similar
- Ideologically similar users share similar content (and ideologically similar content is shared by similar users)

# Latent factors have a probabilistic interpretation

$$J = \|A - UH_uU^T\|_F^2 + \|C - UH_sV^T\|_F^2$$

Latent factors U and V

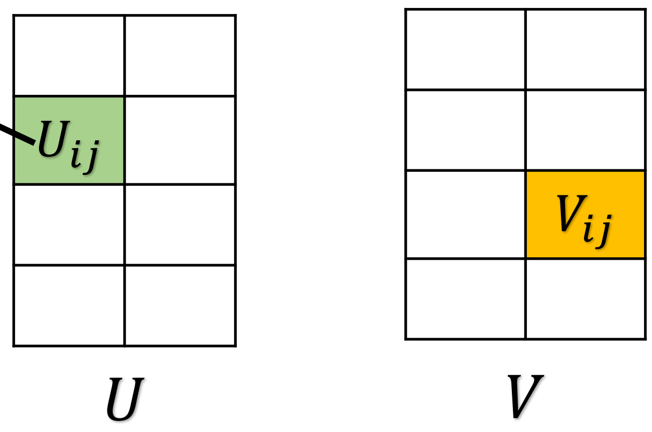


# Latent factors have a probabilistic interpretation

$$J = \|A - UH_uU^T\|_F^2 + \|C - UH_sV^T\|_F^2$$

Latent factors U and V

degree to which user i belongs to ideology j

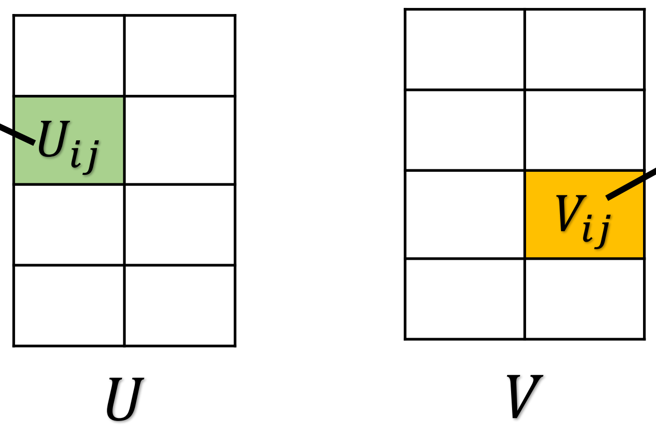


# Latent factors have a probabilistic interpretation

$$J = \|A - UH_uU^T\|_F^2 + \|C - UH_sV^T\|_F^2$$

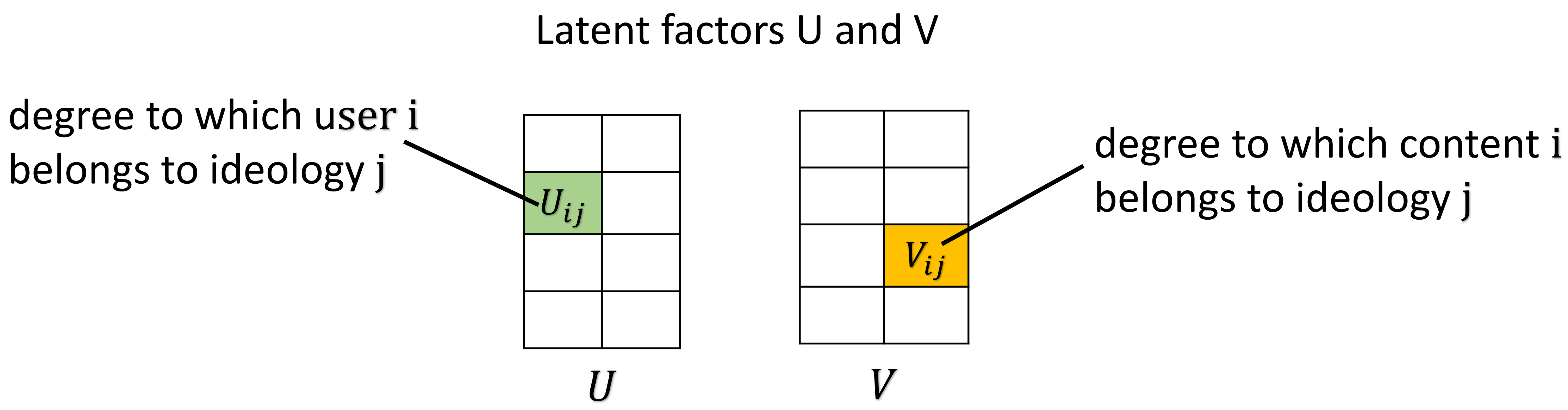
Latent factors U and V

degree to which user i belongs to ideology j



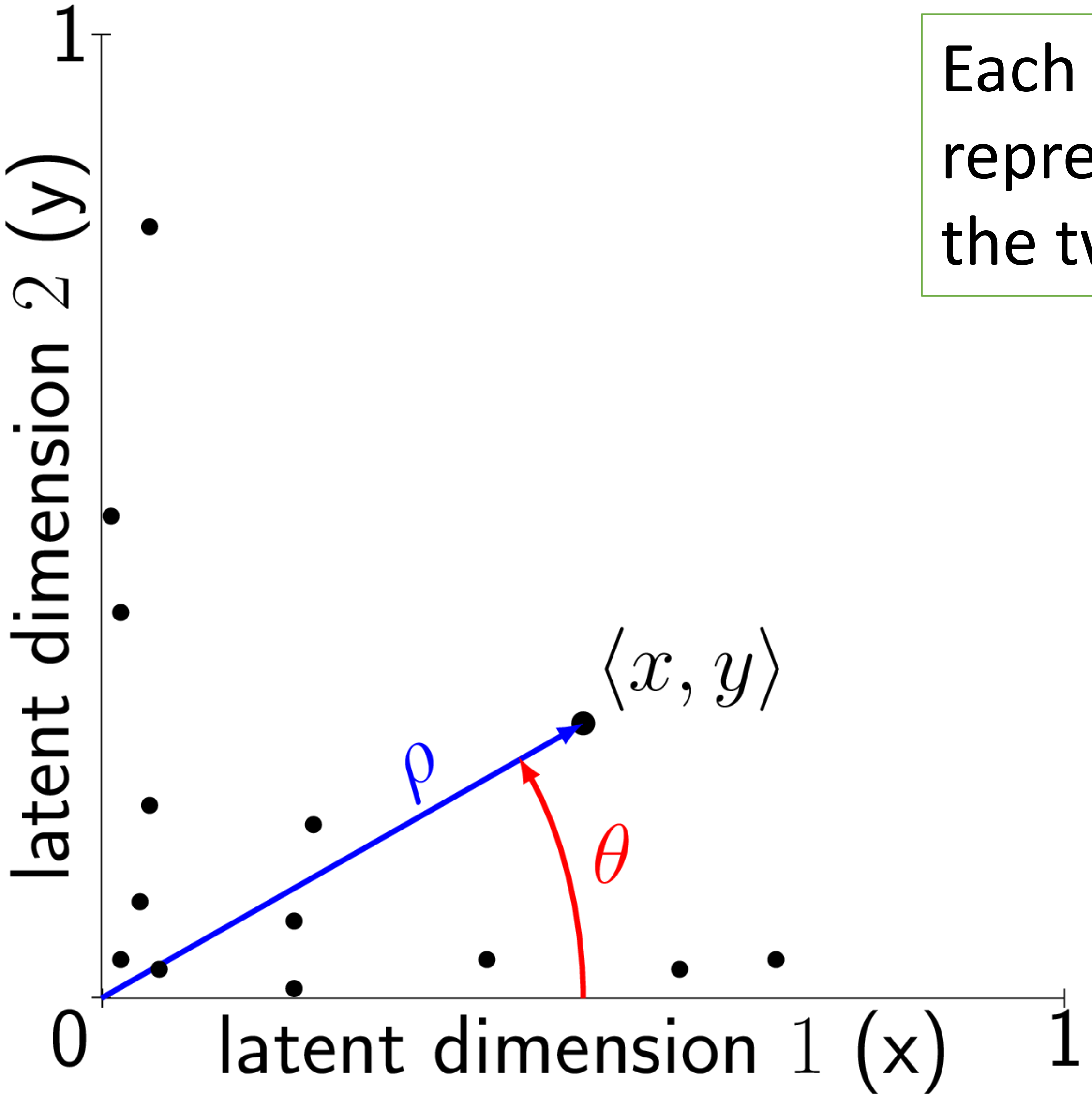

degree to which content i belongs to ideology j

# Latent factors have a probabilistic interpretation




Each row of $U$ and $V$ can be represented as a two dimensional vector $\langle x, y \rangle$

# Estimating Ideological Leaning




Each row of $U$ and $\mathrm{V}$ can be represented as a vector $\langle x, y \rangle$ in the two dimensional space

# Evaluation

# Dataset

Twitter Streaming API (2011 - 2016)

- 7000 users
- 19 million tweets

Three controversial topics

- Gun control
- Abortion
- Obamacare

Ground Truth

- 500 news media channels
- Bayesian point estimate using large annotated data set [Barbera et al]

# Baselines

| | |
|---|---|
| Link (user-user) | • Graph partitioning (Retweet/Follows)<br>• NMF [Lee & Seung] |
| Content (user-content) | • ONMTF [Ding et al]<br>• DMCC [Gu et al] |
| Combined (link+ content) | • IFD / IFD-NGR [Proposed method]<br>• BIAS WATCH [Lu et al]<br>• KULSHRESTHA [Kulshrestha et al] |

# Baselines

No ideology scores

**Link (user-user)**

- Graph partitioning (Retweet/Follows)
- NMF [Lee & Seung]

**Content (user-content)**

- ONMTF [Ding et al]
- DMCC [Gu et al]

**Combined (link+ content)**

- IFD / IFD-NGR [Proposed method]
- BIAS WATCH [Lu et al]
- KULSHRESTHA [Kulshrestha et al]

# Baselines

**Link (user-user)**

- Graph partitioning (Retweet/Follows)
- NMF [Lee & Seung]

**Content (user-content)**

- ONMTF [Ding et al]
- DMCC [Gu et al]

**Combined (link+ content)**

- IFD / IFD-NGR [Proposed method]
- BIAS WATCH [Lu et al]
- KULSHRESTHA [Kulshrestha et al]

No ideology scores

Matrix Factorization based approaches

# Baselines

**Link**
(user-user)

- Graph partitioning (Retweet/Follows)
- NMF [Lee & Seung]

**Content**
(user-content)

- ONMTF [Ding et al]
- DMCC [Gu et al]

**Combined**
(link+ content)

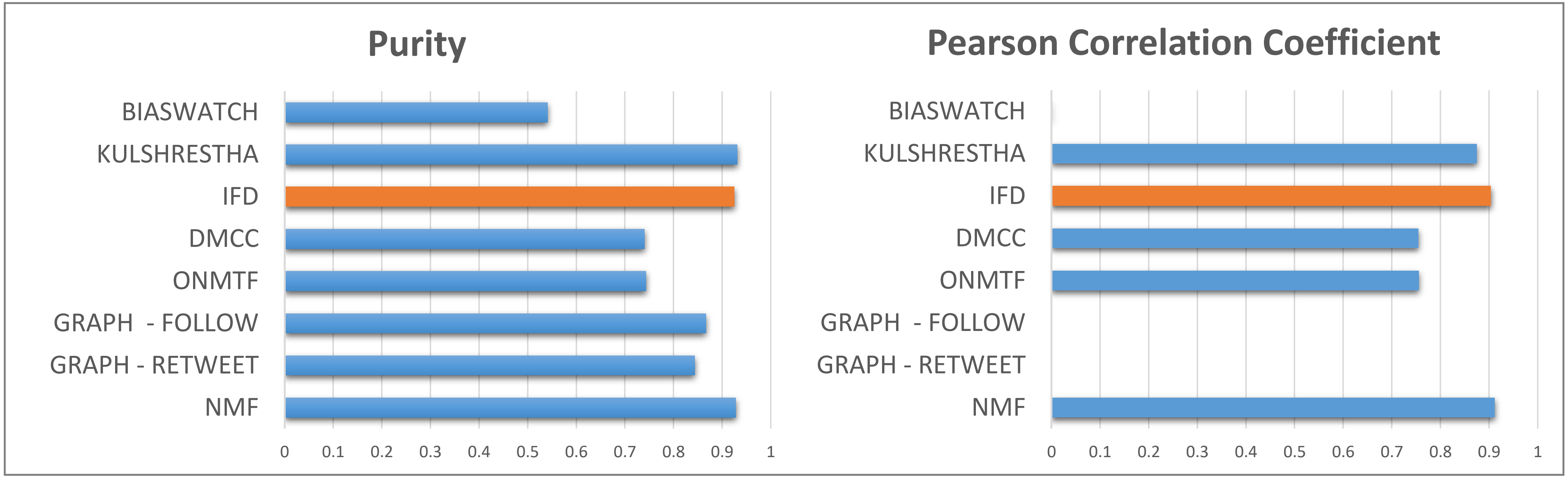
- IFD / IFD-NGR [Proposed method]
- BIAS WATCH [Lu et al]
- KULSHRESTHA [Kulshrestha et al]

No ideology scores

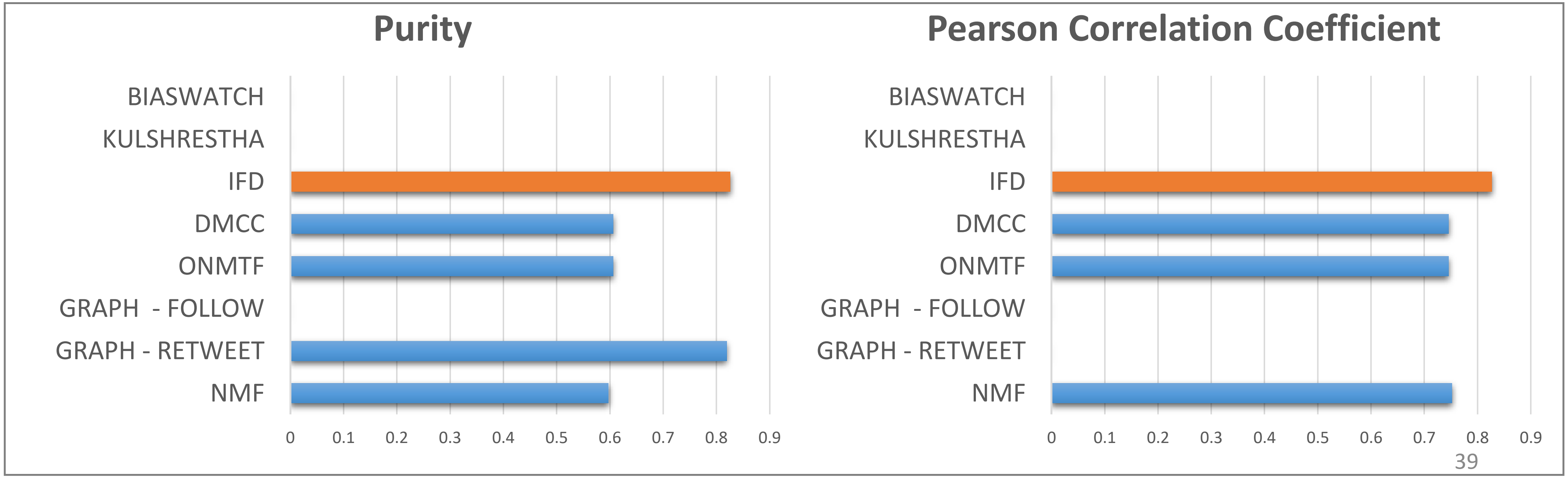
Matrix Factorization based approaches

No ideology scores for media channels

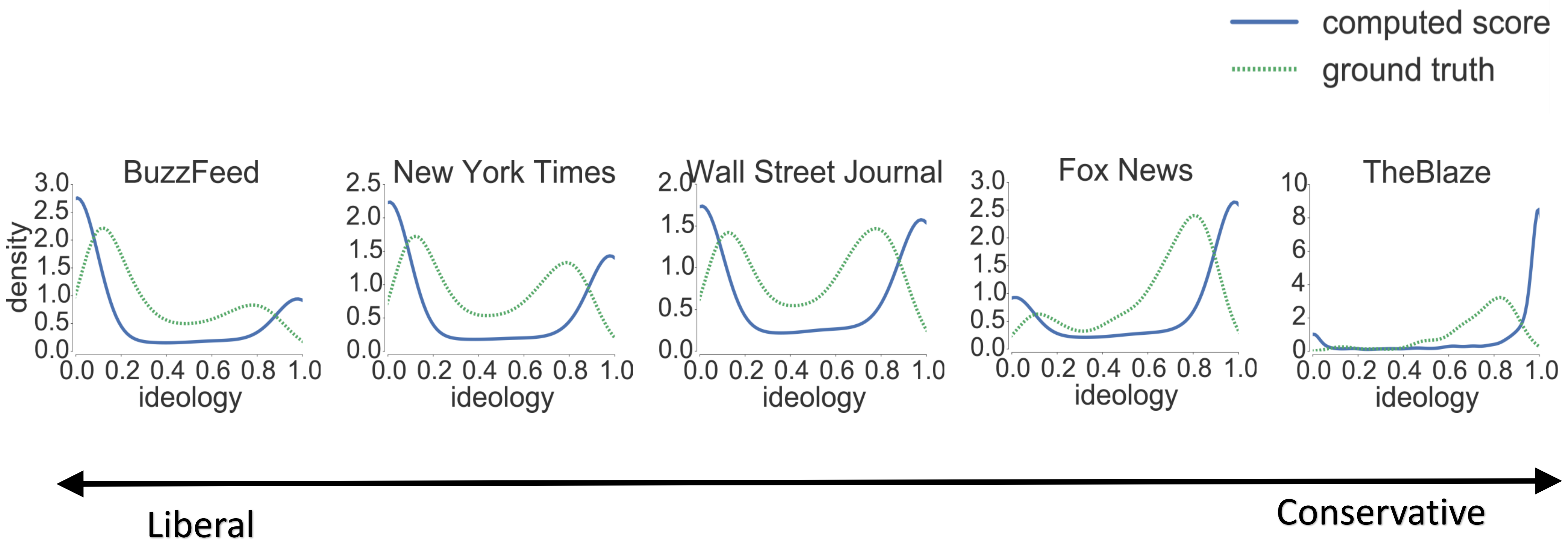
Twitter Users (Ideology)



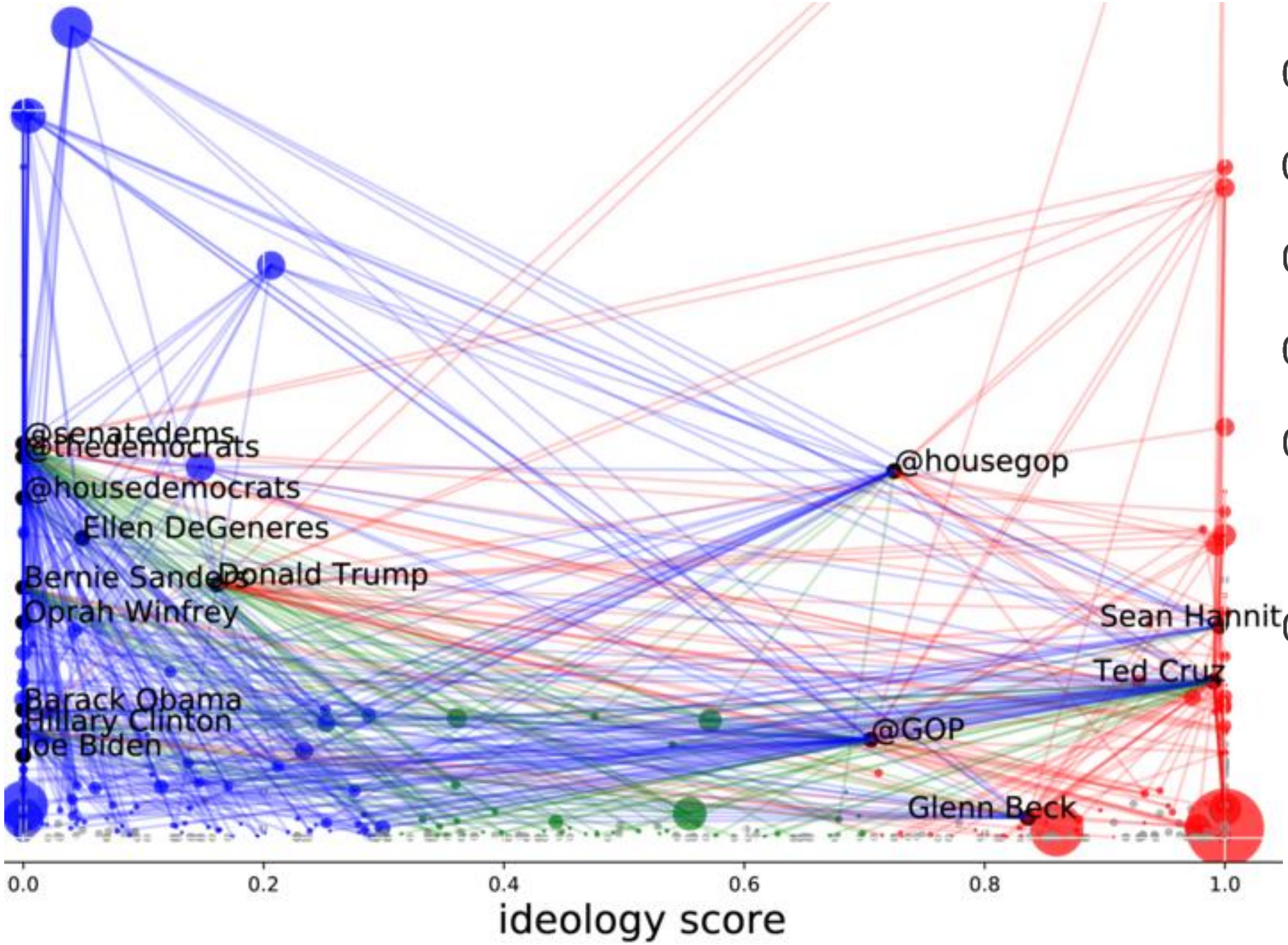

News Media (Ideology)

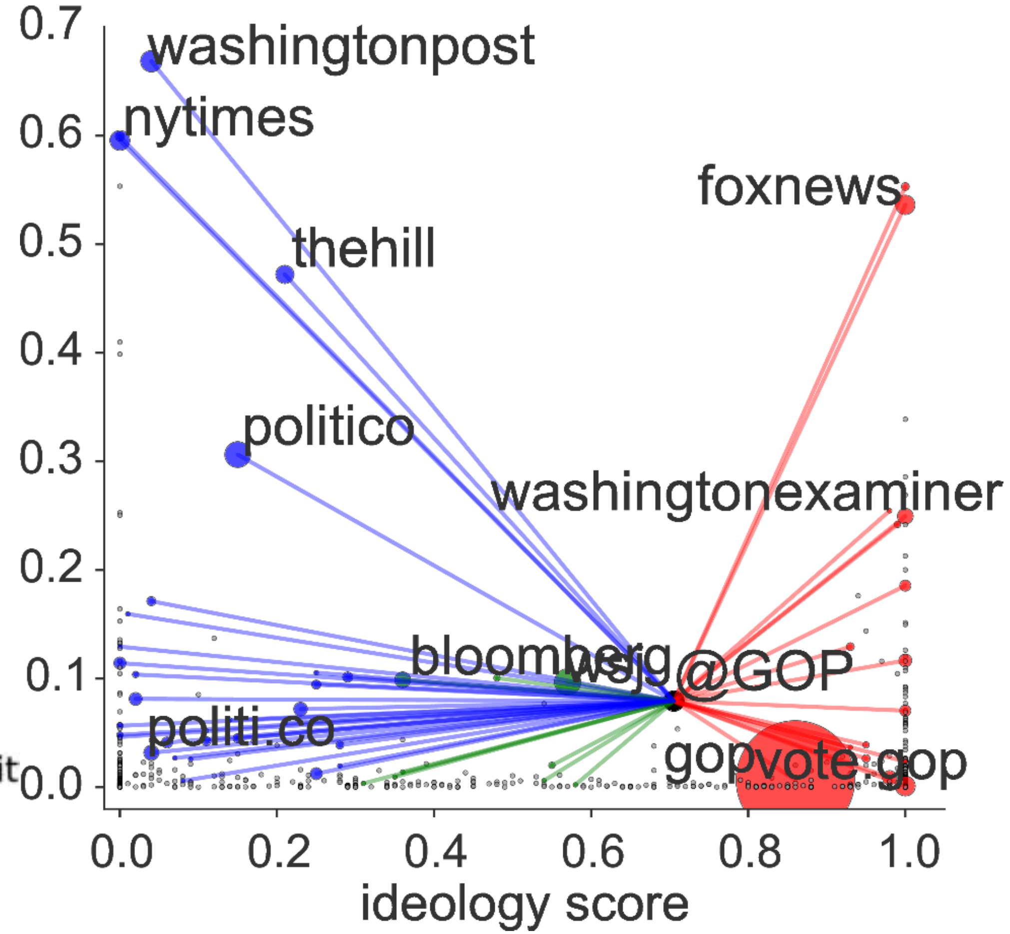


# Estimated Ideology scores of high quality across the ideology spectrum (including center)



# Ideological Latent Space

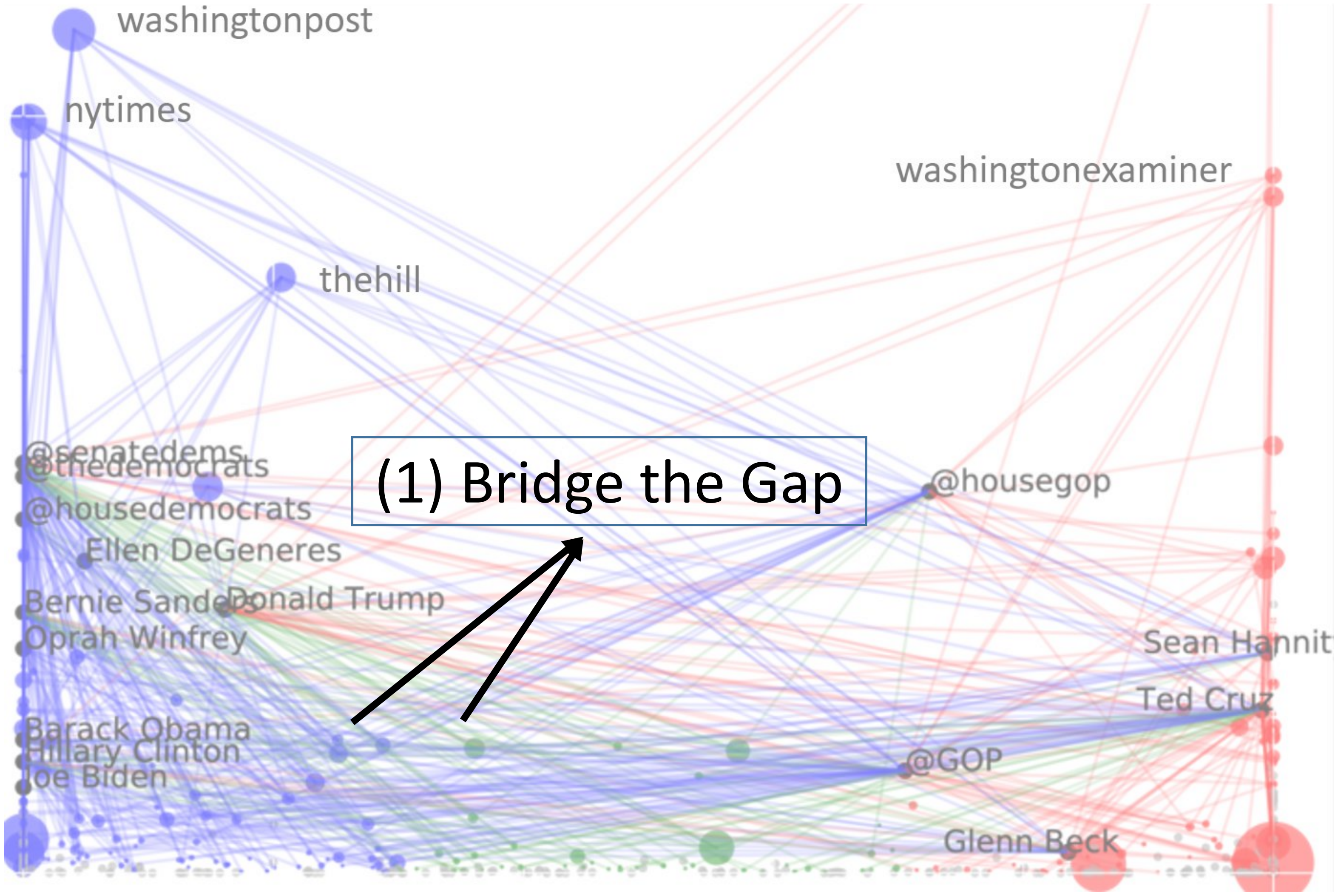


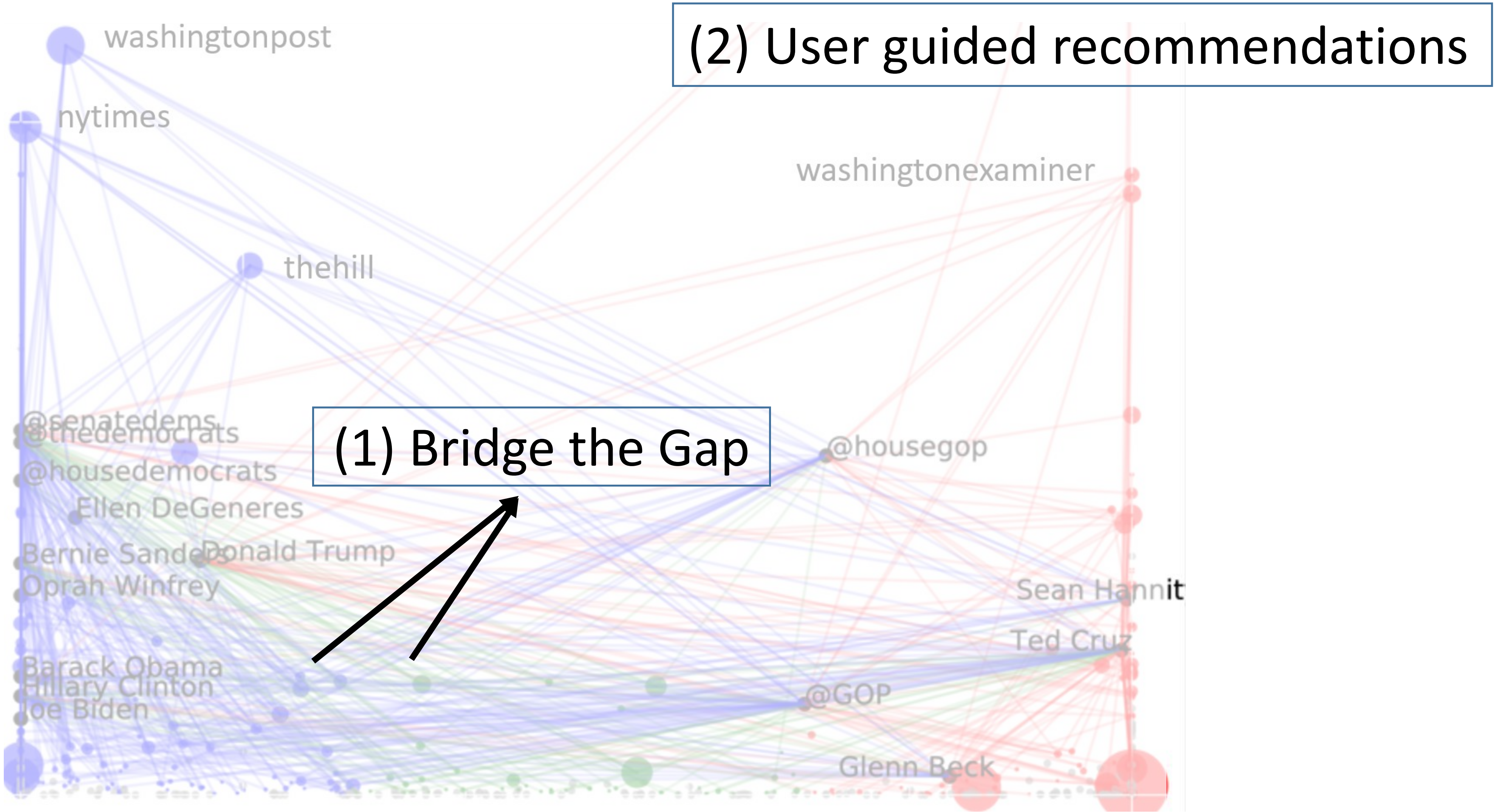




More Interactive Visualizations at http://bit.ly/FilterBubbleDemo

# Motivation (Revisiting)

# We now have access to ideological position of all the users and content

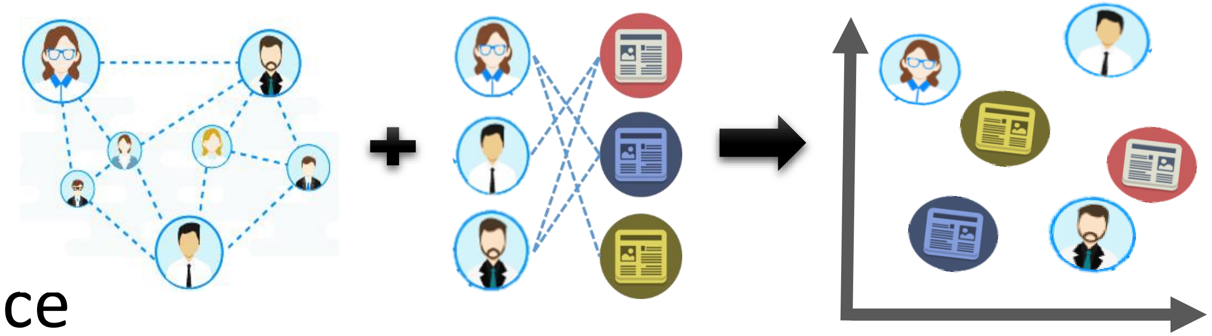



(2) User guided recommendations
washingtonpost
nytimes
washingtonexaminer
thehill
@senatedems
@thedemocrats
@housedemocrats
Ellen DeGeneres
Bernie Sanders
Donald Trump
Oprah Winfrey
Barack Obama
Hillary Clinton
Joe Biden
(1) Bridge the Gap
@housegop
Sean Hannit
Ted Cruz
@GOP
Glenn Beck

# Summary

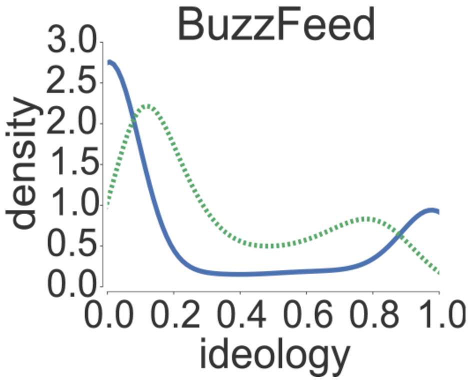
IFD framework

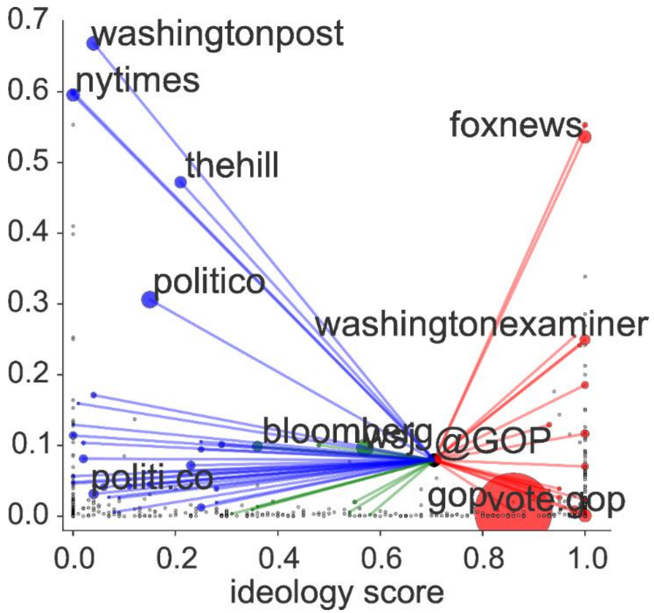
- Combines Link and Content Graph
- Jointly Matrix Factorization
- Shared Multidimensional Latent Space



Compared to Baselines IFD Estimates

- Ideology and Popularity Scores
- Twitter Users and Media Channels
- High Quality



# THANK YOU

# Learning Hidden Manifolds in The Data

Manifold assumption

If two data points $x_i, x_j$ are close to each other in the input space
then their projections in the new basis $u_i, u_j$ are also close.

$$J = \|A - UH_uU^T\|_F^2 + \|C - UH_sV^T\|_F^2 + \underbrace{\boxed{\lambda\, Tr(U^TL_uU) + \lambda\, Tr(V^TL_sV)}}_{\text{Graph regularization constraints [Cai et al]}}$$

where

- $L_u, L_s$ are graph laplacians of the row and column affinity matrices of X
- $Tr(\cdot)$ is trace of the matrix

# Optimization Problem

## IFD (Ideology Factor Decomposition)

$$J = \|A - UH_uU^T\|_F^2 + \|C - UH_sV^T\|_F^2 + \lambda\, Tr(U^T L_u U) + \lambda\, Tr(V^T L_s V)$$

## Constraints

- Bi-orthogonality ($U^T U = I$ ; $V^T V = I$ )
- Non-negativity ($U_+$ ; ${H_u}_+$ ; ${H_s}_+$; $V_+$)

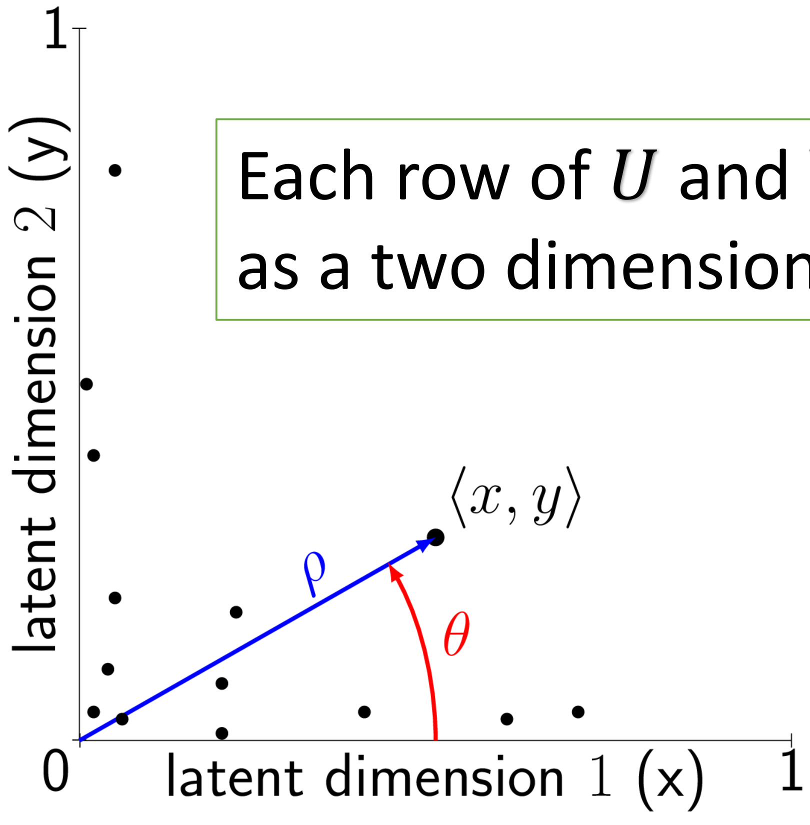
## Solution

- Derive multiplicative update rules for $U, V, H_u$ and $H_s$
- Iterative update algorithm
- Locally optimal solution

# Estimating Ideological Leaning

The latent factors U and V have a probabilistic interpretation:

- $U_{ij}$ : degree to which user i belongs to ideology j
- $V_{ij}$ : degree to which content i belongs to ideology j




Each row of $U$ and V can be represented as a two dimensional vector $\langle x, y \rangle$

Ideology: $i(x, y) = \frac{\theta}{\pi/2} = \frac{\arctan\left(\frac{y}{x}\right)}{\pi/2}$

Popularity: $\rho(x, y) = \sqrt{x^2 + y^2}$

# Estimating Ideological Leaning