

# Purple Feed: Identifying High Consensus News Posts on Social Media

Juhi Kulshrestha

*with*

Mahmoudreza Babaei, Abhijnan Chakraborty,  
Fabricio Benevenuto, Krishna P. Gummadi, Adrian Weller



MAX PLANCK INSTITUTE  
FOR SOFTWARE SYSTEMS



U F *m* G



The  
Alan Turing  
Institute

# News on social media



salon.com

AP



POLITICO

BREITBART

B

Slate

CNN



REUTERS



*DRUDGE REPORT*

The New York Times

# News on social media

Slate



salon.com



The New York Times

AP



POLITICO

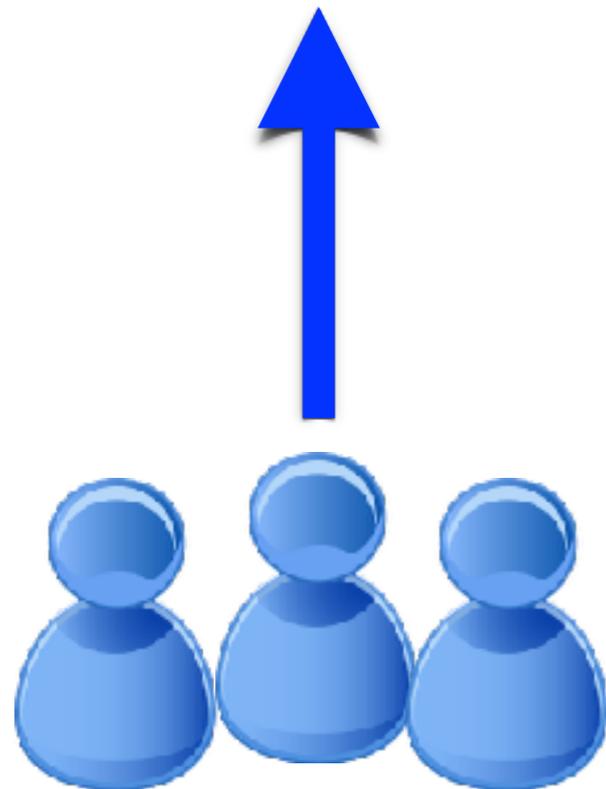
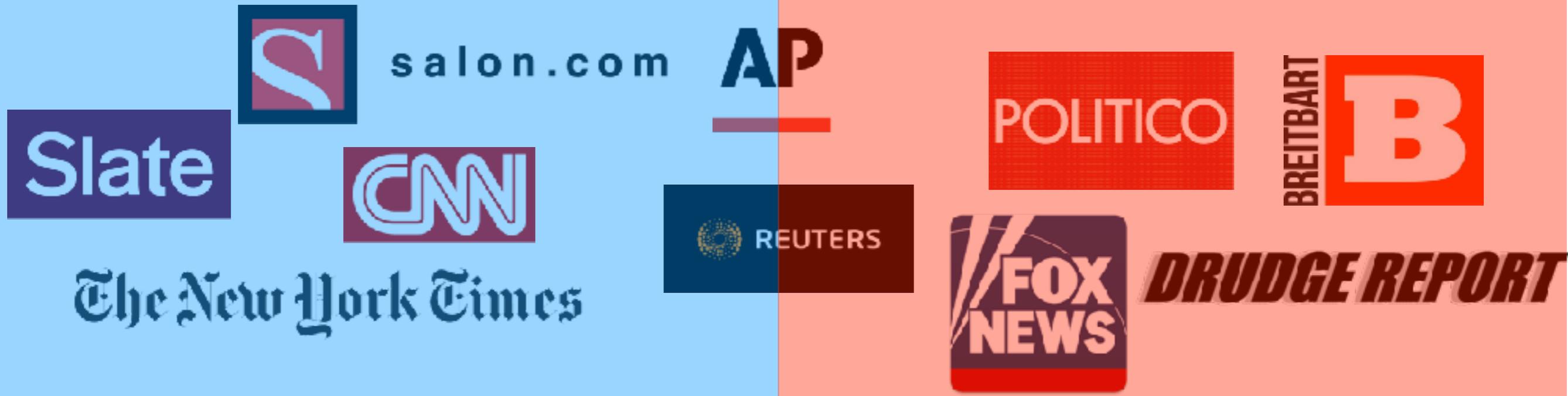


BREITBART



*DRUDGE REPORT*

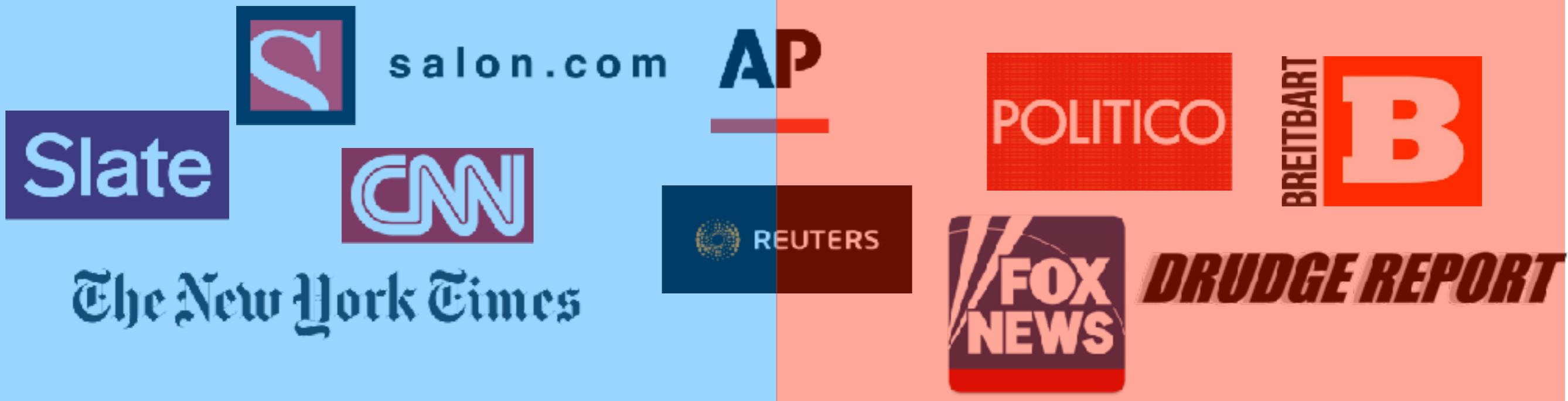
# News on social media



**Selective exposure**



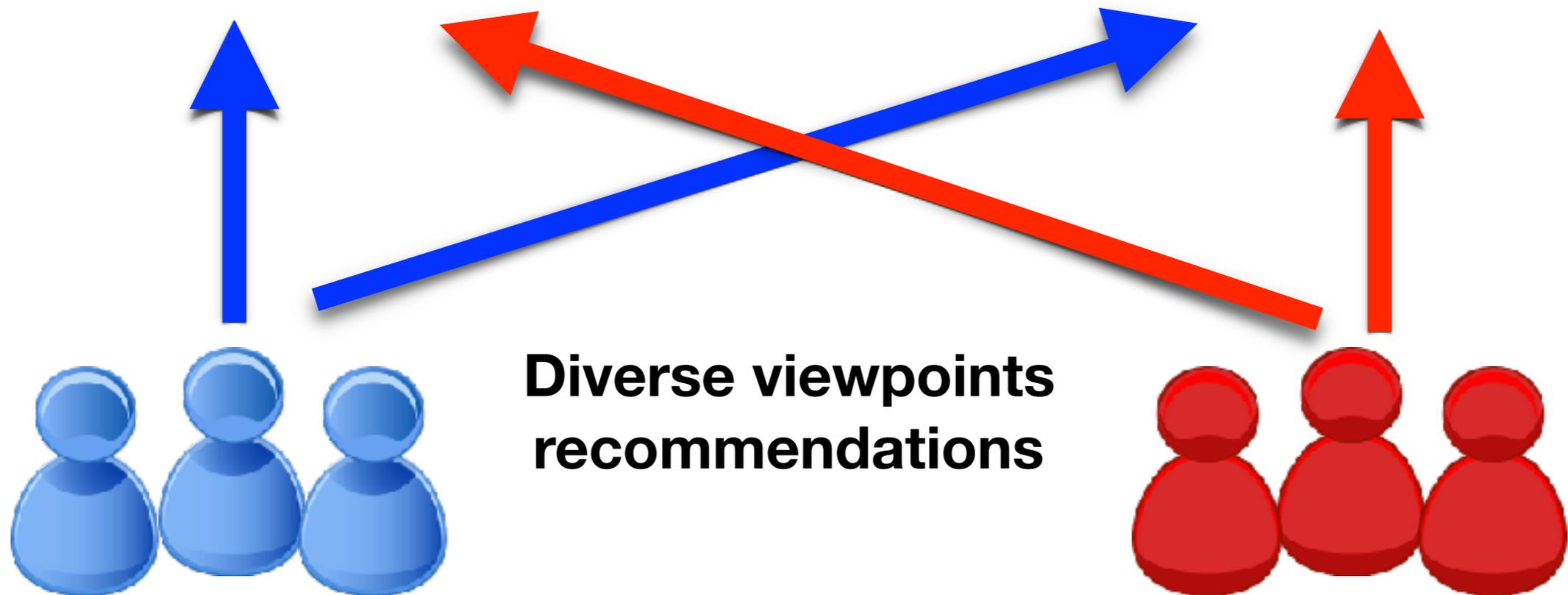
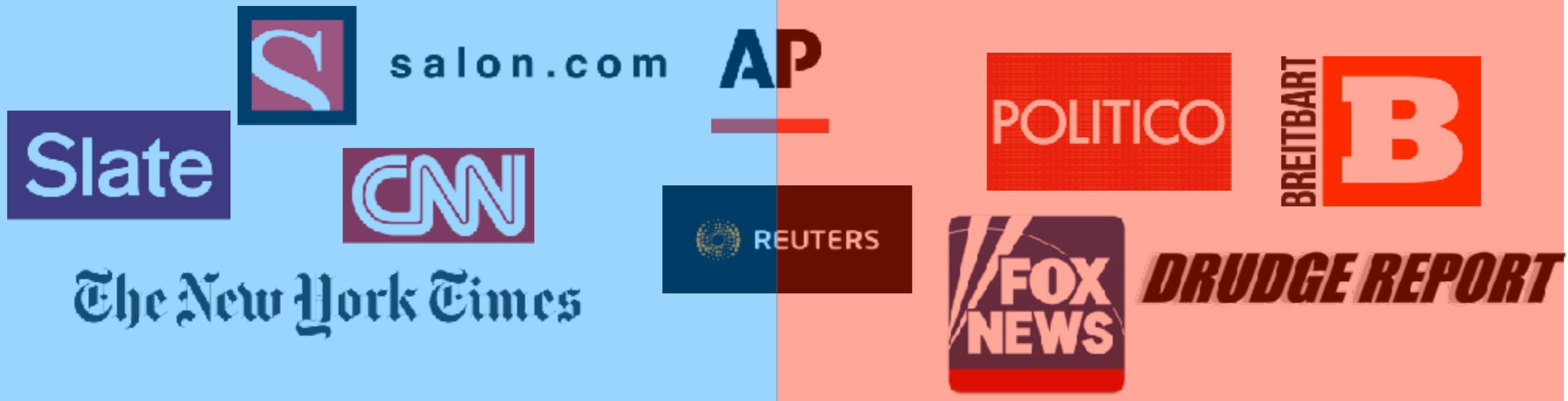
# News on social media



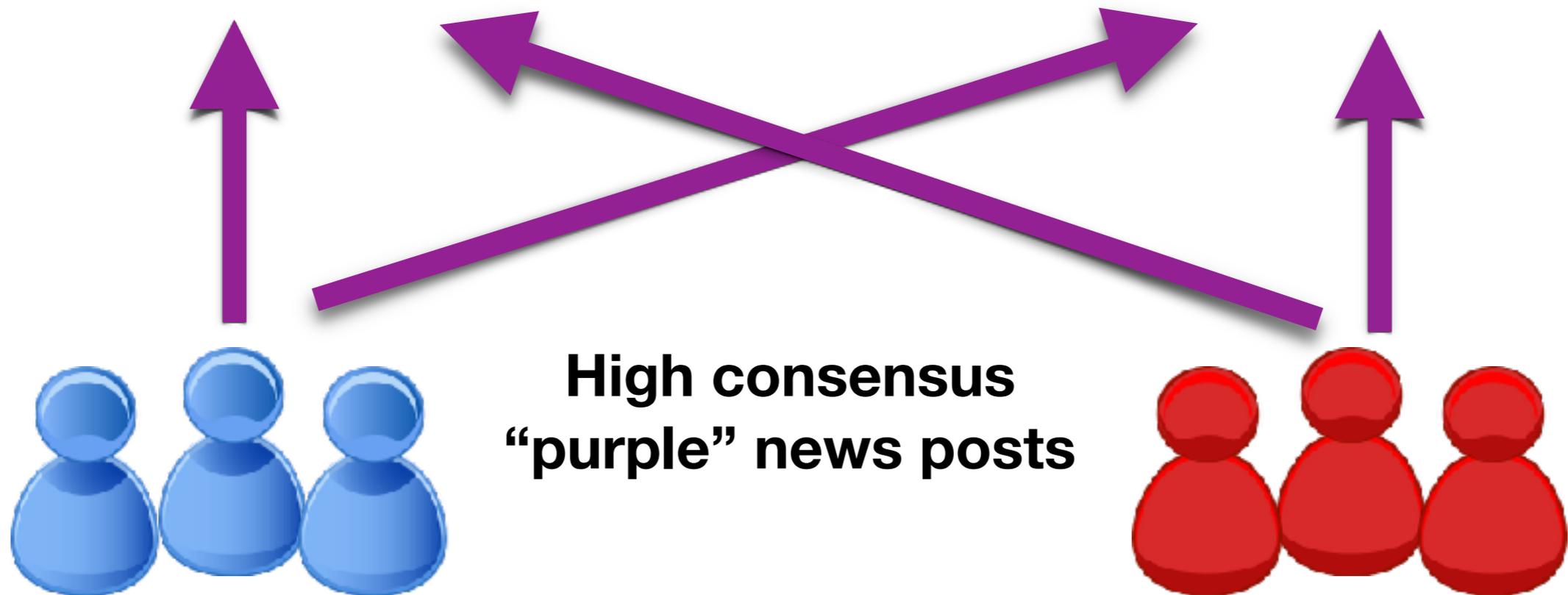
**Selective exposure  
+  
Recommendations**



# News on social media



# News on social media



# High consensus “purple” posts

- A general agreement in readers’ reaction to it, irrespective of their own political leaning

# High consensus “purple” posts

- A general agreement in readers’ reaction to it, irrespective of their own political leaning

$$\textit{Consensus} = 1 - \left| \frac{\#D_{\textit{disagree}}}{\#D} - \frac{\#R_{\textit{disagree}}}{\#R} \right|$$

# High consensus “purple” posts

- A general agreement in readers’ reaction to it, irrespective of their own political leaning

$$\textit{Consensus} = 1 - \left| \frac{\#D_{disagree}}{\#D} - \frac{\#R_{disagree}}{\#R} \right|$$

- AMT experiment to measure consensus of news tweets - Ground truth dataset

# High consensus “purple” posts

- A general agreement in readers’ reaction to it, irrespective of their own political leaning

**Salon: Comey firing coverage shows right-wing media has lost it’s grip on reality [url]**

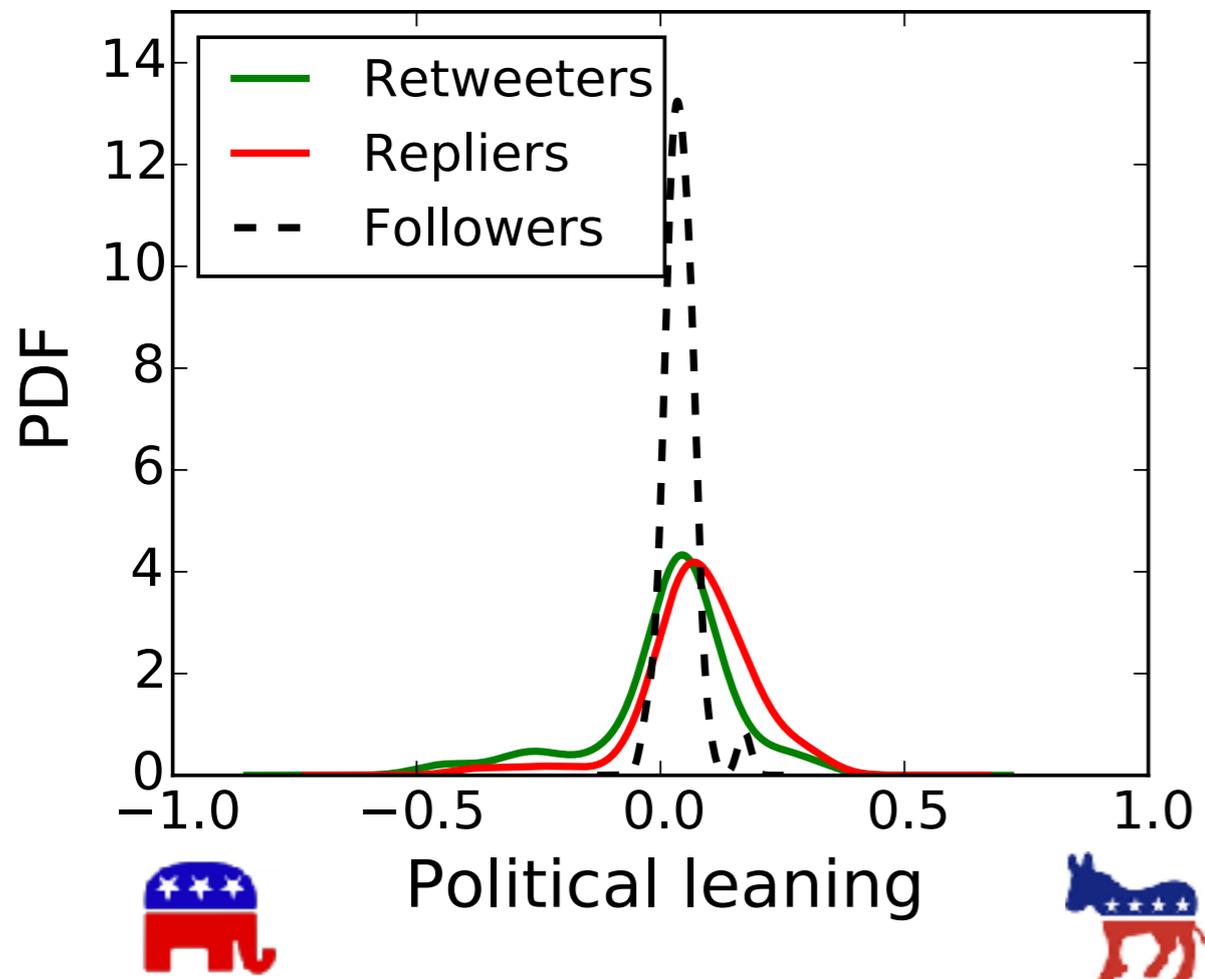
**FoxNews: @seanhannity: “The real reason that President #Trump fired James Comey is because the former @FBI Director was incompetent.” [url]**

**NYTimes: He was fired by President Trump. Where does James Comey go next? [url]**

# Identifying high & low consensus posts

- Prior work - publisher- and content-based features
- Our proposal - audience leaning based features

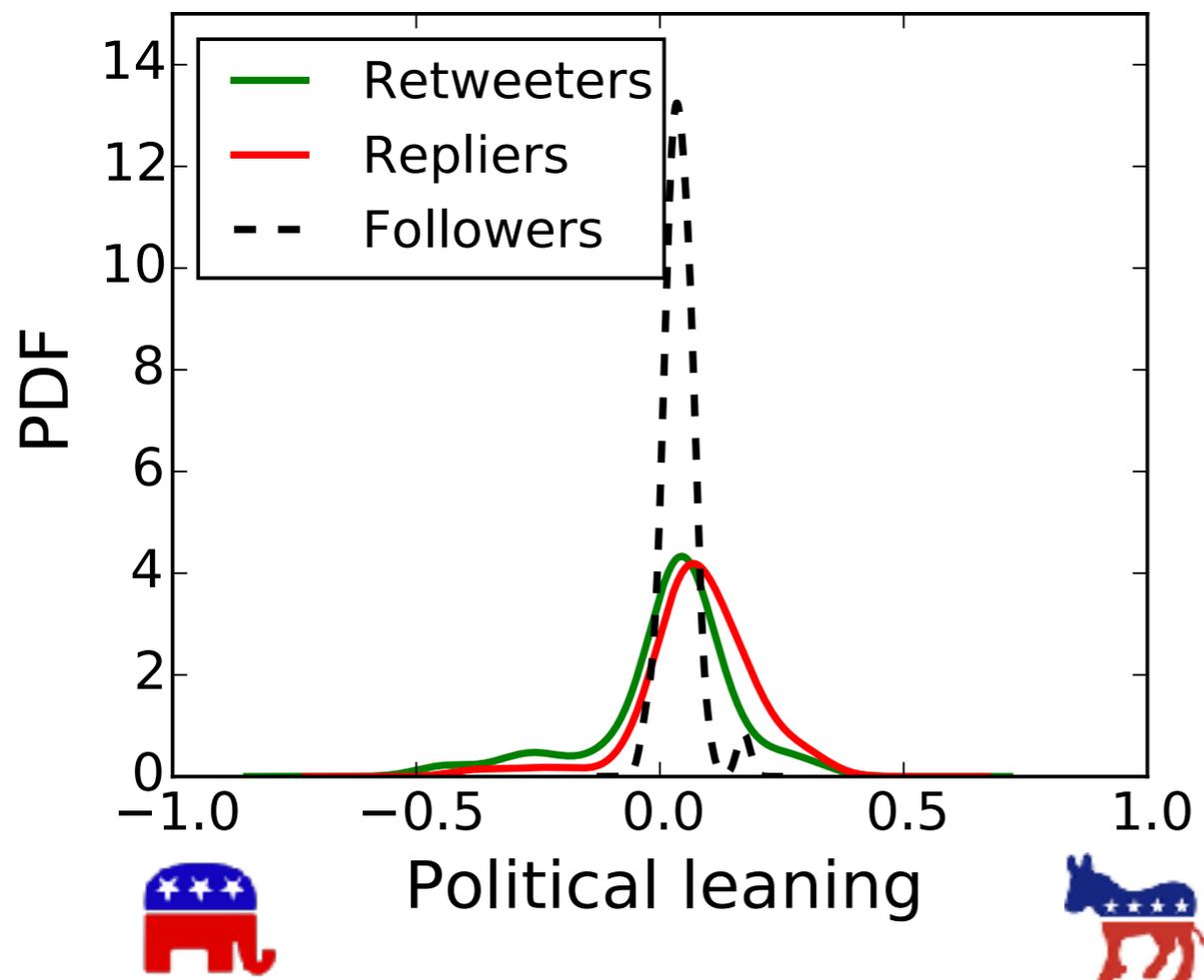
# Audience leaning based features - An intuition



***Reuters: Trump ordered emergency meeting after global cyber attack: official [url]***

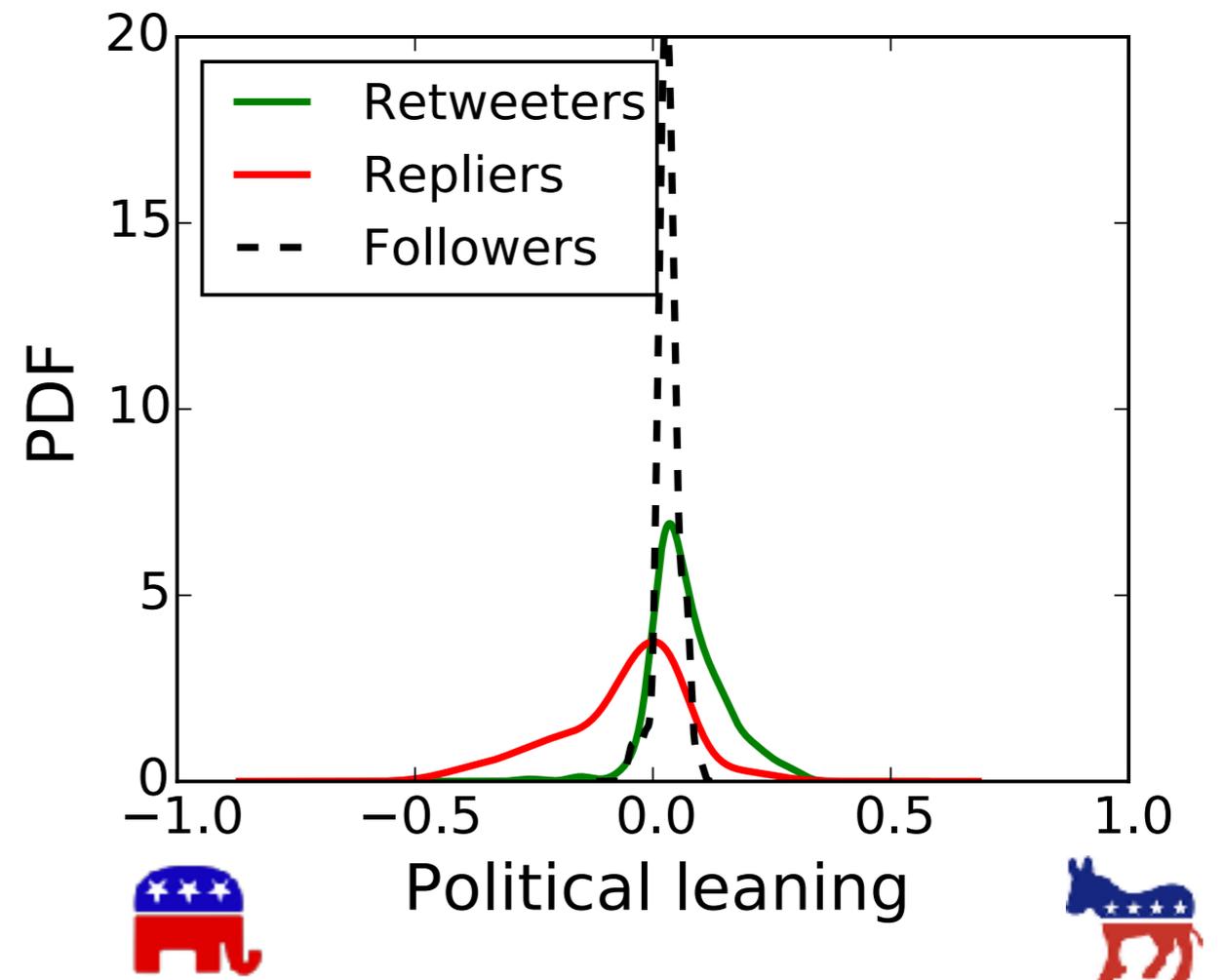
**High consensus post**

# Audience leaning based features - An intuition



***Reuters: Trump ordered emergency meeting after global cyber attack: official [url]***

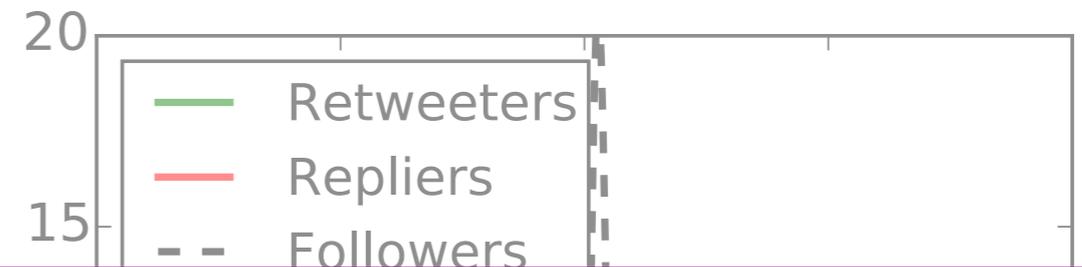
**High consensus post**



***CNN: Michelle Obama criticizes Trump administration's school lunch policy [url]***

**Low consensus post**

# Audience leaning based features - An intuition



PDF

**Audience leaning based features outperform prior publisher and content based features.**

*Reuters: Trump ordered emergency meeting after global cyber attack: official [url]*

**High consensus post**

*CNN: Michelle Obama criticizes Trump administration's school lunch policy [url]*

**Low consensus post**

# Purple Feed: Identifying High Consensus News Posts on Social Media

- Measuring consensus of news posts on social media (AMT survey for creating ground truth dataset)
- Empirical study of consensus of news posts (popularity, topics covered, exposure to ideologically cross-cutting content)
- Audience leaning based features & their experimental evaluation

# Purple Feed: Identifying High Consensus News Posts on Social Media

- Measuring consensus of news posts on social media (AMT survey for creating ground truth dataset)
- Empirical study of consensus of news posts (popularity, topics covered, exposure to ideologically cross-cutting content)
- Audience leaning based features & their experimental evaluation

End

# Empirical study of consensus of news posts

- High & low consensus posts are equally popular
- High & low consensus posts cover similar topics
- High consensus posts provide more cross-cutting exposure to views than low consensus post.

**Details in the paper.**

# Experimental evaluation

- Ground truth dataset - Random news posts by 10 publishers

# Experimental evaluation

- Ground truth dataset - Random news posts by 10 publishers
- Supervised classifiers - Logistic regression, Linear SVM, Naive Bayes, Random Forest

# Experimental evaluation

- Ground truth dataset - Random news posts by 10 publishers
- Supervised classifiers - Logistic regression, Linear SVM, Naive Bayes, Random Forest

**Audience leaning based features outperform publisher and content based features for news post consensus classification task.**

# Experimental evaluation

- Ground truth dataset - Random news posts by 10 publishers
- Supervised classifiers - Logistic regression, Linear SVM, Naive Bayes, Random Forest

**Audience leaning based features outperform publisher and content based features for news post consensus classification task.**

**For more details come to the poster!**

# Experimental Evaluation

<b>Classifier</b>
<b>Logistic Regression</b>
<b>Linear SVM</b>
<b>Naive Bayes</b>
<b>Random Forest</b>

# Experimental Evaluation

Classifier	Different feature categories				
	Publisher based (P)	Tweet based (T)	P and T	Audience leaning based (A)	P, T, and A
Logistic Regression					
Linear SVM					
Naive Bayes					
Random Forest					

# Experimental Evaluation

Classifier	Different feature categories				
	Publisher based (P)	Tweet based (T)	P and T	Audience leaning based (A)	P, T, and A
<b>Logistic Regression</b>	0.58 $\pm$ 0.008	0.58 $\pm$ 0.008	0.68 $\pm$ 0.009	0.72 $\pm$ 0.012	0.72 $\pm$ 0.011
<b>Linear SVM</b>	0.58 $\pm$ 0.008	0.58 $\pm$ 0.008	0.68 $\pm$ 0.009	0.72 $\pm$ 0.012	0.72 $\pm$ 0.011
<b>Naive Bayes</b>	0.59 $\pm$ 0.007	0.57 $\pm$ 0.015	0.60 $\pm$ 0.01	0.66 $\pm$ 0.015	0.66 $\pm$ 0.012
<b>Random Forest</b>	0.58 $\pm$ 0.008	0.57 $\pm$ 0.01	0.64 $\pm$ 0.01	0.67 $\pm$ 0.015	0.67 $\pm$ 0.017

# Experimental Evaluation

Classifier	Different feature categories				
	Publisher based (P)	Tweet based (T)	P and T	Audience leaning based (A)	P, T, and A
<b>Logistic Regression</b>	0.58 $\pm$ 0.008	0.58 $\pm$ 0.008	0.68 $\pm$ 0.009	0.72 $\pm$ 0.012	0.72 $\pm$ 0.011
<b>Linear SVM</b>	0.58 $\pm$ 0.008	0.58 $\pm$ 0.008	0.68 $\pm$ 0.009	0.72 $\pm$ 0.012	0.72 $\pm$ 0.011
<b>Naive Bayes</b>	0.59 $\pm$ 0.007	0.57 $\pm$ 0.015	0.60 $\pm$ 0.01	0.66 $\pm$ 0.015	0.66 $\pm$ 0.012
<b>Random Forest</b>	0.58 $\pm$ 0.008	0.57 $\pm$ 0.01	0.64 $\pm$ 0.01	0.67 $\pm$ 0.015	0.67 $\pm$ 0.017

# Our contributions

- Defined & operationalized consensus of new posts
- Ground truth dataset of high and low consensus news posts on social media
- Empirical analysis of high & low consensus news posts
- Automated identification of high consensus news posts using audience leaning based features