

# Political audience diversity and algorithmic bias in news recommendation

Giovanni Luca Ciampaglia

NECO 2020





# CHEMTRAILS

You think that's condensation you're breathing now?

Copyrighted Material

FLATEARTH SOCIETY



THE EARTH IS FLAT  
QUESTION EVERYTHING

Copyrighted Material

10: 23 2014 Headline: FEMA Ebola Camps: James O'Keefe Exposes Gov't 'Death Cabins'

Search this site..



Ted Cruz



## NATIONAL REPORT

AMERICA'S #1 INDEPENDENT NEWS SOURCE

Sarah Palin



[FRONT PAGE](#) • [POLITICS](#) • [MEDIA](#) • [WORLD](#) • [BUSINESS](#) • [ENTERTAINMENT](#) • [RELIGION](#) • [HEALTH](#) • [SPORTS](#) • [CRIME](#) • [SCIENCE/TECH](#)

[ADVERTISE WITH NR](#)

[FAN MAIL](#)

[MEDIA INQUIRIES](#)

[STAFF](#)

[CONTRIBUTE](#)



RSS



EMAIL



TWITTER



FACEBOOK

# DALLAS CANCELS HALLOWEEN AMID EBOLA CONCERNS



# UNIVERSITY of SOUTH FLORIDA



Don't Shoot

Sponsored

It's not about who is "good" and "bad" it's about what is right!



Don't Shoot  
Community

250,350 people like this.



Musliminst

Sponsored

News, analysis from the Middle East & worldwide, multimedia & interactives, opinions



Al Jazeera English - Live

WWW.ALJAZEERA.COM

Like    Comment    Share



Sponsored

STAND AGAINST  
#POLICEBRUTALITY  
JOIN @COP\_BLOCK\_US



afrokingdom\_

Sponsored

Like Page



You can see an  
of African Ancestors  
slave ships ... more

Suggested Page



Blacktivist

Sponsored

260 Reactions 27 Comments

Black Lives Matter! Say it loud. I'm black and I'm proud!



Blacktivist  
Non-Governmental Organization (NGO)  
388,476 people like this.



Being Patriotic shared their event.

Sponsored

America has always been hinged on hard-working people. If you remove jobs, you'll remove our country from the world map. The state of Pennsylvania rose owing to multiple enterprises mining coal, producing steel, and creating the need for other jobs, groceries, doctors, dentists, insurance, gas, vehicles, mechanics and the list goes on. As far as Mr. Trump pursues the goal of creating more jobs and supports the working class. He said he would put miners back to work. We could ... See More



OCT 2 Miners for Trump: Unity day in Pennsyl...  
Sun 2 PM EDT - Pennsylvania  
77 people interested · 16 people going

Interested



[Learn More](#)



american.veterans Hillary Clinton will never understand what it feels like to lose the person you love for the sake of your country.



Secured Borders

Sponsored

Every man should stand for our borders! Join!



Secured Borders  
News & Media Website  
135,301 people like this.



american.made

Sponsored

This article takes you through some five documented

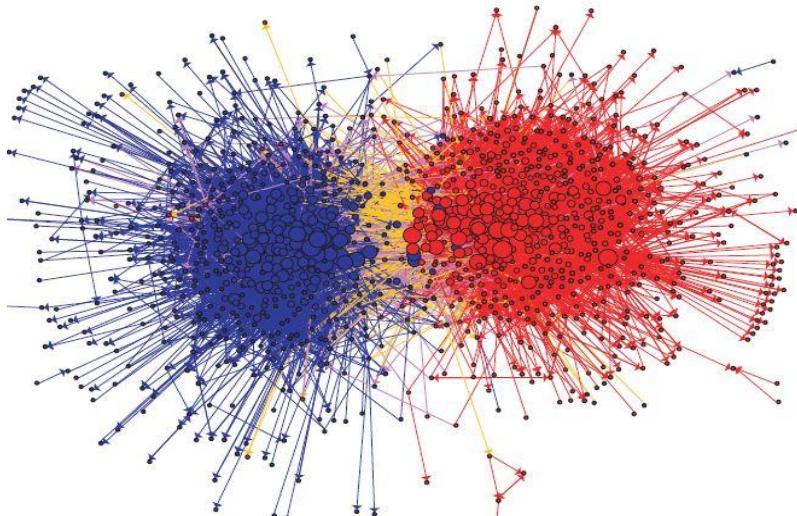
about Terrence Sterling's death that you need to know



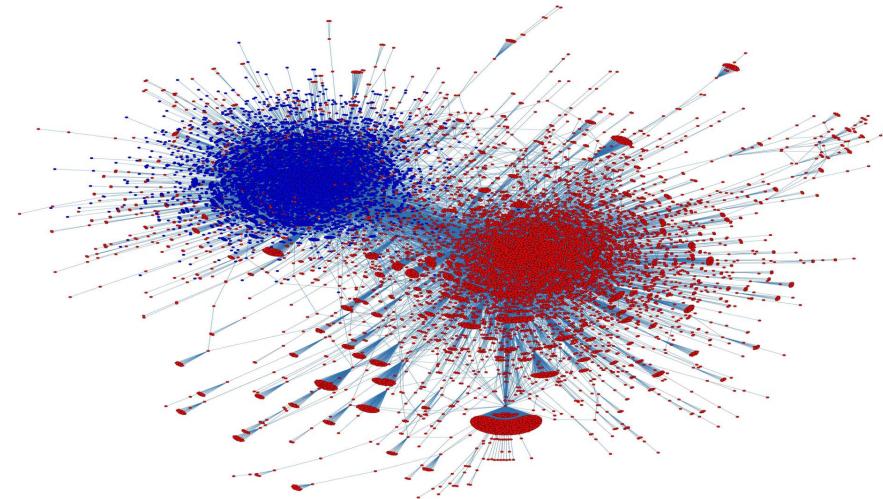
Born Libre

Sponsored

Thank you Sister & Jionists who are  
pared to murder & LIFE? If they  
nothers, set up



Adamic & Glance (2005),  
[Blogs]



Conover et al. (2011),  
[Twitter]

LIBERAL ⓘ

SHOWING POSTS ABOUT:  
"HEALTH CARE"

CONSERVATIVE ⓘ

**ACLU** about 8 months ago

SAY HER NAME:  
Dana Martin  
Jazzaline Ware  
Ashanti Carmon  
Claire Legato... See More



Credit: Governor Tom Wolf/Flickr

ACLU.ORG

**Black Trans Women Are Being Murdered in the Street...**

At least 15 Black trans women have been killed so far in 2019.

139 327 453

**Mom of Diabetics Demands Action from ... U.S. Senator Bernie Sanders**



**DEMANDING ACTION ON RX DRUG PRICING**

**The Daily Caller** about 8 months ago

Kamala flip-flops on health care.



DAILYCALLER.COM

**Kamala Harris Said She's Not 'Comfortable' With Ber...**

'I have not been comfortable with Bernie's plan'

139 327 453

**The Daily Caller** about 8 months ago

**Joe Biden** @JoeBiden

My plan to protect and build on Obamacare will not only provide coverage for uninsured Americans, but it will also make health care more affordable for those who already have coverage.

4:25 PM · 8/17/19 · TweetDeck

Biden: I WILL MAKE THE AFFORDABLE CARE ACT EVEN MORE AFFORDABLE

# Background

- People have strong preferences for **pro-attitudinal content**.
- Low-quality content will generates **higher levels of engagement in like-minded audiences**.
- Newsfeed systems prioritize highly engaging content.
- Fake news websites can abuse **social media recommendation systems** by **targeting** particular audiences.

# Purpose of this study

Can we design news recommendation systems that are robust against echo chambers?

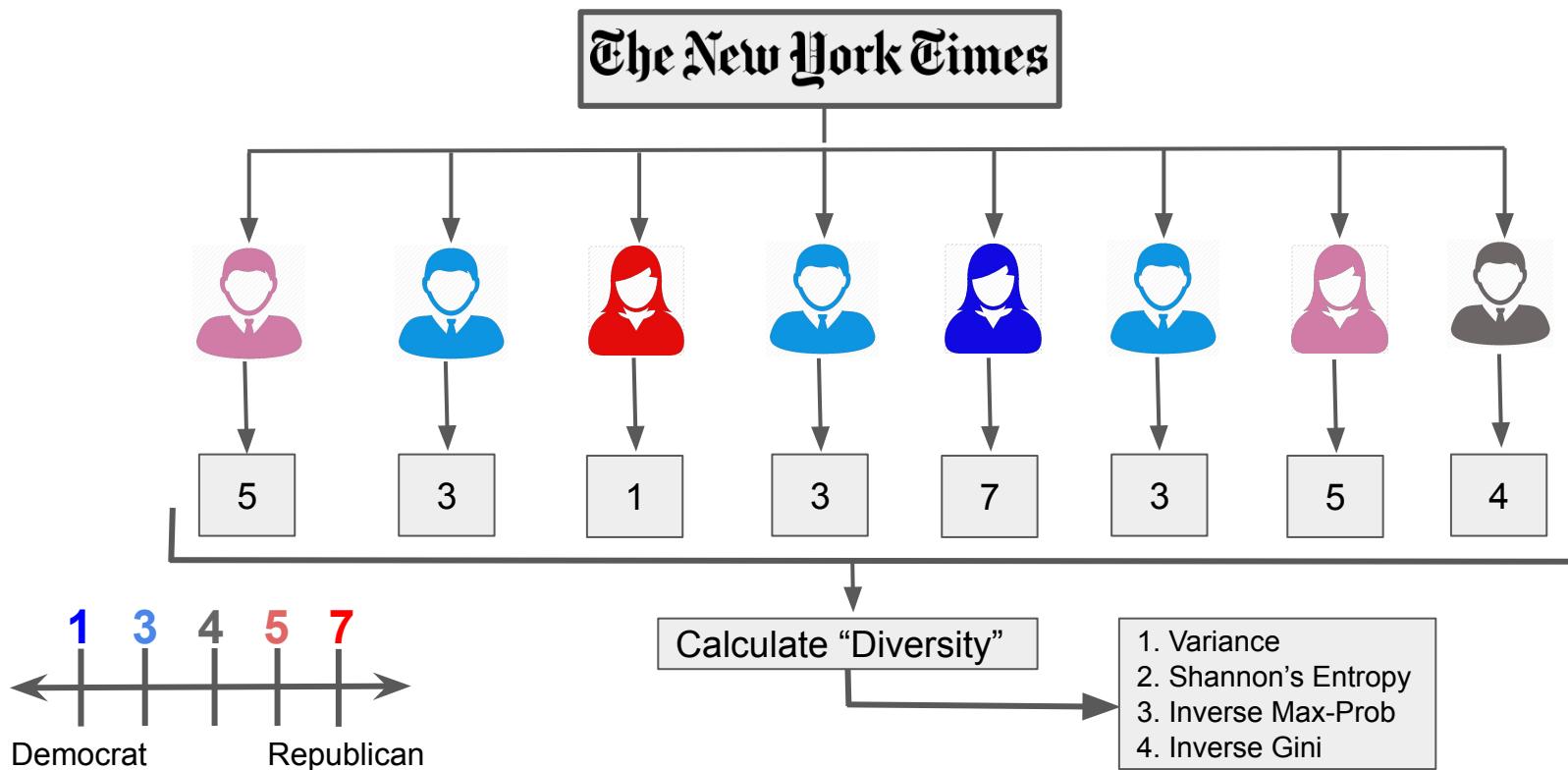
How can news recommendation systems promote sources with higher **journalistic quality** in a polarized environment?

## One possible Solution....

Look for news websites that generate high engagement and also attract a diverse set of audiences.

**Audience Slant Diversity:** The **diversity** in the **ideological slants** of the audience that the news website has managed to attract.

# Empirical Analysis : Audience Slant Diversity



# Research Questions

- ❖ Is there a correlation between audience diversity and news quality?
  - We perform an empirical analysis to determine this.
- ❖ Can we use audience diversity signal to make better recommendations?
  - We evaluate a new variant of the **collaborative filtering** algorithm.

# Data Sources



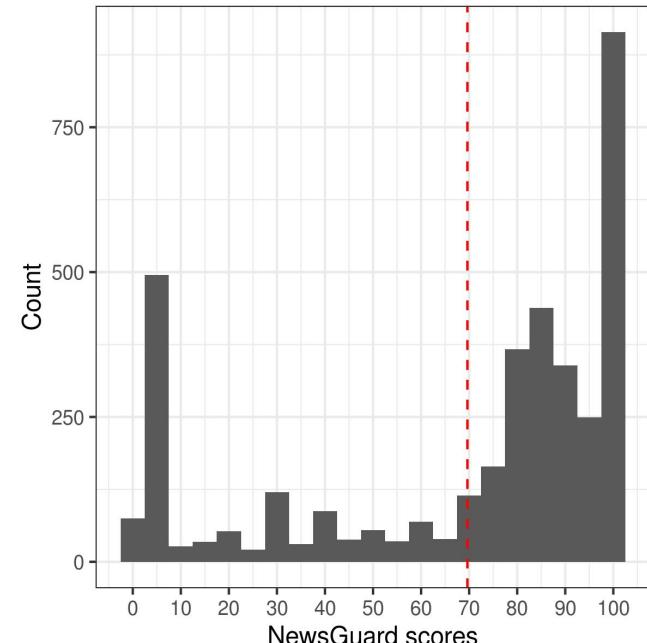
A comprehensive data set of **web traffic history** from 18,298 American citizens



Data set of **credibility ratings** of 3,765 web domains by expert journalists

# NewsGuard

- A list of web domains with reliability ratings.
- Ratings on 100-point scale.
- Coverage : 3,765 domains.
- Rubric:
  - editorial responsibility,
  - accountability,
  - financial transparency, etc..
- Categorical classification of websites also present :
  - Trustworthy, Untrustworthy, Satire, Platform
- Mean reliability score : 69.6



Distribution of NewsGuard scores. The red dashed line is the average score



- A sample of U.S.-based internet users whose web traffic was collected in an anonymized form with their prior consent.
- Total participants: 19,298
- **Participants also reported their partisanship on a seven point scale (where 1 = strong Democrat and 7 = strong Republican)**

Duration	Respondents	Domains	Pageviews
Oct. 07, 2016 – Nov. 14, 2016	3,251	158,706	26,715,631
Oct. 25, 2017 – Nov. 21, 2017	2,100	104,513	14,247,987
Jun. 11, 2018 – Jul. 31, 2018	1,718	108,953	15,212,281
Jul. 12, 2018 – Aug. 02, 2018	2,000	74,469	9,395,659
Oct. 05, 2018 – Nov. 05, 2018	3,332	98,850	19,288,382
Nov. 12, 2018 – Jan. 16, 2019	4,907	117,510	21,093,638
Jan. 24, 2019 – Mar. 11, 2019	2,000	113,700	27,482,462

YouGov Pulse respondent data summary



For each web domain, pool visitors



Remove domains with < 30 visitors



Separate out domain which have  
T/U tag in NewsGuard

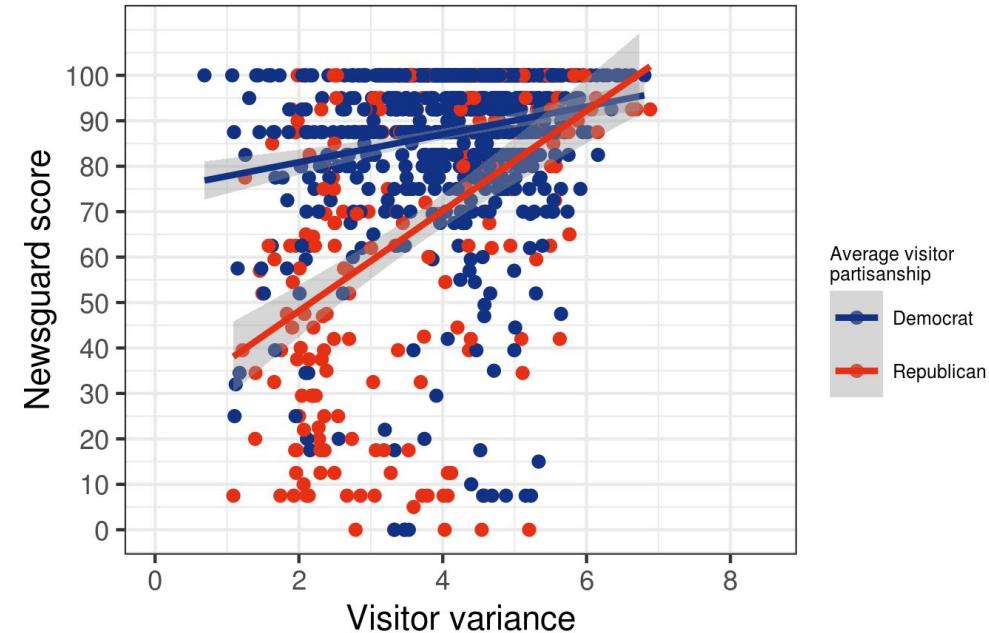
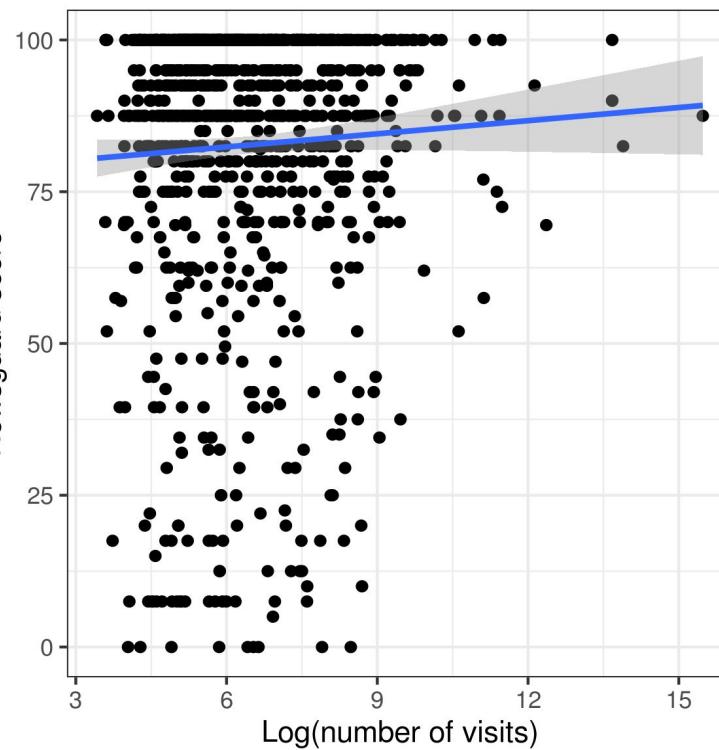


# domains : 1760 # users : 7112

# Empirical Analysis : Diversity Metrics

- Let  $N_k$  be the count of participants who visited a particular domain  $D$  and reported their political affiliation to be  $k$  ( $k = 1, \dots, 7$ ).
- $N = \sum_k N_k$  where  $N$  is the total number of participants who visited that domain.
- Also, let  $s_i$  be the slant of  $i^{\text{th}}$  individual.
- Variance:

$$\sigma^2 = \frac{\sum(s_i - \bar{s})}{N}$$



Pearson Correlation : 0.32

**LIBERAL** ⓘ

ACLU.org about 8 months ago

SAY HER NAME:  
Dana Martin  
Jazzaline Ware  
Ashanti Carmon  
Claire Legato... See More

ACLU.ORG Black Trans Women Are Being Murdered in the Street... At least 15 Black trans women have been killed so far in 2019.

1,655 76 741

DAILYCALLER.COM Kamala Harris Said She's Not 'Comfortable' With Ber...  
'I have not been comfortable with Bernie's plan'

139 327 453

The Daily Caller ⓘ about 8 months ago Tweet

JOE @JoeBiden

My plan to protect and build on Obamacare will not only provide coverage for uninsured Americans, but it will also make health care more affordable for those who already have coverage.

4:25 PM - 8/17/19 · TweetDeck

Biden: I WILL MAKE THE AFFORDABLE CARE ACT EVEN MORE AFFORDABLE

628 5.8K 851

CNSNews.com Elderly Woman Confronts Kamala Harris...  
CNSNews.com

**CONSERVATIVE** ⓘ

DAILYCALLER.COM Kamala flip-flops on health care.

The Daily Caller ⓘ about 8 months ago Tweet

JOE @JoeBiden

My plan to protect and build on Obamacare will not only provide coverage for uninsured Americans, but it will also make health care more affordable for those who already have coverage.

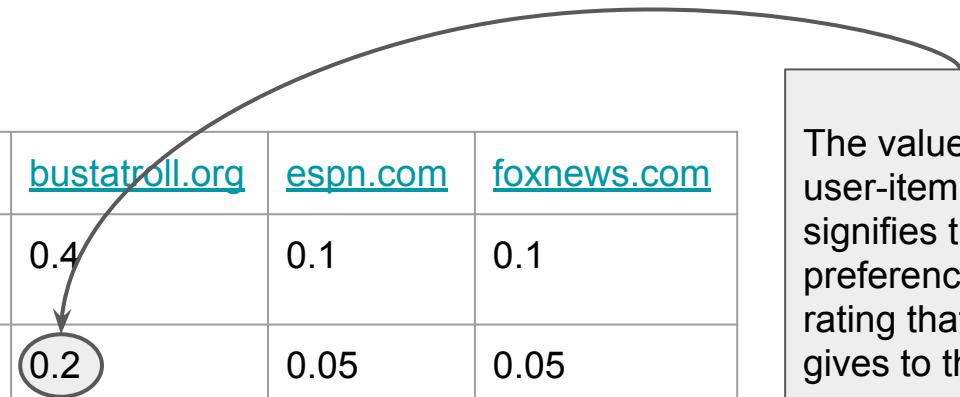
4:25 PM - 8/17/19 · TweetDeck

Biden: I WILL MAKE THE AFFORDABLE CARE ACT EVEN MORE AFFORDABLE

628 5.8K 851

CNSNews.com Elderly Woman Confronts Kamala Harris...  
CNSNews.com

	<a href="#">nytimes.com</a>	<a href="#">cnn.com</a>	<a href="#">bustatroll.org</a>	<a href="#">espn.com</a>	<a href="#">foxnews.com</a>
	0.3	0.1	0.4	0.1	0.1
	0.1	0.7	0.2	0.05	0.05
	0.2	0.4	0.1	0.25	0.05
	0.3	0.1	0.4	0.1	0.1
	0.1	0.7	0.2	0.05	0.05



The values in user-item matrix signifies the preference or rating that the user gives to the item.

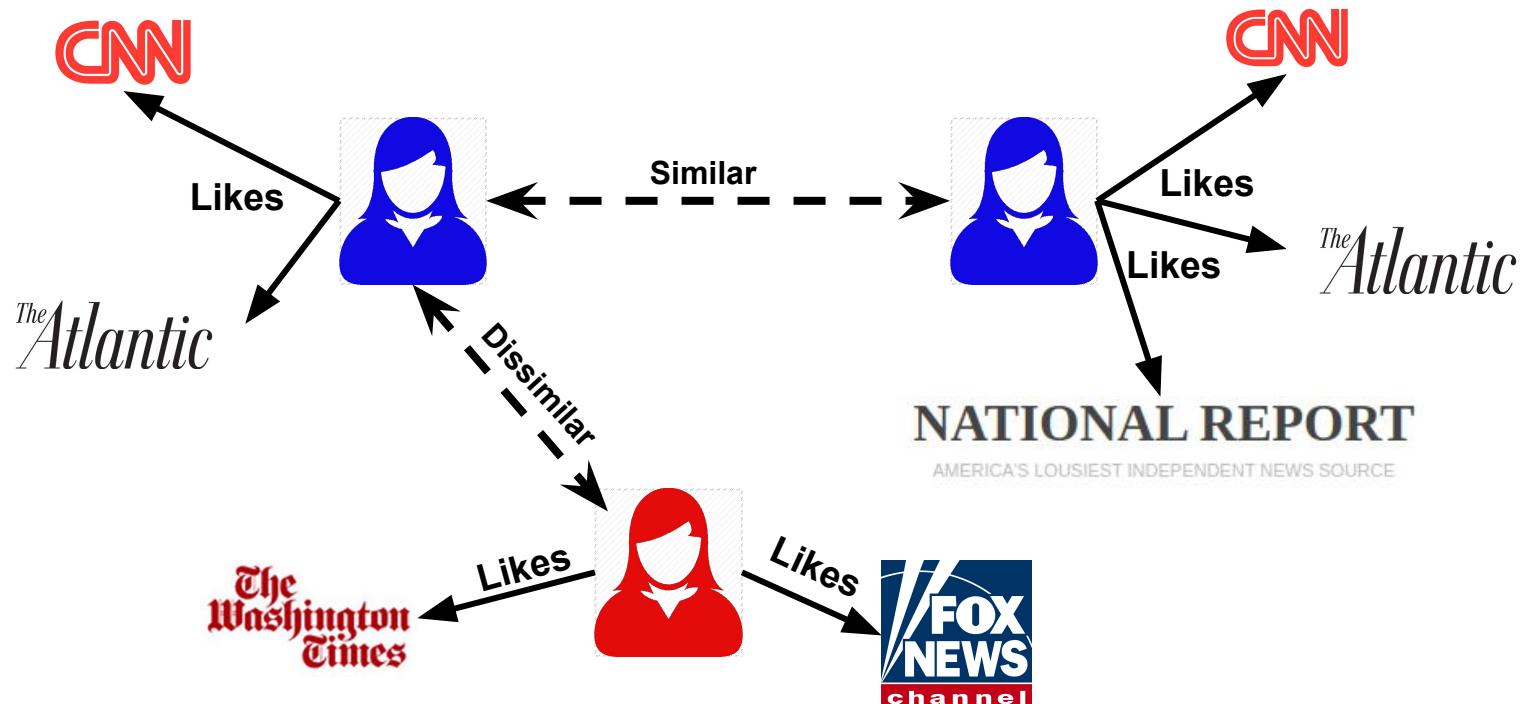
This matrix is partially filled!

# Calculating Ratings:

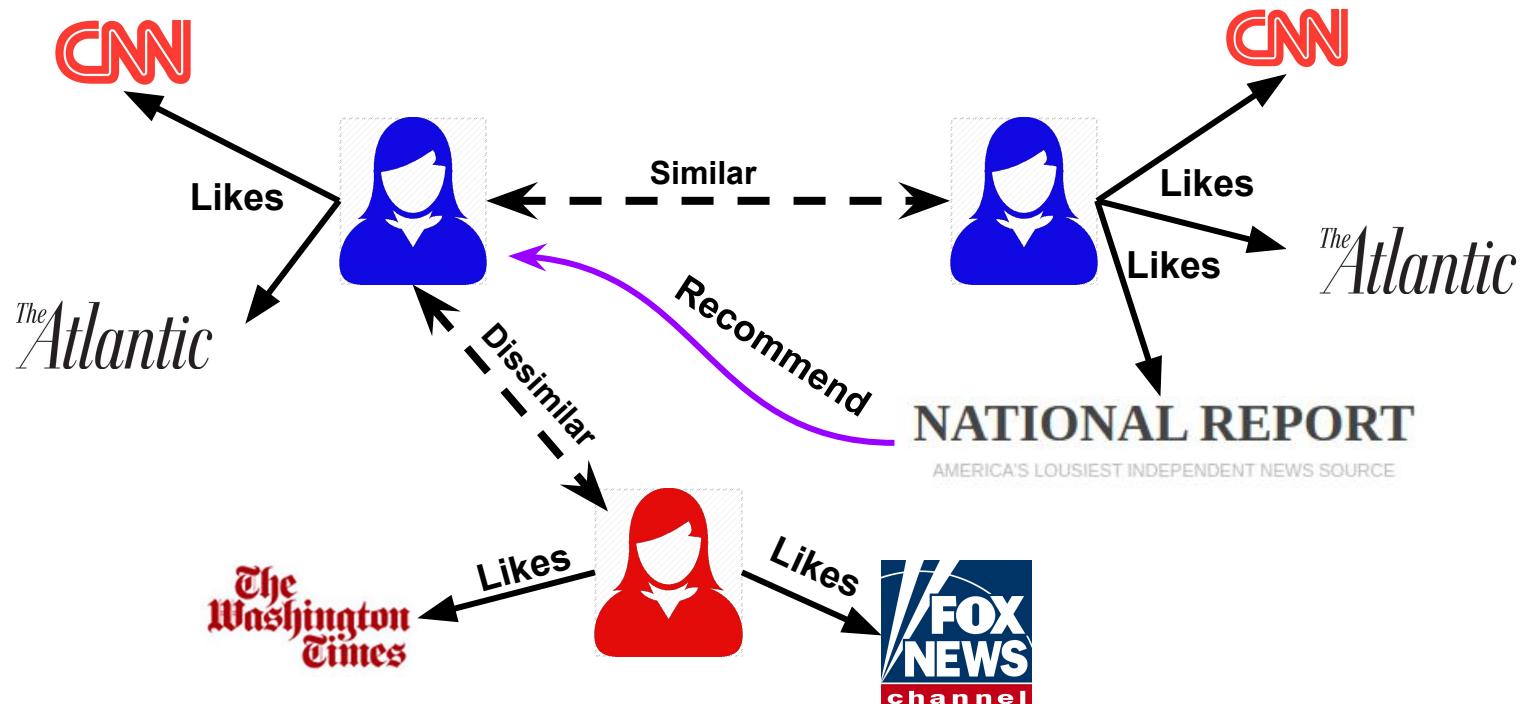
For determining the ratings user  $u$  gives to domain  $d$ , we take the **TF-IDF of the views** that  $u$  has for  $d$ :

$$r_{u,d} = \frac{V_{u,d}}{\sum_d V_{u,d}} \log \left( \frac{V}{\sum_u V_{u,d}} \right)$$

$$V = \sum_u \sum_d V_{u,d}$$



$sim(u, v) = \text{Pearson Similarity}(U_u, U_v) \text{ or Kendall Tau Similarity}(U_u, U_v)$

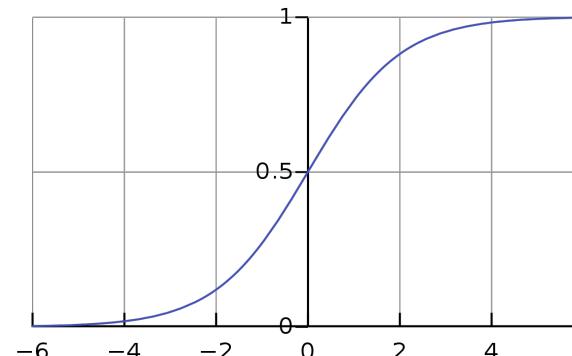


$$r_{u,d} = \bar{r}_u + \frac{\sum_{v \in N(u)} sim(u, v)(r_{v,d} - \bar{r}_v)}{\sum_{v \in N(u)} sim(u, v)}$$

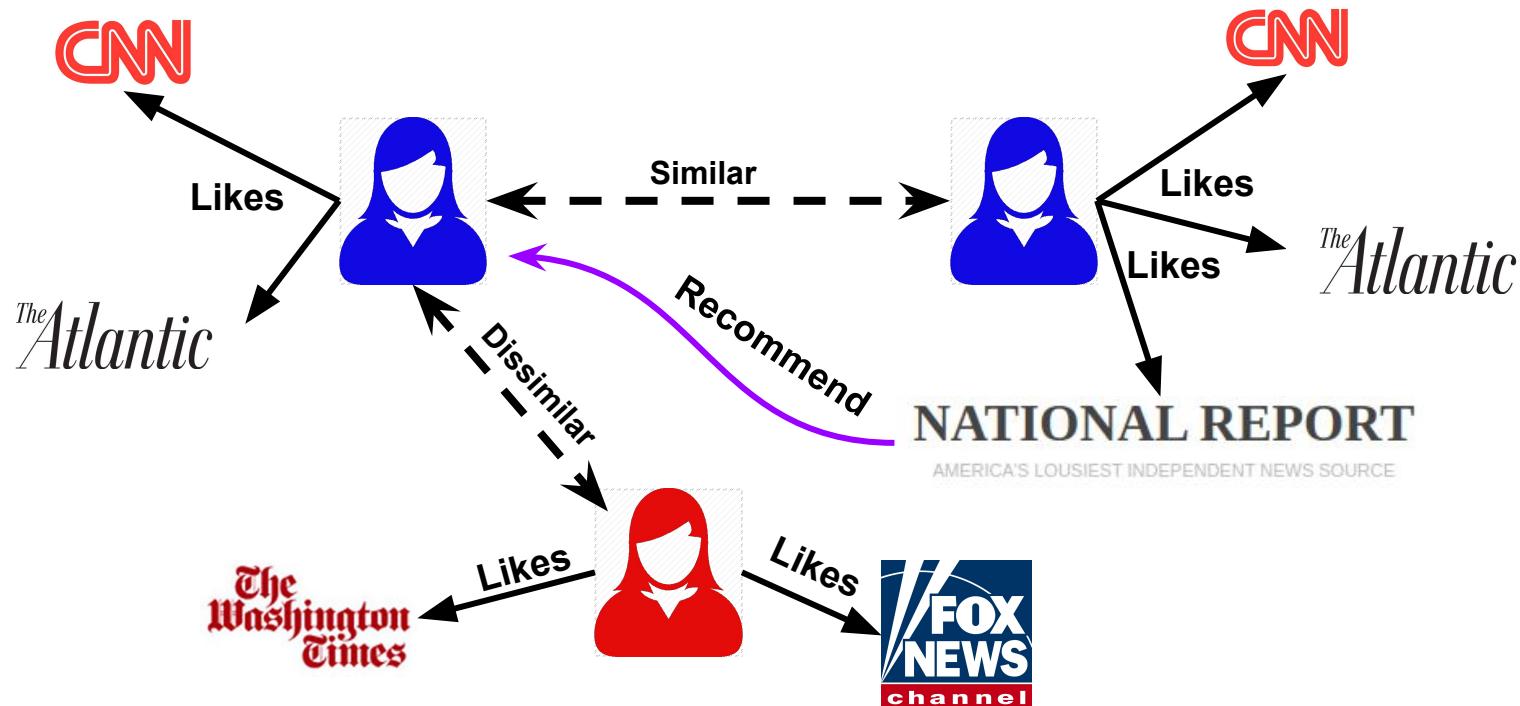
# Incorporating Diversity signal in User based CF

$$\sigma(d) = \frac{a}{1 + e^{-\frac{(var_d - t)}{\tau}}}$$

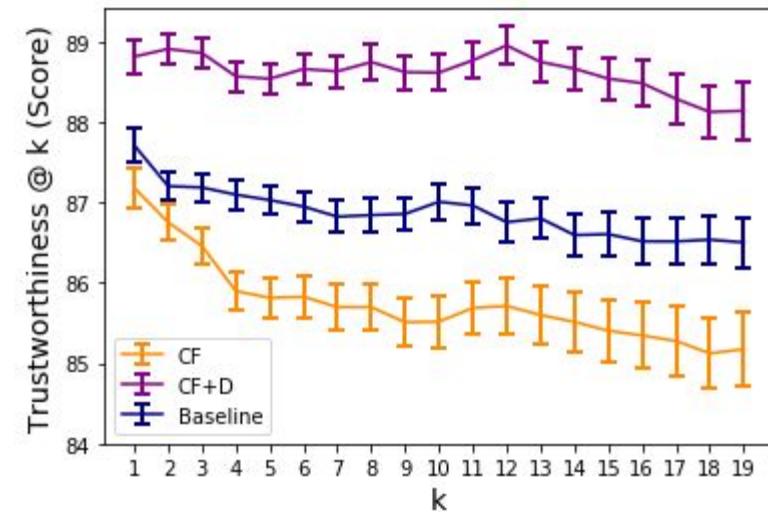
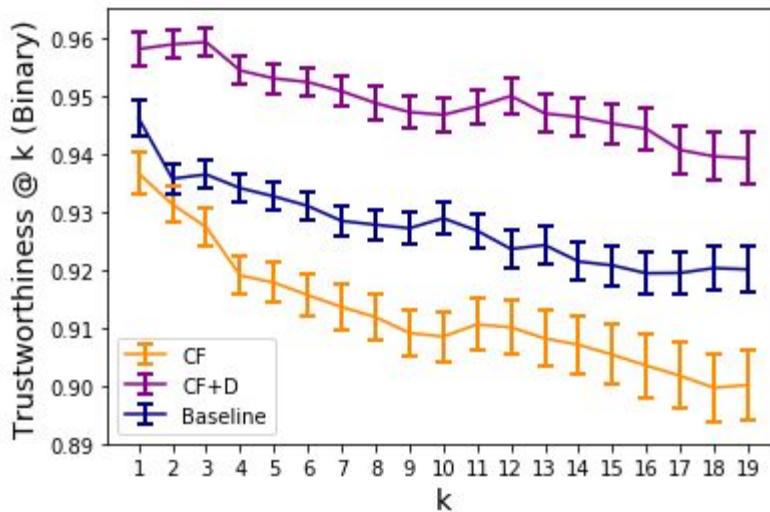
Audience diversity metric of the web domain

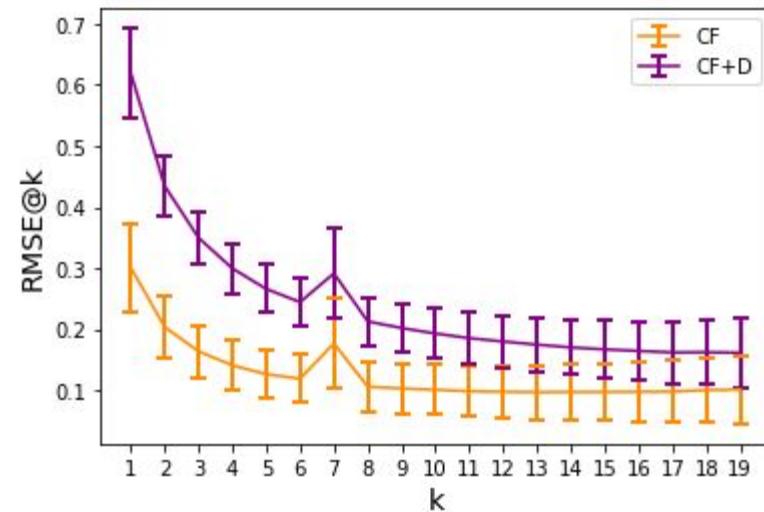
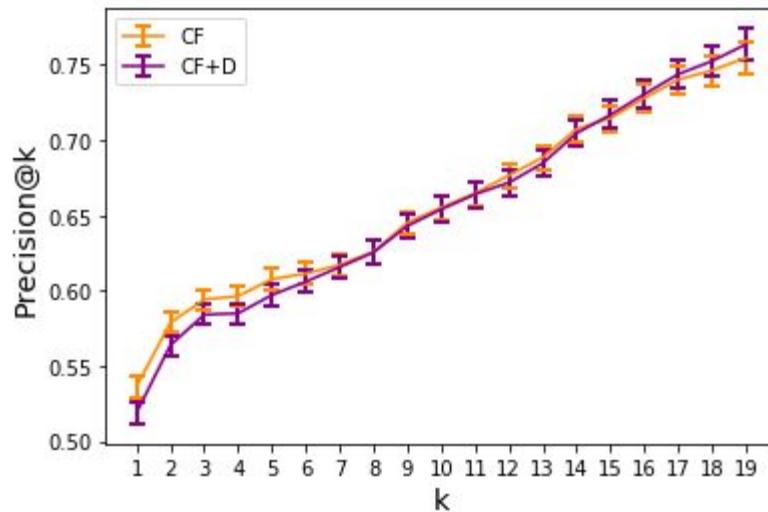


a : upper asymptote  
τ : inverse growth rate  
t : location



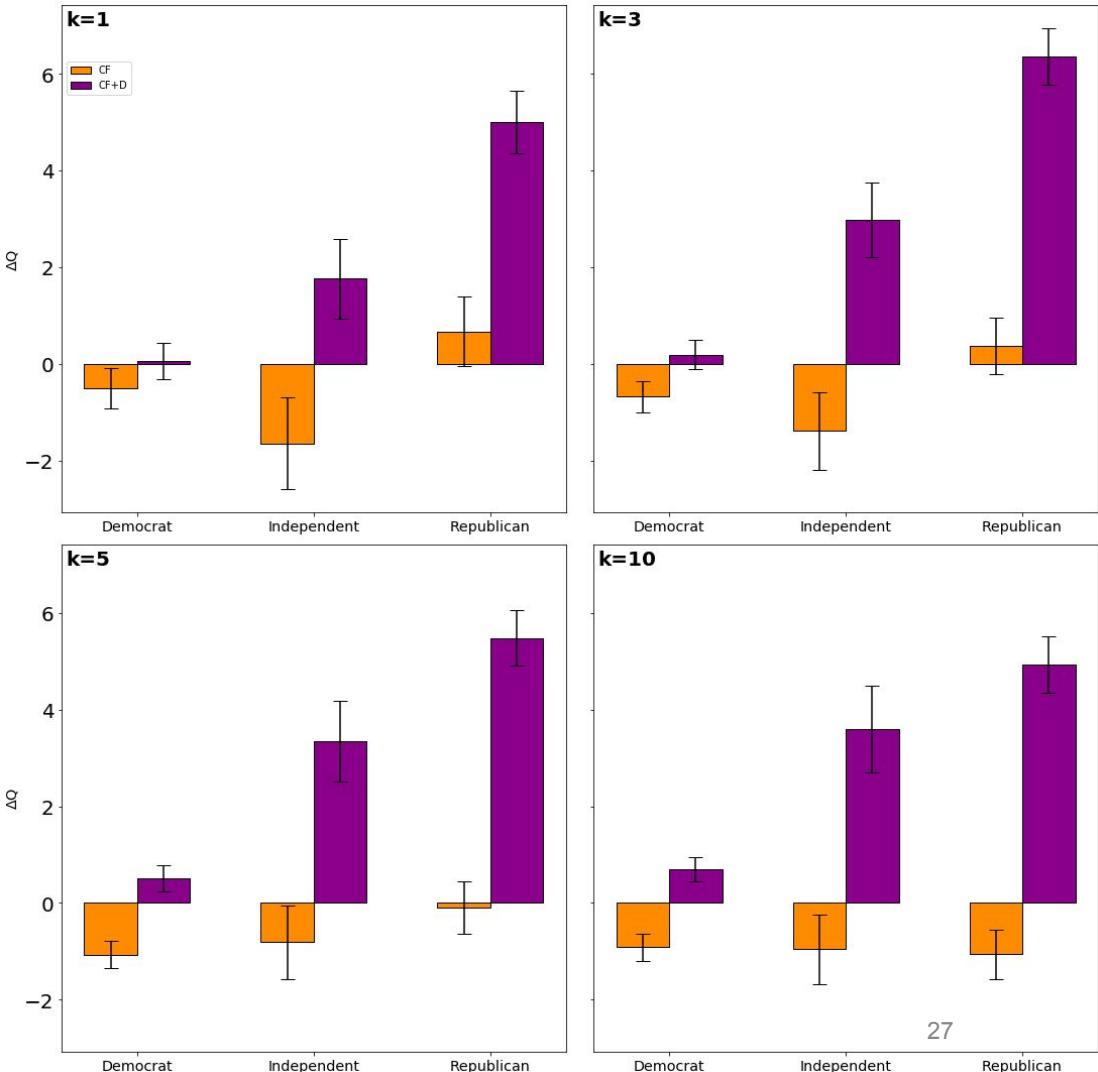
$$r_{u,d} = \bar{r}_u + \frac{\sum_{v \in N(u)} sim(u, v)(r_{v,d} - \bar{r}_v)}{\sum_{v \in N(u)} sim(u, v)} + \frac{a}{1 + e^{-\frac{(var_d - t)}{\tau}}}$$





# Media Diet Effect

Slants from Bakshy et al. 2015



# The Ranking Model

$$P\{X = k\} = \frac{r_k^{-\alpha}}{\sum_h r_h^{-\alpha}}$$

Given two rankings  $\{r_k\}_k$  and  $\{r'_k\}_k$ ,

$$\Delta Q = \mathbb{E}[Q(X)] - \mathbb{E}[Q(X')]$$

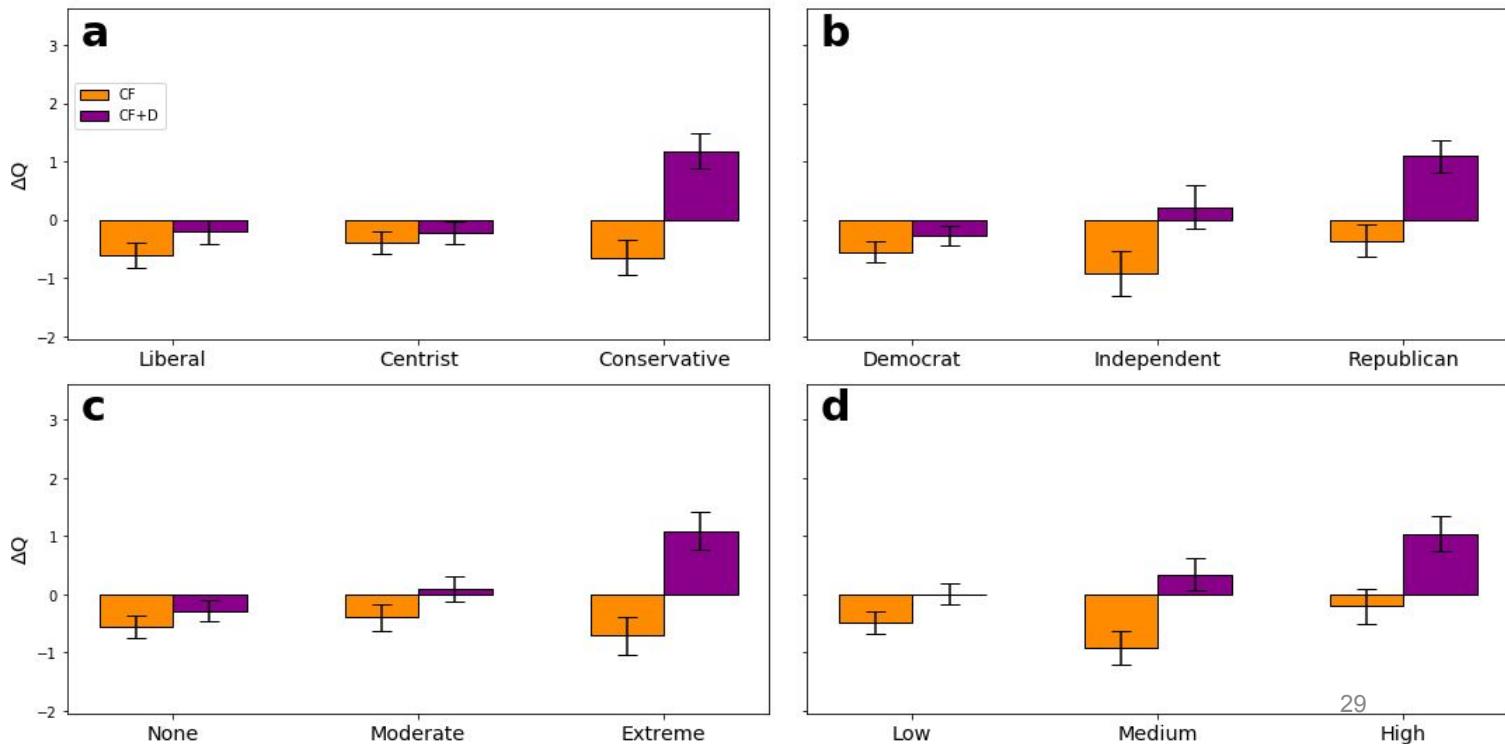
where  $Q(X)$  and  $Q(X')$  is the trustworthiness of item  $X$



1	xyz.com
2	foo.net
3	abc.org
4	bar.us
5	baz.it

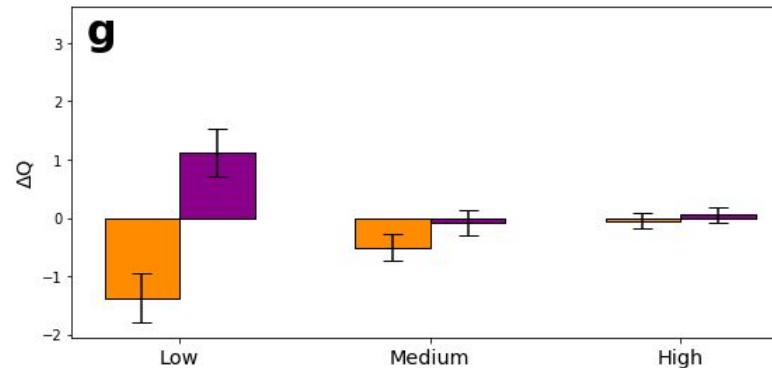
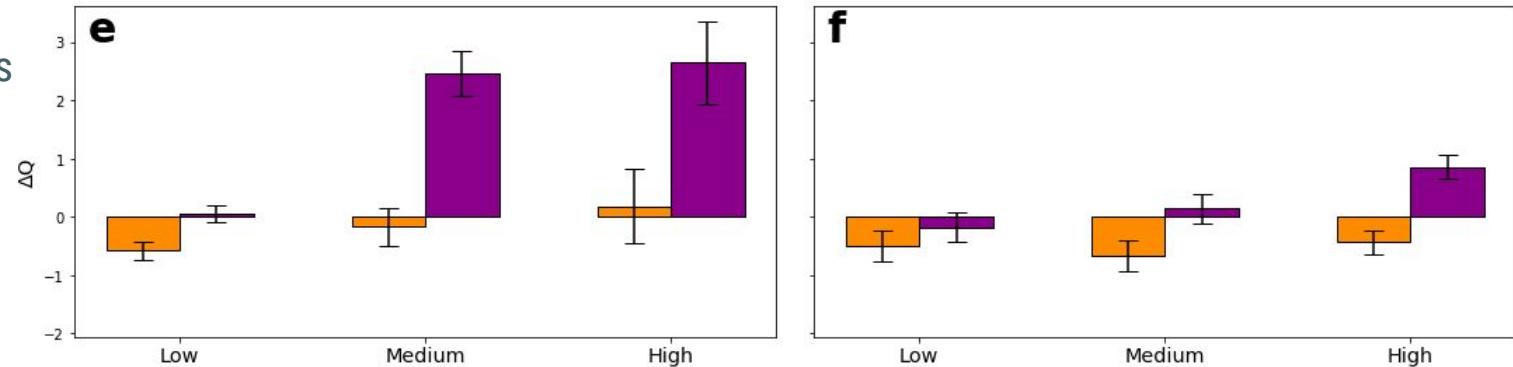
# Stratification

- a) Media Diet
- b) Ideology
- c) Abs. Partisan.
- d) Tot. clicks



# Stratification

- e) Distinct Domains
- f) Top- $n$  similarity
- g) Baseline Trust.



# Summary

- The current news recommendation algorithms are **not able** to select “high quality” news domains, which is being abused by fake news producers.
- **Audience slant diversity** can be a valuable signal for news quality as there is some significant correlation between them.
- Incorporating the diversity signal in recommendation systems can **improve the trustworthiness** of their recommendations while not impacting their precision.

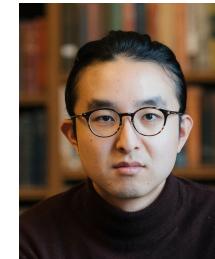
# Acknowledgements



Filippo Menczer



Alessandro Flammini



Shun Yamaya



Brendan Nyhan



Saumya Bhadani



UNIVERSITY OF  
SOUTH FLORIDA

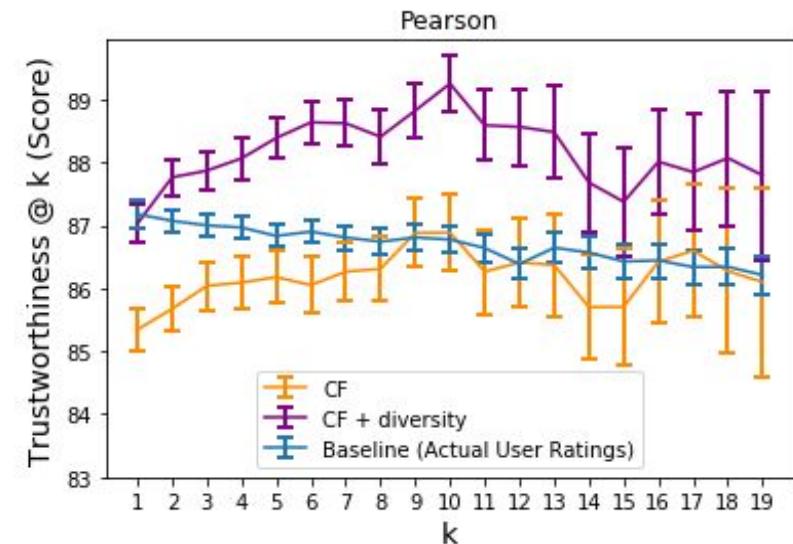
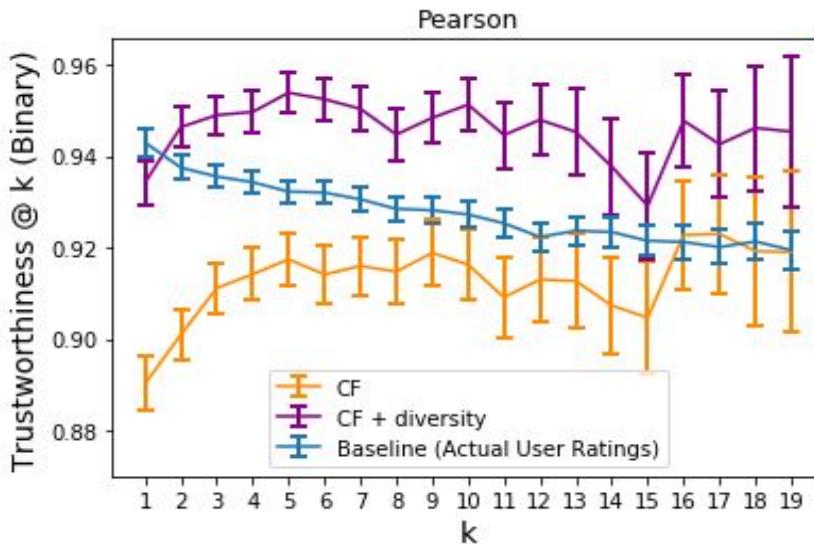


# Thank You!

[glc3@mail.usf.edu](mailto:glc3@mail.usf.edu)

[glciampaglia.com](http://glciampaglia.com)





Diversity metric	Correlation	R <sup>2</sup>	p-value
USER LEVEL			
Variance	0.32	0.10	< 0.01
Entropy (Dir.)	0.21	0.04	< 0.01
Entropy (ML)	0.20	0.04	< 0.01
Entropy (NSB)	0.22	0.05	< 0.01
Inv. MaxProb	0.06	0.00	0.03
Inv. Gini	0.14	0.02	< 0.01