
# Bias and Fairness in Information Filtering Systems

**Registration Seminar**
**Abhisek Dash**
**17CS91R01**

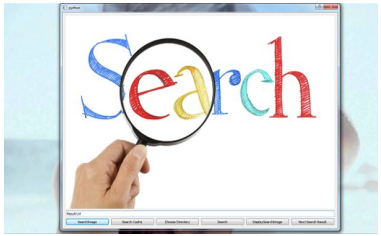
**Supervised by**
**Dr. Saptarshi Ghosh**
**Dr. Animesh Mukherjee**

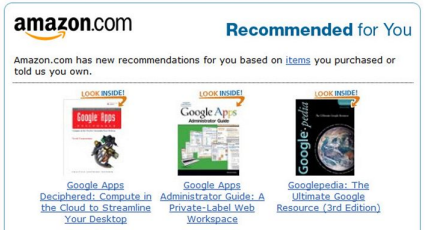
# Information Filtering Systems

A system for monitoring large amount of dynamically generated information and pushing to a user a subset of information likely to be of his/her interest (based on his/her information needs).


Belkin, Nicholas J., and W. Bruce Croft. "Information filtering and information retrieval: Two sides of the same coin?." *Communications of the ACM* 35.12 (1992): 29-38.

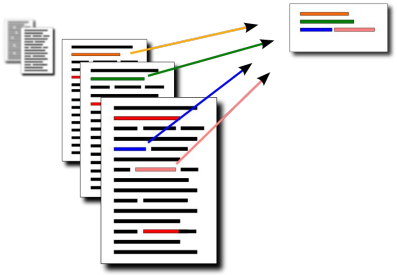
## Search systems



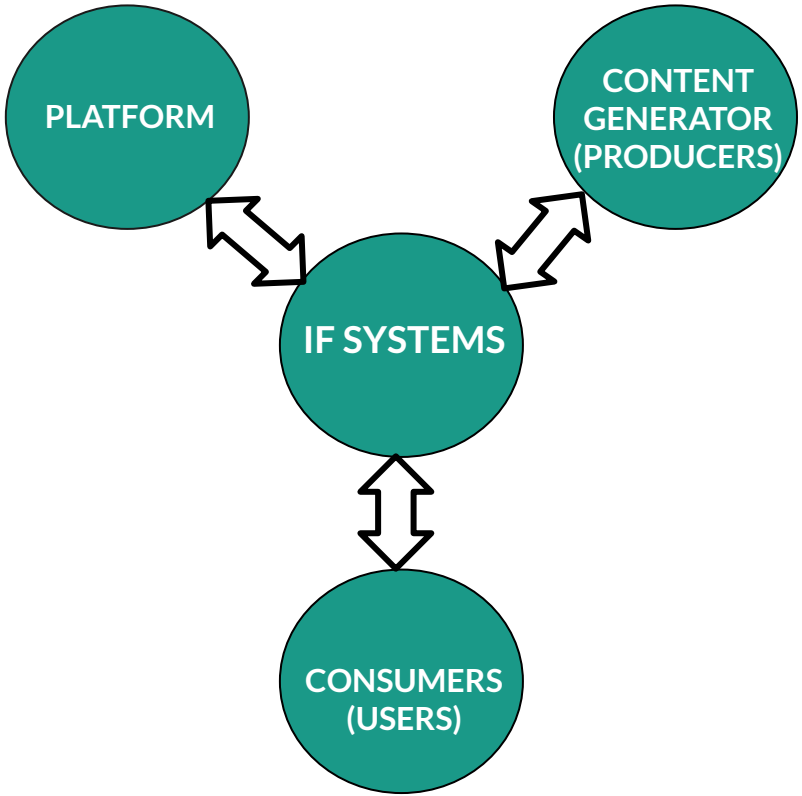

## Recommendation systems

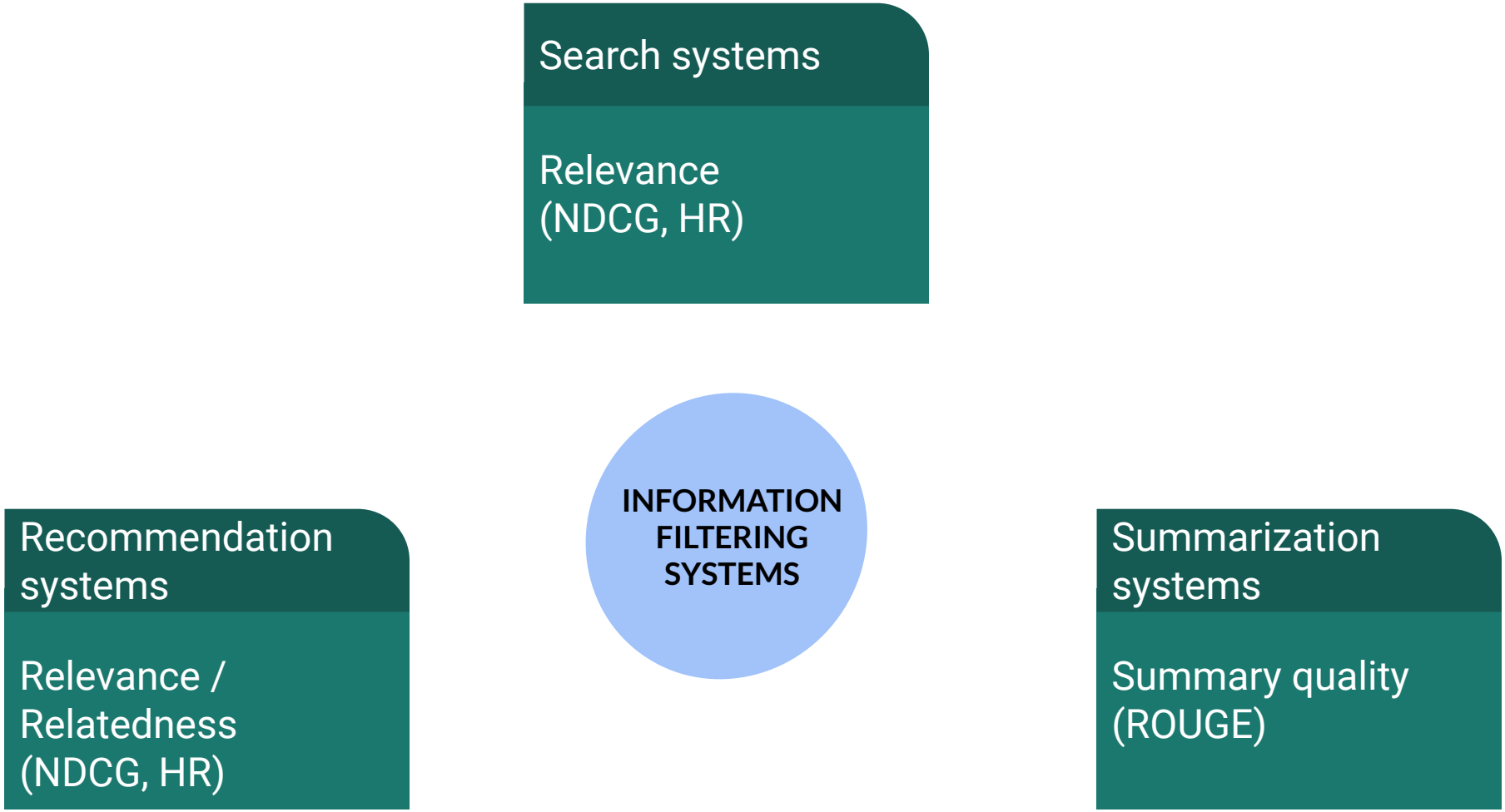
**INFORMATION FILTERING SYSTEMS**



## Summarization systems

# PRIMARY STAKEHOLDERS




Search systems
Relevance
(NDCG, HR)
INFORMATION FILTERING SYSTEMS
Recommendation systems
Relevance / Relatedness
(NDCG, HR)
Summarization systems
Summary quality
(ROUGE)

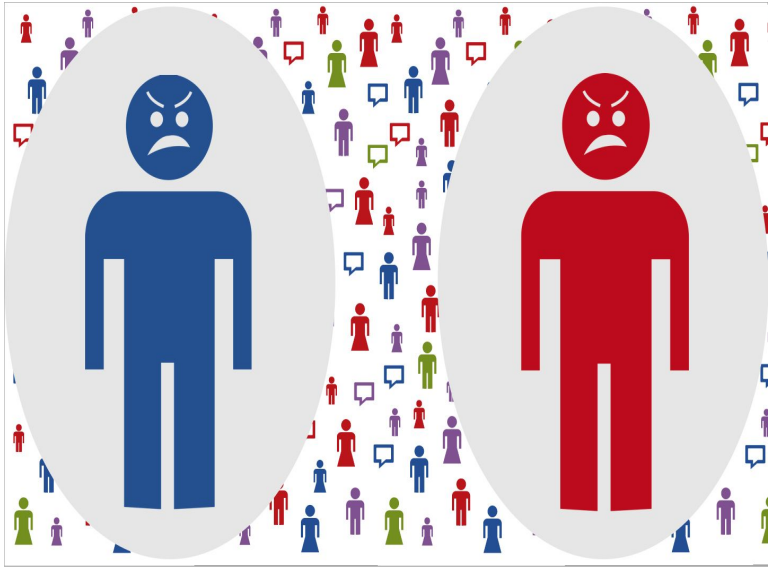
# Inadvertent consequences

- **Discrimination / bias / under representation**

[1] Kulshrestha, Juhi, et al. "Quantifying search bias: Investigating sources of bias for political searches in social media." *ACM CSCW 2017*
[2] Mehrotra, Rishabh, et al. "Auditing search engines for differential satisfaction across demographics." *WWW 2017*

# Inadvertent consequences

- Discrimination / bias / under representation
- **Information segregation**



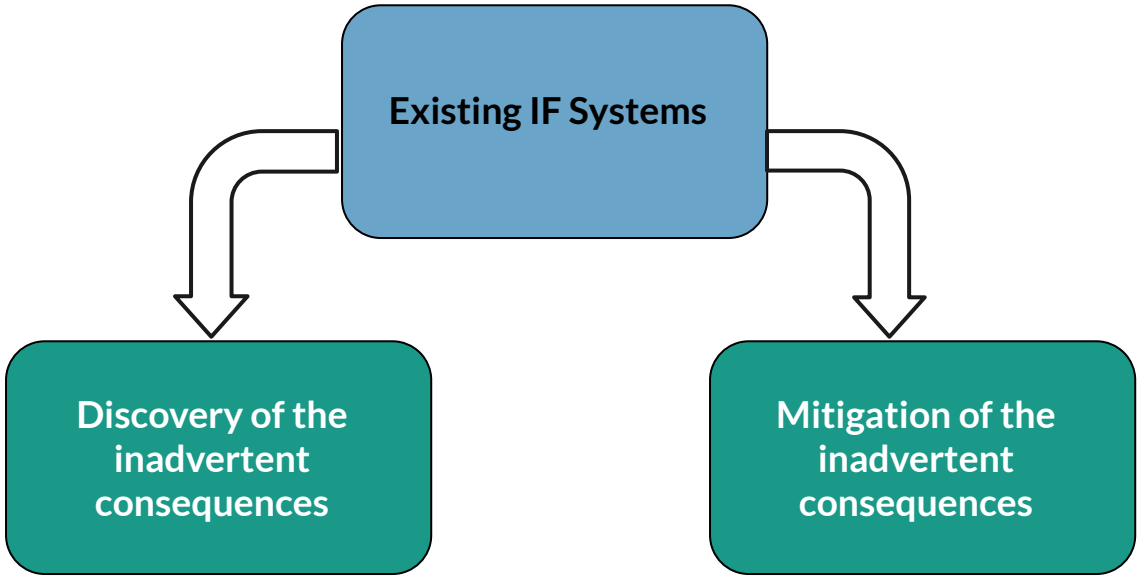
[1] Pariser, Eli. *The filter bubble: What the Internet is hiding from you*. Penguin UK, 2011.
[2] Chakraborty, Abhijnan, et al. "On quantifying knowledge segregation in society." *FATREC workshop, RecSys 2017*.

# Inadvertent consequences

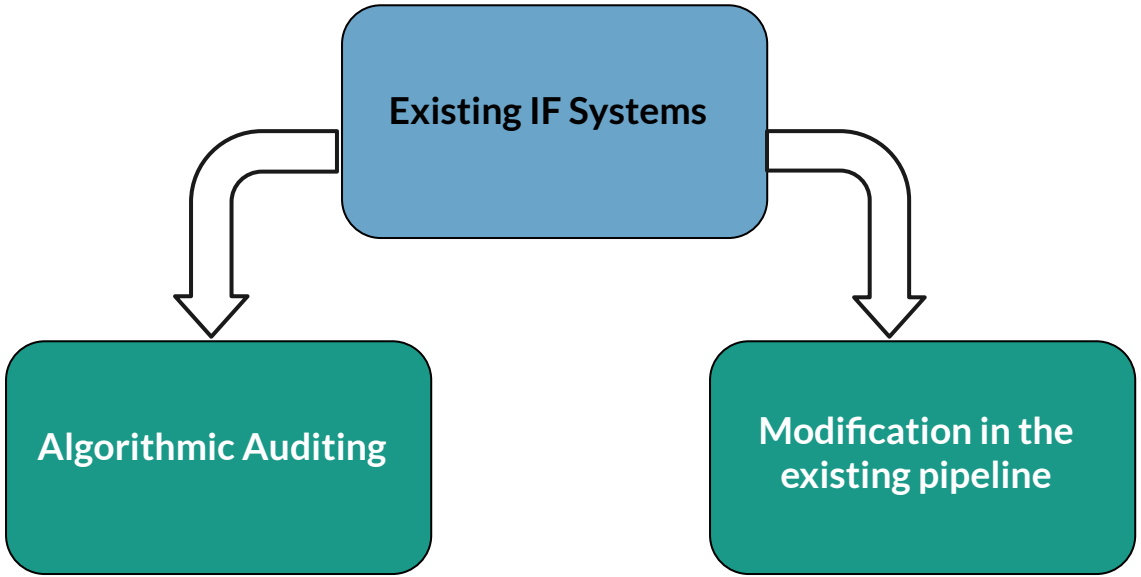
- Discrimination / bias
- Information segregation

**HOW CAN WE ACCOUNT FOR THESE IN THE EXISTING (DEPLOYED) SYSTEMS?**

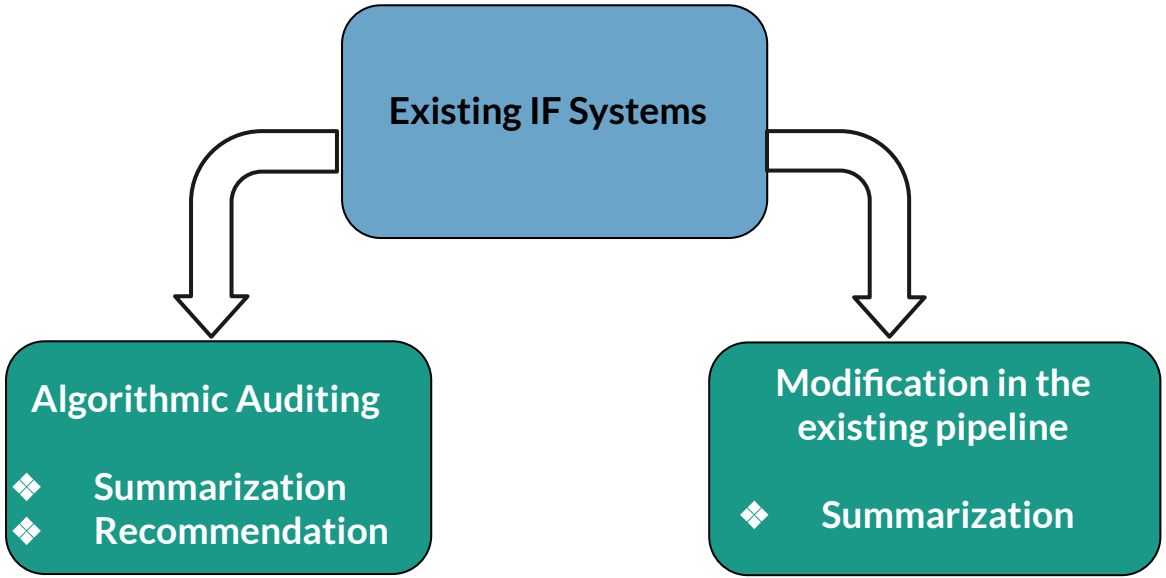
# Need of the hour



# Need of the hour



# Our contributions



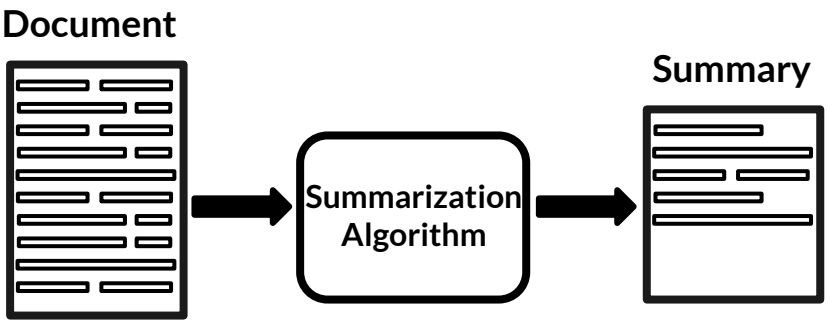
# PART I

## Motivation and Methods for Fairness in Algorithmic Summaries

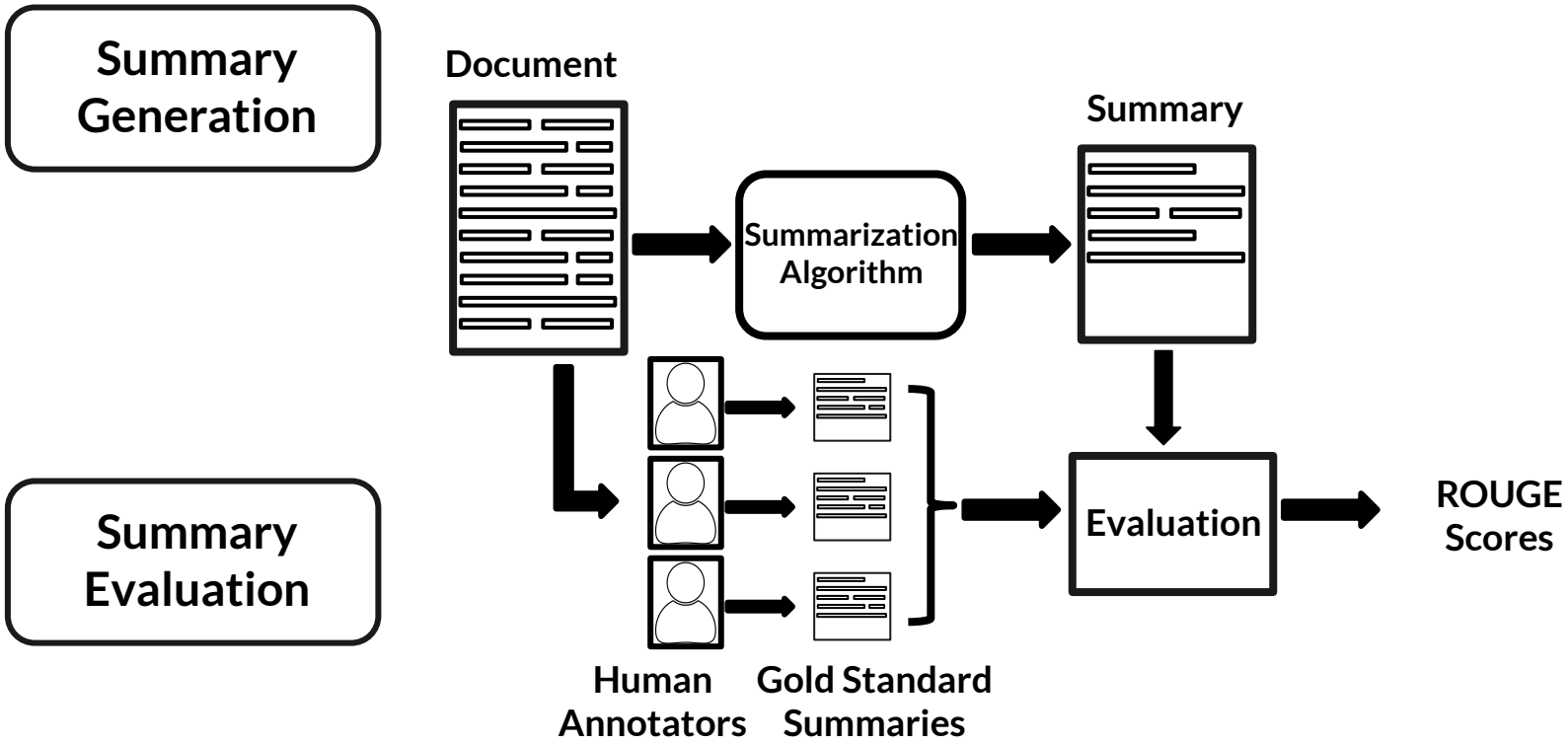
-PACMHCI 2019 (ACM CSCW 2019)

# Generic text summarization framework

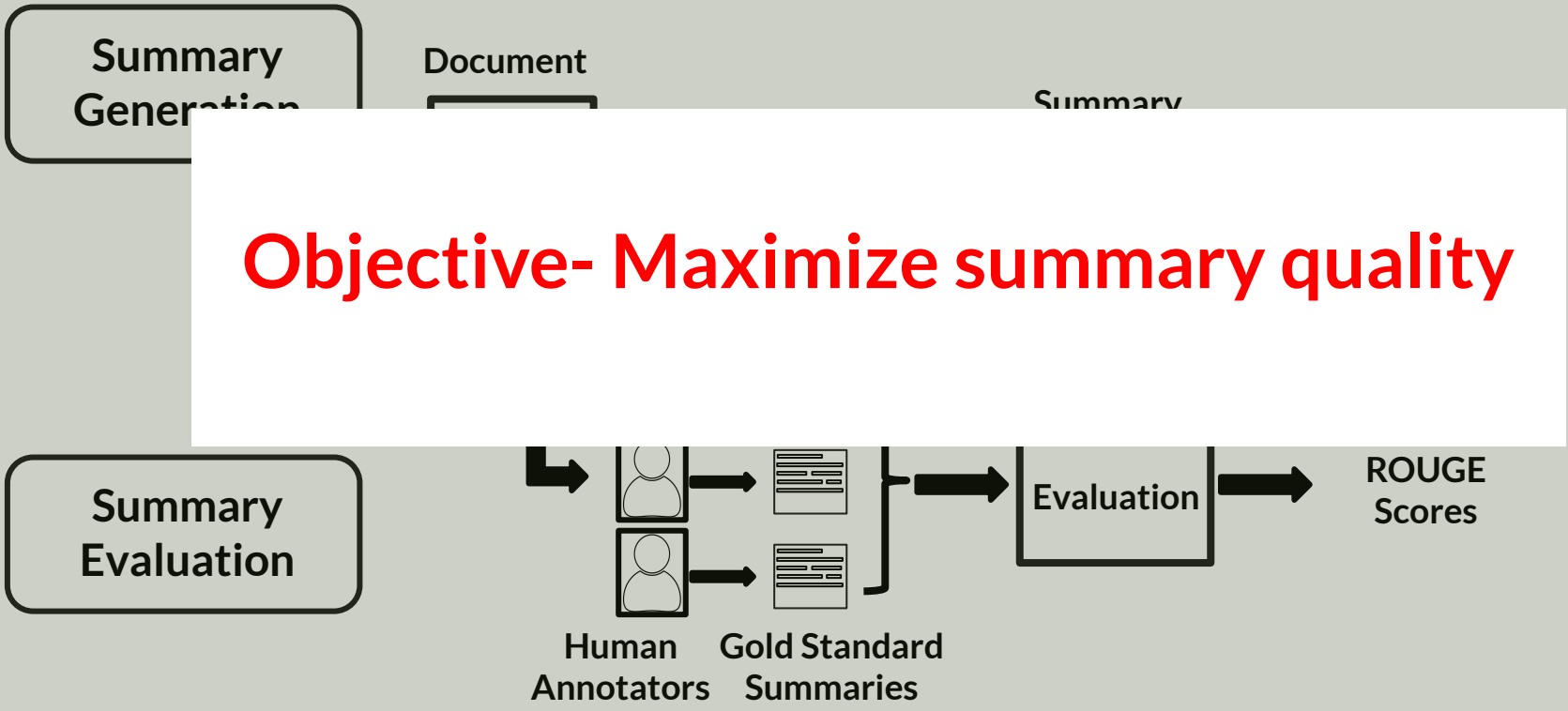
Summary Generation



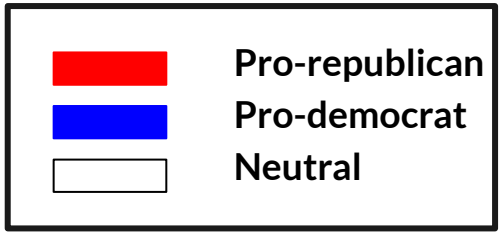
# Generic text summarization framework

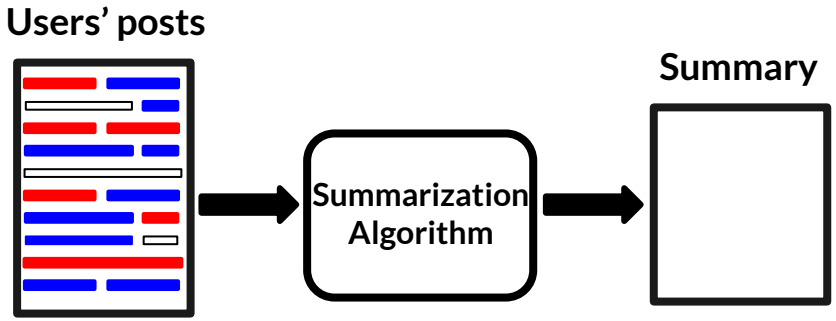


# Generic text summarization framework

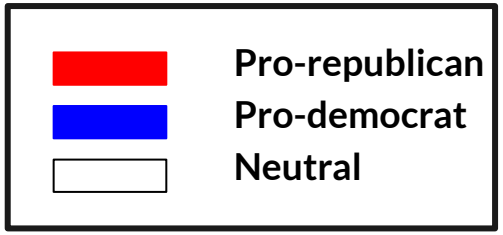


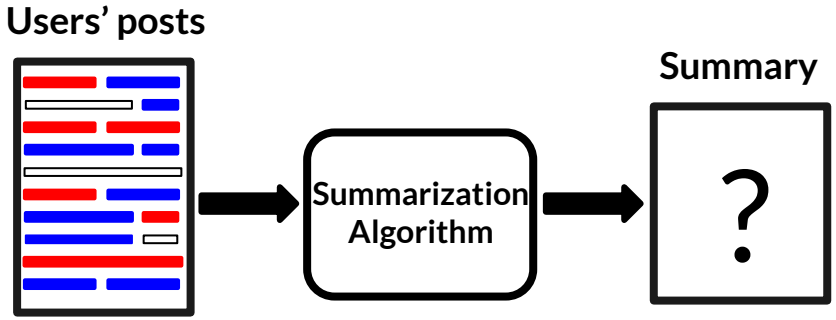
# Heterogenous online content



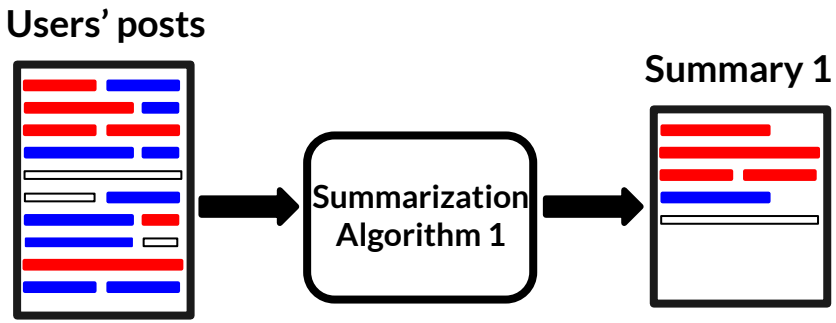
# Heterogenous online content



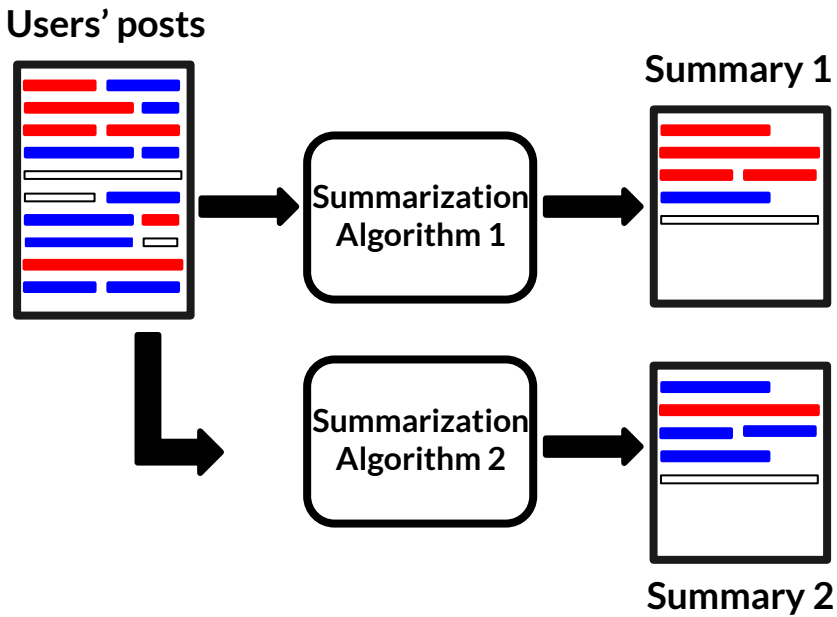




## What should be the color of the summary?

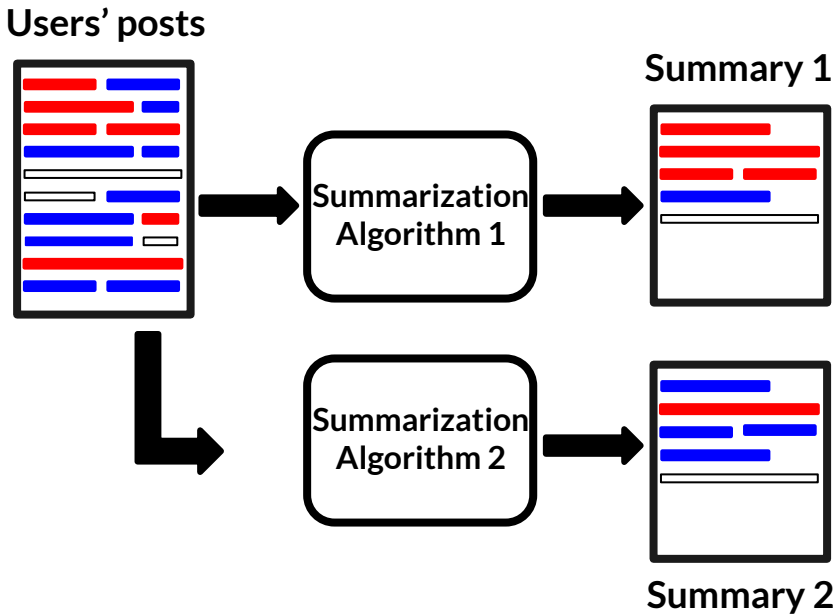
# Why does color of the summary matter?



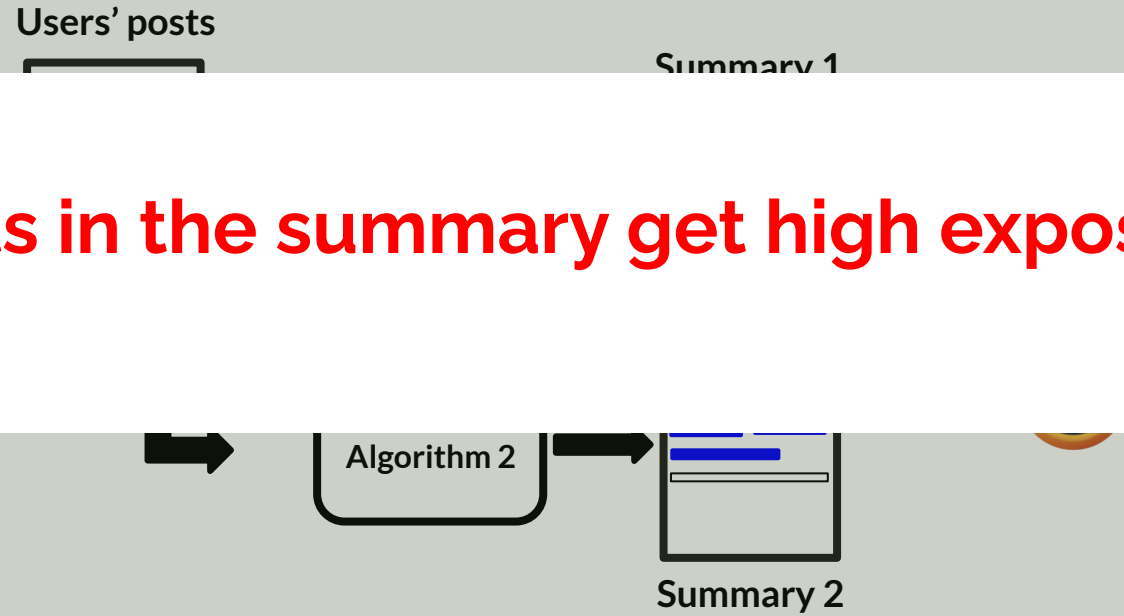
# Why does color of the summary matter?



# Why does color of the summary matter?




# Why does color of the summary matter?




**Posts in the summary get high exposure**

# Two questions

- Do the tweets written by different social groups reflect different opinions?

- Are the tweets written by different social groups of comparable textual quality?

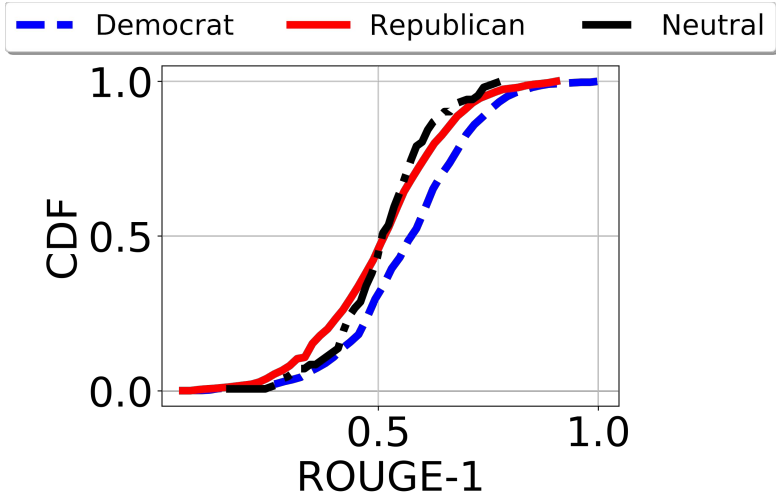
# Two questions

- Do the tweets written by different social groups reflect different opinions? **YES**

- Are the tweets written by different social groups of comparable textual quality?

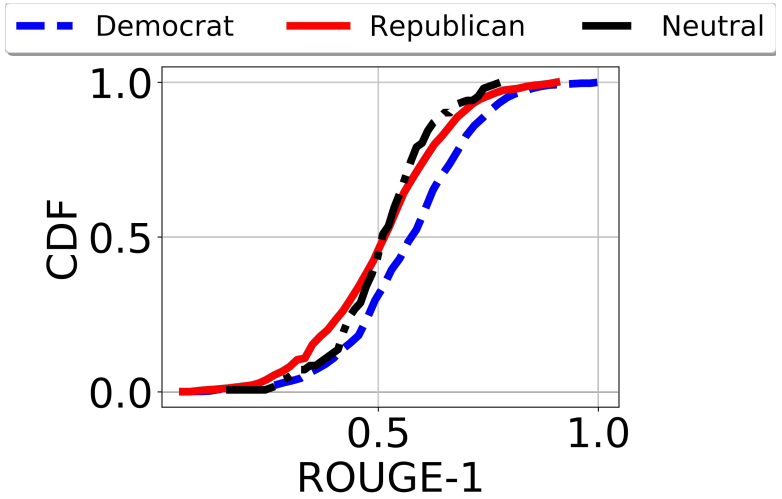
# How to measure textual quality?

- ROUGE-1 of a tweet: fraction of unigrams of the tweet present in gold standard summaries ([0, 1])



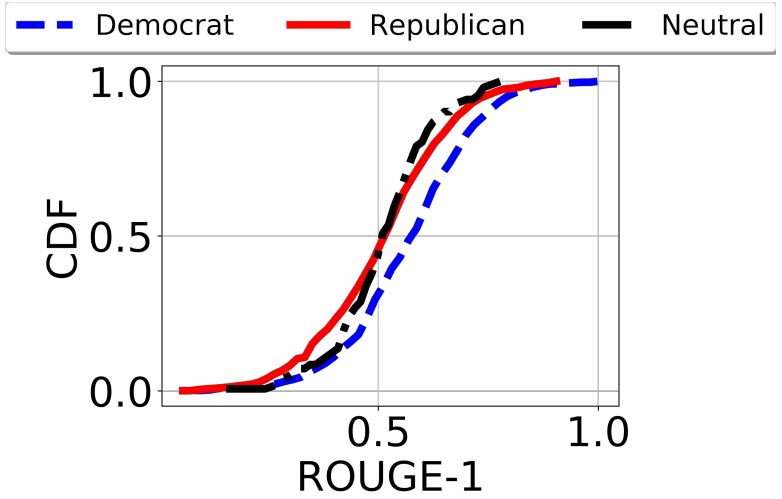
# How to measure textual quality?

- ROUGE-1 of a tweet: fraction of unigrams of the tweet present in gold standard summaries ([0, 1])
- The distributions are very similar.



# How to measure textual quality?

- ROUGE-1 of a tweet: fraction of unigrams of the tweet present in gold standard summaries ([0, 1])
- The distributions are very similar.
- Averages of ROUGE-1 in each class are similar.

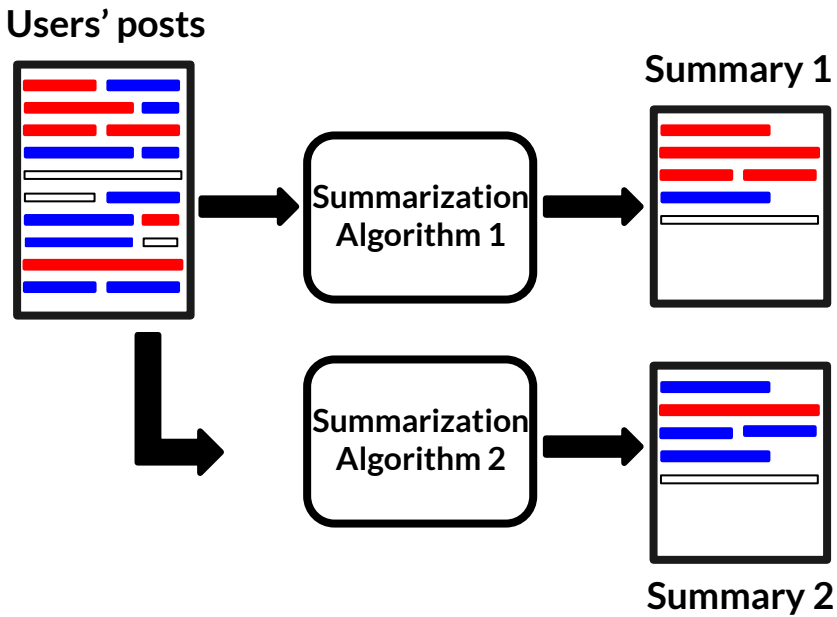


# Two questions

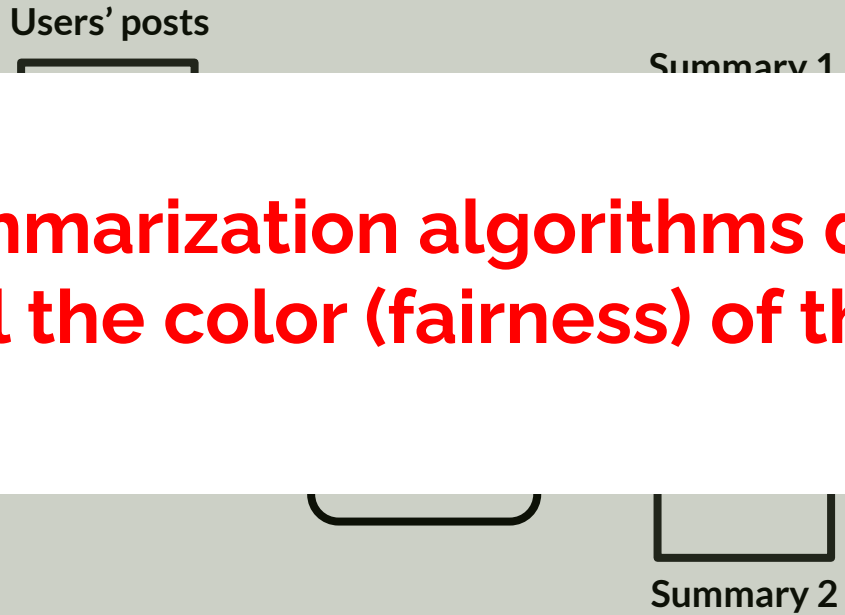
- Do the tweets written by different social groups reflect different opinions? **YES**

- Are the tweets written by different social groups of comparable textual quality? **YES**

# What should be the color of the summary?

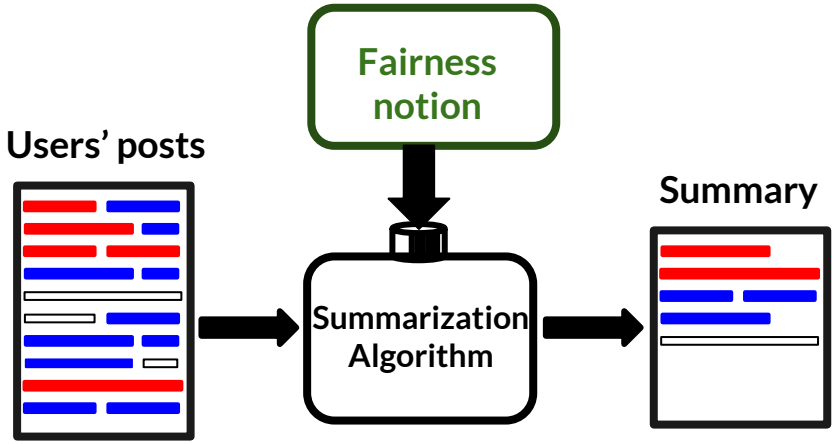


# What should be the color of the summary?

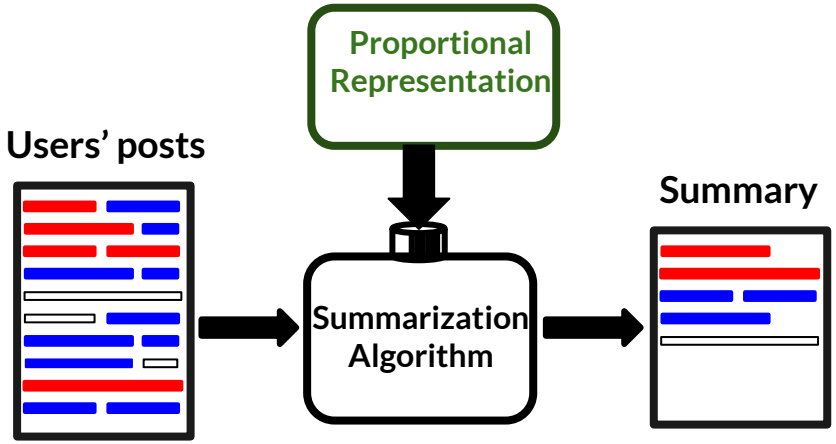



**Existing summarization algorithms do not attempt to control the color (fairness) of the summary**

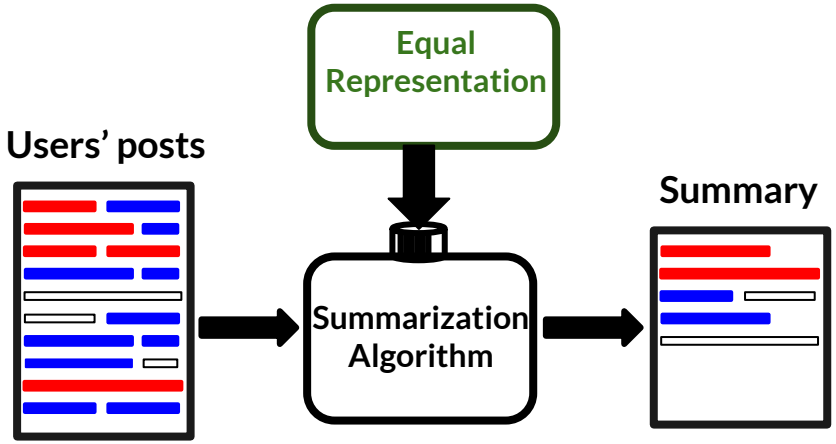
# We propose to include an additional control



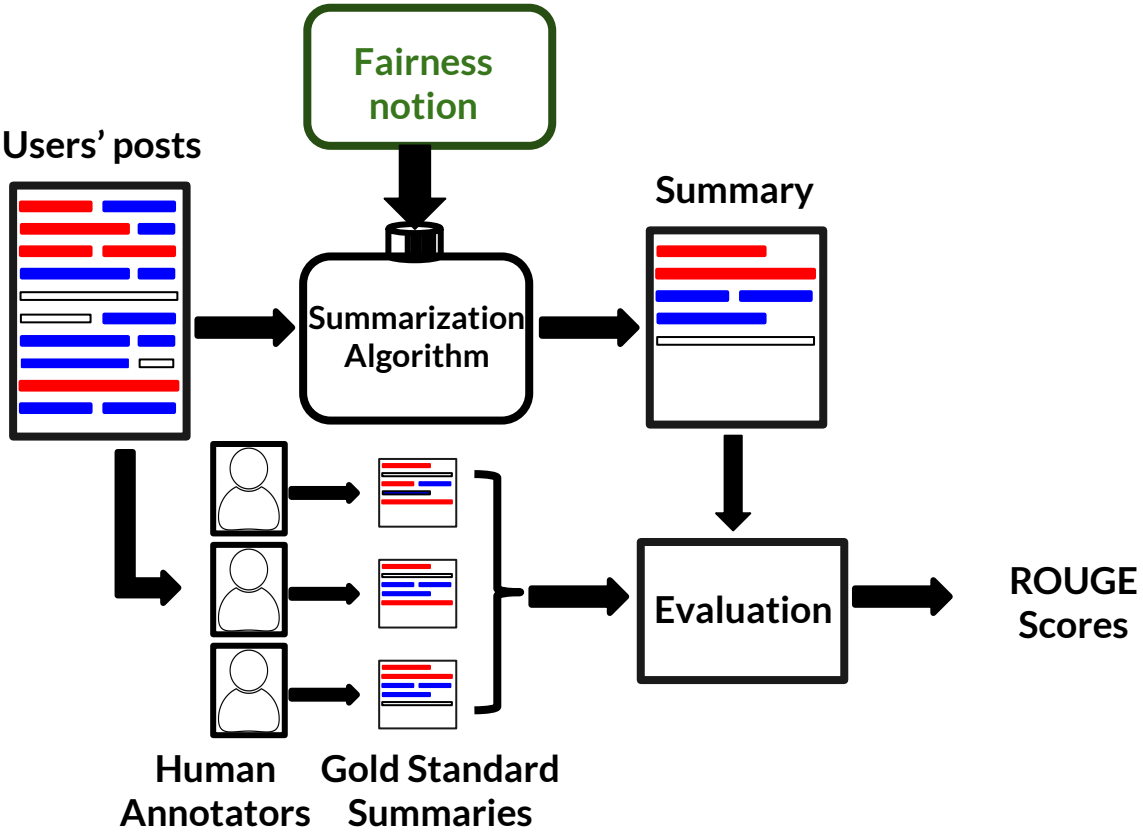
# Proportional Representation



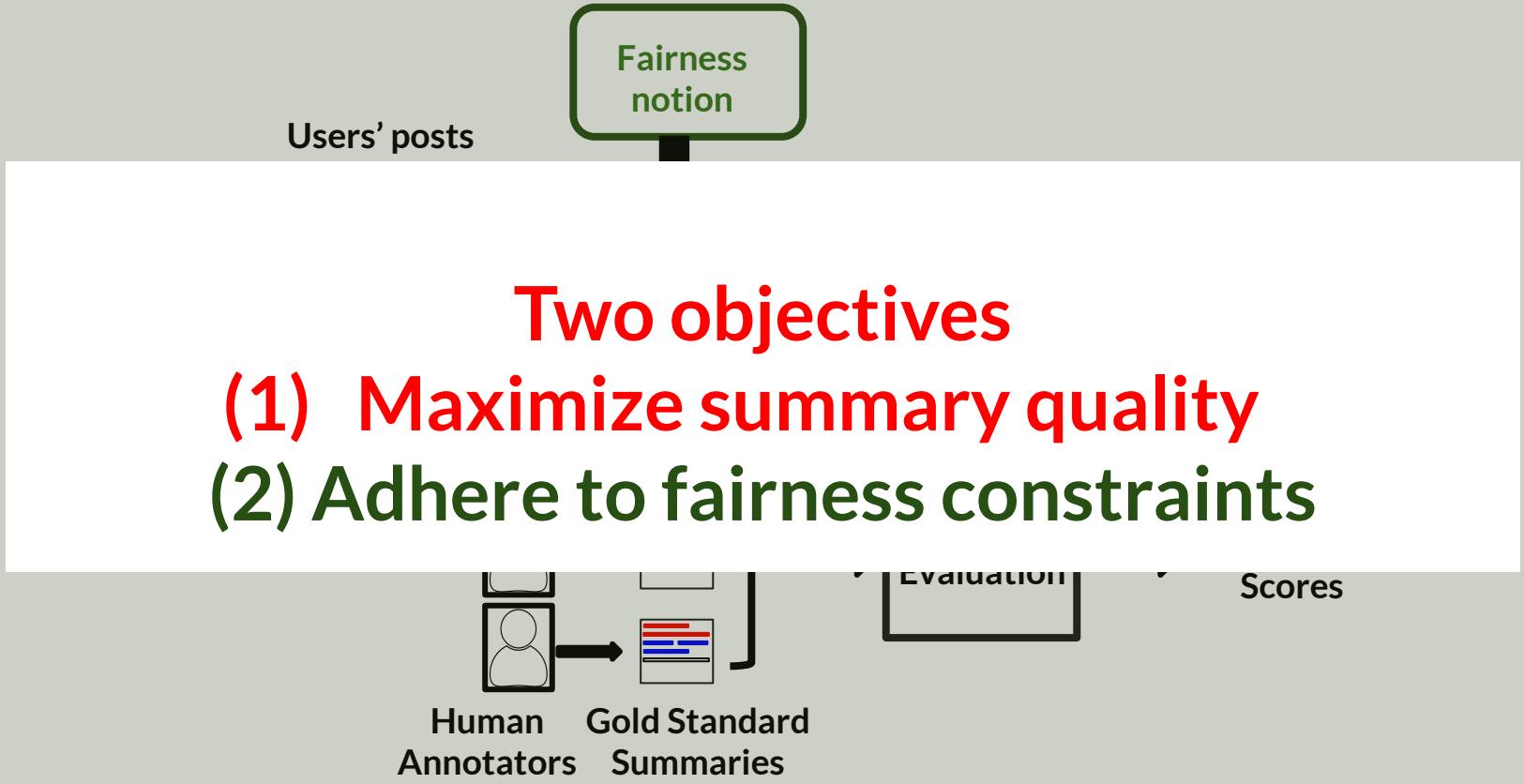
# Equal Representation



# Our framework



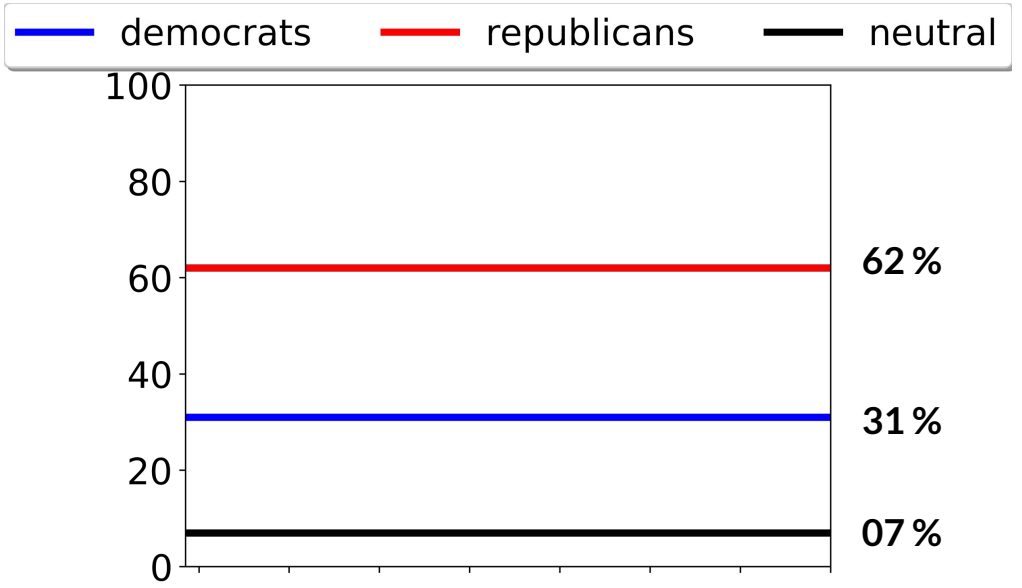
# Our framework



# Do existing algorithms produce fair summaries?

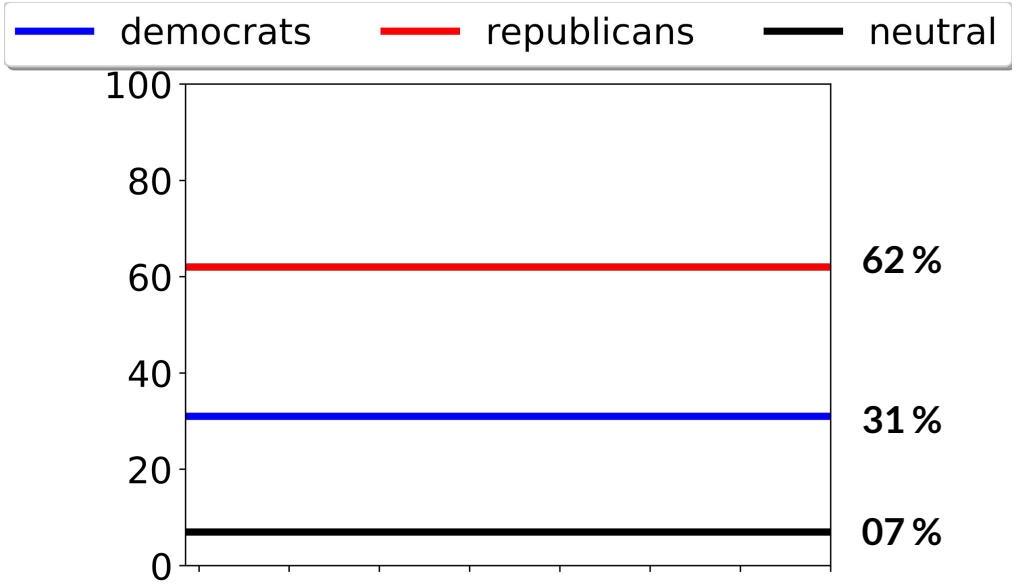
# A sample dataset

- Distributions in input text
  - pro-republican = 62 %
  - pro-democrat = 31 %
  - neutral = 7 %



# A sample dataset

- Distributions in input text
  - pro-republican = 62 %
  - pro-democrat = 31 %
  - neutral = 7 %

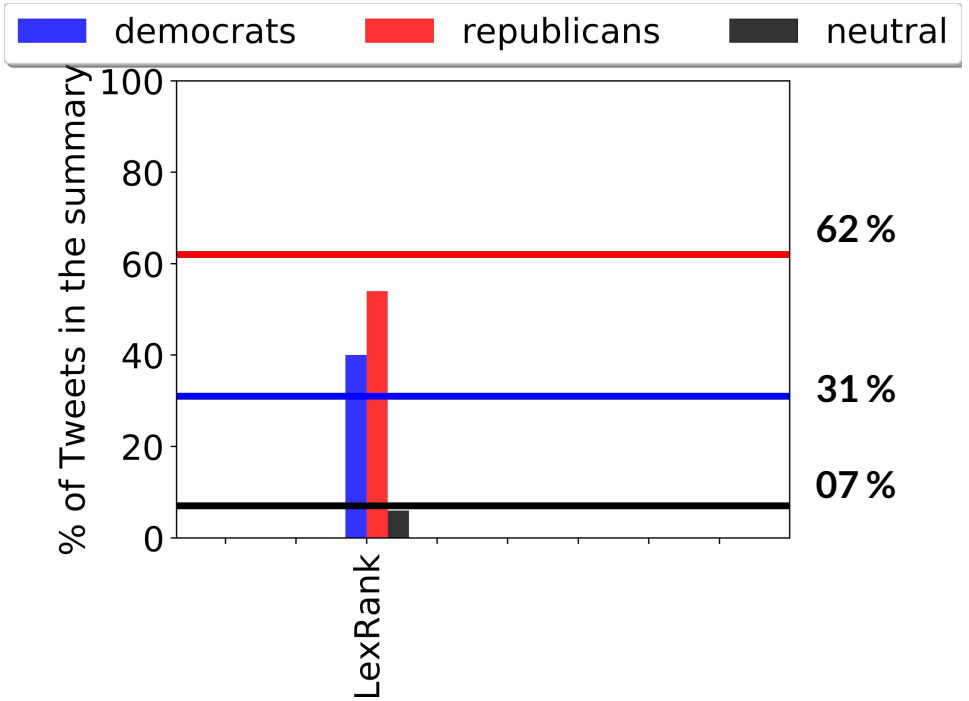



**Does existing algorithms conform to Proportional Fairness?**

# Using an existing algorithm: LexRank

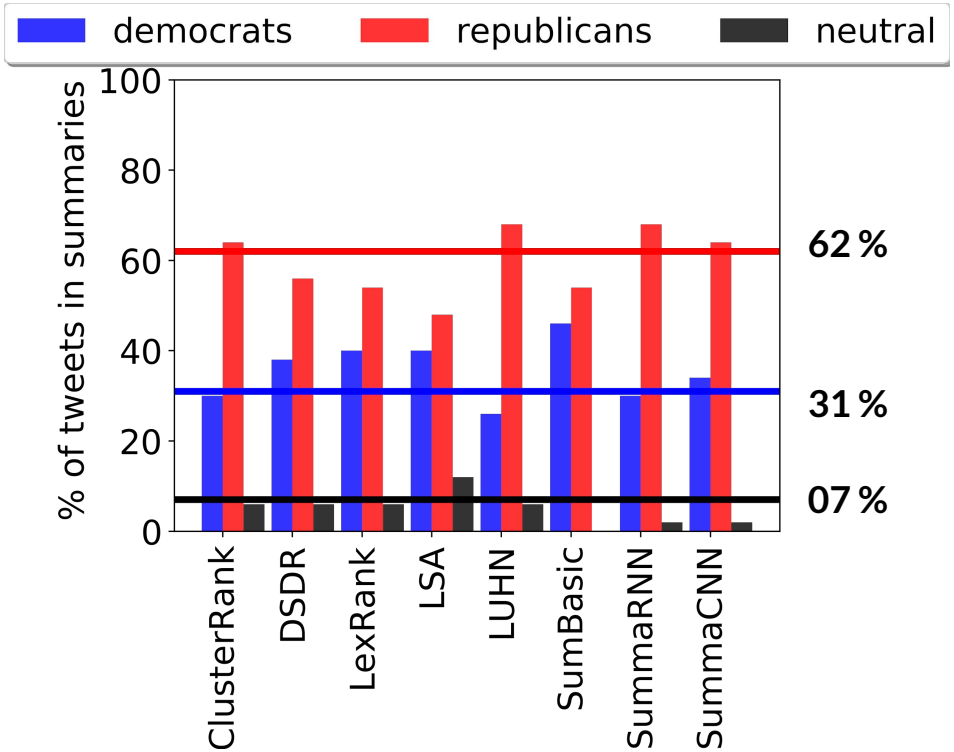
In summary produced by LexRank

- pro-democrat tweets over-represented
- pro-republicans and neutrals are under-represented



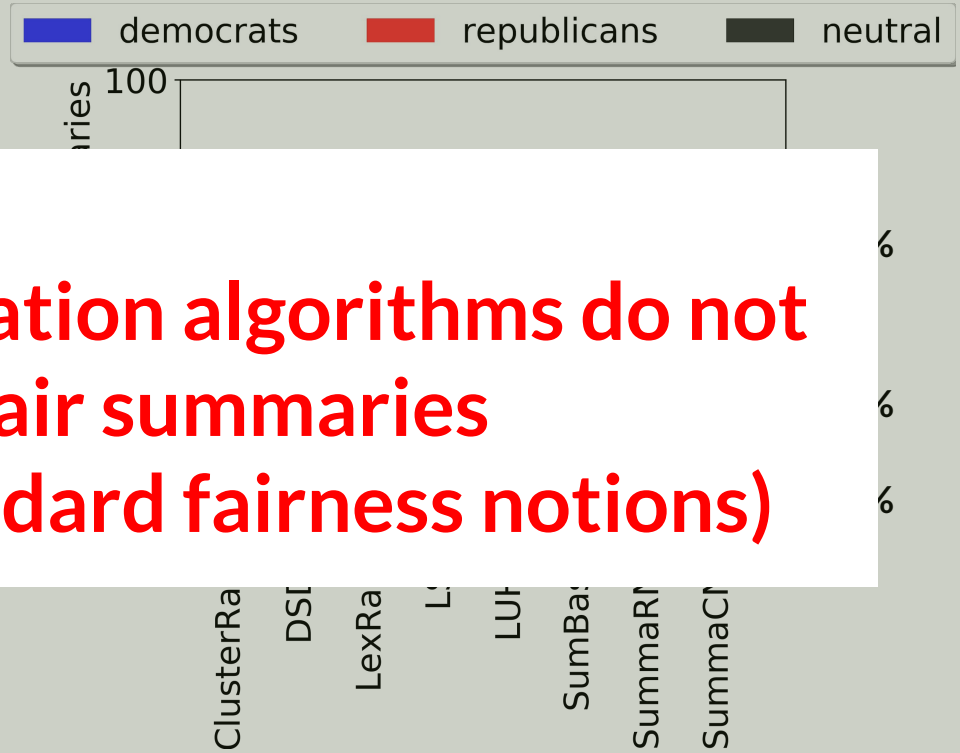
# We analyzed eight summarization algorithms

- Most of the algorithms under-represent tweets from at least 2 groups
- Similar behavior on different datasets



## We analyzed eight s… a…

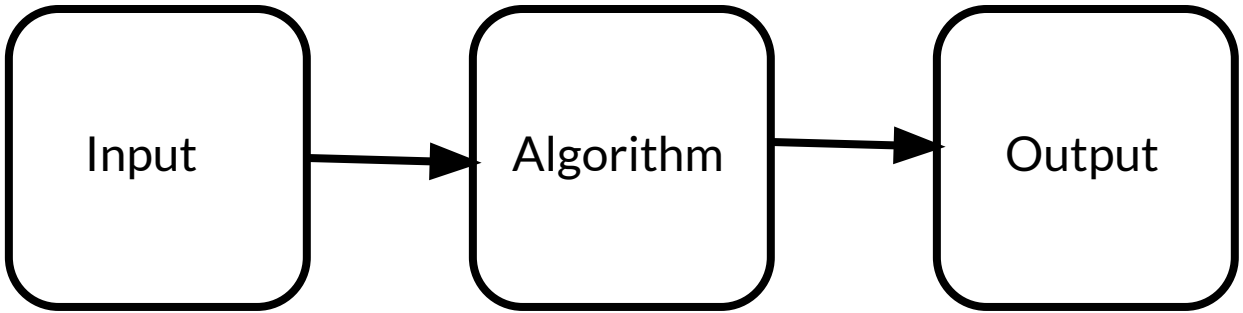
- ❖ M… u… le…
- ❖ Similar behavior on different datasets



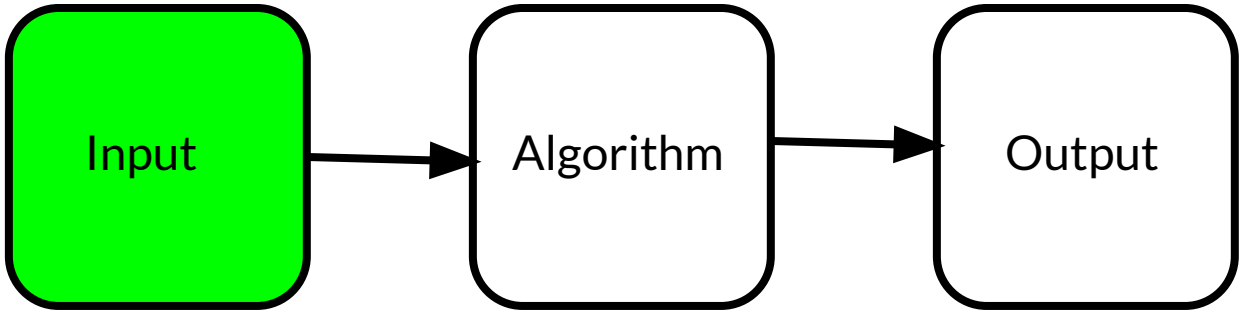

**Existing summarization algorithms do not produce fair summaries (according to standard fairness notions)**

# How to achieve fairness in summarization?

# Standard algorithmic pipeline

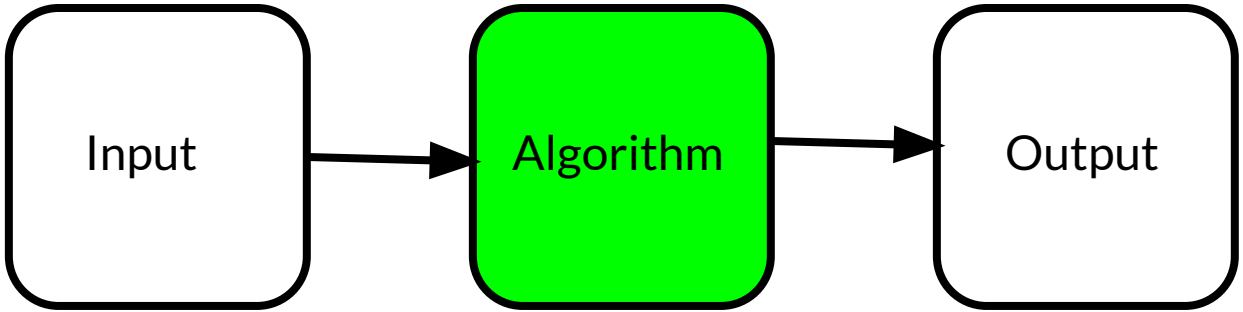


# Pre-processing based algorithm




**ClasswiseSumm**

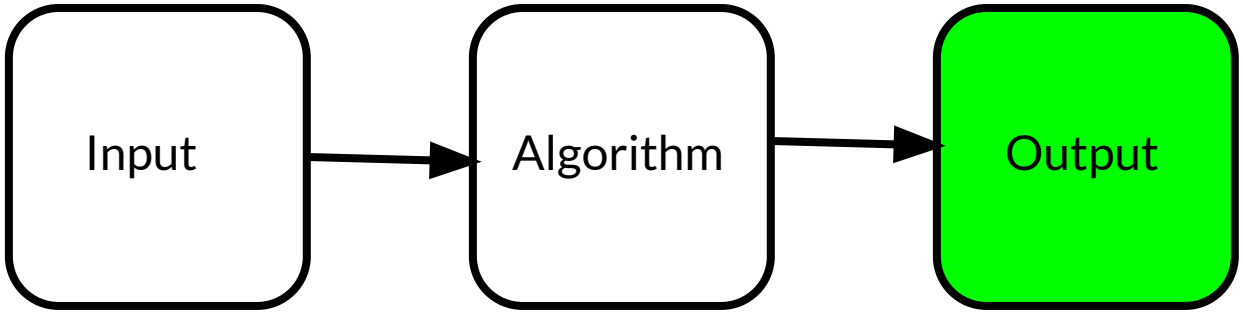
# In-processing based algorithm




**ClasswiseSumm**

**FairSumm**
**Our primary contribution**

# Post-processing based algorithm

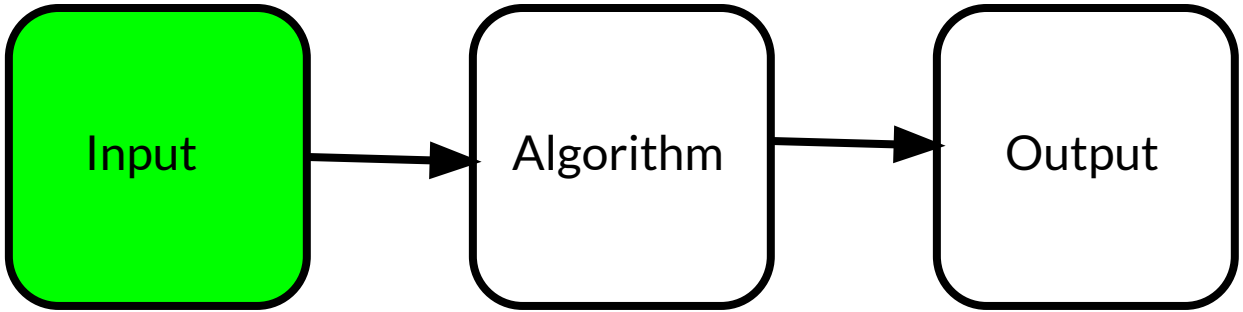



**ClasswiseSumm**

**FairSumm**
**Our primary contribution**

**ReFaSumm**

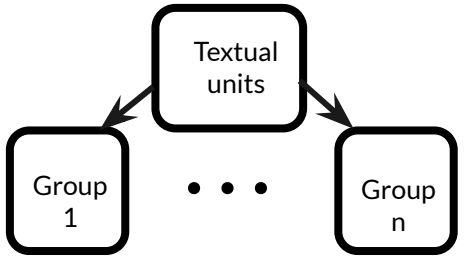
# Pre-processing based algorithm




**ClasswiseSumm**

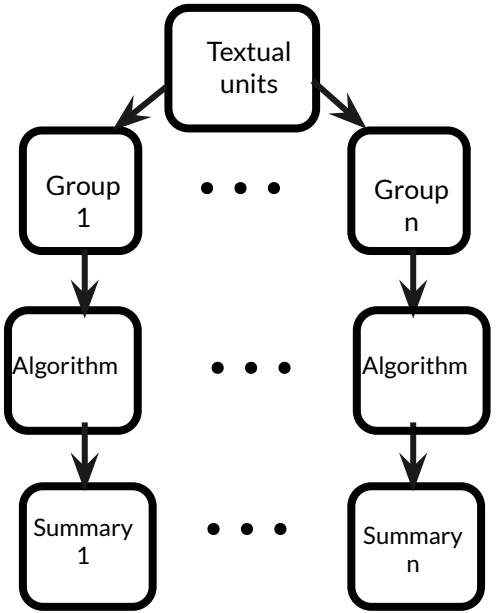
# ClasswiseSumm

1. Split the tweets as per their groups.



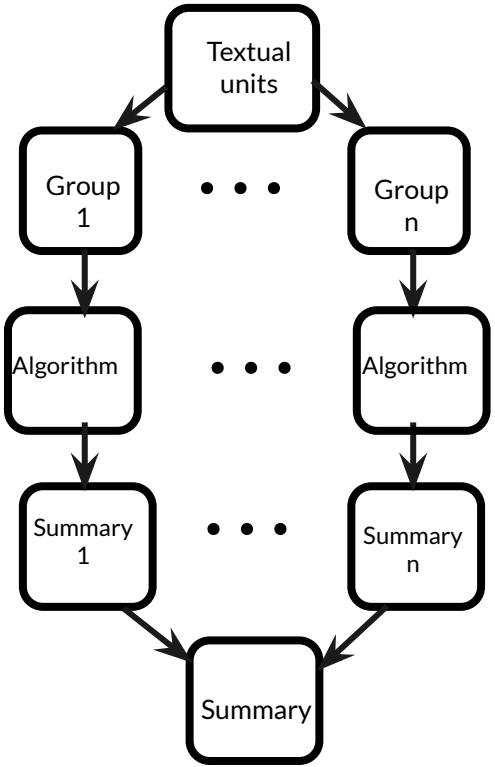
# ClasswiseSumm

1. Split the tweets as per their groups.
2. Summarize each group individually.

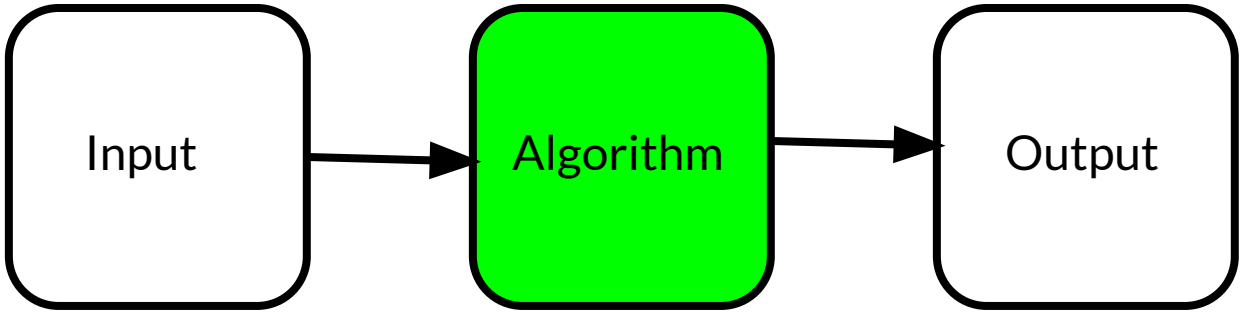


# ClasswiseSumm

1. Split the tweets as per their groups.
2. Summarize each group individually.
3. Combine the summaries to generate the final summary.



# In-processing based algorithm




**FairSumm**
**Our primary contribution**

# FairSumm

- Two important aspects of a good summary
  - Coverage: amount of information covered (*L(S)*) should be high
  - Diversity: avoid redundancy or reward diverse information (*R(S)*)

# FairSumm

- Two important aspects of a good summary
  - Coverage: amount of information covered ($L(S)$) should be high
  - Diversity: avoid redundancy or reward diverse information ($R(S)$)
  - Maximize, $F(S) = \lambda_1 * L(S) + \lambda_2 * R(S)$ → Monotonic non-decreasing submodular function

# FairSumm

- Two important aspects of a good summary
  - Coverage: amount of information covered ($L(S)$) should be high
  - Diversity: avoid redundancy or reward diverse information ($R(S)$)
  - Maximize, $F(S) = \lambda_1 * L(S) + \lambda_2 * R(S)$
- Subjected to the fairness constraint
  - Summary $S$ should satisfy the underlying notion of fairness.

Constrained monotonic non-decreasing submodular function

# FairSumm

- Two important aspects of a good summary
  - Coverage: amount of information covered ($L(S)$) should be high
  - Diversity: avoid redundancy or reward diverse information ($R(S)$)
  - Maximize, $F(S) = \lambda_1 * L(S) + \lambda_2 * R(S)$
- Subjected to the fairness constraint
  - Summary $S$ should satisfy the underlying notion of fairness.
- We solve the above by algorithm proposed by Du et al. to get the final summary. [1]

Constrained monotonic non-decreasing submodular function

[1] Du, MFB Nan, Yingyu Liang, and L. Song. "Continuous-time influence maximization for multiple items." *CoRR, abs/1312.2164* (2013).

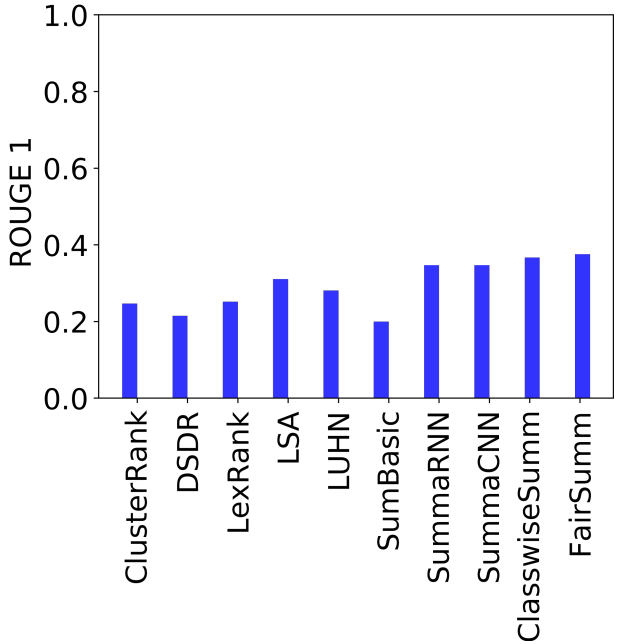
# How good are our fair summarization algorithms?

# Datasets*

- Claritin: opinions of male & female users on a drug Claritin
  - Gender
- US-Election: pro-dem, pro-rep and neutral tweets on US Presidential election 2016
  - Political leaning
- MeToo: opinions of male & female users on MeToo movement
  - Gender

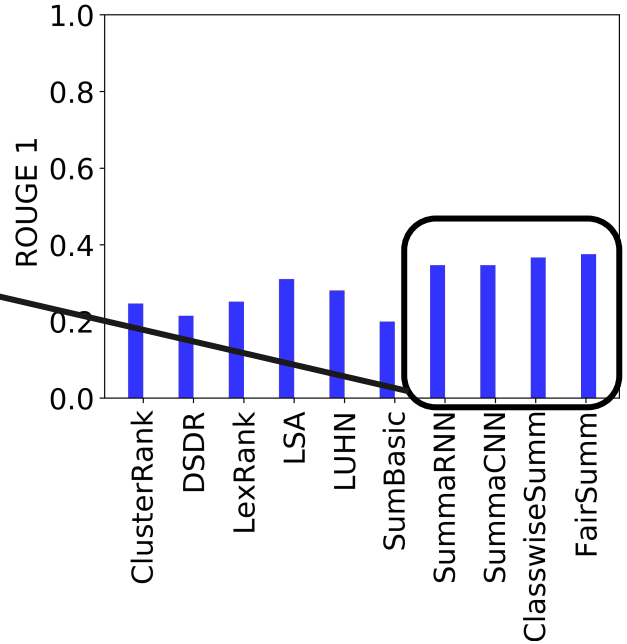
* Available at https://github.com/ad93/FairSumm

# How good are the summaries produced by our algorithms?



# How good are the summaries produced by our algorithms?

- FairSumm and ClasswiseSumm are as good as the state-of-the-art algorithms.



# How good are the summaries produced by our algorithms?

**Introducing fairness does not compromise on summary quality**

# Discussion and future work

- What are acceptable notions of fairness in the context of summarization?
- Who decides the notion of fairness — consumer / producer / platform?
- Fairness w.r.t. Demographics, or fairness w.r.t. Opinions?

# Fairness w.r.t demographics or w.r.t. opinions?

- Distribution of opinions $=$ demographic distribution.

# Fairness w.r.t demographics or w.r.t. opinions?

- Distribution of opinions $\neq$ demographic distribution.
- Perspective based or aspect based summarizations[1] might be the way ahead.

[1] Rudra, Koustav, et al. "Identifying sub-events and summarizing disaster-related information from microblogs." *ACM SIGIR 2018*

# PART II

# Framework for Auditing Recommendation Systems

-IEEE INFOCOM 2019

# Recommendation systems

- **Recommendations are ubiquitous on web.**
- Nudge users to make certain choices.
- Reason for financial benefits to user satisfaction.

# Interstellar (2014)

PG-13 | 2h 49min | Adventure, Drama, Sci-Fi | 7 November 2014 (USA)

8.6/10 1,241,484 Rate This



A team of explorers travel through a wormhole in space in an attempt to ensure humanity's survival.

**Director:** Christopher Nolan
**Writers:** Jonathan Nolan, Christopher Nolan
**Stars:** Matthew McConaughey, Anne Hathaway, Jessica Chastain | See full cast & crew »

+ Add to Watchlist

74 Metascore From metacritic.com | Reviews 3,214 user | 717 critic | Popularity 158 (28)

## More Like This






Add to Watchlist

Next »

**The Matrix** (1999)

R Action | Sci-Fi

8.7/10

A computer hacker learns from mysterious rebels about the true nature of his reality and his role in the war against its controllers.

**Directors:** Lana Wachowski, Lilly W...

# Recommendation systems

- Recommendations are ubiquitous on web.
- **Nudge users to make certain choices.**
- Reason for financial benefits to user satisfaction.

# Recommendation systems

- Recommendations are ubiquitous on web.
- Nudge users to make certain choices.
- **Reason for financial benefits to user satisfaction.**

# Do we know how are we recommended?

# How are we recommended?

- Probably **NO**
- Algorithmic details are unknown.
- User-item interactions are unknown.

# Inadvertent consequences

- Filter Bubble
- Information Segregation

# Potential solution

**Algorithmic auditing** is a research design that has shown promise in diagnosing the unwanted consequences of algorithmic systems.

Sandvig, Christian, et al. "Auditing algorithms: Research methods for detecting discrimination on internet platforms." *Data and discrimination: converting critical concerns into productive inquiry* 22 (2014).

# Algorithmic auditing

- **Why is it the need of the hour?**
- Prerequisites for auditing
- Issues for a third-party auditor

# Algorithmic auditing

- Why is it the need of the hour?
- **Prerequisites for auditing**
  - **Algorithm**
  - **Domain of the algorithm**
  - **User-item interactions**
- Issues for a third-party auditor

# Algorithmic auditing

- Why is it the need of the hour?
- Prerequisites for auditing
- **Issues for a third-party auditor (outside the organization)**
  - **Unavailability of the algorithms**
  - **Unavailability of the user-item interactions**

# Algorithmic auditing

- Why is it the need of the hour?
- Prerequisites for an auditor
- Issues for a third party auditor

**RQ: How does a third party audit a recommendation system without having all the prerequisites?**

# Our contributions

- A Network- framework for third party auditing of Recommendation Systems.

- Useful applications of the framework to explore
  - ➢ Information Segregation

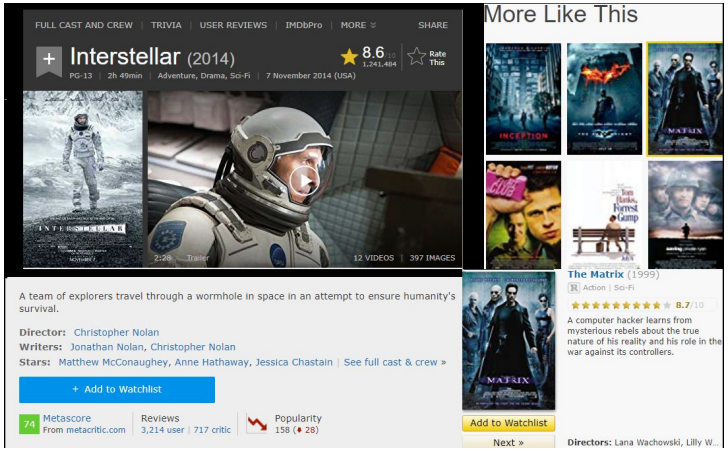
  due to the intervention of such systems.

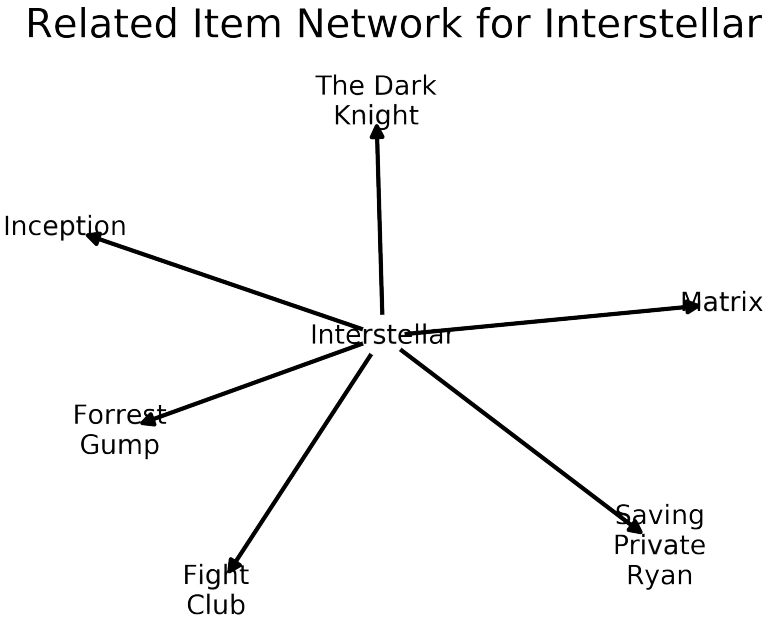
# Framework for auditing recommendation systems

- A directed (un)weighted network,
  - Each node is an item.
  - Directed edge $i \longrightarrow j$, means $j$ is recommended on the page of $i$.

- We name this network as a **“Related Item Network (RIN)”**.

# Related Item Network (RIN)



# Features of Related Item Network

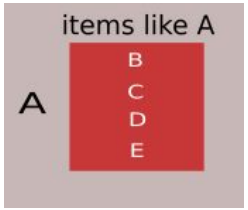
- **Different from item-item similarity graph.**

- All edges can be weighted
  - Sim($i$, $j$)
  - Rank of $j$ in the recommendation list of $i$.

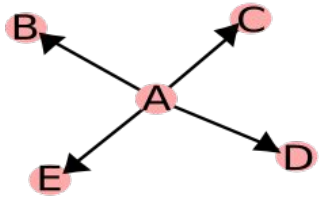
- User browsing can be modelled as random walks.

# User browsing on related item networks (RINs)

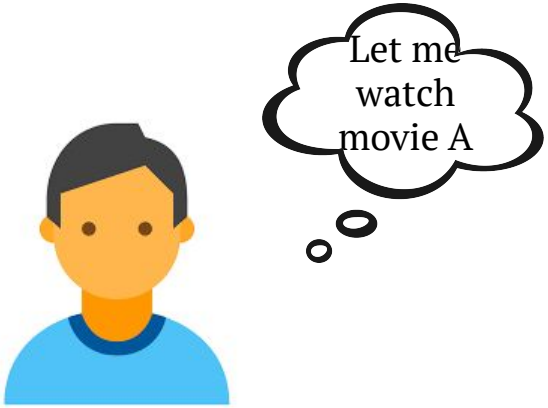
RANDOM USER

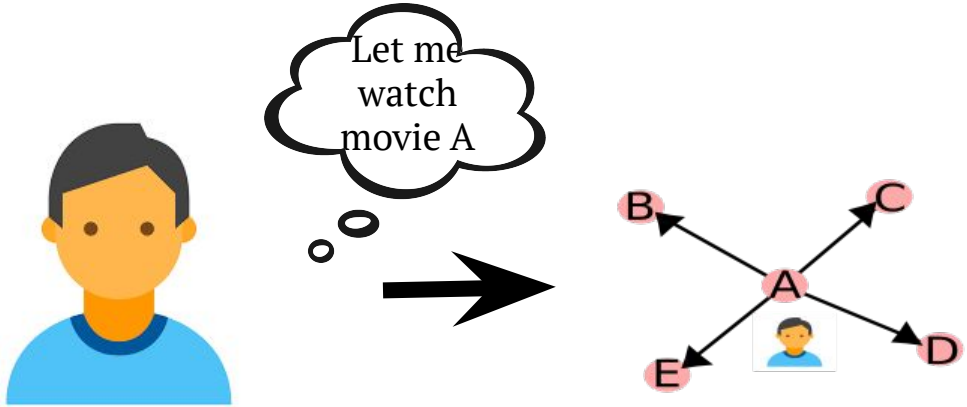


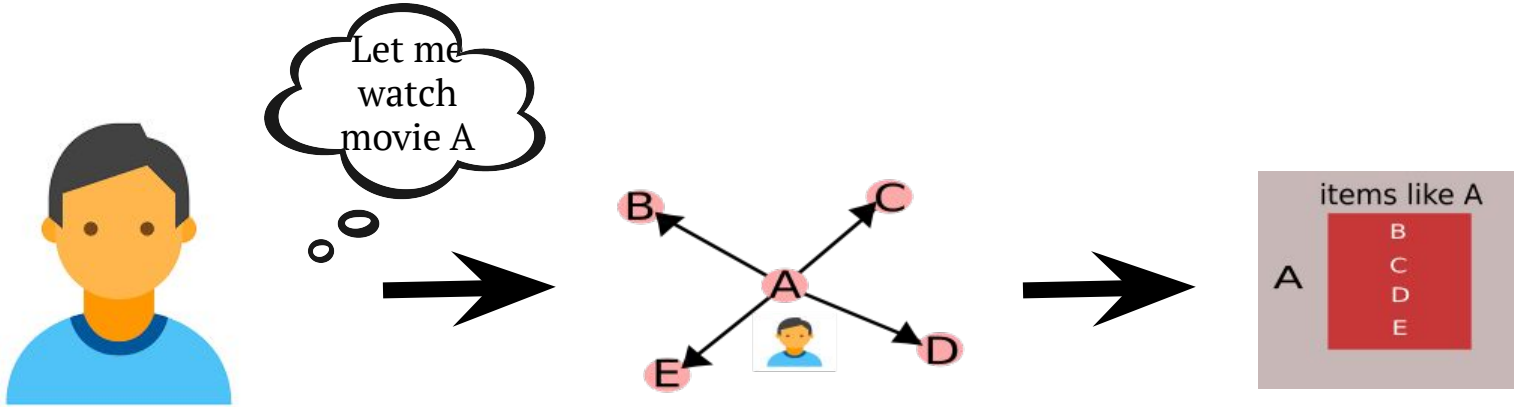

RECOMMENDATIONS ON WEB_PAGE OF A

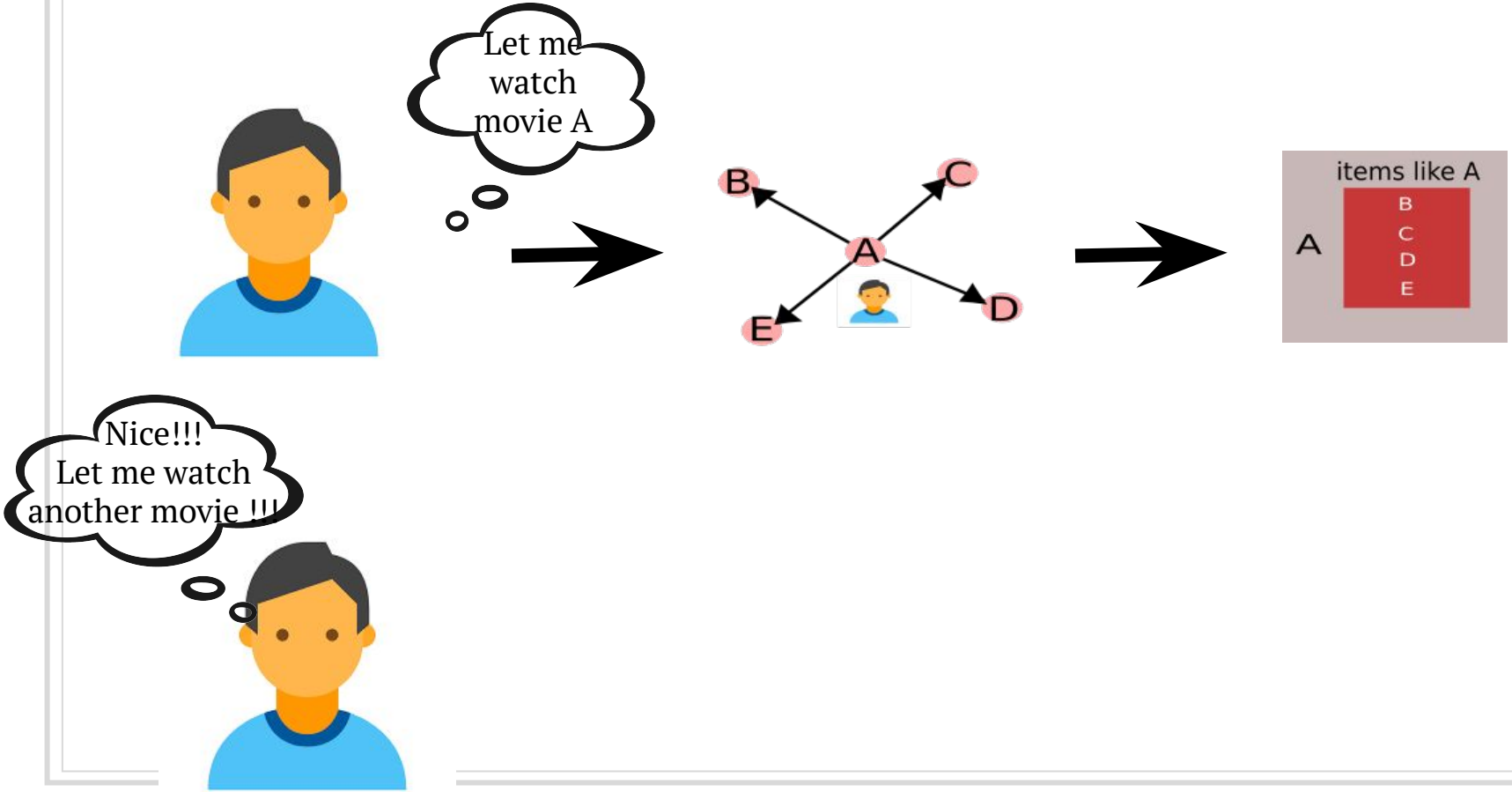


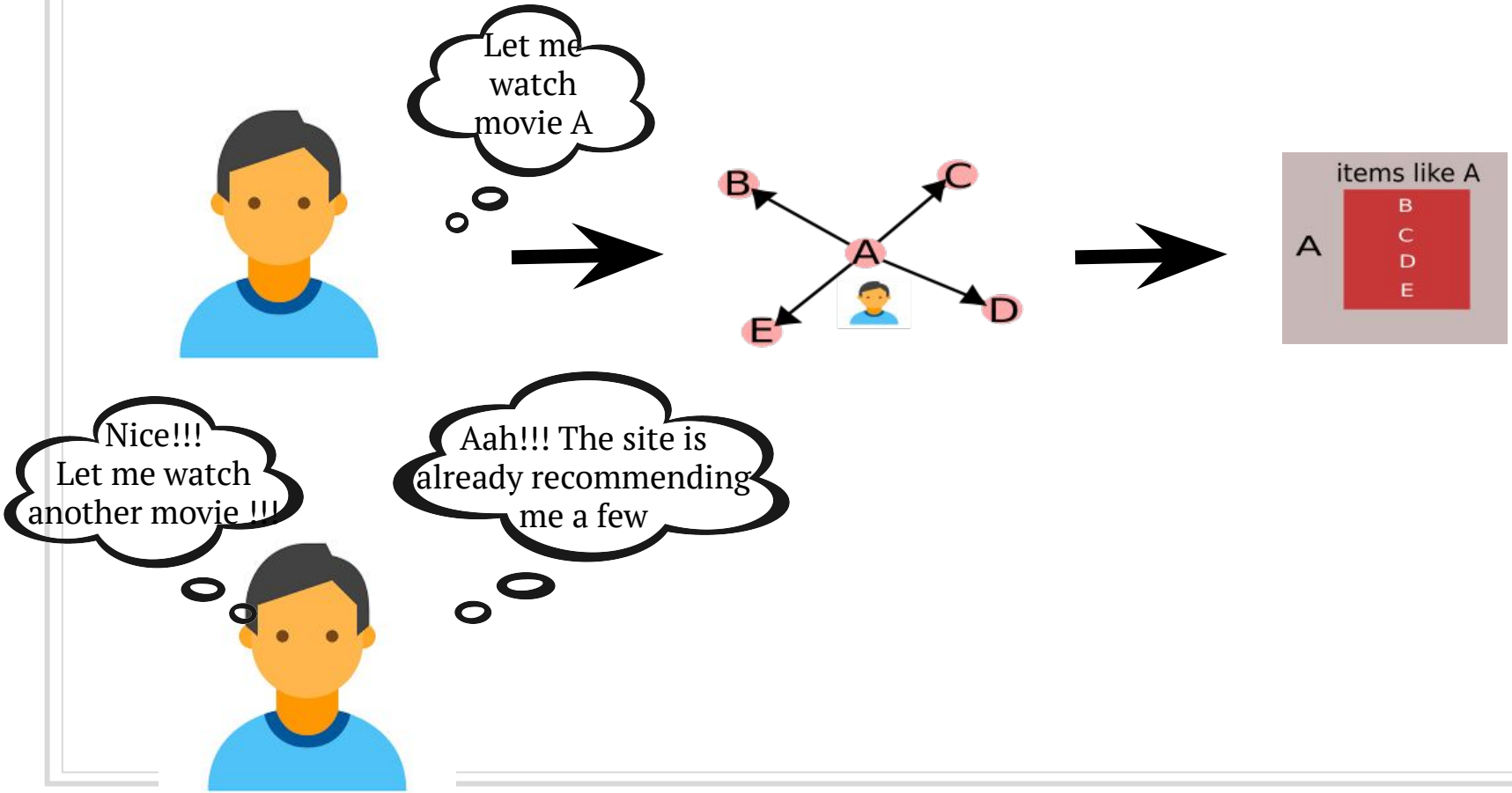

Corresponding RIN

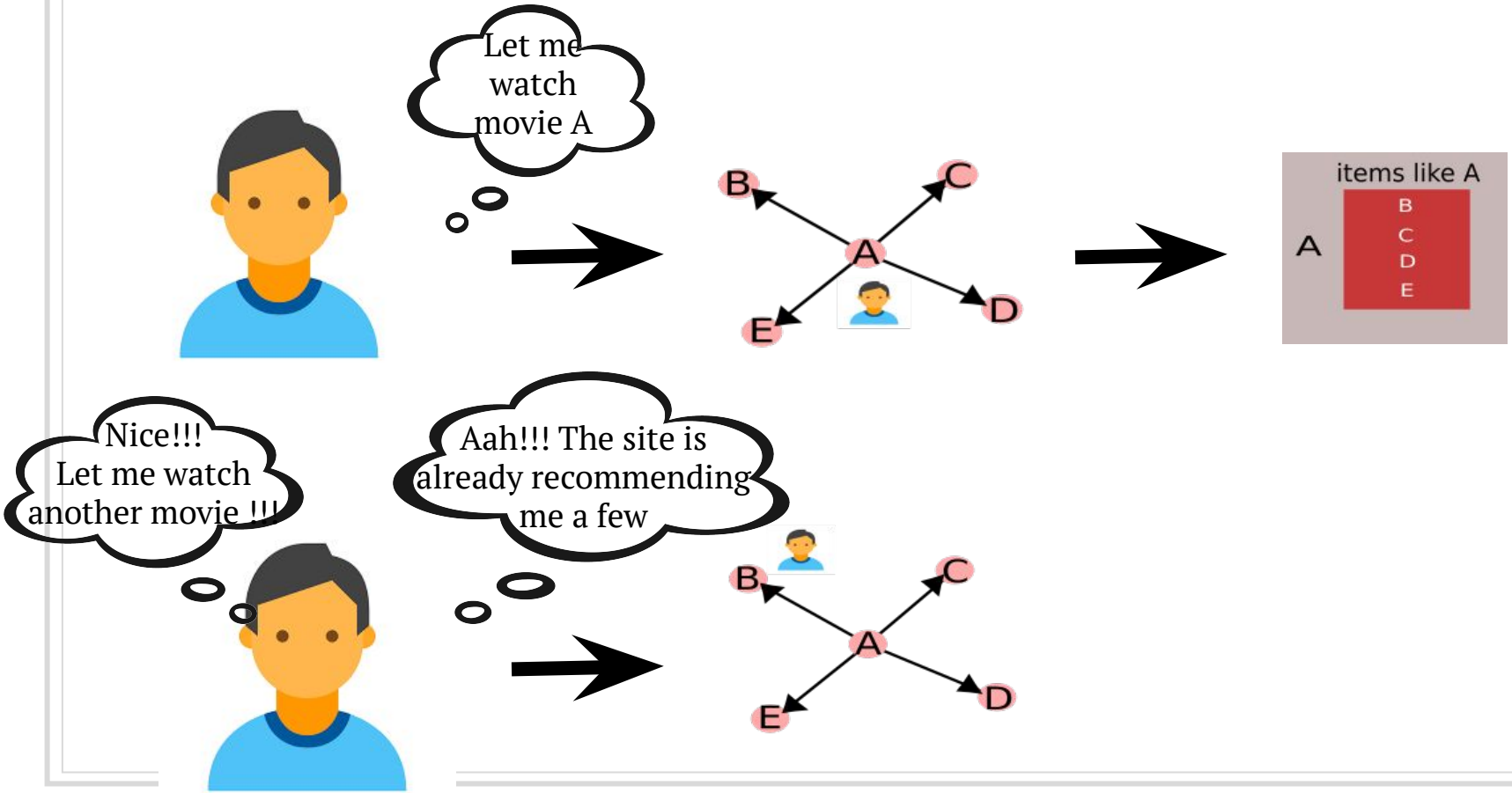

Let me watch movie A

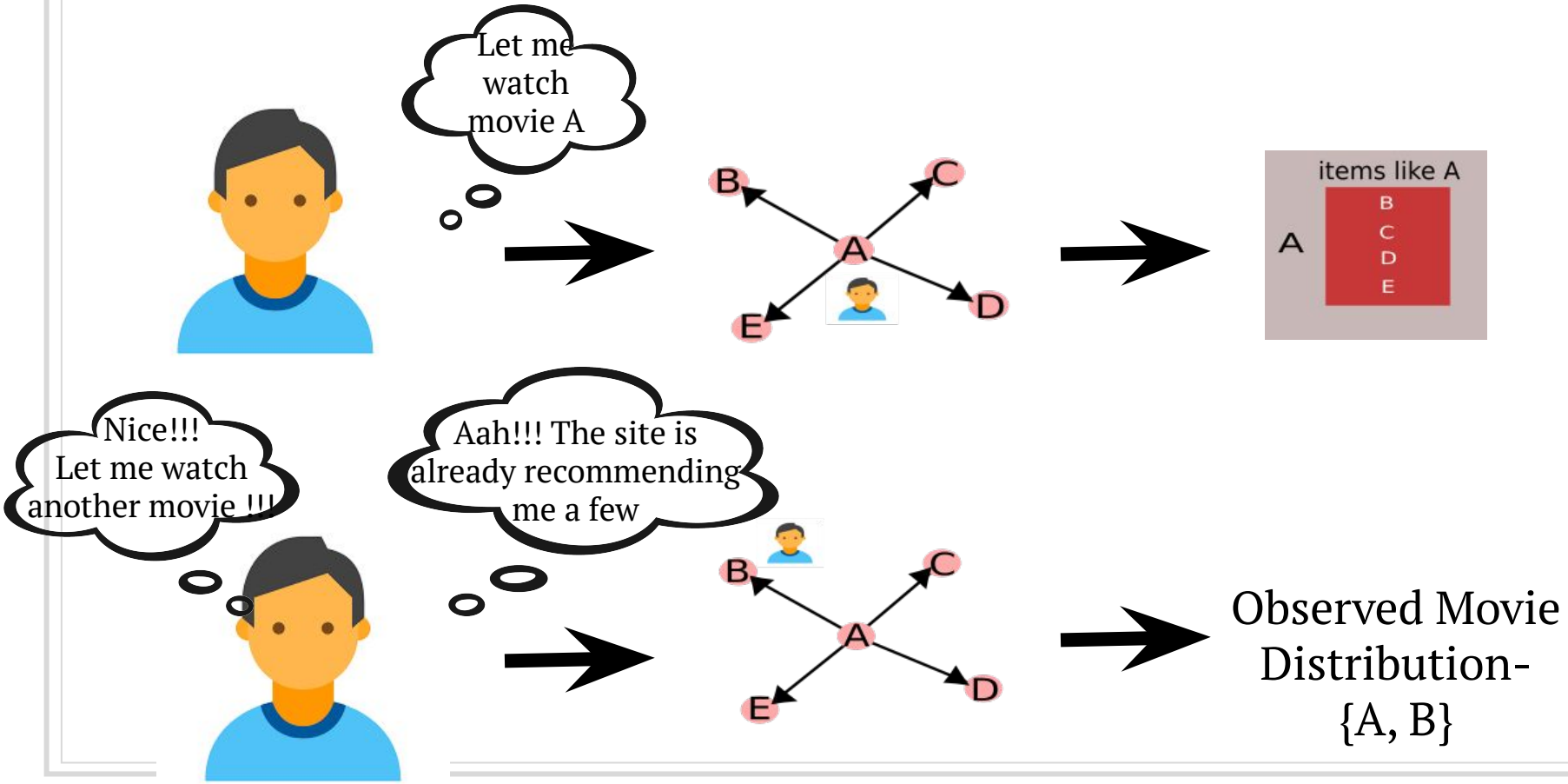

Let me watch movie A
B
C
A
D
E

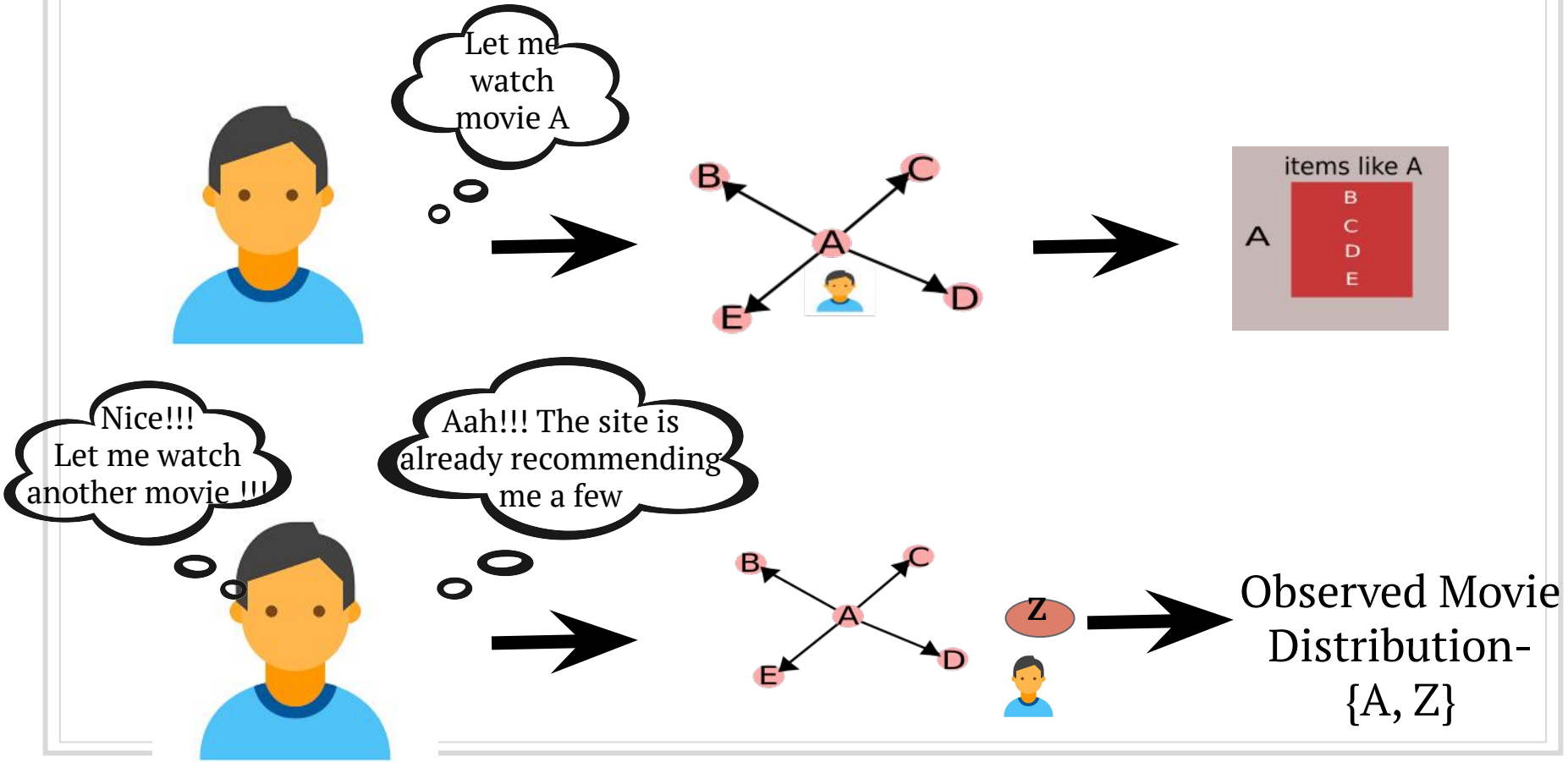

Let me watch movie A
B
C
A
E
D
items like A
A
B
C
D
E


Let me watch movie A
B
C
A
D
E
items like A
A
B
C
D
E
Nice!!! Let me watch another movie !!!


Let me watch movie A
B
C
A
E
D
items like A
A
B
C
D
E
Nice!!! Let me watch another movie !!!
Aah!!! The site is already recommending me a few


Let me watch movie A
B
C
A
D
E
items like A
A
B
C
D
E
Nice!!! Let me watch another movie !!!
Aah!!! The site is already recommending me a few
B
C
A
D
E


Let me watch movie A
B
C
A
E
D
items like A
A
B
C
D
E
Nice!!! Let me watch another movie !!!
Aah!!! The site is already recommending me a few
B
C
A
E
D
Observed Movie Distribution- {A, B}


Let me watch movie A
B
C
A
E
D
items like A
A
B
C
D
E
Nice!!! Let me watch another movie !!!
Aah!!! The site is already recommending me a few
B
C
A
E
D
Z
Observed Movie Distribution- {A, Z}

# Data collections

- Data were collected by a Breadth First Search crawler.
- We seeded the crawler with an initial movie.
- The recommendations were scraped.


172K

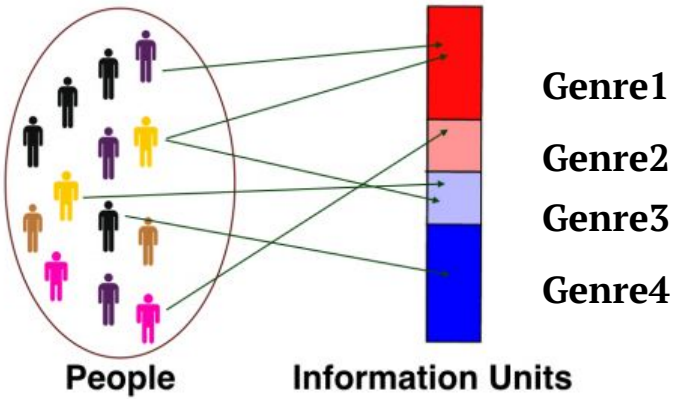

24K


2K

# Information Segregation in Recommendation Systems

# Information Segregation

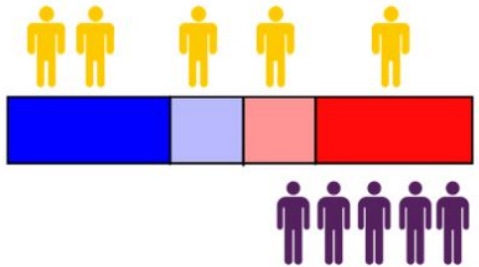
- **It is the property of the process of opinion formation in the society.**
- An *m*-dimensional information space, where
  - Information units are different genres.
  - Each movie is an information source.




P. Dandekar, A. Goel, and D. T. Lee, “Biased assimilation, homophily, and the dynamics of polarization,” PNAS, 2013

# Measure for Information Segregation

- **Evenness:** Captures how uniformly different groups are exposed to different information units.



**Yellow Group**

**Purple Group**



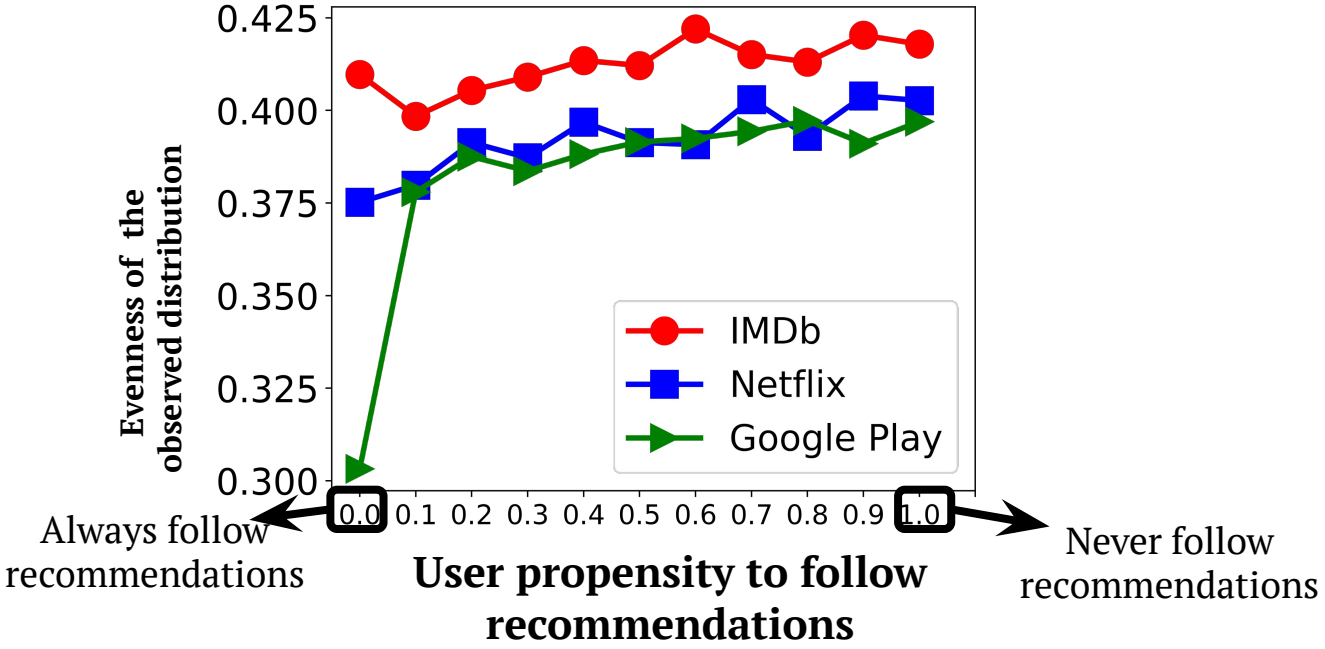
Yellow group is more evenly distributed than purple group.

A. Chakraborty, M. Ali, S. Ghosh, N. Ganguly, and K. P. Gummadi, “On quantifying knowledge segregation in society,” FATREC Workshop on Responsible Recommendation, 2017.

# Experimental Setup

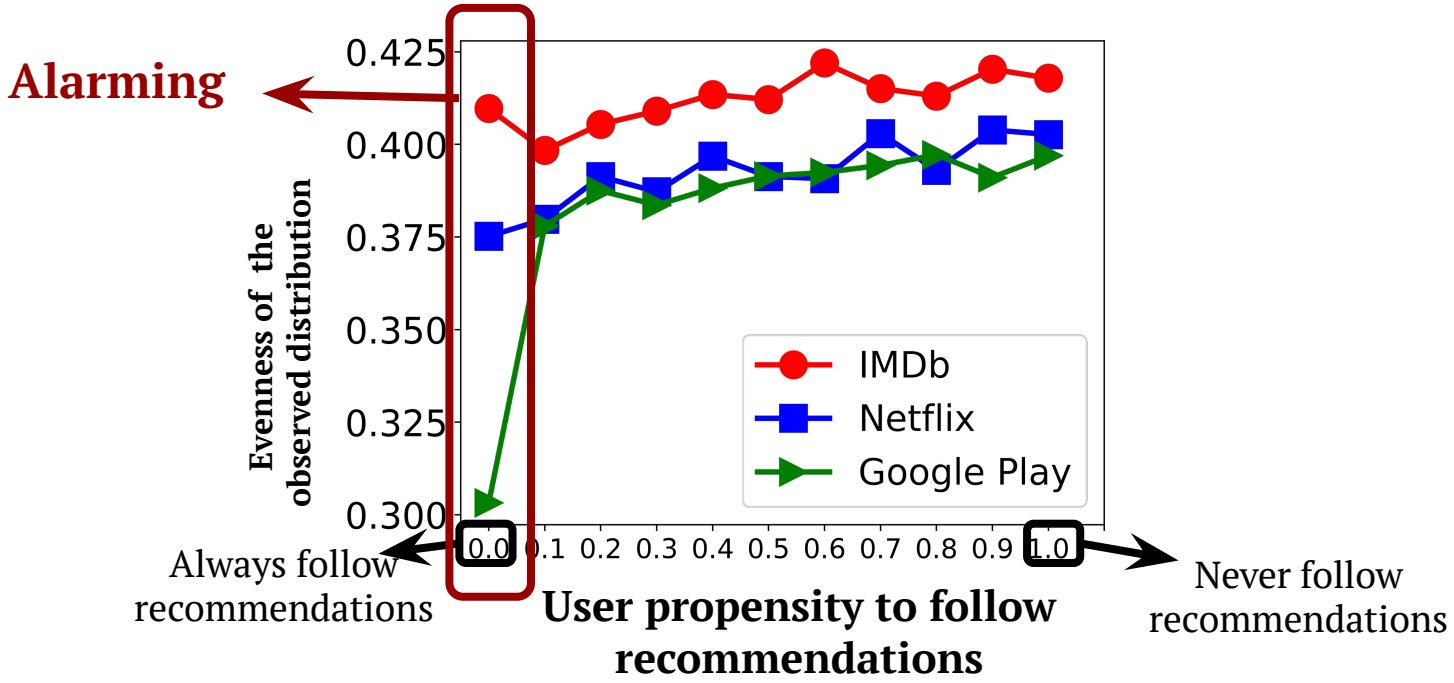
- User browsing by random walks (stated earlier)
- There are 110 groups of 10 random users each.
- Groups have been created as per each random user's
  - Starting movie (10 popular movies)
  - Propensity to follow recommendations (11 different values 0.0, 0.1, 0.2 …, 1.0)

# Extent of Information Segregation




- Evenness slowly increases with decrease in propensity to follow recommendations.
- IMDb has better evenness than Netflix and Google Play.

# Extent of Information Segregation




- Evenness slowly increases with decrease in propensity to follow recommendations.
- IMDb has better evenness than Netflix and Google Play.

# Concluding Remarks

- A simplistic yet useful model for third party auditing.
- How likely is a recommendation system to cause information segregation?

# Future Work

- Our framework can be extended to other domains.
- If enough data are available, personalization can also be incorporated.
- Developing methodology to audit and mitigate for bias in the context of recommendation

# Acknowledgements


- Supervisors
- Collaborators
- CNeRG, IIT Kharagpur
- Dept. of CSE
- TCS Research

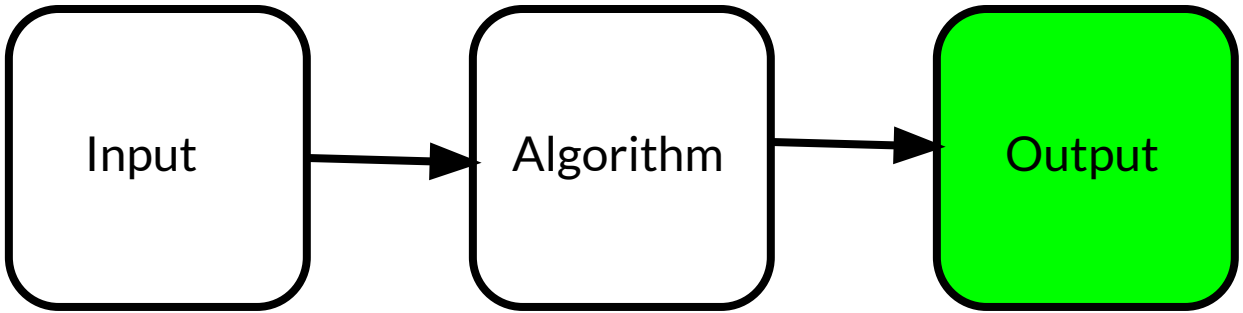
# THANK YOU
# FOR
# YOUR ATTENTION!!!

https://github.com/ad93/

@AbhisekFair

# Backup slides

# Post-processing based algorithm




**ReFaSumm**

# ReFaSumm

- Summarization algorithms generate a ranked list of textual units based on their summary worthiness.

# ReFaSumm

- Summarization algorithms generate a ranked list of textual units based on their summary worthiness.
- Fairly re-ranking them leads to generate fair summaries.

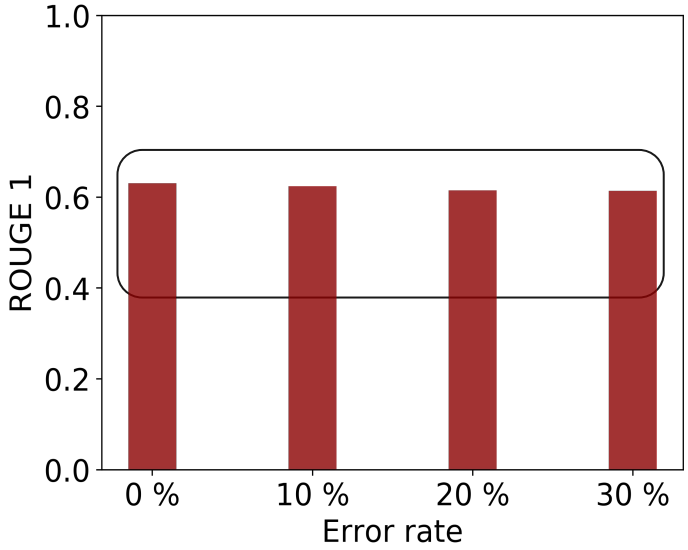
# ReFaSumm

- Summarization algorithms generate a ranked list of textual units based on their summary worthiness.
- Fairly re-ranking them leads to generate fair summaries.
- We have used the FA*IR algorithm for achieving the desired fairness criteria. [1]

[1] Zehlike, Meike, et al. "Fa* ir: A fair top-k ranking algorithm." *In proc ACM CIKM 2017*

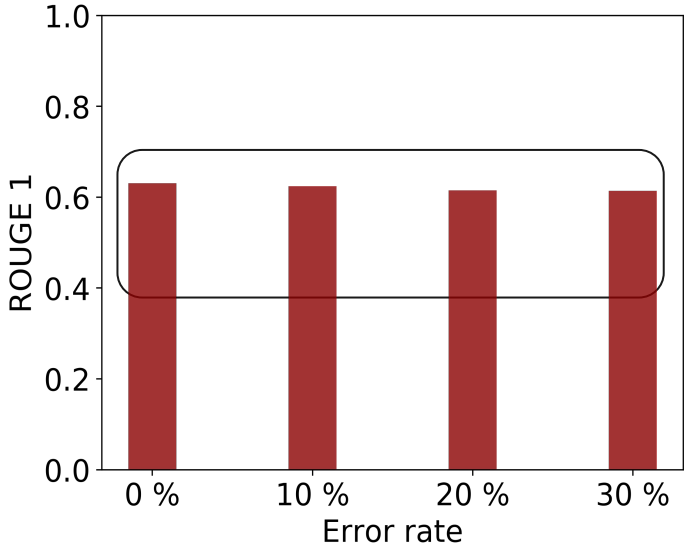
# Results and insights

- The proposed algorithms are immune to degraded information of demographic details.



# Results and insights

- The proposed algorithms are immune to degraded information of demographic details.
- Proposed algorithms are generalizable to different fairness notions.



# Limitations and future work

- Proposed algorithms assume the availability of the class information.
- Who gets to decide what is an appropriate social salient group for an application?
- Fair representation in the opinion space rather than demographic space.

# Diversity

- Generally, diversity is defined as an opposite notion to similarity.
- A set of similar items may not always be useful for a user.
- Many macroscopic measures of diversity are well known in the literature. (e.g. intralist diversity, longtail novelty etc.)

# Where lies the problem?

- Macroscopic measures do not give a clear picture.

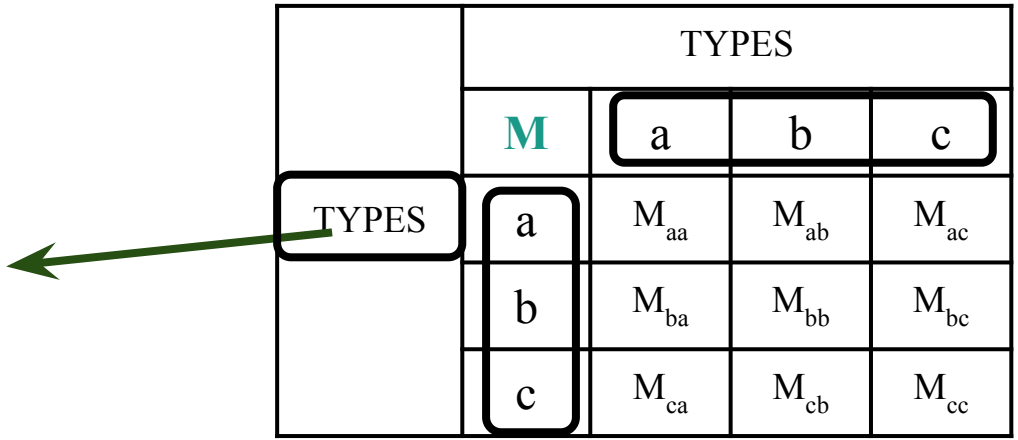
- Diversity or lack of it may not be distributed similarly across different types of items.

# Contingency Matrix

| TYPES | TYPES | | | |
|---|---|---|---|---|
| | **M** | a | b | c |
| | a | $M_{aa}$ | $M_{ab}$ | $M_{ac}$ |
| | b | $M_{ba}$ | $M_{bb}$ | $M_{bc}$ |
| | c | $M_{ca}$ | $M_{cb}$ | $M_{cc}$ |

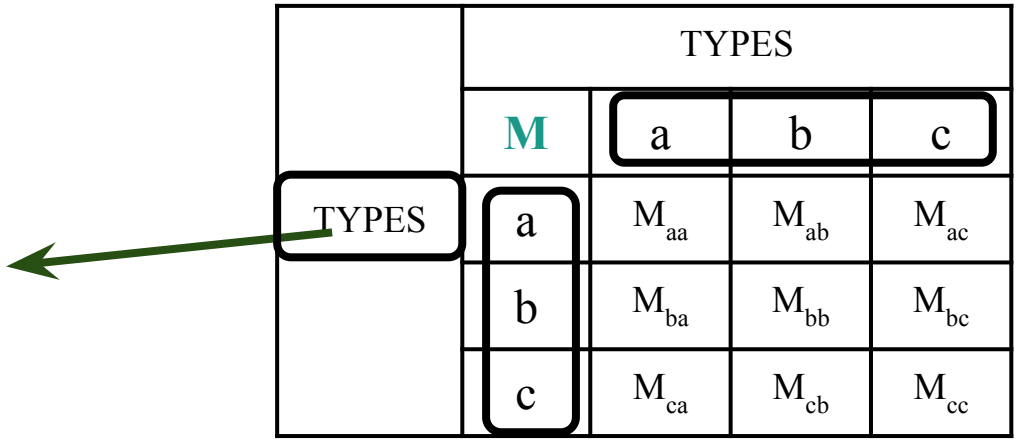
# Contingency Matrix

Application Specific
(Genre/Popularity)



# Contingency Matrix

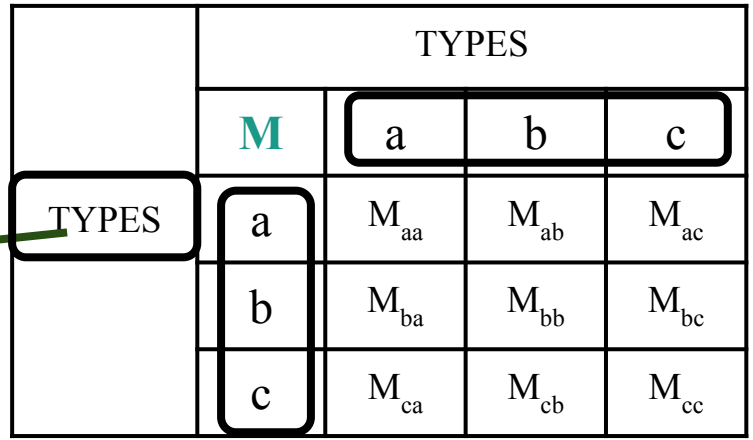
Application Specific
(Genre/**Popularity**)



# Contingency Matrix

Application Specific
(Genre/**Popularity**)

**Indegree** of nodes suggest how often they are being recommended on the platform .

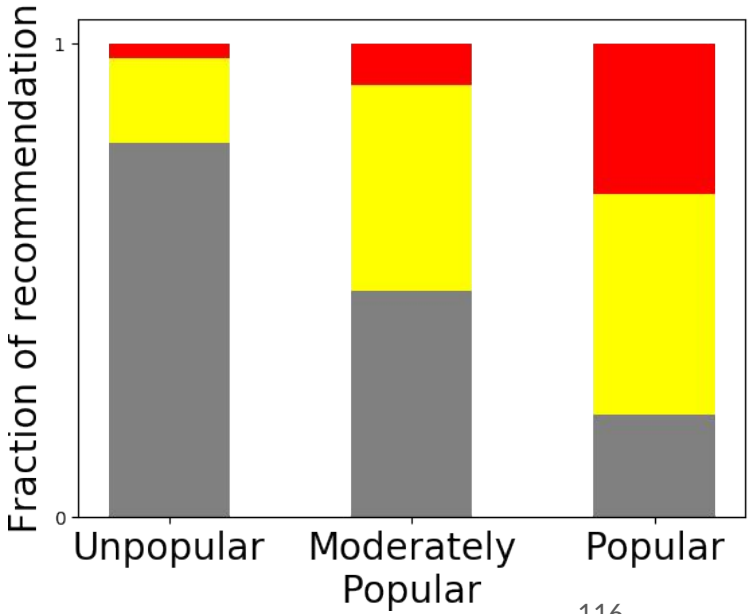


# Contingency Matrix

| TYPES | TYPES | | | |
|---|---|---|---|---|
| | **M** | Unpopular | Moderately popular | Popular |
| | Unpopular | $M_{aa}$ | $M_{ab}$ | $M_{ac}$ |
| | Moderately popular | $M_{ba}$ | $M_{bb}$ | $M_{bc}$ |
| | Popular | $M_{ca}$ | $M_{cb}$ | $M_{cc}$ |

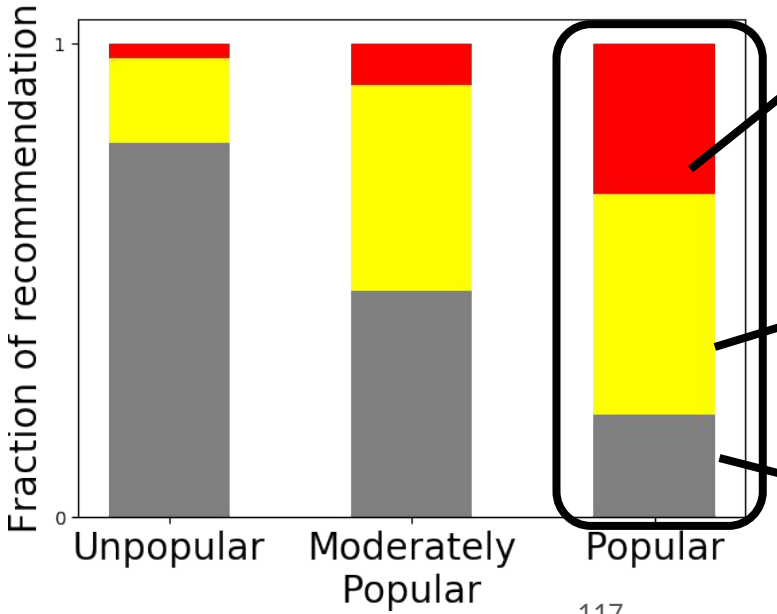
Fraction of recommendations from ‘moderately popular’ movies to ‘popular’ movies..

# Contingency Matrix based on Popularity




**IMDb Recommendation Network**

# Contingency Matrix based on Popularity



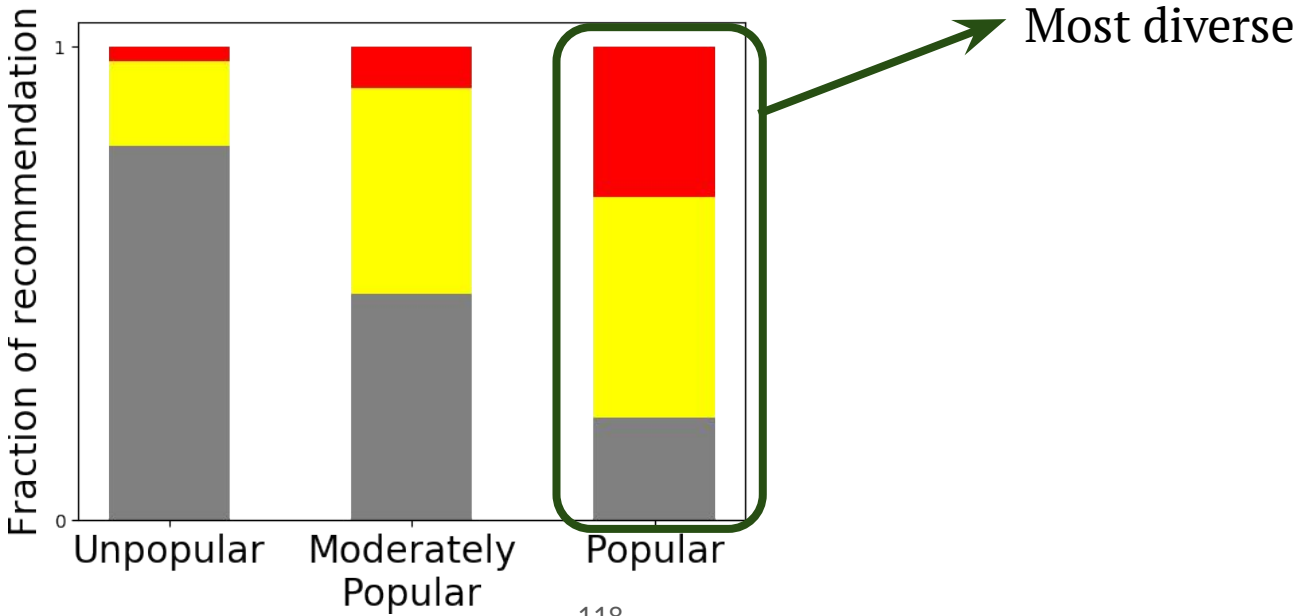

IMDb Recommendation Network

Recommendations from popular movies to popular movies

Recommendations from popular movies to moderately popular movies

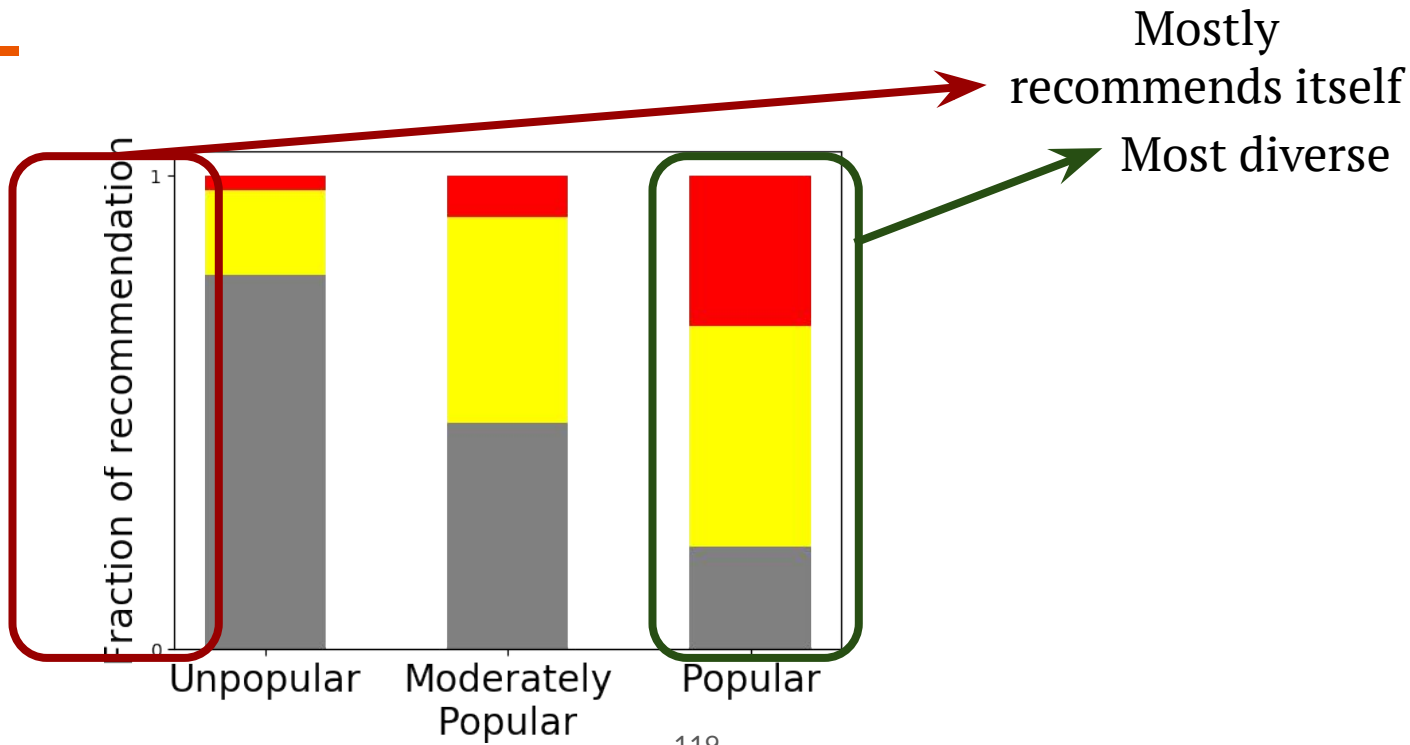
Recommendations from popular movies to unpopular movies

# Contingency Matrix based on Popularity

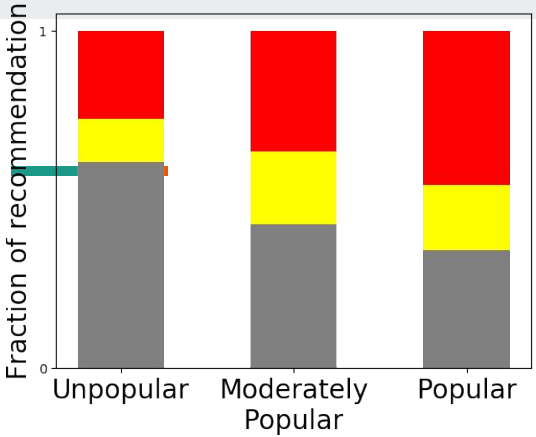



**IMDb Recommendation Network**

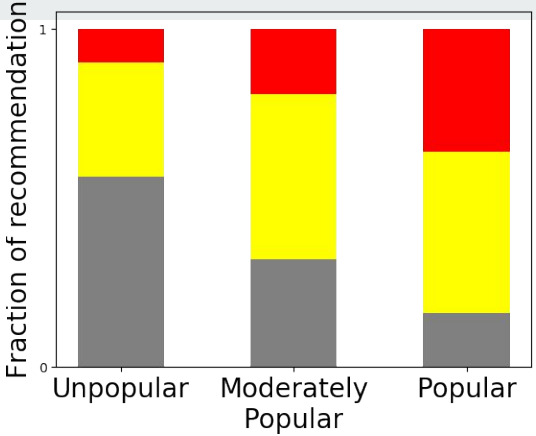
# Contingency Matrix based on Popularity



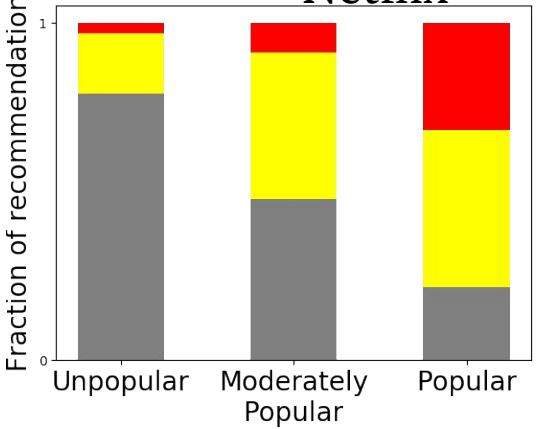

**IMDb Recommendation Network**




Netflix




Google Play




IMDb

High level insight: IMDb is more diverse than Netflix and Google Play.

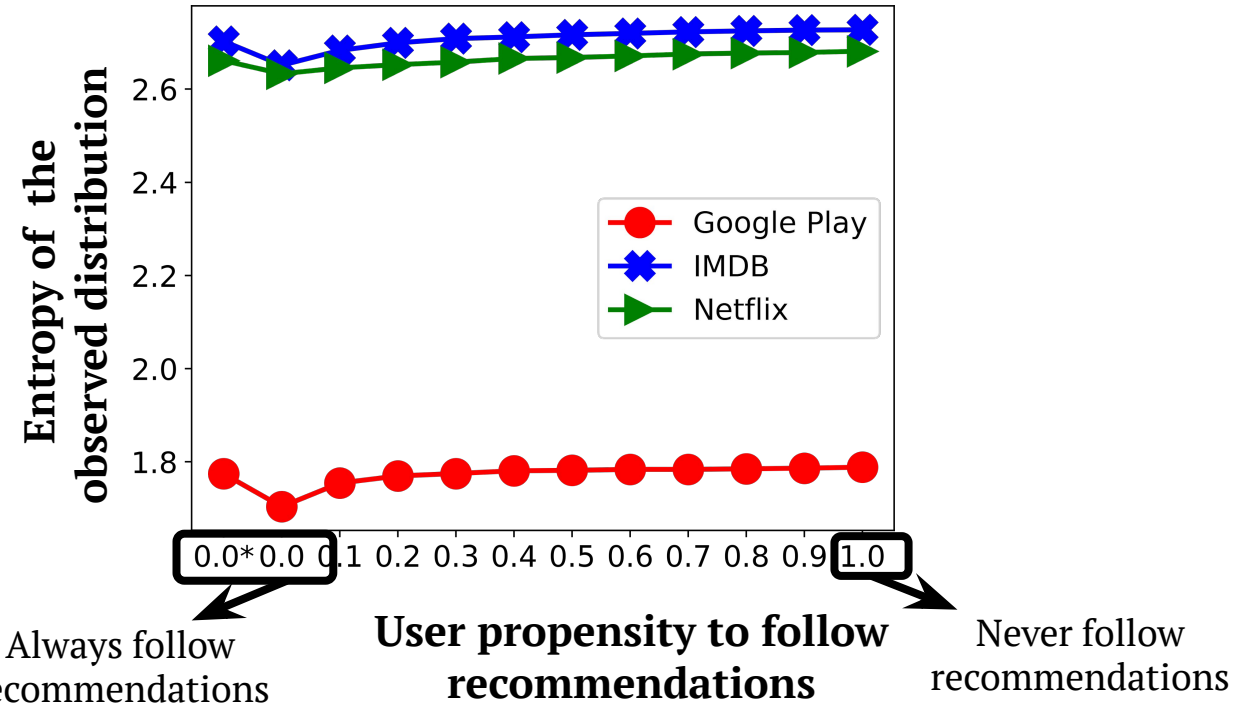
# Random Walk-based Measures

- Recommendation Network + User Browsing
- Shannon entropy of their observed distribution to quantify the diversity.

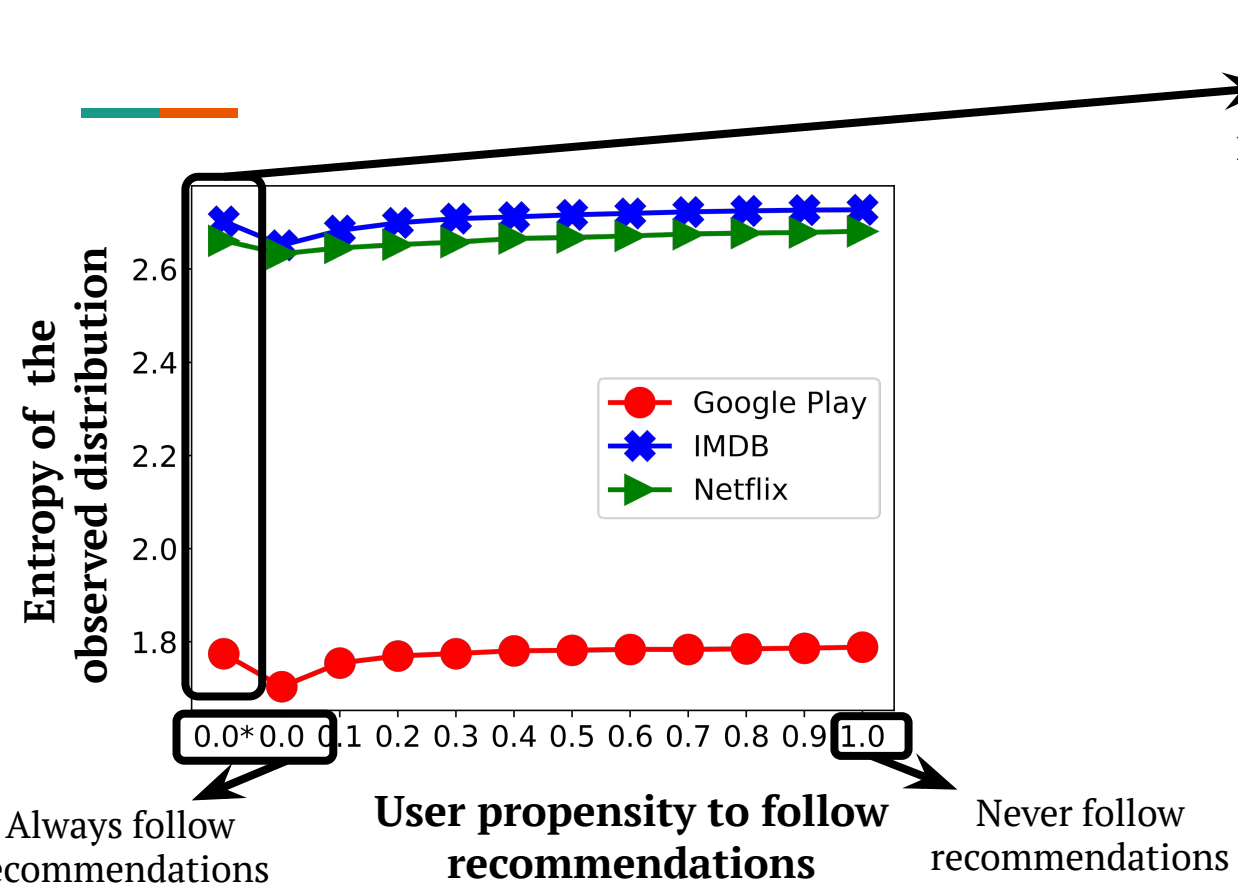
# Random Walk-based Measures

- Recommendation Network + User Browsing
- Shannon entropy of their observed distribution to quantify the diversity.
- Major factor that determines diversity:
  - **Teleportation probability (User Propensity)**

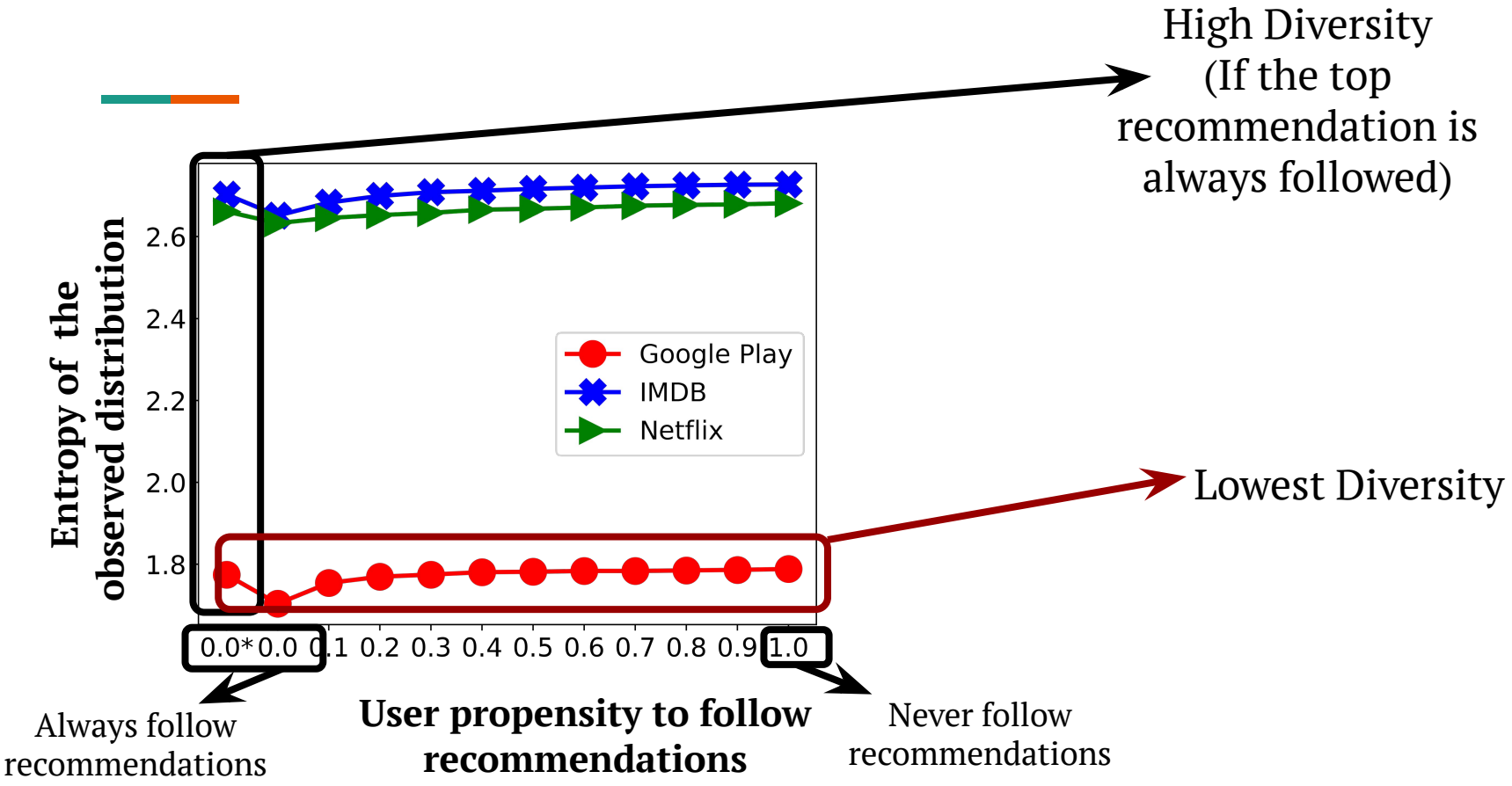
# Variation of Entropy



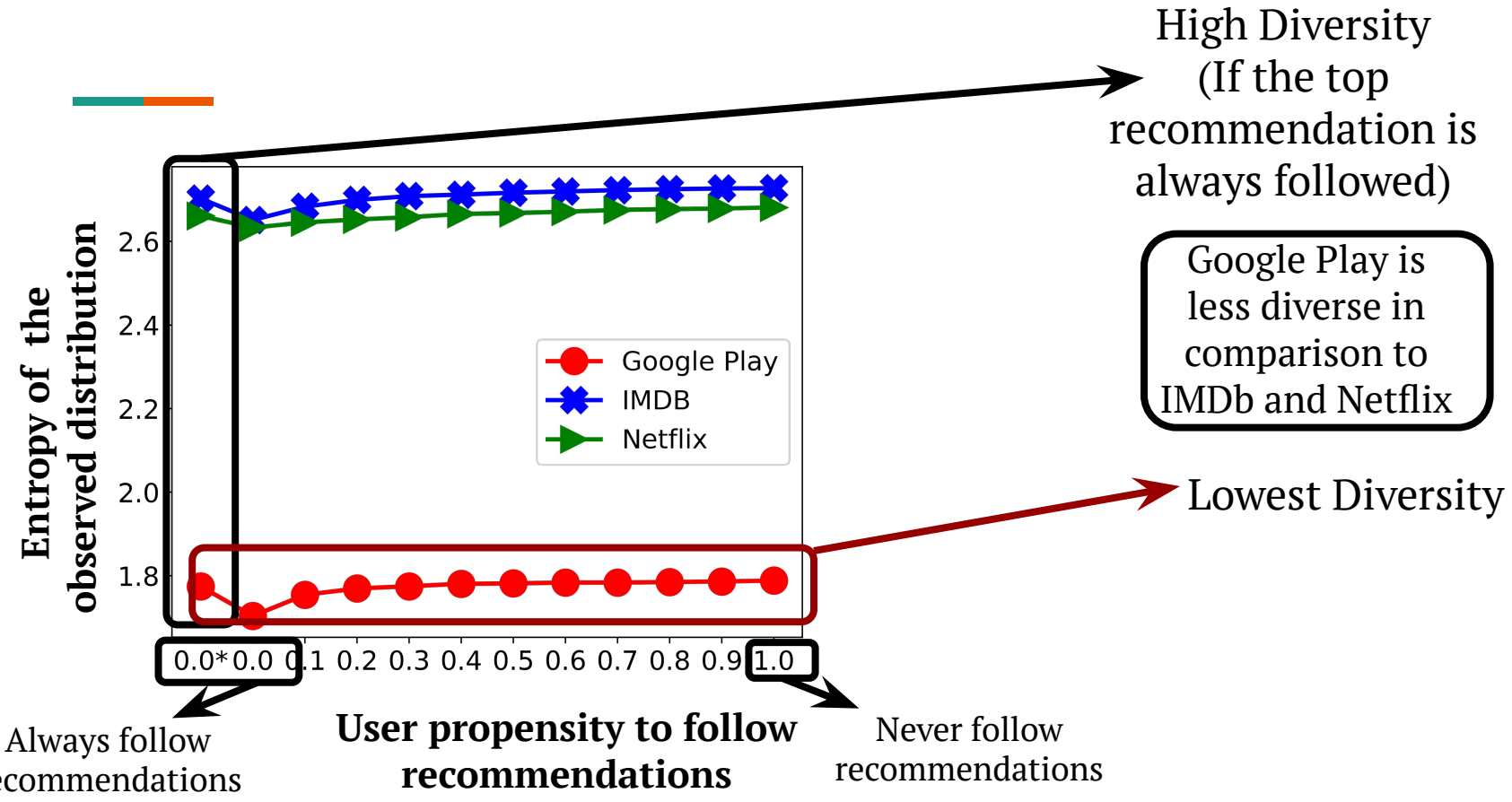
# Variation of Entropy



# Variation of Entropy



# Variation of Entropy

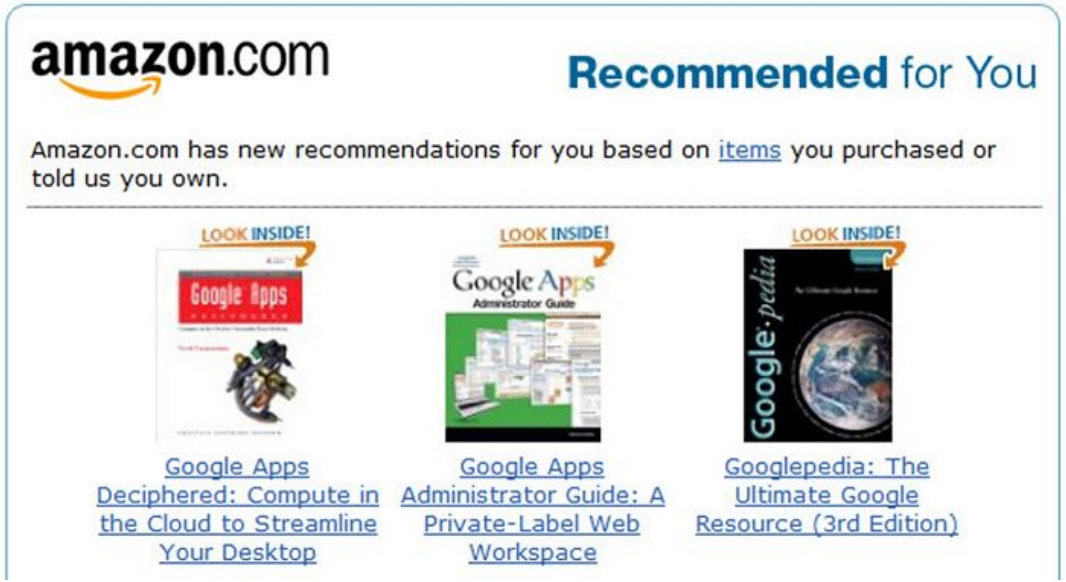


# Information Filtering Systems

- Search system

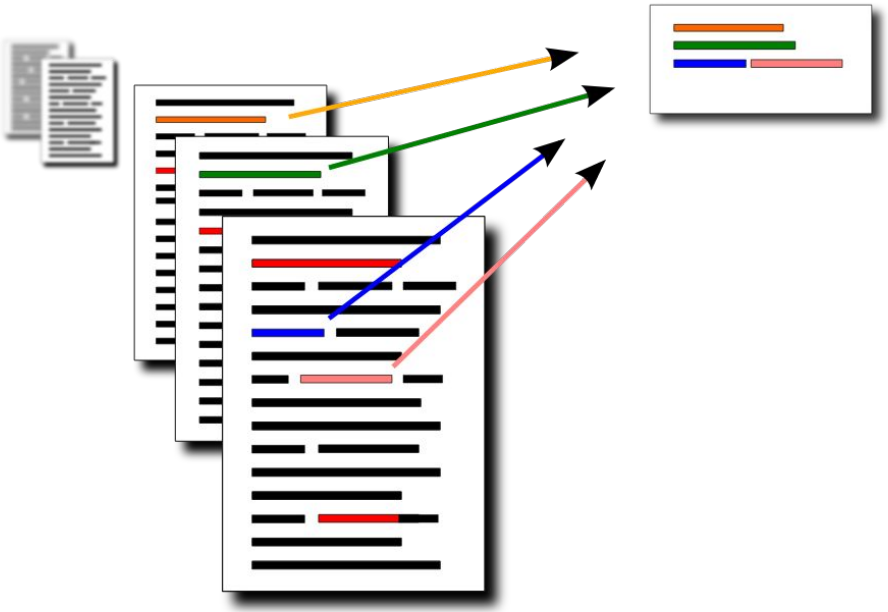
# Information Filtering Systems

- Search system
- Recommendation systems



# Information Filtering Systems

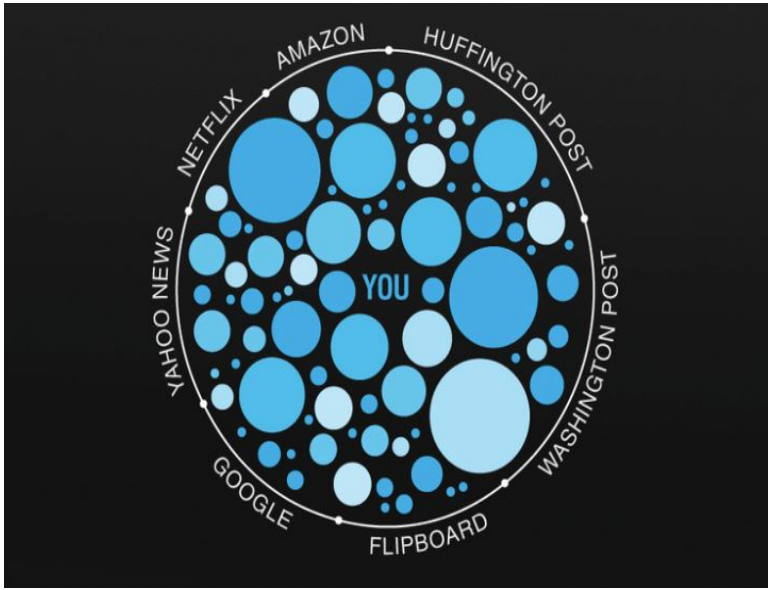
- Search system
- Recommendation systems
- Summarization systems



# Web as a source of information

# Inadvertent consequences

- Discrimination / bias
- **Filter bubble**



# Need for summarization

Explosion in amount of text

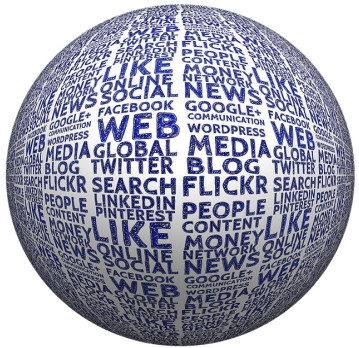



Image courtesy https://libguides.aecc.ac.uk/copyright/socialmedia

# Need for summarization

news360


agolo

inshorts

AYLIEN