UnBias: Emancipating Users Against Algorithmic Biases for a Trusted Digital Economy

Michael Rovatsos reusing lots of slides by
Ansgar Koene, University of Nottingham

Information services

Free to use → no competition on price

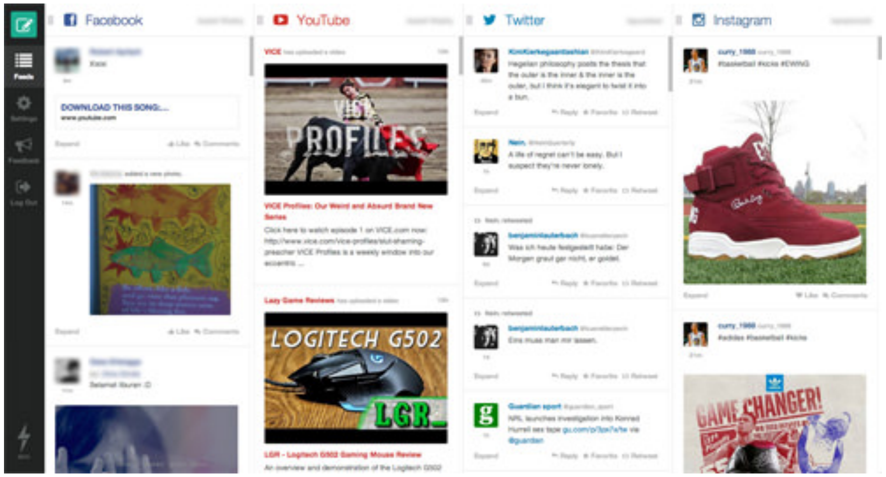
Plenty of assets → no competition on quantity

Competition on quality of service

Quality determined by relevance

Filtering becomes key

Convenience vs. choice



DOWNLOAD THIS SONG:...
www.youtube.com

VICE Profiles: Our Weird and Absurd Brand New Series
Click here to watch episode 1 on VICE.com now: http://www.vice.com/vice-profiles/slut-shaming-preacher VICE Profiles is a weekly window into our eccentric ...

Lazy Game Reviews
LGR - Logitech G502 Gaming Mouse Review
An overview and demonstration of the Logitech G502

KimKierkegaardashian
Hegelian philosophy posits the thesis that the outer is the inner & the inner is the outer, but I think it's elegant to twist it into a bun.

Nein.
A life of regret can't be easy. But I suspect they're never lonely.

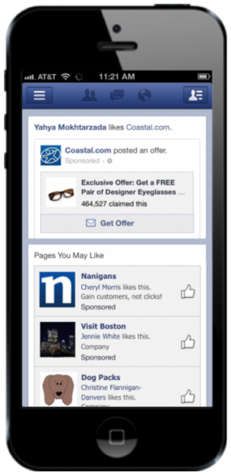
benjaminlauterbach
Was ich heute festgestellt habe: Der Morgen graut gar nicht, er goldet.

benjaminlauterbach
Eins muss man mir lassen.

Guardian sport
NRL launches investigation into Konrad Hurrell sex tape gu.com/p/3pk7a/tw via @guardian

curry_1988
#basketball #kicks #EWING

curry_1988
#adidas #basketball #kicks



Yahya Mokhtarzada likes Coastal.com.

Coastal.com posted an offer.
Sponsored

Exclusive Offer: Get a FREE Pair of Designer Eyeglasses ...
464,527 claimed this

Get Offer

Pages You May Like

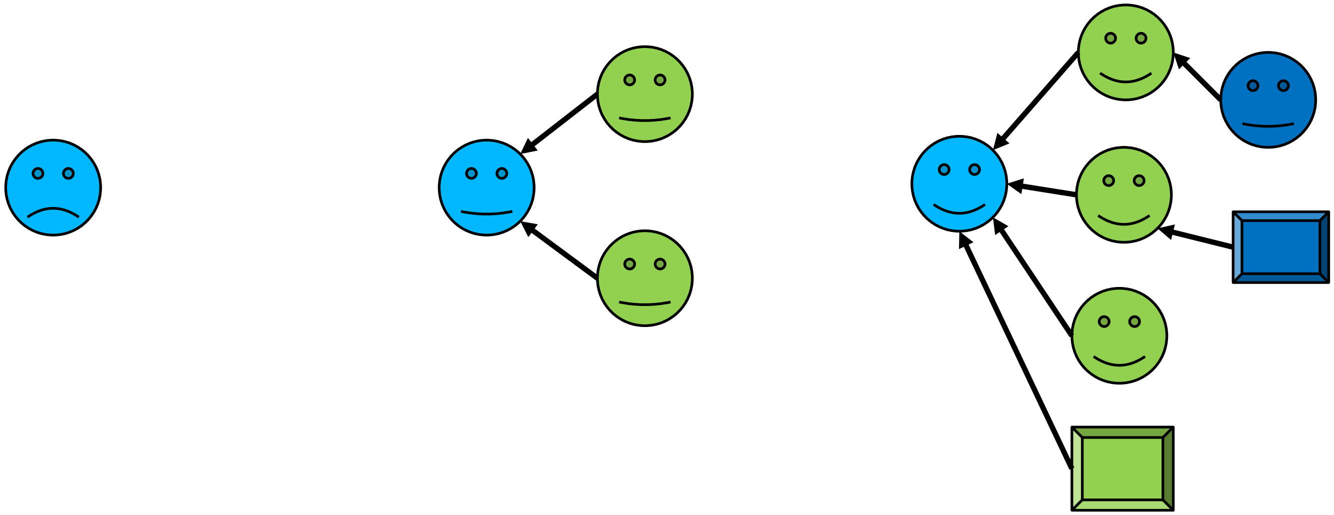
Nanigans
Cheryl Morris likes this.
Gain customers, not clicks!
Sponsored

Visit Boston
Jennie White likes this.
Company
Sponsored

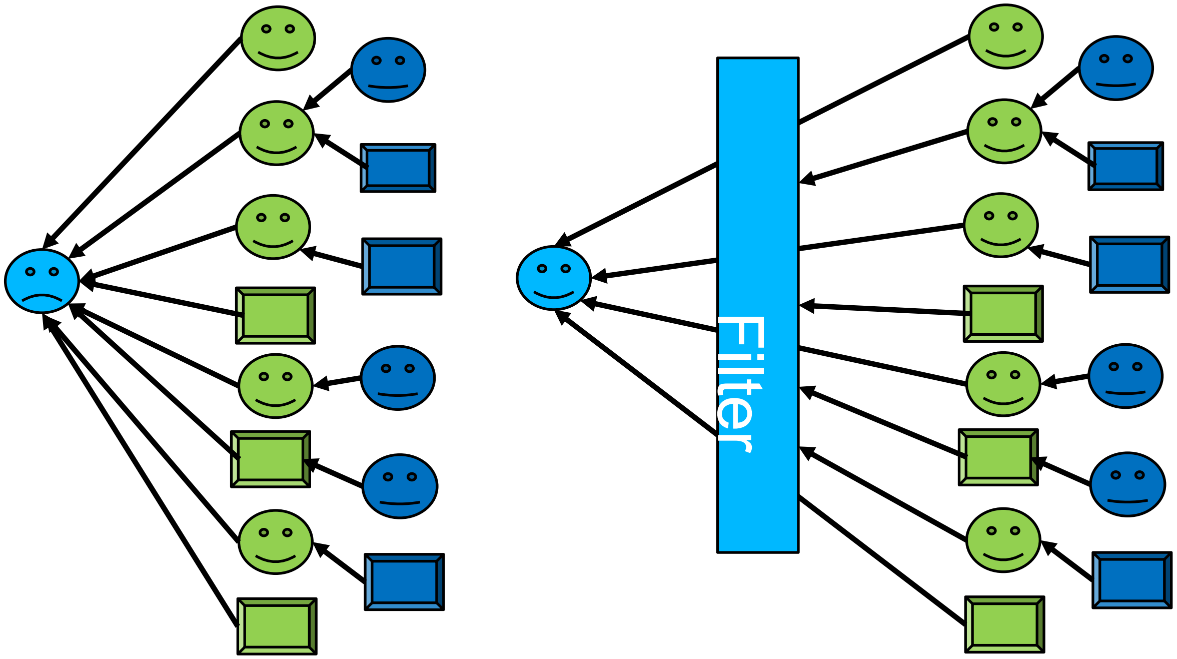
Dog Packs
Christine Flannigan-Danvers likes this.

Social networks

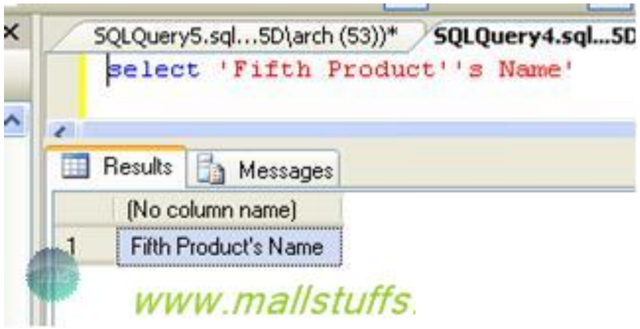
User experience satisfaction on social network sites



Filtering



Propagation of influence



```
select 'Fifth Product''s Name'
```

```
  (No column name)
1 Fifth Product's Name
```

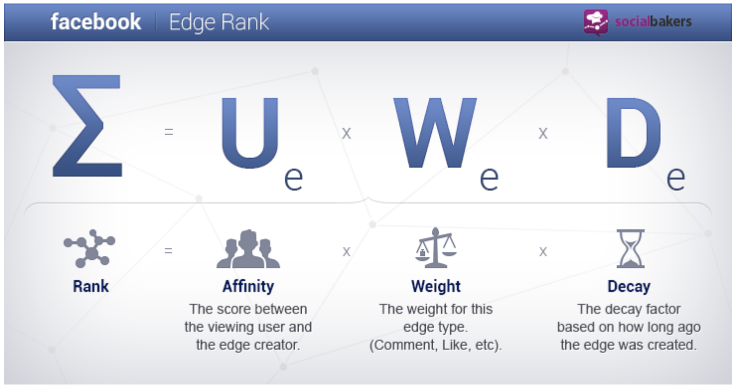
Recommendation

Content-based – **similar items** to those user liked

Collaborative – what **similar users** liked

Community – what people in **same social network** liked

The role of algorithms



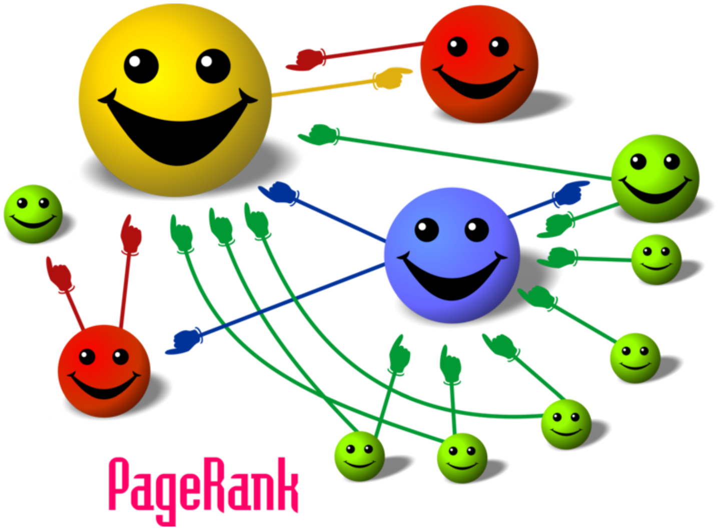
$$\Sigma = U_e \times W_e \times D_e$$

Rank = Affinity x Weight x Decay

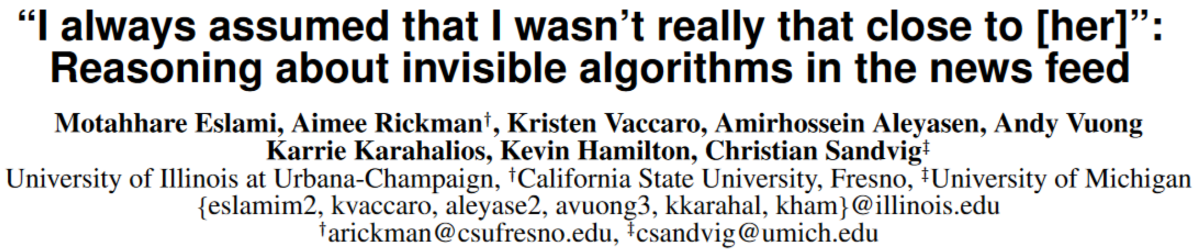
Affinity The score between the viewing user and the edge creator.

Weight The weight for this edge type. (Comment, Like, etc).

Decay The decay factor based on how long ago the edge was created.



User understanding of social media algorithms



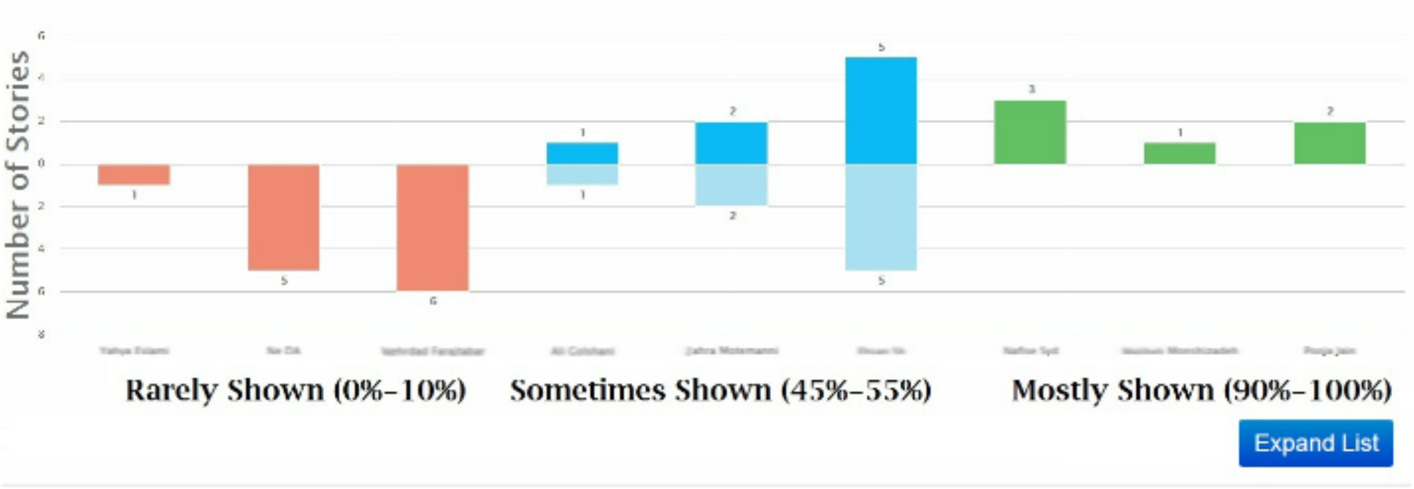
“I always assumed that I wasn’t really that close to [her]”: Reasoning about invisible algorithms in the news feed

Motahhare Eslami, Aimee Rickman[†], Kristen Vaccaro, Amirhossein Aleyasen, Andy Vuong
Karrie Karahalios, Kevin Hamilton, Christian Sandvig[‡]
University of Illinois at Urbana-Champaign, [†]California State University, Fresno, [‡]University of Michigan
{eslamim2, kvaccaro, aleyase2, avuong3, kkarahal, kham}@illinois.edu
[†]arickman@csufresno.edu, [‡]csandvig@umich.edu

More than 60% of Facebook users are entirely unaware of any algorithmic curation on Facebook at all: “They believed every single story from their friends and followed pages appeared in their news feed”.

CHI 2015

Revealing News Feed behaviour



Awareness through experimentation

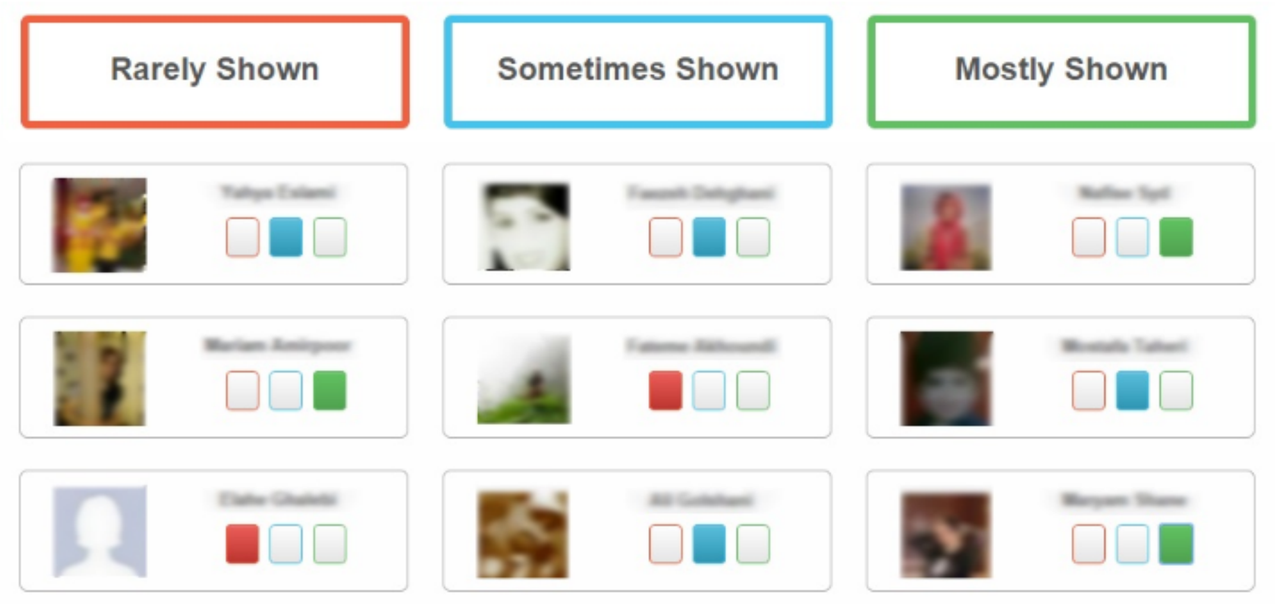
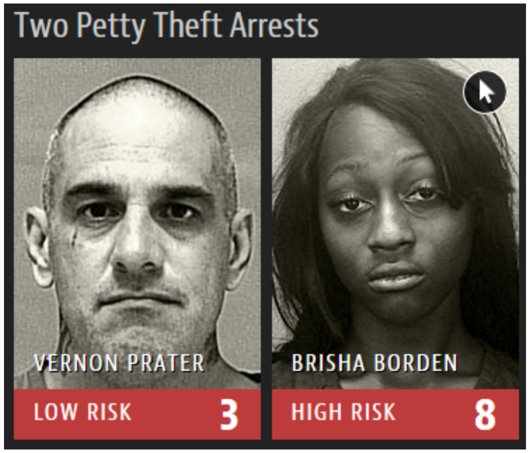


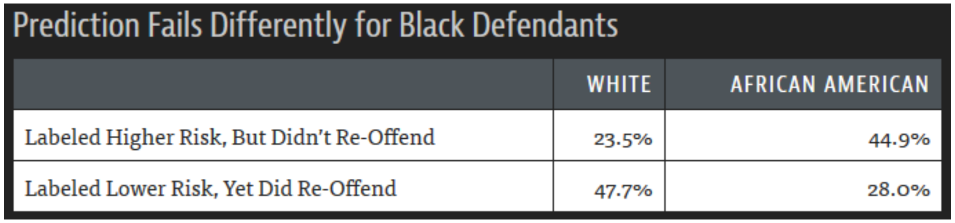
Figure 3. The Friend Rearrangement View. User can move friends between the categories by changing the color of a friend to the destination category's color.

Garbage in, garbage out?

perpetuating the status-quo

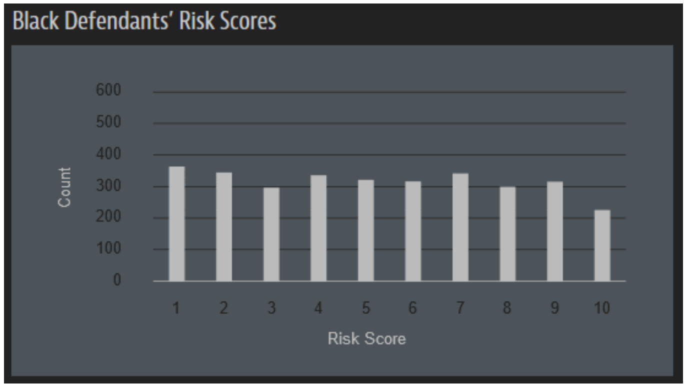


Two Petty Theft Arrests

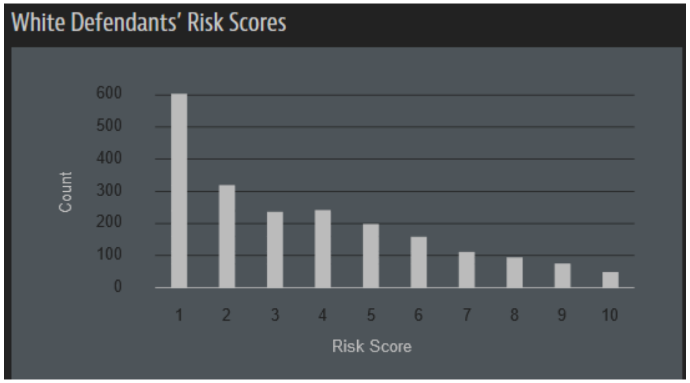


Prediction Fails Differently for Black Defendants

	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%



Black Defendants' Risk Scores



White Defendants' Risk Scores

The risks of algorithmic bias

Information filtering, or ranking, implicitly biases choice behaviour

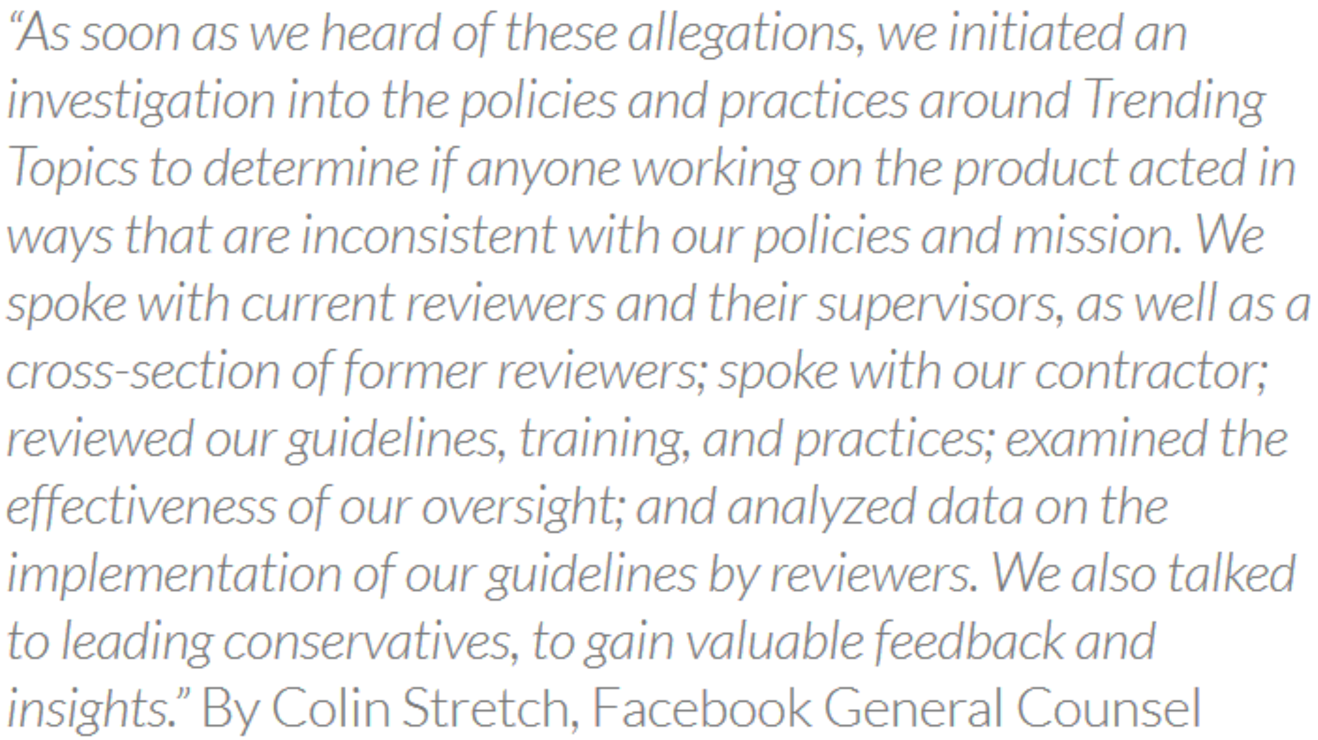
Many online information services are paid for by advertising revenue

Conflict of interest: promote advertisement vs. match user interests

Advertising, which inherently involves manipulation, becomes ubiquitous

Personalized filtering can also be use for political spin / propaganda etc

Trending Topics controversy



“As soon as we heard of these allegations, we initiated an investigation into the policies and practices around Trending Topics to determine if anyone working on the product acted in ways that are inconsistent with our policies and mission. We spoke with current reviewers and their supervisors, as well as a cross-section of former reviewers; spoke with our contractor; reviewed our guidelines, training, and practices; examined the effectiveness of our oversight; and analyzed data on the implementation of our guidelines by reviewers. We also talked to leading conservatives, to gain valuable feedback and insights.” By Colin Stretch, Facebook General Counsel

Q&A with N. Lundblad (Google)

“Human attention is the limited resource that services need to compete for. As long as there exist competing platforms, loss of agency due to algorithms deciding what to show to users is not an issue. Users can switch to other platforms.”

Nicklas Lundblad, Head of EMEA Public Policy
and Government Relations at Google

See also:

2011 FTC investigation of Google for search bias

EU competition regulation vs Google

‘Equal opportunity by design’

“To avoid exacerbating biases by encoding them into technological systems a principle of ‘equal opportunity by design’—designing data systems that promote fairness and safeguard against discrimination from the first step of the engineering process and continuing throughout their lifespan.”

Big Data: A Report on Algorithmic Systems, Opportunities, and Civil Rights“, White House report focused on the problem of avoiding discriminatory outcomes

The UnBias Project

WP1: Co-production of citizen education materials with young people

WP2: Fair algorithms and tools for exposing algorithmic bias

WP3: Qualitative research into of users’ sense-making behaviour

WP4: Developing an information and education governance framework

Two-year project funded by EPSRC (£1.1m budget), collaboration between Nottingham (coordinator), Edinburgh, and Oxford

Challenges relating to data used as inputs to an algorithm

- Poorly selected data, selection bias
- Incomplete, incorrect, or outdated data
- Unintentional perpetuation and promotion of historical biases

Challenges related to the inner workings of the algorithm itself

- Poorly designed matching systems
- Personalisation and recommendation services that narrow instead of expand user options
- Decision-making systems that assume correlation necessarily implies causation
- Data sets that lack information or disproportionately represent certain populations

Concerns regarding personalisation

- **Social consequences**: self-reinforcing information filtering – the 'filter bubble' effect
- **Privacy**: personalisation involves profiling of individual behaviour/interests
- **Agency**: the filtering algorithm decides which segment of available information the user gets to see
- **Manipulation**: people's actions/choices are depend on the information they are exposed to

Challenges related to transparency & accountability

Filter algorithms provide competitive advantage, details about them are often trade-secrets

- Users don’t know how the information they are presented was selected → no real informed consent
- Service users have no ‘manual’ override for the settings of the information filtering algorithms
- It is difficult for service users to know which information they don’t know about because it was filtered

Case study

Fairness in task composition

Combinatorial problem of allocating **groups** of users to **shared** tasks, where task requests come from users

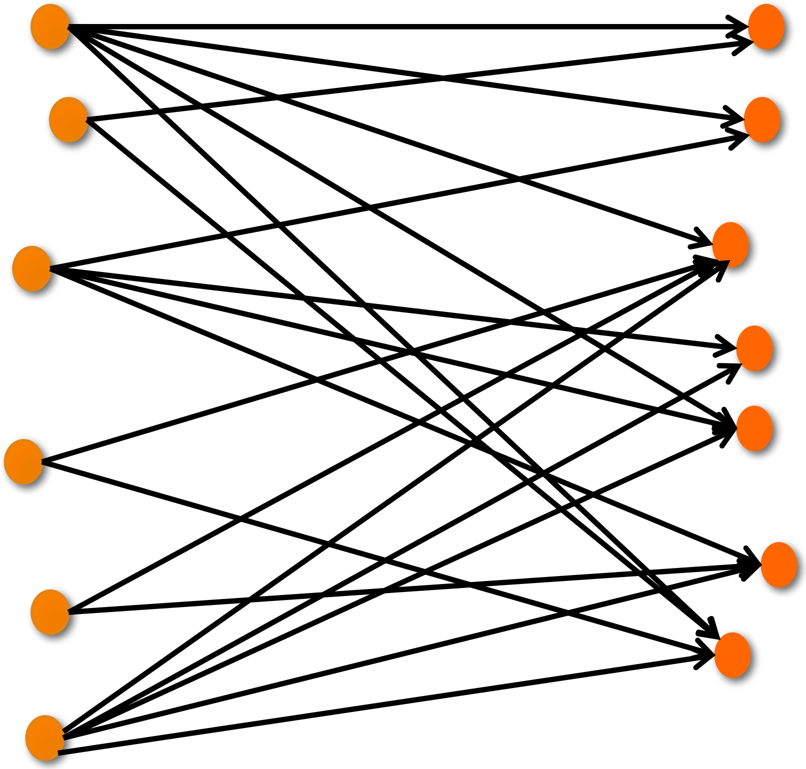
- **Hard constraints** restrict the groupings and task properties that can be realised in principle
- **Soft constraints** determine which coalition structures and task features are preferred by system and/or users

Examples:

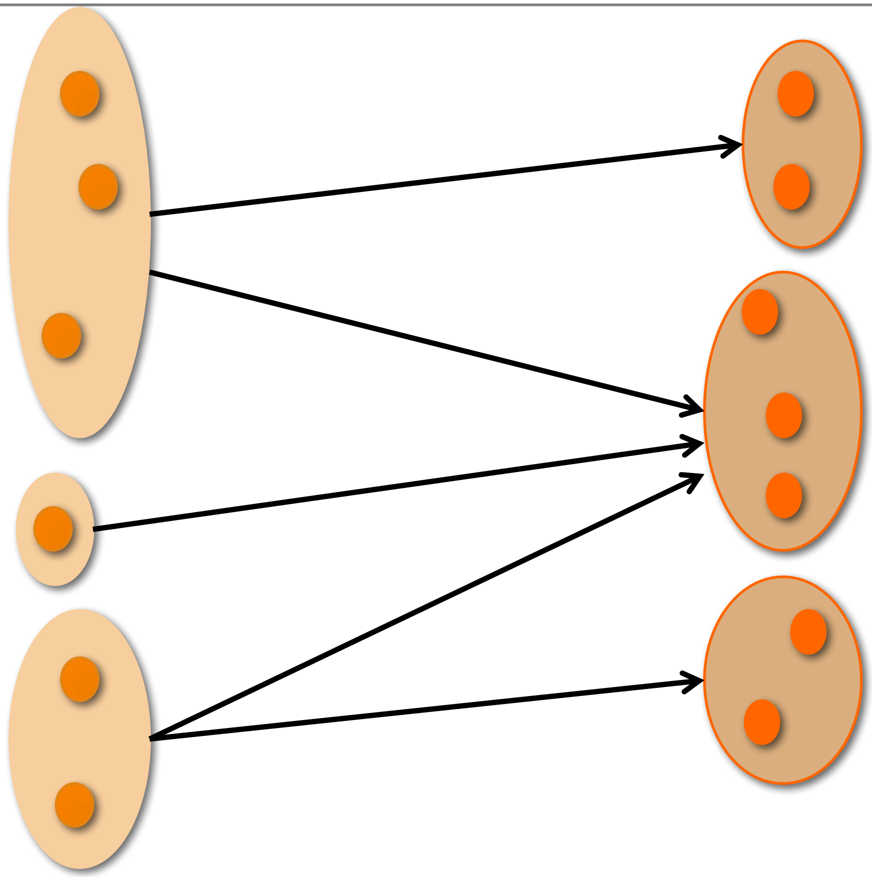
- **Ridesharing** where drivers share their cars with other passengers for a specific journey (our vanilla example)
- Meeting scheduling, citizen science tasks, workforce shift allocation, clinical workflow management, etc etc

Composition

Composition



Composition



Diversity in task composition

In traditional mechanism design, global allocations are computed given individual preferences and global criteria

- E.g. social welfare maximisation, Pareto optimality, strategy-proofness, envy-freeness etc.

Mechanisms are proposed that provably satisfy these properties, solution can therefore be **imposed** on users

Diversity implies that users cannot report all preferences

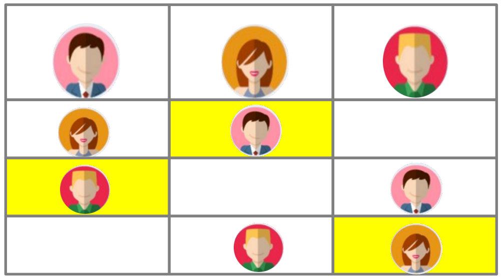
- System never captures all relevant decision variables
- Solutions cannot be computed/considered exhaustively
- Utility of solutions cannot be determined by users a priori

From task allocation to task recommendation

Key problems:

1. How to compute "optimal" **sets** of solutions

2. How to **influence** users' choices



3. How to **learn** users' preferences

User and system utility

User's utility function u_i depends on user’s requirements and preferences

Global (system) utility function depends on social welfare and maximising task completion

$$U_s = \sum_{i \in I} u_i + \sum_{i \in I} \sum_{j \in J} x_{i,j}$$

Computing allocations

MIP*

$\rightarrow V^*$

Objective $max_{a \in A} U_s(a)$

Constraints Hard feasibility constraints

MIP$^{\text{first}}$

$\rightarrow a \in R$

Objective $min_{a \in A} \sum_{i \in I} \sum_{i' \in I | i' > i} |u_i(a) - u_{i'}(a)|$

Constraints MIP*

$U_s(a) \cdot h \geq V^*$

MIP$^{\text{others}}$

$\rightarrow a' \in R$

Objective $min_{a' \in A} \sum_{i \in I} |u_i(a) - u_i(a')|$

Constraints MIP$^{\text{first}}$

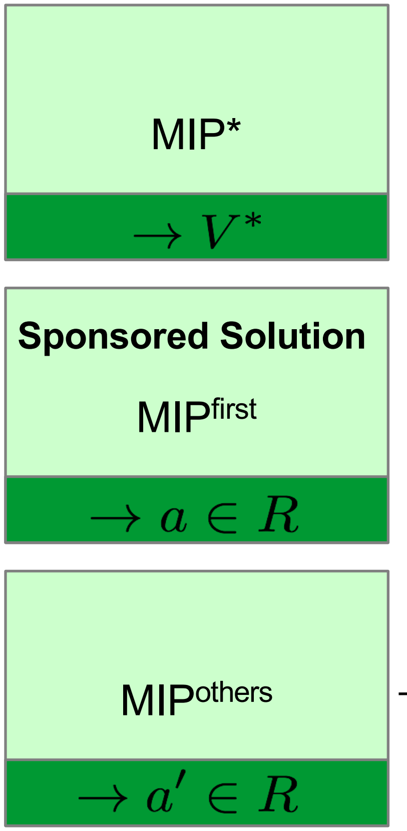
$a' \notin R$

Influencing users

We want to modify users' utility artificially so that their choices lead to a feasible global solution

- Explicit Approaches:
  - Intervention
  - (Possible) future reward
- Implicit Approaches:
  - discounts
  - **taxation**

Taxation



MIP*

$\rightarrow V^*$

Sponsored Solution

MIP$^{\text{first}}$

$\rightarrow a \in R$

MIP$^{\text{others}}$

$\rightarrow a' \in R$

Objective

$$min \sum_{i \in I} |u_i(a) - u_i(a') + \tau_i(a')| + M\Big(\sum_{i \in I} (u_i(a) + \epsilon - u_i(a) + \tau_i(a'))\Big)$$

Constraints

MIP$^{\text{first}}$

$a' \notin R$

Noiseless and Constant Noise Models

$$u_i(a) + \epsilon \geq u_i(a') - \tau_i(a')$$

Logit Model (also goes into objective function)

$$\frac{u_i(a)}{\Big(\sum_{a'' \in R} (u_i(a'') - \tau_i(a'')) + u_i(a') - \tau_i(a')\Big)} \geq \psi$$

Conclusions

Task recommendation scenario exposes challenges related to fairness

What does it mean for algorithms to be “fair”?

How can we map human notions to computational models?

Computational limitations vs. societal demands

Where do the responsibilities lie: algorithm, platform, or users?

This is much bigger in terms of ethics of AI than “killer robots” or “singularity”!