

IN4MATX 153: CSCW

Class 3:
Feeds and their Algorithms

Professor Daniel Epstein
TA Dennis Wang
Reader Weijie Du

Announcements

- Office hours schedule finalized
 - Professor Epstein: We 2:00-3:00, DBH 6093
 - TA Wang: Mo 2:00-3:00, DBH 6121
 - Reader Du: Th 1:00-2:00, DBH 5221
- Start working on your meme assignment, if you haven't already
 - We expect you to iterate, which takes time
- We're working on PollEverywhere integration into Canvas
 - Perusall integration should be up & running, let us know if you have any questions

Today's goals

By the end of today, you should be able to...

- Articulate what a feed is, and how it is represented in different social media sites
- Describe the importance of filtering and sorting algorithms to social media content represented in feeds
- Explain how a typical feed algorithm works, at a high level
- Discuss whether feeds create filter bubbles or echo chambers, particularly around online discourse

Information overload

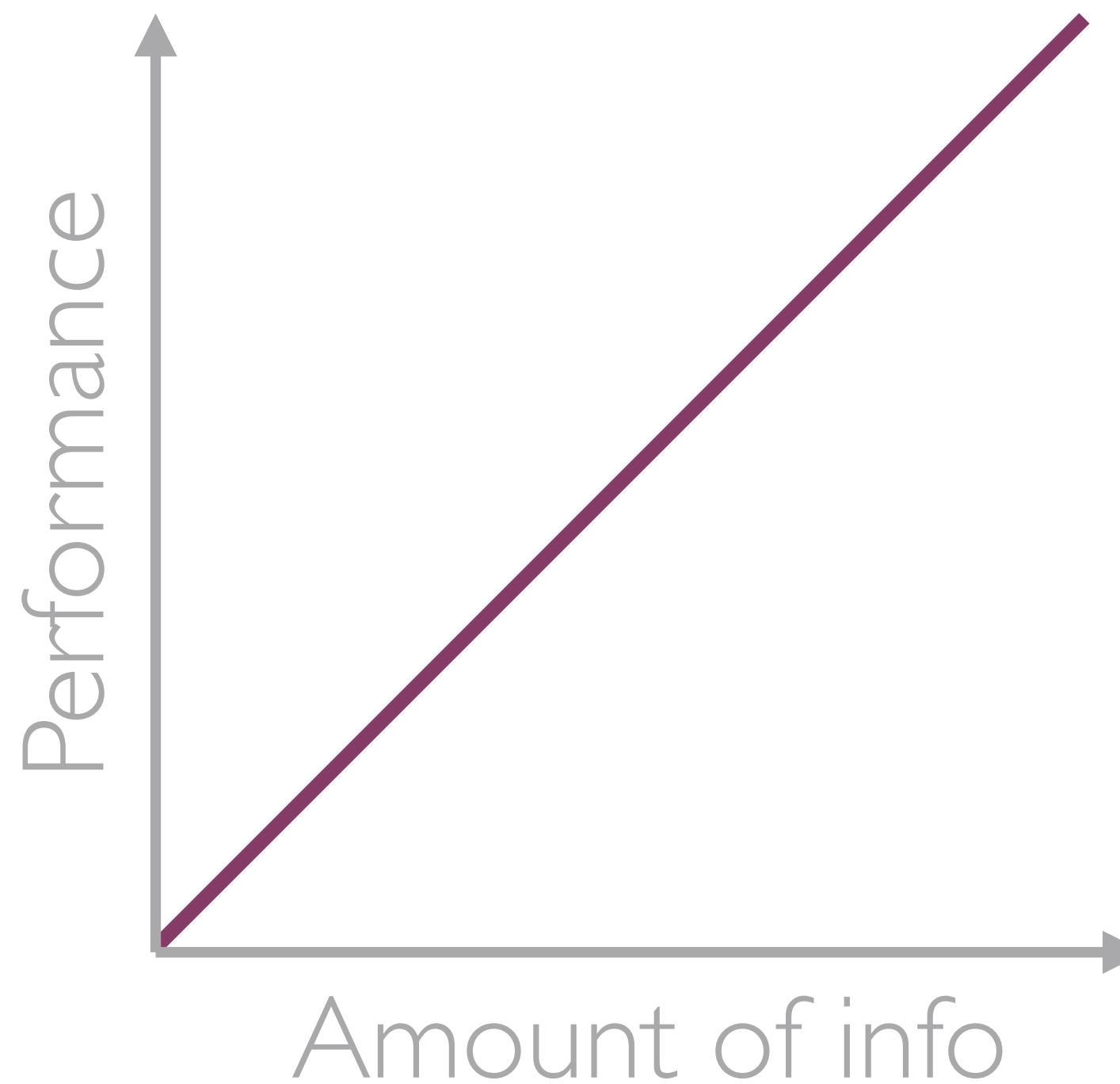
- Our digital tools are constantly bombarding us with information coming through different channels
 - Emails
 - Discord Messages
 - Social media posts
 - News articles
 - ...

Information overload

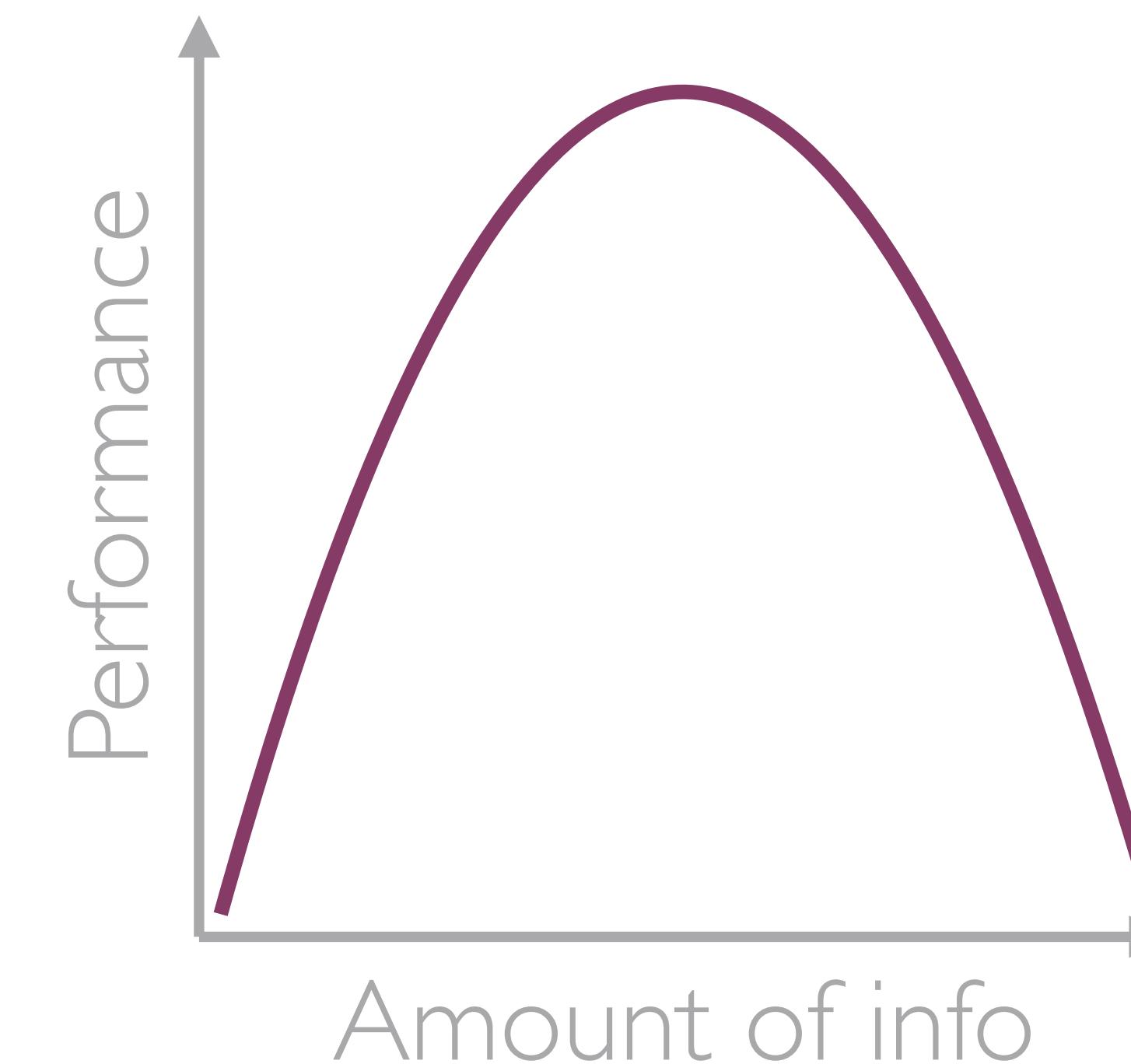
- Our digital tools are constantly bombarding us with information coming through different channels
 - Emails
 - Discord Messages
 - Social media posts
 - News articles
 - ...

Information overload: performance

Humans as information processors



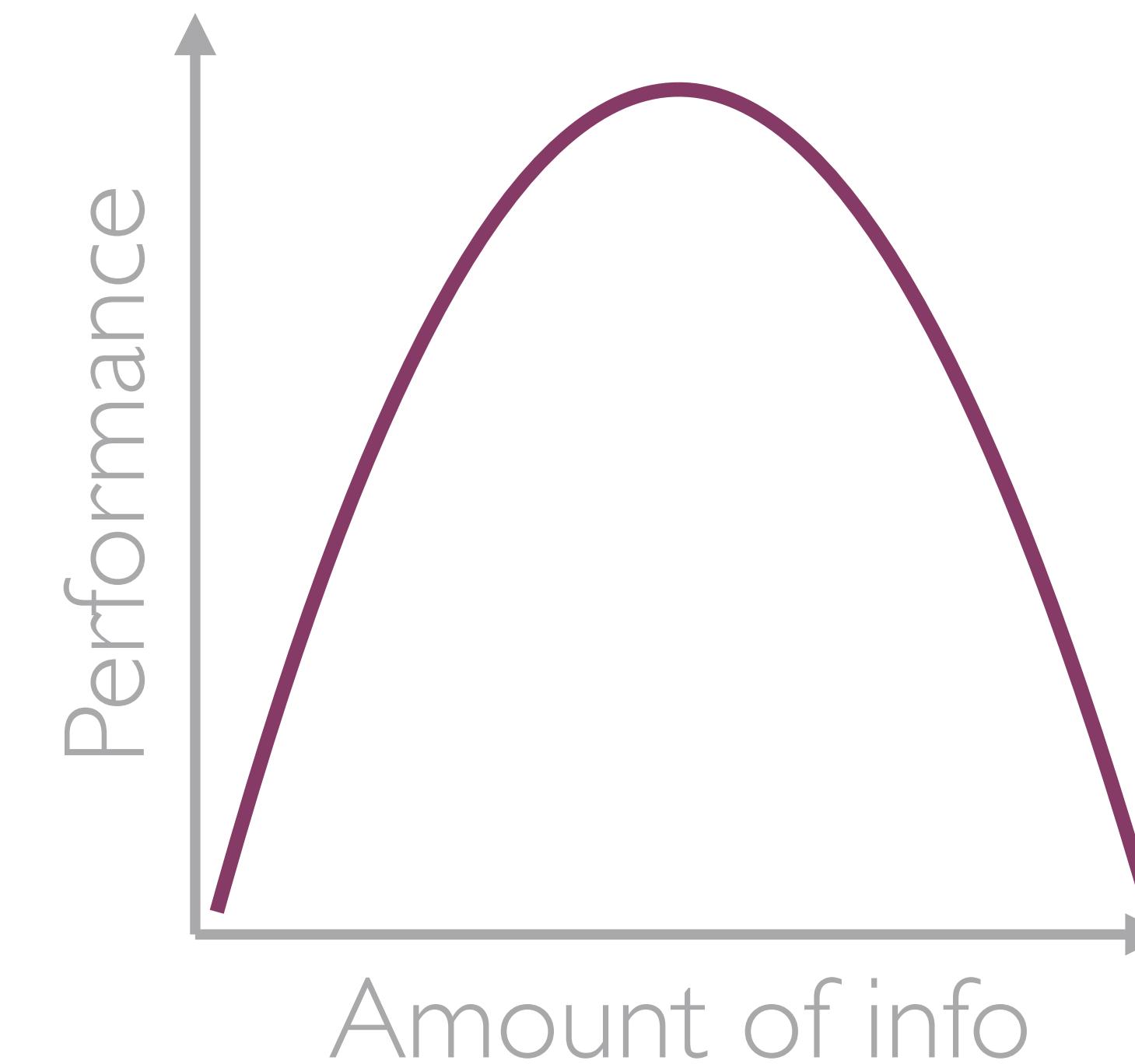
Yerkes-Dodson law



More information = higher performance Too much information overloads us

Information overload: performance

- Human decision-making improves with more content and information
- But past a point, we get saturated, and then performance starts decreasing



Information overload on social media

- As Usenet groups (VERY early online forums) grow in size, members respond to simpler messages, generate simpler responses, and are more likely to leave
- As a subreddit gets larger, members tend to comment on a smaller and smaller proportion of posts
- Fewer than half of Reddit's most popular links get noticed and upvoted the first time they are submitted

Jones, Q., Ravid, G., & Rafaeli, S. (2004). Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration. *Information systems research*, 15(2), 194-210.

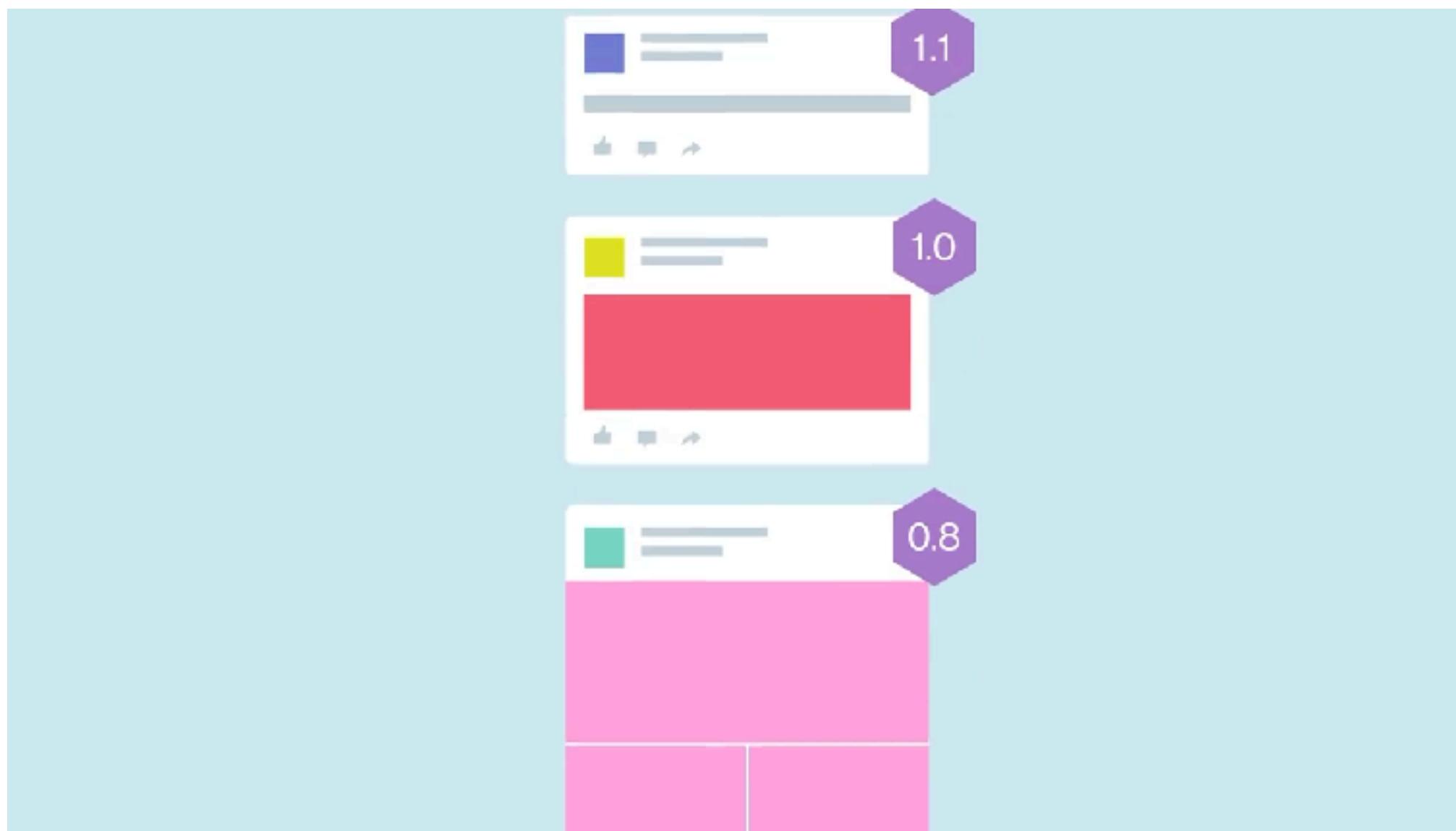
Lin, Z., Salehi, N., Yao, B., Chen, Y., & Bernstein, M. (2017, May). Better when it was smaller? community content and behavior after massive growth. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 11, No. 1, pp. 132-141).

Gilbert, E. (2013, February). Widespread underprovision on reddit. In Proceedings of the 2013 conference on Computer supported cooperative work (pp. 803-808).

How do we design for information overload?

Designing for information overload

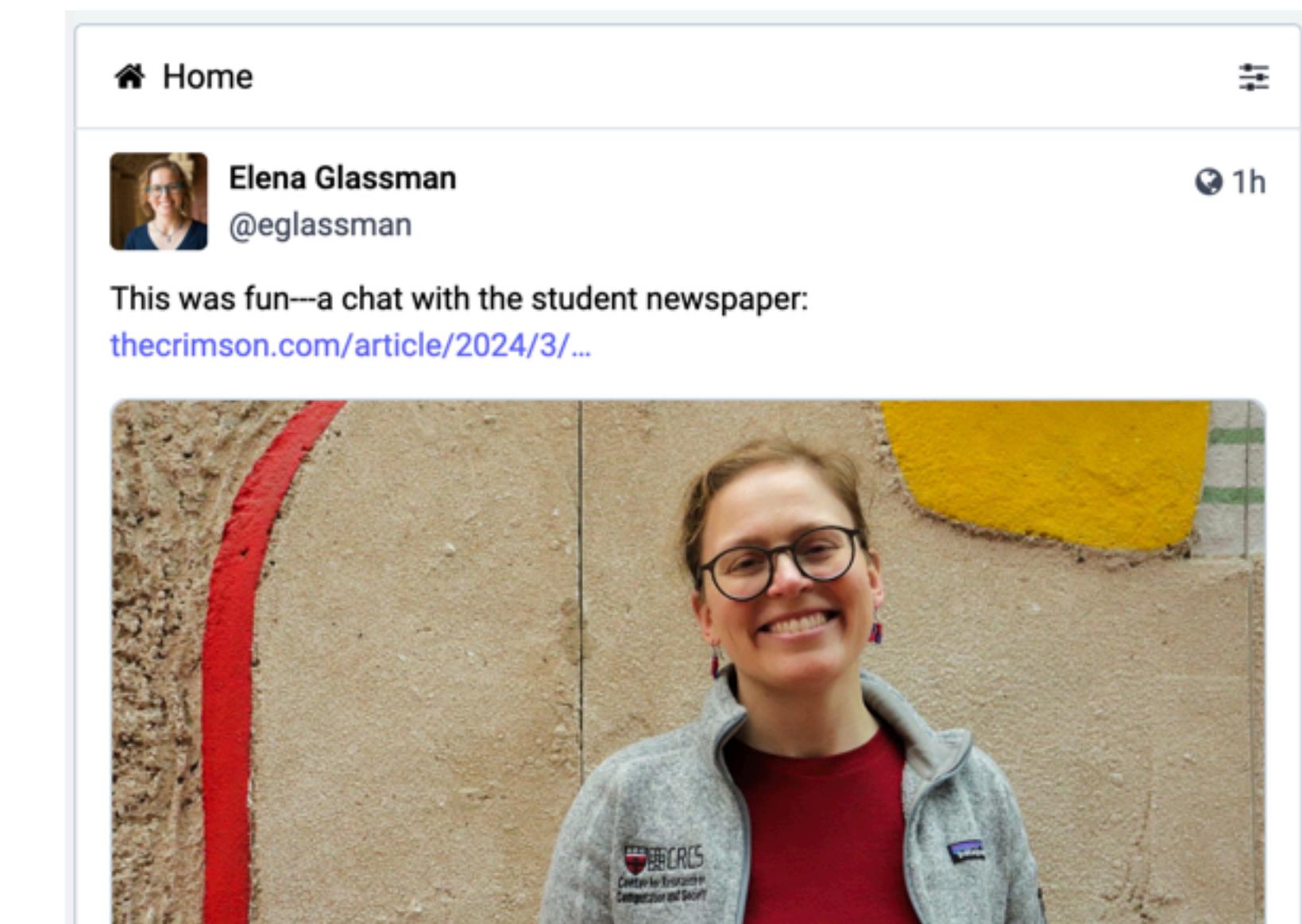
Ranking



Facebook
Instagram
Twitter/X

TikTok
Pinterest
Reddit

Chronological



Mastodon
Email
Slack

Discord
WhatsApp
iMessage

**Discuss: What are the advantages of a
“ranking” approach?**

What are advantages of a "ranking" approach in feeds?

Nobody has responded yet.

Hang tight! Responses are coming in.

**Discuss: What are the advantages of a
“chronological” approach?**

What are advantages of a "chronological" approach in feeds?

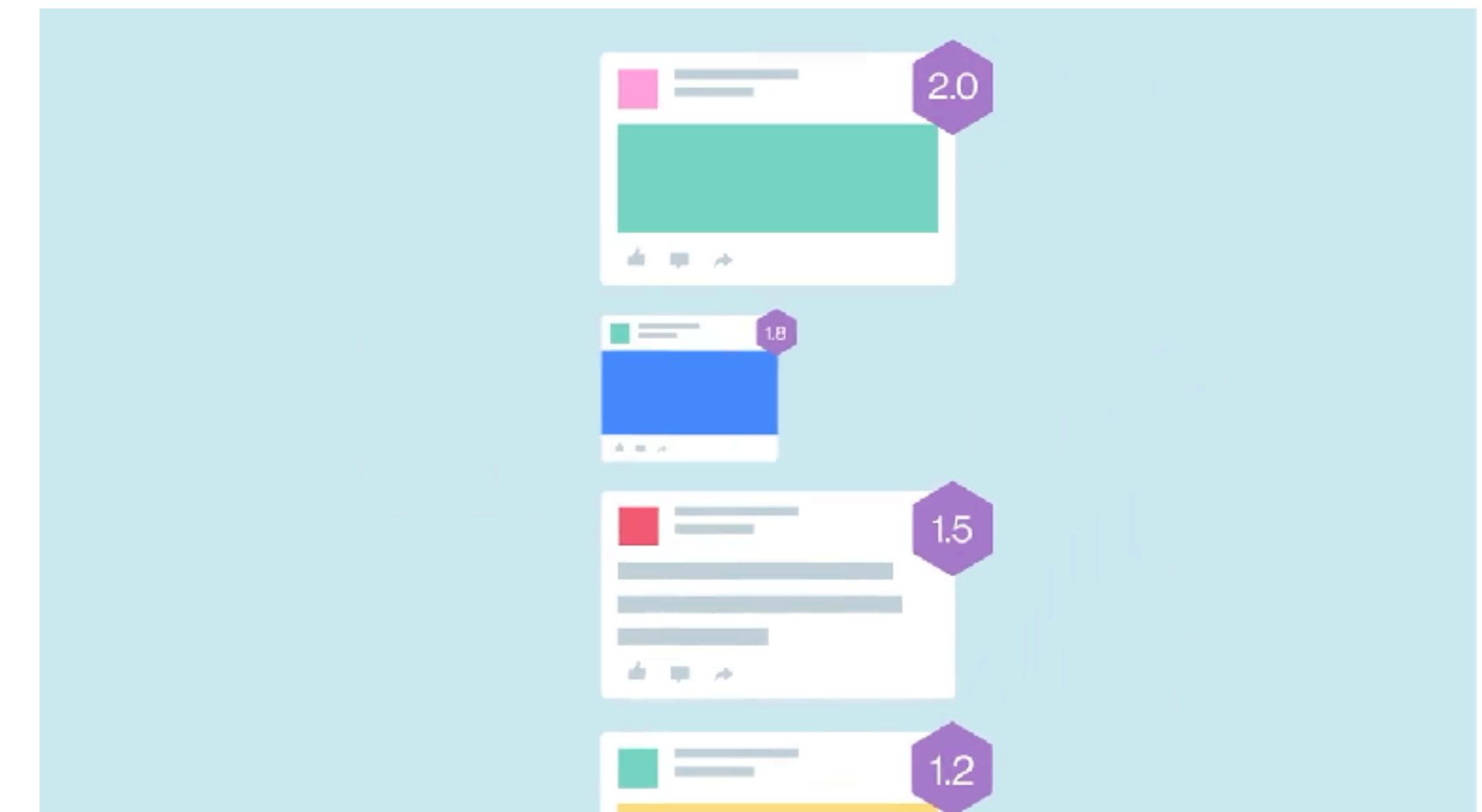
Nobody has responded yet.

Hang tight! Responses are coming in.

Designing for information overload

“Ranking” approach

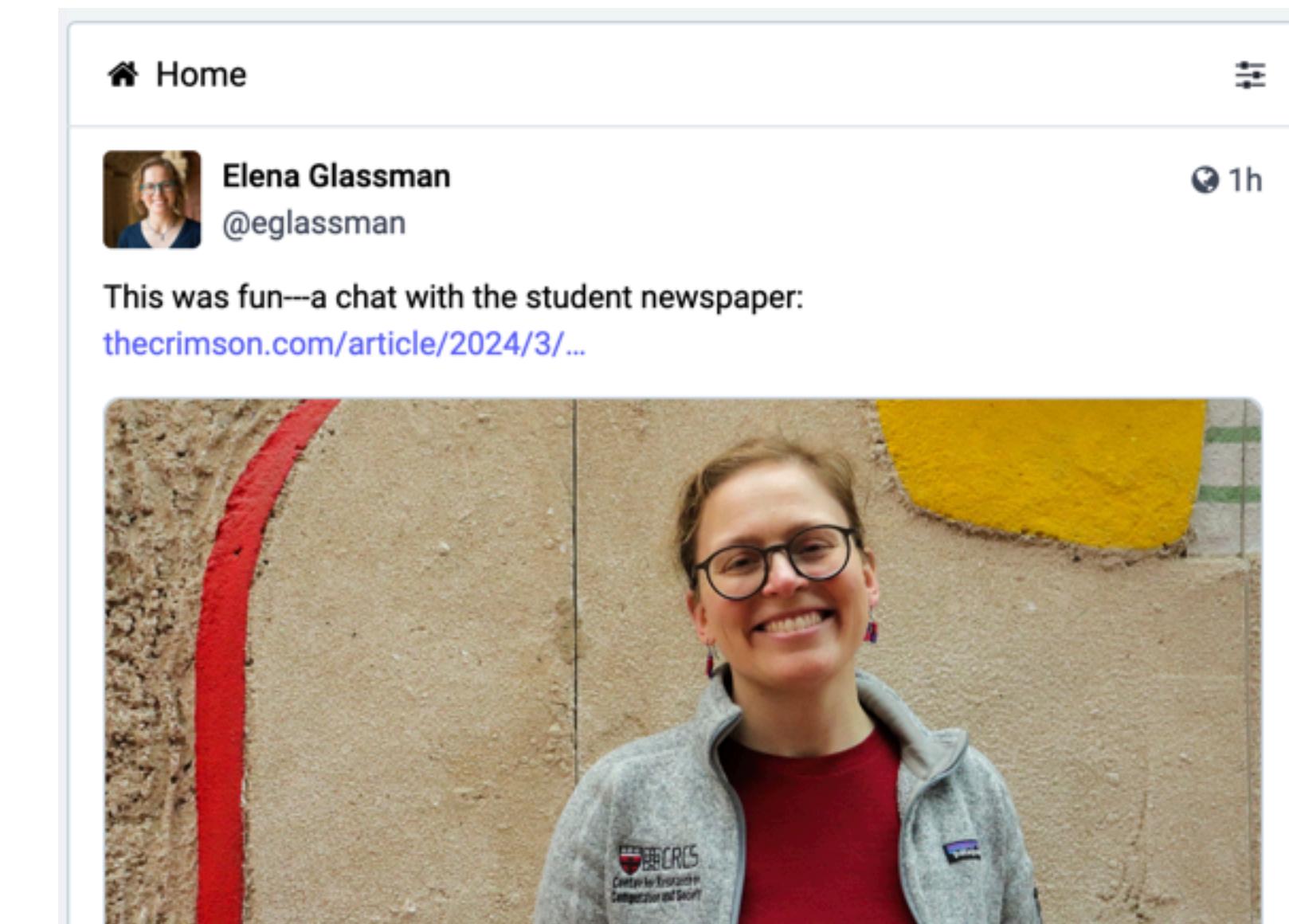
- When it's working well, it can bring what you want to the front
- But, it's unintuitive
- You won't see everything, which others might not expect



Designing for information overload

“Chronological” approach

- Intuitive: follows our expected mental model
- Accounts which post more show up more
 - Burdens the user with decision-making about what to follow or not



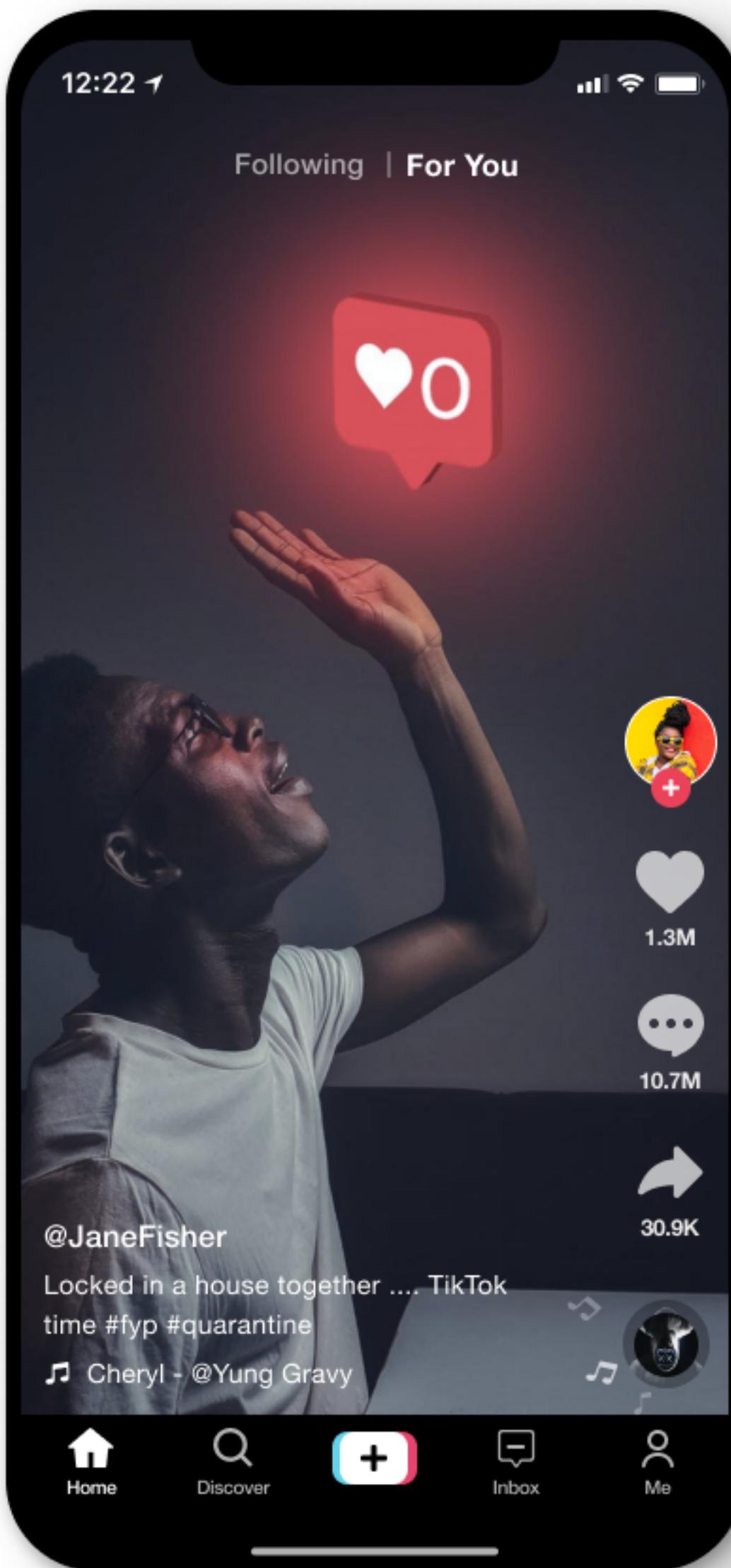
Designing for information overload

- “**Algorithms are unavoidable here.** Even sorting posts by friends in chronological order or videos by overall popularity is algorithmic; and often it is unclear there is a single, simple baseline algorithm.”
 - Dean Eckles, MIT, to the US Senate
 - Hearing on “algorithmic transparency and assessing effects of algorithmic ranking”

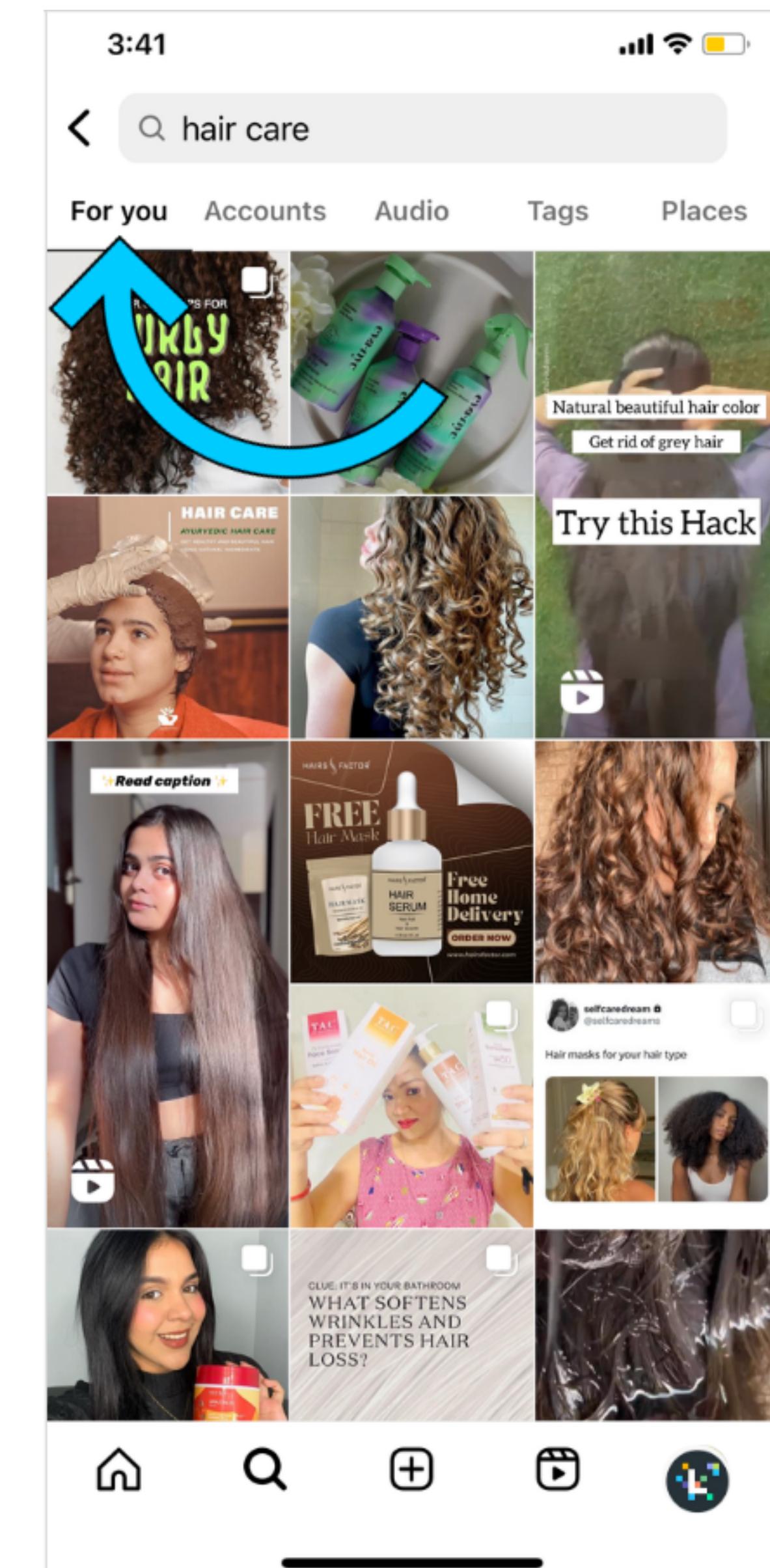
<https://www.commerce.senate.gov/services/files/62102355-DC26-4909-BF90-8FB068145F18>

How algorithmic feeds work

Feeds



A screenshot of the Twitter mobile application. The top navigation bar shows 'Home' and a search bar with the placeholder 'What's happening?'. Below the search bar are icons for photos, GIFs, and other media. A modal window titled 'Home shows you top Tweets first' is displayed, with options to 'See latest Tweets instead' or 'View content preferences'. The main feed shows a tweet from user Tom Neenan (@TNeenan) dated March 1, which reads: 'Important content.' Below the tweet is a large graphic with the text 'GET THE CLEANEST THUMBS IN THE LAND' and a series of eight small images showing hands being washed. The tweet has 3 likes, 20 retweets, 122 favorites, and a share icon. Another tweet from user Jesse Cox (@JesseCox) follows, with a duration of 45s and the caption 'Oof... well the stream is live, but I'm dropping frames like crazy. We'll see how this goes.'

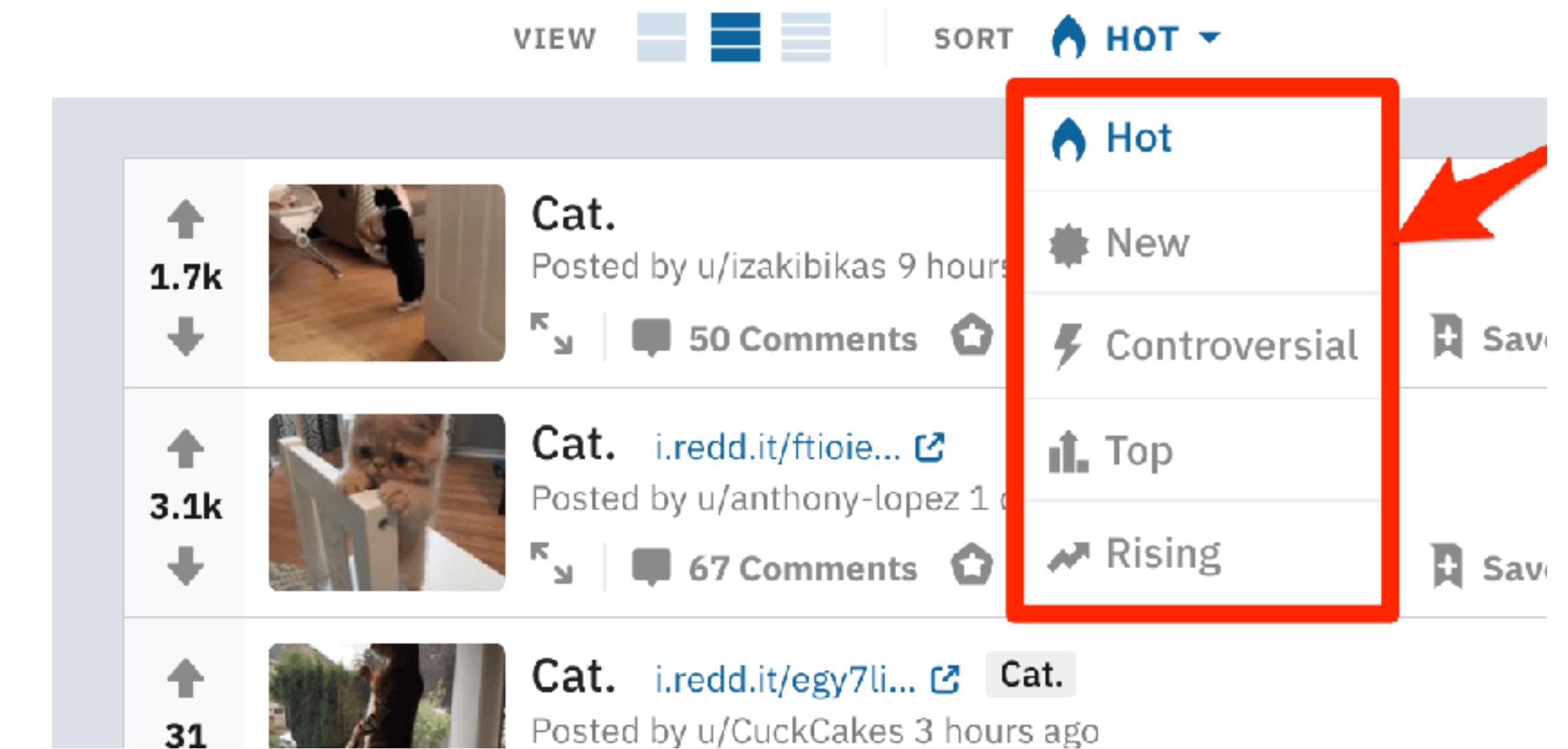


Feeds

- Two approaches:
 - Global ranking
 - Personalized feed

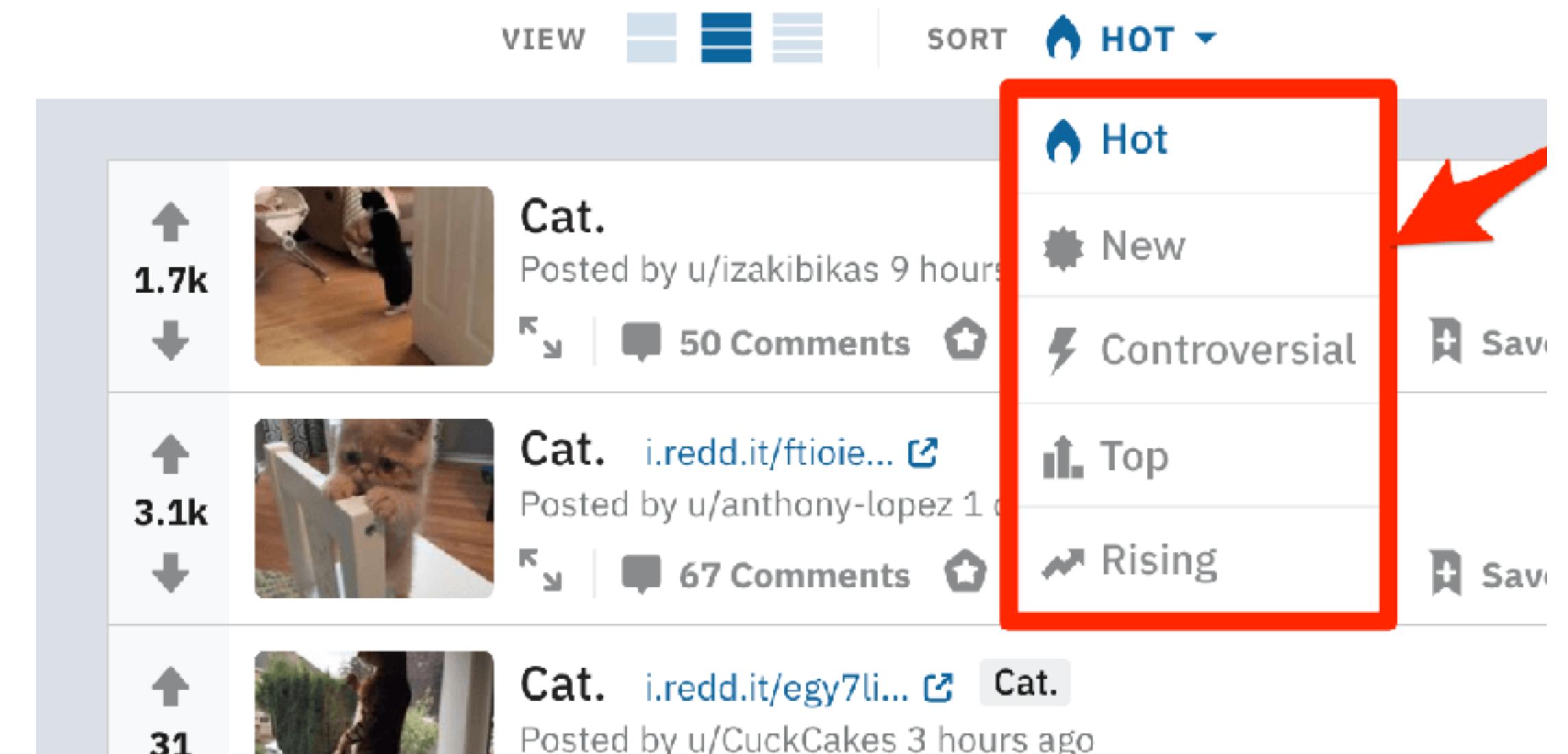
Global ranking

- Bring the most highly-rated posts to the top
- Think Reddit's “hot” ranking. Other sites like Hacker News do similar things



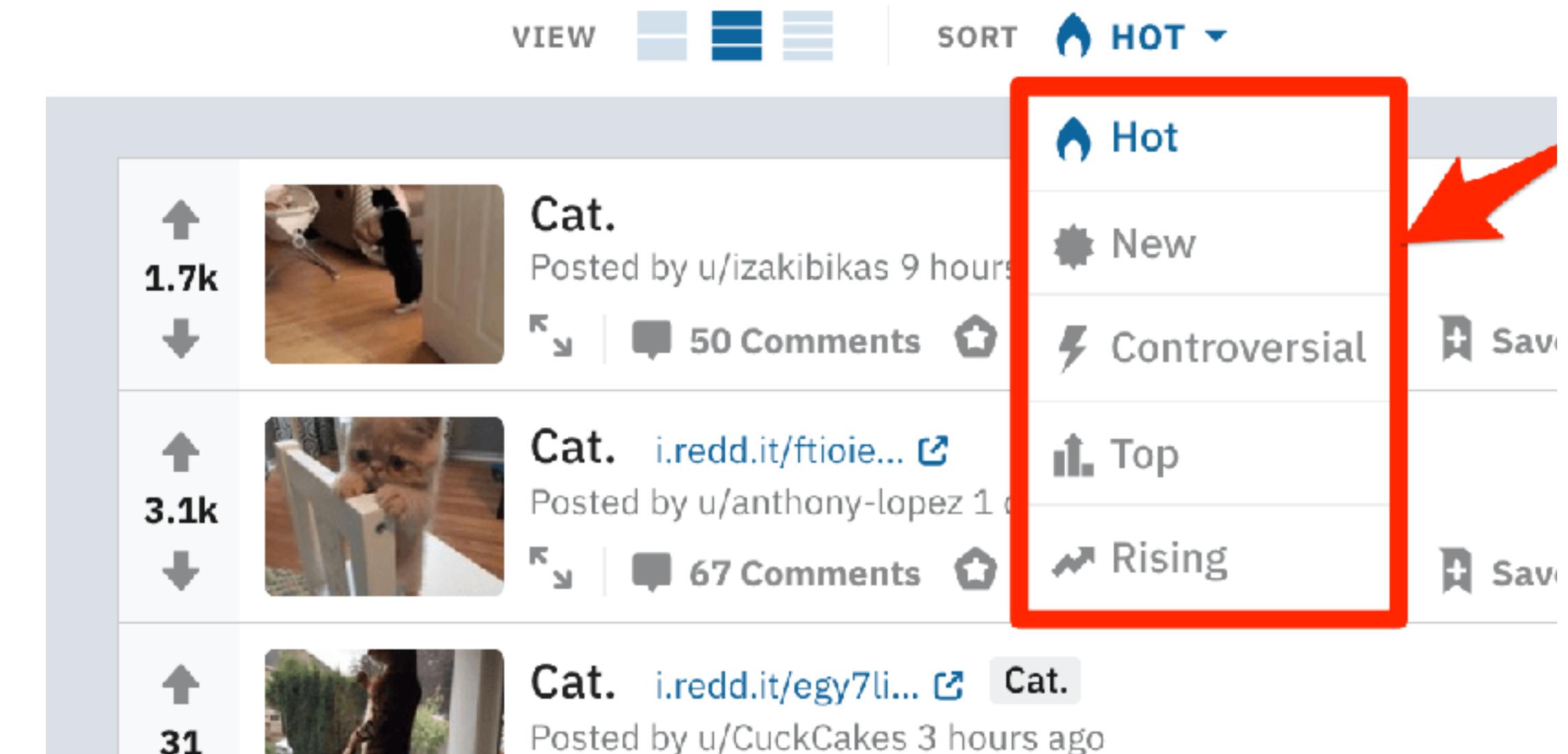
Global ranking

- First attempt: rank posts by the number of upvotes
 - E.g., a post with 1,000 upvotes should be rated more highly than a post with 100 upvotes
- What falls short here?
 - If lots of people saw the post, but a large percentage disliked it



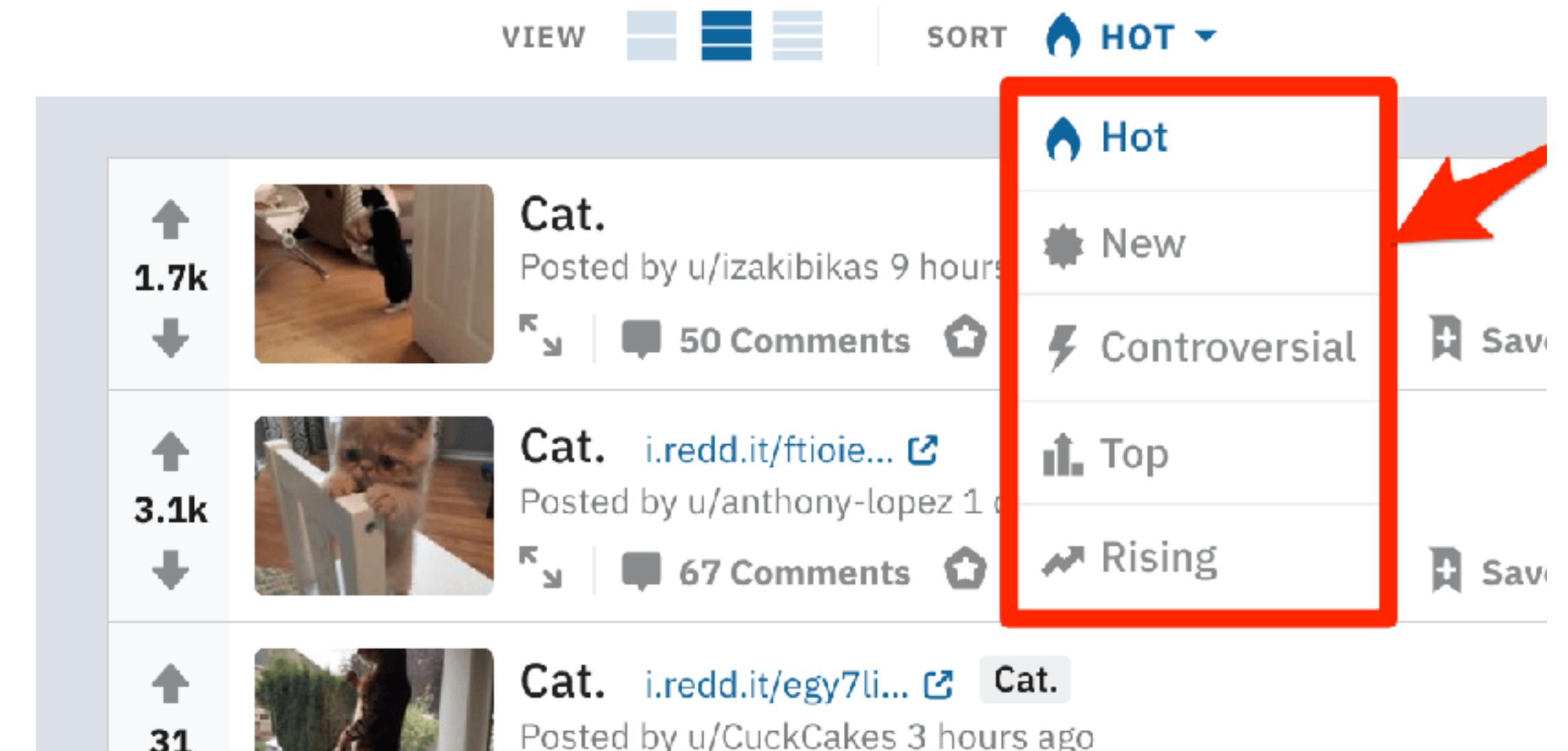
Global ranking

- Second attempt: rank posts by the number of upvotes - the number of downvotes
 - E.g., a post with 100 upvotes and 1 downvote should be rated more highly than a post with 1,000 upvotes and 1,000 downvotes
- What falls short here?
 - This ranking is static, the top posts would rarely change



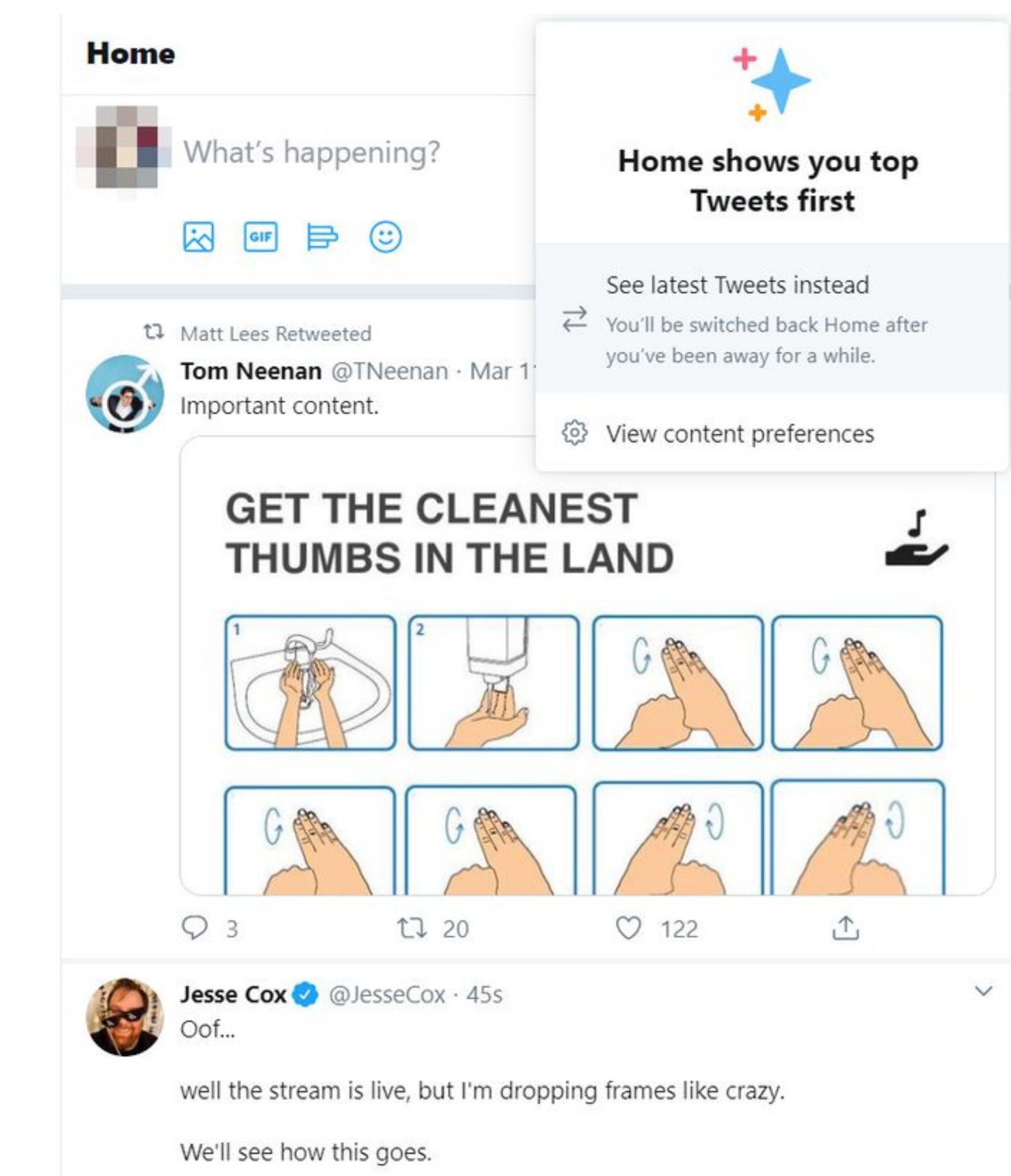
Global ranking

- Final attempt: decay rankings over time
 - $\log(\max(\text{upvotes}-\text{downvotes}, 1))$
 - Why log?
- And, decay the log score over time
 - Platforms do this differently, Reddit linearly decays the ranking over time



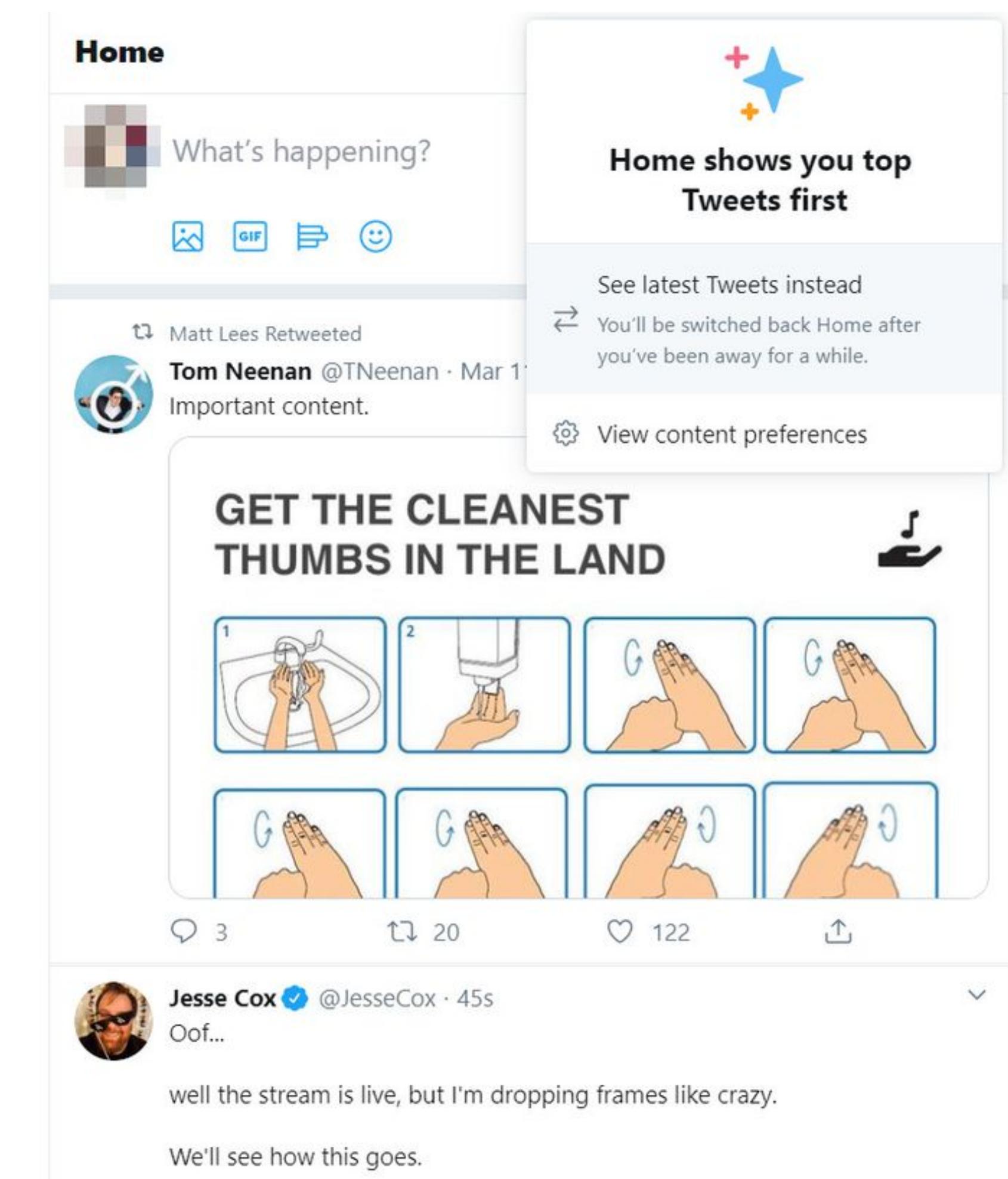
Personalized Feed

- Think TikTok's for you, Facebook or Twitter/X's feed
- Leverages some version of machine learning

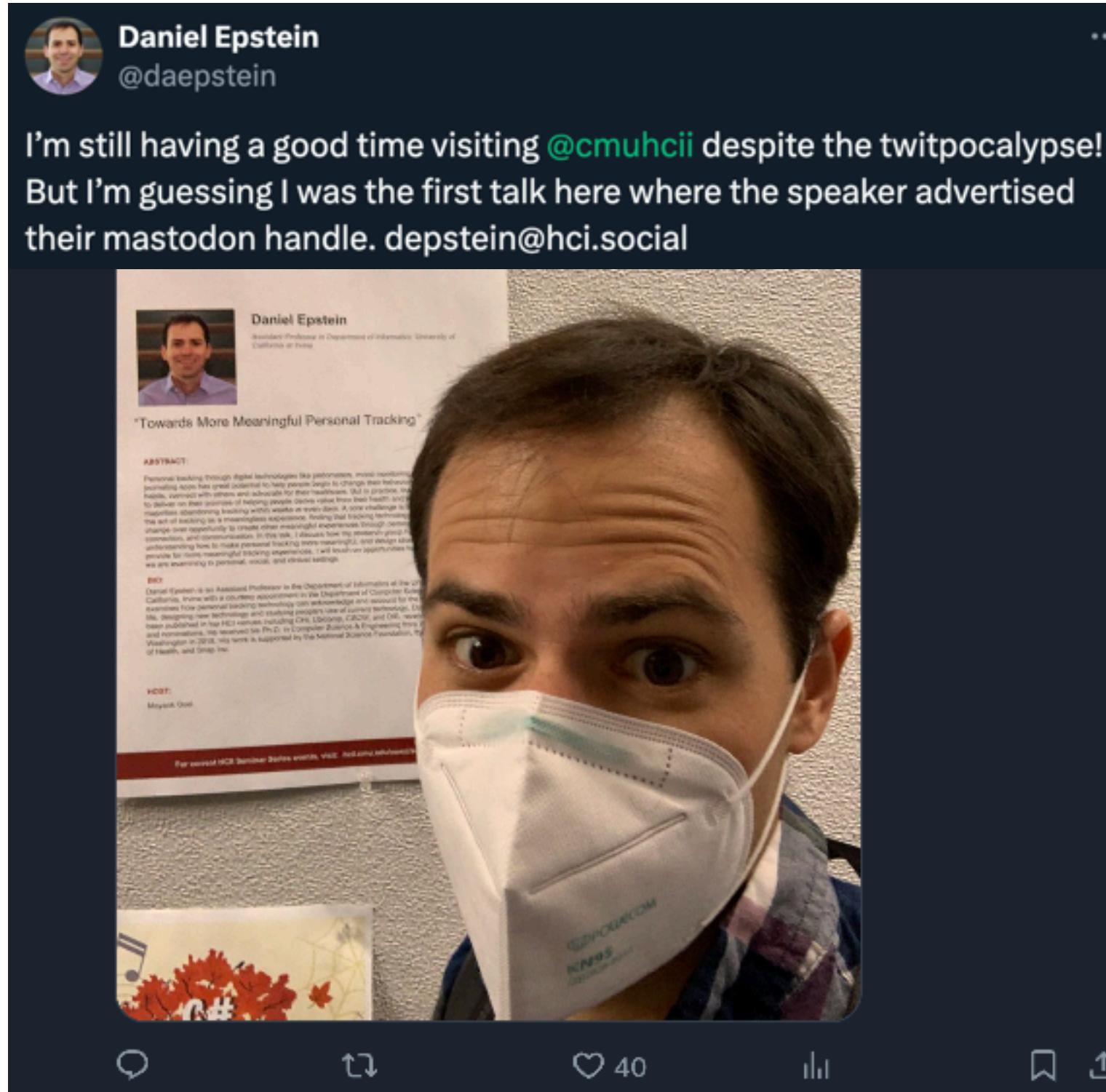


Personalized Feed

- Four steps:
 - Featurize
 - Predict
 - Calculate Objective
 - Rank



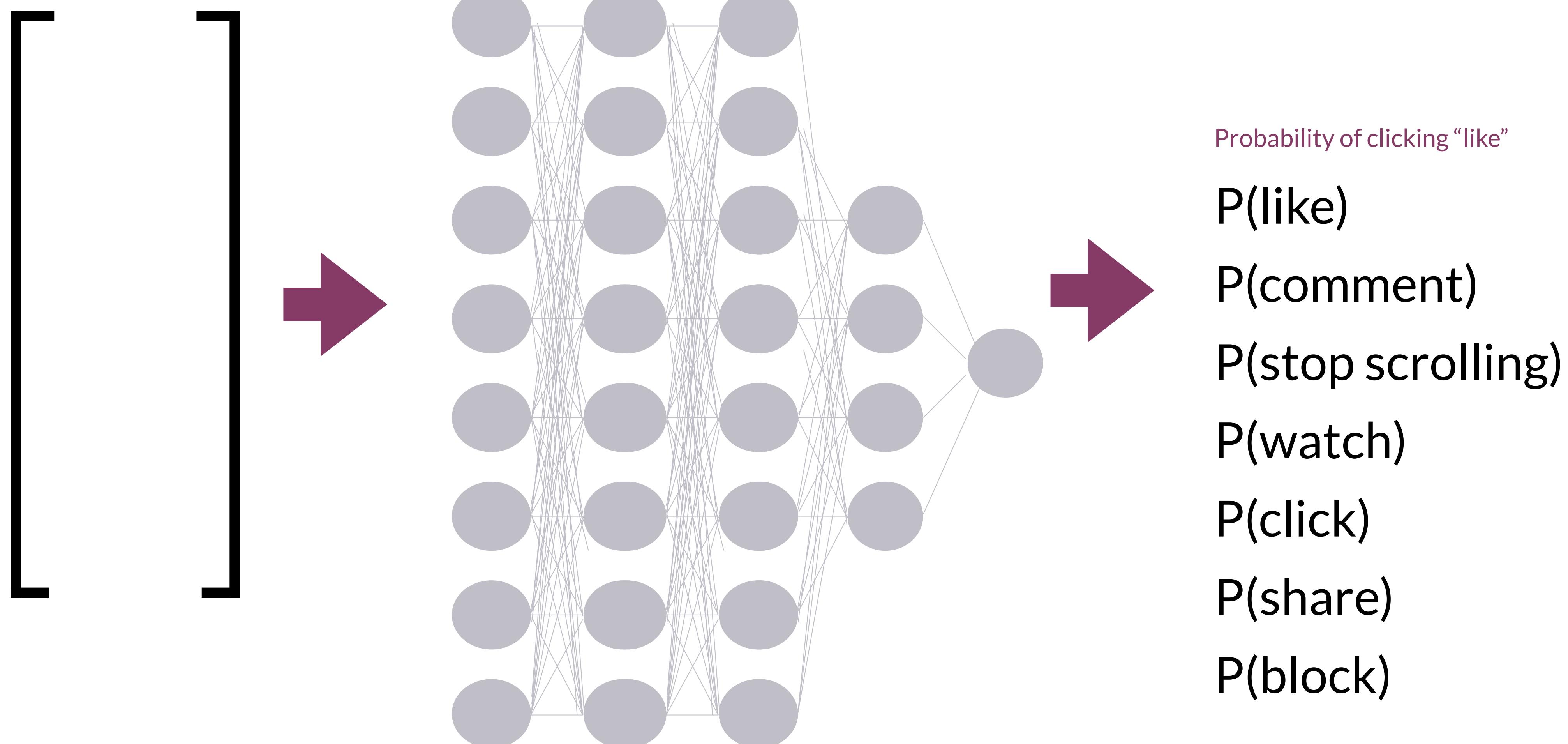
Personalized Feed: Featurize



Tie strength: 8
Contains image: yes
Image contains face: yes
Other objects: mask, paper
Positive keywords: 1
Negative words: 1
Platform: iPhone
Interactions so far: 40
Minutes since post: 52
Internet speed: 10 mbps

...

Personalized Feed: Predict



Personalized Feed: Predict

- How do we train these deep learning algorithms?
- Training data: prior behavior on the platform
 - The system inputs when you scroll, click, etc., and uses it to predict behaviors towards future posts you haven't seen before
 - Or, others like you. If someone else clicked on a similar link, they might share other interests/qualities, and those interests can help predict your interests

Personalized Feed: Calculate objective

- So, what do we do with these predictions?

P(like)

P(stop scrolling)

P(click)

P(block)

P(comment)

P(watch)

P(share)

- We define an **objective function**: a formula to combine and weight the predictions

- How many points do you think each predicted behavior should get?

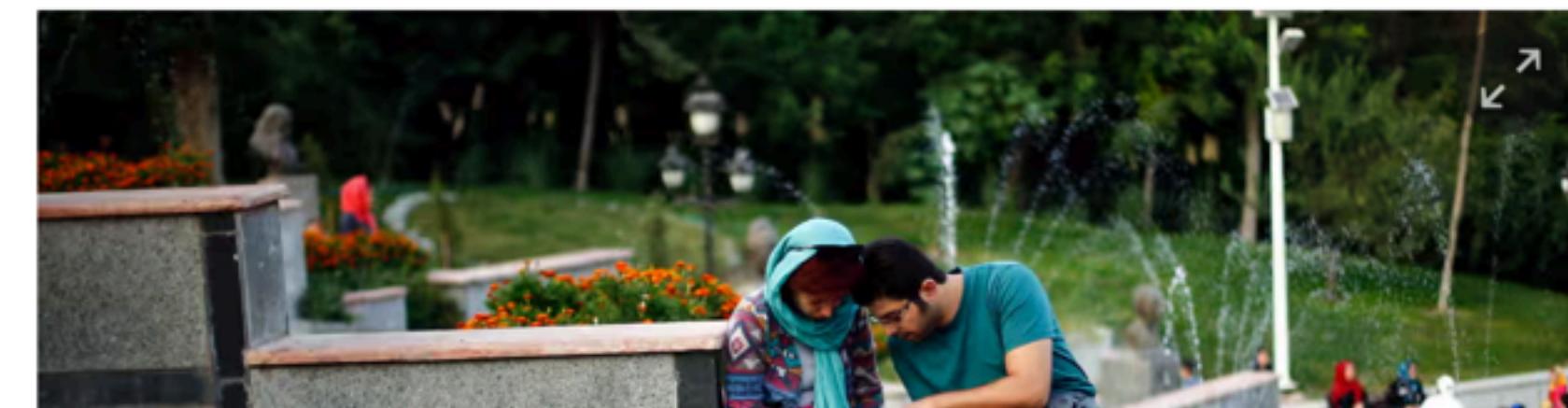
$$\sum_{p \in \text{predictions}} \text{weight}_p \cdot p$$

Quick tangent: objectives in practice

- Facebook: “Meaningful Social Interactions”
 - A weighted average of likes, reactions, reshares, and comments

Facebook overhauls News Feed in favor of 'meaningful social interactions'

Refresh of the News Feed algorithm will de-prioritize content shared by media and businesses in favor of that produced by friends and family, Zuckerberg says



Interaction type	weight
Like	1
Reaction	1.5
Reshare	1.5
Comment	15-20

<https://www.theguardian.com/technology/2018/jan/11/facebook-news-feed-algorithm-overhaul-mark-zuckerberg>
<https://knightcolumbia.org/content/understanding-social-media-recommendation-algorithms>

Quick tangent: objectives in practice

- Twitter's open-sourced algorithm:

75 points if predicted that, if you reply, the author will reply back

27 points if predicted that you'd reply

12 points if predicted that you engage with the author's Twitter profile

1 point if predicted that you'll retweet

0.5 points if predicted that you'll favorite

-74 points if predicted that you'll give negative feedback ("not interested", mute, block)

-369 points if predicted that you'll report it



Jeff Allen
@jeff4llen

...

According to the Heavy Ranker readme, it looks like this is the "For you" feed ranking formula is

Each "is_X" is a predicted probability the user will take that action on the Tweet.

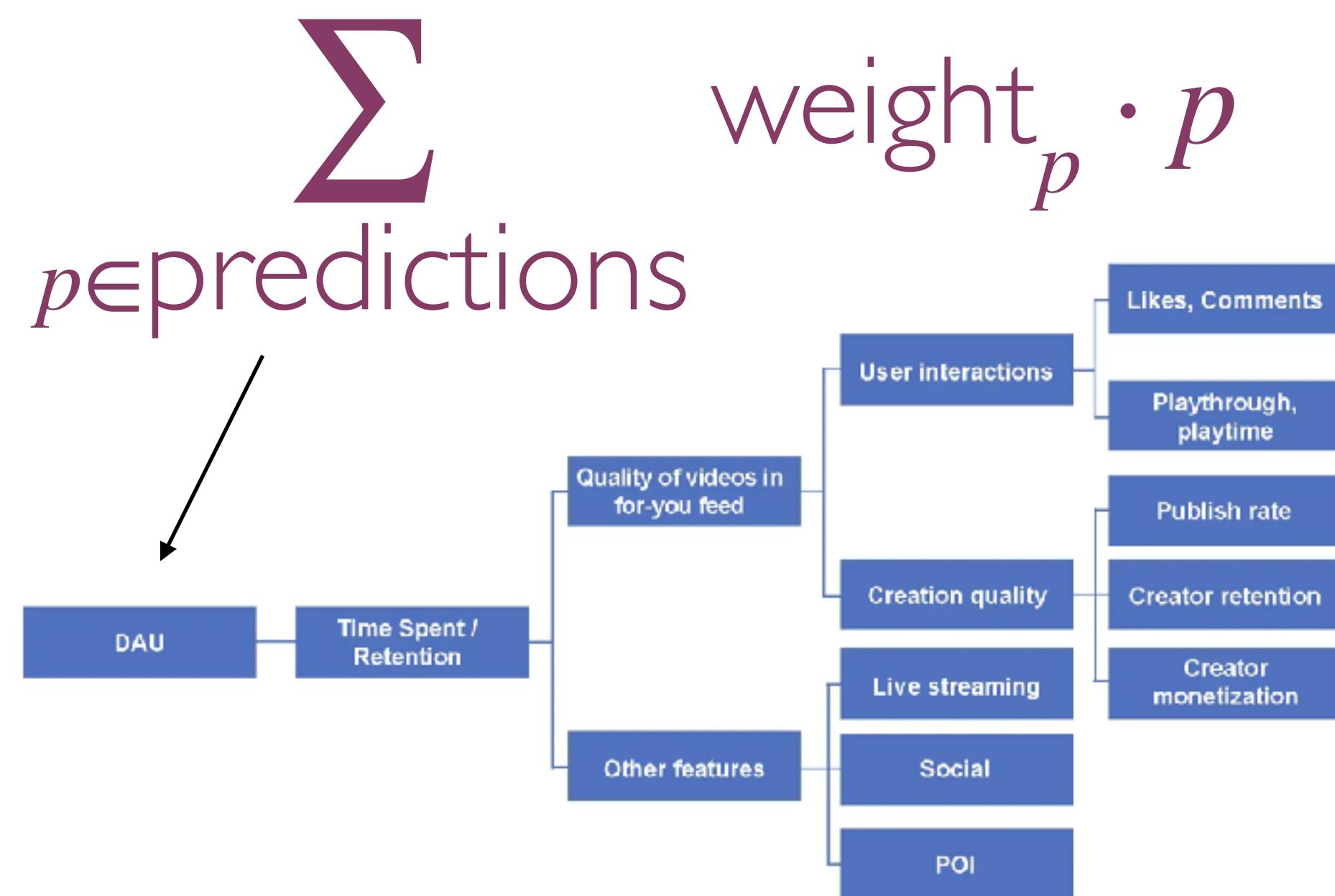
Replies are the most important signal. Very similar to MSI for FB.
github.com/twitter/the-al...

```
Twitter Ranking Score =  
  75 * is_replied_reply_engaged_by_author  
+ 27 * is_replied  
+ 12 * is_profile_clicked_and_profile_engaged  
+ 11 * MAX(  
            is_good_clicked_convo_desc_favorited_or_replied,  
            is_good_clicked_convo_desc_v2  
        )  
+ 1.0 * is_retweeted  
+ 0.5 * is_favorited  
+ 0.005 * is_video_playback_50  
- 74 * is_negative_feedback_v2  
- 369 * is_report_tweet_clicked
```

1:35 PM · Mar 31, 2023 · 56.9K Views

Quick tangent: objectives in practice

- TikTok: some combination of viewing, liking, and commenting



THE MEDIA EQUATION

How TikTok Reads Your Mind

It's the most successful video app in the world. Our columnist has obtained an internal company document that offers a new level of detail about how the algorithm works.

Share full article



<https://www.nytimes.com/2021/12/05/business/media/tiktok-algorithm.html>

Feeds: driven by engagement

- Engagement is typically shorthand for behaviors that the platform can observe, such as likes and comments
- But, optimizing for engagement can create negative outcomes.

**Discuss: what negative outcomes might result from optimizing for engagement?
How might we account for them?**

More recently: global objectives

- Indirect impacts: if we show this post to you, and you leave a comment, will it make a better or worse experience for the person who posted it?
 - Will you post something encouraging?
- Long-term impacts: what impact will seeing this post have on your health and wellbeing?

News Feed quality: We've made several changes to News Feed to provide more opportunities for meaningful interactions and reduce passive consumption of low-quality content — even if it decreases some of our engagement metrics in the short term. We demote things like clickbait headlines and false news, even though people often click on those links at a high rate. We optimize ranking so posts from the friends you care about most are more likely to appear at the top of your feed because that's what people tell us in surveys that they want to see. Similarly, our ranking promotes posts that are personally informative. We also recently redesigned the comments feature to foster better conversations.

More recently: global objectives

- Impacts estimated via survey
 - Inject a survey into a feed to have people rate how a particular piece of content makes them feel, whether it's important, informative, funny, etc.
 - Use that rating to predict perspectives of others
- Feed diversity
 - Penalize content that is too similar to other content in a person's feed



Cheng Zhang, Software Engineer, and Si Chen,
Software Engineer

Personalized Feed: Rank

- Score all posts, and rank posts by their score
- Look at outcomes, see what's going well and poorly, and rethink the cycle
 - Add new features
 - Add new objectives
 - Change the weighting of different objectives

So, how do decisions about the ranking process get made?

Making decisions

- Typically, platforms run an A/B test on a subset of users to see the impact of any ranking changes on evaluation metrics
- In practice, this typically aims to improve on desirable metrics without harming others

So why is TikTok's feed so good?

Why is TikTok's feed so good?

- It's not the algorithm, it's the signals
 - Unlike other platforms, TikTok will serve you content from anyone, not just accounts that you follow
 - TikTok “clusters” users and videos based on their preferences and content
 - This lets it operate more like a recommender system (Netflix) than a social media site
- The algorithm identifies similar videos to those that have engaged a user based on video information, which could include details like captions, hashtags or sounds. Recommendations also take into account user device and account settings, which include data like language preference, country setting, and device type.
 - Once TikTok collects enough data about the user, the app is able to map a user's preferences in relation to similar users and group them into "clusters." Simultaneously, it also groups videos into "clusters" based on similar themes, like "basketball" or "bunnies."
 - Using machine learning, the algorithm serves videos to users based on their proximity to other clusters of users and content that they like.
 - TikTok's logic aims to avoid redundancies that could bore the user, like seeing multiple videos with the same music or from the same creator.

<https://wwwaxios.com/2020/09/10/inside-tiktoks-killer-algorithm>

Does all of this mean that feeds are echo chambers?

Feeds as filter bubbles

- Filter bubbles occur when people are only shown content that they like
 - This is a natural outcome when your platform optimizes for engagement
- Example: YouTube channels that are slightly less mainstream become recommendation gateways to more and more radical channels

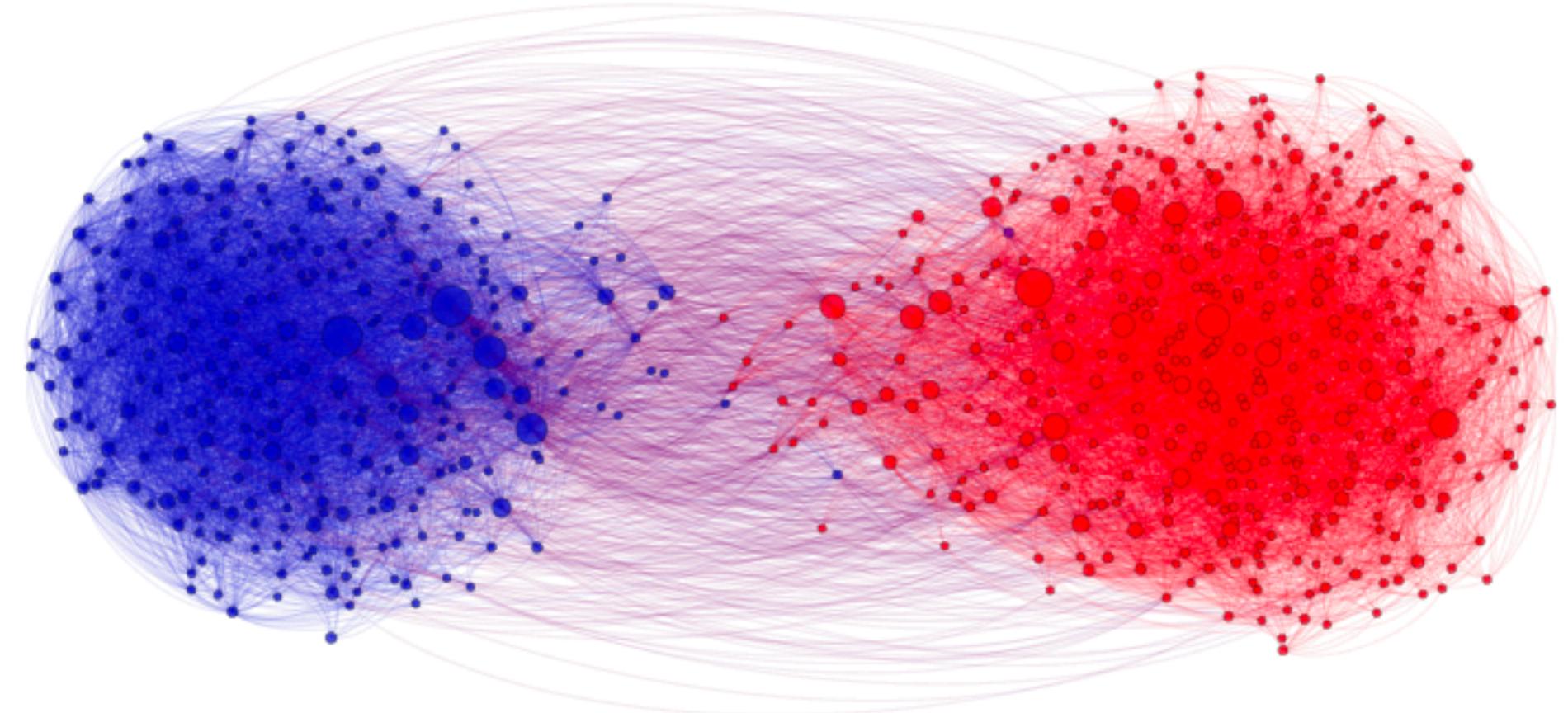


Figure 6: Recommendation graph of YouTube channels.

Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A., & Meira Jr, W. (2020, January). Auditing radicalization pathways on YouTube. In Proceedings of the 2020 conference on fairness, accountability, and transparency (pp. 131-141).

Feeds as echo chambers

- If *your* feed only shows you the things that *you* want to see, and *my* feed only shows me the things that *I* want to see...
 - Won't that result in an echo chamber, where we only hear people who share our opinions?
 - Won't this further polarize our society?

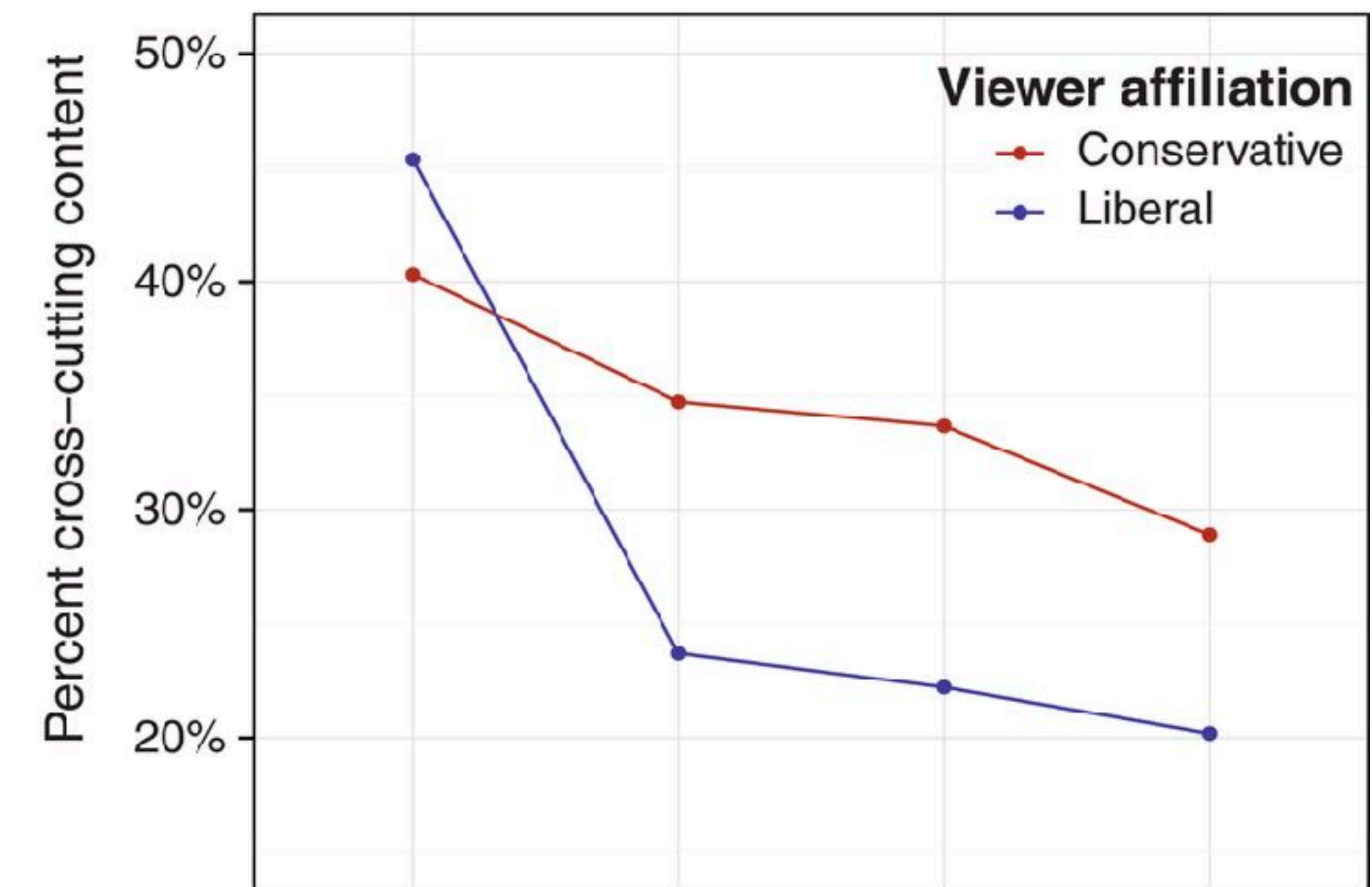


Adamic, L. A., & Glance, N. (2005, August). The political blogosphere and the 2004 US election: divided they blog. In Proceedings of the 3rd international workshop on Link discovery (pp. 36-43).

Feeds as echo chambers

It's more complicated

- From Facebook researchers, a study of log data to understand political news in people's feeds
- The biggest drop is *homophily*: we tend to only friend people who share our views
- The feed itself has a much more minor impact



Baseline: If my inventory feed is everything content from my friends random FB users
Clicked: what my feed actually shows

Feeds as echo chambers

It's more complicated

- People who use social media are exposed to more cross-cutting ideological news than those who don't
- Having people subscribe to news sources that don't align with their views have extremely minimal influence on their perspective on those views
- Instead, we might have it backward.
 - It's not that we're more polarized because social media only exposes us to *similar* viewpoints, but rather because social media exposes us to a *wider variety* of people

Fletcher, R., & Nielsen, R. K. (2017). Are news audiences increasingly fragmented? A cross-national comparative analysis of cross-platform news audience fragmentation and duplication. *Journal of communication*, 67(4), 476-498.

Levy, R. E. (2021). Social media, news consumption, and polarization: Evidence from a field experiment. *American economic review*, 111(3), 831-870.

Törnberg, P. (2022). How digital media drive affective polarization through partisan sorting. *Proceedings of the National Academy of Sciences*, 119(42), e2207159119.

Summary

- One strategy for managing information overload in social media feeds is to filter what content they show to people
 - Global rankings aggregate up/downvotes, and trail off over time
 - Personalized rankings predict on-platform behaviors, assign weights to each predicted behavior to determine a score, and rank according to that score
- A lot has been said about feeds creating filter bubbles and echo chambers. While there are clearly negative outcomes to feeds, the story is more complicated, and our understanding of the topic is still forming

Today's goals

By the end of today, you should be able to...

- Articulate what a feed is, and how it is represented in different social media sites
- Describe the importance of filtering and sorting algorithms to social media content represented in feeds
- Explain how a typical feed algorithm works, at a high level
- Discuss whether feeds create filter bubbles or echo chambers, particularly around online discourse

IN4MATX 153: CSCW

Class 3:
Feeds and their Algorithms

Professor Daniel Epstein
TA Dennis Wang
Reader Weijie Du