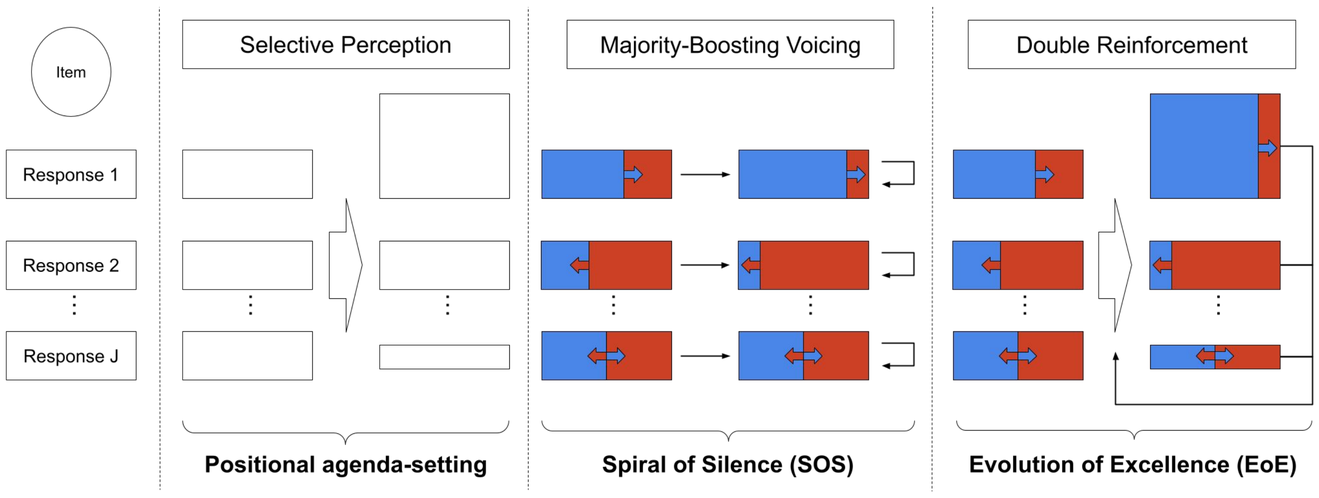
From Popularity to Meritocracy: Information Monopoly and Evolution of Excellence in Online Communities

Moontae Lee, Chang Liu, Ali Tafti

Background:

Helpfulness votes are the currency of evaluation on online platforms. However, they cannot represent the excellence of information in aggregate due to **Position and Herding biases** inherent in voting process.
Additionally, their doubly reinforcing **the-rich-get-richer dynamics** accelerate the popularity of the perceived responses again, transforming excellent information into monopolies of merit.



Contributions:

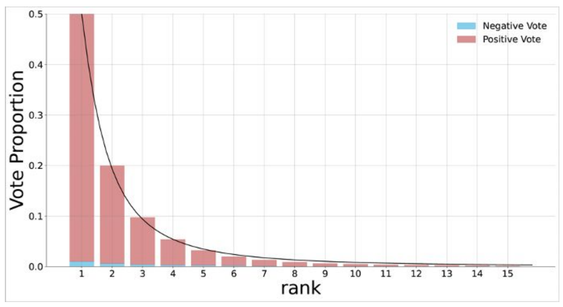
- Proposed the Evolution of Excellence (EoE) framework to explain and disentangle position bias, herding bias, and their mutual reinforcement in online evaluations.
- Developed easy-to-compute quantitative measurements for both biases **without** requiring A/B testing or model training.
- Incorporated historical voting trajectories in the measurement design, instead of only using static snapshot vote data.
- Identified behavioral patterns across different communities and information types, based on their varying sensitivities to the two biases.
- Demonstrated over-year behavioral evolution of individual communities
- Introduced novel metrics to detect information monopoly and assess meritocracy risk, supporting platform-specific interventions.

Data:

Stack Exchange Data Dump
by Stack Exchange, Inc.

- 2008 ~ 2022
- *360 Q&A communities*
- *Largest 120 communities dataset for experiments*
- *Reconstructed vote trajectory of each answer at each time*

Measurement of Position bias: Trendiness

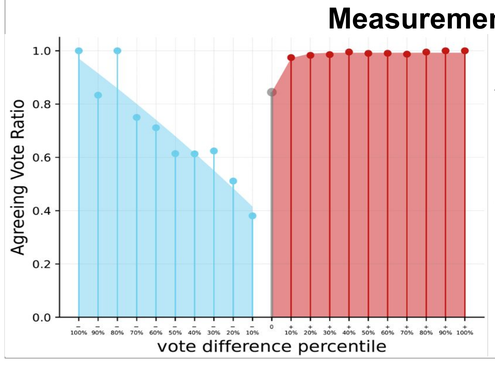


StackOverflow distribution of vote over ranks

To show that votes cast at higher rank is more than that at lower rank.

Fitting function		Trendiness
Simplified power law	$\frac{1}{rank^{b}+1} + c$	**b**

Measurement of Herding bias: Conformity

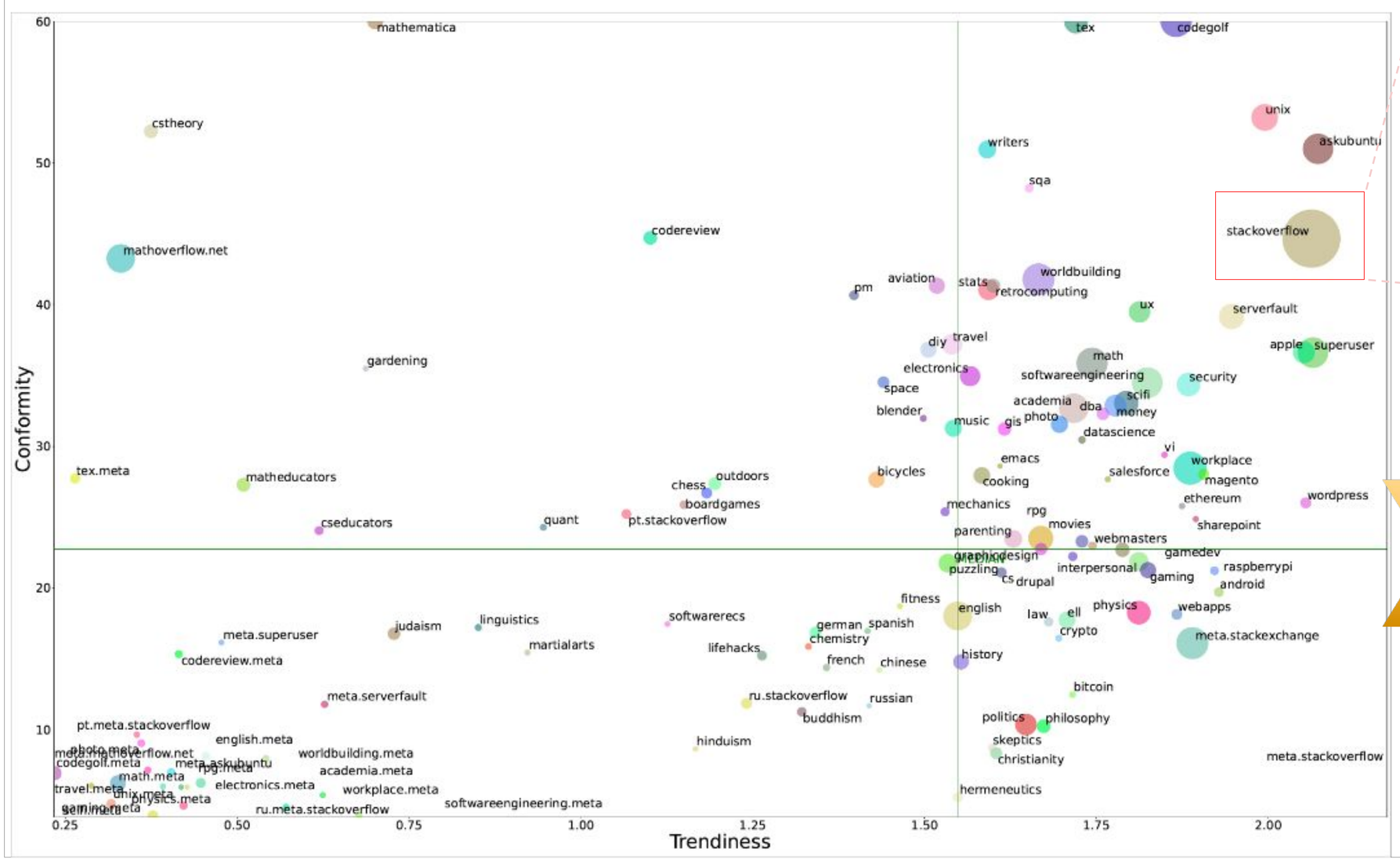


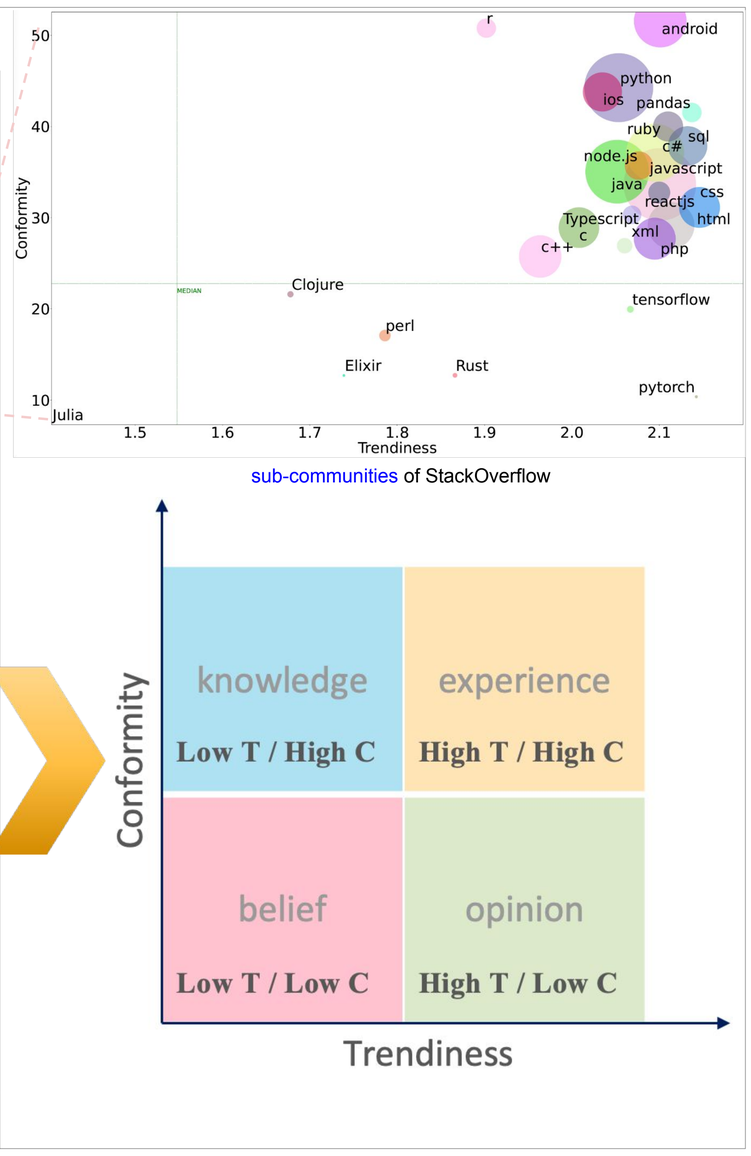
MathOverflow distribution of agreeing vote raio over vote difference scores

To show that when the majority opinion is clearer, there are more agreeing votes.

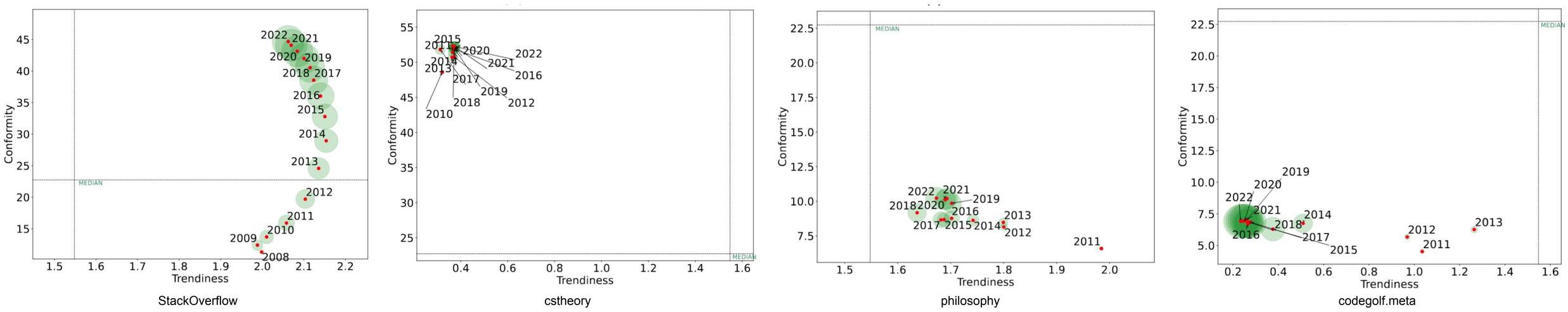
$$Conformity = \left\{ \prod_{i} \prod_{t_i} \frac{P_{ij}^{t_i}(agree\ vote)}{P_{ij}^{t_i}(against\ vote)} \right\}^{1/n}$$

Map the communities into 4 types





Over-year behavioral evolution



Information Monopoly

Information Monopoly as Market Concentration
- **question** : industry
- **answers**: competing companies
- **votes**: product sales
- **aggregate vote proportion**: market share

- Market Concentration Ratio (MCR):

$$MCR(k) = \sum_{j=1}^{k} MS_{ij}$$

- Herfindahl-Hirschman Index (HHI):

$$HHI(k) = \sum_{j=1}^{k} MS_{ij}^{2}$$

Evidence of Meritocracy Risk

If the ranking are based on pure meritocracy without biases, after the monopoly formed, the chance for the dominating answers to receive negative votes should be close to zero. The chance for non-dominating answers to receive positive votes should be very small.

- **MRI (−)** : the Rate of **Negative** votes cast to the **dominating** k answers after no more new answer created.
- **MRI (+)**: the Rate of **Positive** votes cast to the **Non-dominating** k answers after no more new answer created.

Both between [0,1]. Higher value means higher risk of meritocracy

In all 8 communities, top 3 dominating answers have received over 60% of votes showing the monopoly, and there are at least 10% of negative votes cast to dominating answers even after monopoly formed which contradicts the principle of meritocracy.

Community	k=3 MCR	k=3 HHI	k=3 $MRI(-)$	k=3 $MRI(+)$
stackoverflow	0.81	0.30	0.1626	0.2521
askubuntu	0.81	0.28	0.1221	0.2547
mathoverflow	0.65	0.18	0.0967	0.4364
cstheory	0.64	0.16	0.1039	0.4321
politics	0.72	0.22	0.2918	0.4114
philosophy	0.80	0.27	0.2032	0.3319
math.meta	0.59	0.14	0.3211	0.4538
codegolf.meta	0.58	0.14	0.3565	0.4593

Conclusion:

- Our findings can help service providers secure community-specific interventions that reduce the effects of the two biases and promote meritocracy
- For communities highly sensitive to herding bias, we recommend hiding previous vote summaries
- For communities prone to position bias, we suggest varying the display order of answers
- the proposed measurements can be applied to other platforms