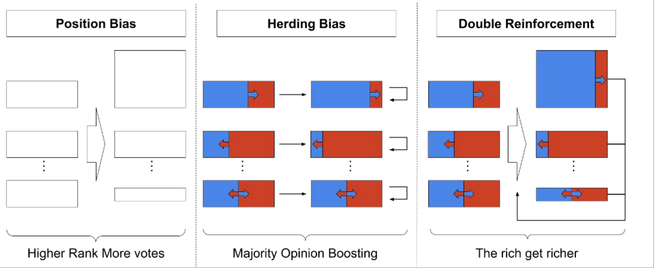
Counterfactual Voting Adjustment for Quality Assessment and Fairer Voting in Online Platforms with Helpfulness Evaluation

Background:

Efficient access to high-quality information is crucial for online platforms. But aggregated helpfulness votes are biased.



Research Goals:

- To mitigate position bias and herding bias at the same time
- To infer fairer qualities of responses with only observational data
- To verify the learned-quality-based ranking is better

Counterfacural Questions:

- **What if the information were presented at different ranks?**
- **What if the information had received different votes?**

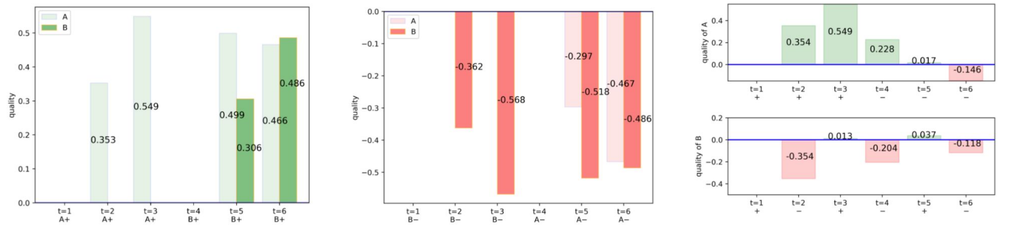
Result summary for 8 representative communities:

Comm		Semi-Synthetic voteDiff	Semi-Synthetic CVP	Semi-Synthetic CVA	Real (comment sentiment) voteDiff	Real (comment sentiment) CVP	Real (comment sentiment) CVA	Real (answer helpfulness) voteDiff	Real (answer helpfulness) CVP	Real (answer helpfulness) CVA
reactjs(SOF)	KT	0.3035	0.4089	**0.4106******	0.0383	0.1301	**0.1524******	0.0805	0.1575	**0.1819*****
	Res	17015	13458	**13364******	7936	6716	**6509******	7514	6551	**6295******
askubuntu	KT	0.2620	0.3603	**0.3691******	0.0266	0.0950	**0.1180****	0.0580	0.1116	**0.1823*****
	Res	10987	9337	**9098******	6769	6245	**6222****	6631	6122	**5753******
mathoverflow	KT	0.2774	0.3759	**0.3902******	0.1034	0.0579	**0.1160**	0.1055	0.0276	**0.1495***
	Res	10438	8232	**8024******	8280	8528	**7940***	8408	8961	**7679*****
cstheory	KT	0.3909	0.4640	**0.4659****	0.0788	0.1275	**0.1317**	0.1464	0.1936	**0.2037**
	Res	2306	1924	**1892******	2032	1872	**1856***	1926	1766	**1735***
politics	KT	0.3779	0.4831	**0.4880******	0.0253	0.1470	**0.1748******	0.0824	0.2138	**0.2573******
	Res	7118	5485	**5359******	9691	8102	**7775******	9114	7463	**6865******
philosophy	KT	0.3484	0.3473	**0.3567**	0.0668	0.0768	**0.0898**	0.1475	0.1530	**0.1654**
	Res	6822	6643	**6566**	6239	6069	**6067**	**5622**	5813	5652
math.meta	KT	0.4469	0.5321	**0.5360*****	0.1052	0.1356	**0.1411**	0.1259	0.1538	**0.1775**
	Res	2253	1869	**1866******	3831	**3465**	3499**	3764	3466	**3425****
codegolf.meta	KT	0.4837	0.5633	**0.5640****	0.1664	0.1075	**0.1070***	0.1797	**0.1509**	0.1567
	Res	1950	1675	**1671*****	3843	3453	**3450****	3713	3430	**3396**

KT: Kendall's tau rank correlation coefficient. Higher KT is better. Res: Sum Squared Residual to diagonal line. Less Res is better.
significance level of CVA better than voteDiff: *($p \le 0.05$) **($p \le 0.01$) ***($p \le 0.001$) ****($p \le 0.0001$)

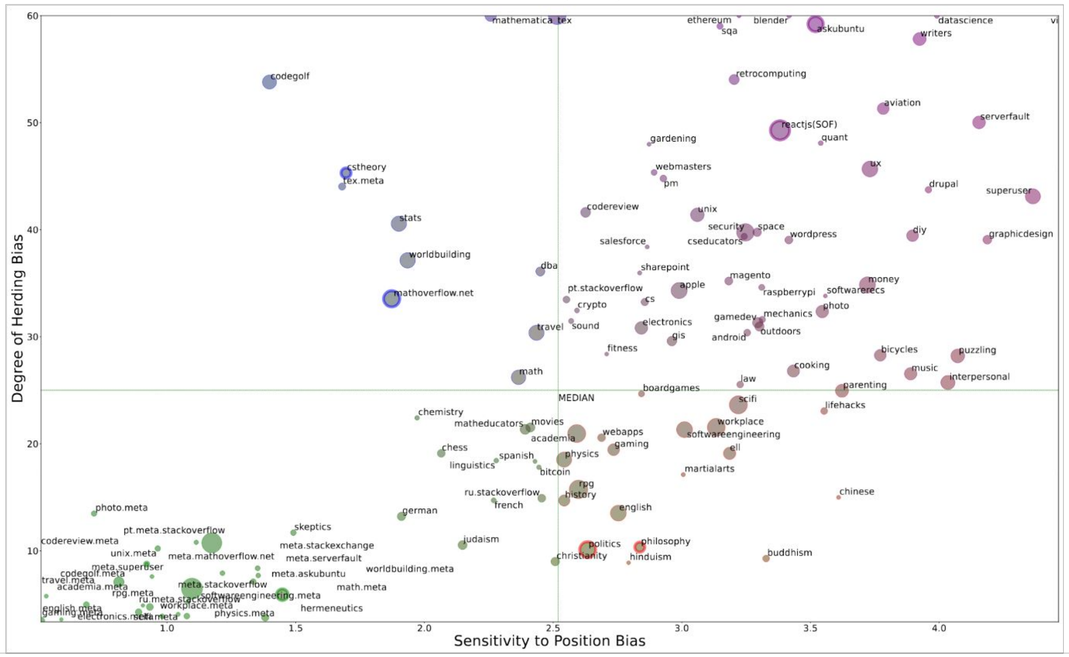
Toy Example Experiment

Example 1a: for position bias							Example 1b: for position bias						Example 2: for herding bias						
Time t	1	2	3	4	5	6	1	2	3	4	5	6	**Time t**	1	2	3	4	5	6
Answer A (rank 1)	+	+	+							−	−	−	**Answer A**	+	+	+	−	−	−
Answer B (rank 2)				+	+	+	−	−	−				**Answer B**	+	−	+	−	+	−



(a) Learned quality over time for Example 1a (b) Learned quality over time for Example 1b (c) Learned quality over time for Example 2

Cross community analysis



Counterfactual Voting Adjustment (CVA) framework:

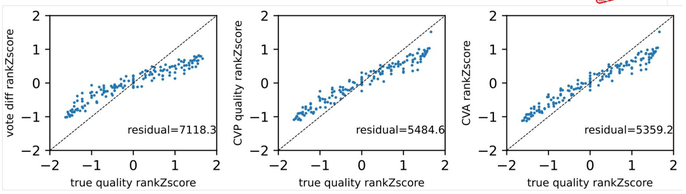
$$\max_{\theta} \log \prod_{t=2}^{T} \text{logit}^{-1}(q_{ij}^{t} + \lambda R_{ij}^{t} + \nu_i L_{ij}^{t} + \beta(\frac{1}{1+D_{ij}^{t}})) - \frac{1}{2}||\boldsymbol{\theta}||_2^2$$

quality | prior vote ratio | relative length | displayed rank

Conclusion:

- Bias mitigation and fairer quality estimation at the same time
- Superior ranking outperforms vote-based and non-causal models
- Managerial insights for both platforms and users
- Broad applicability for other types of platforms
- Future potential to incorporate more features

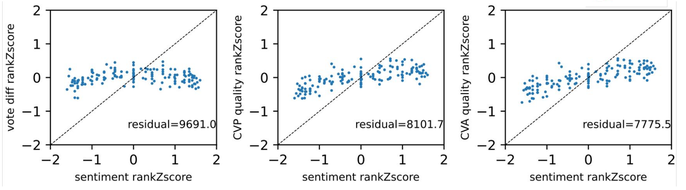
Semi-Synthetic Experiment



KT rank correlation: 0.3779 0.4831 **0.4880**

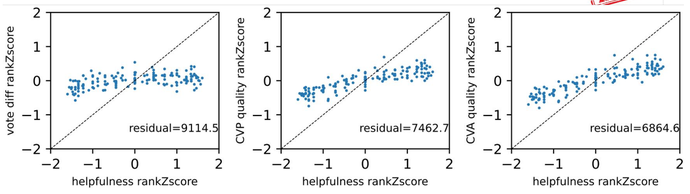
Real data Experiment

Comment Sentiment as ground truth:



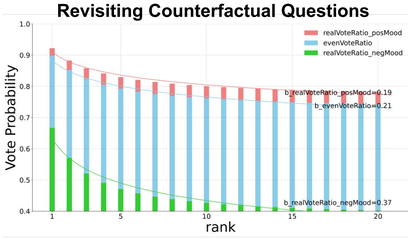
KT rank correlation: 0.0253 0.1470 **0.1748**

GPT-4o evaluated helpfulness score as ground truth:



KT rank correlation: 0.0824 0.2138 **0.2573**

Revisiting Counterfactual Questions



Winrate of the largest 120 communities

		CVA better than vote Diff	CVA better than CVP	CVA better than **both**
sentiment of comments as ground truth	Res	84.17%	90.00%	**78.33%**
	KT	79.17%	86.67%	**74.17%**
helpfulness score of GPT as ground truth	Res	81.67%	94.17%	**77.50%**
	KT	80.83%	91.67%	**75.0%**

https://arxiv.org/abs/2506.21362