Appendix

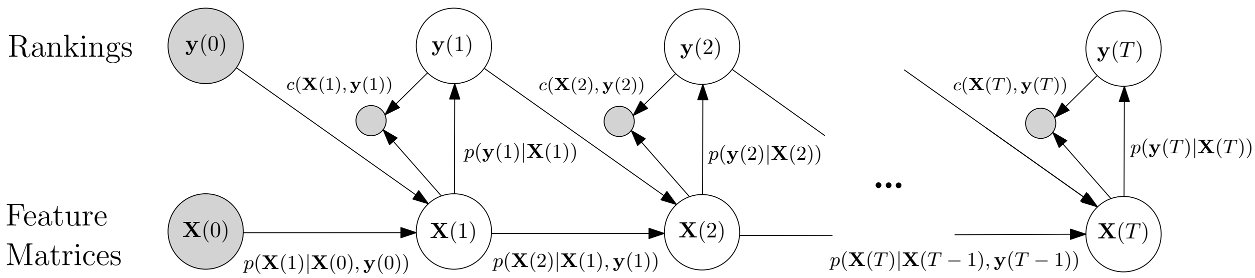
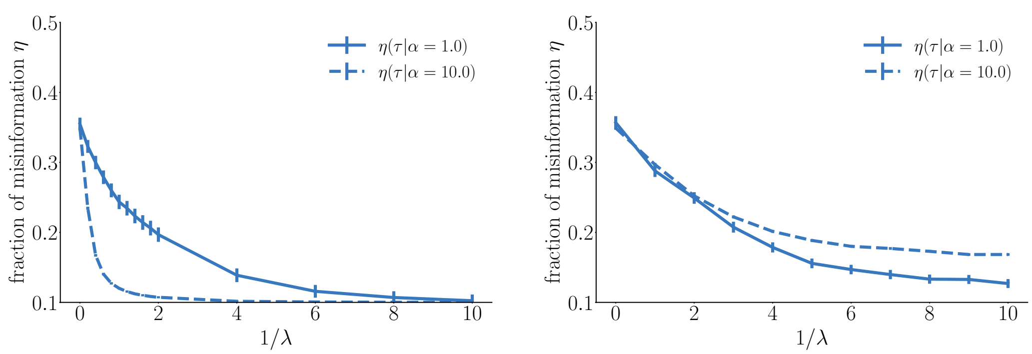


Figure 3: Our joint representation of rankings and user dynamics using Markov decision processes (MDPs). The ranking model $p_\theta(\boldsymbol{y}(t) \,|\, \boldsymbol{X}(t))$ provides a ranking $\boldsymbol{y}(t)$ for a set of items on the basis of the feature matrix $\boldsymbol{X}(t)$ of the items and both the feature matrix and the provided ranking result in a cost to welfare $c(\boldsymbol{X}(t), \boldsymbol{y}(t))$. The distribution of user dynamics $p(\boldsymbol{X}(t+1) \,|\, \boldsymbol{X}(t), \boldsymbol{y}(t))$ determines the feature matrix $\boldsymbol{X}(t+1)$ on the basis of the previous feature matrix $\boldsymbol{X}(t)$ and ranking $\boldsymbol{y}(t)$.

A Representation of rankings and user dynamics

Figure 3 shows the graphical representation associated with the MDP in Eq. 1.

B Additional experiments on synthetic data



(a) Optimal consequential ranking model p^*_θ

(b) Greedy baseline p_{θ_g}

Figure 4: Variation of % of misinformation in the top position for the optimal consequential ranking model p^*_θ, implemented using Algorithm 1, and the greedy baseline p_{θ_g} across both viral and non-viral posts. For the optimal consequential ranking model (panel (a)), as we increase $1/\lambda$, the fraction of misinformation for viral posts on the top 3 positions is significantly lower than the fraction of misinformation for non-viral posts. In contrast, for the greedy baseline (panel (b)), as we increase d, the fraction of misinformation η on the top position does not change significantly with the virality of the posts.

Viral vs. non-viral high risk posts. We investigate whether the optimal consequential ranking model and the greedy baseline treat viral and non-viral posts differently. Intuitively, the ranking model should be more willing to change the rank of high risk viral posts than that of high risk non-viral posts. To confirm this intuition, we compute the fraction of estimated and true misinformation, $\eta(\tau)$ and $\eta^*(\tau)$, in the top position of the rankings over time for both viral ($\alpha = 10$) and non-viral ($\alpha = 0.1$) posts, *i.e.*,

$$\eta(\tau|\alpha) = \frac{\sum_{t=1}^{T} p_{\omega_1}(t)\mathbb{I}(\alpha_{\omega_1(t)} = \alpha)}{\sum_{t=1}^{T} \mathbb{I}(\alpha_{\omega_1(t)} = \alpha)}$$

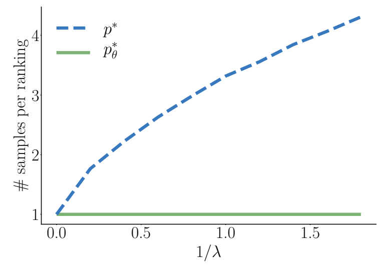


Figure 5: Sample complexity of the optimal consequential ranking model p^*_θ and the P-L consequential ranking model p_θ. The optimal consequential ranking model, implemented using Algorithm 1, becomes computationally prohibitive in terms of # samples needed per ranking as $1/\lambda$ increases and the difference between the original ranking model. This is in contrast with the P-L consequential ranking model, learned using Algorithm 2, which exhibits much better sampling complexity with respect to $1/\lambda$.

Figure 4 summarizes the results. For the optimal consequential ranking model, as we increase $1/\lambda$, the fraction of misinformation η for the viral posts on the top position is significantly lower than the fraction of misinformation for non-viral posts. In contrast, for the greedy baseline, as we increase d, the fraction of misinformation η on the top position does not change significantly with the virality of the posts.

Sampling complexity. We compare the sampling complexity of the optimal consequential ranking model, implemented using Algorithm 1, and the P-L consequential ranking model, learned using Algorithm 2. Figure 5 summarizes the results, which shows that, as $1/\lambda$ grows, it becomes computationally prohibitive to generate optimal consequential rankings using Algorithm 1 due to the growing difference between p^* and p_0. In contrast, the complexity of learning P-L consequential ranking model, using Algorithm 2, stays constant as $1/\lambda$ changes.

C Reddit data preprocessing

In the first set of experiments using Reddit data, we focus on the civility of the comments in each submission. To this aim, we apply sentiment analysis on the text of the comments using the software package *Pattern*[10] and, for each comment, obtain two quantities: mood and polarity. The mood of a comment can take one of the following four values: indicative, imperative, conditional and subjunctive. The polarity of a comment is a real number in $[-1, 1]$, where lower (higher) values indicate more negative (positive) words in the text. Then, we define the incivility score ϕ of a comment as the absolute value of the polarity of the comment if the polarity is negative and the mood of the comment is indicative or imperative and zero otherwise. Finally we apply a uniformly distributed quantile transformer to map the incivility scores to a value in $[0, 1]$ with largest values always mapped to 1. Table 1 provides a few examples of sentences with a high value of ϕ.

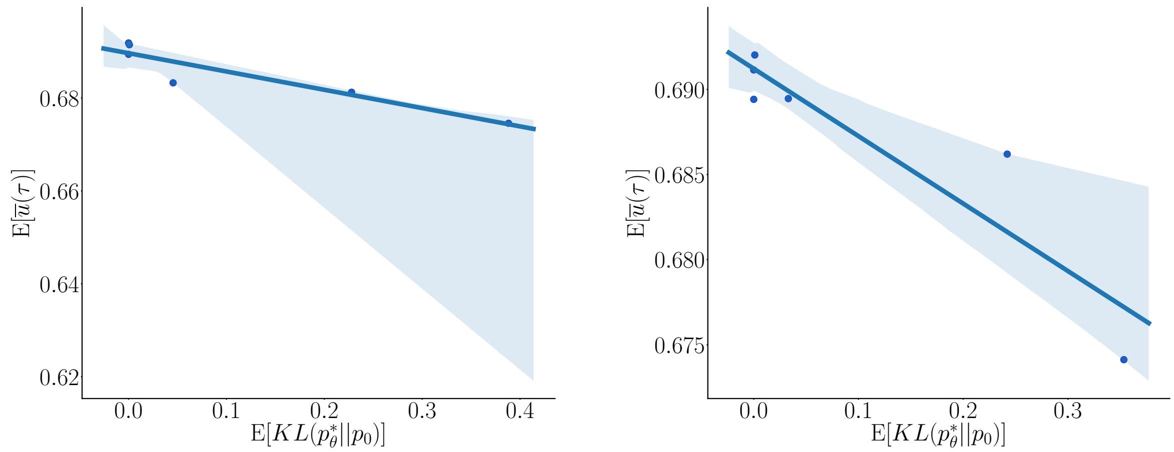
Comment	Incivility (ϕ)
If you once tell a lie, the truth is ever after your enemy.	0.0
I dream of a world where your bigoted stupid ideas don't have the popular shield of faith.	0.1
Shut the f**k up and die already you POS warmongering profiteer.	0.4
Crap? Or pap. Take your pick.	0.8
i blame the evil KOCH BROTHERS!	1.0

Table 1: Examples of sentences with different levels of incivility, as estimated by the feature ϕ. Comments with higher levels of incivility typically correspond to those that use foul language.

[10] https://www.clips.uantwerpen.be/pages/pattern-en

Url	Misinformation (γ)
aids.gov	0.0
pbs.org	0.26
breitbart.com	0.56
lifeisajoke.com	1.0

Table 2: Examples of domains that spread different amounts of misinformation, as estimated by the feature γ.



(a) Misinformation

(b) Incivility

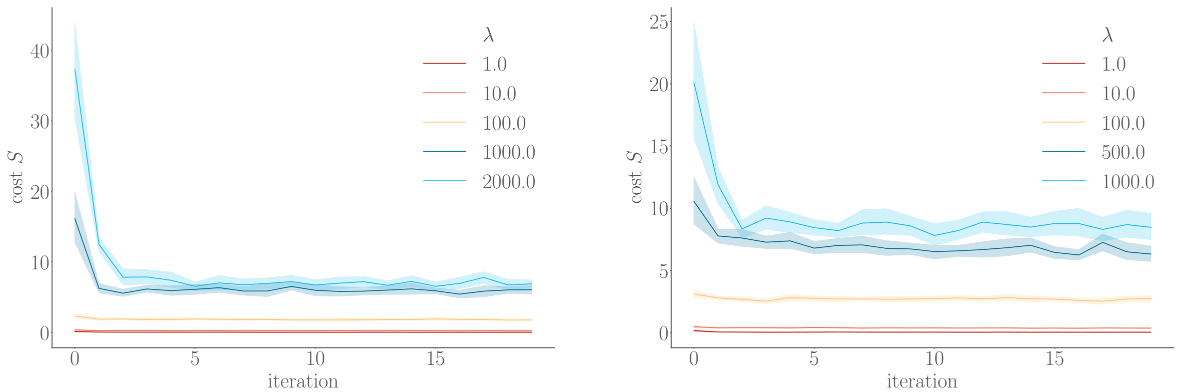
Figure 6: Average immediate utility $\mathbb{E}[\overline{u}(\tau)]$ achieved by the consequential ranking model vs the average Kullback-Leibler (KL) divergence $\mathbb{E}[KL(p_\theta^* \,||\, p_0)]$ between its induced probability $p_\theta^*(\tau)$ and the probability $p_0(\tau)$ induced by the original ranking model. It shows that there is a negative correlation between the immediate utility and the KL divergence for both the experiments.

In the second set of experiments, we focus on the misinformation spread by the comments of each submission. To this aim, we estimate the unreliability score γ for each comment by estimating the average unreliability score of the domains that appeared in each of them, as estimated by aggregating publicly available data from Politifact and Snopes[11]. More specifically, our combined dataset contains fact checking information for $17{,}804$ unique urls from $4{,}540$ unique domains. For each url, it assigns a label that indicates the reliability of its content. We used these labels to assign a numerical unreliability score for each url. More specifically, if the url is labeled as "false", "pants-fire", "mfalse" or "legend", we set the unreliability score to 1. If the url is labeled as "true", "mtrue" or "mostly-true", then we set the unreliability score to -1. And, if the url is labeled using some other label value, we set the unreliability score to 0. We computed an unreliability score for each domain, which measures its level of (un)trustworthiness, by taking the average of the unreliability scores of individual urls from the domain. Then, we define the unreliability score γ of a comment as the average unreliability score of the domain(s) of the link(s) used in the comment if the average is negative and zero otherwise. Here, also note that, if a comment does not contain any links or the domain(s) of the link(s) does not appear in our dataset, we set the unreliability score for that comment to 0. Finally we apply a uniformly distributed quantile transformer to map the unreliability scores to a value in $[0, 1]$ with largest values always mapped to 1. Table 2 provides a few domains with a different values of γ.

D Additional experiment on real data

So far, we have assumed that there is a negative correlation between the immediate utility achieved by the consequential ranking model and the Kullback-Leibler (KL) divergence between its induced probability $p_\theta^*(\tau)$ and the probability $p_0(\tau)$ induced by the original ranking model. Here, we verify this assumption by looking into the variation of $\mathbb{E}[\bar{u}(\tau)]$ with $\mathbb{E}[KL(p_\theta^* \,||\, p_0)]$. Figure 6 demonstrates the results, which show that there is indeed a negative correlation between

[11] https://www.kaggle.com/arminehn/rumor-citation/version/3



(a) Uncivility (b) Misinformation

Figure 7: Average cost S per iteration during training. The results show that, as λ increases, the model takes longer to converge.

the immediate utility and the KL divergence for both the experiments.

Finally, we also compute the average cost S per iteration during training. Figure 7 summarizes the results, which show that, as λ increases, the model takes longer to converge.