Inferring True Voting Outcomes in Homophilic Social Networks

JAAMAS Track

John A. Doucette, Alan Tsang, Hadi Hosseini, Kate Larson and Robin Cohen
jdoucette@ncf.edu, akhtsang@gmail.com, hhvcs@rit.edu, klarson@uwaterloo.ca, rcohen@cs.uwaterloo.ca

ACM Reference Format:
John A. Doucette, Alan Tsang, Hadi Hosseini, Kate Larson and Robin Cohen jdoucette@ncf.edu, akhtsang@gmail.com, hhvcs@rit.edu, klarson@uwaterloo.ca, rcohen@cs.uwaterloo.ca. 2020. Inferring True Voting Outcomes in Homophilic Social Networks. In *Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), Auckland, New Zealand, May 9–13, 2020,* IFAAMAS, 3 pages.

1 INTRODUCTION

Consider a soccer game with spectators in a stadium acting as voters. The spectators are polled to determine whether a ball has crossed a line. Either the ball has crossed, or it has not, but the opinions of individual voters regarding the truth of the matter may differ because of their differing perspectives on the event. Voters positioned far from the event may be unable to accurately assess the outcome compared to those positioned nearby. In practice, however, voters' opinions may not be independently distributed. For example, the voters may talk among themselves before their opinions are gathered. This can distort the distribution of opinions, and introduce correlations into voter reports, preventing recovery of the true outcome. For example, if an announcer states that the ball *did* cross the line, then voters who did not observe this may report this authoritative opinion rather than their own.

Conitzer [2] first questioned whether or not social network structure should inform voting outcomes. In his model, he showed the Maximum Likelihood Estimator (MLE) is exactly equal to majority voting, and does not depend on network structure at all. This in turn led a more complex model [3] where opinions were distributed among *edges* (conversations) of the network. Although this second model produced a tractable estimator, we argue it is not a natural representation opinion spread.

We present a novel approach to objective social choice where agents interact within a social network and alter their opinions based on their peers. We assume that agents are more easily convinced of the truth rather than a falsehood, and produce a computationally efficient, intuitive, and effective method for inferring the true outcome *even after opinions have been altered throughout the social network*. Our main contributions are the following:

- A new model for agent conversations (opinion dynamics) in a social network, for which inference[1] is tractable — the *Correct Conversation model* (CC).
- A theoretical characterization of the model and its properties on simulated and real world datasets to showcase its predictive power over majority voting, under certain conditions.

[1]We use *inference* in the statistical sense: inferring the value of a variable or parameter on the basis of a given set of data. For us, the variable of interest is usually the winner of an election, while the data are usually the votes.

Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), B. An, N. Yorke-Smith, A. El Fallah Seghrouchni, G. Sukthankar (eds.), May 9–13, 2020, Auckland, New Zealand.

2 MODEL

A social network $G = \{V, E\}$ contains n voters V, and m edges where edge $E_{i,j}$ indicates that the opinions of voters V_i and V_j can interact. Let $V_i \in C = \{\psi, \omega\}$ represent voter V_i's vote. We treat the true winner (i.e. ground truth outcome) as a random variable W drawn from C. Voter V_i has *initial* opinion $X_i \in C$, which may be changed by social interactions prior to the reporting stage. Let $N_i = \{V_j \in V | E_{i,j}\}$ be the neighbors of V_i. The process we model is a vote held over the set of alternatives $C = \{\psi, \omega\}$ with true winner $W \in C$, with voters V, connected by edges E in the social network G.

In the style of Conitzer [2], we define the likelihood function $\mathcal{L}(W = \lambda|V)$ for alternative $\lambda \in C$ being the winner, given observed votes V, as follows:

$$\mathcal{L}(W = \lambda|V) \propto \prod_i g(V_i|W = \lambda) h(V_i, V_{N_i}|W = \lambda) \tag{1}$$

Additionally, we assume the tendency for voters to agree with their neighbors can be factored into a product of *individual* such tendencies with each neighboring voter:

$$h(V_i, V_{N_i}|W = \lambda) \propto \prod_{V_j \in V_{N_i}} h'(V_i, V_j|W = \lambda) \tag{2}$$

Finally, we add constraints for g and h' so to bias them toward the correct outcome — both of voters' innate opinions, and tendency toward agreement with neighbors. The latter is the core of our "Correct Conversation (CC)" assumption (voters who talk to one another are more likely to agree on the truth than on a falsehood):

$$g(V_i = \lambda|W = \lambda) > g(V_i \neq \lambda|W = \lambda) \tag{3}$$

$$h'(V_i = \lambda, V_j = \lambda|W = \lambda) > h'(V_i \neq \lambda, V_j \neq \lambda|W = \lambda) \tag{4}$$

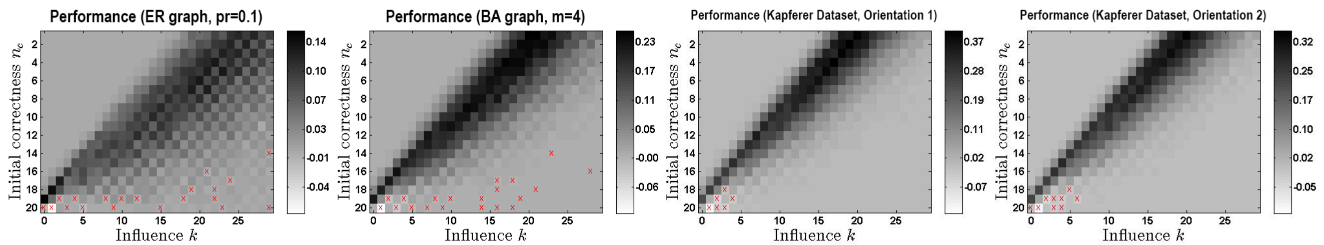
We assume functions g and h' are identical for every voter.

2.1 Finding the Most Likely Alternative

Finding the most likely winner of the election reduces to finding $\arg\max_{\lambda \in C} \mathcal{L}(W = \lambda|V)$. We adopt the notation $g(V_i = \lambda|W = \lambda) \propto p$, $g(V_i \neq \lambda|W = \lambda) \propto 1 - p = \dot{p}$, $h'(V_i = \lambda, V_j = \lambda|W = \lambda) \propto q$, $h'(V_i \neq \lambda, V_j \neq \lambda|W = \lambda) \propto \dot{q}$, and $h'(V_i = \lambda, V_j \neq \lambda|W = \lambda) \propto r$, where $q + \dot{q} + r = 1$. Then our likelihood function can be expressed as follows:

$$\mathcal{L}(W = \lambda|V) \propto p^x \dot{p}^{n-x} q^{2y} r^{2z} \dot{q}^{2(m-y-z)} \tag{5}$$

where n is the number of voters, m is the number of edges between voters in the social network, x is number of votes for λ, y



(a) (b) (c) (d)

Figure 1: *Performance Improvement: (a) ER graphs; (b) BA graphs; networks of (c) and (d) based on Kapferer Tailor Shop data.*

is the number of *concordant edges* (i.e. edges between voters both voting for λ), and z is the number of *discordant edges* (i.e. edges between voters who disagree).

Then, there is an *unambiguous winner* when $x > \frac{n}{2}$ **and** $y > \frac{m-z}{2}$; i.e. **an alternative that captures both a majority of the votes *and* a majority of the concordant edges, is the most likely winner.** When these signals conflict, however, the situation is more complicated and an MCMC approach can be used to maximize the probability of selecting a winner.

2.2 The Homophily Assumption

Unfortunately, the conditions outlined above are not sufficient for the CC model to outperform simple majority. We need to make one additional assumption about how votes are distributed in the network — that they exhibit *homophily*. It is not sufficient that neighbors agree more often than they disagree, but that agreeing nodes are disproportionately clustered together. There are many generative processes which produce such highly homophilic vote distributions on a social network; we examine one now.

Consider a process where we assign each voter an initial opinion, $X_{0,i}$ which is correct with probability p. Then, at each step t, we sample a voter, Swap, with probability proportionate to

$$P(\text{Swap} = i) \propto 1 + \sum_{j \in N_i} I(X_{t,j} = W \wedge X_{t,i} \neq W) \tag{6}$$

where I is the indicator function, and flip her opinion $X_{t,\text{Swap}}$ to the opposite alternative to obtain a new profile X_{t+1}. We repeat this process to perform k swaps.

This process produces homophilic graphs and parallels innovation diffusion mechanisms: "Wrong" voters with many "correct" neighbors are sampled preferentially, and are therefore more likely to be influenced. While flipping the opinion of such a node will increase the discriminatory power of the majority rule by $\frac{1}{|V|}$, it also increases the discriminatory power of counting the concordant edges by far more than $\frac{1}{|E|}$.

3 EMPIRICAL EVALUATION

We evaluate the performance of the CC model on simulated elections on Erdös-Rényi (RE) random graphs [5], and networks based on the Kapferer Tailor Shop dataset [7]. Voters in the network have opinions initialized randomly, with n_c of the n voters starting with the correct opinion, and the remainder receiving the incorrect opinion. Subsequently, voters interact k times based on the dynamics specified in Equation 6. Finally, we aggregate opinions from the networks via two methods: (1) Simple majority, and (2) the CC model described in Section 2.1, using a Metropolis-Hastings algorithm [6, 8] to break ties.

3.1 Results

We measure each model's ability to predict the correct winner and show the *difference* between these values as the *Performance Improvement* as heatmaps in Figure 1. Darker regions correspond to data points where our model performs better than majority voting. The data points where the Correct Conversation model performs *worse* than majority voting are marked with a red X.

In all scenarios, the lighter region in the top left represents cases where both models are incorrect, because most voters begin with the incorrect opinion and have little opportunity to change. Therefore, the performance of both models is exactly zero. In analogous regions in the lower right, most voters have the correct opinion, so both models almost always give the correct answer. With more extreme parameters, majority voting may perform better, but this effect is very small (typically around 1-2%). In between however, there is a critical band where the Correct Conversation model enjoys a considerable advantage: around 15% in ER graphs, and up to 37% in the Kapferer dataset. Interestingly, in the bottom left, just beside the critical band are several scenarios that elicits the worst performance from our model. In these scenarios, almost half of the voters initially begin with the correct opinion, and just enough opinions are revised to produce a small majority. This allows naive voting to produce the correct result, but does not produce sufficient structure in the edges for our model to exploit.

We also explored directed networks based on Barabási-Albert (BA) scale-free random graphs [1]. We find that it performs in a similar manner (with up to 86% performance improvement) when high-degree hub nodes act as information aggregators. Because of the dynamics of Equation 6, these high-degree nodes that are in err are likely to be flipped to the correct opinion. However, it performs poorly in strongly hierarchical networks where high-degree nodes only exert influence on their neighbors. Here, high-degree nodes can easily become misinformed and perpetuate that misinformation to other nodes. We refer the reader to the full paper for additional details and proofs [4].

Future investigation will apply the CC model to more complex graph models, and extend it both to ranked voting rules for multiple alternatives and to a wider range of opinion dynamics.

REFERENCES

[1] Reka Albert and Albert-Laszlo Barabasi. 2002. Statistical mechanics of complex networks. *Reviews of Modern Physics* 74 (2002), 47–97.

[2] Vincent Conitzer. 2012. Should social network structure be taken into account in elections? *Mathematical Social Sciences* 64, 1 (2012), 100–102.

[3] Vincent Conitzer. 2013. The Maximum Likelihood Approach to Voting on Social Networks. In *Proceedings of the 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton-13)*. Allerton Retreat Center, Monticello, IL, USA, 1482–1487.

[4] John A Doucette, Alan Tsang, Hadi Hosseini, Kate Larson, and Robin Cohen. 2019. Inferring true voting outcomes in homophilic social networks. *Autonomous Agents and Multi-Agent Systems* 33, 3 (2019), 298–329.

[5] Paul Erdős and Alfréd Rényi. 1959. On random graphs. *Publicationes Mathematicae Debrecen* 6 (1959), 290–297.

[6] W Keith Hastings. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 1 (1970), 97–109.

[7] Bruce Kapferer. 1972. *Strategy and transaction in an African factory: African workers and Indian management in a Zambian town.* Manchester University Press.

[8] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. 1953. Equation of state calculations by fast computing machines. *The journal of chemical physics* 21, 6 (1953), 1087–1092.