The Social Network Effect on Surprise in Elections

Palash Dey
Indian Institute of Technology Kharagpur
Kharagpur, India
palash.dey@cse.iitkgp.ac.in

Pravesh K. Kothari
Princeton University
Princeton, New Jersey
kothari@cs.princeton.edu

Swaprava Nath
Indian Institute of Technology Kanpur
Kanpur, India
swaprava@cse.iitk.ac.in

Abstract

Elections involving a large voter population often lead to outcomes that surprise many. A better prediction of the true outcome helps reduce the adverse effect of surprise on the economy of a sizable population. This paper starts from the basic observation that individuals in the underlying population build *estimates* of the distribution of preferences of the whole population based on their immediate neighbors in the underlying social network. The outcome of the election leads to a surprise if these local estimates contradict the outcome of the election for some fixed voting rule. To get a quantitative understanding, we propose a novel mathematical model of the setting where the individuals in the population and their connections are described by a random graph with connection probabilities that are biased based on the preferences of the individuals. Each individual also has some estimate of the bias in their connections. The connection model is inspired by the *homophily* effect in social networks.

We show that the election outcome leads to a surprise if the discrepancy between the estimated bias and the true bias in the local connections exceeds a certain threshold, and confirm the phenomenon that surprising outcomes are associated only with *closely contested elections*. We consider large elections with networked voters and compare standard voting rules based on their performance on surprise. Our results show that the rules have different behavior for different parts of the population. It also hints at an impossibility result that any reasonable voting rule will be less surprising for *all* parts of a population. To attest some of our theoretical predictions, we experiment with the large dataset of UK-EU referendum (a.k.a. Brexit).

CCS Concepts

• **Networks → Online social networks**; • **Applied computing → Voting / election technologies**;

Keywords

election, surprise, social networks, voting, prediction

ACM Reference Format:
Palash Dey, Pravesh K. Kothari, and Swaprava Nath. 2019. The Social Network Effect on Surprise in Elections. In *6th ACM IKDD CoDS and 24th COMAD (CoDS-COMAD '19), January 3–5, 2019, Kolkata, India.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3297001.3297002

CoDS-COMAD '19, January 3–5, 2019, Kolkata, India

ACM ISBN 978-1-4503-6207-8/19/01…$15.00
https://doi.org/10.1145/3297001.3297002

1 Introduction

Recent times have witnessed quite a few elections whose outcomes are widely considered as surprises.[1] News reports covered the unprecedented impact on trade, national economies, and job markets because of the results of the elections (e.g., Brexit [39], US presidential elections [35], UK parliamentary election [40, 41] etc.). It was attributed to the fact that many people and the market were unprepared for such an outcome. It has impacted not only the economy and made the stock markets unpredictable, the social impact was also paramount. It was clear that the social connections – either online or offline – and the mass communication media – print or electronic – that are important factors in opinion building, have a localized effect which does not give a holistic idea of the outcome of an election. This effect is more prominent in the online social networks, since communities in social networks inevitably group similar people together and it is easy to ignore biases. Having a large number of friends in an online social network may solidify the belief that the local observation is quite a representative sample than what actually is true. This raises a natural question:

> *"Can the surprise/shock in an election be explained by the social network structure or the biases in the perception of the voters?"*

In this paper, we address this question by proposing a model of the social network formation and voters' perception of the winner. We show that the answer cannot be obtained from an analysis that focuses on only the network structure or only the voter perception. For instance, if we consider only network structure, the following example shows that any perception about the connection probability will always leave at least half the population surprised.

Example 1 (Limitation of a structure-based conclusion). *Suppose in a population of n (even) voters with two candidates (red and blue), $n/2$ are red (meaning they prefer red over blue) and the rest $n/2$ are blue. The voting rule is plurality.*[2] *Suppose the network structure is such that each voter is connected with every other voter that has the same color as hers, but is connected to exactly $n/2 - 1$ voters of the other color. If she perceives the winner just by counting the majority at her own neighborhood, then every voter will 'think' that her favorite candidate wins, and no matter how tie is broken to*

[1] We define *surprise* from the perspective of a voter, and it is the event where the candidate most likely to win according to the voter's estimate did not win. All our conclusions are based on this binary definition of surprise. We note that there can be other notions of surprise that is cardinal and considers the difference of the estimated vote share from the actual, which may lead to a different conclusion.

[2] In plurality voting rule. every voter votes for her favorite candidate and the candidate with most number of votes win.

select the winner, half the population will always be surprised at the outcome.

Clearly, the example can be adapted if the voters discount the number of voters of their own color (given the fact that they are more likely to be connected with a similar colored voter) to yield the same conclusion. Moreover, if there are more than two candidates, an extension of the construction above will lead to a surprise of the voters in the classes where the actual winner (in plurality voting over all voters) is not their favorite candidate.

So, it is clear that a worst case analysis over the social network structure will always lead to surprise in election – which is hardly the case in practice – elections with unsurprising outcomes are in fact quite common. Later in the paper, we discuss how error in voter perception *alone* also cannot give rise to surprise. Our approach takes into account both these factors simultaneously and provides conditions when a typical voter is surprised. In fact, there are some counterarguments claiming that some of these elections cannot be called 'surprising' given a correct model of voter perception (e.g., [27] for Brexit).

We adopt a Bayesian approach that assumes a random generative model of the voters and the social network, and show that an error in estimating the parameters of the generative process may lead to surprises.

1.1 Our Approach and Results

Let us define the voter generation and social network formation process a bit more formally. Consider a set of m candidates and n voters. A class of a voter is identified by a specific linear order over the candidates representing their preference – hence there are $m!$ classes. Each voter is picked *i.i.d.* from a fixed probability distribution of belonging to a class. Once the voters are generated, social network among the voters are formed according to a stochastic block model, which is inspired by the homophily (like minded individuals are more likely to form social links) observed in social networks. This is a general version of an Erdös-Renyi random graph model, where the vertices are partitioned into classes and the edge creation probabilities (which can be different) are defined only among the classes – hence every node of a class connects to every other node in another class with the same probability. In our model, an intra-class connection probability p_{ii} is assumed to be larger than an inter-class connection probability p_{ij} (where i and j are indices for classes). For a specific voting rule r, that aggregates the individual preferences into the choice of a candidate, e.g., plurality, which selects the candidate that maximum number of voters place on the top of their preferences, and a realization of the voters' preferences denoted by the set V, there is a winner which we represent using $w_T(V, r)$. Since every voting rule we consider are anonymous, i.e., winner does not change even if the voter identities are changed, the winner is determined just by the number of voters in each class. Therefore, V in $w_T(V, r)$ can be replaced by $\vec{N} = (N_1, N_2, \dots, N_{m!})$, where N_j is the number of voters in class j. The perceived winner of voter v is dependent on her estimates of the number of voters in different classes, denoted by $\hat{N}^v := (\hat{N}_1^v, \hat{N}_2^v, \dots, \hat{N}_{m!}^v)$, and is given by $w_P(\hat{N}^v, r)$. Voter v is surprised when $w_P(\hat{N}^v, r) \neq w_T(\vec{N}, r)$. We call *surprise* to be the probability of this event. Voter v estimates $\hat{N}_j^v$ by taking the ratio of her observed neighbors of class j with her estimated connection probability with class j. This estimation neutralizes her observation bias had the estimates been perfect.

With this setup, our first result (Theorem 2) shows that for $m = 2$, if a ratio of the estimated connection probabilities stay within a threshold, a voter is *not* surprised with high probability (i.e., surprise asymptotically approaching zero as $n \to \infty$). However, if the threshold is crossed, the voter is surprised *w.h.p.* A corollary of this result is that if the original distribution of the voters was very biased towards one class (*'overwhelming majority for one candidate'*), then, even with erroneous connection probability estimates, a voter will never be surprised *w.h.p.* This result shows that voters' perception error is not solely responsible for surprise. Together with Example 1, we conclude that social connection and voter perception are intertwined reasons for surprise in elections.

The theorem also shows that *surprise is a phenomenon of a closely contested election.*[3] Later, we generalize our results for more than two candidates. As a first approach, we present the case with three candidates in §3.1. However, the method clearly generalizes with similar assumptions to similar conclusions with more candidates. Unlike the case with two candidates, for three candidates, one can consider different voting rules and compare their performances w.r.t. surprise. We consider three prominent voting rules (that are scoring rules). Our next result (Theorem 4) shows that for different classes of voters, different rules perform better in terms of surprise – and hints that there may not be a single surprise-optimal voting rule for all classes of voters. However, we find it interesting that the performance is not proportional to the distribution of the mass in the scoring rules[4] since in certain class of the voters, both plurality and veto perform better than Borda voting. All voting rules are explained when presented.

Though the theoretical results in §3 use the estimates of the connection probabilities and show that the correctness of those estimates w.r.t. the true values may surprise a voter, we do not explicitly mention how the voters arrive at these estimates. In §4, we consider a real dataset (UK-EU referendum, a.k.a. Brexit) and consider a realistic model of network formation and voters' winner anticipation, that is a realistic instantiation of our theoretical model. We investigate the effect of intra and inter-class connection probabilities, and the effect of noisy observation of their estimates on surprise. We find that the conclusions in those results show a resemblance with some of the theoretical predictions.

1.2 Related Work

Online social media is omnipresent in our digital lives, and they have significant influence on public events like national or state elections [7]. The effect has already been observed in the social networks literature [38], and questions have been raised on whether the network structure should be considered in elections [8]. Since people express their opinion on social media, analysis of social networks provides prediction on outcomes [46]. A stream of research investigates how artificial intelligence can change the opinion of the population [4, 33, e.g.].

[3] This is a consequence of our definition of surprise, which is a binary notion as defined before.

[4] A scoring rule is a voting rule where every position in a preference is given a score and the candidate with the maximum aggregate score wins.

On the other hand, public elections are one of the cornerstones of research in *social choice theory*, which deals with decision making with multiple intelligent agents. In the computational social choice and multi-agent systems literature, there had been several notions to measure the 'goodness' of elections. For example, *margin of victory*, defined as the smallest number of voters who can alter the outcome of an election by voting differently [5, 13, 24, 25, 49], provides a quantitative threshold of surprising outcomes in terms of the voter population. A related literature exists for *bribery in election* [6, 20, 28, 31, 37, e.g.], complexity of manipulative attacks [1, 9, 12, 17–19, 21–23, 32, 43, e.g.], and query complexity [11, 15, 16, e.g.].

Surprise in election, to the best of our knowledge, has not been formally studied in either of the above two strands of literature. There is a relevant body of literature on surprise in political economy. Ely et al. [29] formally define *suspense* and *surprise* in a dynamical model and provide a design approach to maximize either of them for a Bayesian audience. Our definition of surprise (the outcome is contrary to a voter's belief) is closely related in spirit, and is adapted to a single-shot decision. Similarly, in sports tournaments, it is important to design the schedule so that the games are highly competitive and results are unpredictable [10, 42]. In fact, *information design*, where a social planner aims to maximize the unpredictability of a contest has been investigated in various contexts (see, e.g., a recent survey by Bergemann and Morris [2]). But in election outcomes stability is of prime importance [26, 44, 45]. The social connection model in our paper is inspired by stochastic block model. This model has a long tradition of study in the social sciences and computer science [34, 36, 48]. Therefore, in this paper, we approach the question of surprise in election using well studied models of social connection and surprise, and introduce a voter perception model to present insightful results.

2 Model

Let $[k] \triangleq \{1, \ldots, k\}$. Let $N = [n]$ be the set of *voters*, and $M = \{a_1, \ldots, a_m\}$ be the set of *candidates*. Every voter has an ordinal preference over the candidates, and we assume that these preference relations are total orders, i.e., transitive, anti-symmetric, and complete. We assume $m << n$, which is representative of real elections. Since the number of preference orders can be at most $m!$, we partition the voters into disjoint *classes* identified by $P_k, k \in C$, with $C = [m!]$ being the indices of the classes. Voters in a given class share the same preference order. Let $\vec{N} := (|P_k|, k \in C)$ denote the vector of the number of voters in each class. With a slight abuse of notation, we will refer to the preference of the voters in P_k also with the same notation.

Every voter is associated with class P_j with probability ϵ_j independently from other voters, where $\epsilon_j \in [0, 1]$, $\forall j \in C$, and $\sum_{j \in C} \epsilon_j = 1$. We assume that the ϵ_j's are unknown to the voters. The association is represented by the mapping $\sigma : N \to C$, which maps the voter identities to the class indices. A random social network is formed with these voters by a stochastic block model which is represented by a $|C| \times |C|$ symmetric matrix $P = [p_{jk}]$, where p_{jk} denotes the connection probability between the classes of voters P_j and P_k. In this connection model, the probability of connections for every voter in a class with every voter in another class is identified by a single parameter, which may change for a different pair of classes. The resulting graph is denoted by $G = (N, E)$, where E is the edge set. The edge creation process is independent among each other and also is independent with the voter-to-class association process. We assume a regularity among the connection probabilities for which we need to define a distance metric.[5] The Kendall-Tau (KT) distance between two preference orderings P_j and P_k is the minimum number of adjacent flip of candidates needed to reach one from the other. Clearly, this is a valid distance metric. We call the p_{jk}'s *regular* if they are monotone decreasing with increasing KT distance between P_j and P_k – which means that the voters with more dissimilar preferences are less likely to be connected. We assume that a voter knows the preferences of her immediate neighbors (on the social network) perfectly, but does know the preferences of the other voters.

A voter $v \in P_j$ estimates these connection probabilities which are denoted by $\hat{p}_{jk}$ for all $k \in C$. We assume that the voters' estimated $\hat{p}_{jk}$'s are also regular. At this point, we do not assume a model on how the voters reach their estimates. In §4, we consider a specific model of estimates for the experiments where voters take weighted average of their own observations and a noisy version of the true global distribution. The next section deals with how the errors in these estimates can affect a voters perception of the winner. We will consider only deterministic voting rules.

Voters' winner perception model: Voter v estimates the number of voters in class P_k by dividing the number of her own neighbors in that class on G, defined as $\text{Nbr}_v^k := \{t : (v, t) \in E, t \in P_k\}$, with her estimated $\hat{p}_{\sigma(v)k}$. Hence voter v's estimated number of voters in class P_k is,

$$\hat{N}_v^k = \begin{cases} \frac{1}{\hat{p}_{\sigma(v)k}} |\text{Nbr}_v^k| & \text{if } k \neq \sigma(v), \\ \frac{1}{\hat{p}_{\sigma(v)\sigma(v)}} |\text{Nbr}_v^{\sigma(v)}| + 1 & \text{otherwise}, \end{cases} \tag{1}$$

Note that if the $\hat{p}_{\sigma(v)k}$'s were accurate, by strong law of large numbers, this estimate gives the right number of voters in each class asymptotically *almost surely*.

The voters now have randomly realized preferences and connections with each other. Also, every voter v has an estimate of the number of voters in different classes, and therefore, under a given (anonymous) voting rule r, her perceived winner is denoted by $w_P(\hat{N}^v, r)$, where $\hat{N}^v := (\hat{N}_1^v, \hat{N}_2^v, \ldots, \hat{N}_{|C|}^v)$. The true winner for the same realization is denoted by $w_T(\vec{N}, r)$. A voter is *surprised* when her perceived winner is different from the true winner, defined formally as follows.

Definition 1 (Event of Surprise). *An* event of surprise *of a voter v for a specific realization of the voter preferences and social graph is the event where the voter's perceived winner is not the true winner, i.e., the event S_v such that,*

$$S_v^r := \{w_P(\hat{N}^v, r) \neq w_T(\vec{N}, r)\}. \tag{2}$$

We will call the probability of this event as *surprise* of voter v under voting rule r, denote by $\mathsf{surp}_v^r := P(S_v^r)$.

[5] A valid distance metric is one that is (1) non-negative, (2) symmetric, and (3) obeys triangle inequality.

Note that, the event of surprise is specific to a voter, but every voter in a given class has same surprise in this model, while voters in different classes may have different surprises for the same parameters.

Metric to compare voting rules: Let the event of some candidate b ($\neq w_T(\vec{N}, r)$) beating the true winner $w_T(\vec{N}, r)$ be defined as $\text{Beat}_v^r(b, w_T(\vec{N}, r)) := \{b \text{ beats } w_T(\vec{N}, r) \text{ in } r\}$. The event of surprise, therefore, can be written as $S_v^r = \cup_{b \neq w_T(\vec{N}, r)} \text{Beat}_v^r(b, w_T(\vec{N}, r))$. For the chosen parameters, define the *most probable false beating candidate* as $b_v^{r*} \in \text{argmax}_{b \neq w_T(\vec{N}, r)} P(\text{Beat}_v^r(b, w_T(\vec{N}, r)))$, with ties broken arbitrarily. Using the union bound and the fact that the probability of an union of events is always larger than that of the largest probability of the individual events, we get,

$$\begin{aligned} P(S_v^r) = \texttt{surp}_v^r \in [\ell_v^r, (m-1)\ell_v^r], \\ \text{where } \ell_v^r = P(\text{Beat}_v^r(b_v^{r*}, w_T(\vec{N}, r))). \end{aligned} \tag{3}$$

It is enough to analyze the event $\text{Beat}_v^r(b_v^{r*}, w_T(\vec{N}, r))$ and consider the quantity $\texttt{MPFB}_v^r := \ell_v^r$, which we will call the *most probable false beating (MPFB)* factor, to compare between different voting rules, since surprise can vary at most by a constant factor of this MPFB factor. In the following sections, we will see that the effect of the number of voters on this factor is in the exponent. Since the number of voters is large, the conclusions on surprise are entirely dictated by the growth or decay of the MPFB factor.

3 Theoretical Results

In this section, we first analyze the setting with two candidates to get a better insight. The set of candidates is $M = \{a_1, a_2\}$ and the classes are $P_1 = a_1 \succ a_2$ and $P_2 = a_2 \succ a_1$. WLOG, we assume that $\epsilon_1 = \frac{1}{2} + \epsilon$ and $\epsilon_2 = \frac{1}{2} - \epsilon$ with $0 < \epsilon < 1/2$. For two candidates, all standard voting rules yield the same winner as the plurality rule, and therefore, we will be considering only plurality in the case of two candidates. We first show that candidate a_1 emerges as winner in plurality *w.h.p.*

Theorem 1. *When voters fall in class P_1 and P_2 w.p. $\frac{1}{2} + \epsilon$ and $\frac{1}{2} - \epsilon$ respectively, with $0 < \epsilon < 1/2$, $P(w_T(\vec{N}, \texttt{Plu}) = a_2) \leqslant e^{-\sqrt{n}/2}$ for sufficiently large n.*

Owing to paucity of space, we refer the reader to the full version of this paper [14] for the proof.

Since candidate a_1 turns out to be the true winner *w.h.p.*, we will consider only the conditional probability that a_2 is the perceived winner given a_1 being the true winner, which will approximately be equal to surprise for large n.

Theorem 2 (Surprise for two candidates). *When voters fall in class P_1 and P_2 w.p. $\frac{1}{2} + \epsilon$ and $\frac{1}{2} - \epsilon$ respectively, with $0 < \epsilon < 1/2$, we have the following.*

For voter v in P_1:

(i) if $\frac{\hat{p}_{11}}{\hat{p}_{12}} > \frac{p_{11}}{p_{12}} \frac{1/2+\epsilon}{1/2-\epsilon}$, then $P(w_P(\hat{N}^v, \texttt{Plu}) = a_2 \mid w_T(\vec{N}, \texttt{Plu}) = a_1) \geqslant 1 - 2e^{-2\left(\frac{\hat{p}_{11}\hat{p}_{12}}{\hat{p}_{11}+\hat{p}_{12}}\right)^2 \sqrt{n}}$ for large enough n; hence, $\texttt{surp}_v^{\texttt{Plu}} \stackrel{n\to\infty}{\longrightarrow} 1$, i.e., voter v is surprised w.h.p.

(ii) if $\frac{\hat{p}_{11}}{\hat{p}_{12}} < \frac{p_{11}}{p_{12}} \frac{1/2+\epsilon}{1/2-\epsilon}$, then $P(w_P(\hat{N}^v, \texttt{Plu}) = a_2 \mid w_T(\vec{N}, \texttt{Plu}) = a_1) \leqslant e^{-2\left(\frac{\hat{p}_{11}\hat{p}_{12}}{\hat{p}_{11}+\hat{p}_{12}}\right)^2 \sqrt{n}}$ for large enough n; hence, $\texttt{surp}_v^{\texttt{Plu}} \stackrel{n\to\infty}{\longrightarrow} 0$, i.e., voter v is not surprised w.h.p.

For voter v in P_2:

(i) if $\frac{\hat{p}_{22}}{\hat{p}_{21}} < \frac{p_{22}}{p_{21}} \frac{1/2-\epsilon}{1/2+\epsilon}$, then $P(w_P(\hat{N}^v, \texttt{Plu}) = a_2 \mid w_T(\vec{N}, \texttt{Plu}) = a_1) \geqslant 1 - 2e^{-2\left(\frac{\hat{p}_{22}\hat{p}_{21}}{\hat{p}_{22}+\hat{p}_{21}}\right)^2 \sqrt{n}}$ for large enough n; hence, $\texttt{surp}_v^{\texttt{Plu}} \stackrel{n\to\infty}{\longrightarrow} 1$, i.e., voter v is surprised w.h.p.

(ii) if $\frac{\hat{p}_{22}}{\hat{p}_{21}} > \frac{p_{22}}{p_{21}} \frac{1/2-\epsilon}{1/2+\epsilon}$, then $P(w_P(\hat{N}^v, \texttt{Plu}) = a_2 \mid w_T(\vec{N}, \texttt{Plu}) = a_1) \leqslant e^{-2\left(\frac{\hat{p}_{22}\hat{p}_{21}}{\hat{p}_{22}+\hat{p}_{21}}\right)^2 \sqrt{n}}$ for large enough n; hence, $\texttt{surp}_v^{\texttt{Plu}} \stackrel{n\to\infty}{\longrightarrow} 0$, i.e., voter v is not surprised w.h.p.

Proof. We prove the result only for the case when $v \in P_2$, since the other case is symmetric. Define $\theta = 1/2 + \epsilon$. Let the random graph formed according to the stochastic model is denoted by $G = (N, E)$. For $i \in [2]$, let X_i be the set of voters denoting the neighbors of v that belong to class P_i. Hence, v's estimated number of voters in classes P_1 and P_2 are $\frac{|X_1|}{\hat{p}_{21}}$ and $\frac{|X_2|}{\hat{p}_{22}} + 1$ respectively. The additional one voter in the estimate of P_2 comes from voter v counting herself. Hence

$$\frac{|X_1|}{\hat{p}_{21}} = \frac{1}{\hat{p}_{21}} \sum_{u \in N} \mathbb{I}(\{(vu) \in E\} \cap \{u \in P_1\}), \tag{4}$$

$$\frac{|X_2|}{\hat{p}_{22}} = \frac{1}{\hat{p}_{22}} \sum_{u \in N \setminus \{v\}} \mathbb{I}(\{(vu) \in E\} \cap \{u \in P_2\}). \tag{5}$$

Taking expectations over these quantities, we get,

$$\begin{aligned} \mathbb{E}\left(\frac{|X_1|}{\hat{p}_{21}}\right) &= \frac{1}{\hat{p}_{21}} \sum_{u \in N} P(u \in P_1) \cdot P((vu) \in E) \mid u \in P_1) \\ &= n\,\theta\,\frac{p_{21}}{\hat{p}_{21}} \qquad \text{and,} \\ \mathbb{E}\left(\frac{|X_2|}{\hat{p}_{22}}\right) &= \frac{1}{\hat{p}_{22}} \sum_{u \in N \setminus \{v\}} P(u \in P_2) \cdot P((vu) \in E) \mid u \in P_2) \\ &= (n-1)\,(1-\theta)\,\frac{p_{22}}{\hat{p}_{22}}. \end{aligned}$$

Define a new random variable, $Z := \frac{|X_2|}{\hat{p}_{22}} + 1 - \frac{|X_1|}{\hat{p}_{21}}$. Its expectation is

$$\begin{aligned} \mathbb{E}Z &= (n-1)(1-\theta)\frac{p_{22}}{\hat{p}_{22}} + 1 - n\theta\frac{p_{21}}{\hat{p}_{21}} \\ &= (n-1)\left(\frac{1}{2} - \epsilon\right)\frac{p_{22}}{\hat{p}_{22}} + 1 - n\left(\frac{1}{2} + \epsilon\right)\frac{p_{21}}{\hat{p}_{21}} \\ &= n\left[\left(\left(\frac{1}{2} - \epsilon\right)\frac{p_{22}}{\hat{p}_{22}} - \left(\frac{1}{2} + \epsilon\right)\frac{p_{21}}{\hat{p}_{21}}\right) + \frac{1}{n}\left(1 - \left(\frac{1}{2} - \epsilon\right)\frac{p_{22}}{\hat{p}_{22}}\right)\right]. \end{aligned} \tag{6}$$

From the definition of Z, it is clear that

$$\Pr(w_P(\hat{N}^v, \texttt{Plu}) = a_2 \mid w_T(\vec{N}, \texttt{Plu}) = a_1) = \Pr(Z > 0).$$

We first consider the case when $\frac{\hat{p}_{22}}{\hat{p}_{21}} > \frac{p_{22}}{p_{21}} \cdot \frac{1/2-\epsilon}{1/2+\epsilon}$.

The first term in the bracket in Equation (6) is negative since $\frac{\hat{p}_{22}}{\hat{p}_{21}} > \frac{p_{22}}{p_{21}} \cdot \frac{1/2-\epsilon}{1/2+\epsilon}$, by assumption. Let $-\ell = \left(\frac{1}{2} - \epsilon\right)\frac{p_{22}}{\hat{p}_{22}} - \left(\frac{1}{2} + \epsilon\right)\frac{p_{21}}{\hat{p}_{21}}$. Hence the whole expression of Equation (6) is negative for $n > \max\{0, \left(1 - \left(\frac{1}{2} - \epsilon\right)\frac{p_{22}}{\hat{p}_{22}}\right)/\ell\} =: n_0$. Hence, $\mathbb{E}Z$ is negative for sufficiently large n. Note from Equations (4) and (5) that Z can also be written as the sum over the differences of the indicator functions. We will use Hoeffding's bound since the random variables in the sum are independent. The maximum of every term in that sum of indicators that represent Z can be $1/\hat{p}_{22}$ and the minimum can be $-1/\hat{p}_{21}$, hence the maximum difference between each of the summands is $(\hat{p}_{22}+\hat{p}_{21})/\hat{p}_{22}\hat{p}_{21}$. Now from Hoeffding's inequality, we have

$$\Pr(Z - \mathbb{E}Z > t) \leqslant e^{-2\left(\frac{\hat{p}_{22}\hat{p}_{21}}{\hat{p}_{22}+\hat{p}_{21}}\right)^2 \cdot \frac{t^2}{n}}. \quad (7)$$

Plugging in $t = n^{3/4}$, we get that the probability of $Z > \mathbb{E}Z + n^{3/4}$ is at most $e^{-2\left(\frac{\hat{p}_{22}\hat{p}_{21}}{\hat{p}_{22}+\hat{p}_{21}}\right)^2 \sqrt{n}}$. Let $n_1 := \inf\{n > 0 : \left(\left(\frac{1}{2}-\epsilon\right)\frac{p_{22}}{\hat{p}_{22}} - \left(\frac{1}{2}+\epsilon\right)\frac{p_{21}}{\hat{p}_{21}}\right) + \frac{1}{n}\left(1-\left(\frac{1}{2}-\epsilon\right)\frac{p_{22}}{\hat{p}_{22}}\right) + \frac{1}{n^{1/4}} < 0\}$. The number n_1 is guaranteed to exist since $\frac{\hat{p}_{22}}{\hat{p}_{21}} > \frac{p_{22}}{p_{21}} \cdot \frac{1/2-\epsilon}{1/2+\epsilon}$, by assumption. Therefore for all $n > n_1$, Z is greater than a negative quantity with probability at most $e^{-2\left(\frac{\hat{p}_{22}\hat{p}_{21}}{\hat{p}_{22}+\hat{p}_{21}}\right)^2 \sqrt{n}}$. Since $\{Z > 0\} \subset \{Z > -ve\}$, we have that $\forall n > n_1$, $\Pr(Z > 0) \leqslant e^{-2\left(\frac{\hat{p}_{22}\hat{p}_{21}}{\hat{p}_{22}+\hat{p}_{21}}\right)^2 \sqrt{n}}$.

We now consider the case when $\frac{\hat{p}_{22}}{\hat{p}_{21}} < \frac{p_{22}}{p_{21}}\frac{1/2-\epsilon}{1/2+\epsilon}$. We leverage the calculations we did for the previous case. Because of the assumption $\frac{\hat{p}_{22}}{\hat{p}_{21}} < \frac{p_{22}}{p_{21}} \cdot \frac{1/2-\epsilon}{1/2+\epsilon}$, $\mathbb{E}Z$ is positive for large n (Equation (6)). Using Hoeffding's inequality, we have

$$\Pr(|Z - \mathbb{E}Z| \leqslant t) \geqslant 1 - 2e^{-2\left(\frac{\hat{p}_{22}\hat{p}_{21}}{\hat{p}_{22}+\hat{p}_{21}}\right)^2 \cdot \frac{t^2}{n}}.$$

Since $\{|Z - \mathbb{E}Z| \leqslant t\} \implies \{Z \geqslant \mathbb{E}Z - t\}$, the probability of the event on the RHS is at least that of the LHS. With the following choice of t, we show that the RHS implies $\{Z > 0\}$ for large n. Plugging in $t = n^{3/4}$ and defining $n_2 := \inf\{n > 0 : \left(\left(\frac{1}{2}-\epsilon\right)\frac{p_{22}}{\hat{p}_{22}} - \left(\frac{1}{2}+\epsilon\right)\frac{p_{21}}{\hat{p}_{21}}\right) + \frac{1}{n}\left(1-\left(\frac{1}{2}-\epsilon\right)\frac{p_{22}}{\hat{p}_{22}}\right) - \frac{1}{n^{1/4}} > 0\}$, which is guaranteed to exist by assumption, we get the desired conclusion for all $n > n_2$. This completes the proof. □

Corollaries. Theorem 2 captures the determining factors for surprise in plurality voting. Few conclusions are in order.

If an agent's estimated $\hat{p}$'s were perfect, then the agent is never surprised *w.h.p.*, since then the ratios will always satisfy the 'not surprised' condition of Theorem 2.

Surprise may happen when ϵ is small, i.e., the winning margin is small. This is because, the surprise-determining thresholds for p_{jj}/p_{jk}s in Theorem 2 are very close to the actual ratios p_{jj}/p_{jk}s and a small error of the voter in estimating these connection parameters may lead to surprise. However, when the winning margin is large, e.g., ϵ is large enough such that $\frac{p_{22}}{p_{21}}\frac{1/2-\epsilon}{1/2+\epsilon} < 1$ and if the $\hat{p}$'s are also regular, i.e., $\hat{p}_{22} > \hat{p}_{21}$, then no agent in P_2 will be surprised. This shows that elections with an overwhelming majority can hardly be surprising. Surprise is a phenomenon only of a *closely contested election.*

3.1 Three Candidates

We now consider the problem with three candidates. In this setting, different voting rules give rise to different winners and therefore it is possible to distinguish them w.r.t. the surprise metric. In this section, we will compare three voting rules, namely plurality, Borda, and veto (explained below), based on the factor $\mathtt{MPFB}_v^r$ (Equation (3)) because of the reason explained right after the equation in §2. A collection of m-dimensional vectors $\vec{s}_m = (\alpha_1, \alpha_2, \ldots, \alpha_m) \in \mathbb{R}^m$ with $\alpha_1 \geqslant \alpha_2 \geqslant \cdots \geqslant \alpha_m$ and $\alpha_1 > \alpha_m$ for every $m \in \mathbb{N}$ defines a voting rule (called scoring rule) — a candidate receives a score of α_i from a vote if it is placed at the i-th position in that vote, and the score of a candidate is the sum of the scores it receives from all the votes. The winners are the candidates with the maximum score. The score vectors for the plurality, Borda, and veto voting rules are $(1, 0, \ldots, 0)$, $(m-1, m-2, \ldots, 1, 0)$, and $(1, \ldots, 1, 0)$ respectively. Scoring rules remain unchanged if we multiply every α_i by any constant $\lambda > 0$ and/or add any constant μ. Hence, we assume without loss of generality that, for 3 candidates, the Borda score vector is $(2/3, 1/3, 0)$ and the veto score vector is $(1/2, 1/2, 0)$ to ensure that $\sum_{i=1}^{3} \alpha_i = 1$ for all the rules.

We chose these three voting rules because (1) they are most frequently used, and (2) the distribution of scores in these rules has wide variety – the whole score concentrated at the top alternative for plurality, (almost) equally distributed for veto, and in between these two extremes for Borda.

For two candidates, we have seen that surprise occurs only in closely contested elections. Hence to compare the voting rules in this section, we consider that the voters are uniformly distributed over the $|C|$ preference classes.

Assumption 1 (Uniform Population). *Every voter belongs to exactly one class of preference in $\{P_k : k \in C\}$ with uniform probability.*

We also assume that the voters' estimates of the connection probabilities are consistently higher than their true values as the KT distance increases between the preference class of the voter and the class of her neighbor, i.e., $p_{ij}/\hat{p}_{ij}$'s are decreasing in $\mathtt{dist}_{\mathsf{KT}}(P_i, P_j)$. The motivation is to capture the fact that people often consider their local neighborhood to be representative of the global population, leading to an uniform $\hat{p}_{ij}$'s for all $i, j \in C$. Since the true connection probabilities are regular, i.e., decreasing in $\mathtt{dist}_{\mathsf{KT}}(P_i, P_j)$, it gives rise to a *monotone estimation error.*

Assumption 2 (Monotone Estimation Error (MEE)). *The ratio of the true connection probability to the estimated one decreases with the KT distance, i.e., $\frac{p_{k\ell}}{\hat{p}_{k\ell}} \geqslant \frac{p_{kp}}{\hat{p}_{kp}}$ when $\mathtt{dist}_{\mathsf{KT}}(P_k, P_\ell) < \mathtt{dist}_{\mathsf{KT}}(P_k, P_p)$ when $v \in P_k$.*

In the proof of our main result in this section, we will use a quantitative version of the central limit theorem due to Berry [3] and Esseen [30]. The following exposition is from Tao [47].

Theorem 3 (Berry-Esseen). *Let X be a random variable with mean μ, unit variance, and finite third moment. Let $Z_n = \frac{\sum_{i=1}^{n} X_i}{\sqrt{n}}$, where X_i's are i.i.d. copies of X. Then we have $\Pr[Z_n > \lambda] = \Pr[G > \lambda] + O(\mathbb{E}|X|^3/\sqrt{n})$, uniformly for all $\lambda \in \mathbb{R}$, where $G \equiv Normal(\mu, 1)$, and the implied constant in $O(\cdot)$ is absolute and does not depend on the distribution of X.*

This theorem gives a quantitative guarantee on the deviation of the cumulative distribution function of the random variable Z_n from that of a normal random variable with mean same as X and unit variance.

With the assumptions as mentioned above, we present our main result for three candidates in the following theorem. Informally, this theorem compares plurality, Borda, and veto voting rules based on MPFB factor. Since $\mathtt{surp}_v^r = \Theta(\mathtt{MPFB}_v^r)$ (Equation (3)), we conclude that a lower MPFB factor gives a lower surprise.

Theorem 4. *Consider $|M| = 3$, and voters are generated from an uniform population. Let v be any voter.*

(i) If v ranks the true winner at the first position, then $\mathtt{MPFB}_v^{\mathtt{Plu}} \leqslant \mathtt{MPFB}_v^{\mathtt{Bor}} \leqslant \mathtt{MPFB}_v^{\mathtt{Vet}}$ w.h.p.
(ii) If v ranks the true winner at the second position, then $\mathtt{MPFB}_v^{\mathtt{Vet}} \leqslant \mathtt{MPFB}_v^{\mathtt{Bor}} \leqslant \mathtt{MPFB}_v^{\mathtt{Plu}}$ w.h.p.
(iii) If v ranks the true winner at the last position, then $\mathtt{MPFB}_v^{\mathtt{Vet}} \leqslant \mathtt{MPFB}_v^{\mathtt{Bor}}$ and $\mathtt{MPFB}_v^{\mathtt{Plu}} \leqslant \mathtt{MPFB}_v^{\mathtt{Bor}}$ w.h.p.

Discussion: This result gives us a fine grained information regarding the performance on surprise of different voting rules in different voter classes. It is also clear that among these standard voting rules there is no single rule that reduces surprise for all sections of voters. But we find it interesting that the performance on surprise is not proportional to the distribution of scores in the rules, since in case (iii), Borda, that has non-extreme distribution of scores performs worse than both the other two rules having extreme score distributions.

Proof. Let $M = \{a_1, a_2, a_3\}$. We label the classes as shown in Table 1. Each voter belongs to class P_k w.p. $1/6$ in the uniform population model (Assumption 1). WLOG, assume that the candidate a_2 wins the election *w.h.p.*, i.e., the overall score is highest for a_2 in every rule, and ties are broken in favor of a_2. Let $(s_1, s_2, 0)$ be a normalized scoring rule vector with $s_1 + s_2 = 1$ and $s_1, s_2 \geqslant 0$. Hence, the vector is $(1, 0, 0)$, $(2/3, 1/3, 0)$, and $(1/2, 1/2, 0)$ respectively for Plu, Bor, and Vet. For a voter v, let $\hat{s}_v(a_1), \hat{s}_v(a_2), \hat{s}_v(a_3)$ be the random variables denoting the estimated scores for the candidates a_1, a_2, and a_3 perceived by v.

Class	Preferences	Class	Preferences
P_1:	$a_1 \succ a_2 \succ a_3$	P_4:	$a_2 \succ a_3 \succ a_1$
P_2:	$a_1 \succ a_3 \succ a_2$	P_5:	$a_3 \succ a_1 \succ a_2$
P_3:	$a_2 \succ a_1 \succ a_3$	P_6:	$a_3 \succ a_2 \succ a_1$

Table 1: Preference classes for 3 candidates

For every rule r and voter v, we are interested in the differences of these estimated scores, i.e., $\hat{s}_v(a_j) - \hat{s}_v(a_2),\ j = 1, 3$, since a positive value of this expression implies that a false beating event has occurred. The maximum probability of these two events is $\mathtt{MPFB}_v^r$.

With the voters' winner perception model, each of these estimated scores of v can be written as a sum over the indicator RVs that another voter belong to a specific preference class and they are connected to v (with appropriate scaling with $\hat{p}_{kl}$ if $v \in P_k$ and the other voter is in P_l). Hence, we can write the difference in the estimated scores as $\hat{s}_v(a_1) - \hat{s}_v(a_2) = \sum_{u \in N \setminus \{v\}} X_{u, a_1 - a_2} + \delta_{v, a_1 - a_2}$ and $\hat{s}_v(a_3) - \hat{s}_v(a_2) = \sum_{u \in N \setminus \{v\}} X_{u, a_3 - a_2} + \delta_{v, a_3 - a_2}$, where we clearly distinguish the contribution of voter v in the differences with the variable $\delta_{v, a_j - a_2},\ j = 1, 3$. We denote the summation on the RHS in each equality with the shorthand $S_{-v, a_j - a_2} := \sum_{u \in N \setminus \{v\}} X_{u, a_j - a_2},\ j = 1, 3$. The expression $X_{u, a_1 - a_2}$ (resp. $X_{u, a_3 - a_2}$) is the indicator random variable denoting voter u's contribution to the difference in the score of a_1 (resp. a_3) and a_2 if u is connected to v. We detail out the exact expressions of $X_{u, a_j - a_2}$ when we consider the following cases.

Case 1: $v \in P_1$ or $v \in P_6$ (i.e., when v ranks the winner at the second position): We only consider $v \in P_1$, since the analysis for $v \in P_6$ is symmetric. For $v \in P_1$, the expression of $X_{u, a_1 - a_2}$ turns out as follows for $u \in N \setminus \{v\}$.

$$X_{u, a_1 - a_2} =$$
$$(s_1 - s_2)\left(\frac{1}{\hat{p}_{11}}\mathbb{I}(\{(u, v) \in E\} \cap \{u \in P_1\}) - \frac{1}{\hat{p}_{12}}\mathbb{I}(\{(u, v) \in E\} \cap \{u \in P_3\})\right)$$
$$+ s_2\left(\frac{1}{\hat{p}_{12}}\mathbb{I}(\{(u, v) \in E\} \cap \{u \in P_5\}) - \frac{1}{\hat{p}_{12}}\mathbb{I}(\{(u, v) \in E\} \cap \{u \in P_6\})\right)$$
$$+ s_1\left(\frac{1}{\hat{p}_{12}}\mathbb{I}(\{(u, v) \in E\} \cap \{u \in P_2\}) - \frac{1}{\hat{p}_{12}}\mathbb{I}(\{(u, v) \in E\} \cap \{u \in P_4\})\right).$$

Note that these are i.i.d. random variables for $u \in N \setminus \{v\}$, whose mean and variances are as follows.

$$\mathbb{E}[X_{u, a_1 - a_2}] = (s_1 - s_2)(p_{11}/6\hat{p}_{11} - p_{12}/6\hat{p}_{12}) \geqslant 0$$

We get the equality due to Assumption 1 and the inequality due to Assumption 2. We also have

$$\mathbb{E}[X^2_{u, a_1 - a_2}] = (s_1 - s_2)^2\left(p_{11}/6\hat{p}^2_{11} + p_{12}/6\hat{p}^2_{12}\right) + (s_1^2 + s_2^2)p_{12}/3\hat{p}^2_{12}.$$

Hence

$$\mathrm{var}(X_{u, a_1 - a_2}) = \mathbb{E}[X^2_{u, a_1 - a_2}] - \left(\mathbb{E}[X_{u, a_1 - a_2}]\right)^2$$
$$= (s_1 - s_2)^2\left(\frac{p_{11}}{6\hat{p}^2_{11}} + \frac{p_{12}}{6\hat{p}^2_{12}} - \left(\frac{p_{11}}{6\hat{p}_{11}} - \frac{p_{12}}{6\hat{p}_{12}}\right)^2\right) + \frac{p_{12}}{3\hat{p}^2_{12}}(s_1^2 + s_2^2).$$

For $u \in N \setminus \{v\}$, define the normalized random variable

$$\bar{X}_{u, a_1 - a_2} = X_{u, a_1 - a_2}/\sqrt{\mathrm{var}(X_{u, a_1 - a_2})}.$$

Clearly, $\mathbb{E}[\bar{X}_{u, a_1 - a_2}] = \mathbb{E}[X_{u, a_1 - a_2}]/\sqrt{\mathrm{var}(X_{u, a_1 - a_2})}$ and $\mathrm{var}(\bar{X}_{u, a_1 - a_2}) = 1$. We can now apply Theorem 3 for large n to get

$$\begin{aligned}
&\Pr[S_{-v, a_1 - a_2} + \delta_{1, a_1 - a_2} > 0 \mid v \in P_1] \\
&= \Pr\left[\frac{S_{-v, a_1 - a_2}}{\sqrt{n\mathrm{var}(X_{u, a_1 - a_2})}} + \frac{\delta_{1, a_1 - a_2}}{\sqrt{n\mathrm{var}(X_{u, a_1 - a_2})}} > 0 \mid v \in P_1\right] \\
&= \Pr\left[G_{v, a_1 - a_2} > -\delta_{1, a_1 - a_2}/\sqrt{n\mathrm{var}(X_{u, a_1 - a_2})}\right] + O\left(1/\sqrt{n}\right) \\
&= \Pr[G_{v, a_1 - a_2} \geqslant 0] + O\left(1/\sqrt{n}\right).
\end{aligned} \quad (8)$$

Where $G_{v, a_1 - a_2}$ is a normal RV with mean $\mathbb{E}[\bar{X}_{u, a_1 - a_2}]$ and unit variance. The last equality follows from the fact that $\Pr\left[0 > G_{v, a_1 - a_2} > -\delta_{1, a_1 - a_2}/\sqrt{n\mathrm{var}(X_{u, a_1 - a_2})}\right] = O(1/\sqrt{n})$ since the length of the interval $\left[-\delta_{1, a_1 - a_2}/\sqrt{n\mathrm{var}(X_{u, a_1 - a_2})}, 0\right]$ is $O(1/\sqrt{n})$, hence the integral of any probability distribution over it is $O(1/\sqrt{n})$.

Similarly, for $u \in N \setminus \{v\}$, $X_{u, a_3 - a_2}$ is defined as follows.

$$X_{u, a_3 - a_2} =$$
$$(s_1 - s_2)\left(\frac{1}{\hat{p}_{12}}\mathbb{I}(\{(u, v) \in E\} \cap \{u \in P_6\}) - \frac{1}{\hat{p}_{12}}\mathbb{I}(\{(u, v) \in E\} \cap \{u \in P_4\})\right)$$
$$+ s_2\left(\frac{1}{\hat{p}_{12}}\mathbb{I}(\{(u, v) \in E\} \cap \{u \in P_2\}) - \frac{1}{\hat{p}_{11}}\mathbb{I}(\{(u, v) \in E\} \cap \{u \in P_1\})\right)$$
$$+ s_1\left(\frac{1}{\hat{p}_{12}}\mathbb{I}(\{(u, v) \in E\} \cap \{u \in P_5\}) - \frac{1}{\hat{p}_{12}}\mathbb{I}(\{(u, v) \in E\} \cap \{u \in P_3\})\right).$$

Taking expectation, we get

$$\mathbb{E}[X_{u,a_3-a_2}] = s_2(p_{12}/6\hat{p}_{12} - p_{11}/6\hat{p}_{11}) \leqslant 0$$

The equality follows due to Assumption 1 and the inequality due to Assumption 2. Performing similar calculation as we did for X_{u,a_1-a_2}, we reach a unit variance normal RV G_{v,a_3-a_2}. However, the mean of G_{v,a_1-a_2} turns out to be larger than G_{v,a_3-a_2}, which lead to the conclusion that for large n

$$\Pr[S_{-v,a_1-a_2} + \delta_{1,a_1-a_2} > 0 \mid v \in P_1] \geqslant \Pr[S_{-v,a_3-a_2} + \delta_{1,a_3-a_2} > 0 \mid v \in P_1]$$

Hence, to find the MPFB factor in this case, we need to compare the probability of Equation (8) among different voting rules. Since, the probability reduces to the tail distribution of G_{v,a_1-a_2} which is a normal RV with unit variance, it is enough to compare the means of G_{v,a_1-a_2} to compare the MPFB factors. Denoting the means of G_{v,a_1-a_2} by μ_v^r for voter v under rule r, we get for large enough n

$$\mu_v^{\texttt{Vet}} \leqslant \mu_v^{\texttt{Bor}} \leqslant \mu_v^{\texttt{Plu}}.$$

Which implies *w.h.p.*

$$\texttt{MPFB}_v^{\texttt{Vet}} \leqslant \texttt{MPFB}_v^{\texttt{Bor}} \leqslant \texttt{MPFB}_v^{\texttt{Plu}}.$$

The analysis for $v \in P_6$ is the same with the roles of candidates a_1 and a_3 being reversed. Hence, we have proved claim (ii) of the theorem.

Case 2: $v \in P_2$ or $v \in P_5$ (i.e., when v ranks the winner at the last position): We adopt a similar calculation as Case 1 to get

$$\mathbb{E}[X_{u,a_1-a_2}] = s_1(p_{11}/6\hat{p}_{11} - p_{12}/6\hat{p}_{12}) \geqslant 0$$
$$\mathbb{E}[X^2_{u,a_1-a_2}] = s_1^2(p_{11}/6\hat{p}^2_{11} + p_{12}/6\hat{p}^2_{12}) + p_{12}/3\hat{p}^2_{12}((s_1 - s_2)^2 + s_2^2)$$

With notations similar to Case 1, we denote the mean of the normalized variance normal RV G_{v,a_j-a_2} by μ^r_{v,a_j-a_2}, $j = 1,3$, for the differences of estimated scores of voter v between candidates a_j and a_2, $j = 1,3$. Hence

$$\mu^r_{v,a_j-a_2} = \mathbb{E}\left[\frac{S_{-v,a_j-a_2}}{\sqrt{n\text{var}(X_{u,a_j-a_2})}} \mid v \in P_2\right],\ j = 1,3.$$

With similar computations, we get for voter $v \in P_2$

$$\max\{\mu^{\texttt{Plu}}_{v,a_1-a_2}, \mu^{\texttt{Plu}}_{v,a_3-a_2}\} = \max\{\mu^{\texttt{Vet}}_{v,a_1-a_2}, \mu^{\texttt{Vet}}_{v,a_3-a_2}\} \leqslant \max\{\mu^{\texttt{Bor}}_{v,a_1-a_2}, \mu^{\texttt{Bor}}_{v,a_3-a_2}\}$$

Which implies *w.h.p.*

$$\texttt{MPFB}_v^{\texttt{Vet}} \leqslant \texttt{MPFB}_v^{\texttt{Bor}} \quad \text{and} \quad \texttt{MPFB}_v^{\texttt{Plu}} \leqslant \texttt{MPFB}_v^{\texttt{Bor}}.$$

The case for $v \in P_5$ is same with the roles of candidates a_1 and a_3 reversed. Hence, we have proved claim (iii) of the theorem.

Case 3: $v \in P_3$ or $v \in P_4$ (i.e., when v ranks the winner at the first position): With notations similar to Case 1, and denoting the mean of the normalized variance normal RV G_{v,a_j-a_2} by μ^r_{v,a_j-a_2}, $j = 1,3$, for the differences of estimated scores of voter v between candidates a_j and a_2, $j = 1,3$, we have

$$\mu^r_{v,a_j-a_2} = \mathbb{E}\left[\frac{S_{-v,a_j-a_2}}{n\sqrt{\text{var}(X_{u,a_j-a_2})}} \mid v \in P_3\right],\ j = 1,3.$$

With similar computations, we get for voter $v \in P_3$

$$\max\{\mu^{\texttt{Plu}}_{v,a_1-a_2}, \mu^{\texttt{Plu}}_{v,a_3-a_2}\} \leqslant \max\{\mu^{\texttt{Vet}}_{v,a_1-a_2}, \mu^{\texttt{Vet}}_{v,a_3-a_2}\} \leqslant \max\{\mu^{\texttt{Bor}}_{v,a_1-a_2}, \mu^{\texttt{Bor}}_{v,a_3-a_2}\}$$

Which implies *w.h.p.*

$$\texttt{MPFB}_v^{\texttt{Plu}} \leqslant \texttt{MPFB}_v^{\texttt{Bor}} \leqslant \texttt{MPFB}_v^{\texttt{Vet}}.$$

The case for $v \in P_4$ is same with the roles of candidates a_1 and a_3 reversed. Hence, we have proved claim (i) of the theorem. □

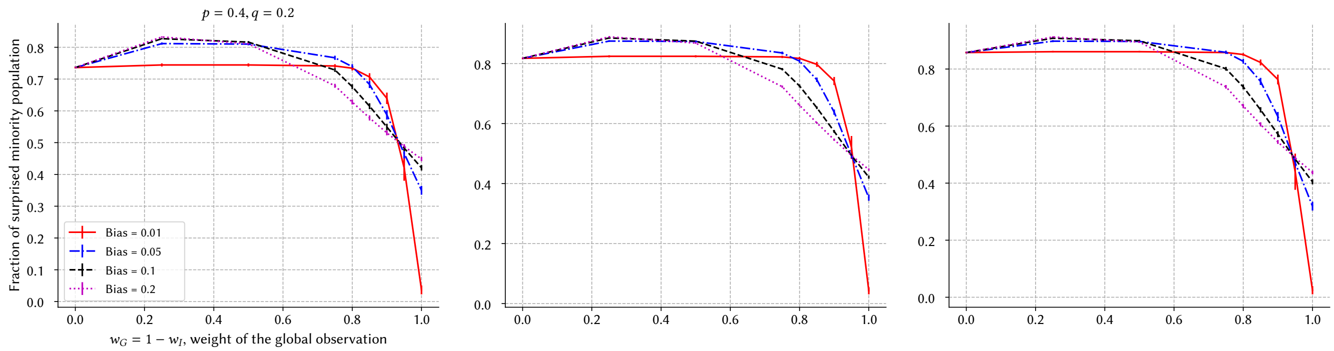
4 Empirical Results

Our theoretical results in §3 use some simplifying assumptions in the interest of a cleaner analysis. Firstly, we assumed that the voters have an estimate of the connection probabilities, though we do not explicitly mention how the voters arrive at these estimates. In practice, voters anticipate a winner by implicitly estimating the number of voters voting in favor of the candidate versus voting against him. There are typically two major sources of information to a voter: first, via her own neighbors in the (online/offline) social network, and second via the public broadcasting media – print or electronic. Secondly, our connection model was following the stochastic block model that only depends on the voters' preferences and had no dependence on the voters' geographical locations. In this section, we relax these two simplifying assumptions and from an empirical viewpoint try to see if the broad theoretical predictions hold.

We instantiate the voting population with a real election dataset. We construct the social network of voters depending on their preferences and geographical locations to make the network more realistic. We capture a voter v's estimates of the population of different classes by taking a weighted average of (1) voter v's individual observation, i.e., the number of voters of different classes in v's immediate neighborhood and (2) a noisy version of the global (true) number of voters in each class. Effects (1) and (2) capture a voter's private and public observations respectively and give a realistic view of opinion forming.

Datasets: We use the UK election dataset of EU referendum (popularly known as Brexit)[6]. The dataset is publicly available and gives the total count of votes cast by the UK voters that voted either 'remain' (R) or 'leave' (L) the EU. The data consists of approximately 33 million valid votes and is partitioned across 382 regions within the UK. Each region is identified with the name of the town, city, or county. We will refer to this dataset as Brexit dataset. We have used another dataset[7] to find the latitude and longitude of these regions. Since the location dataset gives the latitude-longitude of a town and the voting constituencies are collection of a number of them, we have averaged over the towns in a region to find the approximate centroid of the region. There were few locations (about 18%) whose information were not available in the location dataset, we have filled in their location to be the centroid of all the available locations in the dataset. The Brexit data are suitable for our experiment, since (a) it has only two candidates for which we have a simple yet insightful theoretical result (Theorem 2), (b) it is large enough to draw conclusions on large-scale elections, and (c) the election was closely contested, (51.9% for L and 48.1% for R).

[6] https://goo.gl/MtTdIT
[7] https://www.townslist.co.uk/



(a) $p = 0.4$, $q = 0.2$ **(b)** $p = 0.6$, $q = 0.2$ **(c)** $p = 0.8$, $q = 0.2$

Figure 1: Surprise in Brexit for different intra and inter-class connection probabilities (legends same as Figure 1a).

Approach: In each location, based on the total number of voters and their votes, we re-created the voters. The connection follows a random graph model where the probability of connection between two voters is the average of (a) p_1, which is decreasing in the geographical distance between the voters, and (b) p_2, which is p if both voters are from the same class, and q, otherwise (with $p \geqslant q$). This relaxation from the theoretical model allows for a social connection where two individuals are geographically close despite having different political opinions.

In this social network, voters perceive the outcome of the election according to effects (1) and (2) as explained before. For the individual observation (effect 1), we assume that a voter can perfectly observe the true voting preferences of her immediate neighbors in the graph. The number of voters that voted R or L in the immediate neighborhood gives a distribution of the R and L voters in the neighborhood including herself. For the global observation (effect 2), we add a zero mean truncated Gaussian noise to the *true* distribution of the votes – the truncated Gaussian is set such that after the addition of noise, the resulting noisy distribution still remains a valid one, i.e., no probability mass goes negative. We call the variance of the truncated Gaussian the *bias* of this observation. The voter combines these two distributions with weights w_I for the noise-free individual distribution and w_G for the noisy global distribution. Her perceived winner is the one that has larger mass among the two outcomes in the weighted sum distribution.

Due to the massive scale of the dataset, which takes significant time to run a single experiment, we have sampled 10,000 votes uniformly at random and created a sub-election. In this sub-election, every individual attempts to connect to 500 other individuals picked uniformly at random. In this discussion, we consider the surprise of the voters in the minority class of this sub-election (i.e., the voters whose favorite candidate does not win – hence they get surprised when they perceive this candidate to be the winner).

Results: We have three independent parameters that give rise to surprise: (1) the weight on global observation w_G (w_I is fixed given this), (2) the bias on this observation, and (3) the choices of p and q. To show how these parameters affect surprise, we plot the fraction of surprised minority population versus w_G for different choices of observation bias of the global distribution. Figure 1a shows such a plot for a specific choice of p and q. Figures 1a to 1c show similar information when p/q increases.

Observations. Some results support our theoretical predictions, even after relaxing our assumptions on network formation and voter estimates. (i) When the ratio p/q is large, the surprises are large too. A large p_{22}/p_{21} implies that more $\hat{p}_{22}, \hat{p}_{21}$ satisfies the condition of surprise in part 2 of Theorem 2 (here $p_{22} = p, p_{21} = q$) – giving rise to a higher surprise. (ii) More bias in the observation leads to a higher surprise (note, e.g., when w_G is close to 1.0 in Figure 1a). A larger bias gives rise to a larger chance that the ratio of the estimated number of people voting *for* and *against* her favorite candidate, which is same as the ratio of the estimated p and q, will be different from p/q – thereby making the condition of surprise in part 2 of Theorem 2 getting satisfied more likely.

However, we also find few observations surprising. The downward trend of the curve was expected with more weight on global information – but when there is noise in the global information, there is an increase and dip in the surprise. Also, each curve shows a cross-over region, where mixing a more noisy global observation gives a lower surprise.

5 Discussion

This paper gives a quantitative understanding of why and how surprise can occur in elections. The results for more than two candidates hint that possibly no single voting rule can reduce the surprise for all sections of the voters. Our empirical results complement our assumption on voter's estimates of connection probabilities.

Several interesting future works can spawn out of this work. The network structure here is assumed to be of stochastic block model. However a more natural connection model in social networks follow power law distribution. We believe that if the stochastic model of the connections is known to the agents, surprise will exhibit a similar phenomenon when the estimates of the connection model is mistaken by the agents beyond a certain point. Also, it is intuitive that if an agent can observe the preferences of friends' friends and so on, they are less likely to be surprised – an extreme case of this is when an agent can observe all other agents' preferences (if the network is strongly connected). A future result may capture this phenomenon quantitatively.

Though a more fine-grained model of voter perception and surprise will help better understand the phenomenon, we believe that a formal study of surprise in elections is essential for mitigating it, and our work contributes with a starting model in that direction.

References

[1] J. Bartholdi, C. A. Tovey, and M. A. Trick. 1989. The Computational Difficulty of Manipulating an Election. *Social Choice and Welfare* 6 (1989), 227–241.
[2] Dirk Bergemann and Stephen Morris. 2017. Information design: A unified perspective. (2017).
[3] Andrew C Berry. 1941. The accuracy of the Gaussian approximation to the sum of independent variates. *Transactions of the american mathematical society* 49, 1 (1941), 122–136.
[4] Alessandro Bessi and Emilio Ferrara. 2016. Social bots distort the 2016 US Presidential election online discussion. (2016).
[5] Arnab Bhattacharyya, Palash Dey, and David P. Woodruff. 2016. An Optimal Algorithm for l1-Heavy Hitters in Insertion Streams and Related Problems. In *Proc. 35th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2016, San Francisco, CA, USA, June 26 - July 01, 2016*. 385–400. https://doi.org/10.1145/2902251.2902284
[6] Robert Bredereck, Piotr Faliszewski, Rolf Niedermeier, and Nimrod Talmon. 2016. Large-scale election campaigns: Combinatorial shift bribery. *Journal of Artificial Intelligence Research* 55 (2016), 603–652.
[7] Michael P Cameron, Patrick Barrett, and Bob Stewardson. 2016. Can social media predict election results? Evidence from New Zealand. *Journal of Political Marketing* 15, 4 (2016), 416–432.
[8] Vincent Conitzer. 2012. Should social network structure be taken into account in elections? *Mathematical Social Sciences* 64, 1 (2012), 100–102.
[9] V. Conitzer, T. Sandholm, and J. Lang. 2007. When Are Elections with Few Candidates Hard to Manipulate? *J. ACM* 54, 3 (2007), 1–33.
[10] Dmitry Dagaev and Alex Suzdaltsev. 2018. Seeding, Competitive Intensity and Quality in Knock-Out Tournaments. *Journal of Combinatorial Optimization* 35, 1 (January 2018), 170–188.
[11] Palash Dey. 2017. Query Complexity of Tournament Solutions. In *Proc. 31-st AAAI Conference on Artificial Intelligence*. 2992–2998.
[12] Palash Dey. 2018. Manipulative elicitation - A new attack on elections with incomplete preferences. *Theor. Comput. Sci.* 731 (2018), 36–49. https://doi.org/10.1016/j.tcs.2018.03.028
[13] Palash Dey and Arnab Bhattacharyya. 2015. Sample Complexity for Winner Prediction in Elections. In *Proc. 14th International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2015, Istanbul, Turkey, May 4-8, 2015*. 1421–1430. http://dl.acm.org/citation.cfm?id=2773334
[14] Palash Dey, Pravesh K. Kothari, and Swaprava Nath. 2018. Surprise in Elections. (2018). https://arxiv.org/abs/1801.10110 ArXiv preprint, arXiv:1801.10110.
[15] Palash Dey and Neeldhara Misra. 2016. Elicitation for Preferences Single Peaked on Trees. In *Proc. 25th International Joint Conference on Artificial Intelligence*. 215–221.
[16] Palash Dey and Neeldhara Misra. 2016. Preference Elicitation for Single Crossing Domain. In *Proc. 25th International Joint Conference on Artificial Intelligence*. 222–228.
[17] Palash Dey and Neeldhara Misra. 2017. On the Exact Amount of Missing Information that Makes Finding Possible Winners Hard. In *Proc. 42nd International Symposium on Mathematical Foundations of Computer Science, MFCS 2017, August 21-25, 2017 - Aalborg, Denmark*. 57:1–57:14. https://doi.org/10.4230/LIPIcs.MFCS.2017.57
[18] Palash Dey, Neeldhara Misra, and Y. Narahari. 2015. Detecting Possible Manipulators in Elections. In *Proc. 14th International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2015, Istanbul, Turkey, May 4-8, 2015*. 1441–1450. http://dl.acm.org/citation.cfm?id=2773336
[19] Palash Dey, Neeldhara Misra, and Y. Narahari. 2016. Kernelization complexity of possible winner and coalitional manipulation problems in voting. *Theor. Comput. Sci.* 616 (2016), 111–125. https://doi.org/10.1016/j.tcs.2015.12.023
[20] Palash Dey, Neeldhara Misra, and Y. Narahari. 2017. Frugal bribery in voting. *Theor. Comput. Sci.* 676 (2017), 15–32. https://doi.org/10.1016/j.tcs.2017.02.031
[21] Palash Dey, Neeldhara Misra, and Y. Narahari. 2017. Parameterized Dichotomy of Choosing Committees Based on Approval Votes in the Presence of Outliers. In *Proc. 16th Conference on Autonomous Agents and MultiAgent Systems*. 42–50.
[22] Palash Dey, Neeldhara Misra, and Y. Narahari. 2018. Complexity of manipulation with partial information in voting. *Theor. Comput. Sci.* 726 (2018), 78–99. https://doi.org/10.1016/j.tcs.2018.03.012
[23] Palash Dey and Y. Narahari. 2014. Asymptotic collusion-proofness of voting rules: the case of large number of candidates. In *Proc. 13th International conference on Autonomous Agents and Multi-Agent Systems*. 1419–1420.
[24] Palash Dey and Y Narahari. 2015. Estimating the Margin of Victory of an Election Using Sampling.. In *IJCAI*. 1120–1126.
[25] Palash Dey, Nimrod Talmon, and Otniel van Handel. 2017. Proportional Representation in Vote Streams. In *Proc. 16th Conference on Autonomous Agents and MultiAgent Systems*. 15–23.
[26] Michael Dummett and Robin Farquharson. 1961. Stability in voting. *Econometrica: Journal of The Econometric Society* (1961), 33–43.
[27] The Economist. 2016. Who said Brexit was a surprise? https://www.economist.com/blogs/graphicdetail/2016/06/polls-versus-prediction-markets. [Online; accessed 3-September-2017].
[28] E. Elkind, P. Faliszewski, and A. Slinko. 2009. Swap Bribery. In *Proceedings of the 2nd International Symposium on Algorithmic Game Theory (SAGT)*. 299–310.
[29] Jeffrey Ely, Alexander Frankel, and Emir Kamenica. 2015. Suspense and surprise. *Journal of Political Economy* 123, 1 (2015), 215–260.
[30] Carl-Gustaf Esseen. 1942. *On the Liapounoff limit of error in the theory of probability*. Almqvist & Wiksell.
[31] P. Faliszewski, E. Hemaspaandra, , and L. A. Hemaspaandra. 2006. The Complexity of Bribery in Elections. In *Proceedings of the 21st AAAI Conference on Artificial Intelligence (AAAI)*. 641–646.
[32] Piotr Faliszewski, Edith Hemaspaandra, and Lane A Hemaspaandra. 2014. The complexity of manipulative attacks in nearly single-peaked electorates. *Artificial Intelligence* 207 (2014), 69–99.
[33] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2016. The rise of social bots. *Commun. ACM* 59, 7 (2016), 96–104.
[34] Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. 1983. Stochastic blockmodels: First steps. *Social networks* 5, 2 (1983), 109–137.
[35] The Independent. 2016. President Donald Trump: The most shocking results night in US history âĂŞ in pictures. http://www.independent.co.uk/news/world/americas/us-elections/donal-trump-us-president-hilary-clinton-election-2016-esults-night-in-pictures-a7407156.html. [Online; accessed 3-September-2017].
[36] Brian Karrer and Mark EJ Newman. 2011. Stochastic blockmodels and community structure in networks. *Physical Review E* 83, 1 (2011), 016107.
[37] Nicholas Mattei, Maria Silvia Pini, Francesca Rossi, and K Brent Venable. 2013. Bribery in voting with CP-nets. *Annals of Mathematics and Artificial Intelligence* 68, 1-3 (2013), 135–160.
[38] Panagiotis T Metaxas and Eni Mustafaraj. 2012. Social media and the elections. *Science* 338, 6106 (2012), 472–473.
[39] BBC News. 2016. Brexit: Europe stunned by UK Leave vote. http://www.bbc.com/news/uk-politics-eu-referendum-36616018. [Online; accessed 3-September-2017].
[40] CNBC News. 2017. Election shock? Why you now need to sit up and take note of UK politics. https://www.cnbc.com/2017/05/31/election-shock-why-you-now-need-to-sit-up-and-take-note-of-uk-politics.html. [Online; accessed 3-September-2017].
[41] Vox News. 2017. The shocking UK election results, explained. https://www.vox.com/world/2017/6/9/15767522/uk-election-results-hung-parliament. [Online; accessed 3-September-2017].
[42] Jarrod Olson and Daniel F Stone. 2014. Suspense-optimal college football play-offs. *Journal of Sports Economics* 15, 5 (2014), 519–540.
[43] David C Parkes and Lirong Xia. 2012. A Complexity-of-Strategic-Behavior Comparison between Schulze's Rule and Ranked Pairs.. In *AAAI*.
[44] Prasanta K Pattanaik. 1973. On the stability of sincere voting situations. *Journal of Economic Theory* 6, 6 (1973), 558–574.
[45] Ariel Rubinstein. 1980. Stability of decision systems under majority rule. *Journal of Economic Theory* 23, 2 (1980), 150–159.
[46] Harald Schoen, Daniel Gayo-Avello, Panagiotis Takis Metaxas, Eni Mustafaraj, Markus Strohmaier, and Peter Gloor. 2013. The power of prediction with social media. *Internet Research* 23, 5 (2013), 528–543.
[47] Terence Tao. 2010. 254A, Notes 2: The central limit theorem. https://terrytao.wordpress.com/2010/01/05/254a-notes-2-the-central-limit-theorem/#more-3281. [Online; accessed 27-August-2017].
[48] Stanley Wasserman and Katherine Faust. 1994. *Social network analysis: Methods and applications*. Vol. 8. Cambridge university press.
[49] Lirong Xia. 2012. Computing the margin of victory for various voting rules. In *Proceedings of the 13th ACM Conference on Electronic Commerce*. ACM, 982–999.