ABSTRACT

Title of dissertation: THE PHYSICS OF IDEAS: INFERRING THE MECHANICS OF OPINION FORMATION FROM MACROSCOPIC STATISTICAL PATTERNS

Keith Burghardt, Doctor of Philosophy, 2016

Dissertation directed by: Professor Michelle Girvan
Department of Physics

In a microscopic setting, humans behave in rich and unexpected ways. In a macroscopic setting, however, distinctive patterns of group behavior emerge, leading statistical physicists to search for an underlying mechanism. The aim of this dissertation is to analyze the macroscopic patterns of competing ideas in order to discern the mechanics of how group opinions form at the microscopic level.

First, we explore the competition of answers in online Q&A (question and answer) boards. We find that a simple individual-level model can capture important features of user behavior, especially as the number of answers to a question grows. Our model further suggests that the wisdom of crowds may be constrained by information overload, in which users are unable to thoroughly evaluate each answer and therefore tend to use heuristics to pick what they believe is the best answer.

Next, we explore models of opinion spread among voters to explain observed universal statistical patterns such as rescaled vote distributions and logarithmic vote correlations. We introduce a simple model that can explain both properties, as well

as why it takes so long for large groups to reach consensus. An important feature of the model that facilitates agreement with data is that individuals become more stubborn (unwilling to change their opinion) over time.

Finally, we explore potential underlying mechanisms for opinion formation in juries, by comparing data to various types of models. We find that different null hypotheses in which jurors do not interact when reaching a decision are in strong disagreement with data compared to a simple interaction model. These findings provide conceptual and mechanistic support for previous work that has found mutual influence can play a large role in group decisions. In addition, by matching our models to data, we are able to infer the time scales over which individuals change their opinions for different jury contexts. We find that these values increase as a function of the trial time, suggesting that jurors and judicial panels exhibit a kind of stubbornness similar to what we include in our model of voting behavior.

THE PHYSICS OF IDEAS: INFERRING THE MECHANICS OF OPINION FORMATION FROM MACROSCOPIC STATISTICAL PATTERNS

by

Keith Burghardt

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2016

Advisory Committee:
Professor Michelle Girvan, Chair/Advisor
Professor Edward Ott
Professor Subramanian Raghu Raghavan
Professor William Rand, Co-Advisor
Professor Rajarshi Roy

Dedication

For my friends and family.

Acknowledgments

I want to offer my gratitude to everyone who made this work possible.

First, I must thank Michelle Girvan and William Rand, who have been excellent mentors, having guided my research and encouraged me through both research challenges and successes. I have to single Michelle out for her advice on everything from data visuals to job offers, which has aided in both my research and in my future endevors.

Next, I would like to thank my co-workers, such as David Darmon, Emanuel F. Alsina, and collaborators, such as Kristina Lerman, among many others whom I have had the pleasure of working with and talking to. It is from their expertise that I have gained more insight into both my own research, and research far outside of my field.

Third, I would like to thank Kevin Cornell for inspiring me to become a physicist, with his excellent teaching and charisma, and S. James Gates, whose research inspired me to go to the University of Maryland.

Fourth, I must thank staff at the university, such as Paulina Alajandro, Edward Condon, Jessica Crosby, and Tom Gleason, among many others, who have helped in more ways than I could mention.

Finally, I must thank my parents, John Burghardt and Elaine Byergo, and anyone whom I am may have missed, for their support and encouragment.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Chapter 2

β	Regression coefficient (n.b. this definition is different for Chapter 3)
CDF	Cumulative Distribution Function
CV	Cross Validated
D	Deviance
D_{fit}	Deviance of fitted model
D_{null}	Null deviance
LASSO	Least Absolute Shrinkage and Selection Operator. L^1-norm regression
p	Statistical probability
Q&A board	Question answering board

Chapter 3 & Appendix A

A_{xy}	x, y coordinate of an adjacency matrix
α	Scale-free degree distribution coefficient ($p(k) \sim k^{-\alpha}$)
β	Persuasiveness (n.b. this definition is different for Chapter 2)
$C(r)$	Correlation as a function of Euclidean distance
CCIS	Competing contagion with individual stubbornness model
D	Diffusion coefficient
d	Dimension
D_k	Diffusion coefficient for degree k
δ	Recovery rate
$\delta\rho_k$	Rate at which all nodes with degree k change their state
ΔP	Difference between the fraction of individuals with one of two opinions
Δt	Timestep size
f	Fraction of edges rewired in a model
FC	Fortunato and Castellano model [2]
FPA	Fokker-Planck Approximation
$I^{(A)}(t)$	$\sum_{\tau'} \mu\tau' \rho^{(A)}(t, \tau')$
IP	Invasion process [10]
k	Degree of a node
k_{min}	minimum degree in a network
$\langle k^a \rangle$	Mean of a network's a^{th} degree moment
L_k	Lowering operator
$\mathcal{L}$	Likelihood function
λ_1	The largest eigenvalue
M	Degree distribution-depenent factor for consensus to be reached A.34
MLE	Maximum Likelihood Estimation
μ	Stubbornness rate
N	Number of nodes (individuals)
N_{eff}	Effective number of nodes (individuals) 3.4
n_i	Number of opinions in each state
N_{op}	Number of opinionated individuals
$\eta(x)$	State of node x
$\eta_x(y)$	State of system after changing a node x
$P(0)$	Initial fraction of individuals seeded with an opinion
$P^{(A)}(t)$	Fraction of individuals with opinion A as a funtion of t
PTP	Pre-trial publicity
Q	Number of opinions
r	Euclidean distance
R_k	Raising operator
ρ	$\frac{1}{N}\sum_x \eta(x)$
$\rho^{(A)}$	Opinion A density as a function of t and τ
ρ_k	$\frac{1}{N}\sum_{x'} \eta(x')$, where x' are all nodes with degree k
S	β^{-1}
s_i	State of each node
SIS	Susceptible-Infectious-Susceptible epidemic model
t	Time
T_{cons}	Consensus time
T_{eq}	Equilibrium time
TLA	Transportation–Like Approximation
τ	Time the most recent opinion has been kept

Chapter 4 & Appendix B

α_{25}	Rate for juries to stop deliberating when verdict is not guilty and $> 75\%$ are in the majority
α_{75}	Rate for juries to stop deliberating when verdict is guilty and $> 75\%$ are in the majority
α_G	Rate for judges to stop deliberating when verdict is guilty
α_I	Rate for judges to stop deliberating when verdict is not guilty
BMM	Block maxima model
CA #	San Fransisco County, California jury dataset with # jurors per case [11]
ECHR	European Court of Human Rights
KS	Kolmogorov-Smirnov test
MVM	Majority voter model [6]
NE	Douglas County, Nebraska jury dataset [12, 13]
OR #	Multnomah County, Orgeon jury dataset with # jurors per case [14]
T_{delib}	Deliberation time
TT Null	Two-timescale block maxima null model
T_{trial}	Trial time
τ	Mean time for an individual to change their opinion (n.b. this definition is different for Chapter 3)
ρ_g	Fraction of individuals whose final vote is guilty in a case
VM	Voter model [7, 8]
WA	Thurston County, Washing jury dataset [12, 13]

Chapter 1: Introduction

The opinions of groups are important to decision making, e.g., in elections, trials by jury, or even crowdsourcing platforms. Therefore opinion dynamics, the study of how opinions form in groups, may help explain and predict dynamics at the national level, such as who becomes president, as well as at the very local level, such as whether a defendant will be found guilty or not. In a different context, opinion dynamics can help us understand how crowds determine the quality of answers in online question and answer boards, such as in Stack Exchange. These boards allow users to choose the best answers to questions about everything from programming in C to learning a natural language, and with an aggregation of experts that helps to make a decision on the best answer, one will presumably find a very good answer quickly. All of these systems feature group-level decisions. Groups, especially large ones, have been known to find answers more accurately than individuals [15], therefore we might expect democracies to find better leaders than any political expert, or crowds to find the best answers, in aggregate. Social influence, however, is known to have a large effect on crowd behavior [16–19], suggesting that these groups may not arrive at an optimal decision. Several factors besides the the the quality of competing ideas can heavily affect which idea is chosen by a group. Modeling how

ideas form is therefore critical to understanding influence and exploring the complex effects it can have on groups.

Why should physics be used to understand these dynamics? Early on, opinion dynamics was noted by some physicists as being "out of the realm of physics" [20]. Even today, with many publications in the field, one may question whether a simple model can adequately capture why opinions form or ideas spread, because each person may decide to follow an idea by weighing options based on their unique past experiences.

To answer these critiques, we can take a step back and examine what aggregate statistics on opinion dynamics suggest. For example, the correlations between the percentage of individuals who vote for a Democratic or Republican presidential candidate [21], or the correlations among individuals who turnout for elections [3] falls as approximately the logarithm of the distance between individuals. In comparison, the population correlation exhibits a significantly different form, at least in the US [22]. We also see surprisingly universal distributions in the vote-share a candidate receives [1, 2] and a similar scaling collapse of time distributions for juries and judicial panels to reach decisions (Chap. 4). Our goal in this dissertation to analyze the macroscopic patterns of competing ideas in order to infer the mechanics of group opinion formation at the microscopic level.

There is a complication when modeling opinion dynamics, however. Previous work strongly questioned whether homophily, where individuals with similar opinions become friends, can be distinguished from influence, where individuals adopt an opinion from their neighbors [23]. We are therefore motivated to find out how in-

fluence can be quantified, or if influence can ever be distinguished from independent decision making. Chapter 4 explores signatures of mutual influence among individuals in group decision dynamics. We find that the distribution of deliberation times and final votes provides evidence that influence may play a role in the dynamics, therefore we have taken some initial steps to distinguish independent decision making from influence and quantify this effect.

With these points in mind, we explore opinion dynamics in three ways. First, we model how individuals in aggregate choose a higher quality answer over lower quality ones across all question answering boards in Stack Exchange (Chap. 2). (This research has been submitted to KDD and posted online as a pre-print [24].) Next, we create a model to reproduce long range vote correlations and vote share distributions (Chap. 3). (This research has been published in Physical Review E [25].) In Chapter 4, we use survival analysis to help distinguish which model can best capture the dynamics of juries and judicial panels. (This research, in preparation, is in collaboration with Michelle Girvan and William Rand). Further evidence supporting the modeling assumptions made in Chap. 3 are found, as well as statistics that suggest juries exhibit unusually complex dynamics. Finally, we summarize what we have found in the concluding chapter (Chap. 5).

Chapter 2: The Myopia of Crowds: A Study of Collective Evaluation on Stack Exchange

2.1 Introduction

Are crowds wiser than informed individuals? Generally speaking, a crowd's collective opinion—whether through votes, likes, or thumbs up/down—is often used to rank order items in crowdsourcing systems, which determines how much attention they receive [26], as well as users' incentives for participating [27]. The assumption is that collective opinions outperform individual experts, an observation long seen in a variety of contexts [28–30], even when they are less-informed than the experts. Recent evidence, however, has shown that the collective decision of the crowd is not foolproof. One known limitation, for example, is social influence, which biases individual judgments and degrades crowd performance [31], obscuring the underlying quality of choices [32]. We try to answer whether crowd wisdom limitations affect a common crowdsourcing application, question answering boards.

We carry out an empirical study of Stack Exchange[1], a network of more than a hundred question answering (Q&A) communities, where millions of people post

[1] http://stackexchange.com

questions on a variety of topics, and others answer them asynchronously. Like other Q&A sites, such as Quora and Yahoo! Answers, Stack Exchange has a number of features for enhancing collaborative knowledge creation. In addition to asking and answering questions, users can evaluate answers by (1) *voting* for them, and (2) askers can *accept* a specific answer to their question. The votes, in aggregate, reflect the crowd's opinion about the quality of content, and are used by Stack Exchange to surface the right answers. They also provide a lasting value to the community [33], enabling future users to identify the most helpful answers to questions without asking the questions themselves.

We find that the number of answers users parse through can dramatically affect how users choose answers, including a greater reliance on heuristic-like answer attributes, potentially limiting the usefulness of question answering boards. In addition, we find behavior biases allow for users to choose answers in an increasingly predictable way, as the number of answers increases, running counter to our intuition that increasing the numbers of choices makes user decisions less predictable.

Alternatively, work also addresses some of the challenges of data heterogeneity. Large-scale datasets of human behavior, such as this one, provide new opportunities to study decision-making processes in crowdsourcing systems. In contrast to laboratory studies, which typically involve dozens of subjects, behavioral data are collected from millions of people under real-world conditions. Mining observational behavioral data, however, presents significant computational and analytic challenges. Human behavior is noisy and highly heterogeneous: aggregating data to improve the signal-to-noise ratio may obscure underlying patterns in heterogeneous data and even lead

to nonsensical conclusions about human behavior [34]. We discover that splitting data by the number of answers addresses one of the larger sources of user heterogeity, potentially providing greater predictive power in future models.

We use penalized regression to uncover factors associated with users' decisions to vote for or accept answers on all Stack Exchange communities. To partly control for heterogeneity, we split data by community type (technical, non-technical, meta) and leave out the largest community to check the robustness of results. In all cases, behavior was qualitatively the same and quantitatively similar. We find that the a significant source of behavioral heterogeneity is the number of existing answers to questions. To account for this, we separate data according to the number of answers questions have at the time that a user makes a decision about which answer to vote for (or accept).

We find that a few answer attributes are important in our regressions, including the order in which the answer appears, its share of words compared to the other available answers to the question, and whether it was accepted by the asker. This appears to imply that users rely on simple heuristics to choose an answer based on its rank, how much screen space it occupies, or whether it was approved by others. These heuristics may be useful proxies for answer quality, but our work suggests otherwise. For example, voters are more likely to choose an accepted answer after it has been accepted than before. Although answer acceptance is often viewed as a standard of answer quality [35–37], the only discernable difference in an answer after acceptance is a signal that the asker chose this answer, suggesting users view acceptance as a useful signal about quality, but are less able to discern that quality

on their own.

We also find that heuristics better explain user behavior as the number of available answers to a question grows. Two different explanations are feasible. First, as the number of answers to a question grows, users may become less willing to thoroughly evaluate all answers, instead increasingly relying on cognitive heuristics when choosing an answer. A similar effect exists in other domains. For instance, information overload impacts consumer's choice of products [38] and the spread information in online social networks [39, 40]. An alternative explanation is that later voters are different and happen to rely more on cognitive heuristics compared to people who vote early. This view is potentially supported by the observation that users who answer early in a question's life cycle on Stack Overflow, a programming-related community on Stack Exchange, have higher reputation than users who answer later [33]; therefore, time acts as a potential source of heterogeneity. In either case, the finding that voters rely more on heuristics as the number of answers grows points to a limitation of the "wisdom of crowds" effect on Stack Exchange: crowd's judgments become less reliable as proxies of quality as questions accumulate answers.

The rest of the chapter is as follows. In the related work section, we review work related to our current analysis, while, in the materials and methods section, we discuss our data and ways in which we analyze it. Next, in the results section, we discuss our main findings. Finally, in the conclusion section, we review our findings, discuss future work, and discuss ways to improve upon question answering sites.

2.2 Related Work

Prior research on Q&A sites has shown that a variety of attributes can provide useful insights into content quality [37, 41–43]. For example, Kim and Oh [36] examined how users evaluate information in Yahoo! Answers forums, by examining the comments askers leave on answers. They found socioemotional-, content-, and utility-related criteria are dominant in the choice of the best answer, and found users evaluate information based not only upon the content, but also on cognitive and collaborative aspects. Adamic et al. [44] conducted a large-scale network analysis of Yahoo! Answers, trying to predict which answers would be judged best and found that, for both technical and non-technical sites, answer length and the number of other answers the asker has to choose from are the most significant features to predict the future best answer. A preference for longer answers, however, diminishes with the number of answers. One limitation in these previous studies, however, is in assuming that the answer an asker chose was the "best" answer, and did not correct for asker biases when choosing any answer.

Several authors [35–37] used logistic regression to determine which attributes best describe high quality answers, although, again, it is often assumed that a"high quality" answer is one an asker accepts, a conclusion that our work casts doubt on. Other works have examined the impact of answer order on answer quality. Anderson et al. [33] found that early answers in Stack Overflow (the Stack Exchange community that deals with programming questions) tend to be posted by expert users with higher reputation, and subsequent answers come from lower reputation users. While

the first answer tends to be more appreciated by the asker, the longer a question goes unanswered, the less likely that an answer will eventually be accepted. Similarly, Rechavi and Rafaeli [45] concluded that askers use response time as a parameter at evaluation time. However, this hypothesis was refuted in other works. Shah [46] analyzed the responsiveness in Yahoo! Answers forums, finding that more than 90% of the questions receives an answer within an hour. However, satisfactory answers may take longer, depending on the difficulty of the questions. Interestingly, our work, discussed in the Results section, suggests that answer age and chronological order are not particularly important attributes for askers or voters. In part this is because high reputations answerers do not strongly affect whether an answer gets voted on (not shown). Older answers, however will accumulate more votes and will therefore be more likely to be voted on , but the main driver appears to by answer attributes not directly dependent on time.

Unlike previous studies, we examine how voting may be affected by various answer attributes. This is an an important area to study, because people often use votes as a signal of the best answer to a particular problem. One previous study that also attempted to tackle this problem deduced a set of possible factors that indicate bias in user voting behavior [47]. They provided a method to calibrate the votes inside Q&A sites, principally based on the average value of the answer and the average vote value received in the answerer history. This type of calibration is useful to restrict the effects of users who are trying to game the system, or to signal the reputation of answerers. Our work, however, answers a different set of questions: we want to find the role heuristics play in answer evaluation, how voter

and asker behaviors differ, and what drives heterogeneity within voter and asker populations. The role of heuristics in human decisions has been studied by behavioral economics [48, 49], but, to the best of our knowledge, our work is the first that investigates the potential impact of heuristics on the performance of crowdsourcing systems.

2.3 Data and Methods

Stack Exchange launched in 2008 with Stack Overflow, its first Q&A community for computer programming questions. Over time, Stack Exchange has added more communities covering diverse topics:

- 49 *Technical* communities on topics, such as Programming, Server Faults, Information Security, Apple, Android and Ubuntu;
- 33 *Culture and recreation* communities, e.g., English Language Learners, Bicycles, Videogamers Platforms, and Anime & Manga;
- 17 *Life and Arts* communities on topics related to the everyday life: e.g., Cooking, Photography, DIYers, and Movies & TV;
- 16 *Science* communities, e.g., Mathematics, Statistics, Biology, and Philosophy;
- 4 *Business* communities on topics, such as Bitcoin, Project Management, and Finance;

There is a *meta* board for each community where users discuss the workings and policies of the community: e.g., in Meta Stack Overflow users discuss the policies of Stack Overflow rather than computer programming itself. Posts that are overly subjective, argumentative, or likely to generate discussion rather than answers, are removed from the website.

A user can post a question, which may receive multiple answers from different people, as shown in Fig. 2.1. The asker can *accept* an answer, which generally signifies that the asker finds it helpful. Regardless of acceptance, others can *vote* an answer up (or down) if they think that it provides helpful (or irrelevant) information. By upvoting more helpful answers, a community collectively curates the information for both askers and future users interested in the same topic. The difference between the up and down votes is the *score* of the answer. Answers with higher scores are shown at the top of the list of answers to the question, so that they are easier to find (answers with the same score are shown in random order). Figure 2.1 shows an example question with answers, score for both answers and question, the time the answer was submitted, answerer's reputation, and whether the answer was accepted by the asker.

For our study we used anonymized data consisting of all user contributions to Stack Exchange from 2009 until September 2014[2]. The data contain information about five million posts (questions and answers) and 23 million votes. In particular, we used the data of 250 communities, including information related to the posts: the ID of the post, creation date, type of post (question or answer), ID of the relative

[2]https://archive.org/details/stackexchange

Figure 2.1: A screenshot of a Stack Exchange web page, showing a question (at top) and answers listed below in default order. The score next to the answer (red box), is defined by upvotes minus downvotes, and the green checkmark (blue box) denotes that the answer was accepted by the asker. We also consider other factors, including the times the question was asked (green box) and the answer was provided, as well as the answerer's reputation (purple box).

question (in case of answers), the ID of the eventually accepted answer (in case of questions), and the content of the post; and related to the history of the votes made on each single post: the type of vote (up, down votes, or acceptance), the ID of the related post, and the time of assignment. In addition, we considered the information related to the users, such as the ID of the user, creation date (date of the sign up), and reputation. Particularly, we calculated the reputation of the users at the moment they asked or answered a question, considering the rules of Stack Exchange[3].

Each question on these communities received almost three answers, on average. The "Programming Puzzles & Code Golf" community had the highest number of average answers per question at 8, while the "Magento" site had an average of only 1. About 10% of the questions went unanswered, and 42% received only one answer. Only the questions that received two or more answers were included in our study.Figure 2.2 shows the complementary cumulative distribution of the number of answers posted for each question on technical, non-technical, and meta communities. We considered an average of 11k answers per community, although this varied significantly. Technical communities had on average twice as many answers as non-technical communities, and an order of magnitude more than meta communities, which were not broad in appeal. Askers accepted an answer 59% of the time in technical communities, and 59% in non-technical communities, but, curiously, only 37% of the questions in meta boards were similarly accepted. Answers with votes consist of 86% (std 7%) of the total, but this too varied across site types. 78% of the

[3]http://meta.stackexchange.com/how-does-reputation-work

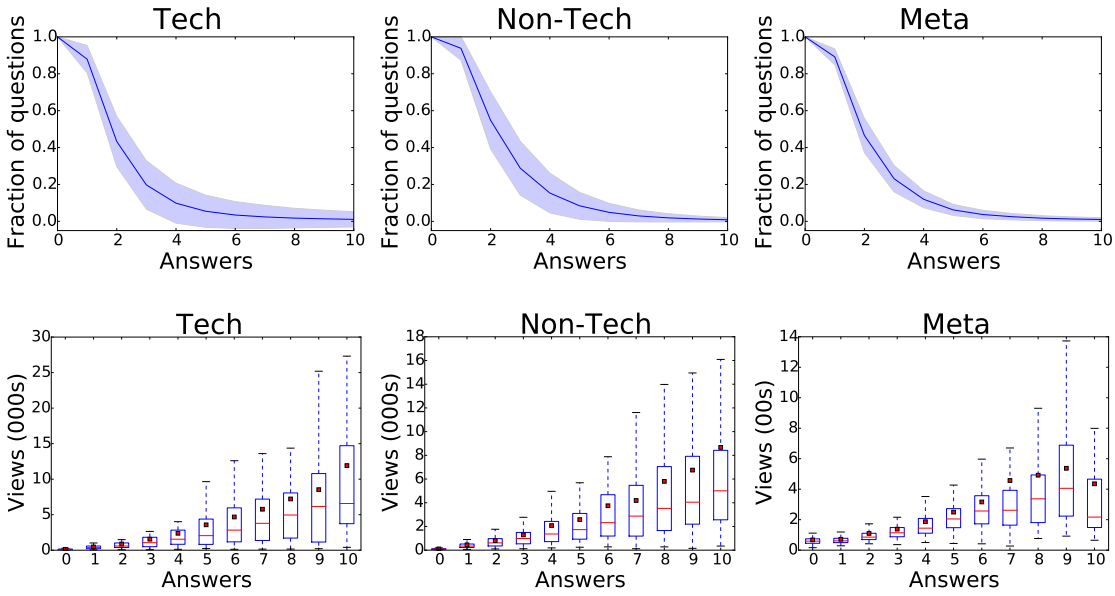

Figure 2.2: (Top row) Complementary cumulative distribution of the final number of answers posted in reply to a question as of August, 2014, on (a) technical, (b) non-technical, and (c) meta sites. Shaded areas correspond to the standard deviation in the distributions. (Bottom row) Number of views per question (in August, 2014) as a function of the number of answers on (d) technical, (e) non-technical, and (f) meta sites. Boxes indicate 50% confidence intervals, with a red line to indicate the median view count, and a red dot to represent the mean viewcount.

answers on technical sites have votes, versus 86% in non-technical and 88% in meta communities (all differences are statistically significant with $p < 10^{-3}$ using t-tests). The median time to obtain the first answer was 2.77 hours, and the eventually accepted answer, 4.47 hours. As the community matures, the questions become more complex, which attracts the attention of users who may focus on different facets of the problem, posting multiple good answers for the same question.

2.3.1 Logistic Regression

We use logistic regression to understand which factors drive user actions on Stack Exchange. Because our data is highly multi-dimensional, and some attributes are strongly correlated with others, we use LASSO penalized regression, where parameters are determined by maximizing the likelihood function with the addition of a penalty to avoid overfitting [50]. The value of this penalty was adjusted such that the deviance from 10-fold cross-validation (CV) was minimzed. As a check, we did the same fits with a different type of penalty, ridge regression, and found the behavior to be qualitatively the same. The fitting was performed with the R package "glmnet" [51], which allows for fast and accurate determination of regression coefficients, β.

We checked the robustness of our results by omitting data from the largest community for each board type (meta, non-technical, and technical), and re-determining the regression parameters. The qualitative results were unaffected, and quantitatively, the results were very similar. For the rest of the chapter, we focus on LASSO penalized regressions with all boards included.

2.3.2 Deviance Ratio

We use deviance ratio to determine how well the model fits the data. The deviance ratio is reminescent of R^2, although it is used for models that maximize the likelihood function rather than minimize the mean squared error.

The deviance ratio is defined as:

$$R_{dev} = 1 - \frac{D_{fit}}{D_{null}}, \tag{2.1}$$

where

$$D = -2\left\{\log[p(y|\hat{\theta}_0)] - \log[p(y|\hat{\theta}_s)]\right\} \tag{2.2}$$

In this case, $\log[p(y|\hat{\theta}_s)]$ is the log-likelihood of the saturated model, with one degree of freedom per observation, while $\log[p(y|\hat{\theta}_0)]$ is the log-likelihood for the fitted model. D_{fit} is for the best fit model, while D_{null} is the null intercept model. A careful observation reveals D is simply $-2\times$ (the log likelihood ratio), therefore the deviance ratio tells us how much of the likelihood ratio for the null model can be explained with a fitted model. Errors for this value are defined in the next section.

2.3.3 Error

The uncertainty in β and the deviance ratio (shaded regions in subsequent figures) is defined as the range of values such that, by changing the LASSO regression bias, the mean 10-fold CV error (in this case, the deviance) is within one standard deviation of the minimum mean CV error. This spread of values is the clearest way we are aware of to show parameter uncertainty or sensitivity, because LASSO regression, like all penalized regression methods, does not have a standard method to caluculate uncertainties with high dimensional data [52].

2.3.4 Attributes and Normalization

We use the following answer attributes in analysis:

1. answerer's *reputation* at the time the answer was created,
2. mean rate of *reputation increase* over time,
3. answer's Flesch Reading Ease [53], or *readability*, score,
4. answerer's *tenure* (i.e., time since joining the site) at the time of the answer,
5. number of *hyperlinks* per answer,
6. binary value denoting whether the *answer was eventually accepted* (for voting only),
7. answer *score* before each vote,
8. default *web page order* for an answer (i.e., its relative position),
9. *chronological order* of an answer (whether it was first, second, third, etc.),
10. *time* since an answer was created, or its age
11. *number of words* per answer,
12. answer's *word share*, that is the fraction of total words in all answers to the question.

Answerer reputation [35], Flesch readability, and word count [54] were used in previous works as measures of answer quality, and often a "high quality" answer was at least in part defined as the accepted answer [35–37]. To adeqately compare datasets, we removed all data where the question was not eventually accepted within the collection timeframe. Qualitatively, voters and askers in unaccepted questions had similar behavior to those in accepted questions, but quantitatively, we found variations on regression coefficients, potentially suggesting voters behave differently in this hold-out set. We also consider an answer's rank in the list of answers (what

we refer to as web page order) and score, because these variables affect how much attention the answer receives [18, 32, 55]. The other attributes were also examined as additional factors that could affect how answers are voted or accepted. These were, however, not found to significantly affect the results.

There is large variability in attribute values within and across the attributes. To account for the variability, we *normalize* all attributes by mapping them to their associated cumulative distribution function (CDF). CDF normalization is non-parametric and accounts for the distribution of attribute values. An advantage of this normalization is that outliers have a minimal effect because values are evenly spread and bounded between 0 and 1. Normalization allows us to compare the relative importance of different attributes by comparing their regression coefficients. For web page order attribute, we divided by the number of answers available, which is equivalent to a CDF for the number of answers equal to 2, 3, etc., while for all other attributes, we used the CDF across all answers on all Stack Exchange communities.

To verify the selected attributes, we checked each attribute to make sure correlations with other attributes were reasonably low, and, if they were greater than 0.7, we checked whether removal of the attribute increased the CV error significantly. This correlation condition seems very liberal, but we wanted to include as many attributes used in previous literature as possible, and then use penalized regression to appropriately reduce the effect of colinearity. To check if this affected our results, we separately removed the most important attributes: webpage order, wordshare, and whether the answer was eventially accepted. Broadly, we found the results

were qualitatively the same, although when webpage order was removed, the score became a more important attribute.

2.4 Results

We analyze Stack Exchange data to understand what attributes are strongly associated with the decision to vote for, or accept, an answer. To do this, we find all attribute values just before an answer was voted for (or accepted), and then estimate attribute coefficients for a logistic regression model.

2.4.1 Taming Heterogeneity

Automatically uncovering homogeneous populations within heterogeneous observational data remains an open research challenge. In our study of Stack Exchange, we used exploratory data analysis to identify potential sources of heterogeneity. For example, users who are interested in technical topics (e.g., programming) may be driven by different factors to contribute to Stack Exchange than those who are interested in non-technical subjects (e.g., cooking), or governance (meta boards). To account for this source of heterogeneity, we split the data by the type of board—technical, non-technical and meta—and run regression analysis separately on each dataset. We further split data by whether the asker eventually accepted an answer in our observation window, how an answer is chosen (vote versus accept), and the number of answers, but find that the greatest source of heterogeneity is the number of answers a question has at the time the user votes for or accepts it.

In the next two sections, we discuss our findings in greater detail, including the implications of the most important attributes, and the reasons for the heterogeneity in our data. Regression fits suggest users who vote when many answers are visible strongly depend on a small set of heutistic-like attributes compared to users who vote when there are few. Furthermore, askers are found to depend on heurstics much more than voters, which undermines the assumption that accepted answers are probably one of the best answers [35–37]. Overall, we find evidence that the wisdom of crowds in Stack Exchange boards could be reduced by the number of answers to a question, and the role of the user.

2.4.2 Answer Attributes and Behavior

We take logistic regressions for votes cast before any answer was accepted, votes after an answer was accepted, as well as accepted answers. The average and variance of the regression parameters across $2-20$ answers are shown in Figure 2.3. Because all attributes were normalized, the larger the value, the more the respective parameter affects user behavior, relative to all others in the regression.

We find that web page order and word share are the two most important factors for users to choose an answer (Fig. 2.3). Because strong correlations may affect the coefficient of a particular attribute in penalized regression, we also remove each attribute separately (not shown), and find the CV error decreases the most when the highest-coefficient attributes are removed, thus validating the usefulness of CDF normalization.

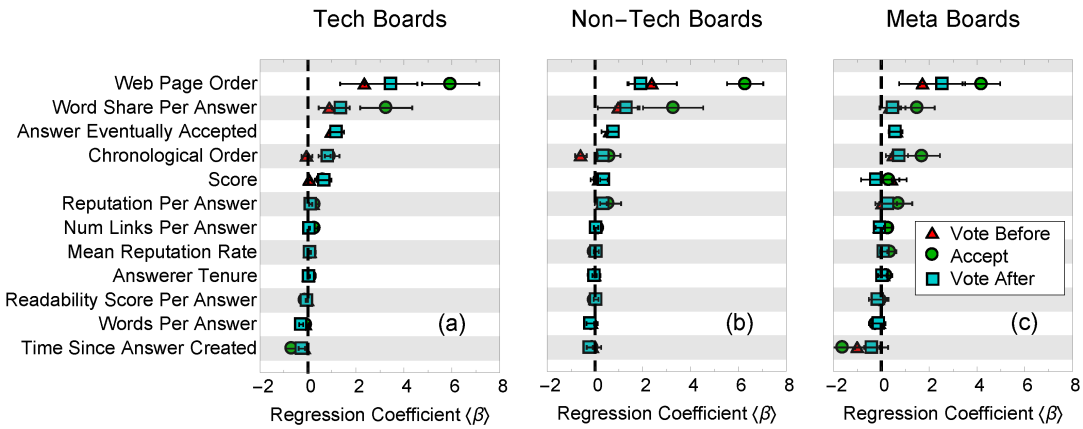

Figure 2.3: Regression coefficients for answerers to accept (green circles) and voters to vote for an answer both before (red triangles) and after (blue squares) an answer is accepted on (a) technical, (b) non-technical, and (c) meta boards, averaged over the number of available answers from 2-20. Higher values indicate a stronger relationship between attributes and user behavior (voting or accepting an answer). Error bars indicate the variance of these values as the number of answers increases.

These findings alone are not necessarily surprising. We know from previous research that people's choices are biased by the rank order of items [18, 55, 56]. Word share is potentially correlated with higher answer quality, because relatively long answers may be more informative, or they may just be easier to see (take up a large portion of the web page space). We notice that both of these regression coefficients are even higher for askers than voters, across different board types, already suggesting a surprising degree of heterogeneity. Other factors, however, such as an answerer's reputation or tenure, how thoroughly an answer is documented with hyperlinks, how easy it is to read (readability), etc., do not seem to play a big role in users' choices of which answers to vote or accept.

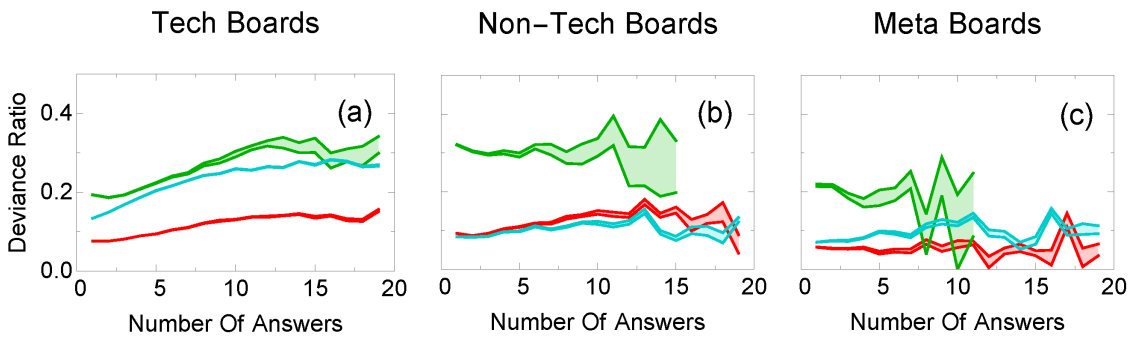

Figure 2.4: The deviance ratio (fraction of deviance explained by the model) for votes before acceptance (red), answer acceptance (green), and votes after acceptance (blue), for (a) technical, (b) non-technical, and (c) meta boards, with 2 to 20 answers. The shaded region represents the uncertainty in our values (see Section 2.3). Askers have a larger deviance ratio, and therefore appear to be better modeled by our regressions, compared to answerers. Furthermore, the deviance ratio of voters tends to increase with the number of answers, suggesting increasing agreement with our model.

2.4.3 Behavior vs Number of Answers

What is more surprising than the overall size of the regression coefficients, however, is that the largest coefficients, e.g., for web page order and word share, change substantially as the number of available answers to a question increases (Fig. 2.6 and Fig. 2.7). Furthermore, the models describe the data increasingly well (Fig. 2.4) and reduce predictive error (Fig. 2.5). In other words, users' future decisions appear to be increasingly dependent on these attributes. This is also seen when we remove each attribute and check the resulting CV error of the model

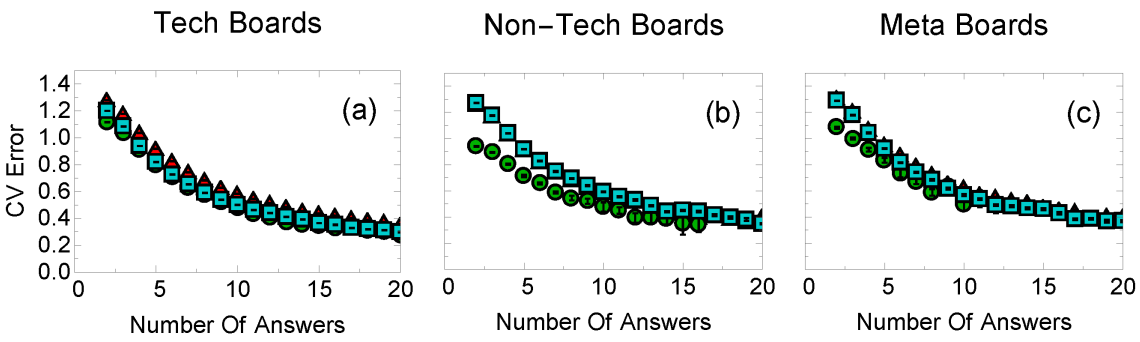

Figure 2.5: Mean deviance from 10-fold CV for votes before acceptance (red), answer acceptance (green), and votes after acceptance (blue), for (a) technical, (b) non-technical, and (c) meta boards, with 2 to 20 answers. The deviance for votes before and after acceptance almost completely overlap. Askers have a lower prediction error compared to voters, but the most significant drop in deviance for all users occurs when the number of answers increases.

(not shown). We find that removing attributes, such as whether the answer was accepted or its web page order, would increasingly impact the CV error of voters as the number of answers grow.

A number of plausible explanations exist:

- The subsequent answers improve upon the previous answer, or
- Some unknown confounding variable affects both the number of answers as well as user behavior, or finally
- User behavior changes as a function of the number of available answers.

According to the first hypothesis, the last answer may be such an improvement on the previous ones that users will "flock" to it. Therefore, it should be no surprise

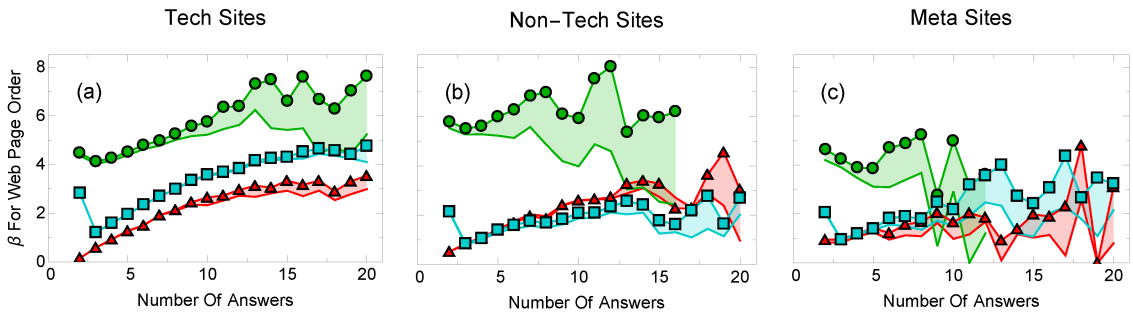

Figure 2.6: Web page order regression coefficients for voting before (red triangles) and after (blue squares) an answer is accepted, as well as accepting an answer (green circles) for (a) technical, (b) non-technical, and (c) meta boards, with 2 to 20 answers. The shaded region represents the uncertainty in our values (see Section 2.3). Users increasingly depend on the web page order of an answer as the number of answers increases.

that as the number of answers increases, changes in votes are seen. In theory, this should be captured by a significant dependence on answer's chronological order: voters should prefer newer answers to older ones. In practice, this does not seem to be the case. The dependence on chronological order is relatively small (Figure 2.3), and furthermore decreases with the number of answers (Figure 2.8), which is exactly the opposite of what should be expected if this hypothesis were true.

The second hypothesis says that the number of answers and the behavior of the user both correlate to something else entirely; the results presented so far could be strongly affected by some confounding variable. For example, [33] finds that the reputation of later answerers on Stack Overflow, a technical board within Stack Exchange devoted to programming questions, is lower than the reputation of earlier

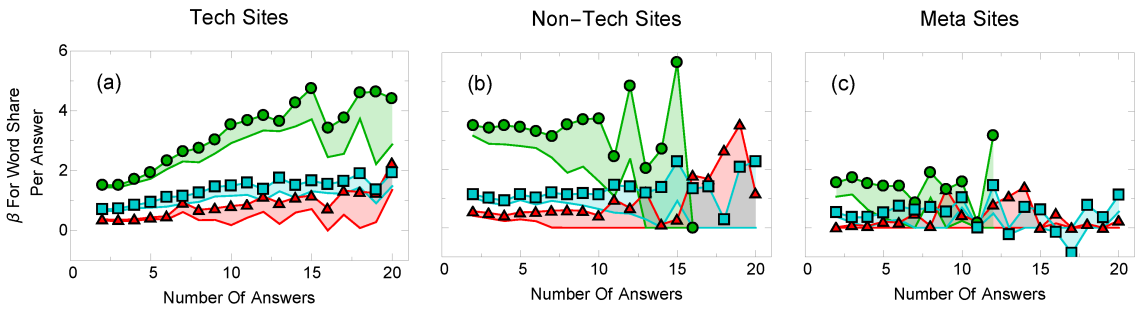

Figure 2.7: Word share regression coefficients for voting before (red triangles) and after (blue squares) an answer is accepted, as well as accepting an answer (green circles) for (a) technical, (b) non-technical, and (c) meta boards, with 2 to 20 answers. The shaded region represents the uncertainty in our values (see Section 2.3). Across all baords, voters appear increasingly likely to choose answers that take up a relatively large amount of web page space as the number of answers grows.

answerers. If later voters similarly differ in reputation or some other attribute, this could potentially explain our results. We call this the "lazy voter" hypothesis, because later voters may simply be "lazier" and rely on heuristics to a greater extent. It is curious, however, that voter behavior does not seem to be significantly affected by the age of the answer, based on our regressions, and instead on the shear number of answers, as time progresses.

The last hypothesis is that users behave differently as the number of answers grows. Economics and psychologists believe that people usually do not have the time, nor inclination or cognitive resources, to process all available information, but instead, employ heuristics to quickly decide what information is important. This phenomenon, known as "bounded rationality" [48,49], profoundly affects what

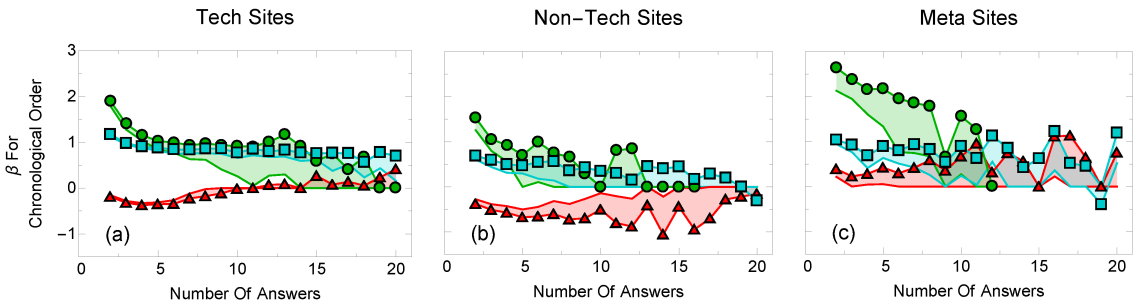

Figure 2.8: Chronological answer order regression coefficients for voting before (red triangles) and after (blue squares) an answer is accepted, as well as accepting an answer (green circles) for (a) technical, (b) non-technical, and (c) meta boards, with 2 to 20 answers. The shaded region represents the uncertainty in our values (see Section 2.3). For all boards, there is a decreasingly significant dependence on the order in which answers appear. For askers and voters after acceptance, newer answers are preferred, while, for voters before acceptance, older answers are preferred.

information people pay attention to and the decisions they make [57]. Our results suggest that rather than thoroughly evaluating all available answers to a question on Stack Exchange, users employ cognitive heuristics to choose the "best" answer. These heuristics include choosing top-ranked answer (Fig. 2.6) or one that occupies more screen space (Fig. 2.7). These heuristics become more pronounced when the volume of information (number of available answers) grows.

Instead of being a cognitive heuristic, word share could plausibly reflect answer quality: high quality answers may be wordy. Interestingly, however, the regression coefficient for the number of words for each answer (rather than its share of words)

is slightly negative, suggesting users overall prefer somewhat shorter answers if they prefer anything at all. It is intuitive that longer answers are more salient and catch a user's eye, especially when there are many answers.

Whether second or third hypothesis is true, our observation of a strong dependence of votes and accepts on the number of available answers suggests a strong limitation of crowdsourcing answer quality: collective judgment of quality may change with the number of answers, which is especially noticable with popular, and presumably important, questions which have many answers available (Fig. 2.2).

We see further evidence of the final two arguments in Figure 2.9, where we plot the regression coefficients for accepting an answer as a function of number of answers for voters before, and after, an answer is accepted.

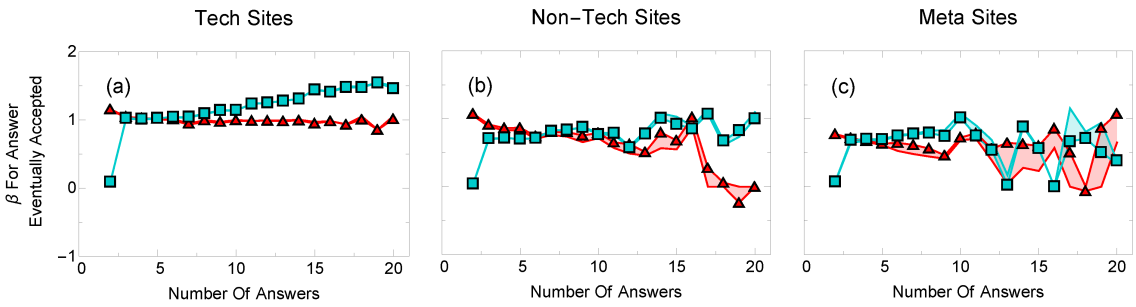

Figure 2.9: Regression coefficients for voting on an (eventually) accepted answer before (red triangles) and after (blue squares) that answer is accepted for (a) technical, (b) non-technical, and (c) meta boards, with 2 to 20 answers. The shaded region represents the uncertainty in our values (see Section 2.3). There is a large and increasing vote dependence on the accepted answer once the asker accepts it, compared to before the answer is accepted, meaning the signal that this answer is accepted appears to have a statistically significant effect on voter behavior.

We find that voters are more likely to choose an answer that is eventually accepted (the regression coefficients are positive), but, curiously, voters are even more likely to choose the answer *after* it is accepted as the number of answers increase (the regression coefficient is usually even higher, and increases with the number of answers). In other words, although answer quality does not change before or after acceptance, users are more likely to vote on whatever the asker chooses, especially as the number of answers increases. This could either be due to "lazy voters", who appear later on, when the number of answers is high, or because voters are overwhelmed by the number of answers.

Finally, askers are much better modeled by our regressions compared to voters (Fig. 2.4) and similarly, are more predictable (Fig. 2.5). To better understand what we are seeing, we must understand Stack Exchange's rules. Namely, voters need a reputation above 15 in order to vote, which becomes a barrier to entry: typically, users must have provided answers or questions in the past that others upvoted in order to be able to vote. Askers on the other hand require less reputation. Presumably they rely more on heuristics than voters, because they are less able to recognize the correct answer. This is important because accepted answers have been used as a gold standard of answer quality in previous research [35–37], but, if askers strongly rely on heuristics like answer rank order, this puts into question whether accepted answers are the best standard. Instead, we find that highly voted answers may be a better standard, because voters appear to depend less on heuristics.

2.5 Conclusion

We analyzed user activity over a five year period on 250 Q&A communities on the Stack Exchange network. The goal of our study was to understand what factors influence users to vote for, or accept, particular answers. Analysis from our models of voter and asker behavior suggest that Stack Exchange users rely on simple cognitive heuristics to choose an answer to vote for or accept, especially as the number of answers available increases. First, model parameters describing the dependence of behavior on answer's web page order and word share increase with the number of available answers. Such dependence would not necessarily exist if web page order and word share were merely proxies for answer quality. Second, askers appear to rely more on heuristics compared to voters, who need higher reputation and therefore may be more proficient Stack Exchange users. This suggests that answer acceptance might not be the best proxy for answer quality. Finally, voters are more likely to vote for an answer after it is accepted than before that very same answer is accepted as the number of available answers grow. Not only does acceptance appear to change a user's judgment of answer quality, it appears to become an increasingly strong bias with the number of answers.

The behaviors we describe are consistent with, but not proof of, bounded rationality, in which decision-makers employ cognitive heuristics to make quick decisions instead of evaluating all available information [57, 58]. Moreover, people tend to use heuristics to cope with the "cognitive strain" of information overload [59]. Psychologists and behavioral scientists have identified a wide array of cognitive heuris-

tics, which introduce predictable biases into human behavior. Social influence, *aka* "bandwagon effect", is one such heuristic: people pay attention to the choices of others [32]. We find, however, that this affect is not very significant in Stack Exchange. Another important heuristic for online activity is "position bias" [60]: people pay more attention to items at the top of the list or the screen than those below [55]. Position bias, or rank order, plays a large effect in user choices even after accounting for item quality [55,61], which is in agreement with the results presented here. Alternative explanations of our results, however, are plausible. For example "lazy" (more heuristically driven) voters might arrive later, after a question has many answers.

No matter which explanation holds, however, our work offers a cautionary note to designers of crowdsourcing systems, such as Stack Exchange: collective judgments about content quality are not necessarily accurate.To partly address this problem, the order in which answers are presented to users could be randomized, or questions could be closed to voting after some time.

Our work makes a number of methodological contributions valuable to the Data Science community. First, we use CDF normalization to make all variables commensurate. While this is a nonlinear transformation, it accounts for the distribution of variable values in the dataset, which reduces the influence of outliers and allows for fair comparison of heterogeneous variables. Also, we handled behavioral heterogeneity by splitting by board type and number of answers. To check robustness of regression results, we used two types of penalized regression and "leave out the largest board" analysis. These methods can be applied to model other heterogeneous behavioral data. Finally, we measured the uncertainty in parameter

coefficients as the range of coefficients, due to varying the penalization in our regressions, such that the CV error is within one standard deviation of the minimum mean error. We are unaware of alternative methods to accurately display the uncertainty in measurements from penalized regressions, and the uncertainties from our regressions appear reasonable.

Our analysis of observational data cannot completely control for the some of the known (and unknown) covariates that can affect our conclusions. For example, we cannot completely separate the effects of cognitive heuristics from those of answer quality. A necessary step in future research is to conduct a laboratory study to control for variation in answer quality, similar to previous studies [55, 61], to quantify the degree to which crowds are "myopic." Despite known limitations, our work highlights the benefits of using data mining to understand and predict human behaviors, and may provide insight into improving the quality and performance of crowdsourcing systems.

Chapter 3: Competing Opinions and Stubbornness: Connecting Models to Data

3.1 Introduction

The study of opinion dynamics, which has received considerable attention from statistical physicists, network scientists, and social scientists [10, 62–68], explores the dynamics of *competing* ideas or opinions via interactions between individuals. Example application areas include voting patterns [1, 2, 7, 8, 10, 66, 69–72], product competition [73], and the spread of cultural norms and religions [74–76]. The goal of our work is to gain new insights into opinion dynamics by introducing a well-motivated model that can simultaneously describe multiple empirical observations which have previously been explained by several different models.

A variety of models have been proposed to explain individual features of opinion dynamics observed in empirical data. For example, some models have focused on producing nonconsensus in equilibrium [63, 65, 77], while others can reproduce observed vote distributions [2, 70], or long-range vote correlations [21]. Because we believe these observations are all fundamentally related, we introduce a new model, called the Competing Contagions with Individual Stubbornness (CCIS)

model, which can robustly explain the above behaviors using agent-based dynamics designed to mimic observed human behaviors. Not only does the CCIS model match the aforementioned observations with consistent parameter values, it is general enough to incorporate a wide array of plausible factors affecting the success of opinions in the real world, allowing for agents with a neutral state, opinions that are stronger than others, and opinions that may be introduced after an earlier opinion has spread through a population. Here, for simplified modeling and analysis, we focus on the case of equal strength opinions introduced at the same time and leave these other cases for future work.

In the CCIS model, at any given time point, individuals can either be in a neutral state or in one of Q different opinion states. Opinions can change over time as individuals try to "convince" others in their social network to adopt their opinion. In our model, individuals exhibit "stubbornness," meaning that the longer an opinionated individual keeps his or her opinion, the less likely they are to switch to a new one. This property has been seen empirically in previous studies [78]. We distinguish this from other models in which individuals resist changes in their opinion independent of time, e.g., [66, 70–72, 79, 80]. Within the CCIS model, individuals that have held on to their opinion for a long time will eventually completely lose the ability to be convinced by one of their neighbors to adopt a different opinion. However, all opinionated individuals move back to the neutral state at a constant rate, which is designed to allow for a large fraction of "independent" voters, as is the case for the United States electorate [81]. Once an individual becomes neutral, they can switch opinions to any of their neighbors', which creates longer timescale

opinion dynamics.

The remainder of the chapter is structured as follows. We first describe related work (Section 4.2) and then provide the details of our model and algorithm implementation (Section 3.3), before comparing the results of our model to empirical data (Section 3.4). We then analyze the dynamics of our model using a series of approximations (Section 3.5) and numerically study the consensus time outside of the parameter ranges for which our analysis is valid (Section 3.6). Finally, we conclude with a discussion of future work (Section 4.6).

3.2 Related Work

In this section, we review the empirical studies that motivate the CCIS model and we discuss related models.

In recent years, large sets of empirical data have allowed researchers to better observe collective social dynamics [1, 2, 21, 82–85], leading to new insights in the field. We first focus on two themes that have received recent attention: candidate vote distributions [1, 2, 21] and spatial vote correlations [21].

Two important studies on election data from several countries demonstrate that vote distributions, when rescaled by Q/N, where Q is the number of candidates and N is the number of voters, often collapse to a universal distribution (see inset of Fig. 3.2) [1, 2]. Two recent models have been proposed to explain this behavior [2, 70].

A model by Fortunato and Castellano [2] assumes that voters are convinced to

vote for a specific candidate unique to each of Q social networks, with no interaction between voters of opposing candidates. While the model provides good agreement with vote distribution data and demonstrates how "word-of-mouth" or contagion-style spreading can play an important role in observed voting patterns, it cannot capture one important feature of real elections − that candidates seem to often compete for a common set of voters [86–88]. Hence, we believe that a model with competing opinions on a single network, such as the one introduced in this chapter, is needed to for a more complete picture of how individual level dynamics can translate to observed voting patterns.

Another model by Palombi and Toti, which does include interactions between supporters of different candidates, yields qualitative agreement with empirical data on vote distributions by assuming a network of interactions with significant structure (non-overlapping cliques connected by sparse random links) as well as a distribution of zealots (unwavering candidate supporters) that is related to the underlying clique structure of the network. By contrast, our goal is to show agreement with empirical data on both vote distributions and voter correlations using a somewhat more generic network of interactions and without imposing any connection between candidate preferences and network placement for any individuals. The contagion-inspired framework of our CCIS model, e.g., the inclusion of a neutral state and a tunable transmissability parameter, gives it the flexibility to match the two aforementioned empirical patterns of interest while simultaneously remaining relatively simple.

Recent empirical studies have shown that the spatial correlation of vote-shares

in United States elections and the spatial correlation of turn-out rates in European elections decreases as the log of the distance between two voting districts [3,21]. This contrasts to correlations of spins in many statistical mechanic spin models, which decrease as a power law or exponentially with distance [89], but is a prediction of some spin (or opinion) models, such as the VM, at an arbitrary, fixed time [7,8,90].

In addition to matching these empirical patterns by yielding spatial opinion correlations that decrease as the log of the distance between individuals (in the case of networks with significant spatial structure), the CCIS model shares other important features with the well-studied VM. In the VM, at each time step, an individual chooses to adopt the opinion of one of their randomly chosen neighbors [7,8]. In the basic CCIS model, opinions also change via interactions with neighbors, but instead of interacting with one neighbor at a time, individuals try to persuade all their neighbors simultaneously, similar to the approach used in the aforementioned Fortunato and Castellano [2] paper. In Section 3.5, we also consider CCIS type dynamics for the situation in which, as in the VM, interactions at each time step are focused on an pair of connected individuals instead of one individual and all of their neighbors.

The CCIS model also has important similarities to the well-studied Susceptible-Infected-Susceptible (SIS) model from epidemics. In the SIS model, individuals exist in only one of two states: "susceptible" and "infected," and infections propagate via contacts between infected and susceptible individuals, with infected individuals eventually recovering to the susceptible state. The SIS model can be applied to the study of opinion dynamics, but, because the basic model is an explicitly a two-

state model, it can only be used to explore how a single opinion (contagion strain) propagates through a neutral (susceptible) population, and the SIS model must be modified to explore the competition dynamics among multiple opinions.

A few recent studies have modeled the coexistence of two contagion strains on networks with SIS-like models [91–97]. Typically, in these models, individuals can only switch from one strain to another if they recover first [92, 95, 96], or else two strains can cohabit a single individual but interact on coupled networks [94]. In the CCIS model, however, individuals can switch directly between opinions instead of first moving to the recovered state, and all opinions propagate on a single network. Furthermore, no individual can have more than one opinion at any time. These are realistic assumptions for opinion dynamics, because individuals can directly switch between opinions more easily than they might directly switch between diseases, and would be unlikely to hold contrasting opinions at the same time. We note, however, that across a wide parameter space in our model, one opinion eventually dominates (e.g., Eq. 3.4 and Figs. 7, 8, & 9), while the contagion models described above have large parameter regimes where two contagions can stably coexist. In Section 3.5, we discuss in more detail how the CCIS model approaches consensus.

The CCIS model is further distinguished from the VM and SIS model by having individuals exhibit stubbornness [98] (similar model assumptions are made in other works [69,99–101]). In our definition of stubbornness, individuals increasingly resist changing their opinion, in contrast to other models where individuals resist changes in their opinion independent of time, e.g., [66, 70–72, 79, 80]. In pre-trial publicity (PTP) experiments [78], the correlation between the jury decision and the PTP

opinion is stronger when individuals are exposed to PTP more than a week before the mock trial compared to when the exposure happens closer to the start of the trial. This provides some evidence that individuals change their resistance to alternative opinions, but not necessarily monotonically with time. Further evidence from voter data is currently lacking and is an important area for future study. Nonetheless, the initial evidence from jury experiments and the strong agreement to data we find with our current model is suggestive that stubbornness may play an important role in the dynamics of opinions. We also note that stubbornness is similar to the primacy effect, well studied in psychology [102, 103], in which the first idea someone hears is favored regardless of its validity. That effect, however, deals only with the ordering of choices and does not take into account the time intervals between choices.

The CCIS model is designed to offer a more general framework than many previous models. It allows for different opinions to be more or less likely to be adopted relative to each other, for individual opinions to be more or less likely to exhibit stubbornness, for some opinions to be introduced at later times than others, and for individuals to exist in a neutral state. These additions give it the flexibility to capture a variety of situations. For simplicity, we focus on the case of opinions with equal strengths and individuals with identical stubbornness parameters.

3.3 Model Details

In this section, we describe the dynamics of the CCIS model in detail (see Fig. 3.1 for a schematic). The model operates on a network with N nodes, in which

the state of each node, i, is $s_i \in \{0, 1, 2, ..., Q\}$, where Q is the total number of opinionated states and 0 corresponds to the neutral state. For ease of analysis, we study the case in which interactions between individuals occur on a fixed, unweighted network.

At time $t = 0$, n_0 (possibly 0) nodes are in state 0, n_1 (again, possibly 0) are in state 1, etc., such that $n_0 + n_1 + ... + n_Q = N$. We leave open the possibility for new opinions to be added at arbitrary times in the simulation. However, in this paper, we focus on the case where at $t = 0$, $n_1 = n_2 = ... = n_Q$ (and therefore all opinions are simultaneously introduced).

Algorithmically, we implement the model as follows:

1. Pick a random opinionated node i (i.e, a node not in state 0)
   (a) Revert i's state to 0 with probability $\frac{\delta}{1+\delta}$
   (b) Otherwise pick each of i's neighbor at random:
      i. Convert any neutral (state 0) neighbor to state s_i with probability β
      ii. Convert any contrary opinionated neighbor j to state s_i with probability $\max\{\beta(1 - \tau_j\mu), 0\}$, where τ_j is the time since node j adopted its current opinion.
2. Count the number of opinionated individuals, $N_{op} = N - n_0$, and repeat from step 1 with time incremented by $\Delta t = N_{op}(1 + \delta)^{-1}$.

Here, for simplicity, we assume that the persuasiveness of each individual, β, the recovery rate, δ, and the stubbornness, μ, does not depend on which opinion is held,

but there may be situations for which these parameters should be differentiated according to opinion. We implement stubbornness in the following way: the effective persuadability of a node j by a neighbor with a contrary opinion, $\beta(1-\tau_j\mu)$, decreases linearly in time until $\tau_j = \mu^{-1}$, at which point individual j's opinion remains fixed unless j moves to the neutral state, which occurs at rate δ. A natural alternative to our implementation of stubbornness is to construct an effective persuadability that decreases exponentially, $\beta\exp(-\tau_i\mu)$. We choose the linear form for its simplicity, but we expect similar dynamics for the two cases.

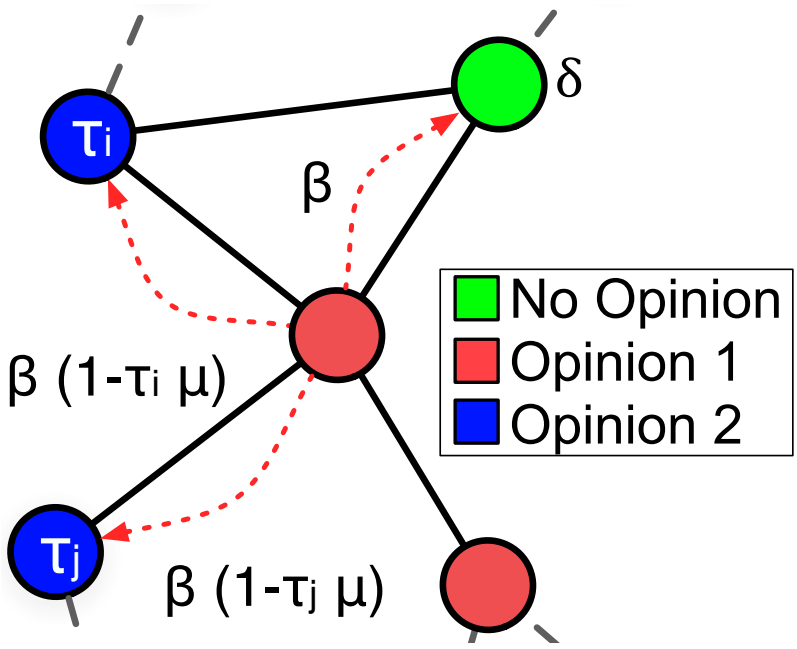

Figure 3.1: The schematic of our model. Arrows indicate attempts to convince neighboring individuals, with probabilities for success appearing next to each arrow. The length of time the nodes have held their current opinion is indicated by the text inside the node.

Note that at each time step, Δt, is normalized such that N_{op} node-node interac-

tions take place, and δN_{op} of the opinionated nodes recover, after a time $\sum_i \Delta t_i = 1$. Holding N_{op} constant for each time step, $\Delta t = ((1+\delta)N_{op})^{-1}$ and the recovery probability is $\delta/(1+\delta)$. This method is based upon a similar approached used for the SIS model to approximate continuous time dynamics [104].

We include the recovery rate in our model to allow for a large fraction of individuals to remain neutral over long time scales. This is motivated in part by the empirical observation that a significant fraction of Americans remain unaffiliated with any political party, and that this fraction is stable over the timescale of years [81], yet in individual elections, these "independents" frequently vote for candidates with party affiliations, and hence can be thought of as having adopted the party "opinion" over short timescales. Additional elements of realism, such as mass media [105], party affiliation [9], and variations in the recovery rate, have been left out of this model for simplicity, and may be important for future study.

3.4 Agreement With Data

In this section, we show that the CCIS model can reproduce two empirical observations: (1) distributions of votes received by candidates, when appropriately rescaled, follow a nearly universal function [1, 2] and (2) correlations between voters decrease only logarithmically as a function of distance [3, 21]. We find agreement between the CCIS model and both empirical observations using spatially extended networks with heavy-tailed degree distribution (a reasonable model for social networks [106, 107]). In agreement with Fortunato and Castellano [2], we find that

a heavy-tailed degree distribution is important for matching the opinion model's distribution to the empirical vote distribution data. We emphasize that the spatial component (meaning that nodes preferentially connect to others that are spatially close) is necessary to create spatial correlations that match empirical observation. The networks are created as follows: all nodes are embedded on an $\sqrt{N} \times \sqrt{N}$ two-dimensional grid with periodic boundary conditions. The out-degree, $k_i \geq k_{min}$, is chosen from from a power law degree distribution, $p(k) \sim k^{-\alpha}$ with minimum degree k_{min}, which is specified so that the desired average degree, $\langle k \rangle$, is reached. Directed links from node i to the k_i nearest (in grid-space) other nodes are then created. A fraction f of edges are then rewired at random to add noise to the network. A more detailed description of the network is given in Appendix A.1.

3.4.1 Voter Scaling

As Fig. 3.2 shows, the CCIS model with appropriate parameter choices can closely match empirical vote distributions rescaled by Q/N. We simulate each election one time for each set of parameters to test how well our model can typically follow the empirical data, and each election is run on a spatially distributed scale-free graph (as described above) with N and Q the same as empirical data to account for finite size effects. We vary the initial fraction of individuals seeded until the model fits the distribution from Poland's 2005 elections (which has the largest number of elections). All other simulation parameters are fixed to reasonable values: $\beta = 0.1$, $\langle k \rangle = 10$, $\mu = 1$, $\delta = 0$, and $\alpha = 2.01$ (see Appendix A.1 for details regarding the fit

and the robustness of the results to changes in the parameters).

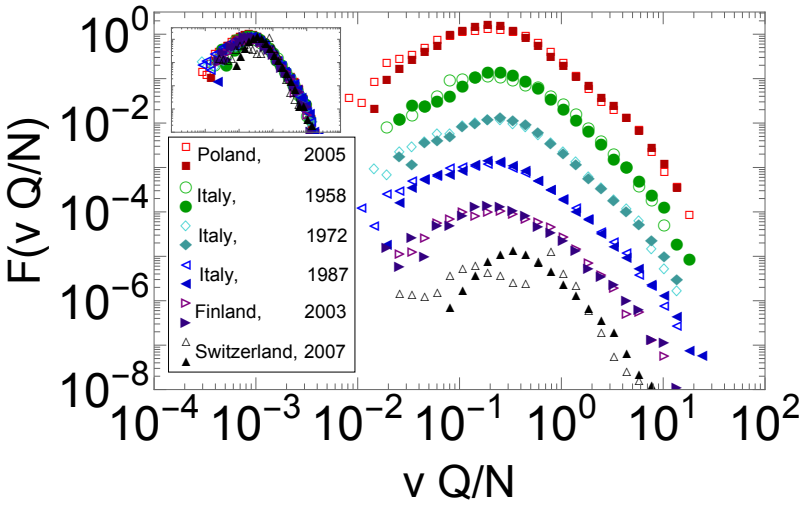

Figure 3.2: A comparison of scaled vote distributions between the CCIS model (closed markers) and elections (open markers) (data from [1]), in which data is shifted down by decades for clarity (inset shows the original data collapse). Here v corresponds to the number of votes, with the number of candidates, Q, and size of the population, N, equal to the empirical data values. The initial fraction seeded with a preference to a candidate is fitted to the scaled vote distribution of Poland's 2005 elections. All other parameters are fixed.

The simulation results plotted are for networks without random rewiring (i.e., $f = 0$), but we find similarly good fits for larger values of f. In the simulations, $\mu > 0$ and $\delta = 0$ in order to reach a non-consensus equilibrium, because otherwise we would have to stop the simulation at some arbitrary time before consensus is reached. These same parameters were used to fit all the other countries' elections.

Overall, we find good fits between our model and voter data as long as $\mu > 0$, and the distribution is sufficiently heavy tailed, i.e. the magnitude of the degree

distribution exponent is small ($\alpha < 3$). See Fig. A.2 in Appendix A.1 for a detailed analysis of the robustness of the fit to parameter variation. Our findings suggest that both individual stubbornness and heavy-tailed degree distributions in social networks [106] may be important underlying drivers of the generic behaviors observed in opinion dynamics.

The reason for the strong fit in Fig. 3.2 is in part because our model appears to follow a nearly universal distribution when each vote is rescaled by Q/N, like the empirical data from the elections it attempts to model. Of the elections modeled, we find that only Switzerland's diverges significantly from our model due to its unusual "double-hump" distribution, plausibly because votes are swayed by the local language differences (primarily French and German).

Agreement between the model and empirical data (Fig. 3.2) is also possible when the initial fraction of individuals seeded, $P(0)$, is 100% if the persuasiveness of each individual, β, is adjusted to 0.65 (see Fig. 3.3). In this case, because $\delta = 0$, no individual ever reaches the neutral state. Despite the fact that agreement with data can be achieved without the inclusion of a neutral state, we believe that such a state is important because most voters start out with little knowledge of the candidates.

One natural way to seed opinions when explaining the candidate vote distribution is to assume that only one individual has an initial vote preference: the candidate himself. This creates a poor fit for our model (not shown), possibly suggesting that the initial spreading process differs from the one that takes over after a short time.

Our work is influenced by the Fortunato and Castellano (FC) model (intro-

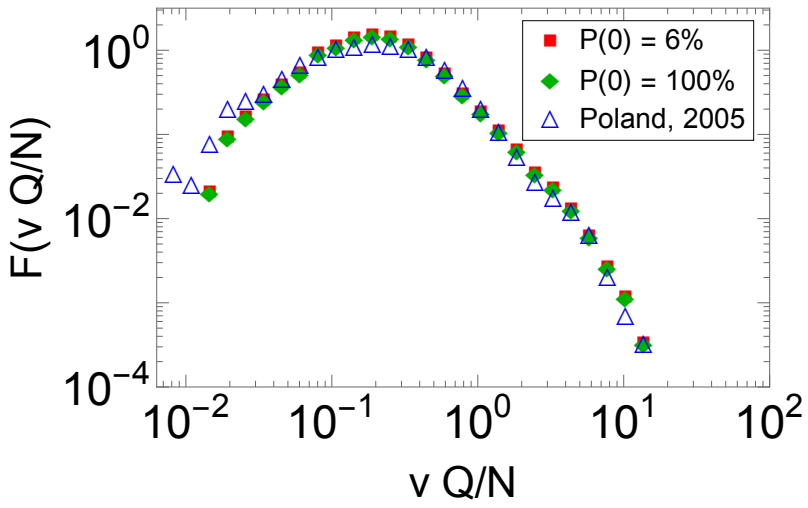

Figure 3.3: A comparison of the best fits between the CCIS model and the 2005 Poland elections with a fraction of nodes initially seeded with an opinion ($P(0)$) equal to 6% and 100% (see List of Abbreviations for the definitions of parameters). The parameters are the same except when $P(0) = 6\%$, $\beta = 0.1$, and when $P(0) = 100\%$, $\beta = 0.65$.

duced in Section 4.2), which was developed to describe the same distribution data [2]. In both the FC and CCIS model, individuals try to persuade neutral neighbors in the network at some rate. Opinions do not compete in the FC model, but instead spread within isolated networks, meaning that each of the Q candidates convince voters to vote for him or her by word of mouth to their friends, which then spreads to their friends' friends, etc. In this scenario, an individual only decides whether or not to vote for one specific candidate and never decides between candidates. The CCIS model is designed to capture a more realistic scenario in which candidates compete for the same set of voters [86–88]. We directly compare our model to the FC model in Fig. 3.4. Both models create similar fits, based on the log-likelihood

function, with neither being significantly better.

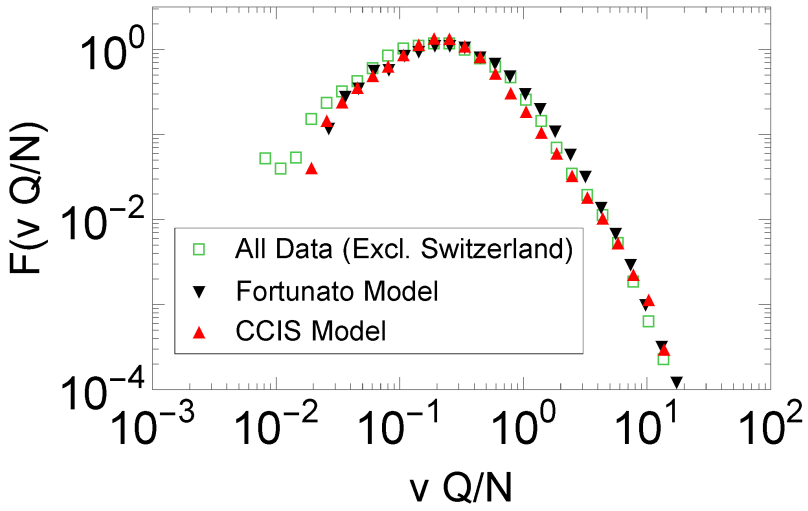

Figure 3.4: We plot the distribution across all elections shown in Fig. 3.2, excluding Switzerland, and compare our fit to the fit of the FC model [2].

3.4.2 Spatial Correlation

Next, we show that the CCIS model creates correlations that decrease logarithmically with distance, as seen in empirical studies [3, 21]. This behavior is not unique to our model because many models can create logarithmically decreasing correlations as they approach the VM Universality Class [108] in some special parameter range. We find it important, however, that our model is the first model we are aware of that can reproduce both the previously mentioned vote distributions, and this behavior, especially over a wide set of parameters. In comparison, the FC model [2] assumes non-interacting opinions on random graphs, and the Palombi and Toti model [70] assumes opinions interact on non-spatially distributed cliques with edges connected randomly between them, so votes are uncorrelated in space.

Analysis of the observed logarithmic correlations in the CCIS model are discussed in the next section. Simulations, however, suggest the most important property in our model to reproduce the empirical observations is a spatial structure in our social network, whether the network is a lattice, small-world (random rewiring), or the current scale-free spatial network. Therefore, this property is very general, and should be generically seen in empirical data.

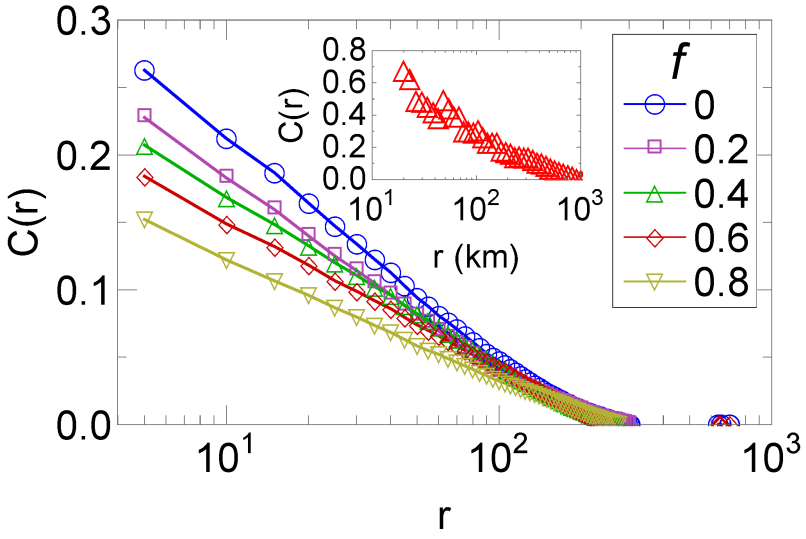

Figure 3.5: The correlation as a function of distance for the CCIS model (where nodes are separated by a unit 1 distance on a 10^6 node network). The CCIS model parameters are the same as in Fig. 3.2 except here $Q = 2$ and 100% of nodes seeded. f fraction of edges are randomly rewired on a scale-free spatially distributed graph ($f = 0$ corresponds to the network in Fig. 3.2) showing that the logarithmically decreasing correlations are robust. Inset: similar correlations are seen for data from the year 2000 United States Presidential election [3].

Figure 3.5 shows results from simulations of our model on spatial scale-free networks with 10^6 nodes and the same model parameters as in Fig. 3.2 (if $f = 0$).

The figure also shows results from simulations for which a fraction, f, of edges were randomly rewired. The rewiring process reduces the spatial features of the graph by creating long-range ties that significantly reduce the mean geodesic distance between points. Even with large f, however, we still see strong qualitative agreement with empirical data.

We note, however, that while empirical voting patterns are consistent with the CCIS model operating on a spatially-extended network, we cannot rule out the possibility that the empirical correlation data is the result of self-segregation, e.g., that "Republicans" move to "Republican" counties. Additional data is necessary to differentiate these two potential explanations for spatial correlations in voting behavior.

3.5 Analysis

In this section, we analyze the dynamics of our model to better understand the behaviors it is capable of producing. To do so, we simplify the model in three different ways, allowing us to probe the dynamics more thoroughly than any single approximation.

First, to probe the spatial correlation behavior discussed in the previous section we explore the limit in which our model simplifies to a diffusion process. Second, we explain how opinion sizes change in time with a transport-like equation, which assumes individuals mix homogeneously in an infinitely large network and tracks the time evolution of the density of individuals who have held a specified opinion for

designated length of time. Finally, we use the Fokker-Planck equation to explore, for the case $\mu = \delta = 0$ (i.e., no stubbornness and no recovery), how our model reaches opinion consensus for finite systems with heterogeneity in the connectedness of individuals. Under the Fokker-Planck approximation (FPA), we handle heterogeneity in the number of connections but we do not capture spatial effects or incorporate stubbornness and recovery, motivating all three separate types of analysis.

3.5.1 Spatial Correlations

Spatial correlations between opinions in the CCIS model decrease logarithmically over a wide parameter space (see Fig. 3.5). We can demonstrate this spatial correlation behavior analytically for the continuum limit of the CCIS model seeded with two opinions (and no neutral individuals) on a lattice grid, for the case $\mu = \delta = 0$. Because $\delta = 0$, nodes do not independently change to any other state, and furthermore, because $\mu = 0$, the probability of each node changing their state is $(number\ of\ opposing\ neighbors)/[(2d)^2\beta]$ at any timestep, where $2d$ is the degree of a d-dimensional lattice. In comparison, the two-opinion VM assumes that agents are convinced by a random neighbor's opinion at each timestep [7,8], or equivalently, the probability of any node changing their state is $(number\ of\ opposing\ neighbors)/(2d)$, therefore, in this parameter range, the CCIS kinetics is exactly the same as the VM, with time scaled by $2d\beta$.

The VM can be approximated as a diffusion process in the continuum limit [90],

meaning the correlation as a function of time, t, can be expressed as:

$$C(r) \sim \begin{cases} 1 - \frac{r}{\sqrt{Dt}} & d = 1 \\ 1 - \frac{\log(r)}{\log(\sqrt{Dt})} & d = 2 \\ r^{2-d} & d \geq 3 \end{cases}, \tag{3.1}$$

in which $D = d$, r is the distance between nodes, and nodes are separated from their neighbors by a distance of one unit. Eq. 3.1 is the same for the CCIS model in this limit, with $D = (2d^2\beta)^{-1}$ to reflect the rescaling of time. The spatial correlation between opinions in the CCIS model therefore decreases as $\log(r)$ for fixed time in this limit.

3.5.2 Transport-Like Approximation (TLA)

Next, we try to better understand how opinions change in time in the CCIS model. We present a partial differential equation similar to the transport equation, to describe the dynamics of the CCIS model in the mean field. This approximation, which we discuss in more detail in Appendix A.2, holds for all β, $\mu > 0$, and $\delta = 0$:

$$(\partial_t + \partial_\tau)\rho^{(A)}(t,\tau) = -\Theta(1-\tau\mu)(1-\tau\mu)\beta k \rho^{(A)}(t,\tau) \sum_{B \neq A} P^{(B)}(t). \tag{3.2}$$

Here, $\rho^{(A)}(t,\tau)$ is the density of individuals at time t that have have opinion A for a time τ. The above equation says that $\rho^{(A)}(t,\tau) \to \rho^{(A)}(t+\Delta t, \tau+\Delta t)$, and change to an opinion $B \neq A$ at a rate $\beta(1-\tau\mu)$. If $\tau\mu > 1$, the right hand side is 0 due

to the Heaviside step function, Θ. The boundary condition (not shown) describes the gain in new individuals (increase in $\rho^{(A)}(t,0)$) via conversion of individuals who were neutral or of an opposing opinion, allowing $P^{(A)}(t) = \int \rho^{(A)}(t,\tau')d\tau'$ to remain constant in equilibrium. Agreement between the equation and simulations is poor when $\mu = 0$, because, after being stochastically pushed out of equilibrium, the system quickly approaches consensus. Similar results are seen when $\delta > 0$, after incorporating a few additional terms. We will discuss how to analyse the dynamics when $\delta > 0$ in the next section. However, excellent agreement between theory and simulations is observed in Fig. 3.6 when $\delta = 0$ and $\mu > 0$.

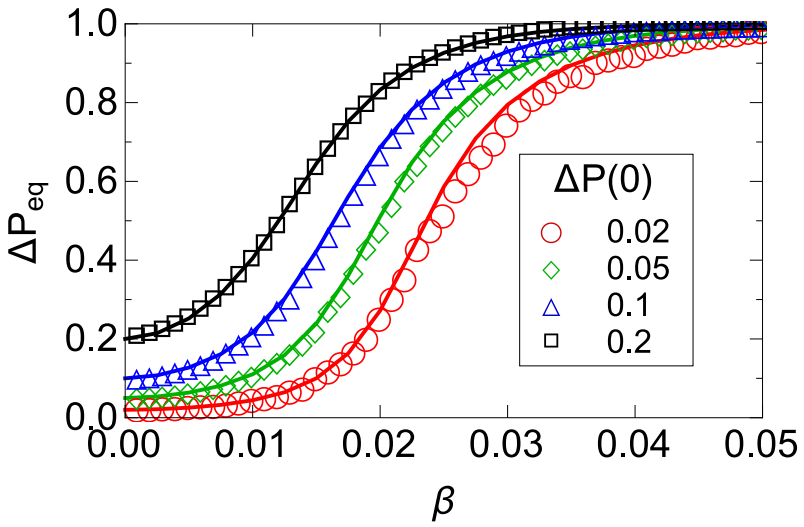

Figure 3.6: The difference in equilibrium opinion densities, $\Delta P \equiv |P^{(1)} - P^{(2)}|$, as a function of β between theory (solid lines) and simulations, where $\delta = 0$ and $\mu = 0.2$. $\Delta P = 0$ corresponds to a 50/50 split in opinions while $\Delta P = 1$ corresponds to complete consensus. Simulations are on networks $N = 10^5$ and degree $k = 10^2$.

3.5.3 Fokker-Planck Approximation of the CCIS Model

We can also analyze the model when $\mu = \delta = 0$, with the Fokker-Plank Approximation (FPA). The main difference between the FPA and the TLA is that the FPA takes into account the size of the system, and degree heterogeneity of a random graph, but does not incorporate the effects of stubbornness or recovery. Under this approximation, links randomly rewire, so we have no spatial information about the network, and cannot say anything about spatial correlations. It is therefore a powerful theory but only for specific network topologies. Our analysis may be improved upon, by modeling bipartite networks, networks with strong cliques, or using a more accurate pair approximation [10,109,110], but our goal here is to derive simple expressions that can describe some of the most interesting behavior. We give the details of the FPA in Appendix A.3 and describe the main results here.

Consensus time, T_{cons}, is found to be finite and scales in non-trivial ways with the network topology and the persuasiveness parameter β. If ρ are the fraction of individuals with one of two opinions, we find that

$$\frac{\rho(1-\rho)}{N_{eff}} \frac{\partial^2 T_{cons}}{\partial \rho^2} = -1, \tag{3.3}$$

where N_{eff} is the effective size of the network:

$$N_{eff} = \begin{cases} \frac{N}{\beta^2 \langle k^2 \rangle} & Outward\ Process \\ \frac{N}{\beta^2 \langle k \rangle^2} & Neutral\ Process \\ \frac{N}{\beta^2 \langle k^2 \rangle} & Inward\ Process \end{cases}, \tag{3.4}$$

and where $\langle k \rangle$ and $\langle k^2 \rangle$ are the first and second moments, respectively, of the network

degree distribution. Solving Eq. 3.3, we find that $T_{cons} \sim N_{eff}$ (see Appendix A.3 for derivation).

In Eq. 3.4, the outward process is where an opinion spreads from an individual to its neighbors (which is assumed in the basic CCIS model). More generally, there are two other ways the opinion could spread: (1) the neutral process is where an opinion spreads between two individuals on a random link, and (2) the inward process is where opinions spread from neighbors to an individual.

We now discuss comparisons between simulations and theory for the outward process (in Appendix A.3, we compare T_{cons} in simulations to an equivalent T_{cons} theory for the neutral and inward processes).

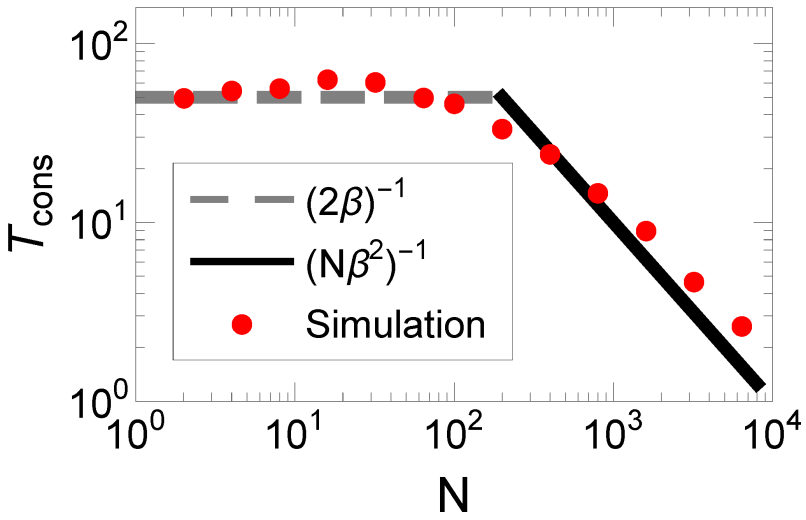

Figure 3.7: Mean consensus time versus N for a complete graph with $\beta = 10^{-2}$. Theory is the dashed gray line $T_{cons} \sim (2\beta)^{-1}$for small N, and the black line $T_{cons} \sim (N\beta^2)^{-1}$ for large N. This figure contrasts significantly with the IP model, which predicts that $T_{cons} \sim N$.

When $\delta = 0$, $\mu = 0$, and $\beta k = 1$, the CCIS model is similar to the invasion

process (IP) [10], in which a neighbor is randomly chosen to have the same opinion as the root node [10]. In the true IP, $T_{cons} \sim N\langle k^{-1}\rangle\langle k\rangle$, but in the CCIS model, $T_{cons} \sim \frac{N}{\langle k^2\rangle}$ for large N. The discrepancy is due to a fixed fraction of neighbors, $1/\langle k\rangle$, being changed in the CCIS model, instead of exactly one in the IP. Interestingly, this implies that $T_{cons} \sim (N\beta^2)^{-1}$ in a complete graph, which we observe in Fig. 3.7, while in the IP, $T_{cons} \sim N$ for $N \geq 10$ (not shown). In the CCIS model, we find that, for small N, the consensus time is roughly $(2\beta)^{-1}$, the mean time for consensus to be reached between two nodes. The crossover to the asymptotic limit is when $T_{cons} = (2\beta)^{-1} = (N\beta^2)^{-1}$ or $N = 2/\beta$. In conclusion, although some of the scaling behavior resembles previous work on the VM, we make predictions that are completely distinct from previous VM-like models. This discrepancy has the potential to be tested in a social experiment by observing the time to consensus in small groups, because the difference is apparent even for small N. We leave this for future work.

3.6 Consensus Times For $\delta > 0$

Finally, we numerically study T_{cons} for $\delta > 0$, where the previous analysis breaks down, in two ways. Figure 3.8 illustrates how the consensus time depends on the recovery rate δ when $\mu = 0$. Figure 3.9 shows how the consensus time depends on the stubbornness rate μ for different values of δ. Note that "consensus" here refers to the state in which at most one opinion remains. Thus the consensus state may contain a mixture of opinionated and neutral individuals, as long as all opinionated

individuals hold the same opinion.

Fig. 3.8 shows that the consensus time decreases with δ.Because the expected number of opinionated individuals at any given time decreases as δ increases, the time it takes for the opinionated individuals to reach consensus is also shorter. For this reason, we hypothesize that $T_{cons} \sim N_{eff} \sum_A P^{(A)}$, with N_{eff} as defined previously. In other words, we generalize Eq. 3.4 and claim $N_{eff} \sum_A P^{(A)}$ is the new effective size of the network which we leave for future work to explore more deeply.

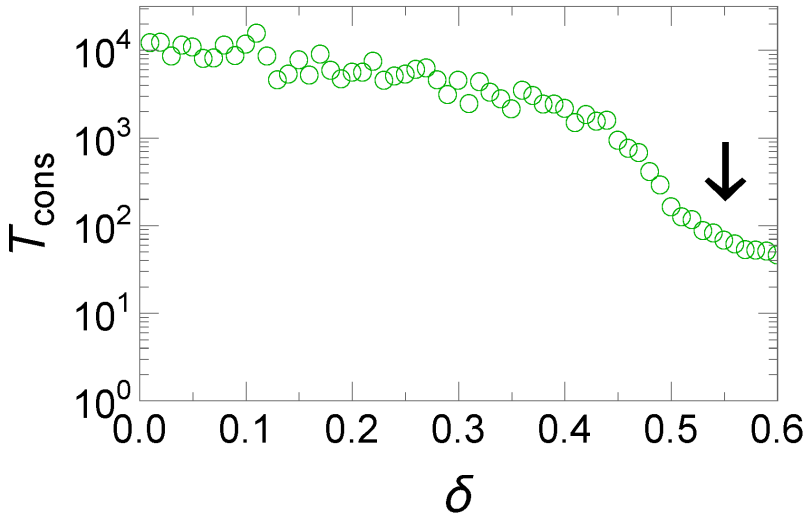

Figure 3.8: The consensus time versus δ with $\mu = 0$, and $\beta = 0.05$, on a $\langle k \rangle = 10$ Erdos-Renyi Network. The arrow indicates the critical point (calculated using SIS model analysis [4]) of the CCIS model, above which all individuals quickly approach the neutral state. We note that the consensus time appears to decrease monotonically with δ. The initial condition is a 50/50 mixture of opinions 1 and 2.

In Fig. 3.9, we plot T_{cons} versus μ for various values of δ to understand how our model more generally reaches consensus for finite networks. First, we find that

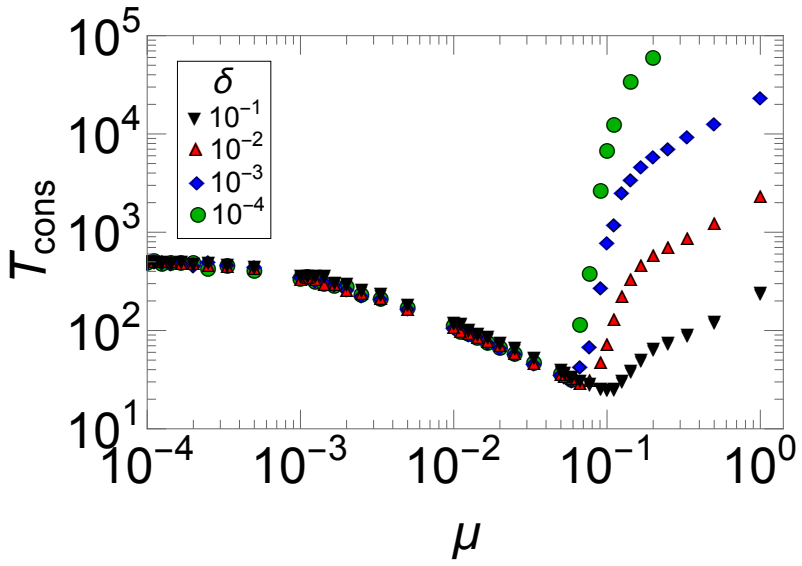

Figure 3.9: Mean consensus time for varying μ and δ on $\langle k \rangle = 10$, $N = 10^4$ Poisson networks with $\beta = 0.5$. A minimum in the consensus time is observed for $\mu \approx 0.1$, while analysis of model behavior for $\mu > 0.1$ reveals that $T_{cons} \sim \log(N)\delta^{-1}$.

$T_{cons} \sim \log(N)\delta^{-1}$ for small δ and $\mu > 0.1$, which, in this limit, is in agreement with previous analysis [98]. The behavior of T_{cons} versus μ demonstrates interesting parallels to other models [69, 98, 99] (Fig. 3.9), whereby at a non-trivial value of $\mu = \mu_c(\delta)$, the consensus time reaches a minimum, and at larger values of μ the consensus time increases significantly. This may generically imply that large groups reach consensus relatively quickly if individuals are moderately resistant to changing their opinion.

3.7 Conclusion

In conclusion, we have introduced a model of opinion dynamics that agrees with current empirical data and exhibits interaction dynamics based upon real hu-

man behavior.

In addition, because our model makes few assumptions, it may plausibly explain a range of behaviors, which future empirical investigations may be able to corroborate. For example, the model can be used to explore the "viral" spread of competing products, in which stubbornness is mapped to increasing brand loyalty [111, 112]. In this case, the brand-share distribution might be similar to Fig. 3.2.

Future work is necessary, however, to model opinions with greater realism. As mentioned previously, this model might benefit from additional realistic assumptions. For example, mass media could be added, because it can be more influential than individual persons. Similarly, we could add party affiliation, which may bias which candidate(s) individuals initially prefer, or are likely to support in the future [9]. Additionally, the recovery rate could be tied to an individual's stubbornness, instead of constant as we assume here for simplicity.

In addition, one could model heterogeneous stubbornness, either at the opinion level (as our model assumes) or individual level, because some individuals appear to stubbornly hold on to an idea, while others may shift their stance more readily. This is known to add greater realism to opinion dynamics because the most stubborn individuals possible, known as "zealots" in previous literature, can help push the political preference in a two party system near the 50/50 mark, alike to what we observe in the CCIS model [66, 71, 72]. Expanding on previous work, we expect that adding heterogeneous stubbornness to our model can further slow down or stop consensus and potentially create better agreement with data. In addition,

we assume agents linearly increase their resistance to alternative opinions in time. This is not necessarily true because PTP a day before a trial produced a negative correlation between the biased news and the juror decision, while PTP exactly a week before a trial is not statistically significant [78]. A non-linear or non-monotonic stubbornness may significantly change the dynamics.

Finally, this work assumes that all opinions are equally strong and spread at the same time, but this is not necessarily true in reality, which we discuss briefly in Section 3.3. MySpace started before Facebook, for example, and therefore more people initially preferred MySpace [91]. Facebook was later seen as a preferred option, however, and eventually dominated social media at the expense of MySpace and similar platforms. Future work should therefore allow for a first-mover advantage [113] and opinions that are stronger or weaker than others to better capture reality.

3.8 Acknowledgments

We would like to thank Dr. Arnab Chatterjee for pointing us to the election distribution data. Our work is supported by DARPA under contract No. N66001-12-1-4245 and No. D13PC00064.

Chapter 4: Opinion Dynamics in Juries and Judicial Panels

4.1 Introduction

Opinion dynamics, the study of how competing ideas spread through social interactions, has a long history in mathematics [7, 8], physics [75, 79, 80, 108, 114], linguistics [115,116], and computer science [91,117–119]. Recent work has focused on whether competing opinion models can successfully describe the statistical patterns observed in empirical data, such as the lack of consensus among groups [65, 77], universal distributions of votes [1, 2, 25], and the long-range correlations between voters [21, 25]. Typically, these kinds of studies focus on a large population of individuals, using an assumed [2,65,77], or partially inferred [21] interaction network, and a single minimal model is used to explain the data, rather than a class of plausible models, although exceptions to this rule exist [120].

In this chapter, we study how juries and judicial panels arrive at their verdicts. When these groups hear a case, they deliberate amongst themselves and tend to avoid influence from outsiders. After a period of time, they vote on one of two verdicts, "guilty" or "innocent". It is therefore safe to assume that juries and judicial panels strictly interact with each other, if anyone, which reduces the potential complication of trying to model influence from outside the group.

One feature of the process we wish to capture through modeling is how the time to reach a decision varies across cases. We note that different opinion models have very different time distributions when reaching a verdict [121], even though the average or variance in their distributions may be similar. Matching the model's time distribution to the data can allow us to systematically test what models can fit the data best, based on methods introduced in previous work [122, 123].

Our findings contrast with previous work on opinion dynamics, which typically study the interaction of thousands or even millions of individuals [1, 2, 21, 124], and do not model the time for groups to reach a (possibly split) decision in empirical data. Despite that, several studies have looked at the times to reach consensus [9, 10, 69, 98, 109, 110, 121, 125] or even the asymptotic consensus time distributions [124], which partly inspired our current work.

An important limitation in our data is that each group of jurors and judges observes a different trial, and there is a diversity of lawyers, defendants, evidence and so on between trials that can potentially affect how cases are decided. To address this issue, we first explore whether the data is too diverse to be described by a consistent mechanistic model. To do this, we examine whether we see consistent patterns in the data even once we control for diverse aspects. For instance, we split the data by the number of jurors or judges, the trial time (when available) and the final vote (when available, e.g., 10-2 in favor of a guilty verdict). Across these different categories, we see some statistically significant differences in the distributions, but we also see common patterns in the data, to be discussed shortly, therefore many common properties in this data appear resilient across trial and jury attributes. This

provides a strong hint that a consistent mechanistic model may be able to describe all our data.

Before undertaking our modeling goals, we first analyzed the empirical data for notable patterns. We find that the mean deliberation time, $\langle T_{\text{delib}} \rangle$, typically appears to scale with the trial time, T_{trial}, as $\langle T_{\text{delib}} \rangle \sim (T_{\text{trial}})^{1/2}$, and notably it scales sublinearly across all datasets. If T_{trial} is a proxy for the amount of information gathered, and $\langle T_{\text{delib}} \rangle$ a proxy for the amount of information individuals argue over, this scaling would suggest that a smaller relative share of the information is deliberated on, as the trial time increases. We also observe that the system size does not scale strongly with the number of jurors or judges, which is counter to many intuitive models [10, 125].

In order to identify what model features are most important for capturing the data, we study a set of plausible models of our data: the majority voter model (MVM) [6], three non-interacting null models, in which jurors make decisions independently of each other, and models in which the deliberation time has an exponential tail. Non-interacting null models reflect the hypothesis that jurors might reach a decision without significant influence or consultation from each other, while exponential tail models represents a wide class of competing models where opinion consensus is rare, e.g., [6, 9]. We find that none of these models can describe the deliberation time distributions. By ruling out simple non-interacting null models and showing that an interacting model does describe the data, our studies suggest that jurors influence each other to reach a decision.

We find that we can modify the MVM to more successfully describe the delib-

eration time and final vote distributions under the assumption that groups stop at a constant rate once enough people are in the majority for a verdict to be reached (75% for juries, and a simple majority for judges in the ECHR). Alternative non-interacting null models, in comparison, are unable to qualitatively match the final vote statistics. Our results provide mechanistic support for previous work on group decision formation [16–19] suggests the accuracy of group judgements is harmed by social influence. Finally, by assuming the timescale for individuals to change their opinions is determined by a random walk, we can successfully recreate the observed scaling between deliberation time and trial time.

The rest of the chapter is organized as follows. First, we discuss related work in Section 4.2). We then discuss patterns seen across our datasets in Section 4.3. Next, in Section 4.4 we discuss in detail how we model the data, which models disagree with data, and how models in agreement with data hint at mechanisms for group opinion formation. Finally, we summarize our results, and present future work in Section 4.6.

4.2 Related Work

Here, we discuss previous work on how the decision time and decision quality is affected by various factors for juries. We also discuss work that has more generally tried to understand how collective human behavior can emerge from individual interactions, before discussing opinion dynamics research in physics and mathematics that has tried to quantify these behaviors. Finally, we discuss how the time to reach

a decision is closely related the study of survival analysis.

First, we want to better understand how our work relates to previous work on juries and judicial panels. After the US Supreme Court case *Williams v. Florida* 399 U.S. 78 (1970), in which jury sizes of less than 12 individuals were ruled constitutional, sociologists became very interested in whether a smaller jury still provided a fair trial to defendants [126–129]. Researchers found, for example, that juries with twelve members achieved consensus less frequently than three-member juries [129] (based on a lab experiment) or six-member juries [127] (based on a meta-analysis of experiments), and that larger groups generally took longer to reach consensus [127].

In addition, Devine *et al.* explored how the distribution of times to reach a decision varied with external factors, such as size, number of counts, civil-versus-criminal trials, unanimity, and the verdict [127]. As might be expected, a higher number of counts, criminal trials, and non-unanimous verdicts all take longer to be decided.The authors also find that, for civil trials, innocent verdicts take longer than guilty verdicts, although guilty verdicts are more common. This property is also seen in many datasets we explore as well. Our work differs from that of Devine *et al.*, however, in exploring several newer datasets, including court cases from the ECHR, and most importantly, in modeling the dynamics of the decision process to understand why particular universal properties are observed.

Unlike previous work, we try to understand the mechanisms behind the formation of jury and judicial verdicts. We find evidence that decisions depend strongly on the influence individuals have on each other, rather than only on the compelling evidence of the court case.

In addition to the aforementioned studies of juries, our work also builds upon general models of collective behavior. Statistical patterns in language diffusion [116] [1], elections [21], and the spread of opinions on a cosmetic website [130] all appear to be well-modeled by variants of this model.

We also explore the utility of the MVM, a well-studied model of opinion dynamics [6], for explaining the empirical patterns we observe. In this model, at each time step, a randomly chosen individual adopts the opinion of the majority of his/her neighbors with probability p (and hence adopts the minority opinion with probability $1-p$). individuals pick the majority opinion of their neighbors, and then change their opinion at a constant rate. Although the dynamics appear similar to the VM, the MVM creates very different predictions. For example, unlike in the VM, the two-opinion MVM has a critical value of p, above which a majority of individuals settle on one of the two opinions, while below this point, opinions are equally split.

There have also been a number of empirically-based studies exploring how individual opinion formation is influenced by others. For example, some previous work has discussed the strength social influence has on individual decisions [16–19, 24], or has inferred the mechanism of opinion formation by studying the influence neighbors have on a user [131]. Furthermore, a recent paper explored how to infer the best model of opinion formation, based on how users influence each other in a cosmetic website [130]. More specifically, data from a popular Japanese cosmetics website was modeled with a VM-like model that weighted a decision based on the

[1]The similarity between the cited model and the VM was brought to my attention by Prof. Maxi San Miguel.

current and past opinions of neighbors. It appears that a power-law weighting, where the weight of older opinions falls as t^a, and where t is time, and a is a constant, best describes this data. Further research has tried to capture general statistical patterns, such as the distribution of votes [2, 21, 25], correlations between votes [21, 25], and the propagation of competing hashtags on Twitter [118, 132].

We distinguish our study from previous work by inferring the opinion dynamics of groups through easy-to-measure macroscopic variables, such as the distribution of times to reach a decision. Importantly, in juries and judicial panels, this method allows us to understand dynamics far better than we might otherwise, because, in an offline setting, it is difficult to know what each individual's opinion is at a given time, and therefore we cannot directly determine the mechanism through which individuals influence each other. For this reason, however, our method to determine how opinions form could be applied to more general settings than many previous experiments.

4.3 Data

The goal of this chapter is to gain insights into the dynamics of competing opinions by modeling data from juries and judicial panels. In this section, we discuss the data that we aim to model.

Before we describe the features of the data we study, we first discuss the features we seek in an ideal data set to use for building modeling models of competing ideas. Namely, we believe the ideal data set should have some or all of the following

qualities:

1. Well-defined competing opinions (i.e., each individual has a precisely known opinion)

2. A closed environment with Isolated groups of individuals, uninfluenced by outside environmental factors

3. Randomized groups (i.e., not self-selected)

4. No arbitrary time limit before agreement has to be reached

Well-defined opinions would be opinions that can be quantified. Asking an individual their stance on abortion, for example, may not be well-defined, because the topic is too broad to usually get a single opinion. We should not confuse a well-defined opinion with a discrete opinion, however. One could quantify an opinion to be on a continuous range, a la the Deffuant model [133].

Groups must be isolated from outside factors like media, or friends, in order to reduce the influence they may have on group decisions, enabling better control over the experimental conditions. With isolated groups, we will have some idea of who influences whom, but if the influence is compounded by unknown sources, then modeling the dynamics becomes more difficult.

Groups must be randomized to partially account for homophily, where individuals choose to contact those with similar opinions. This has been known to cause biases in studies on influence [134], and a paper by Shalizi and Thomas [23] suggested that homophily and influence are extremely difficult to distinguish in many

systems. Even after randomizing links, individuals may still live in cities that are spatially close, and therefore some opinions may be more similar than expected by chance [21], but the effect can at least be reduced by making connections between strangers.

Any time limit, we believe, could bias how opinions form. In a time-limited study, a group may want to reach any collective decision before the allotted time, rather than the best decision through a careful discussion. There are realistic instances when groups make decisions with in a strongly limited timeframe, e.g., collaborative decisions made before a deadline, thus motivating studies on these dynamics in the future, but we want to first focus on a simpler system in this work.

Despite the relatively strict goals we set, we find that juries and judicial panels can create strong candidate datasets for modeling. First, the decisions are binary, i.e., a juror typically votes that a defendant is either guilty or not. Second, juries are typically randomized groups of strangers who are not allowed to be influenced by any outside factors, like the media, and deliberations typically take place in in one sitting in an isolated room, which reduces the chance for outside influence. In addition, there is no strong limit to the deliberation times, therefore the distribution of times before agreement is reached only happens at the groups choosing.

There are, however, some less-than-ideal properties of our data. First, the decisions made by juries and the ECHR are case-dependent, a verdict for one case should not necessarily be the same as the verdict in another case. Similarly, some cases appear to simply take less time to deliberate on average, which affects how we can model the time distributions before a decision is reached. That said, despite

the richness of individual cases, in aggregate, broad statistical patterns appear. Our analysis finds there are statistically significant differences in behavior within datasets, but the qualitative findings, and agreement with our models is broadly unchanged.

Next, our data is limited to small groups (e.g., ≈ 10 individuals), and jurors do not necessarily reach consensus. Since the US Supreme Court case *Apadaca v. Oregon* 406 U.S. 404 (1972), state court cases did not require unanimity from juries for a verdict to be reached, and similarly, the ECHR only requires a simple majority for a case to be decided. Recent theoretical models, however, have typically assumed complete consensus can be reached and that groups are very large groups (e.g., sufficiently large as to study using finite scaling analysis) [9, 10, 25, 69, 98, 121, 135]. Therefore, a direct comparison between the time distributions in empirical data and previous models is not necessarily easy. We find in the next section, however, that incorporating the final vote into our models allows us to more directly compare models to data, and to gain more insight into how group decisions are reached than previous work.

Our jury data is taken from Multnomah County, Oregon [14], San Fransisco County, California [11], Thurston County, Washington, and Douglas County, Nebraska [12, 13], as well as from the European Court of Human Rights (ECHR) [5] (see Appendix B.1 and B.2 for details). We will call these datasets OR, CA, WA, and NE, respectively, and split this data by the number of jurors in each case (see Appendix B.2). Therefore, the OR 6 and OR 12 datasets corresponds to the OR data with 6 jurors and 12 jurors, respectively. In all our datasets except for those

from WA and NE, the deliberation time and final vote are known, which can affect each other. However, unique to the CA, WA, and NE datasets, is information on the trial time, which we discovered can also strongly affect when decisions are reached.

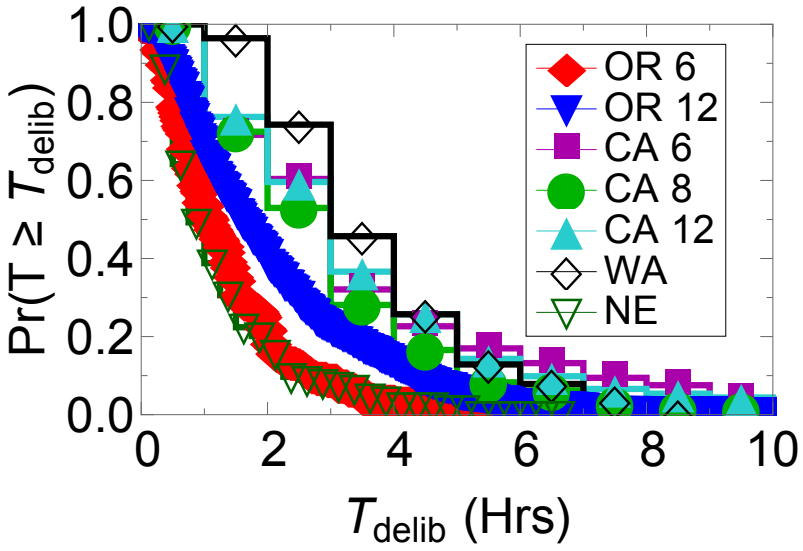

Figure 4.1: Fraction of juries still deliberating versus time in the jury datasets. Shown are datasets from CA, OR, WA, and NE.

We first observe that the fraction of juries still deliberating as a function of time varies between datasets (Fig. 4.1), but they have comparable median times (1-2 hours), and an exponential-like tail. It may be intuitive to expect that similar processes underly the dynamics of all datasets.

Second, we find significant heterogeneity in the final vote for juries (Fig. 4.2). Surprisingly, very few cases have a final vote of between 25% and 75% guilty. Looking more closely at the data, we find that the 25% and 75% fractions correspond to the proportion of guilty votes necessary for a verdict to be reached [11, 14]. It would apparently suggest that juries deliberate until enough individuals are in the

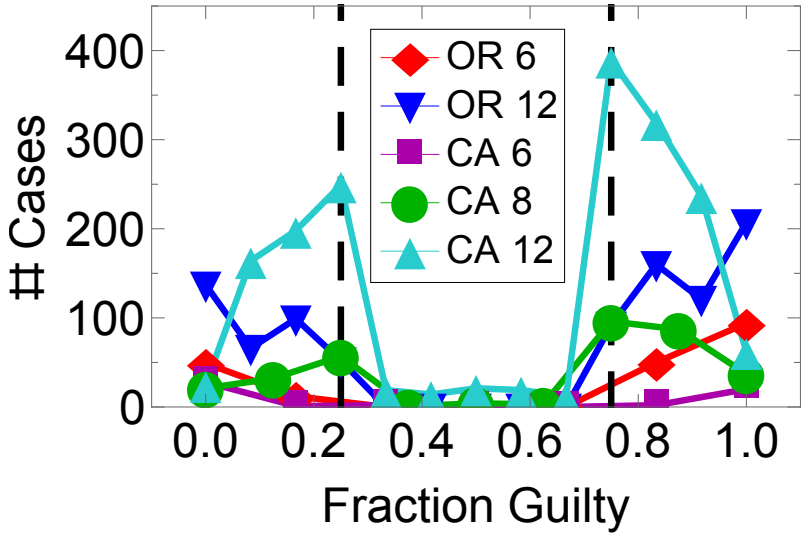

Figure 4.2: Number of Jury Cases as a function of vote (NE and WA data does not provide the final vote). We notice the cases are highly peaked near the critical fractions of guilty voters, 25% and 75%. Between those ranges, where very few cases lie, juries are considered hung. This data may suggest that juries skew their votes to reach a verdict.

majority to justify stopping the deliberation. We show that this hypothesis agrees with our mechanistic model in the next section.

Next, we explore how the deliberation times differ (Fig. 4.3). We notice a strong correlation between the mean deliberation time and the trial time in CA[2], WA, and NE datasets. Notably, this appears to roughly follow the line $(T_{\text{trial}})^{1/2}$, suggesting, that despite the heterogeneity in the datasets, there is a common un-

[2]Unlike the WA and NE datasets, the trial time in the CA dataset was recorded with a resolution of days and not hours. WA and NE datasets did, however, list both days and hours a jury was in trial. From this data, we found that a day in a trial took approximately 4 hours, which was our conversion for CA data as well.

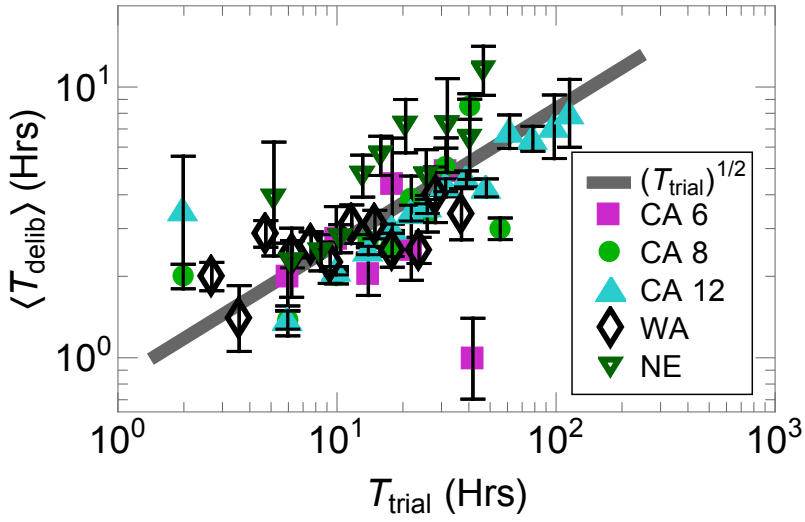

Figure 4.3: The mean time to reach a verdict versus trial time, with error bars corresponding to 90% confidence intervals. Interstingly, we see a strong trend in the data which roughly follows $(T_{\text{trial}})^{1/2}$. The single outlier for the CA 6 data was estimated from only two datapoints, and is therefore only a product of how we bin the data.

derlying mechanism for how long deliberation usually takes, which we discuss in greater detail in the next section. Fitting the datasets separately, we find that all datasets except the WA dataset [3] has a scaling in line with our hypothesis, most notably with the CA 12 dataset, one of the largest, scaling as $T_{\text{delib}} \sim (T_{\text{trial}})^{0.53\pm0.03}$. We also find that the deliberation time is not strongly dependent on the number of jurors (Fig. B.1a in Appendix B.3), in contrast to predictions made from the VM [7, 8], and variants of the VM [9].

Finally, we notice that the time to reach a verdict is strongly dependent on the final vote (Fig. 4.4a), although the trial time and final vote are not strongly

[3] For this dataset, we find $T_{\text{delib}} \sim (T_{\text{trial}})^{0.19\pm0.07}$

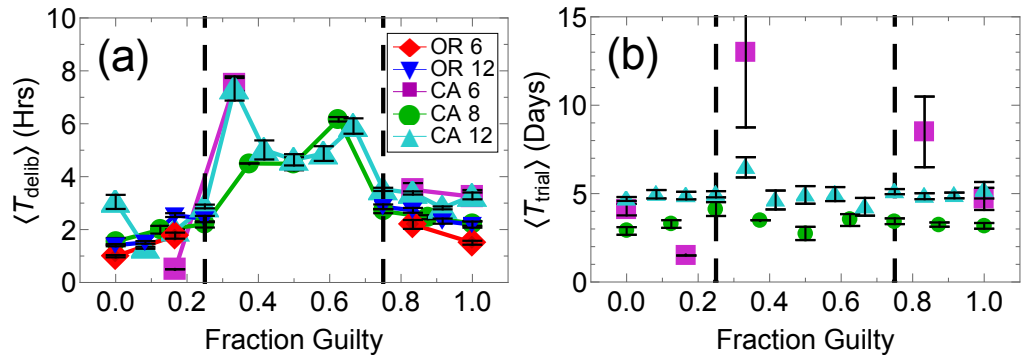

Figure 4.4: (a) The mean deliberation time versus the final vote, with error bars corresponding to 90% confidence intervals. Dashed line corresponds to the intervals that juries are hung. The mean deliberation time peaks just after, and just before, respectively, the critical fractions for juries to reach a verdict. (b) The mean trial time versus final vote with the same legend as (a), with error bars corresponding to 90% confidence intervals. Although Figs. 4.3 and 4.4 show that the final vote and trial time can affect the deliberation time, these two properties do not appear to be strongly correlated with each other.

correlated with each other (Fig. 4.4b). Generally, with less unanimity, there is more time needed for a verdict to be reached, a finding also seen in the ECHR dataset (not shown). However, unique to the jury data, when juries are hung, there is an enormous amount of time needed to reach a verdict (as much as 8 hours on average for hung juries, versus $2-3$ hours for a jury verdict. If we image the critical fractions between juries that reach a verdict and are hung as critical points, in statistical mechanical language, it would appear that the deliberation time acts as an order parameter, which peaks near this critical point. This is known in statistical me-

chanics as "critical slowing down". We stress this data is suggestive and not a proof of critical dynamics, however if our finding is true, then the mechanics of opinion formation near this critical point could approach an opinion dynamics universality class, where the statistical behavior looks similar even when the detailed mechanics of opinion dynamics differ from dataset to dataset. Future work is necessary to find whether this behavior is seen in other datasets, and to model the dynamics near this phase.

4.4 Modeling Data

The goal of our models is to address the following questions:

- Can we distinguish whether the similarity of opinions among jurors results from independent opinion formation versus influence?
- Why are decisions typically in non-consensus?
- Why does the trial time affect the time to reach a decision?
- Most fundamentally, how can we understand social systems without the need for expensive and time-consuming social experiments?

To begin answering the final question, recall that once we successfully model the dynamics, we can vary parameters at will, and make predictions of how new behavior arises without needing to carry out social experiments, a motivation similar to the motivations of modeling disease epidemics [136]. More practically, once we understand how individuals influence each other, we can understand how to promote

wanted ideas, and suppress unwanted ideas, from spreading, without expensive social experiments. This has been a driver of recent work in idea spread literature [137, 138], but future work should extend the work to competing idea spread.

In order to simplify the problem at hand, we do not try to capture all of the statistical patterns in the data, and instead focus on key features while ignoring, for the time being, some other properties. First, we do not model hung juries. Although their deliberation time behavior appears unusual (Fig. 4.4), the number of juries that end up being hung is a small fraction of the entire dataset. This will be important to model in the future however, if we want to ask, for example, whether the requirement for unanimity in jury data necessarily creates more hung juries, or whether the mechanics of juries approaches a universality class near these points. We also choose not to look at the varying amount of influence individuals may have. We find in the OR dataset, for example, that juries are much more likely to agree with the majority of other jurors, compare to the typical juror (not shown). This may suggest the foreman is highly influential, or is heavily persuaded by the other jurors.

What we wish to model instead are the deliberation time distributions, especially as a function of trial time. We will show, however, that this is insufficient information for our minimal models to adequately model data. We will next model both the deliberation time and final vote to understand why decisions are not strictly in consensus, and find surprisingly good qualitative agreement with data. First, we compare well-motivated models against each other and find what model best describes the opinion dynamics of juries and judicial panels.

We test several models, including:

- Exponential tail distribution models
- Maximum of random variables with exponential distributions
- Maximum of random variables with Gaussian distributions
- A simplified MVM

In all cases, parameter estimations are found by maximum likelihood estimation.

Models with exponential tail distributions are surprisingly common, e.g., when verdicts take so long to be reached that they can be approximated by the first passage time of a Poisson process. This may be seen, for example, in the MVM, with a non-zero flip rate [6], and in variants of the VM [9]. This should be typically seen in models with a non-consensus equilibrium point.

We find, however, agreement is very poor with most of our data (Table 4.1), especially when the dataset is large, such as the CA 12, or OR 12 datasets. Furthermore, recall that it takes more time to reach a verdict when the final vote is non-consensus than consensus (e.g., 9-3 versus a 12-0 decision), seen in Fig. 4.4. This is opposite of what we would expect if juries are in non-consensus more than consensus, in which it is rarer for near-consensus to be reached.

The maximum of exponential or Gaussian distribution models assume that the time to reach a verdict occurs when then final juror or judge makes a decisions, where each individual's decision is a Gaussian or exponential random variable (which are plausible and common distributions in nature, but not necessarily the only

distributions possible). We call these models the Block Maxima Models (BMM), and is one of the simplest non-interacting null models, where individuals make a decision independently of one another. The algorithm is as follows:

1. Pick N random variables from an exponential (Gaussian) distribution, where N is the number of jurors.

2. The deliberation time for this set of random variables is the maximimum of the set.

3. The entire distribution from this model can be constructed after running steps 1 and 2 many (e.g., 10^6) times.

Similarly, we could create a more complicated null model where the exponential distribution differs depending on whether the juror's verdict is guilty or innocent, which we call the Two-Timescale (TT) Null model. The final vote can then be picked from a binomial distribution based on the overall fraction of individuals who vote guilty. This model is therefore able to predict both a time and vote distribution which we can compare to data (Table 4.1 and Fig. 4.5). We focus on a model with exponential time distributions because the exponential BMM, a simplified version of this model, fits the data much better than the Gaussian BMM, based on the log-likelihood ratio.

Surprisingly, we find poor agreement to the null hypotheses, leading us to rule all these models out (see Appendix B.4). Furthermore, we will provide evidence that a simple model, which includes influence from neighbors, can better match these mechanics. This would suggest that jury verdicts are shaped by mutual influence.

Dataset	**#Jurors**	**N**	T_{trial} (Hrs.)	N_{bin}	P-Values, Significant When $p > 0.1$, (# Parameters)					
					Time Distributions					Time & Vote
					Exp. Tail (1)	BMM: Exp. (1)	BMM: Gauss (2)	TT Null (2)	MVM (1)	Modified MVM (3)
CA	6	53	8-88 (All)	53	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	0.06	$< 10^{-3}$	2×10^{-3}
	8	338	6-10	170	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
			10-19	121	**0.24**	$< 10^{-3}$	$< 10^{-3}$	**0.13**	3×10^{-3}	8×10^{-3}
			19-34	30	-	**0.11**	$< 10^{-3}$	**0.13**	**0.21**	**0.57**
	12	1726	6-10	498	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
			10-19	646	0.03	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
			19-34	383	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
			34 - 61	104	0.02	6×10^{-3}	$< 10^{-3}$	**0.34**	**0.12**	2×10^{-3}
			61-110	37	**0.38**	$< 10^{-3}$	$< 10^{-3}$	6×10^{-3}	7×10^{-3}	$< 10^{-3}$
OR	6	207	-	207	0.03	$< 10^{-3}$	$< 10^{-3}$	**0.71**	$< 10^{-3}$	1×10^{-3}
	12	951	-	951	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
WA	6[a]	141	3-6	29	-	0.07	1×10^{-3}	**0.12**	**0.70**	-
			6-10	45	**0.72**	**0.96**	2×10^{-3}	**0.99**	**0.39**	-
			10-19	40	**0.30**	**0.68**	8×10^{-3}	**0.70**	**0.78**	-
NE	12[a]	135	6-10	38	**0.26**	$< 10^{-3}$	$< 10^{-3}$	**0.16**	**0.11**	-
			10-19	44	**0.53**	2×10^{-3}	$< 10^{-3}$	**0.23**	0.07	-

Table 4.1: P-values for jury models (see Appendix B.4 for details). "-" in cell: we have too few cases to determine p-values.

[a]Estimates based on [13].

Previous work suggested that independent decision making (specifically homophily) and influence can be indistinguishable in certain circumstances [23, 134], but the disagreement with our null model may provide a way to distinguish these two mechanisms. We cannot rule out that a more complicated model without influence could produce these same statistics, although we will show that a simple influence model provides surprisingly strong qualitative agreement.

Finally, we model the deliberation times with a simplified MVM, which is much like the original MVM [6], except we remove the noise term. We assume jurors and judges interact on a complete network, meaning each individual can influence, and be influenced by, every other individual whom they deliberate with. We assume a particular node is picked at random and chooses the majority opinion at a rate τ^{-1}. As mentioned previously, an MVM with a noise term might create an exponential tail, which we have already ruled out as an appropriate model, therefore it would seem that the MVM model in the noiseless limit is the best way to adequately model jury decision times. If we ignore vote data, however, it would appear that agreement with the dynamics is poor (Table 4.1), so we instead will modify this model to incorporate both the vote and time distributions.

Importantly, this "modified MVM", can help us better understand why consensus does not appear in juries or the ECHR. First, we start with a jury where votes are evenly split 1/2 guilty and 1/2 innocent. Next, let ρ_g be the proportion of people who ultimately vote guilty, across the whole data set of interest. The algorithm for the modified MVM is as follows:

1. At the first step: pick a innocent-voting individual with probability ρ_g, and a guilty-voting individual with probability $1 - \rho_g$)

2. If the first individual picked is a non-guilty voter, the jury will eventually reach a guilty consensus if not stopped at an earlier time (see step below). Once a majority and minority opinion are established, individuals each choose the majority opinion at a rate τ^{-1}.

3. For juries, once 75% or more individuals choose a guilty verdict, at a rate α_{75}, deliberation stops, and a verdict is reached. Similarly, if 25% or less individuals choose a guilty verdict, deliberation stops at rate α_{25}. For the ECHR, verdicts stop at a rate α_G or α_I, when the fraction of guilty verdicts is over 50% or $\leq 50\%$, respectively.

The motivation for the various parameters is the following. Guilty verdicts are seen in roughly 60% of jury cases, and 90% of ECHR cases. We therefore need more individuals to pick a guilty verdict over a non-guilty one to match the vote histograms. Second, we notice in our jury data that innocent verdicts can take more time than guilty verdicts, and that consensus does not always happen. We therefore have deliberations stop at a fixed rate that depends on whether the jury verdict is guilty or innocent. We can interpret these various rates as the probabilities that juries decide to take a vote and stop when they are in they have a supermajority. Finally, we allow individuals to change their opinion at a rate, τ^{-1}, where τ is the mean time to change their opinion. Although τ itself could vary depending on whether the verdict is guilty or innocent, we keep this parameter constant to

simplify the dynamics.

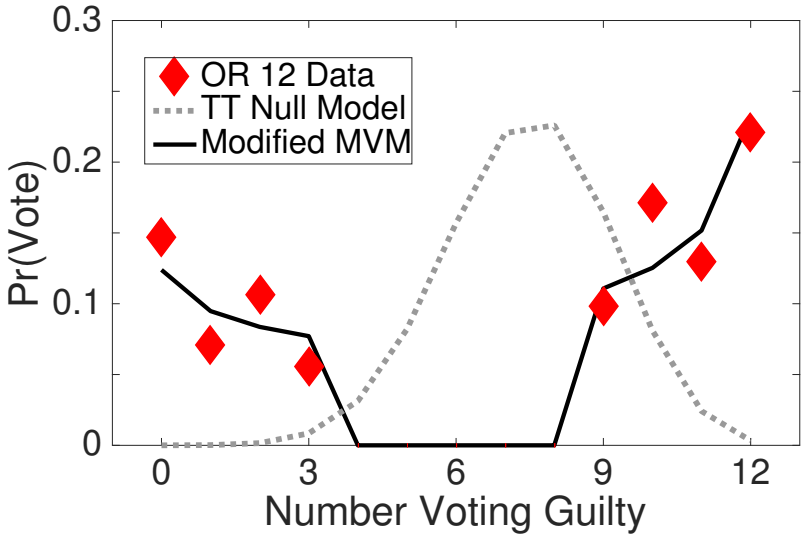

Figure 4.5: A plot of the OR 12 vote histogram, the TT Null model's vote distribution, and the modified MVM's best fit distribution, which is in better agreement with data.

We find agreement with data is still poor (Table 4.1), at least in part because the requirements for good agreement are much stronger. Namely, we test the model to a 2-dimensional Kolmogorov-Smirnov (KS) test, where the model must match both the vote and time distributions for each dataset. Despite the quantitative disagreement, we find strong qualitative agreement to the data (Fig. 4.5), which suggests that we are approaching the true dynamics.

Next, we try to better understand the dynamics by observing how these parameters change with external factors, like the trial time and number of jurors. We first observe that $\tau \sim (T_{\text{trial}})^{1/2}$ (Fig. 4.6), which implies something stronger than Fig. 4.3: not only does the overall time to consensus increase with trial time, but the time for each individual to switch their opinion increases as well. We therefore

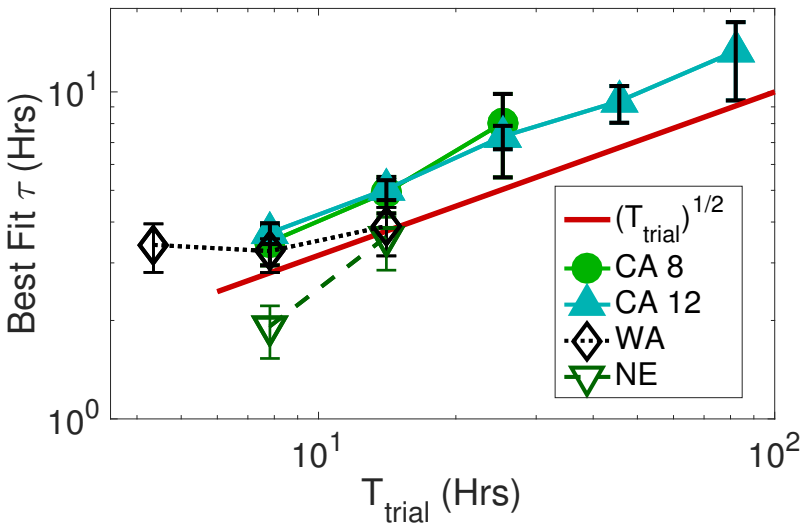

Figure 4.6: A plot of the best fit mean time to change ones opinion τ versus the trial time, with bands corresponding to 90% confidence intervals. The solid line is $(T_{\text{trial}})^{1/2}$ to guide the eye. We find that the rate at which opinions change depends strongly on the trial time.

wish to model why τ scales the way it does.

A simple model is to assume that τ changes for each individual as a random walk over the course of the trial with a reflecting boundary at $\tau = 0$. Simple math implies that $\langle\tau\rangle \sim (T_{\text{trial}})^{1/2}$. This argument, however, implies τ is not a constant for all jurors in a single case. We therefore compared models where τ was constant, and τ followed the random walk distribution, but find that both fit our data equally well, based on the log-likelihood ratio (not shown).Our data is therefore consistent with this argument. We can interpret the random walk as the amount of information each individual accumulates (remembers). The more information they have, the less easily they will change their opinion when faced with a contrary opinion. This mechanistic explanation is alike to a previous model, called the CCIS model [25],

which was found to strongly agree with empirical data, such as universal scaled-vote distributions, and long-range correlations between voters.

Finally, we explore in greater detail how α_{25} and α_{75} varies across datasets. Fig. 4.7 suggests that the rate individuals stop for innocent verdicts, α_{25}, is higher than the rate to to stop for guilty verdicts, α_{75}, and appears to decrease slightly with the trial time. We find, however, that $\alpha_{25} < \alpha_{75}$ in the OR 6 dataset, and there are too few cases in the CA 6 dataset to make any conclusion (not shown). The OR 12 datasets, however, (not shown) is consistent with the findings in Fig. 4.7. Whether $\alpha_{25} > \alpha_{75}$ or not appears to depend on whether the guilty verdicts take more or less time. These results may therefore help explain why guilty verdicts take longer to decide than innocent verdicts: jurors decide to stop more often when they know that the verdict will be guilty.

4.5 Extending Results To Judicial Panels

We find that many of our results can be extended to judicial panels in the ECHR. For example, the fraction of cases under deliberation as a function of time (Fig. 4.8) qualitatively resembles Fig. 4.1.

This is unexpected because the ECHR is composed of sitting judges from various countries in Europe, while the data we have explored so far is of juries in the US composed of randomly selected civilians. Furthermore, while juries typically deliberate continuously, judges in the ECHR deliberate on select dates that can be separated by several months. Finally, judges do not need a supermajority to reach

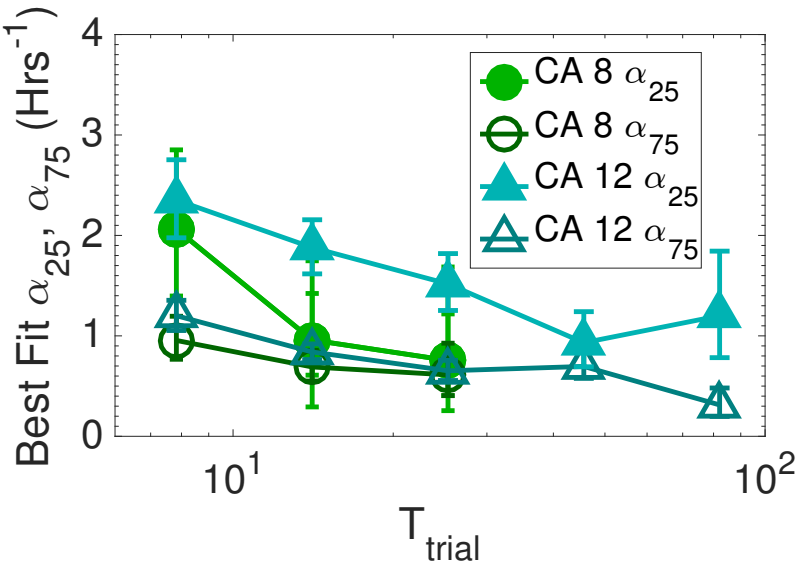

Figure 4.7: Plots of α_{25} and α_{75} across various datasets. We notice that, in this data, $\alpha_{25} > \alpha_{75}$, consistent with the idea that juries take longer to make a guilty verdict.

a decision, but instead only require a simple majority. Similar results may therefore suggests a common underlying mechanism of deliberation.

One limitation of this data, however, is that we only know the *days* that deliberation has occurred, rather than hours or minutes. We therefore assume that deliberation occurs 8 hours a day. We expect similar results for our models regardless of this assumption.

Another limitation is that, unlike jury cases, several counts against the defendant are brought before the ECHR, which are then ruled on. We cannot determine the length of time to rule on each count, although, to our surprise, we see negative correlations between the number of cases and the total amount of deliberation, which may need to be explored in the future. We included each count as a separate

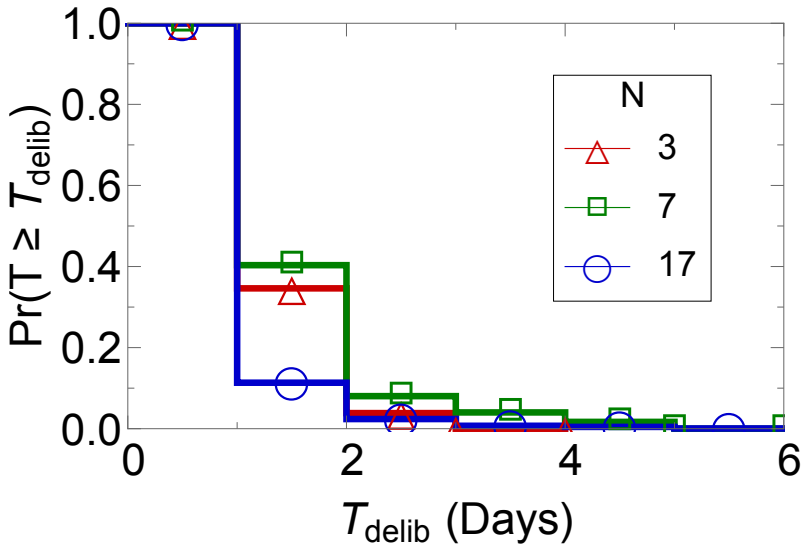

Figure 4.8: Example plots of the fraction of cases under deliberation as a function of time for the ECHR, with the number of judges (N) equal to three, seven, and seventeen, corresponding to a committee, chamber, and grand chamber, respectively [5]. We find that N can vary from three to roughly twenty three.

vote, with a time equal to the total deliberation time (in other words, we assume that judges deliberate on all counts at once, and that no single count takes up a significant share of the vote). Alternatively, we could have averaged the votes across all counts, but this ignores the fact that many counts in the same case were unanimously accepted or rejected, while an average would view those unanimous verdicts as non-unanimous decisions. Our method, however, is meant to directly compare the ECHR and jury data, where votes are discrete and are typically only for one count.

Finally, we do not have any data on the time judges were in a trial, which we found in Fig. 4.6 can strongly affect the dynamics.

Through a careful comparison of models to data (Table 4.2), we find that, in agreement with Table 4.1, the time distribution data is generally in disagreement with the models, unless the number of cases is very small. If there is greater agreement to models than Table 4.1, however, it would likely be because of the very coarse time binning, which is known to strongly affect the amount of information a distribution can contain [123].

Because judges only require a simple majority to make a ruling, we modified the MVM further such that groups stop deliberating when they have a simple majority, therefore $\alpha_{25} \rightarrow \alpha_I$ and $\alpha_{75} \rightarrow \alpha_G$. We further assume a split decision is an innocent verdict.

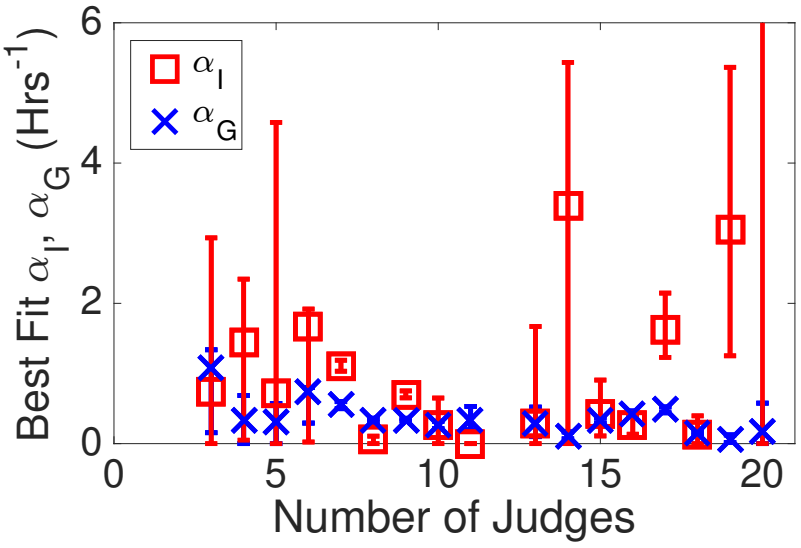

Figure 4.9: Plot of α_I and α_G for the ECHR dataset.

Alike to Table 4.1, we also see in Table 4.2 poor quantitative fits to the newly modified MVM. However, qualitatively, it fits far better than the null hypothesis we proposed, alike to Fig. 4.5 (not shown). Unlike Fig. 4.7, however, there is no

consistency in the values of α_I and α_G, probably because a relatively small share of votes are non-consensus.

4.6 Conclusion

The goal of this chapter was to better understand how competing ideas spread in groups using the time distribution for verdicts in criminal court cases.We find that models where individuals make decisions independent from each other disagrees with our data. On the other hand, models where individuals are influenced by each other agrees well, at least qualitatively, with our data, when we take into account the final vote of each individual.

We found a novel way to measure the influence individuals have on each other in a way that needs a minimal amount of interaction with the group itself. This is important because we cannot directly study the dynamics of real juries, but can, to our surprise, indirectly infer the dynamics by using the statistics of when they finish deliberating. Our findings also suggest that the deliberation time scales with $(T_{\text{trial}})^{1/2}$, which similarly affects how quickly individuals can change their opinion when deliberating with each other. When modeling this behavior, we find that the timescales for individuals to change their opinion scales similarly. This ideas of "stubbornness", where individuals are increasingly likely to keep their old opinion is seen in a previous model which found agreement with scaled vote distributions and spatial correlations between voters [25]. We therefore have more evidence that stubbornness can explain many statistical patterns in opinion dynamics.

Finally, our modeling results are in surprising agreement with data from the ECHR. This is notable because the ECHR in many ways breaks from the properties of juries, e.g., groups are not randomly assigned, and deliberation can occur over the course of many days, therefore any significant agreement between the datasets suggests common mechanics underly each.

There is a significant amount of future work, however. First, we appear to have evidence of a critical point in the dynamics of juries when less than 75% of individuals are in the majority. If this is a true critical point, we should see this behavior in similar scenarios, where a proxy for the control parameter, the final vote, makes the deliberation time, an order parameter, peak at a a critical point. Furthermore, the peak should vary strongly with system size, due to finite scaling effects. If this is a true critical point, then we should expect broad agreement in statistical patterns near that point, consistent with an opinion dynamics universality class.

Second, we need to find a model of the dynamics of juries that quantitatively agrees with data. There are a number of plausible steps forward, for example by modeling how the details of the trial may affect the deliberation. A simpler model may even be able to explain how deliberation time depends on the final vote (Fig. 4.4).

Finally, future research should extend our work to predict how opinions will change over time. Our work can only say what the statistics of should look like, rather than how each jury's decision will unfold. Being able to predict decisions will represent a great advancement in our understanding of opinion dynamics, because

we can directly test our model to data.

4.7 Acknowledgements

KB would like to thank Walter Fontana and the Santa Fe Institute for bringing to his attention related work in biology and engineering on the dynamics of failures. KB would also like to thank Nicholas Pace from the RAND corporation for helping procure the San Fransisco, California dataset. Finally, KB would like to thank HURIDOCS for providing the ECHR dataset.

#Judges	**N**	P-Values, Significant When $p > 0.1$, (# Parameters)					
		Time Distributions (# parameters)					Time & Vote
		Exp. Tail (1)	BMM: Exp. (1)	BMM: Gauss (2)	TT Null (2)	MVM (1)	Modified MVM (3)
3	22	-	-	-	-	-	0.09
4	9	-	-	-	-	-	**0.18**
5	6	-	-	-	-	-	0.05
6	25	-	**0.50**	**0.12**	**0.98**	**0.84**	**0.30**
7	1054	$< 10^{-3}$	2×10^{-3}		0.01	$< 10^{-4}$	$< 10^{-3}$
8	1342	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-3}$
9	2497	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-4}$	0.02	2×10^{-3}	$< 10^{-3}$
10	267	**0.46**	**0.48**	0.06	**0.61**	**0.14**	$< 10^{-3}$
11	67	**0.16**	**0.94**	**0.68**	**0.29**	**0.78**	0.03
12	43	-	**0.81**	**0.99**	**0.58**	**0.60**	**0.27**
13	21	-	-	-	-	-	$< 10^{-3}$
14	185	0.01	$< 10^{-3}$	$< 10^{-4}$	**0.72**	$< 10^{-4}$	$< 10^{-3}$
15	970	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-4}$	2×10^{-3}	$< 10^{-4}$	$< 10^{-3}$
16	2815	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-4}$	2×10^{-3}	$< 10^{-4}$	$< 10^{-3}$
17	1767	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-3}$
18	727	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-3}$
19	100	**0.79**	$< 10^{-3}$	6×10^{-4}	**0.13**	3×10^{-3}	$< 10^{-3}$
20	130	**0.57**	$< 10^{-3}$	$< 10^{-4}$	0.05	$< 10^{-4}$	$< 10^{-3}$

Table 4.2: P-values for ECHR models (see Appendix B.4 for details). "-" in cell: we have too few cases to determine p-values.

Chapter 5: Conclusion

In conclusion, we have explored how physics can help us model competing ideas in groups. Specifically, we have analyzed empirical data on macroscopic patterns of opinion formation in order to infer features of individual-level models.

First, we find that the number of answers users see in question answering boards can strongly affect both the predictability of user behavior, and the likelihood users will choose answers based on heuristics, such as whether an answer is listed at the top. The universality of this behavior across boards suggests a common mechanism, but the data cannot tell us whether this behavior is due to information overload, or if older questions tend to gather more heuristically driven users. Regardless of the mechanism, however, we find strong constraints in the power of crowdsourcing information as the number of answers to a question increases.

Next, we find that a simple model can explain the universal vote-share distribution in Europe and long range spatial correlations of voters in the US and Europe. The model is robust to variations in parameters, but two general principles are necessary for the model to fit the data: the contact network must be heavy tailed, and individuals must increasingly resist adopting alternative ideas (i.e., they exhibit stubbornness). The former principle is interesting because many social net-

works appear to be heavy tailed, e.g., the social networks of Twitter [139]. The inclusion of stubbornness in the model is motivated by work on pre-trial publicity (PTP) [78], where older information appears to be preferred over newer information, but we have no way of observing stubbornness using the data collected in Chapter 3.

To more directly measure stubbornness, and other modeling assumptions, we explored the dynamics of juries in the US and judicial panels in the ECHR. Using survival analysis, we find that the time (and vote) distributions of juries can determine which models are in agreement with our data, and, of those in agreement, which model fits best. Importantly, we rule out classes of models, such as those with exponential tail time distributions, in order to narrow down what models can plausibly explain our data. We find that a modified MVM can qualitatively explain our data well, including the vote and time distributions, in stark contrast to non-interacting null models, which perform poorly in both regimes. From fits of this model, we find that the rate an individual changes his or her opinion decreases with trial time, in agreement with our hypothesis of stubbornness made in Chap. 3. The model's agreement with data suggests that an interacting or influence model may better describe the mechanism of opinion formation in juries and judicial panels. This result helps address issues raised earlier that influence and independent decision making are indistinguishable [23, 134], although we cannot rule out that a more complicated model without influence could produce these same statistics.

It is clear that opinion dynamics could strongly affect dynamics across a diverse range of topics. Our goal in the future, however, is to better quantify the dynamics,

especially by experimental verification of the findings in Chap. 2 and better modeling in Chap. 4. We will begin answering what drives heuristically driven dynamics in question answering boards with an experiment in Amazon's Mechanical Turk, where we can independently vary attributes such as the number of answers, the answer order, and the number of votes each answer receives. In addition, we want to directly test how individuals change their opinion over time, and derive a model in quantitative agreement with jury and judicial data in Chap. 4.

Chapter A: CCIS: Modeling and Analysis

A.1 Fitting the CCIS Model to Data

In this section we describe in more detail how the CCIS model is fit to empirical vote distribution data and correlation data

A.1.1 Network Model

To match the model to data, we use a spatially distributed network, which creates a non-zero spatial correlation, and we find that we need a scale-free distribution to best match scaled vote distribution data. Adding both of these properties to a single network, however, is not just convenient, but realistic. For example, we could try to run a model on the most natural spatial network: a grid. In a grid, individuals only interact if they are spatially close, but previous work on the "six degrees of separation" between two randomly chosen individuals [140, 141] and "weak ties" between socially disparate individuals [142], suggests that ties can exist between individuals who are spatially separated by large distances. Furthermore, unlike grids, the degree distribution of many social networks is a power law [106].

Combining all these properties, we can create spatial scale-free networks, such

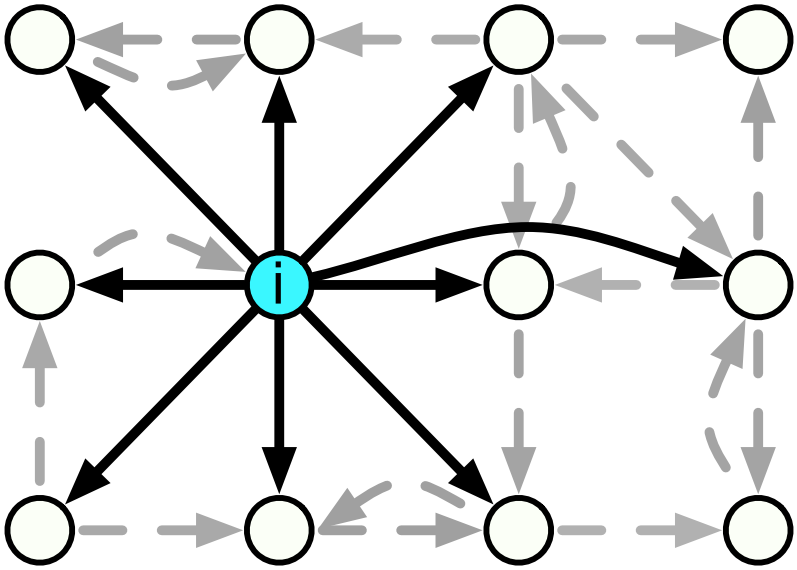

Figure A.1: A schematic of the network chosen to fit our models to empirical data. All nodes have a scale-free out-degree distribution whereby a node i with degree k_i (in this example, $k_i = 9$) is then connected to its nearest neighbors.

as the one in Fig. A.1. Nodes have an out-degree k chosen from a scale-free distribution, and are placed on a grid with unit distance. Each node is then connected to their nearest neighbors, although to test the robustness of our results, a fraction f of are randomly rewired. As f increases, the model makes similar fits to the vote distribution data but the spatial correlation decreases. To keep $\langle k \rangle$ constant for fixed degree distribution $p(k > k_{min})$, we change the proportion of nodes with degree k_{min} until we have the appropriate $\langle k \rangle$. The directed nature of the network reduces the chance of multi-edges or self loops, and it seems to be a reasonable assumption that people with a lot of connections broadcast their opinion to a wide audience without as much attention paid to the ideas of those same individuals.

A.1.2 Fitting Model Parameters

Next we discuss how our model is fit to data. The Poland 2005 data set is chosen due to the large number of elections (593, versus $\sim 200 - 400$ for other countries). In our simulations, seeded individuals are equally split among the various candidates, but variations in seeding should create similar results. Maximum Likelihood Estimation (MLE) is used to determine the appropriate seeding fraction.

The model has no readily apparent closed-form solution, and a Kernel Density Estimator for the model greatly over-estimates the probability for small x_i, therefore we approximate the probabilities with log-binned histograms (the widths, however, do not seem to change the best fit parameter significantly).

A.1.3 Parameter Values

In the FC model, only the candidate has an initial preference of whom to vote for, while in our model, we assume a set percentage of individuals have an initial preference to some candidate. The CCIS model creates a poorer fit when Q individuals are seeded (not shown), but seeding a fixed percentage seems to be an equally realistic assumption if we imagine that a small percentage of voters are initial strong supporters of the candidates.

We can also let the fraction seeded be 100%. Holding $\mu = 1$, the best fit β value is 0.65, with a fit similar to Fig. 3.2 (see Fig. 3.3). We choose to seed less than the total population, because it seems reasonable that at some starting point, not everyone is aware of the candidates.

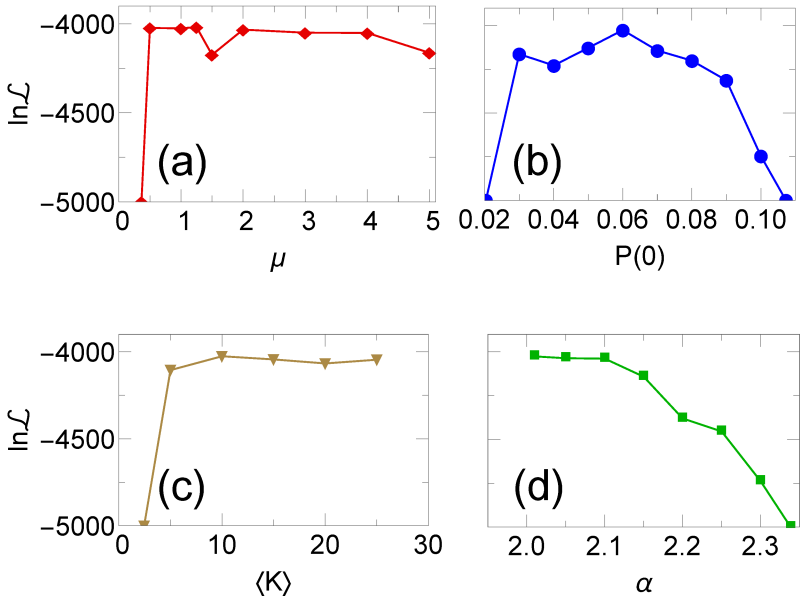

Figure A.2: The log-likelihood function versus (a) μ, (b) the fraction of individuals seeded, $P(0)$, (c) $\langle k \rangle$, and (d) α. Not shown in (d) is the log-likelihood of a 10-regular spatial graph (-10771), which is far below the current y-axis scale. Arrows indicate the chosen values for our fit. $\mathcal{L}$ varying by less than 100 does not look appear visually different from our fit.

To fit our model to the distributions, we set β to 0.1, μ to 1, and $\langle k \rangle$ to 10, but variations in these values do not significantly affect our results (see Fig. A.2 in Appendix A.1, where we hold all parameters fixed, except for the given parameter plotted). We also fix $\delta = 0$ in order for the distribution to remain fixed in equilibrium. The MLE for alpha, however, varies depending on the type of network chosen. For example, while $\alpha = 2.01$ creates a good fit with our current model network (spatially-extend scale-free), $\alpha = 2.5$ creates a good fit on an undirected scale-free network with no spatial structure. Whatever the optimal α, however, we find that a wide distribution (e.g., $\alpha < 3$) works best, when fitting to data. A Poisson or k-regular graph, for example, never appears to fit well with data, regardless of the other parameter choices.

We have more freedom to vary all parameters if our only goal is to create vote correlations similar to empirical data (Fig. 3.5). The roughly logarithmically decreasing correlation with distance is observed for many values of $\delta \geq 0$, $\mu \geq 0$, $\beta > 0$, $P(0) > 0$ and $\alpha > 2$. Just one example are the parameters chosen in Fig. 3.5.

A.1.4 Determining The Spatial Correlation

We finally mention how the correlations in Fig. 3.5 are calculated. To be consistent with previous work [3,21] and Fig. 3.5a, we define the normalized correlation in our figures as:

$$C(r) = \frac{\langle P_i^{(1)} P_j^{(1)} | d_{ij} \approx r \rangle - \langle P_i^{(1)} \rangle^2}{\sigma^2_{P^{(1)}}}, \tag{A.1}$$

in which $P_i^{(1)}$ is the fraction of voters for candidate 1 within a small region (which we choose to be 5×5 node squares), $\langle P^{(1)}\rangle$ is the average fraction of voters for candidate 1, and $\sigma^2_{P^{(1)}}$ is the variance in vote distribution across all regions. $\langle P_i^{(1)}P_j^{(1)}|d_{ij}\approx r\rangle$ is the 2-point correlation function between regions whose centroid is a distance $r\pm1/2$ from each other.

A.2 Derivation of the Transport-Like Approximation

In this appendix, we use the TLA to understand the initial jump in the opinion densities (see Figs. 6 & 12).

Our model can be described by the following equation in the mean field:

$$(\partial_t+\partial_\tau)\rho^{(A)}(t,\tau) = -\delta\rho^{(A)}(t,\tau)-\Theta(1-\tau\mu)(1-\tau\mu)\beta k\rho^{(A)}(t,\tau)\sum_{B\neq A}P^{(B)}(t), \quad \text{(A.2)}$$

with the boundary conditions:

1. $\rho^{(A)}(t,\infty)=0$,
2. $\rho^{(A)}(t,0)=\delta(0^+)[\beta kP^{(A)}(t)\tilde{P}(t)+\beta kP^{(A)}(t)\sum_{B\neq A}\int_0^{\mu^{-1}}(1-\tau'\mu)\rho^{(B)}(t,\tau')d\tau']$,
3. and $\rho^{(A)}(0,\tau)=f(\tau)$,

(see Fig. A.3 for a visual representation) where t is time, τ is the time an individual has had their most recent opinion, $P^{(X)}$ is the fraction of individuals with opinion X at time t, $\rho^{(X)}$ is the density of individuals with opinion X at time t who have kept their opinion for a time τ (variables and parameters are also defined in Table I). Finally, $\tilde{P}(t)=1-\sum_X P^{(X)}(t)$. The right hand side of the equation describes

the ability of individuals to recover as well as the ability to change opinions. We can interpret the boundary conditions as:

1. Normalizability

2. An increase in the infection density due to neutral neighbors and opinionated neighbors.

3. Initial conditions

We focus on the simpler case of $\delta = 0$ for our analysis because adding δ to the equation numerically does not seem to affect consensus, while, in simulations, consensus happens quickly. Future analysis of perturbations around equilibrium, however, may give us better insight into what happens in simulations. We do know, however, that when $\delta \ll 1$, the equation can be simplified to the one seen in [98], were they find, to use our notation, $T_{cons} \sim \delta^{-1}$, in agreement with our own simulations (not shown).

The simplified equation is:

$$(\partial_t + \partial_\tau)\rho^{(A)}(t,\tau) = -\Theta(1-\tau\mu)(1-\tau\mu)\beta k \rho^{(A)}(t,\tau) \sum_{B \neq A} P^{(B)}(t), \tag{2}$$

with the same boundary conditions.

We first try to understand the transient "jump" in the fraction of individuals following a given opinion on a timescale that is in many cases much smaller than the time to reach consensus. We wish to understand the the equilibrium fraction of individuals with a given opinion, and the time to reach equilibrium.

We find strong agreement between theory and simulations for the equilibrium fraction of individuals in each opinion (Fig. 3.6), especially when $\langle k \rangle \geq 10^2$. For fixed networks with $\langle k \rangle < 10^2$, the equilibrium values are on average below theoretical values, plausibly because individuals are less connected to their neighbors, and thus less influenced by them, than the mean field theory assumes. To find agreement with simulations, we numerically determined equilibrium values by stepping forward the equation using the forward Euler method.

This method is inherently sensitive to the timestep width, Δt, especially when $\Delta P_{eq} \approx 0.5$, therefore we find the equilibrium value can be more accurately determined by varying the timestep width and, via linear regression, determining the asymptotic limit for the equilibrium as $\Delta t \to 0$ (Fig. A.4). This seems to reduce our statistical error to less than 0.5% compared to as much as $1 - 6\%$, and is in excellent agreement with the simulations.

Next, we determine the time to reach equilibrium. We discretize τ, following [98], to derive a set of equations that we linearize around a fixed point to determine the scaling of the transient time (Eq. A.10 & A.11). Our approximations are only accurate for $\mu \ll 1$, but seem to be qualitatively similar to numerical data for $\mu \sim O(1)$. We define the following macroscopic variables:

$$P^{(1)}(t) = \sum_{\tau'} \rho^{(1)}(t, \tau'), \ \textit{and} \ P^{(2)}(t) = \sum_{\tau'} \rho^{(2)}(t, \tau'), \tag{A.3}$$

in which $\sum_{\tau'}$ is shorthand for $\sum_{\tau'=0}^{\infty}$. If we let $\Omega[|]$ be the conditional probability

function, and $\dot{x} \equiv \frac{d}{dt}x$, then Eq. 3.2 becomes (for $\tau > 0$):

$$\begin{aligned} &\dot{\rho}^{(1)}(t,\tau) \\ &= \Omega[\rho^{(1)}(t,\tau)|\rho^{(1)}(t,\tau-1)]\rho^{(1)}(t,\tau-1) - \rho^{(1)}(t,\tau) \end{aligned} \tag{A.4}$$

Expanding these variables out, we find that:

$$\dot{\rho}^{(1)}(t,\tau) = (1+\beta k\{[P^{(1)}(t)+\mu(\tau-1)]P^{(2)}(t)-1\})\rho^{(1)}(t,\tau-1) - \rho^{(1)}(t,\tau), \tag{A.5}$$

and for $\tau = 0$:

$$\dot{\rho}^{(1)}(t,0) = \beta k P^{(1)}(t)[P^{(2)}(t) - I^{(2)}(t)] - \rho^{(1)}(t,0). \tag{A.6}$$

With an equivalent set of equations for $\rho^{(2)}(t,\tau)$ and

$$I^{(1)}(t) = \sum_{\tau'} \mu\tau'\rho^{(1)}(t,\tau'),\ \ I^{(2)}(t) = \sum_{\tau'} \mu\tau'\rho^{(2)}(t,\tau'). \tag{A.7}$$

From the above results we can sum $\rho^{(1)}(t,\tau)$ to find the equations for the macroscopic variables:

$$\dot{P}^{(1)}(t) = \beta k[I^{(1)}(t)P^{(2)}(t) - I^{(2)}(t)P^{(1)}(t)]. \tag{A.8}$$

To lowest order in μ, we also find that:

$$\begin{aligned} \dot{I}^{(1)}(t) \approx\ & \mu(1-\beta k)P^{(1)}(t) \\ & +\beta k[\mu P^{(1)}(t)^2 + I^{(1)}(t)P^{(1)}(t) - I^{(1)}(t)]. \end{aligned} \tag{A.9}$$

These equations are solvable by expanding around the solution $P^{(1)} = P^{(2)} = 1/2$ and $I^{(1)} = I^{(2)} = \mu[1/(\beta k) - 1/2]$ to first order. The resulting largest eigenvalue is

$$\begin{aligned} &\lambda_1 \approx \\ &\frac{-\beta k - 4\mu + 2\beta k\mu + \sqrt{16\beta k\mu + (\beta k + 4\mu - 2\beta k\mu)^2)}}{4}. \end{aligned} \tag{A.10}$$

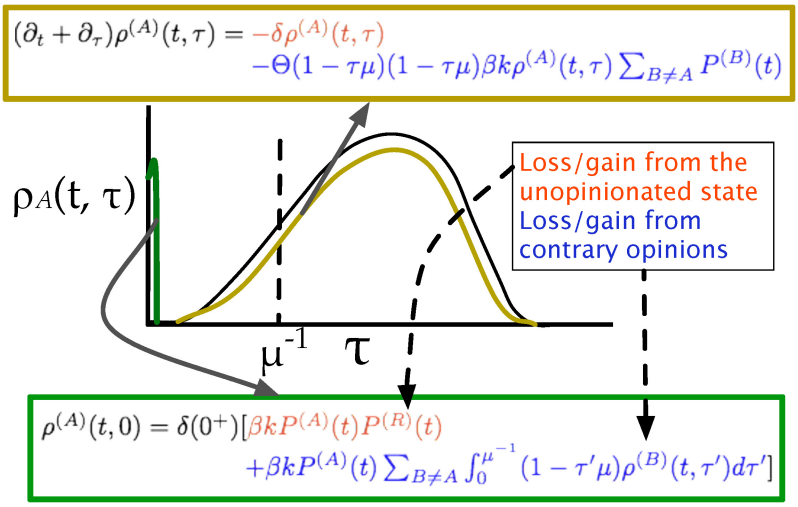

Figure A.3: Schematic of the scalar variable in Eq. A.2 as a function of time, t, and time since opinion adoption, τ.

The time to reach equilibrium, T_{eq}, is:

$$T_{eq} = \nu\frac{\log(N)}{\lambda_1}, \tag{A.11}$$

where ν is a fitting parameter found to be 1.26 ± 0.04 from simulations. When $\beta k = 1$, this eigenvalue should agree exactly with the value cited previously [98], but we find disagreement by an overall prefactor of $1/4$ which, at least to our knowledge, may have been missed in the previous work. Figure A.5 shows how simulations agree with theory. We notice disagreement is most significant when μ approaches 1, and β is small (e.g., $\beta = 0.1$).

A.3 Scaling of effective network size

In this appendix we derive Eq. 3.4 using a FPA, which is distinct from the TLA in Appendix A.2. Our derivation is heavily based on the derivation of consensus

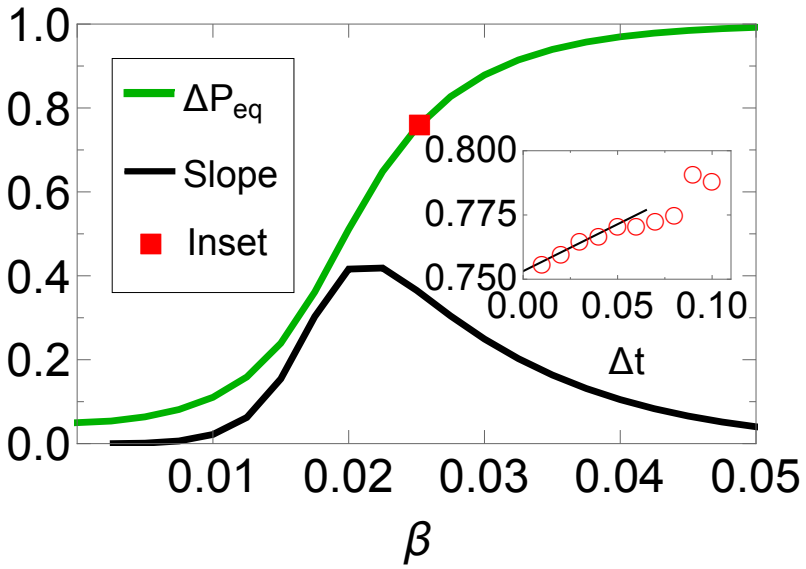

Figure A.4: Details regarding the theory curve of Figure 3.6. (Inset) For each value of β, we vary the timestep width for Eq. 3.2 (Δt), and find the resulting equilibrium value. $\Delta P_{eq}(\Delta t \to 0)$ is estimated via linear regression. (Main figure) Plotting $\Delta P_{eq}(\Delta t \to 0)$ and slope for $\Delta P(0) = 0.05$, we find the slope, seen in the inset, is greatest when $\Delta P_{eq} \approx 0.6$, implying the error from using a single value of Δt would have been largest in this range.

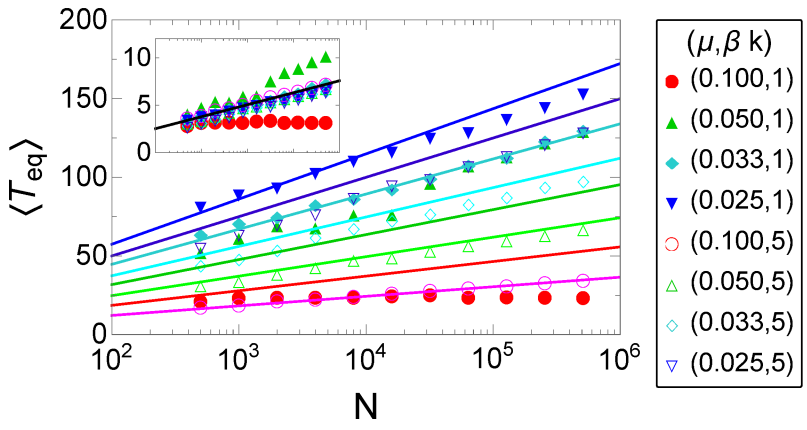

Figure A.5: T_{eq} versus N for various μ and βk (simulations on k-regular random graphs,with $k = 10$). Inset shows collapse when T_{eq} is rescaled by λ_1, with the best fit slope equal to ν in Eq. A.11.

times for the VM and Invasive Process by Sood, Antal, and Redner [10].

A.3.1 Derivation

We use the same conventions as in that paper, except the transition probability scaling factor, S, the degree distribution, $p(k)$, and associated moments, $\langle k^m \rangle = \sum_k p(k) k^m$. Note that we assume for now that $\beta < 1/\langle k \rangle$ for the inward spreading process (opinions spread inward toward an individual).

Let $\eta(x)$ be the state of a node x on a network with adjacency matrix A_{xy} and order N. Assuming 2 opinions and that μ and $\delta \to 0$, we have a two-state system. Using the conventions of Sood, Antal, and Redner [10] the opinions of the two-state system are "0" or "1" instead of "1" or "2". We stress that the 0 state *is*

an opinionated state. Lastly η_x is the state of the system after changing a node x:

$$\eta_x(y) = \begin{cases} \eta(y), & y \neq x \\ 1 - \eta(x), & y = x \end{cases}. \tag{A.12}$$

The transition probability at node x is therefore:

$$\mathbf{P}(\eta \to \eta_x) = \sum_y \frac{A_{xy}}{NS}[\Phi(x,y) + \Phi(y,x)], \tag{A.13}$$

in which:

$$S = \beta^{-1}, \tag{A.14}$$

and

$$\Phi(x,y) = \eta(x)[1 - \eta(y)]. \tag{A.15}$$

We further assume a mean field solution, in which the adjacency matrix becomes the average adjacency matrix:

$$A_{xy} \to \langle A_{xy} \rangle = \frac{k_x k_y}{\langle k \rangle N}. \tag{A.16}$$

Instead of individual states $\eta(x)$, we can instead focus on ρ_k, the density of states with degree k:

$$\rho_k = \frac{1}{N} \sum_{x'} \eta(x'). \tag{A.17}$$

Here, x' is the sum of all nodes with degree k. To clarify the below equations, we also define a variable ω:

$$\omega = \frac{1}{N\langle k \rangle} \sum_x k_x \eta(x) = \frac{1}{\langle k \rangle} \sum_k k p(k) \rho_k. \tag{A.18}$$

Next, we define our raising and lowering operators for ρ_k, which defines the probability of increasing or decreasing ρ_k by a small increment:

$$\rho_k \to \rho_k^{\pm} \equiv \rho_k \pm \delta\rho_k, \tag{A.19}$$

in which

$$\delta\rho_k = \begin{cases} \frac{N\langle k^2\rangle}{S\langle k\rangle p(k)} & Outward\ Process \\ \frac{N\langle k\rangle}{Sp(k)} & Neutral\ Process \\ \frac{Nk}{Sp(k)} & Inward\ Process \end{cases} \tag{A.20}$$

The change in ρ_k is proportional to the probability that an individual of degree k changes his or her opinion in a given time step. This probability scales as $\frac{\langle k^2\rangle}{\langle k\rangle}$, $\langle k\rangle$, and k for the outward, neutral, and inward processes, respectively.

The raising operator is defined as:

$$\mathbf{R}_k = \mathbf{P}(\rho_k \to \rho_k^+) = \sum_{x'}\sum_{y} \frac{k_{x'}k_y}{S\langle k\rangle N^2}\Phi(y,x). \tag{A.21}$$

With simplification, this yields

$$\mathbf{R}_k = \frac{\omega}{S}p(k)k(1-\rho_k). \tag{A.22}$$

Similarly, for the lowering operator:

$$\mathbf{L}_k = \mathbf{P}(\rho_k \to \rho_k^-) = \frac{\rho_k}{S}p(k)k(1-\omega). \tag{A.23}$$

The exit probability, ξ_1, defined as the probability for all nodes to reach state one in equilibrium, is the same for all cases

$$\xi_1 = \langle\rho\rangle \equiv \sum_k \rho_k p(k), \tag{A.24}$$

and similarly, that $\langle\rho\rangle$ (or magnetization, if this were a spin system) is a conserved quantity. The reason is because

$$\begin{aligned} \langle\Delta\eta(x)\rangle &= [1-2\eta(x)]\mathbf{P}(\eta\to\eta_x) \\ &= [1-2\eta(x)]\sum_y \frac{A_{xy}}{NS}[\Phi(x,y)+\Phi(y,x)], \end{aligned} \tag{A.25}$$

$$\Delta\langle\rho\rangle = \sum_x \langle\eta(x)\rangle =\sim \sum_{x,y}[\eta(x) - \eta(y)], \tag{A.26}$$

which is trivially 0. We note that this argument is exact (not a mean field approximation) and is independent of the method in which opinions spread (at least, again, assuming $\beta < 1/\langle k\rangle$ for the inward dynamics). The time to consensus is

$$\begin{aligned} T_{cons}(\{\rho_k\}) &= \sum_k \Delta t_k \\ &+[\mathbf{R}_k(\{\rho_k\})T_{cons}(\rho_k^+) + \mathbf{L}_k(\{\rho_k\})T_{cons}(\rho_k^-)]\cdot \\ &+[1 - \sum_k \mathbf{R}_k(\{\rho_k\}) + \mathbf{L}_k(\{\rho_k\})]T_{cons}(\{\rho_k\}) \end{aligned} \tag{A.27}$$

The average number of interactions per timestep is:

$$\Delta t_k = \begin{cases} p(k)\frac{\langle k^2\rangle}{\langle k\rangle SN} & Outward\ Process \\ p(k)\frac{\langle k\rangle}{SN} & Neutral\ Process \\ \frac{k}{SN} & Inward\ Process \end{cases} \cdot \tag{A.28}$$

We expand to second order in $\Delta\rho_k$ and find that

$$\sum_k v_k \frac{\partial T_{cons}}{\partial \rho_k} + D_k \frac{\partial^2 T_{cons}}{\partial \rho_k^2} = -1, \tag{A.29}$$

in which

$$v_k \equiv \frac{\Delta\rho_k}{\langle\Delta t\rangle}(\mathbf{R}_k - \mathbf{L}_k) \to 0. \tag{A.30}$$

As is shown in in the original voting model paper [10], this value reaches 0 for time $T_{cons} \sim O(1)$ which is much smaller than the next term:

$$D_k \equiv \frac{(\Delta\rho_k)^2}{\langle\Delta t\rangle}\frac{(\mathbf{R}_k + \mathbf{L}_k)}{2}. \tag{A.31}$$

A change of variables implies that:

$$\frac{\partial T_{cons}}{\partial \rho_k} = \frac{\partial T_{cons}}{\partial \rho}\frac{\partial \rho}{\partial \rho_k} = p(k)\frac{\partial T_{cons}}{\partial \rho}, \tag{A.32}$$

therefore

$$\sum_k \frac{M}{2\langle k\rangle N S^2}(\omega + \rho_k - 2\omega\rho_k)p(k)\frac{\partial^2 T}{\partial \rho^2} = -1. \tag{A.33}$$

in which

$$M = \begin{cases} \langle k^2\rangle & Outward\ Process \\ \langle k\rangle^2 & Neutral\ Process \\ k^2 & Inward\ Process \end{cases}. \tag{A.34}$$

This can be made into a more compact form, noting that ρ is conserved and $v_k \to 0$, $\omega \to \rho$. :

$$\frac{\rho(1-\rho)}{N_{eff}}\frac{\partial^2 T_{cons}}{\partial \rho^2} = -1.$$

[Eq. 3.3 from the main text], where N_{eff} is as follows:

$$N_{eff} = \frac{NS^2}{\langle M\rangle} = \begin{cases} \frac{N}{\beta^2\langle k^2\rangle} & Outward\ Process \\ \frac{N}{\beta^2\langle k\rangle^2} & Neutral\ Dynamics \\ \frac{N}{\beta^2\langle k^2\rangle} & Inward\ Dynamics \end{cases}.$$

[Eq. 3.4 from the main text].

We find that this equation simplifies down to (24) in [10], noting the boundary condition, $T_{cons}(0) = T_{cons}(1) = 0$ in which:

$$T_{cons}(\rho) = N_{eff}\left[(1-\rho)ln\frac{1}{1-\rho} + \rho ln\frac{1}{\rho}\right], \tag{A.35}$$

implying that $T_{cons} \sim N_{eff}$.

As we discuss shortly, if $\beta > 1/\langle k\rangle$ in the inward-spreading case, we have VM dynamics, and the mean field consensus time replaces β with $1/\langle k\rangle$. Furthermore, this approximation breaks down for small $\langle k\rangle$ and small β, in which we show in

Section 3.5 that the consensus time scales as β^{-1}. Future work could improve the accuracy of the current results with a pair approximation theory [109, 110].

This paper mainly focuses on the outward process, but we have also compared theory and simulation for the other processes by varying β and $\langle k^2 \rangle$, in Poisson and scale-free networks, while setting δ and μ to 0. First, we observe the dependence on $\langle k^2 \rangle$ by simulating the models on scale-free networks.

In a scale-free network, $\langle k^2 \rangle$ diverges with network order, N:

$$\langle k^2 \rangle \sim \begin{cases} N^{3-\alpha} & \alpha < 3 \\ \log(N) & \alpha = 3 \\ O(1) & \alpha > 3 \end{cases} . \tag{A.36}$$

Therefore, for outward and inward dynamics:

$$T_{cons} \sim \begin{cases} O(1) & \alpha < 2 \\ \log(N)^2 & \alpha = 2 \\ N^{2(\alpha-2)/(\alpha-1)} & 2 < \alpha < 3 \\ N/\log(N) & \alpha = 3 \\ N & \alpha > 3 \end{cases} . \tag{A.37}$$

A.3.2 Agreement With Simulations

Fig. A.6 compares outward process simulations to the FPA (inward processes simulations are similar, due to the equilvalent scaling). Although a finite size transient impedes this scaling behavior for $N \leq 10^4$, we still see agreement for large

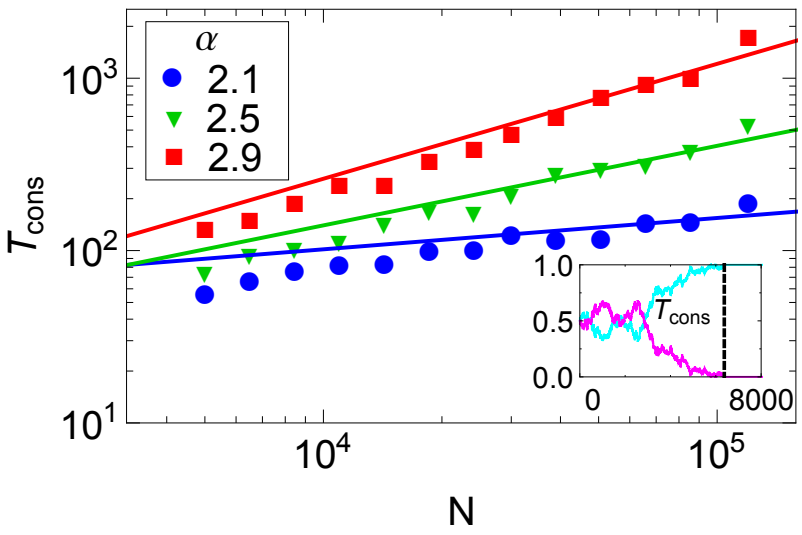

Figure A.6: Mean consensus time, T_{cons}, for scale-free networks with $\beta = 0.5$, $\delta = 0$, and $\mu = 0$. Inset is one example of consensus with $P^{(1)}(t)$ and $P^{(2)}(t)$. Using the FPA, the expected fit (solid lines) is Eq. 3.4. Simulations are averaged over 10 networks (30 networks for $3 \times 10^4 \leq N < 10^5$, and 20 networks for $N = 1.2 \times 10^5$) with 100 trials per network.

enough networks. For Poisson networks, we see $T_{cons} \sim \langle k^2 \rangle^{-1} = (\langle k \rangle^2 - \langle k \rangle)^{-1}$ in the inset of Fig. A.7.

The inward-spreading dynamics closely parallel the outward spreading dynamics when $\beta^2 \langle k^2 \rangle < 1$. On the other hand, when β is large enough, each node is, on average, infected by multiple nodes at each timestep, although, by the end of the timestep, only one opinion is chosen. This maps exactly onto the VM, and therefore so does the consensus time (Fig. A.8). Setting the model's mean field consensus time equal to the VM consensus time implies that $\beta_c = \langle k \rangle^{-1}$ is the critical value between CCIS and VM dynamics [1]. Neutral spreading (not shown), on the other

[1] We should mention that more accurate methods for determining the mean consensus time exist

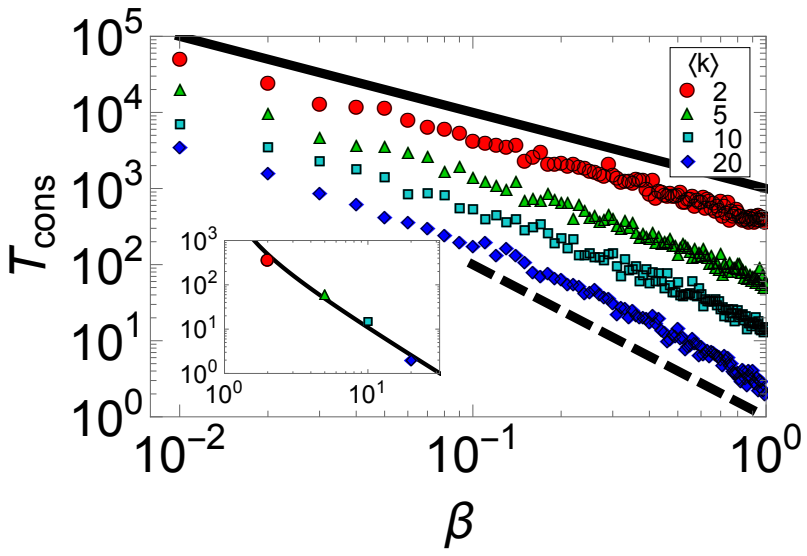

Figure A.7: Mean consensus time for varying β (δ and $\mu = 0$) on 1000 node Poisson networks with different average degree, $\langle k \rangle$. Inset: consensus time versus average degree for $\beta = 0.99$. Simulations are averaged over 30 networks. The theory are the dashed lines ($T_{cons} \sim \beta^{-2}$ when β and $\langle k \rangle$ large, and $T_{cons} \sim \beta^{-1}$ in the opposite limit) and solid line in the inset ($T_{cons} \sim \langle k^2 \rangle^{-1}$).

hand, breaks with the other spreading methods by only depending on the first degree moment, and is therefore mostly independent of the network's degree distribution in the mean field.

Finally, we check whether $T_{cons} \sim \beta^2$ for each process (Figs. A.7 and A.8). Agreement with theory is closest when $\beta \sim O(1)$ and $\langle k \rangle \sim 10 - 20$. When $\langle k \rangle$ approaches 1 or $\beta \ll 1$ we see that $T_{cons} \sim \beta^{-1}$. The reason is as follows: the number of nodes convinced at each timestep in this limit is very low (i.e., 2 with probability $\beta^2 \approx 0$, 1 with probability β, and 0 with probability $1 - \beta$), therefore, the time until a given node is convinced is a geometric process:

for the VM, as explained further in [109, 110].

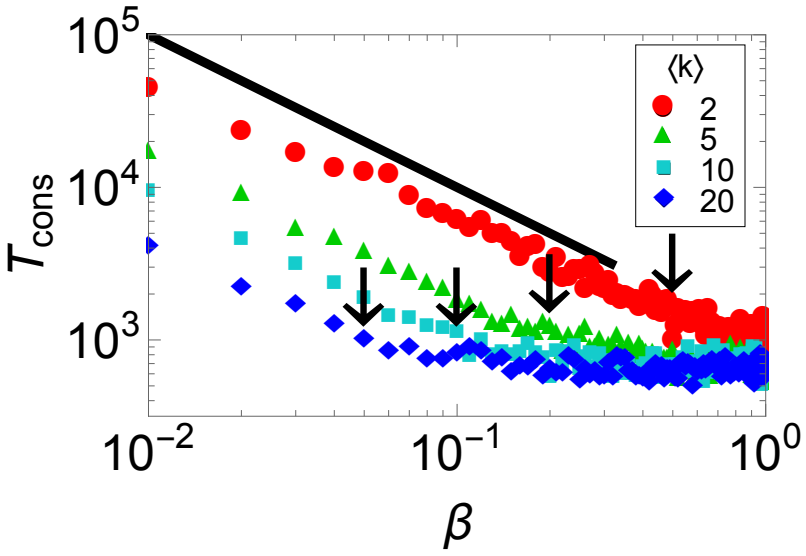

Figure A.8: Mean consensus time versus β for Poisson graphs in which we use the inward infection method. Theory is the black line $T_{cons} \sim \beta^{-1}$ and arrows indicate when $\beta\langle k\rangle = 1$, whereby we transition to true VM dynamics, which is independent of β.

$$p(t) = \beta(1-\beta)^{t-1}, \tag{A.38}$$

which would imply the average time until a node is convinced is β^{-1}. The consensus time would scale similarly.

Chapter B: Jury Data: Collecting, Parsing, and Modeling

In this section, we describe how we found our data, how it was parsed and split, and how it was compared to models.

B.1 Gathering Data

Data from Oregon and California juries came from the Inter-university Consortium for Political and Social Research [11,14]. Data from Nebraska and Washington juries was used with permission from the authors of the studies [12,13].

B.2 Cleaning Data

After we gathered data, we found that each set of data had to be cleaned in a different way.

For the Oregon, Washington, and Nebraska data, we discovered some case data was redundant. Namely we had information on multiple jurors who were in the same cases, so we only recorded the first instance for each case number. Furthermore, in the Oregon data, the times individuals were deliberating was not always recorded, so we removed these cases (3 cases in the OR 6 data, and 10 cases in the OR 12 data, or $< 2\%$ of cases in the respective datasets).

For the California data, there were some complications when parsing data. Namely, some deliberation times were recorded in days (which we did not record), and others in hours (which we did). Furthermore, some cases did not record the trial time or the final vote, and were therefore removed because both case attributes are known to strongly affect the deliberation time. In total, we removed 3184 cases in this manner. Furthermore, in 124 of the cases, multiple votes were cast to decide whether the defendant was innocent or guilty. We found these cases had statistically significantly higher deliberation times ($\approx$ 5 hours instead of the more typical 3-4 hours). We therefore removed those from analysis as well. We find, however, that this particular data subset suggests one vote to the next varies fairly continuously, and in rarer instances, votes can flip from a mostly guilty verdict to an innocent one. These properties were not captured in our models, because we were unsure of how universal these properties were. Among the cases that were deemed applicable, we noticed that an unusually large proportion of cases were marked "12 hours" compared to 10, 11, 13 hours, etc. This may imply that some approximately day-long cases were labeled 12 hours, and therefore, we havw no idea of the true deliberation time. For this reason, we removed all 12-hour cases from our data. We believe this may have been a conservative assumption, but in not removing these cases, we believe the dataset would have been poorer.

Data from the ECHR was gathered by HURIDOCS [143], spanning the beginnings of the court in 1959 until roughly 2014. The data was a 1.7 GB CSV file seemingly without a consistent tag to denote the beginning of one document or end of another. We therefore used common key words to determine the start of a docu-

ment, and found the number of judges, final votes, and dates of deliberation, while being mindful to only capture this metadata within this document. To complicate matters further, some cases were written entirely in french, which was a language unfamiliar to the authors. Therefore we used a few keywords found in french documents, but by no means was this search exhaustive. In total, we were successfully able to parse 12,076 cases, or approximately 40% of all cases.

B.3 Splitting Data

Given the various attributes, we had to decide how to best split up our data. If we split our data too finely, e.g., all cases with a trial time of 1 day, 2 days, 3 days, etc., we risked having too little data to be useful. If we coarsely split our data, then we risk modeling the several heterogeneous processes. To find the best solution, we first split by the number of judges/juries.

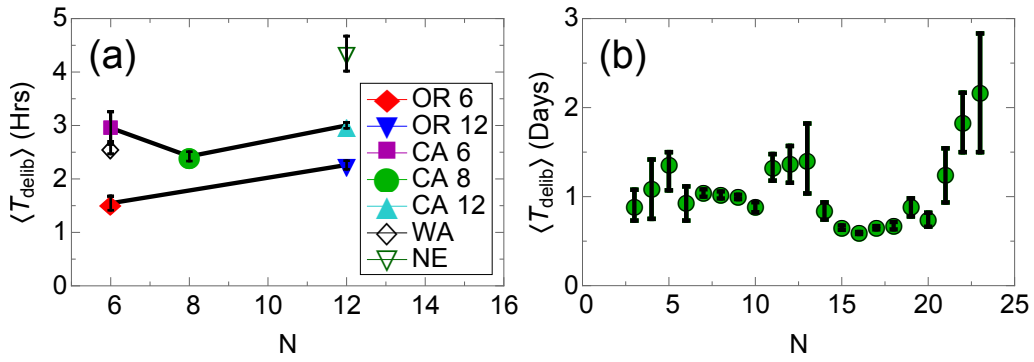

Figure B.1: Mean deliberation time versus the number of (a) jurors or (b) judges. Error bars correspond to 90% confidence intervals. We notice little dependence on N, in contrast to many models [6–9]/

As seen in Fig. B.1a, for juries and, and Fig. B.1b for the ECHR, this has

some effect on the deliberation time. It is notable that the deliberation time does not increase strongly with the number of individuals, in contrast to, e.g., [7,8,69,98,121, 125]. In these models, typically timesteps are normalized such that N interactions happen in one unit time. Therefore, a constant deliberation time here would mean, the consensus time in models would have to decrease as $1/N$, where N is the number of individuals, which is a prediction made in [25] for all-to-all networks.

As Fig. 4.3 in the main text shows, we similarly find a strong dependence on the trial time. We therefore split up data with logarithmic bins. We fairly arbitrarily chose bin sizes of between $1 - 1.8$ hours, between $1.8 - 1.8^2$ hours, etc., such that the data was fairly evenly spread out.

B.4 Comparing Models To Data

To fit models to data, we used maximum likelihood estimation, which is a first-principles approach to finding the optimal point estimate of parameter values. We then bootstrapped data 1000 times (300 in the case of the ECHR modified MVM model) to find confidence intervals in these estimates.

To determine the exponential tail, however, I first find the bin that minimizes the Kolmogorov-Smirnov (KS) test between the exponential distribution and the data, based on previous work on distributions in empirical data [122, 123]. This is the minimim bin, after which the data is fit to an exponential.

We test how well the models agree with data by bootstrapping the estimated distribution, and finding how often the KS test between the bootstrapped model,

and the corresponding fitted model of the bootstrap data, is larger than the KS test between the data and the model, following the methods of [122, 123]. If the KS test deviates more from from data than our model in $< 10\%$ of the time, then we rule out this model. Usually one uses 5% significance to rule whether data is statistically significant, but we are trying to answer a slightly different question: whether the data tends to agree with a particular model. In this case, 10% is high enough that we do not risk creating too many false positives. When incorporating both vote and time distributions, we use a two-dimensional version of the KS test, based on [144], with code adapted from [145].

If p-values are high, however, this test fails to distinguish which model fits the data better. To find the best model, we use the log-likelihood ratio test, where the log-likelihood ratios are normalized in the following way:

$$\mathcal{R} \to \mathcal{R}/\sigma N^{-1/2} \tag{B.1}$$

where σ is the standard deviation of $\log(p(x_i|\lambda))$, the log of the probability that a random valuable would be x_i in the model, and N is the number of data points. For large N, this approaches a normal gaussian distribution. If the log-likelihood ratio deviates from 0 more than 1.96 standard deviations ($p < 0.05$) then we can say that one of the two models matches the data statistically significantly better than the other. In our data, however, we could not always distinguish which model fit better, potentially because the data was so strongly binned.

Bibliography

[1] A. Chatterjee, M. Mitrovic, and S. Fortunato. Universality in voting behavior: An empirical analysis. *Sci. Rep.*, 3(1049), January 2013.

[2] S. Fortunato and C. Castellano. Scaling and universality in proportional elections. *Phys. Rev. Lett.*, 99(13):138701, September 2007.

[3] C. Borghesi, J. Raynal, and J. Bouchaud. Election turnout statistics in many countries: Similarities, differences, and a diffusive field model for decision-making. *PLoS One*, 7(5):e36289, May 2012.

[4] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86(14):3200–3203, April 2001.

[5] 2016.

[6] Thomas Liggett. *Interacting Particle Systems*. Springer-Verlag, New York, NY, 1985.

[7] P. Clifford and A. Sudbury. A model for spatial conflict. *Biometrika*, 60(3):581–588, March 1973.

[8] R. A. Holley and T. M. Liggett. Ergodic theorems for weakly interacting infinite systems and the voter model. *Ann. Probab.*, 3(4):643–663, August 1975.

[9] Naoki Masuda, N. Gibert, and S. Redner. Heterogeneous voter models. *Phys. Rev. E*, 82:010103(R), 2010.

[10] V. Sood, T. Antal, and S. Redner. Voter models on heterogeneous networks. *Phys. Rev. E*, 77(4):041121, April 2008.

[11] Terence Dunworth and Nicholas Pace. *JURY VERDICTS DATABASE FOR COOK COUNTY, ILLINOIS, AND ALL COUNTIES IN CALIFORNIA, 1960-1984.* RAND Corporation, Santa Monica, CA, 1992.

[12] John Gastil, E. Pierre Deess, and Phil Weiser. Civic awakening in the jury room: A test of the connection between jury deliberation and political participation. *J. Polit.*, 64(2):585–595, May 2002.

[13] V. Hans, J. Gastil, and T. Feller. Deliberative democracy and the american civil jury. *J. Empirical Legal Stud.*, 11(4):697–717, December 2014.

[14] Bernard Grofman. *Multnomah County [Oregon] Jury Project, 1973-1976.* Inter-university Consortium for Political and Social Research, Ann Arbor, MI, 1984.

[15] F. Galton. Vox populi. *Nature*, 75:450–451, 1908.

[16] Jan Lorenz, Heiko Rauhut ad Frank Schweitzer, and Dirk Helbing. Why amazon's ratings might mislead you: The story of herding effects. *Big Data*, 2(4):196–204, 2014.

[17] Jan Lorenz, Heiko Rauhut ad Frank Schweitzer, and Dirk Helbing. How social influence can undermine the wisdom of crowd effect. *PNAS*, 108(22):9020–9025, 2011.

[18] G. Stoddard. Popularity dynamics and intrinsic quality in reddit and hacker news. In *Proceedings of the Ninth International AAAI Conference on Web and Social Media*, pages 416–425, 2015.

[19] M. Salganik, P. Dodds, and D. Watts. Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311:854–856, 2006.

[20] J. A. Holyst, K. Kacperski, and Frank Schweitzer. Social impact models of opinion dynamics. *Annual Reviews of Computational Physics*, pages 253–273, 2001.

[21] J. Fernandez-Gracia, K. Suchecki, J. J. Ramasco, M. SanMiguel, and V. M. Eguiluz. Is the voter model a model for voters? *Phys. Rev. Lett.*, 112(15):158701, April 2014.

[22] Lazaros K. Gallos, Pablo Barttfeld, Shlomo Havlin, Mariano Sigman, and Hernán A. Makse. Collective behavior in the spatial spreading of obesity. *Sci. Rep.*, 2:454 EP –, 2012.

[23] Cosma Shalizi and Andrew Thomas. Homophily and contagion are generically confounded in observational social network studies. *Sociol. Method. Res.*, 40:211–239, 2011.

[24] Keith Burghardt, Emanuel F. Alsina, Michelle Girvan, William Rand, and Kristina Lerman. The myopia of crowds: A study of collective evaluation on stack exchange. *arXiv: 1602.07388*, 2016.

[25] Keith Burghardt, William Rand, and Michelle Girvan. Competing opinions and stubbornness: Connecting models to data. *Phys. Rev. E*, 2016.

[26] Arpita Ghosh and Patrick Hummel. A game-theoretic analysis of rank-order mechanisms for user-generated content. *Journal of Economic Theory*, 154:349–374, 2014.

[27] Shaili Jain, Yiling Chen, and David C Parkes. Designing incentives for online question and answer forums. In *Proceedings of the 10th ACM conference on Electronic commerce*, pages 129–138. ACM, 2009.

[28] James Surowiecki. *The wisdom of crowds*. Anchor, 2005.

[29] Aniket Kittur and Robert Kraut. The wisdom of reluctant crowds. In *Proceedings of the 43rd Hawaii International Conference on System Sciences (HICSS)*, HICSS, 2010, pages 1–10, Honolulu, HI, 2010. IEEE.

[30] Hailiang Chen, Prabuddha De, Yu (Jeffrey) Hu, and Byoung-Hyoun Hwang. Wisdom of crowds: The value of stock opinions transmitted through social media. *Rev. Financ. Stud.*, 27(5):1367–1403, 2014.

[31] Jan Lorenz, Heiko Rauhut, Frank Schweitzer, and Dirk Helbing. How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences*, 108(22):9020–9025, May 2011.

[32] Matthew J. Salganik, Peter S. Dodds, and Duncan J. Watts. Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762):854–856, 2006.

[33] Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. Discovering value from community activity on focused question answering sites: a case study of stack overflow. In *KDD '12 Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 850–858. ACM, 2012.

[34] J. W. Vaupel and A. I. Yashin. Heterogeneity's ruses: some surprising effects of selection on population dynamics. *The American Statistician*, 39(3):176–185, 1985.

[35] Chiraq Shah and Jefferey Pomerantz. Evaluating and predicting answer quality in community qa. In *Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval*, pages 411–418. ACM, 2010.

[36] Soojung Kim and Sanghee Oh. Users relevance criteria for evaluating answers in a social q&a site. *J. Assoc. Inf. Sci. Techno.*, 60(4):716–727, 2009.

[37] Eugene Agichtein, Carlos Castillo, Debora Donato, Aristides Gionis, and Gilad Mishne. Finding high-quality content in social media. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, WSDM '08, pages 183–194. ACM, 2008.

[38] Benjamin Scheibehenne, Rainer Greifeneder, and Peter M Todd. Can there ever be too many options? a meta-analytic review of choice overload. *Journal of Consumer Research*, 37(3):409–425, 2010.

[39] Nathan O. Hodas and Kristina Lerman. The simple rules of social contagion. *Scientific Reports*, 4, 2014.

[40] Manuel G. Rodriguez, Krishna Gummadi, and Bernhard Schoelkopf. Quantifying information overload in social media and its impact on social contagions. In *Proceedings of Eighth International AAAI Conference on Weblogs and Social Media*, March 2014.

[41] Mohan John Blooma, Dion Hoe-Lian Goh, and Alton Yeow-Kuan Chua. Predictors of high-quality answers. *Online Information Review*, 36(3):383–400, 2012.

[42] Xin-Jing Wang, Xudong Tu, Dan Feng, and Lei Zhang. Ranking community answers by modeling question-answer relationships via analogical reasoning. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 179–186. ACM, 2009.

[43] Tetsuya Sakai and Ruihua Song. Evaluating diversified search results using per-intent graded relevance. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 1043–1052. ACM, 2011.

[44] Lada A Adamic, Jun Zhang, Eytan Bakshy, and Mark S Ackerman. Knowledge sharing and yahoo answers: everyone knows something. In *Proceedings of the 17th international conference on World Wide Web*, pages 665–674. ACM, 2008.

[45] Amit Rechavi and Sheizaf Rafaeli. Not all is gold that glitters: Response time & satisfaction rates in yahoo! answers. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*, pages 904–909. IEEE, 2011.

[46] Chirag Shah. Measuring effectiveness and user satisfaction in yahoo! answers. *First Monday*, 16(2), 2011.

[47] Bee-Chung Chen, Anirban Dasgupta, Xuanhui Wang, and Jie Yang. Vote calibration in community question-answering systems. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 781–790. ACM, 2012.

[48] Daniel Kahneman. Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, 93(5):1449–1475, 2003.

[49] Herbert Alexander Simon. *Models of bounded rationality: Empirically grounded economic reason*, volume 3. MIT press, 1982.

[50] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer, 2009.

[51] J. Friedman, T. Hastie, and R. Tibshirani. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.*, 33:1–22, 2010.

[52] J. J. Goeman. L1 penalized estimation in the cox proportional hazards model. *Biometrical Journal*, 52(1):70–84, 2010.

[53] J P Kincaid, Robert P Fishburne, Jr, Richard L Rogers, and Brad S Chissom. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. Technical report, U.S. Navy, 1975.

[54] L. Ponzanelli, A. Mocci, A. Bacchelli, M. Lanza, and D. Fullerton. Improving low quality stack overflow post detection. In *Software Maintenance and Evolution (ICSME), 2014 IEEE International Conference on*, pages 541–544. IEEE, 2014.

[55] Kristina Lerman and Tad Hogg. Leveraging position bias to improve peer recommendation. *PLoS ONE*, 9(6):e98914, 2014.

[56] Nick Craswell, Onno Zoeter, Michael Taylor, and Bill Ramsey. An experimental comparison of click position-bias models. In *Proceedings of the international conference on Web search and web data mining*, WSDM '08, pages 87–94, 2008.

[57] Daniel Kahneman. *Thinking, fast and slow.* Farrar, Straus and Giroux, 2011.

[58] Herbert A. Simon. Designing organizations for an information rich world. In Martin Greenberger, editor, *Computers, communications, and the public interest*, pages 37–72. Baltimore, 1971.

[59] Ronald N. Taylor. Psychological determinants of bounded rationality: Implications for decision-making strategies. *Decision Sciences*, 6(3):409–429, 1975.

[60] Stanley L. Payne. *The Art of Asking Questions.* Princeton University Press, 1951.

[61] Tad Hogg and Kristina Lerman. Disentangling the effects of social signals. *Human Computation Journal*, 2(2):189–208, 2015.

[62] K. Sznajd-Weron and J. Sznajd. Opinion evolution in closed community. *Int. J. M. P. C*, 11(6):1157–1165, September 2000.

[63] S. Galam. Contrarian deterministic effect: the "hung elections scenario". *Physica A*, 333:453–460, 2004.

[64] S. Gekle, L. Peliti, and S. Galam. Opinion dynamics in a three-choice system. *Euro. Phys. J. B*, 45(4):569–575, June 2005.

[65] J. Shao, S. Havlin, and H. E. Stanley. Dynamic opinion model and invasion percolation. *Phys. Rev. Lett.*, 103(1):018701, July 2009.

[66] E. Yildiz, D. Acemoglu, A. Ozdaglar, A. Saberi, and A. Scaglione. Discrete opinion dynamics with stubborn agents. *Oper. Res.*, 1(4):19:1–30, 2011.

[67] R. Hegselmann and U. Krause. Opinion dynamics and bounded confidence models, analysis, and simulation. *JASSS*, 5(3):1–33, 2002.

[68] T. Lux. Rational forecasts or social opinion dynamics? identification of interaction effects in a business climate survey. *J. Econ. Behav. Organ.*, 72(2):638–655, 2009.

[69] Z. Wang, Y. Liu, L. Wang, Y. Zhang, and Z. Wang. Freezing period strongly impacts the emergence of a global consensus in the voter model. *Sci. Rep.*, 4:3597 EP –, January 2014.

[70] F. Palombi and S. Toti. Stochastic dynamics of the multi-state voter model over a network based on interacting cliques an zealot candidates. *Journal of Statistical Physics*, 156:336–367, 2014.

[71] S. Galam and F. Jacobs. The role of inflexible minorities in the breaking of democratic opinion dynamics. *Physica A*, 381:366–376, July 2007.

[72] M. Mobilia, A. Petersen, and S. Redner. On the role of zealotry in the voter model. *J. Stat. Mech.-Theor. E.*, 2007, August 2007.

[73] K. Sznajd-Weron and R. Weron. How effective is advertising in duopoly markets? *Physica A*, 324(1):437–444, 2003.

[74] P. Holme and M. E. J. Newman. Nonequilibrium phase transition in the coevolution of networks and opinions. *Phys. Rev. E*, 74(5):056108, 2006.

[75] R. Axelrod. The dissemination of culture the dissemination of culture: A model with local convergence and global polarization. *J. Conflict Resolut.*, 41(2):203–226, April 1997.

[76] D. H. Zanette and S. C. Manrubia. Vertical transmission of culture and the distribution of family names. *Physica A*, 295(1):1–8, June 2001.

[77] Q. Li, L. A. Braunstein, H. Wang, J. Shao, H. E. Stanley, and S. Havlin. Non-consensus opinion models on complex networks. *J. Stat. Phys.*, 151(1-2):92–112, April 2013.

[78] N. M. Steblay, J. Besirevic, S. M. Fulero, and B. Jimenez-Lorente. The effects of pretrial publicity on juror verdicts: A meta-analytic review. *Law Human Behav.*, 23(2), 1999.

[79] Serge Galam and Serge Moscovici. Towards a theory of collective phenomena: Consensus and attitude changes in groups. *Eur. J. Soc. Psychol.*, 21(1):49–74, 1991.

[80] Serge Galam. Rational group decision making. a random field ising model at t=0. *Physica A*, 238(1-4):66–80, 1997.

[81] Pew Research. Party affiliations and election polls, August 2012.

[82] R. Bond, C. Fariss, J. Jones, A. Kramer, C. Marlow, J. Settle, and J. H. Fowler. A 61-million-person experiment in social influence and political mobilization. *Nature*, 489:295–298, September 2012.

[83] D. Romero, B. Meeder, and J. Kleinberg. Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on twitter. In *Proceedings of the 20th International Conference on World Wide Web*, pages 695–704, New York, NY, USA, 2011. ACM.

[84] N. Hodas and K. Lerman. How visibility and divided attention constrain social contagion. In *Proceedings of the 2012 ASE/IEEE International Conference on Social Computing*, pages 249–257. IEEE, 2012.

[85] R. Erikson and C. Wlezien. The timeline of presidential election campaigns. *J. Polit.*, 64(4):969–993, November 2002.

[86] R. Congleton. *The Median Voter Model*, pages 707–712. Springer US, New York, NY, USA, 1 edition, 2004.

[87] D. Black. One the rationale of group decision-making. *J. Polit. Econ.*, 56(1):23–34, 1948.

[88] R. Holcombe. An empirical test of the median voter model. *Econ. Inq.*, 18(2):260–274, April 1980.

[89] L. Onsager. Crystal statistics. i. a two-dimensional model with an order-disorder transition. *Phys. Rev.*, 65(4):117–149, February 1944.

[90] P. Krapivsky, S. Redner, and E. Ben-Naim. *A Kinetic View of Statistical Physics*. Cambridge University Press, 2010.

[91] B. A. Prakash, A. Beutel, R. Rosenfeld, and C. Faloutsos. Winner takes all: Competing viruses or ideas on fair-play networks. In *Proceedings of the 21st international conference on World Wide Web*, pages 1037–1046, New York, NY, USA, 2012. ACM.

[92] F. DarabiSahneh and C. Scoglio. Competitive epidemic spreading over arbitrary multilayer networks. *Phys. Rev. E*, 89(6):062817, 2014.

[93] C. Poletto, S. Meloni, V. Colizza, Y. Moreno, and A. Vespignani. Host mobility drives pathogen competition in spatially structured populations. *PLoS Comput. Biol.*, 9(8):e1003169, 2013.

[94] J. Sanz, C. Y. Xia, S. Meloni, and Y. Moreno. Dynamics of interacting diseases. *Phys. Rev. X*, 4(4):041005, 2014.

[95] Gabriel Leventhal, A. L. Hill, Martin A. Nowak, and S. Bonhoeffer. Evolution and emergence of infectious diseases intheoretical and real-world networks. *Nature Comm.*, 6, 2015.

[96] J. Humplik, A.Hill, and M. A. Nowak. Evolutionary dynamics of infectious diseases in finite populations. *J. Theor. Biol.*, 360:149–162, 2014.

[97] M. Girvan, D. S. Callaway, M. E. J. Newman, and S. H. Strogatz. Simple model of epidemics with pathogen mutation. *Phys. Rev. E*, 65(3):031915, March 2002.

[98] H. U. Stark, C. J. Tessone, and F. Schweitzer. Decelerating microdynamics can accelerate macrodynamics in the voter model. *Phys. Rev. Lett.*, 101(1):018701, June 2008.

[99] L. Dall'Asta and C. Castellano. Effective surface-tension in the noise-reduced voter model. *Europhys. Lett.*, 77(6):60005, 2007.

[100] Andre C. R. Martins and Serge Galam. The building up of individual inflexibility in opinion dynamics. *PRE*, 87:042807, 2013.

[101] Andre Martins. Continuous opinions and discrete actions in opinion dynamics problems. *Int. J. Mod. Phys. C*, 19:617–624, 2008.

[102] A. Mantonakis, P. Rodero, I. Lesschaeve, and R. Hastie. Order in choice: Effects of serial position on preferences. *Psychol. Sci.*, 20(11):1309–1312, 2009.

[103] Joan B. Kessler. *The Jury System In America, A Critical Overview.* Sage Publications, 1975.

[104] M. Boguna, C. Castellano, and R. Pastor-Satorras. Nature of the epidemic threshold for the susceptible-infected-susceptible dynamics in networks. *Phys. Rev. Lett.*, 111(6):068701, August 2013.

[105] Maxwell McCombs. *Setting the Agenda: The Mass Media and Public Opinion.* Polity Press, 2004.

[106] A. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[107] D. Watts and S. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, June 1998.

[108] I. Dornic, H. Chaté, J. Chave, and H. Hinrichsen. Critical coarsening without surface tension: The universality class of the voter model. *Phys. Rev. Lett.*, 87(4):045701, July 2001.

[109] F. Vazquez and V. Eguiluz. Analytical solution of the voter model on uncorrelated networks. *New J. Phys.*, 10, June 2008.

[110] E. Pugliese and C. Castellano. Heterogeneous pair approximation for voter models on networks. *Europhys. Lett.*, 88(5), December 2010.

[111] A. Chaudhuri and M. B. Holbrook. The chain of effects from brand trust and brand affect to brand performance: The role of brand loyalty. *J. Marketing*, 65(2):81–93, April 2001.

[112] J. Jacoby and D. B. Kyner. Brand loyalty vs. repeat purchasing behavior. *J. Marketing Res.*, 10(1):1–9, February 1973.

[113] M. B. Lieberman and D. B. Kyner. Montgomery. First mover advantages. *Strategic Management Journal*, 9:41–58, 1988.

[114] L. Frachebourg and P. L. Krapivsky. Exact results for kinetics of catalytic reactions. *Phys. Rev. E*, 53(4), 1995.

[115] L. Steels. A self-organizing spatial vocabulary. *Artificial Life*, 2(3):319–332, April 1995.

[116] D. Abrams and S. Strogatz. Linguistics: Modelling the dynamics of language death. *Nature*, 424(6951):900–900, August 2003.

[117] X. Wei, N. Valler, B. A. Prakash, I. Neamtiu, M. Faloutsos, and C. Faloutsos. Competing memes propagation on networks: A case study of composite networks. *Comput. Commun. Rev.*, 42(5):5–12, 2012.

[118] L. Weng, A. Flammini, A. Vespignani, and F. Menczer. Competition among memes in a world with limited attention. *Scientific Reports*, 2(335):1–8, 2012.

[119] J. Gleeson, J. Ward, K. P. O'Sullivan, and W. T. Lee. Competition-induced criticality in a model of meme popularity. *Phys. Rev. Lett.*, 112(4):048701, January 2014.

[120] Baruch Barzel, Yang-Yu Liu, and Albert-László Barabási. Constructing minimal models for complex system dynamics. *Nature Comm.*, 6:7186, 2015.

[121] J. Yang, I. Kim, and W. Kwak. Optimization of consensus time by combining the voter and the majority voter models on scale-free networks. *Europhys. Lett.*, 88(2):20009, 2009.

[122] Aaron Clauset, Cosma Rohilla Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51:661–703, 2009.

[123] Yogesh Virkar and Aaron Clauset. Power-law distributions in binned empirical data. *Ann. Appl. Stat.*, 8(1):89–119, 2014.

[124] William Pickering and Chjan Lim. Solution of the voter model by spectral analysis. *Phys. Rev. E*, 91:012812, January 2015.

[125] P. L. Krapivsky and S. Redner. Dynamics of majority rule in two-state interacting spin systems. *Phys. Rev. Lett.*, 90(23):238701, 2003.

[126] M. J. Saks and M. W. Marti. A meta-analysis of the effects of jury size. *Law Human Behav.*, 21(5):451–466, 1997.

[127] D. J. Devine an L. D. Clayton, B. B. Dunford, R. Seying, and J. Pryce. Jury decision making: 45 years of empirical research on deliberating groups. *Psychol. Public Pol. L.*, 7(3):622–727, March 2000.

[128] T. Brunell, C. Dave, and N. Morgan. Factors affecting the length of time a jury deliberates: Case characteristics and jury composition. *Rev. Law Econ.*, 5(1):555–578, 2009.

[129] N. L. Kerr and R. J. MacCoun. The effects of jury size and polling method on the process and product of jury deliberation. *J. Pers. Soc. Psychol*, 48(2):349–363, 1985.

[130] Masahiro Kimura, Kazumi Saito, Kouzou Ohara, and Hiroshi Motoda. Opinion formation by voter model with temporal decay dynamics. *Lecture Notes in Computer Science*, 7524:565–580, 2012.

[131] A. Das, S. Gollapudi, and K. Munagala. Modeling opinion dynamics in social networks. In *WSDM '14 Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, pages 403–412, 2014.

[132] Oren Tsur and Ari Rappoport. Don't let me be #misunderstood: Linguistically motivated algorithm for predicting the popularity of textual memes. In *Ninth International AAAI Conference on Web and Social Media*, 2015.

[133] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch. Mixing beliefs among interacting agents. *Adv. Complex Syst.*, 3(01n04):87–98, 2000.

[134] Sinan Aral, Lev Muchnik, and Arun Sundararajan. Distinguishing influence-based contagionfrom homophily-driven diffusion in dynamic networks. *PNAS*, 106(51):1544–21549, 2009.

[135] D. Volovik and S. Redner. Dynamics of confident voting. *J. Stat. Mech.-Theory E.*, 2012.

[136] Eric T. Lofgren, M. Elizabeth Halloran, Caitlin M. Rivers, John M. Drake, Travis C. Porco, Wan Yang Bryan Lewis, Alessandro Vespignani, Jeffrey Shaman, Joseph N. S. Eisenberg, Marisa C. Eisenberg, Madhav Marathe, Samuel V. Scarpino, Kathleen A. Alexander, Rafael Meza, Matthew J. Ferrari, James M. Hyman, Lauren A. Meyers, and Stephen Eubank. Opinion: Mathematical models: A key tool for outbreak response. *PNAS*, 111(51):18095–18086, 2014.

[137] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse. Identification of influential spreaders in complex networks. *Nature Physics*, 6:888–893, 2010.

[138] F. Morone and H. A. Makse. Influence maximization in complex networks through optimal percolation. *Nature*, 524:65–68, 2015.

[139] J. Leskovec. Stanford large network dataset collection, 2014.

[140] S. Milgram. The small world problem, May 1967.

[141] J. Leskovec and E. Horvitz. Planetary-scale views on an instant-messaging network. In *Proceedings of the 17th international conference on World Wide Web*, pages 915–924, 2008.

[142] M. Granovetter. The strength of weak ties. *Am. J. Sociol.*, 78(6):1360–1380, May 1973.

[143] Anamaria Berea and Rizwan Ahmed Mohammed. Huridocs/teradata cares data dive in collaboration with datakind.

[144] G. Fasano and A. Franceschini. A multidimensional version of the kolmogorov-smirnov test. *Monthly Notices of the Royal Astronomical Society*, 225:155–170, 1987.

[145] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C*. Cambridge University Press, Cambridge, UK, 1992.