Group Decision Making and Preference Learning on Social Networks

by

Amirali Salehi-Abari

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Computer Science
University of Toronto

Abstract

Group Decision Making and Preference Learning on Social Networks

Amirali Salehi-Abari
Doctor of Philosophy
Graduate Department of Computer Science
University of Toronto
2016

This thesis focuses on exploiting the dynamics and correlations of preferences over social networks for developing efficient group decision making systems. One of the main challenges in any group decision problem is learning the individual preferences upon which decisions are based. It is the position of this thesis that social networks—by capturing preference correlations across individuals induced by social interactions—provide a natural and informative platform for preference learning. By mathematical modelling of preference dynamics and correlations over social networks, we focus on developing efficient algorithms for group decision making and recommendations, with less required user data, and lower cognitive and communication burden.

We introduce *empathetic* frameworks in which individuals derive utility based on both their own intrinsic preferences (or happiness) and empathetic preferences, determined by the satisfaction of their acquaintances. After theoretically analysing this framework, we develop a scalable algorithm for group recommendation, and empirically demonstrate its performance.

To capture the correlation of preference rankings on social networks, we introduce a network formation model called *ranking networks* in which the similarity of two individuals' rankings determines the chance they are connected to each other. After a thorough theoretical analysis, we use a special instance that we call *preference-oriented social networks*, for group decision making when faced with missing preferences. We develop algorithms to predict unknown individual preferences given known preferences of others in the social network and to make effective group decisions with partial preferences. Our empirical results demonstrate that incorporating social ties can significantly improve predictions and group decision making.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my advisor Craig Boutilier. Through my studies, I was fortunate to benefit immensely from his guidance and intellectual support. I could not have hoped for a better supervisor, mentor, researcher, and role model. His impact on my career has gone beyond this thesis as he has helped me improve my critical thinking skills, communication skills, professionalism, and ambition in conducting meaningful and practical research. His patience, tenacity, and conscientious nature demonstrate the excellence of his character.

I am very grateful of my excellent supervisory committee members Allan Borodin, Peter Marbach, and Rich Zemel. They supported my research by sharing their helpful comments, thoughts, and perspectives. They all have heavily contributed to my professional development in many ways. Special thanks to Allan for being an excellent mentor in the last one and half year of my studies; his thoughtful, diligent, and hardworking character and attitude have touched me. I am also very thankful of Peter Marbach for cultivating my curiosity and passion in research, from the beginning and throughout of Ph.D. studies at UofT. Also, I would like to thank Rich Zemel for being always available to talk and meet with me about my research. Also, I am very thankful to my external examiner, Kate Larson, for providing very thoughtful comments on this thesis.

Thanks should go to my fellow lab mates and graduate students who have made my time at the University of Toronto a meaningful and memorable experience: Tyler Lu, Joel Oren, Xin Sui, Joanna Drummond, Andrew Perrault, Laurent Charlin, Milad Eftekhar, Mohammad Norouzi, Aida Nematzadeh, Jake Snell, Yujia Li, Jamie Ryan Kiros, and Ozan Erdem. I also offer my thanks to the administrative and computing staff in the AI group, Luna Keshwah and Relu Patrascu, for their wonderful job keeping things running smoothly in the AI group.

Finally, I offer my heartfelt thanks to my family for their everlasting intellectual and emotional support. My parents have given me all a child can wish for; I am so grateful for them and I owe all of my education and success to their unconditional support. I especially thank my caring brother Omid Abari who has always been a great source of wisdom and is supportive of my success. I am truly in debt of my lovely wife and best friend Julie Thorpe who has always been a great encouragement, source of inspiration, open-minded listener, indefatigable supporter, and even skillful editor.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Recent advances in computational networked systems (e.g., Internet, Web, email, online social network, and social media) have revolutionized our daily lives and interactions, where information and people are united over the globe. Due to these high-impact technological developments, a new class of *socio-computational* challenges has emerged, requiring new models of computation designed around the social, informational, psychological, and even economic needs of people.

With the explosive growth of choices and information, we have more freedom to choose at the expense of dealing with many overwhelming (minor or major) decisions in our day-to-day lives, thus resulting in more frustration and less satisfaction for us—"more is less" [316]. As such, the crucial role of *decision-support systems* (e.g., recommender systems) and *navigational tools* (e.g., search engines) is inevitable in the near future, helping to inform us of and navigate decisions and information space.

Of practical import and aligned with my research interests are *social choice (or group decision) problems*, which address the problem of choosing an *outcome* or a *decision*, from a set of *alternatives*, for a group of individuals (or agents) who have their own personal *preferences* over the set of alternatives. Social choice problems in the physical and online worlds are prevalent: for instance, when a group of friends chooses a movie to watch, a resort at which to vacation, or a restaurant at which to eat; when a company (e.g., Facebook) decides on assigning various advertisements to its users based on the relevance of ads and user preferences; selecting a policy for an online system or a nation given individuals' preferences, etc.

Social choice theory has its root in political science and economics. Of special interest to these fields is the study of *voting rules* that take individual preferences as input, and output a recommended option or a decision, which reflects some notion of consensus. However, nowadays, addressing the computational challenges in social choice is indispensable for computer scientists, since many applications which deal with user preferences to make a decision (e.g., recommender systems, search engines, advertisement and marketing mechanisms, etc.) have some sort of social choice problem at their core.

One of the main endeavours in any social choice problem is understanding individual preferences upon which decisions are made or recommended. This understanding might be gained by *preference learning* from historical user interaction data, or by *preference elicitation* via asking relevant queries which result in revealing user preferences. In addition to understanding user preferences, preference learning and elicitation have another common objective of *reducing the cognitive and communication burden* imposed on users, by requiring as little information as possible to make (close to) optimal decisions. The latter is essential to the practicality of decision-support systems with the goal of filtering out irrelevant options and information.

1.1 Thesis Vision and Statement

It is the position of this thesis that social networks provide a natural and informative platform for preference learning and elicitation and consequently social choice problems. Social networks play a crucial role in facilitating a variety of social and economic interactions [187, 88], including discovery of job opportunities [171] and the products we consume [154], how we vote [54], and how we cooperate [136]. It is widely recognized that the behaviors of individuals in a network are correlated with those of their friends or other social connections. These correlations are usually explained by two sociological phenomena: *homophily* or *social selection*—the tendency of people to associate with others whom they perceive to be similar to themselves—and *social influence*—in which people tend to adopt the properties and attitudes of those to whom they are connected. Because of this, and the increasing availability of user behavioral data, it is essential to study the interplay of social network structure and individual behaviour, attitudes and preferences.

In this light, specific questions direct the research in this thesis on social choice problems in social networks: (i) What dynamics dictate how individual preference become correlated in a social network over time (see Chapter 3)? (ii) How should one mathematically model such correlations and dynamics (see Chapter 3 and Chapter 4)? (iii) Can such correlations and dynamics be harnessed for more efficient preference learning and elicitation, and consequently more effective group decision making? (see Chapter 3 and Chapter 5)

Thesis Statement. *This thesis intends to advance the understanding and mathematical modelling of preference dynamics and correlations over social networks and to exploit the computational and predictive power of these models to develop efficient algorithms for decision making and recommendations, while requiring less user data, and imposing lower cognitive and communication burden on users.*

1.2 Thesis Organization and Contributions

After reviewing the related work and background in Chapter 2, three chapters follow containing the primary contributions of the thesis. Respectively, these chapters present models and algo-

rithms for *empathetic preferences*, *ranking networks*, and *preference-oriented social networks*. The contributions of these chapters are outlined below. Finally, Chapter 6 concludes the thesis and present possible future research directions.

Chapter 3: Empathetic Social Choice on Social Networks. In Chapter 3, we introduce the *empathetic social choice framework* [312] in which agents derive utility (or benefit) based on both their own intrinsic preferences (or happiness) and empathetic preferences, determined by the satisfaction of their acquaintances (e.g., friends voting for a vacation spot or a movie) while considering both their own and others' satisfaction. Empathetic utility in this sense reflects the fact that a person's happiness may be influenced by the happiness of others with whom they are connected [135]. We study the theoretical conditions under which such empathetic preferences are well-defined (i.e., converge to a fixed-point). We show that consensus decision making (or group recommendation) problems with empathetic preferences can be recast as a weighted preference aggregation over intrinsic preferences alone, when weights are determined by social network structure. We develop two scalable iterative algorithms for consensus decision making. We also generalize our empathetic framework to accommodate other social choice problems (e.g., constrained resource allocation, etc.) and show how some of our theoretical results (e.g., fixed-point convergence) still hold. We also demonstrate that, in the general empathetic framework, these other social choice problems can be viewed as using weighted preference aggregation. We develop a scalable iterative method for estimating those societal weights, which serves as a building block for solving other social choice problems. Through our empirical experiments, we demonstrate the value of accounting for empathetic preferences in group decisions as well as the performance of our algorithms. Our results confirm that neglecting empathy usually yields sub-optimal group decisions which degrade the well-being of the group members.

Chapter 4: Ranking Networks. To capture the correlation of ranking preferences (or general rank data) on social networks, in Chapter 4 we introduce a network formation model called *ranking networks* [311], in which the similarity of two entities' rankings over a set of options determines the chance they are connected to each other. We theoretically analyze general topological properties of this model, demonstrating that it exhibits some commonly observed properties of real-work networks such as a small diameter, the existence of a giant connected component, and shrinking diameter. We also derive closed-form formulae for estimating degree distribution, edge density, and clustering coefficients under this model, but show that their computations are very expensive. To counteract this, we develop easy-to-compute approximations for the special class of *distance-based ranking models*, while studying properties that emerge among networks in this class. We show how our approximations can be used for efficient model learning when faced with missing data. Through empirical experiments, we demonstrate the effectiveness of our approximation and learning methods.

Chapter 5: Group Recommendation on Preference-Oriented Social Networks. In Chapter 5, we address how to use *social network structure* to support more accurate infer-

ence of preference rankings and to make group decisions when some individual preferences are unknown. To this end, we introduce the *preference-oriented social networks (POSN)* model [313]—an instance of ranking networks—to capture and exploit the correlation of preferences over social networks. We exploit this model to infer unobserved individual preferences given observed preferences of others in the social network. Intuitively, if we know something about the preferences of your friends, family or colleagues—or their friends, etc.— we should be able to more accurately predict your preferences if any degree of homophily or social influence is present in the underlying social network. We develop a Markov chain Monte Carlo method for inferring unobserved preferences and evaluate the ability of our methods to predict unobserved preferences and support effective group decision making with partial preferences. We also propose some enhancements to our inference and group decision making methods which improve both the speed and efficacy of of our original methods. These enhancements build upon on a simple and seemingly natural assumption that not everyone has the opportunity to be acquainted with everyone else in a society. Using various datasets (e.g., Flixster, Irish election, etc.), we compare our group recommendation methods to different benchmarks which neglect the information contained in the social network. Our empirical results demonstrate that accounting for social ties can significantly improve predictions and group decision making/recommendation.

1.3 Bibliographic Notes

As Chapters 3–5 are based on the work coauthored with my thesis advisor Prof. Craig Boutilier [312, 311, 313], I would like to acknowledge that I have borrowed some text from our original papers, jointly written with him.

Chapter 2

Background and Related Work

This chapter is organized as follows. Social choice, and related computational problems and models, are reviewed in Sec. 2.1. Probabilistic models of preferences, and their related learning and inference problems, are discussed in Sec. 2.2. Topological properties, dynamics and formation models of social networks, along with related computational problems, are outlined in Sec 2.3.

2.1 Social Choice

People frequently have to arrive at joint (or collective) decisions, despite the fact that they might possess conflicting preferences or attitudes. In our day-to-day lives, we regularly participate in collective decision making processes, ranging from low-stakes or routine decisions (e.g., which movies to watch, at which restaurant to eat dinner, etc.) to high-stakes decisions (e.g., which president to elect, which economic policy to pass, etc.). Our participation in making a collective decision might be in the form of either an informal negotiation process (e.g., discussing with our friends to watch a specific movie) or a carefully specified protocol (e.g., vote for a preferred political candidate in an election). *Social choice theory*—a sub-field of economics and political science—provides a mathematical framework to study and develop mechanisms for collective decision making. Using this framework, one can thoroughly analyse decision making processes or protocols and their efficiency for making the "right" decision.

Generally speaking, social choice addresses the problem of choosing an outcome or a decision, from a set of outcomes or alternatives, for a group of individuals (or agents) who have their own personal preferences over the set of alternatives.. We present a formal definition of preferences and then explain a variety of social choice problems and applications.

2.1.1 Preferences

Agent preferences over alternatives can be encoded as either rankings or utilities, called *ordinal preferences* and *cardinal preferences*, respectively. We consider a set of alternatives, options, or

items $\mathcal{A} = \{a_1, \ldots, a_m\}$ and a set of individuals (or agents) $\mathcal{N} = \{1, \ldots, n\}$. A strict preference relation $\succ$ over $\mathcal{A}$ is a binary, *transitive*, *complete*, *asymmetric* relation. A *weak preference* relation $\succsim$ allows individuals to have ties in their preferences. However, we focus on strict preferences and note that most of our results can be readily extended to weak preferences.

Given a strict preference relation, one can linearly order alternatives in $\mathcal{A}$ with respect to $\succ$. In this sense, the preference of individual i over $\mathcal{A}$ can be presented in the form of a *strict preference ordering* $\succ_i$ (i.e., a permutation of $\mathcal{A}$). Let $\Omega(\mathcal{A})$ denote the set of all possible permutations or preferences over $\mathcal{A}$. Strict preferences can also be viewed as a *ranking*, that is, a bijection $r : \mathcal{A} \to \{1, \cdots, m\}$ mapping each alternative to its rank. The ranking indicates the "place" an alternative holds in the preference ordering (i.e., rank 1 is the most preferred, rank 2 is second most, and so on). Consequently, $r^{-1}(i)$ denotes the i^{th} ranked alternative in ranking r. The collection of agent preferences $P = \{r_1, \cdots, r_m\}$ is referred to as *preference profile*. Similarly, the set of all possible preference profiles is denoted by $\Omega(\mathcal{A})^n$.

Agent preferences can also be represented as utilities or cardinal preferences in which an agent's preference is specified by a utility function $u : \mathcal{A} \to \mathbb{R}$. Utilities are numerical values associated with the alternatives. For any preference ordering, one always can find a set of consistent utilities which do not violate the underlying ordering. Cardinal or ordinal preferences each offer certain advantages over the other. A utility function encodes more information about individuals' preferences (i.e., it is more expressive), including the intensity of preferences and representing preferences in the form of a probability distribution; nonetheless, in the absence of common numeric value such as money, interpersonal comparison between utilities is quite problematic and controversial [177, 42, 41]. This—together with the fact that much of the social choice literature deals with rankings rather than utilities—motivates us to focus on ordering/ranking preferences in this thesis.

2.1.2 Preference Aggregation

Preference aggregation, the problem of aggregating individual preferences into one collective *social preference* relation, is arguably the most fundamental question in social choice theory. Preference aggregation has a variety of applications in the social sciences, economics, political science, and even computer science.

When there are only two alternatives, *majority rule*—built upon common sense and several axiomatic characterizations [263]—suggests that society prefer a to b if and only if there are more individuals who strictly prefer a to b than b to a. In order to generate a social preference relation, one can naturally extend majority rule to more than two alternatives by considering the majority rule for all possible pairs of alternatives. However, Condorcet noted that the concept of social preference can be problematic for more than two alternatives by demonstrating that majority rule can result in a cycle—known as the *Condorcet paradox*. Arrow [13] studied the problem of preference aggregation in-depth and in a systematic manner. Arrow's famous impossibility theorem states that there is no general rule which can aggregate preferences over

more than two alternatives while three natural, sensible properties are met simultaneously. While this impossibility result seem discouraging at the first glance, it has provided a valuable road map for further research.

One practical aggregation rule is *Kemeny's rule* which finds a collective social ranking (sometimes referred to as the *Kemeny consensus*) with the minimum total number of disagreements of voters on relative ranking of any pairs of alternatives. In other words, it minimizes the total *Kendall's* τ distance to all individuals' preferences. Kendall's τ distance of two rankings r and r' is the number of pairwise disagreement of alternatives in those two rankings:

$$d_\tau(r, r') = \sum_{k \neq l} I[r(a_k) > r'(a_l) \text{ and } r(a_l) < r'(a_k)]. \tag{2.1}$$

Kemeny's rule is widely used in practical settings such as rank aggregation for information retrieval applications (see Section 2.2.3 for details).

There are many interpretations of Kemeny's rule: one can characterize it using maximum likelihood estimation [364, 366] assuming that there exists a "correct" reference ranking of the alternative while individuals' preferences (or ranking) are noisy estimates of this reference ranking. Kemeny's rule has attracted considerable amount attention in computational social choice literature. It is NP-hard to compute the Kemeny consensus [28, 117]. Despite this computational difficulty, heuristic optimization techniques have been developed to effectively approximate it. *Local Kemenization* [117] starts with a randomly chosen ranking, then iteratively swapping the order of adjacent alternatives in order to decrease the *Kemeny cost*; this process continues until no more improvements are possible.

Preference aggregation methods are not only limited to Kemeny's rule. There is a class of preference aggregation methods using *positional scoring* rule to rank the alternatives based on their accumulated scores over all rankings (or preferences). Positional scoring rules will be discussed in-depth in Sec. 2.1.3.

2.1.3 Voting

A natural and common approach for deciding among a pool of options is to vote over them and then choose the winning option using some *voting rule*.

Voting Rules

The *voting rule* takes as input preference profiles (i.e., the collection of agent preferences) and returns the winning alternative. Condorcet [107] proposed that a voting rule should always select the winner which is preferred by majority of voters in any possible pairwise comparison, if such an alternative exists. Such an alternative is called *Condorcet winner* and a voting rule which selects the Condorcet winner, if one exists, is called *Condorcet consistent*. We now discuss a variety class of voting rules adopting the taxonomy presented in [63].

Positional Scoring Rules. Under a positional scoring rule, an alternative in each preference (or ranking) is assigned a score based on its position in the ranking. The scores of each alternative over all rankings are then added to form its cumulative score. The winner is the alternative with the highest cumulative score. Generally speaking, any non-increasing function $f : \{1, \cdots, m\} \to \mathbb{R}$, mapping possible rank positions into numerical scores, can be positional scoring rule as long as $f(1) > f(m)$. Under this definition of scoring rules, Fishburn [130] has shown that no positional scoring rule is Condorcet consistent. Nonetheless, positional scoring rules are widely-deployed due to their simplicity, computational tractability, and effectiveness in practice. We here review several common positional scoring rules: plurality, Borda, k-approval, and veto.

Plurality scoring assigns 1 to the first ranked alternative and 0 to all other alternatives. Thus, using plurality, the winning alternative is the one ranked first most often. Note that voters need not specify their entire ranking for the plurality rule to be used. A common criticism of plurality is that it fails to assign partial credit for being ranked second, third, etc. The *Borda rule* [106] addresses this issue by assigning a score of $m - i$ to the alternative ranked i^{th} in a ranking. The Borda rule selects the alternative with highest total Borda score as the winner. Due to its simplicity and advantageous axiomatic characterizations (for example see [363]), this rule has been widely deployed in a range of applications from university council elections to sport team rankings to political elections.

Under the *anti-plurality* or *veto* rule, the last ranked alternative in a ranking receives 0 while the other alternatives are assigned 1; therefore, the winner is the alternative that appears least often as the least-preferred alternative in the individual rankings. *k-approval* is a generalization of plurality where the alternatives in the first k positions of a ranking receive 1 whereas the other alternatives get 0.

Condorcet Extensions. We here review several Condorcet consistent voting rules, which select the Condorcet winner if there is one. As mentioned earlier, the Condorcet winner is that alternative which beats all other alternative in pairwise majority comparisons. *Dodgson's rule* chooses a candidate that is "closest" to being the Condorcet winner, where the notion of distance for each alternative is defined based on the minimum number of swaps one must perform on adjacent alternatives (in a given preference profile) to make that alternative a Condorcet winner. The alternative with the smallest score (i.e., minimum number of swaps) is the consensus winner. Deciding if an alternative is a consensus winner under Dodgson's rule is computationally intractable [28, 180]. Moreover, some other computational properties of Dodgson's rule such as approximability [78] and fixed-parameter tractability [34] have been studied. *Young's rule* shares some similarity with Dodgson's rule, but uses a different notion of distance to the Condorcet winner.

Other Voting Rules. We here review some other voting rules that have some practical

importance.

Single Transferable Vote (STV) is a popular runoff voting procedure which starts by searching for the alternative that is ranked least often as the first-ranked alternative, then removes it from all rankings in the preference profile. STV continues this elimination process until only one alternative is left (the consensus winner). STV has been deployed in various elections around the world including parliamentary elections in Australia and municipal elections in Cambridge, MA.

The *Bucklin Rule* starts by looking at first-ranked alternatives in the preference profile to determine whether an alternative has appeared more than $\frac{n}{2}$ times; if so, that alternative is the winner. If not, it extends its pool by considering the alternatives in first and second place in the preference profiles; if there is an alternative with more than $\frac{n}{2}$ appearances, it is the winner. Otherwise, it continues expanding the pool in the same manner until one alternative appears more than $\frac{n}{2}$ times. If two or more alternatives cross the threshold of $\frac{n}{2}$ in the same round, ties are usually broken by the margin with which they have passed the threshold.

Approval voting [61, 62] is closely related to positional scoring rules (especially to k-approval). In approval voting, every individuals can approve any number of alternatives and consequently the alternative with the highest number of approval is selected as the consensus winner.

Manipulation

Real-world applications of social choice (e.g., elections) require individuals to report their preferences over a set of alternatives. One practical problem is that individuals might be incentivized to misreport their true preference in order to induce the selection of a more desirable alternative. This behaviour is called *strategic manipulation*. One fundamental question is whether there exist voting rules resistant to manipulation; this property is usually referred as *strategy-proofness*. The answer is "no" as proven by Gibbard and Satterthwaite [153, 315] who show that any strategy-proof voting rule with some sensible properties, which deals with at least three alternatives, has to be dictatorship. In other words, assuming some sensible conditions for desirable voting rules, any voting rule is manipulable for some preference profile when there are at least three alternatives. However, this theorem does not hold, if one considers certain restrictions on preference profiles. For example, one very specific restriction on the allowable preferences, called *single peakedness* [275], is enough to guarantee the existence of strategy-proof rules.

As it is theoretically impossible to prevent manipulation in general, one possible remedy is to make manipulation difficult. This difficulty can take the form of computational difficulty for finding the "correct" manipulation strategy. Following the work of Bartholdi *et al.* [30], computer scientists have investigated how to exploit computational hardness as a barrier against manipulation [93, 97, 124]. Computing manipulations is known to be NP-hard for several rules, including STV [29]. The computational complexity of coalitional manipulation [294], in which group of agents collaborate to change the outcome of voting in their favor, is NP-

hard for maxmin [358] and Borda [35, 105] and Copland [128]. State-of-the-art research on harnessing computational hardness in preventing manipulation is surveyed in [129]. However, these computational complexity results should be interpreted cautiously as the hardness comes from worst-case scenarios. This means that there are some settings (e.g., preference profiles) under which manipulation of a specific voting rule is computationally intractable whereas for most (common) settings, the manipulator might be able to efficiently compute a beneficial manipulation strategy. Much recent research on manipulation has shown that some common voting rules in the average case given certain preference distributions are manipulable [95, 300, 141, 346]. Also, some recent studies have relaxed the common assumption that manipulators have full knowledge of other voters (for example, see [250]).

2.1.4 Resource Allocation

Many practical applications in social choice require mechanisms to allocate a collection of resources (or tasks) to the individuals (or agents) under some notion of fairness. The huge literature on *welfare economics* is dedicated to determining the efficient and fair allocation of resources in society. However, we here very briefly review main fundamental concepts of this field mostly from an algorithmic perspective.

Generally speaking, resource allocation of indivisible goods involves in assigning items from a fixed finite set to the agents in a society or group. The set of items assigned to an agent is called that agent's *bundle*. An *allocation* is a specific distribution of resources amongst individuals (i.e., the collection of agent bundles). In resource allocation problems, an agent's preference is usually abstracted by a *valuation function* (or *utility* function), mapping each bundle to a numerical value. Under the *no-externalities* assumption, an agent only has preferences over its own assigned bundle and does not care about bundles that other agents have received. However, it is possible that agents care about other individuals' assignments in addition to their own bundle. This type of preference is usually referred as an *allocative externality*.[1] Another common assumption for valuation functions is *monotonicity* or *free disposal*: for any two sets of goods $S^{'}$ and S, in the case that $S \subset S^{'}$, the value of $S^{'}$ is at least that of S. Resource allocation can be made in either a *centralized* or *decentralized* fashion. In centralized problems, a central authority is in charge of allocating resources to the agents. In distributed settings, agents iteratively communicate their preferences, negotiate, and exchange goods with each other.

The quality of an allocation is usually evaluated based on its *fairness* and/or *efficiency*.

[1] Recently, the settings with externalities have attracted considerable attention in computational social choice literature [223, 21, 96, 110, 64, 332]. Conitzer and Sandholm introduced a general representation for settings with externalities and studied the complexity of efficient computation of optimal outcomes under various restrictions with the presence of externalities [96]. Externalities are studied widely in advertisements [21] and recently in sponsored-search auctions of web advertisements [6, 202, 151, 152, 162]. The main idea here is that the attention that users give to an ad depends on which other ads are displayed simultaneously. Spiteful bidding in auctions [64, 332] in which a spiteful agent benefits by degrading other competitors' profits or utilities is another example of externalities.

Pareto optimality or *Pareto efficiency* is one widely-recognized efficiency criteria. An allocation is Pareto optimal if there is no other allocation that is preferred by some agents and leave no other agents worse off. Deciding if an allocation is Pareto optimal is coNP-complete [84]. The most common condition for fairness is *envy-freeness*. An allocation is envy-free if no agent prefers any other agent's bundle over its own bundle.

A class of quality metrics are defined based on the concept of a *collective utility function* or *social welfare function*, mapping the set of individual valuations into a numerical value which represents overall societal satisfaction. The most commonly deployed collective utility function is *utilitarian* social welfare, simply summing up all individuals' utilities. In this setting, the efficiency of an allocation is usually measured/judged by the extent to which utilitarian social welfare is maximized. This collective utility function favors efficiency of allocations at the cost of neglecting fairness. A more fairness-oriented cumulative utility function is *egalitarian social welfare*, which is given by the lowest (i.e., minimum) utility of any individual under the current allocation. The *Nash product* is another cumulative utility function, multiplying all agents' utilities together, and provide a compromise between egalitarian and utilitarian functions. This function favors both increases in overall utility (i.e, efficiency) and decreases in utility inequalities. An overview of resource allocation for indivisible goods can be found in the survey by Chevaleyre *et al.* [83].

Assignment problems [70], widely studied in computer science and operations research, are closely related to the resource allocation of indivisible goods. In the classical assignment problem, there are a set of agents and a set of items with the same size. Agents have preferences or valuations over items. The goal is to match each person with each item so as to maximize the total commutative valuations. An assignment problem, actually, can be viewed as special case of resource allocation with some extra constraints: the set of resources and the set of agents have the same cardinality, each agent bundle must have only one item.

2.1.5 Stable Matching

At the intersection of game theory and social choice, matching problems have drawn an enormous amount of attention due to its prevalent applications in our day-to-day lives [307, 215]; e.g., matching students to colleges [145], matching medical students to residency programs [308, 309], or matching kidney donors to recipients [1, 17]. In the *stable marriage problem*, there are two sets of agents (colloquially, men and women) where each agent in one set has preferences over the agents in the other set. The goal is to find a *stable* matching between agents of two sets. The stability concept usually reflects the lack of incentive for agents to change their matched partners.

The existence of stable matchings no longer holds in general if one considers *externalities* in preferences, for example, if individuals not only have preferences over the possible partners but also have preferences over who is matched to whom. Recently, the literature on stable matchings with externalities has been growing rapidly [174, 353, 46, 210, 209, 314, 120].

2.1.6 Multi-winner Elections and Segmentation Problems

Motivated by the application of selecting members for a committee (e.g., electing parliament members or selecting board of directors for a company), multi-winner problems focus on selecting a *subset* of alternatives (or candidates) given a preference profile. Indeed, one can view a multi-winner election problem as a voting problem with combinatorial alternatives, corresponding to any feasible subset of candidates (or items) in the multi-winner problem. This perspective results in huge combinatorial choice space, thus complicating both decision making and the expression of preferences. As such, specialized methods and algorithms for solving the multi-winner elections have been developed, which operate on preferences expressed over the original alternatives (rather than the subsets of alternatives).

Multi-winner problems can be solved in less principled ways using preference aggregation methods. The general idea here is to use preference aggregation methods—for example those relying on positional scoring rule—to compute a social preference ranking and then select the top k alternatives as the winners of the original multi-winner election problem. However, this ad hoc approach might output non-optimal solutions. A more principled approach to multi-winner elections are those built upon the concept of *proportional representation*. Roughly speaking, the goal is to select an alternative or candidate for a winner committee based on the fraction of individuals (or voters) who have highly supported that alternative. In this sense, when electing k members as political representatives, the number of elected candidates who belong to a particular political party must be proportional to the number of individuals who voted for that party.

One formulation of proportional representation is given by Chamberlin and Courant [80] in which each individual receives satisfaction or utility only from her most preferred alternative amongst selected candidates (or alternatives) on the *slate*. Potthoff and Brams [299] generalized the Chamberlin-Courant model, allowing an agent to be represented by several candidates in the winning slate. Lu and Boutilier [245] introduced *budgeted social choice*—a more general framework built upon the Chamberlin-Courant model—in which there is a unique cost for adding each alternative to the slate and also a unit cost for the association of an alternative with an agent. They show that the multi-winner elections problem is NP-hard in their framework and develop approximate (greedy) algorithms.

The class of *segmentation problems* [211] is closely related to multi-winner election problems. Each alternative is modelled as a real-valued vector of l attributes and each agent possesses a l-vector of attribute weights. Each agent utility for an alternative is the dot product of its weight vector and the alternative's attribute vector. The objective is to select k alternatives so as to maximize utilitarian social welfare when each agent's utility comes only from its most preferred alternative on the slate.

2.1.7 Preference Elicitation

Eliciting individual preferences for making optimal group decisions presents certain practical difficulties. By operating on all individuals' complete preferences, social choice methods usually elicit or access more information than required to determine a consensus option or make an optimal decision. Eliciting less information from individuals not only reduces the imposed communication and cognitive burden but also mitigates privacy concerns. Although, for many voting schemes, winners cannot be determined without (almost) full preference information in the worst case [92, 94], the development of preference elicitation methods for social choice problems has recently drawn considerable attention due to its promising practical importance [193, 247, 248, 115, 288, 249].

Kalech et al. [193] developed several heuristic methods for preference elicitation using the concept of *possible and necessary winners* and evaluate their performance thorough empirical experiments. Both the possible winner and necessary winner problems [217, 224, 357] take a partial preference profile (i.e., the collection of incomplete preferences) and alternative a as inputs. The possible winner problem determines whether there exists a completion of the partial preference profile under which a is the winner; whereas the necessary winner problem determines whether a is the winner under all possible completions of the partial preference profile. Applying the notion of minimax regret for vote elicitation, Lu and Boutilier [247] propose incremental vote elicitation mechanisms which not only guide the elicitation process, but also compute worst case bounds on the quality of approximate winners given partially elicited preferences. Minimax regret is adopted for incremental preference elicitation in other social choice problems such as stable matching problems [115] and multi-winner elections [249].

2.2 Ranking and Preference Learning

Understanding and modelling individual preferences is a fundamental component of many practical applications including recommender systems, computational advertising, and decision support systems. Focusing on rank data for representing user preferences, we first review several distributional ranking models which are popular in psychology and psychometrics, and then briefly discuss machine learning methods for rankings.

2.2.1 Distance Metrics on Rankings

When one is interested in measuring to what extent two rankings are similar (or close) to each other, a variety of well-known distance metrics for rankings or permutations can be applied [111]. We briefly describe several of the more common distance metrics. A natural set of (spatial) distances are "d_p distances", where

$$d_p(r, r') = \sum_{i=1}^{m} |r(a_i) - r'(a_i)|^p$$

for $p \in [1, \infty)$. The well-known *footrule* ($p = 1$) and *Spearman* distances ($p = 2$) are instances of this. *Hamming* distance can be defined as the number of disagreement between ranking vectors:

$$d_H(r, r^{'}) = |\{a | r(a) \neq r^{'}(a), a \in \Omega(\mathcal{A})\}|.$$

Kendall's τ distance $d_\tau(r, r^{'})$ is used in a variety of contexts, including psychometric [259] and social choice settings (e.g., Kemeny's aggregation rule discussed in Sec.2.1.2) , where it measures the number of pairwise swaps needed to transform r to r' (see Eq. 2.1 for its formal definition).

2.2.2 Ranking Distribution Models

Distributional models of rankings were originally developed in psychometrics and statistics, and are now widely used in machine learning and information retrieval [242] due to their practical applications in web search, recommender systems, etc. Some commonly used examples are Mallows [258], Plackett-Luce [251], Bradley-Terry [59], and many others (see [259] for an overview). Generally speaking, ranking models take two different forms. *Process approaches* mathematically model the behavioural, physical, psychological, or neurological process that occurs when a person ranks a set of objects (e.g., Thurnstone's discriminal process [335], or Placket-Luce's choice process [251]). An alternative is a more *data-driven approach* which parametrically describes the distribution of rankings (e.g., ϕ-Mallows [258]).

Thurstone Model

Thurstone [335] was a pioneer of process modelling, introducing the *discriminal process*. Each object i has tangible and measurable physical feature z_i; then the ranking over objects is generated based on the order of the z_is' magnitudes. Here, the probability of a ranking is the probability of those z_is' magnitudes which result in the corresponding ranking. Thus, the model parameters are those parameters required for describing the distribution of features $\mathbf{z}$. Thurstone [335] originally proposed using a multivariate Gaussian distribution over $\mathbf{z}$. Later, Daniels [104] considered another variant of this model by allowing the z_is in $\mathbf{z}$ to be independent of each other. Henry [181] and Starn [330] employ a Gamma distribution. Sometimes, the Thurstone model and its variants are referred to as *random utility models* in the economics and marketing literature (see, for example, [345, 265]).

Distance-Based Models

Distance-based ranking distributions [131, 259] have ranking probabilities that decrease with increasing distance from a "modal" or "reference" ranking $\sigma \in \Omega(\mathcal{A})$. This is sometimes based on the assumption that there is a "correct" or "popular" ranking (i.e., the modal ranking) and other rankings are expected to be more or less close to this modal ranking. As these models

are usually exponential family models, the probability of ranking r is given by

$$\mathbb{P}(r|\sigma,\omega) = \frac{1}{\psi(\omega)}\exp(-\omega d(r,\sigma)), \tag{2.2}$$

where $\omega \in [0,\infty)$ is a *dispersion parameter*, $\psi(\omega)$ is a normalizing constant, and $d(r,\sigma)$ is the distance between r and σ. As $\omega \to \infty$, P becomes concentrated at the reference ranking σ, whereas for $\omega = 0$, P is the uniform distribution over $\Omega(\mathcal{A})$. Distance-based ranking models differ in the choice of distance metrics. The widely-used Mallows ϕ-model is an example of such a model with d being Kendall's τ distance d_τ. Letting the dispersion parameter $\phi = e^{-\omega}$, $\mathbb{P}(r|\phi,\sigma) = \frac{1}{Z(\phi)}\phi^{d_\tau(r,\sigma)}$ where $Z(\phi) = \prod_{i=0}^{m-1}\sum_{k=0}^{i}\phi^k$.

Paired Comparison Models

Several models use as their building blocks pairwise comparisons rather than rankings. As each ranking can be presented in the form of a set of pairwise comparison, these comparison models have been widely deployed for modelling rankings. The most general pairwise comparison model is due to Babington-Smith [328], which is parametrized by a $\binom{m}{2}$-vector $\mathbf{p}$ with indices ij, $i < j$ where p_{ij} represents the probability that item i is preferred to the item j. Given these probabilities, the model samples independently from all $\binom{m}{2}$ possible comparisons. If the comparisons are consistent (i.e., transitivity holds), they form the sampled ranking; otherwise they are discarded and the process is repeated. As the Babington-Smith model has many parameters especially for large m, Bradley-Terry [59] introduce a variant with m parameters by setting, $p_{ij} = \frac{v_i}{v_i+v_j}$ where v_i is the positive constant weight associated with object i. The ϕ-Mallows model (mentioned above) is also a special case of Babington-Smith with $p_{ij} = \frac{e^{-\omega I[i>j]}}{1+e^{-\omega}}$.

Multistage Models

The general idea behind the *multistage models* (sometimes referred as "choice processes") is simple: the "best" object is ranked first, then the best from the remaining object ranks second and so on. Plackett [296] and Luce [251] proposed a multistage model of rankings, motivated by horse races. Each horse (or object) i is associated with a positive weight v_i capturing the probability i wins the race. The ranking process successively selects an item (proportional to its corresponding weight) to place it in the highest available position. The general model is given by

$$\mathbb{P}(r|\mathbf{v}) = \prod_{i=1}^{m}\frac{v_{r^{-1}(i)}}{\sum_{j=i}^{m} v_{r^{-1}(j)}},$$

where $r^{-1}(i)$ represents the item ranked i^{th} in ranking r. Luce [251] established a choice-based axiomatic foundation for this model. It is clear that the most probable ranking under this model is the one which sorts the items in descending order of their associated weights. Different multistage models have been proposed (e.g., Flinger and Verducci [132]) in which the

probability of selecting the winner in each stage is different that those of Plackett-Luce model.

Mixture Models

One common way to extend the simpler models (e.g., ϕ-Mallows or Placket-Luce) is to assume that there are some types/groups of individuals where each is modelled by a simple ranking model. Most often, each group uses a same parametric family of simple models but with various parameters. More specifically, mixture models assume that rankings come from l distinct *latent* groups, where each group i generates a ranking with probability π_i. The probability of observing ranking r under a mixture model is given by

$$\mathbb{P}(r|\boldsymbol{\pi}) = \sum_{i=1}^{l} \pi_i P_i(r),$$

where $P_i(r)$ denotes the probability distribution of rankings within group i. For example, one can consider the mixture of ϕ-Mallows model by defining the number components l, the probability of each component π_i, and the reference ranking σ_i and dispersion parameter ω_i of each component.

Spatial Models

Spatial models [297, 123, 47, 182, 336, 298] or *unfolding models* [98, 31, 339, 340], developed originally in social psychology and psychometrics, are widely deployed in political science, economics, and consumer choice to capture the relationship between voters (or consumers) and candidates (or products). These models encapsulate individuals, objects, and their relations in a joint multidimensional (usually Euclidean) space.[2] More precisely, m objects are placed in unidimensional or multidimensional scale. Each individual is assumed to possess an *ideal-point* in this common joint space. The ranking of objects for each individual is formed deterministically (or stochastically) by taking objects inversely proportional to their distance; that is, the closer an individual is to an object, more likely the object is to be preferred by the individual.

Blokland-Vogelesang [339, 340] pioneered the use of ϕ-Mallows models for capturing error in unfolding models. In this model, each individual has a modal ranking in the admissible set but her preferences arise from the ϕ-Mallows model with the corresponding modal ranking. The overall ranking distribution can be presented as a mixture of ϕ-Mallows models. Gormley and Murphy [166] employ the Placket-Luce choice model for accommodating error in spatial model.

2.2.3 Learning and Inference Problems

There has been growing body of research in modelling and learning of user preferences in machine learning literature, especially in the context of recommender systems [325]. In parallel,

[2]Unfolding models were originally developed for the unidimensional scaling space by Coombs [98]; later, Bennett and Hays [31] generalized it to the multidimensional space.

the literature on learning-to-rank—at the intersection of information retrieval and machine learning—focuses on how to learn ranking models to successfully rank relevant information (e.g., web pages) given users' queries. We briefly review these related works while focusing on those algorithms and methods that deal with rankings.

Preference Learning

Recently, researchers in machine learning and statistics have exploited distributional models of rankings for modelling and learning of preferences in a society [276, 166, 267, 73, 246, 227]. We here briefly review several of these approaches.

Murphy and Martin [276] employ a mixture of distance-based ranking models to describe individual preferences (in the form of full rankings) from a heterogeneous population. They use the *expectation maximization (EM)* algorithm with *maximum likelihood estimation (MLE)* in the M-step for learning the model parameters; moreover, they study the problem of selecting the optimal number of mixture components. Similarly, Busse *et al.* [73] use EM to learn a mixture of ranking models for partial preferences of the top-t type (i.e., individuals have ranked their t favourites out of m items). Focusing on top-t partial preferences and using a *Dirichlet process mixture*, Meila and Chen present the non-parametric Bayesian treatment of mixture of ϕ-Mallows models [267]. They also study two variants of Gibbs sampling inference techniques for estimating the posterior distribution. Lu and Boutilier [246] relax the restriction on t-type partial rankings by representing partial rankings as pairwise comparisons; they also developed algorithms for learning a mixture of Mallows models using pairwise comparisons while introducing new sampling mechanisms, called *generalized repeated insertion sampling model (GRIM)*. More recently, the learning of the mixtures of distance-based ranking models with the generalized weighted distance metric has been studied [227].

Gormley and Murphy [167] develop learning algorithms for mixtures of both Placket-Luce models and Benter models [32] (i.e., a dampening variant of Placket-Luce model) while focusing on top-t preferences. Their algorithms are applied to explore rank data from two Irish elections. A spatial model combined with Placket-Luce model is deployed for exploring voting data [166]. In this model, both voters and candidates are located in the same multidimensional space where the votes (i.e., rankings) are derived from a Plackett-Luce model. Gormley and Murphy [166] also discuss methods for choosing the dimensionality of the spatial space and learning algorithms. Guiver and Snelson [173] introduced a message-passing algorithm using *expectation propagation* for inferring parameters of a Plackett-Luce ranking distribution. More recently, Azari *et al.* [18] studied conditions on exponential families of random utility models (e.g., Thurstone and its variants) under which fast inference within a Bayesian framework is possible.

Rank Aggregation and Learning-To-Rank

The problem of finding the most "correct" ranking or best *objective* ranking is sometimes referred to as *rank aggregation* (e.g., ranking basketball teams according to their championship

chance). Dwork *et al.* [117] study the problem of aggregating a set of web page rankings into a higher quality ranking, motivated by meta-search engine applications. Using the Kemeny objective cost, they show that rank aggregation is NP-hard. They introduce local search algorithms for approximating aggregate rankings. Cohen *et al.* [91] study a rank aggregation problem in which a set of items (e.g., web pages) and a set of noisy rankings over items (e.g., ranking of experts on web pages) is given and the "correct" ranking of items is desired as an output. They first formulate a supervised learning problem to learn a *preference function*, capturing the probability that each item beats another item in a pairwise comparison. They introduce an online weight learning algorithms, similar in spirit to the weighted majority algorithm [240]. More recently, Volkovs and Zemel [344] developed a learning algorithm to find "correct" or "consensus" rankings from *partial pairwise comparisons* while introducing their own multinomial distributional model over pairwise comparisons.

The literature on rank aggregation is related to *learning-to-rank* problems. The learning-to-rank problem is a supervised learning problem with various practical applications ranging from information retrieval to natural language processing (see [235, 236] for an overview). In the context of information retrieval, the primary goal is to learn a feature-based *ranking function* which ranks the relevant information (e.g., documents or web pages) given a user's query. The training data consists of a set of queries where each query is accompanied by a set of retrieved documents and their corresponding "relevant" labels or scores, indicating the relevant degree of document to a given query. One can categorize the learning-to-rank algorithms to *pointwise*, *pairwise*, and *listwise*. Ignoring the group structure of documents and queries, pointwise methods (e.g., PRank [100]) transform the learning-to-rank problem to standard classification or regression learning to directly find a scoring function whose scores are utilized for ranking documents. In contrast, pairwise methods (e.g., RankNet [69] and RankBoost [140]) convert the learning-to-rank problem to (comparison) pairwise classification and regression while minimizing the relative pairwise misclassification errors. Listwise approaches (e.g., AdaRank [360], SoftRank [334] and BoltzRank [343]) treat the learning-to-rank problem in more natural way by taking each query and its retrieved documents with their relevance scores as an instance during training and learn a ranking function by minimizing some evaluation metrics.

Collaborative Ranking

Collaborative filtering (CF) [68, 220, 239], widely deployed in recommendation systems such as Amazon, Youtube, and Netflix, is a mechanism for making predictions about a user's preference using the preferences of many other "similar" users. Most recommender systems employ collaborative filtering methods to recommend a (small) set of ranked items to a user (e.g., recommending a set of books in Amazon or a set of movies in Netflix). Recently, in the machine learning literature, considerable attention is given to a new generation of these methods, referred to as *collaborative ranking* [352, 241, 23, 342]. As opposed to the traditional collaborative filtering methods which focus on predicting *ratings* of items for a given user, these methods

predict the rankings of items (usually top-t preferences) for a given user. The most common technique in this work is to transform the collaborative ranking problem into a learning-to-rank problem by extracting [342] or learning [23] some user-item features. Balakrishnan and Chopra [23] use learning-to-rank algorithms on user-item features created by concatenating the latent representations of each user and each item. They learn the latent representations of users and items by using *probabilistic matrix factorization (PMF)* [270]. Volkovs and Zemel [342] extract some user-item features for each pair of users and items from the ratings of "similar" users. Then they deploy these user-item features in classical learning-to-rank algorithms to learn rankings of each user over the items.

Label Ranking

Traditional classification problems can be generalized as the problem of *label ranking* when the complete "relevance" ranking of all predefined class labels is requested as a prediction rather than only one single class label. In other words, the label ranking is the problem of learning a function, mapping instances to a complete ranking over some predefined class labels (see [341] for a survey on the topic). Recently, considerable attention is given to label ranking methods, built upon the distributional models of rankings (e.g., the use of ϕ-Mallows [81] and Plackett-Luce [82]).

2.3 Social Networks

Social networks pervade our day-to-day lives. They play a central role in wide variety of social, economical, and political interactions [187, 118, 88] including which job opportunities are provided to us [171], how diseases spread [283, 127], which products we consume [154], how we vote [277, 329, 54], how cooperative we are [136], or even how much weight we gain or lose [87].

In the remainder of this section, we first explain various metrics and representations of social networks while reviewing some topological properties of networks. We then present several statistical models that have been proposed to study how networks form. Afterwards, we look into the dynamics of networks which explain the dissemination of information, the choice of behaviour by people, and how disease spreads in the social networks. Finally, we review some computational problems in social networks.

2.3.1 Structural Metrics and Properties

A social network is composed of a set of individuals and their relationships/interactions with each other. A network can reflect variety of different types of relationships; for example, who is friend with whom, who communicates with whom, who trusts whom, who interacts with whom, etc. We usually represent a social network as a graph $G = (V, E)$, where V is a set of nodes (or

individuals) and E is a set of edges that captures a specific relationship (e.g., friendship, co-authorship, etc.) between the nodes. The edges in a graph can be *directed* or *undirected* based on whether the order of nodes are important on the underlying relationship; for example, who-trusts-whom networks are directed but co-authership networks are undirected. A social network can also be represented by its adjacency matrix $\mathbf{A} = [a_{ij}]$ where a_{ij} represents the weight (or presence) of the edge between nodes i and j (or from i to j for directed graphs). To classify, compare, and characterize different large-scale networks, various statistical measures and metrics have been developed for capturing the structural properties of networks and providing meaningful insight into network structures. These structural metrics—some of which have been originally developed in the *social network analysis* literature [317, 347]—have established the quantitative foundation for the study of networks in network science [279, 26, 280, 284, 7, 45]. One can divide structural metrics into two categories. *Micro-level measures* (e.g., centrality) describe how a node relates to the network whereas *macro-level measures* (e.g., degree distribution and diameter) provide a more holistic perspective on networks.

Centrality Measures

A large body of literature deals with the concept of *centrality*, addressing the question of which nodes are the most "important" or "central" in a network. Of course, centrality or importance can be defined in various ways, consequently many centrality metrics have been proposed (see, [279, 187, 118] for an overview). *Degree centrality* is probably the simplest centrality metric in which the degree of a node (in-degree or out-degree of a node for directed networks) represents its level of importance or centrality. However, degree centrality ignores many interesting aspects of the network.

Closeness centrality measures how close a given node is to other nodes in the network. A simplest form of closeness centrality is the inverse of the average geodesic distance (i.e., the average shortest path length) of a given node to all other nodes. The other variants can be defined using the harmonic mean distance (i.e., the average of inverse distances) or using *decay* parameters which penalize longer paths.

Betweenness centrality, originally proposed by Freeman [138], measures the number of times that a node i acts as a bridge on the shortest paths between all other nodes (i.e., the number of times that the shortest paths between all other pairs of nodes pass through node i). When there is some sort of flow in the network (e.g., passing messages, spreading diseases, trading between individuals, etc.), this measure is of interest when addressing the question of the importance of nodes in facilitating that flow. *Random-walk betweenness* [286] is based on the number of times that random walks pass a given node rather than counting all shortest paths passing through a given node.

The class of *prestige-based and eigenvector-based centrality* measures is built upon the premise that a node's importance is determined by the extent to which its neighbours are important. Using this self-referential concept of "importance," Bonacich [53] proposed *eigen-*

vector centrality in which a node's centrality score is its corresponding element in the network eigenvector which has the highest eigenvalue (right eigenvector with the highest eigenvalue for directed graphs). In a similar spirit, in *Katz centrality* [197], each node receives a small constant centrality score and a fraction of its neighbours centrality scores. Google's PageRank [289, 225] is a variant of both eigenvector centrality and Katz centrality, where each node's *pagerank* consists an equally shared fraction of the total pagerank and the summation of its neighbours' pagerank divided by their degree. For directed networks, Kleinberg [213] put forward the idea of *authority centrality* and *hub centrality* which quantify the importance of nodes in two distinct roles, being an *authority* and a *hub*, respectively. In the context of information networks (e.g., world wide web), authorities are those documents (or web pages), containing the information on a topic of interest whereas hubs point to these authorities.

Edge Density and Degree Distribution

The *Edge density* (sometimes referred to as *network density*) of a network is the relative fraction of existing edges in the network to all possible edges (or equivalently, the average degree divided by $n-1$). Given this definition, a full-connected network has edge density one. Generally speaking, when the edge density is close to zero, the network is very *sparse*, as opposed to *dense* networks with edge density close to one. Real-world social and biological networks are usually sparse, with low edge density [187, 284, 7].

One of the most fundamental topological properties of a network is its *degree distribution*, describing the relative frequency of node degrees. More precisely, $\mathbb{P}(d)$ is the fraction of nodes that have degree d (or the probability that a randomly selected node has degree d) under distribution P. It has been observed that many real-world networks (e.g., transportation networks, computer networks, world wide web, etc.) have degree distributions that obey a power-law distribution [25, 7, 284] with various exponents. In contrast, social networks are not well-explained by power-law distributions, especially *acquaintance* social networks [10, 338]. However, one might consider *collaboration networks*, in which an edge represents the collaboration of two individuals in an activity (e.g., being co-actor in a movie, or co-auther of a paper), as a type of social network with power-law distribution [303, 25] (sometimes, with an exponential cutoff [281]).

Giant Component, Diameter, and Average Path Length

One can use various measures to quantify and understand the connectivity of a network. We here briefly review some of these.

A (strongly-connected) *component* in a network is a maximal subset of nodes which are *strongly-connected* to each other (i.e., each node in the component is accessible by some path from any other). Real-world social networks often include a large component, referred to as the *giant component*, which usually contains more than half of the nodes in the network. The

giant component is observed in many real-world social, economical, biological, and technological networks (see, e.g., [284, 7]).

One of the mostly celebrated phenomena in social networks is the *small-world* effect, the discovery that the geodesic distance (path length) between nodes tends to be relatively small. *Diameter*—the length of the longest shortest path between any pair of nodes—and the *average shortest path length (ASPL)* are common metrics for capturing the path lengths in a network. Milgram [269] was a pioneer in experimenting with small world effects by discovering that the diameter of real-world social networks is around 6, thus resulting in popularizing the concept of "six degrees of separation". The small world effect is observed in the collaboration networks of actors [350] (with mean path length of 3.7) and co-authorship networks of mathematicians [172], physicists [285], and economists [170] (with average path length of 7.5, 5.9, and 9.5 respectively). Facebook possessed an average path length of 4.74 over 721 million users in 2011 [19]. Many real-world networks' diameters tend to shrink over the time as the network grows [232].

Transitivity and Clustering Coefficient

One of the predominant properties of social networks is *transitivity*. Treating connectivity in social networks as a transitive relation, if u is connected to v and v is connected to w, then u must be connected w. In common jargon, this means that "the friend of my friend is also my friend." While *perfect* transitivity yields a fully-connected graph or a network with components of cliques, it is rarely observed in real-world (social) networks. However, *partial* transitivity can be very useful: the fact that u is acquainted with v, and v knows w makes it much more probable that u knows w. To quantify the extent of transitivity of a network, one can use the *clustering coefficient*, defined as the fraction of the number of triangles over the total number of length-two paths. In social network jargon, a *closed triad* or triangle is a triple of mutually connected nodes. Social networks demonstrate a high clustering coefficient (e.g., 0.2 for an actor collaboration network [284], 0.09 for a biologist co-authorship collaboration network [281], and 0.16 for a university email network [119]). The high clustering coefficient in social networks can be partially explained by *triadic closure* [302], indicating that two people with a common friend are more likely to be acquainted with each other than two individuals without any common friend.[3]

Watts [349] defines the *local clustering coefficient* of a node to be the fraction of the *pairs* of neighbouring nodes that are themselves connected. Watts and Strogatz [350] suggest computing the clustering coefficient of a network as the mean of the local clustering coefficients over nodes

[3]One should be cautious in judging the transitivity of networks from their clustering coefficients. A large clustering coefficient is not necessarily an indication of the presence of transitivity or a "triadic closure" process, especially when the edge density is high. Because of this, the literature has chosen to compare the clustering coefficient with edge density to understand the significance of its value. The choice of edge density as a benchmark is not arbitrary and mainly motivated by the fact that the edge density and clustering coefficient is the same in purely random graphs. As such, comparing edge density with clustering coefficient tells us whether some special process (e.g., triadic closure) is taking place in the network.

rather than as defined above. The local clustering coefficient is a good indicator of *structural holes* in a network, which help explain on how information diffuses and how innovations arise and propagate over social networks [71]. When a node's neighbours are missing many edges between themselves, the missing edges are referred as *structural holes*. In this sense, structural holes lie mainly around nodes with low local clustering coefficient.

Correlations: Homophily and Assortativity

Correlations of nodes' attributes over networks can be taken into account in various ways; however, we here mainly focus on *homophily* or heterophily which captures the similarity or dissimilarity of connected nodes' internal attributes (e.g., race, gender, religion, preference, etc.) and *assortativity* which captures correlation patterns over the degrees of connected nodes.[4]

Homophily and Heterophily. The term *homophily* refers to the fact that people are more prone to associate with others whom they *perceive* to be similar to themselves in some way (e.g., age, race, education, language, etc.). While Burton [72] first noted the homophily effect and coined the well-known phrase *birds of a feather* in the network context, research on homophily has been widely studied in the social sciences (see McPherson *et al.* [266] for an overview). The impact of homophily in information diffusion [188, 74] and opinion formation [60, 160, 161] in social networks is also well studied. The opposite of homophily is *heterophily* in which individuals have strong tendency to interact with those who are different than them. The most apparent examples of heterophilous social networks are sexual contact networks, where the majority of sexual relationships are between people of opposite sex.

Assortativity. In networks with *positive assortativity*, nodes with relatively high degrees have a tendency to be connected to other high-degree nodes. Newman [282, 284] put forth the hypothesis that positive assortativity is a property of social networks by examining the degree correlation of nodes in some social networks. In constrast, technological and biological networks (e.g., Internet, power grid, neural networks) often exhibit negative assortativity, in which high-degree nodes are be linked to lower-degree nodes [282, 284]. Similarly, some economic networks exhibit negative assortativity (e.g., trading networks [323]). A related concept to positive assortativity is the presence of *core-periphery* structure in social networks, with a core of highly connected nodes and a periphery of less connected nodes [67].

2.3.2 Dynamics, Processes, and Learning

In most social and economic networks, there are dynamical processes where topological properties of the network play a role in dynamic behaviour of social and economic systems. These

[4]The terminology adopted here is more consistent with that of Jackson [187] as opposed to that of Newman [279] which refers to homophily by *assortative mixing* and to our "assortativity" by "assortative mixing by degree."

dynamics take various forms ranging from propagation of fads, fashion, technologies (see [154]) to formation of cultural and ideological opinions and beliefs, to movement synchronization in the crowds. The rough (and overlapping) categorization of such processes follows.

Contagion, Diffusion, and Epidemics

Social contagion (or sometimes, *emotional contagion* [179]) usually refers to the tendency of individuals to follow or mimic others' ideas, emotions, or habits [89, 88]. Information diffusion [306, 331] or rumor spreading processes [271, 102] are also considered as special type of social contagion, which captures the tendency of individuals to propagate knowledge, innovations, or rumours.[5] The use of the term of contagion is not arbitrary and of course evokes the concept of contagious diseases and their spreading patterns over social networks [221, 272, 292, 291, 293, 26]. The literature on information diffusion and social contagion also recognizes the similarities between social contagion processes and *epidemiological processes* [101, 102, 257], thus usually adopting epidemiological models as a basis for information diffusion models.[6] The interplay between network structure and diffusion is of practical import [326, 229]. As an example, network structure is observed to play a central role in "world-of-mouth" advertisement [326] and "viral marketing" [229].

Cascading Behaviour and Network Games

An individual's behaviour is enormously impacted by that of one's peers [88, 154]. Psychologists, socialists, and economists have attempted to understand and model how individuals' decisions and behaviour are influenced by those of their friends. Models of human behaviour on social network can be categorized into *stochastic* and *game-theoretic* models (see [187] for an overview). The stochastic approach considers the probability of an individual choosing an action (or behaviour) conditional on her neighbors' actions. One technique is to use Markov chains to model human behaviour, defining a *state space* of all possible combinations of individual actions while a weighted social network captures the probability of *imitation* of behaviour of their neighbors (self-loops represent the probability of taking action independently).[7] Given this behavioural Markov model, one can use steady-state probabilities to characterize individual behaviours. In addition to the intractability of finding the joint distribution, the other criticisms of this approach is that individuals are "backward looking" into the actions of their neighbors in the previous state. These shortcomings have motivated the development of game-theoretic models.

Network games capture the interdependency between the actions, decisions, and behaviours of individuals and those of their neighbours by modelling each interaction as a *game* [147, 190,

[5]The terms *information diffusion* or diffusion of innovations have used interchangeably in the literature.

[6]Recent research has questioned the validity of epidemic-like model of diffusion [155]. Unlike epidemic processes, most of cascades of information or adoption are observed to be small and terminate within one degree of the initial influencer or adopter [155].

[7]Ignoring the social network structure, Markov models of human behaviour are used in [207].

189, 362, 187, 168, 75, 365]. The games between individuals can be represented by *graphical games* [198], with an agent's *payoff* depending on its own and neighbours' actions.[8] As such, a player's behaviour is related to that of its neighbors, which depends on that of their neighbours' neighbours, and so on, thus resulting in cascading behaviour over the social network. Two classes of games have been studied in network games settings [187]. In a game with *strategic complementaries*, an agent's incentive for taking a specific action increases when more friends choose that action (e.g., choosing compatible operating systems, deciding whether to attend a party, etc). In contrast, a player's incentive decreases as more friends take similar actions in games with *strategic substitutes* (e.g., public good services). Usually with some additional assumptions, optimal strategies (e.g., Nash equilibrium strategies) in network games with strategic complementaries translate to simple *threshold rules*, determining whether a player adopts a behaviour given that a weighted fraction of its neighbors have adopted that behaviour (see [118] for an example how the Nash equilibrium strategy of a coordination game translates into a simple threshold rule, which has been studied for computational problems on technology adoption [200]). Different analyses of behaviour on network games can be conducted when the underlying games are precisely defined. One can predict an individual's behaviour by examining (Nash) equilibrium behaviour [187]. The relationship between equilibria and network structure can also be analysed [187, 168]. To understand cascading behaviour and diffusion dynamics, the *best-response* behaviour of individuals over time has been studied [44, 125, 273]. Jackson and Yariv [190] studied behaviour cascades and the stability of equilibria in more realistic models by introducing heterogeneity among agents.

Opinion Formation and Social Learning

Our friends, family, and acquaintances play a central role in the formation of our opinions and attitudes. Real-life examples are ubiquitous, ranging from our cultural attitudes to religious belief to judgements. Early work [226, 196] studies the formation of opinion in voting (and other decisions) and provides the foundation for identification of *opinion leaders*. Nowadays, the literature on *opinion formation* and *social learning* in social networks is rich (see, e.g., [187, 3]). This work mainly focuses on: (a) whether and how individuals with different opinions converge to a consensus; (b) how quickly individuals learn from each other; (c) which individuals have the greatest impact on the opinions of others or on the emerging consensus; and (d) whether social learning processes converge to "correct" beliefs. Two modelling approaches are common. *Bayesian learning* models provide firm normative foundations; however they are usually intractable and complicated to analyse when addressing questions of type (b) and (c). The alternative is more tractable *naïve learning rules* (e.g., weighted average updating [108]) which usually lack a firm theoretical foundation.

Observational Bayesian learning studies (see, e.g., [22, 146, 2]) usually investigate whether agents collectively learn and converge to the action with highest payoff by observing their

[8] Network games are closely related to the concept of *externalities* [121] discussed in Sec.2.1.4

neighbours' outcomes and actions when all are faced with the same type of uncertainty. The main premise is that an agent doing worse than one of its neighbours should realize this and switch to the neighbours' action with a better payoff. The main drawback of observational learning is that network structure does not come into play in analysing these models; thus, the impact of network structure on opinion formation process or identification of opinion leader cannot be fully understood.

The most celebrated weighted updating model of opinion formation is due to DeGroot [108]. Equipped with a real-valued opinion variable, each node iteratively updates its opinion by weighted averaging its neighbours' current opinions, with weights arising from underlying weighted directed social network. Different variations of DeGroot's model include a model with time-varying weights [109], a model with self-loops [142] (for always mixing some portion of initial opinion in updates), and others [143, 176, 139]. To guarantee the convergence of the updating process, Markov chain conditions of *aperiodicity* (i.e., the greatest common divisor of all directed cycle lengths is 1) and *irreducibility* (i.e., any node is reachable from any other node with non-zero probability) are usually assumed [199, 268]. Golub and Jackson [157] studied a slightly weaker set of assumptions for convergence of opinion formation processes; they also characterized the conditions under which a society converges to a "correct" opinion. The speed of convergence to consensus opinion is slower with increased homophily, where homophily is measured by proposed *spectral homophily* measure [160, 161]. More recently, Lobel and Sadler [243] discovered that network density is a factor in determining how homophily/hetrophily impacts the speed of convergence. They showed that for sparse networks, homophily is actually beneficial for social learning whereas hetrophily (i.e., diversity of preferences) is beneficial for dense networks.

2.3.3 Stochastic Network (Formation) Models

Of particular interest in network science [279, 26, 280, 284, 7, 45] is the development of network formation models that explain the emergence of common structural properties of real-world networks (e.g., [25, 350]). These mathematical network models also provide a theoretical and empirical framework for further investigation and understanding of the interplay between processes in complex systems and the structure of interacting entities. Generally speaking, network formation models can be divided into *strategic network models* and *random network models*. The former, mostly popularized in the economics literature, studies how networks form in some game-theoretic setting by understanding the incentive of individuals in making connections to each other [186, 187]. We here focus on the random network models which consider some stochastic process in the formation of edges between nodes.

Random Graphs: The Erdős-Rényi Model

Erdős and Rènyi [126] were among the pioneers of random network models. Their random graph model $G(n, p)$ assumes each edge between any pair of nodes occurs with independent

probability p. The random graph theory literature is rich [52, 116], where various structural properties of the Erdős-Rènyi model and its variants are studied. Under $G(n, p)$, node degrees exhibit a binomial distribution. Keeping mean degree constant and letting n grows, the degree distribution transforms to a Poisson distribution. The diameter can be approximated by $\frac{\ln n}{\ln c}$ where $c = p(n-1)$ is the expected degree. The clustering coefficient is p, as is edge density. This, in addition to several other properties (e.g., a lack of community structure and a Poisson degree distribution), distinguish $G(n, p)$ from real-world social networks, which tend to have a high clustering coefficient. In spite of its implausibility in representing real-world network, $G(n, p)$ is a very useful benchmark for understanding real-world networks and their underlying formation processes.

The Small-World Model and Its Variants

The *small-world model* [350] addresses the low clustering coefficient issue of random graphs in an elegant way. The model starts with a highly regular and clustered graph, then rewires each edge with probability p by uniformly random choosing a node as a new end point. The main point here is that by rewiring a few links in a highly clustered graph, we can create a network with short diameter but still maintain a high clustering coefficient. A more mathematically tractable variant, for each edge, *adds* a new edge between two randomly selected nodes, with probability p [287]. Many other variants have been proposed and studied (see, e.g., [11, 192, 214]).

Preferential Attachment and Its Variants

Of special interest are *growing network models*, capturing mechanisms by which networks grow. One popular model in this class is the *preferential attachment (PA)* model for scale-free networks [25]. It starts with n_0 initial nodes, then adds nodes in turn, with a new node connected to $k \leq n_0$ existing nodes. With probability proportional to its degree, each node is selected as a neighbour of recently added node. The emergent network has a power-law degree distribution (a fat-tailed distribution). The clustering coefficient is not sufficiently large in these networks for modelling real-world social networks. To address this, an extension is proposed by Holme and Kim [185] to incorporate a *triadic closure process* in preferential attachment. Other variants of PA include preferential attachment by fitness and popularity [39], and models in which nodes optimize certain trade-offs between popularity and similarity [290].

Spatial and Latent Space Models

In the class of *spatial (or latent space) networks* [27, 183], nodes have a set of real-valued, binary, or integer-valued *latent variables*, with the probability of an edge forming between two nodes determined by their attributes. In other words, it is the value of nodes' attributes which control the degree distribution, clustering, or small-world property of the observed network.

The *Waxman random graph model* [351]—predominantly used to model the internet and

other computer networks—distributes n nodes uniformly at random on the plane; then two nodes are connected with a probability decreasing exponentially with Euclidean distance between them. Hoff *et al.* [183] develop a similar model where nodes are points in a d-dimensional Euclidean "social space." The *random dot product model* [367] is somewhat distinct in that the distance between nodes is given by the inner product of their position vectors. The *hidden variable model* [48, 49] is a generalization of the Waxman model: nodes are equipped with a hidden (real-valued or integer) random variable drawn independently from a specified distribution. Two nodes are connected according to a symmetric probability function over node attributes. Serrano *et al.* [322] slightly modify this approach to introduce a general class of models based on a hidden metric space where nodes—located at a specific point in this space—are connected with a probability determined by a *connection probability function* over node distances.

The *random geometric graph (RGG)* model [295, 103] places n nodes uniformly at random in the d-dimensional unit cube, then connects nodes that are closer than some threshold parameter. Recently, geometric graphs in hyperbolic space have been studied [222]. The *multiplicative attribute graphs (MAG) model* [203, 206] generalizes the Kronecker graph model [230, 256], assigning each node a vector of categorical (integer-valued) attributes, and using a link-affinity matrix to capture connection probabilities based on attribute interactions.

Exponential Random Graph Models

A generalization of the Erdős-Rènyi model—mostly used for statistical analysis of observed social networks [165, 305]—is *exponential random graph model* [184], also known as Markov graphs [137], or p*networks [348, 165, 305]. This model defines a probability distribution $\mathbb{P}(G)$ over all possible graphs with n vertices in the exponential form of $\mathbb{P}(G) = \frac{1}{Z}e^{H(G)}$, where Z is a normalizing constant and $H(G) = \sum_i \beta_i x_i(G)$ is the *graph Hamiltonian*. Here, β_i represents a free model parameter associated with the particular measure of interest $x_i(G)$. Measures of interest can be any combination of topological measures (e.g., edge density, degree distribution, etc) or even a probability distribution of attributes.

2.3.4 Computational and Learning Problems

There has been growing interest in computational, learning, and inference problems on social networks ranging from social choice problems to optimization of contagion processes to inferring and predicting network structure.

Social Choice and Networks

Decision making on social networks in the presence of externalities has recently attracted considerable attention. The literature has tackled to various computational social choice problems such as stable matching [46], coalition formation [65, 66], voting [50, 51, 8, 337], auction design [175], and resource allocation [36] on social networks in the presence of allocative network ex-

ternalities.

Stable Matching and Coalition Formation. Bodine-Baron *et al.* [46] study stable matchings (e.g., of students to residences) with *peer effects*: these local social network externalities reflect the fact that students prefer to be assigned to the same residence as their friends. Brânzei and Larson address coalition formation on social networks where agent utility for a coalition depends on either her *affinity weights* with others in the coalition [65]; or her closeness centrality measure [66], as explained in Sec. 2.3.1.

Voting. Boldi *et al.* [50, 51] study *delegative democracy* on social networks, where an individual can either express her preferences directly, or delegate her vote to a neighbor. The weights of delegated votes are exponentially dampened by an attenuation factor that reduces the weight of a vote as it passes from one person to another. In a game-theoretic framework, Alon *et al.* [8] studied the "herding effect" in sequential voting procedure over only two alternatives, with agents having private preferences over alternatives and experiencing disutility if the winner is not the one that they vote for. Mechanism design for approval voting in social network has been studied in [9], where an edge in the underlying social network represents "who approves of whom." Recently, Tsang and Larson [337] study strategic voting on social networks with the presence of homophily. By observing their direct neighbours' ballots, voters can strategize their votes. The analyses suggest that homophily reduces the frequency of strategic voting.

Resource Allocation. Bhalgat *et al.* [36] focus on utilitarian social welfare maximization in unit-demand resource allocation problem with the the presence of positive externalities arising from social networks. In their model, each agent's overall utility is the *multiplication* of its *intrinsic valuation* function—mapping each alternative to a utility value—and its *externalities function*. Their externalities functions map the number of an agent's neighbours with the same assigned alternative to a numerical value.[9]

Optimization Problems and Social Dynamical Processes

There has been growing interest in optimization problems that exploit the underlying dynamics within social networks. Of special interest is the *influence maximization* problem, which finds a small subset of individuals who could maximize the spread of a fad, fashion, information, or particular behaviour (such as adopting a novel technology) [200, 274, 37, 79, 56, 169, 86]. Another interesting and relevant problem is that of *outbreak detection* [233] with the goal of "efficiently" and in a timely manner detecting the spread of disease, news, etc. on social networks.

[9]There has been growing interest in price setting [16, 76, 77] or strategic marketing [178, 38] over social networks, which can be viewed as decision making problems over social networks.

Influence or Cascade Maximization. In *influence maximization*, first formalized by Domingos and Richardson [113, 304], the goal is to find k optimal "seeds" or initial adopters, that yield the maximum possible cascade of behaviour (e.g., adoptation of a technology) over a social network. Assuming either the *independent cascade model* (similar to the models in Sec. 2.3.2) or the *linear threshold model* (as explained in Sec. 2.3.2), Kempe *et al.* [200, 201] show that the influence maximization problem is NP-Hard. They propose a natural greedy algorithm with a guaranteed $(1 - 1/e)$ approximation bound. Different variations of influence maximization are studied, including those in a *competitive setting* where multiple technologies compete to initiate a large cascade by targeting different sets of individuals [37, 79, 56, 169], and those which target influential groups of individuals [122]. In a competitive variant of influence maximization, Borodin *et al.* [55] study the strategic behaviour and incentive of companies to misreport their budget constraints in a mechanism design problem. They propose polynomial-time approximate strategy-proof mechanisms to address this problem.

Recently, Chierichetti *et al.* [86] studied the scheduling problem for cascade maximization in social networks. Assuming that individuals make decisions sequentially over a given social network with the presence of positive externalities, the goal is to find the "optimal" ordering of individuals whose sequential decision making induces the maximum cascade (i.e., the maximum spread of a technology). Assuming an Erdős-Rènyi model and considering a society with two agent types, namely "imitators" (blindly imitating friends' decisions according to their internal policy), and "informed adopters" (making a informed decision to adopt a technology), Zhang and Marbach [369] show that an imitator agent's "optimal" policy is to adopt a technology as soon as two neighbours have already adopted it.

Learning and Inference on Social Networks

The literature on inference and learning problems on social network is rapidly growing. Various learning and inference problems on structures and dynamical processes of social networks are being studied, dealing with both *complete* and *missing* data settings.

Link Prediction. The inference problem of predicting which future interactions between members of a society are likely to happen is the *link prediction problem* [244, 237]. A tightly related problem in complex networks is *link mining*, with the goal of inferring which link exist but are not observed [244, 150, 90]. Indeed, there is no clear-cut difference between these two literatures as both attempt to predict the edge occurrence between a pair of nodes using some *proximity* concept, arising from their local structural information [237, 370] or internal associated attributes [333]. Recently, Backstrom and Leskovec [20] proposed a *supervised random walk* algorithm for link prediction that exploits a node's attribute and structural information. An interesting variation of the original problem is predicting the signs of the social relationship in addition to predicting link occurrence [231], where the sign captures either being "friend" or "foe." See surveys by Lü and Zhou [244] and Getoor and Diehl [150] for an introduction to this

literature. A related problem is *community detection* in social networks [134, 279], where a *community* or *cluster* is a collection of nodes highly connected to each other but loosely connected to the other nodes. Another problem related to link prediction problem is the *perfect network formation problem* in the peer-to-peer literature (see, for example, [354, 355, 356]) where each node has a noisy observation of "optimal" edges in a "ground truth" network. The goal is to design a distributed (or sometimes, centralized) algorithm for discovering nodes' optimal links.

Inferring Diffusion Networks and Cascade Paths. Generally speaking, it is more straightforward to observe an individual's state (e.g., whether it is infected, has purchased a product, or has received some information) than the causal effect involved in the change of individual's state (e.g., who infects whom, who influences whom to buy a product, who listens to whom). As such, the social networks over which diffusion or propagation occur might be unobserved or unknown. The main problem here is to infer and discover the underlying hidden network of diffusion [5, 164, 163, 278]. Researchers have used various machine learning approaches to tackle this problem. One approach is to formulate the problem as a supervised classification problem [5]. The more statistically sound approach is to infer latent social network using some generative probabilistic models, capturing the correlation of individuals' state changes over time [164, 163, 278]. This approach is related to the structure learning in directed probabilistic graphical models [144, 216, 149].

A closely related problem to inferring diffusion networks is the problem of *inferring cascade paths*, with the goal of reconstructing (or sometimes characterizing) the tree structure of a partially observed cascade [238, 85, 159, 4, 114]. In contrast to inferring diffusion networks, the focus here is on understanding and reconstructing a large-scale cascade's pathway (e.g. chain petition letters [238]) rather than understanding who influences or infects whom.

Latent Nodal Attribute Inference. By adopting spatial network formation models (see Sec. 2.3.3), a body of research has focused on inferring latent attributes of nodes when social network structure is (fully) observed [183, 204, 205].[10] Hoff *et al.* [183] study inference and learning techniques on a spatial latent network model in which nodes are characterized by points in a fixed d-dimensional Euclidean "social space". Using *maximum likelihood estimation (MLE)* and *Markov Chain Monte Carlo (MCMC)*, the main objective is to infer the positional attributes of individuals in the latent space given fully observed social structure. By adopting the *multiplicative attribute graph (MAG)* model [206] (as explained in sec. 2.3.3), Kim and Leskovec [204] model a social network with nodes possessing categorical attributes. They develop a scalable *variational expectation maximization* method for (a) learning the model's parameters when the network structure and node attribute are given, and (b) inferring a node's latent variables and model parameters when the network structure is observed. In other work

[10] Less attention, it seems, is given to partially observed networks. Indeed, this is an appealing, practical direction for further exploration but may pose considerable computational challenges.

with similar aims, Kim and Leskovec [205] consider each node as belonging to multiple (possibly overlapping) latent groups which govern the social network structure and the nodes' categorical attributes; the MAG model is employed to capture the interplay of individuals' latent groups in network formation. In addition to learning model parameters and inferring a node's latent groups, they investigate the performance of their model for three applications: *missing feature prediction*, *link prediction*, and *supervised node classification.*

In the machine learning literature, there are related research areas, closely related to node's attribute learning and inference problems in social network, but not explicitly explored for this purpose. In *collective classification of networked data* [321], the goal is to *collectively* predict the labels of nodes (e.g, documents) in a graph (e.g. citation networks, or the web) while exploiting correlations between internal features of nodes and their connectivity to each other.[11] In *active learning in graphs or network data* [40, 255, 371, 324], the general goal is to minimize the cost of node labelling (usually, by selecting the minimum number of nodes required to be labelled) in order to have acceptably "accurate" collective classification of network data.

Collaborative Filtering and Social Networks. Collaborative Filtering (CF) methods which exploit social networks for rating prediction have recently become popular (see for example, [252, 191, 253, 361, 156, 260, 219] and [359] for a recent survey). These methods—along with traditional CF methods—fall into two broad categories, *memory-based* [156, 260, 219] and *model-based* [252, 191, 253, 361] approaches. In memory-based approaches, the social network structure is usually taken into account when computing the pairwise similarity scores (or trust values) between users [156, 260]. These scores are then used for prediction of missing ratings. Model-based approaches focus largely on latent space probabilistic models in which users and items are embedded in a low-dimensional latent feature space, and ratings are generated by combining these feature vectors while accounting for social network structure.

[11] One can readily observe the similarity of collective classification and the latent group membership problem [205] explained above; however, this connection has not yet been noted in the literature.

Chapter 3

Empathetic Social Choice on Social Networks

This chapter focuses on the problem of *consensus decision making (or group recommendation)* on social networks in the presence of *empathy*. Specifically, the goal is to select a single option from a set of options for some group connected by a social network, e.g., a local constituency electing a political representative, or friends selecting a vacation spot or a movie. While individuals have personal *intrinsic utility* over the options, we also incorporate a novel form of *empathetic utility* on social networks: the utility (or satisfaction) of an individual with an alternative a is a function of both her intrinsic utility for a and her *empathetic utility* for the "happiness" of her neighbours with the selected option. Empathetic utility in this sense reflects the fact that a person's happiness may be influenced by the happiness of others with whom they are connected [135]. This inherent interdependency of agent utilities is captured in related models in the economic literature, including those that deal with empathy, envy, and other forms of "other regarding" preferences [254, 208, 41]

We consider two varieties of empathetic preference. In our *local empathetic model*, the utility of individual i for alternative a combines her intrinsic preference for a with the *intrinsic* preference of i's neighbours for a, where the weight given to j's preference depends on the strength of the relationship of i with j (and also the intrinsic/extrinsic tradeoff made by i). For instance, i may be willing to sacrifice some of her intrinsic preference for a restaurant if her colleagues are happier with the cuisine; and she defers more to her closer friends. In our *global empathetic model*, i's utility for a depends on her intrinsic preference and the *total utility* of her neighbours for a (not just their intrinsic preference): she wants her neighbours not only to be satisfied with a, but to have high utility, which depends on the utility of *their* neighbours, and so on. For example, in political voting, i may have a mild preference for a over b, but if b is strongly preferred by her closest neighbours and also by their neighbours and many others in the community, she might prefer to see b elected so she won't have grumpy neighbours for the next five years. Companies linked in complex supply chain may care about the success of their

suppliers and customers, and hence consider adopting industry-specific or economic policies in that light.

We describe methods for computing optimal options under the local and global models where by an optimal option, we mean an option that maximize social welfare (see details below). The former, unsurprisingly, corresponds to a simple form of *weighted preference aggregation*, or voting in which each agent implicitly "delegates" a portion of her vote to her neighbors. The latter, because individual utilities are interdependent—indeed, utility spreads much like PageRank values [289]—requires the solution of a linear system to determine the optimal option. We describe (mild) conditions under which such fixed points exist, and show that it too results in a form of weighted voting. We develop two scalable iterative algorithms for consensus decision making. We also generalize our empathetic framework to accommodate other social choice problems (e.g., constrained resource allocation, matching, etc.) and show how some of our theoretical results (e.g., fixed-point convergence) still hold. We also demonstrate that, in the general empathetic framework, these other social choice problems can be viewed as using weighted preference aggregation, and that our general empathetic model is not be subsumed by allocative externalities models. Our experiments show that the algorithm converges quickly in practice. Experiments also demonstrate the effectiveness of our algorithms and show that, in some settings, ignoring empathetic preferences results in incorrect decisions and high social welfare loss.[1]

3.1 Social Empathetic Models

This section outlines our basic social choice model, then describes our model of empathetic preference on social networks, and discusses related work.

3.1.1 The Social Choice Setting

Apart from empathetic preferences on a network, the social choice framework we adopt is standard. We assume a set of alternatives or options $\mathcal{A} = \{a_1, \ldots, a_m\}$ and a set of agents $\mathcal{N} = \{1, \ldots, n\}$. Each agent j has *intrinsic preferences* over $\mathcal{A}$ in the form of either a (strict) preference ranking $\succ_j^I$ or a (cardinal) utility function u_j^I (see Section 2.1.1). We describe preferences in terms of utility functions, but discuss below how to interpret voting procedures within our model.

Our goal is to select a consensus option $a^* \in \mathcal{A}$ that implements some social choice function f relative to the preferences of $\mathcal{N}$. For example, if agent utilities are dictated solely by intrinsic

[1] Computational models of empathy may prove relevant in online social applications, to address a recently observed decline in empathy among young adults in which online social networks and media may have a role [218].

preference and f is (utilitarian) *social welfare maximization (SWM)*, we select

$$a^* = \arg\max \sum_j u_j^I(a). \tag{3.1}$$

We assume that only one social welfare maximizing alternative exits for ease of exposition. If multiple winning alternatives exist, the tie can be broken based on some predefined policy (e.g., some lexicographic order). If preferences are given by preference rankings, f might correspond to some voting rule.[2]

3.1.2 Empathetic Preference on Social Networks

We depart from standard social choice by considering *empathetic preferences*, in which the preferences of one agent are dependent on those of other agents. We consider the specific case in which these influences are induced by connections in a social network (though the notion need not be confined to networks). We focus on utility functions rather than preference rankings, since these allow the direct expression of quantitative tradeoffs between intrinsic and empathetic preference.

We assume a directed weighted graph $G = (\mathcal{N}, E)$ over agents, with an edge jk indicating that j's utility is dependent on its neighbor k's utility, with the strength of dependence given by edge weight w_{jk}. A loop jj indicates that j's utility depends on her own intrinsic preferences (at certain points below we assume that all such loops exist). We assume $w_{jk} \geq 0$ for any edge jk, and refer to this as the *nonnegativity* assumption. The implication of the nonnegativity assumption is that an individual's overall utility cannot degrade when others' utilities increase, and cannot increase when other utilities decrease. This rules out the possibility of "jealousy" or "envy." For any individual $j \in \mathcal{N}$, we assume j's out-going edge weights are *normalized*: $\sum_k w_{jk} = 1$. We usually refer to this as the *normalization* assumption. This mild assumption is mainly to facilitate our theoretical analyses. In practice, as long as weights are bounded above, they can be always normalized when the positive self-loop assumption (explained below) is present. We treat missing edges as having weight 0, thus represent G with a weight matrix $\mathbf{W} = [w_{ij}]$. We generally think of these edges as corresponding to some relationship in a social network; see Fig. 3.1(a) for an illustration. For example, the edge between individuals 2 and 3 has weight 0.7, thus 2's overall utility heavily depends on the 3's utility (or satisfaction). We further note that all edge weights are positive and the normalization assumption holds for all nodes in this example.

We first consider pure consensus social choice scenarios in which a single option a is selected but will later generalize our framework to accommodate combinatorial domains. We take j's utility for a to be a linear combination of its own intrinsic preference for a and the empathetic preference derived from each of its neighbors $j \in \mathcal{N}$ where weights determine the relative

[2]Our model applies directly to more general social choice problems, such as assignment problems with network externalities, matching, etc., without difficulty.

importance of each neighbor. (General non-linear models are possible also.) Letting $e_{jk}(a)$ denote the *empathetic utility derived by j from k*, define j's utility $u_j(a)$ to be

$$u_j(a) = w_{jj}u_j^I(a) + \sum_{k \neq j} w_{jk}e_{jk}(a). \tag{3.2}$$

The ratio of w_{jj} to $\sum_{k \neq j} w_{jk}$ captures the relative importance of intrinsic and empathetic utility to j. Our framework does not *impose* empathetic preference: fully self-interested agents are represented by self-loops of weight 1.

We consider two ways of defining empathetic preferences. In the *local empathetic model*, we define $e_{jk}(a) = u_k^I(a)$; i.e., j's utility for a combines the intrinsic utilities of each of its neighbors (including itself if $w_{jj} > 0$):

$$u_j(a) = \sum_k w_{jk}u_k^I(a). \tag{3.3}$$

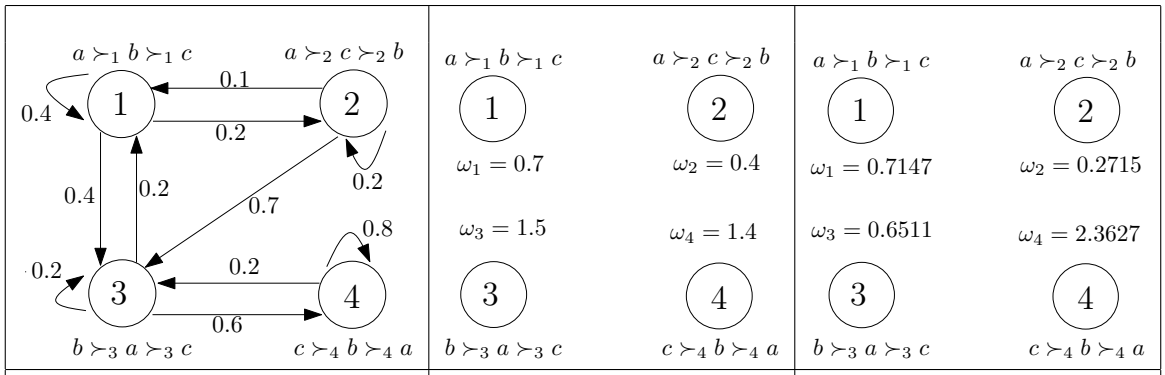
This model reflects agents j who are concerned about the "direct" preference of their neighbors k for a; but the fact that k's utility may depend on k's *own* neighbors does not impact j. Consider a family deciding on a movie: the preferences of certain family members (e.g., parents) for a film may depend on the preferences of others (e.g., children, whom they want to ensure are entertained).

In the *global empathetic model*, we define $e_{jk}(a) = u_k(a)$, so that k's total utility for a—which *may depend on k's neighbors*—influences j's utility for a, giving rise to

$$u_j(a) = w_{jj}u_j^I(a) + \sum_{k \neq j} w_{jk}u_k(a). \tag{3.4}$$

Here j's utility for a depends on the *utility*, not just intrinsic preferences, of its neighbors. For example, a voter may care about the overall satisfaction of her neighbors when voting for a political representative, but recognize that their satisfaction also depends on *their* neighbors, etc. Companies linked in complex supply chain may care about the success of their suppliers and customers, and consider adopting industry-specific or economic policies in that light. In the global model, the circular dependence of utilities requires a fixed point solution to the linear system of equations in Eq.3.6, defined using Eq. 3.4 for all $j \in \mathcal{N}$.

Correlations of behavior and/or preferences among agents connected in social network is widely accepted, and can be explained by a variety of mechanisms [118, 187] *information diffusion*, *network externalities*, or *homophily* (see Section 2.3 for details). Our empathetic model is somewhat different in that a person's *intrinsic* preferences over options $\mathcal{A}$ are not presumed to be correlated with their neighbors, but their *revealed* preferences might be: their choices (or stated utilities) reflect some consideration, however determined, of their neighbors' preferences.



(a) network and preferences | (b) local empathetic weights | (c) global empathetic weights

Figure 3.1: A social network with ranked preferences, with weights under the local and global empathetic model. Using Borda or plurality-based utility, the consensus winner is different in each model: a under intrinsic; b under local empathetic; c under global empathetic.

3.1.3 Weighting Agent Intrinsic Utilities

In realistic social choice situations, agents with empathetic preferences must often perform sophisticated reasoning about not only their own intrinsic preferences, but also those of their neighbors. Even in the local setting, expressing preferences (e.g., voting) is difficult since agents usually have incomplete (and in some cases, *no*) information about the preferences of their friends, neighbors, or colleagues. For instance, consider a group of friends who plans to decide on a movie to watch, from a list of twenty movies available in a theatre and assume each agent has on average 5 friends. In addition to reasoning about her own preferences over those 20 movies, an agent is required to understand and learn her friends' preferences over all of those movies ($5 \times 20 = 100$ cases, on average). This learning or reasoning might involve a considerable amount of communication and deliberation, even if an agent has few friends. The global empathetic setting is even more complex, since an agent is required to reason about her neighbors' connections as well as their intrinsic/empathetic tradeoffs.

In our models, preference aggregation and optimization are simpler: agents need only specify their *intrinsic preferences*, as is standard in social choice, and the *empathetic weights* they assign to their acquaintances. In social scenarios, this can remove a considerable informational and cognitive burden from agents who might otherwise be required to explicitly compute or otherwise determine their total utility for alternatives. In other settings, agents might not wish to reveal their preferences to their neighbors, but still want their neighbors to obtain a favorable result (e.g., companies voting on economic policy who are linked together in supply chain relationships which correlate their stability or profitability). Fortunately, given a network G, consensus decision making with empathetic preferences can be recast as a *weighted preference aggregation problem over intrinsic preferences alone.* This eases the burden on agents, and also allows one to recast the problem as simple weighted voting, or weighted (utilitarian) SWM, rendering the decision making process fully transparent. We present the model using SWM

(but draw connections to weighted voting later.)

In the local model, determining the weights associated with each agent's intrinsic preference is straightforward. Assume network weights $\mathbf{W}$. Let $\mathbf{u}(a)$ be the n-vector of agent utilities to be computed as a function of the corresponding vector $\mathbf{u}^I(a)$ of intrinsic utilities for a fixed alternative a. By Eq. 3.3, $\mathbf{u}(a) = \mathbf{W}\mathbf{u}^I(a)$. Letting $\boldsymbol{\omega} = \mathbf{e}^\top \mathbf{W}$ (where $\mathbf{e}$ is a vector of ones), the *(local) social welfare* of a is:

$$sw_l(a, \mathbf{u}^I) = \boldsymbol{\omega}^\top \mathbf{u}^I(a). \tag{3.5}$$

Thus SWM under the local model is simply the weighted maximization of intrinsic preferences, where the weight of j's intrinsic utility ω_j is the sum of its incoming edge weights. Fig. 3.1(b) shows the weights derived for each agent under the local model. For example, $\omega_3 = 1.5$ which derives from the summation of incoming edge weights for node 3: $\omega_3 = 0.2+0.4+0.7+0.2 = 1.5$. Note that ω_2 is the smallest and ω_3 is the largest among the societal weights in this example. This implies that the agent 3's intrinsic utility and agent 2's intrinsic utility have the most and least impact on social welfare, respectively, under the local model.

When intrinsic preferences are represented by rankings, one can adopt any scoring rule (e.g., Borda, plurality, k-approval, etc.) to transform intrinsic ranking preferences to intrinsic utilities. Indeed, using score-based voting rules, we can readily interpret this model as a form of *empathetic voting*, where the weight one assigns to a neighbor can be interpreted as the extent to which one would trade off one's own preferences with that neighbor's intrinsic satisfaction with the winning alternative.

We note that the optimal decision may be different in the local model than when only intrinsic preferences are used. Referring to Fig.3.1(a) and (b), we show how the decision is different for the intrinsic model and local model under two different scoring rules plurality and Borda. In the intrinsic model, the social welfare values for alternatives a , b, c under plurality are $sw_i(a) = 1 + 1 + 0 + 0 = 2$, $sw_i(b) = 1$, and $sw_i(c) = 1$, respectively. Under Borda, the social welfares in the intrinsic model is $sw_i(a) = 2 + 2 + 1 + 0 = 5$, $sw_i(b) = 1+0+2+1 = 4$, and $sw_i(c) = 0+1+0+2 = 3$. Thus, a wins in the intrinsic model under both plurality and Borda scoring. However, b wins in the local model for both scoring rules. One can compute the social welfares for the local model as follows: for plurality, social welfares are $sw_l(a) = \omega_1 + \omega_2 = 0.7 + 0.4 = 1.1$, $sw_l(b) = \omega_3 = 1.5$, and $sw_l(c) = \omega_4 = 1.4$; and for Borda, $sw_l(a) = 2\omega_1+2\omega_2+\omega_3 = 3.3$, $sw_l(b) = \omega_1+2\omega_3+\omega_4 = 5.1$, and $sw_l(c) = 0+\omega_2+0+2\omega_4 = 3.2$.

Things are more subtle in the global empathetic model. Computing the utility vector $\mathbf{u}(a)$ for alternative a requires solving a linear system of equations defined by:

$$\mathbf{u}(a) = (\mathbf{W} - \mathbf{D})\mathbf{u}(a) + \mathbf{D}\mathbf{u}^I(a), \tag{3.6}$$

where $\mathbf{D}$ is the $n \times n$ diagonal matrix with $d_{jj} = w_{jj}$. Each equation of this linear system corresponds to an individual (and is simply Eq. 3.4). A unique solution of this system is not

guaranteed to exist;[3] however, in addition to our assumptions above of *non-negativity* (i.e., $\mathbf{W} \geq \mathbf{0}$) and *normalization* (i.e., $\sum_k w_{jk} = 1$ for all j), a third mild condition on the social network $\mathbf{W}$ is sufficient to ensure a unique fixed point solution, namely, *positive self-loop*: $w_{jj} > 0$ for all j. This assumption implies that each individual is not completely "selfless". In other words, each agent cares, at least to some extent, about his/her own intrinsic preferences.

Proposition 3.1.1 (Fixed-point Utility)**.** *Assuming nonnegativity, normalizaton, and positive self-loop, Eq. 3.6 has a unique fixed-point solution* $\mathbf{u}(a) = (\mathbf{I} - \mathbf{W} + \mathbf{D})^{-1}\mathbf{D}\mathbf{u}^I(a)$.

(Proofs of all theoretical results of this chapter are included in Appendix A.) As in the local model, SWM in the global model can be seen as weighted maximization of intrinsic preference:

Corollary 3.1.1. *In the global empathetic model, (global) social welfare of alternative a is given by* $sw_g(a, \mathbf{u}^I) = \boldsymbol{\omega}^\top \mathbf{u}^I$ *where* $\boldsymbol{\omega}^\top = \mathbf{e}^\top(\mathbf{I} - \mathbf{W} + \mathbf{D})^{-1}\mathbf{D}$.

Once again, in (scoring-rule based) voting contexts one can view the global empathetic model as trading off one's own satisfaction for a winning alternative with the "overall" satisfaction of one's neighbors (not merely their intrinsic preference). Fig. 3.1(c) illustrates the distinctions: when either Borda or plurality is used as scoring rule in this example, alternative c wins under the global model but a wins in the intrinsic model and b wins in the local model as discussed above. One can compute social welfares to verify why c is the winner under the global model. For plurality, social welfares are $sw_g(a) = \omega_1 + \omega_2 = 0.7147 + 0.2715 = 0.9962$, $sw_g(b) = \omega_3 = 0.6511$, and $sw_g(c) = \omega_4 = 2.3627$. For Borda, $sw_g(a) = 2\omega_1 + 2\omega_2 + \omega_3 = 2.6235$, $sw_g(b) = \omega_1 + 2\omega_3 + \omega_4 = 4.3796$, and $sw_g(c) = 0 + \omega_2 + 0 + 2\omega_4 = 4.9969$. We discuss weight computation for the global model in Sec. 3.2.

Nonnegativity, normalization, and positive self-loop are not the only (collective) conditions under which fixed-point utilities are guaranteed to exist. Viewing the network as a Markov chain, if one assumes the chain is *aperiodic* (i.e., the greatest common divisor of all directed cycle lengths is 1) and *irreducible* (i.e., any node is reachable from any other node with non-zero probability) [199, 268], fixed point solutions are also guaranteed to exist. One might also adopt a weaker variation of our assumptions by imposing positive self-loop only on a subset of agents: this would suffice if all closed strongly-connected partitions of the network have at least one node with positive self-loop [158]. However, we believe our assumptions—and in particular, positive self-loop over all individuals—are appropriate in most social choice settings.

3.1.4 Related Models and Concepts

The term empathy is used in several different ways in the literature [133]. Sometimes it refers to "seeing the world through the eyes of others" without being affected by this view, and such preferences [41] or "extended sympathy" [319, 15] is used to frame interpersonal comparison of

[3] Consider two individuals j and k, with $w_{jj} = w_{kk} = 0$, $w_{jk} = w_{kj} = 1$, $u_j^I(a) = 0.1$, and $u_k^I(a) = 1$. The induced system does not have a unique fixed-point solution.

utilities [177, 41]. However, our model is more consistent with an affective understanding of another, and having concern for that person's welfare [228], or having "other-regarding" preferences [208]. Empathy has recently drawn attention in neuroeconomics and social neurosience [327] to study the extent people can place themselves in the position of others and share another's feelings. This further motivates computational study of empathy and its application to social choice.

The impact of others' actions and utilities is considered in some economic models (see, e.g., accounts of envy, sympathy/empathy in various contexts [254, 208, 41]). Most closely related to our work is the model of Maccheroni *et al.* [254], who establish the axiomatic foundations of interdependent "other-regarding" preferences in which the outcome experienced by others affects the utility of an agent. In their general formulation, the utility of an agent for an act incorporates both its subjective expected utility for that act and an *expected externalities function* over the agent's perceived social value of its own act and others' acts. While the general form of these externalities can model our notion of empathy, the specific axioms proposed for the application of their model (e.g., their anonymity axiom prevents the agent from distinguishing *which* of its peers attains a specific outcome) preclude its direct application to our setting.

Our work bears some connection to models of *opinion formation* and *social learning* in social networks (see Section 2.3.2 for a review). Our empathetic model can be viewed mathematically as a special case of a general model due to Friedkin and Johnson [142] (other special cases include [108, 139]). However, our goal in empathetic social choice is of course different: we capture preference interdependence in our model as a form of empathy, and focus on algorithms and mechanisms to implement a social choice function, not propagate beliefs.

Empathetic utilities can also be viewed as a form of network externality in an agent's utility function, though unlike typical models of externalities, an agent's utility depends on the utility of her neighbors for the chosen alternative rather than the behavior of, or the (direct) allocation made to, her neighbors (see Section 2.3.4 for a review of related work on this topic). We discuss these distinctions further in Sec. 3.3.

Our empathetic model bears some resemblance to certain centrality measures in social and information networks, which use (self-referential) notions of node *importance*. Some well-known examples include *eigenvector centrality* [53], *hubs and authorities* [213] and *PageRank* [289] (see Section 2.3.1 for details). Apart from conceptual differences and the fact that we address *decision* (social choice) problems, a key technical distinction is the use of self-loops in our empathetic model, which allows each node to contribute intrinsic utility to its fixed-point value.

3.2 Computing Winners

To compute the social welfare maximizing alternative, in both the local and global empathetic models, recall that social welfare can be expressed as $sw(a, \mathbf{u}^I) = \boldsymbol{\omega}^\top \mathbf{u}^I(a)$ for a suitable weight

vector $\boldsymbol{\omega}$. Given vectors $\mathbf{u}^I(a)$ for any $a \in \mathcal{A}$, we can compute the optimal option

$$a^* = \arg\max_{a \in A} \boldsymbol{\omega}^\top \mathbf{u}^I(a),$$

in $O(nm)$ time where n comes from the computation of social welfare for each option and m comes from iterating over all options. So we focus on: (i) computation of $\boldsymbol{\omega}$ in each model; and (ii) for the global model, a method for computing a^* without full computation of $\boldsymbol{\omega}$.

In the local model, $\boldsymbol{\omega}^\top$ requires only a single vector-matrix multiplication, $\boldsymbol{\omega}^\top = \mathbf{e}^\top \mathbf{W}$, taking $O(n^2)$ time. However, social networks are generally extremely sparse, with the number of incoming edges to any node j bounded by some small constant. In such sparse networks, $\boldsymbol{\omega}$ can be computed in $O(n)$ time since ω_j is simply the sum of j's incoming edge-weights; and a^* can be determined as above in $O(nm)$ time. Thus the complexity of computing optimal alternatives in the local empathetic model is no different than that of straightforward SWM or (weighted) voting.

In the global model, $\boldsymbol{\omega}^\top$ has a more complicated expression: $\boldsymbol{\omega}^\top = \mathbf{e}^\top \mathbf{A}^{-1}\mathbf{D}$ where $\mathbf{A} = \mathbf{I} - \mathbf{W} + \mathbf{D}$ (see Cor. 3.1.1). The difficulty lies largely in matrix inversion: $\mathbf{A}^{-1}$ can be computed via Gauss-Jordan elimination, which has complexity $O(n^3)$. This implies that straightforward computation of a^* requires $O(n^3 + nm)$ time. In general, matrix inversion is no harder than matrix multiplication [99, Thm. 28.2], but its complexity cannot be less than $O(n^2)$ since all n^2 entries must be computed. Therefore, straightforward computation of a^* in the global model cannot have complexity less than $O(n^2 + nm)$.

For large n (e.g., voting in large cities, Facebook, Twitter), algorithms that scale linearly (or better) in n are needed. Many iterative methods have been proposed for matrix inversion and solving linear systems (e.g., Jacobi, Gauss-Siedel, etc.) which have $O(n)$ complexity (in sparse systems) per iteration and tend to converge very quickly in practice. We now describe our technique that exploits a standard Jacobi method for computing a^* in the global model.

We consider a simple iterative method for computing $\mathbf{u}(a)$. Let $\mathbf{u}^{(t)}(a)$ be the vector of the estimated utilities of a after t iterations.

Theorem 3.2.1. *Consider the following iteration:*

$$\mathbf{u}^{(t+1)}(a) = (\mathbf{W} - \mathbf{D})\mathbf{u}^{(t)}(a) + \mathbf{D}\mathbf{u}^I(a). \tag{3.7}$$

Assuming nonnegativity, normalizaton, and positive self-loop, this method converges to $\mathbf{u}(a)$*, the solution to Eq. 3.6.*

For each $j \in \mathcal{N}$, the method computes:

$$u_j^{(t+1)}(a) = w_{jj} u_j^I(a) + \sum_{k \neq j} w_{jk} u_k^{(t)}(a), \tag{3.8}$$

where $u_j^{(t)}(a)$ is agent j's estimated utility for a after t iterations. This scheme has a natural

interpretation in terms of agent behavior: suppose that each agent repeatedly observes her friends' revealed utilities, and updates her own utility for various options in response. This process will converge—even if the updates are "asynchronous." Under this iterative process, the local empathetic model provides a first-order approximation to the global model if we simply let $u_k^{(0)}(a) = u_k^I(a)$. In other words, for this initialization, after the first iteration, we have computed the utilities for the local model. Critically, the error in the estimated utilities at the t^{th} iteration can also be bounded:

Theorem 3.2.2. *In the iterative scheme above,*

$$\left\|\mathbf{u}(a) - \mathbf{u}^{(t)}(a)\right\|_\infty \leq (1-\tilde{w})^t \left\|\mathbf{u}(a) - \mathbf{u}^{(0)}(a)\right\|_\infty,$$

where $\tilde{w} = \min_{1\leq i\leq n} w_{ii}$.

Hence, societies in which individuals have self-loops with relatively large weight (i.e., less empathy) converge to fixed-point utilities faster than societies with greater empathy (our empirical results below support this).

This error bound allows one to bound the error in *estimated social welfare* if the utilities of all options are estimated this way. Let $sw^{(t)}(a) = \sum_j u_j^{(t)}(a)$.

Theorem 3.2.3. *Assume* $u_j^I(a), u_j^{(0)}(a) \in [c,d]$, *for all* j *where* $c \leq d$ *are constants. Then* $\left|sw(a) - sw^{(t)}(a)\right| \leq n(d-c)\,(1-\tilde{w})^t$, *for all* t, *under the conditions above, where* $\tilde{w} = \min_{1\leq i\leq n} w_{ii}$.

As a result, we know that (under the same assumptions):

Proposition 3.2.4. *If* $sw^{(t)}(b) - sw^{(t)}(a) \geq 2n(d-c)\,(1-\tilde{w})^t$ *then* $sw(b) > sw(a)$.

We exploit Prop. 3.2.4 in a simple algorithm, that we call *iterated candidate elimination (ICE)*, for computing a^*. The intuition is simple: we iteratively update the estimated utilities of the subset $C \subset \mathcal{A}$ of options that are non-dominated, and gradually prune away any options that are dominated by another until only one, a^*, remains. ICE first initializes $C = \mathcal{A}$ and $u_j^{(0)}(a) = c$ for all $j \in \mathcal{N}, a \in \mathcal{A}$. An iteration of ICE consists of: (1) updating estimated utilities using Eq. 3.8 for all j and $a \in C$; (2) computing estimated social welfare of each $a \in C$; (3) determining the maximum estimated social welfare $\hat{sw}^{(t)}$; (4) testing each $a \in C$ for domination, i.e., $\hat{sw}^{(t)} - sw^{(t)}(a) \geq 2n(d-c)\,(1-\tilde{w})^t$; and (5) eliminating all dominated options from C. The algorithm terminates when one option a^* remains in C. The pseudo-code for the algorithm is provided in Algorithm 1. ICE runs in $O(tm|E|)$ time, where t is the number of iterations required; and if the number of outgoing edges is bounded, $O(tmn)$. As we demonstrate below in our experiments, ICE converges quickly in practice.

Algorithm 1: Iterated Candidate Elimination (ICE)

input : Social graph G, intrinsic utilities $u_i^I(a) \in [c,d]$, $\forall i \in \mathcal{N}$ and $\forall a \in \mathcal{A}$.
output: Consensus winner a^*.
Initialize $u_i^{(0)}(a) \leftarrow c$, $\forall i \in \mathcal{N}$ and $\forall a \in \mathcal{A}$;
// C is the possible winner candidate set
$C \leftarrow \mathcal{A}$;
$\tilde{w} = \min_{1 \le i \le n} w_{ii}$;
$t \leftarrow 0$;
while *size*$(C) > 1$ **do**
  $t \leftarrow t+1$;
  foreach $a \in C$ **do**
    $sw^{(t)}(a) \leftarrow 0$;
    foreach $j \in \mathcal{N}$ **do**
      $u_j^{(t)}(a) \leftarrow w_{jj} u_j^I(a) + \sum_{k:e_{jk} \in E, j \neq k} w_{jk} u_k^{(t-1)}(a)$;
      $sw^{(t)}(a) \leftarrow sw^{(t)}(a) + u_j^{(t)}(a)$;
  $\hat{sw}^{(t)} \leftarrow \max_{a \in C} sw^{(t)}(a)$;
  foreach $a \in C$ **do**
    if $\hat{sw}^{(t)} - sw^{(t)}(a) \ge 2n(d-c)(1-\tilde{w})^t$ **then**
      $C \leftarrow C - \{a\}$
return $a^* \in C$

3.3 A Generalized Empathetic Framework

Our empathetic model and its theoretical results applies directly to more general social choice problems (e.g., assignment, allocation, segmentation, etc.) problems. We present a reformulation of our model for these applications in this section. We let $x_i \in \mathcal{A}$ represent the assignment (or label) of individual i. This x_i is the alternative (or option) that is assigned (or matched) to an individual i. Let $\mathbf{x} = (x_1, \cdots, x_n)$ be the vector of all assignments. To demonstrate that generalized empathetic model can accommodate combinatorial problems (such as assignment, allocation, segmentation, or matching), we first show how utilitarian social welfare over the assignment vector can be decomposed as a weighted summation of intrinsic utilities under both the local and global models. Various social choice problems can be modelled by maximizing utilitarian social welfare under an appropriate set of constraints.

Letting $e_{jk}(x_k)$ denote the *empathetic utility derived by j from k when x_k is assigned to k*, define j's utility $u_j(x_j|\mathbf{x}_{-j})$ to be

$$u_j(x_j|\mathbf{x}_{-j}) = w_{jj} u_j^I(x_i) + \sum_{k \neq j} w_{jk} e_{jk}(x_k), \tag{3.9}$$

where $\mathbf{x}_{-j}$ denotes the vector of assignments for all individuals except j. In the *local empathetic model*, we define $e_{jk}(x_k) = u_k^I(x_k)$; i.e., j's utility combines the intrinsic utilities of each of its

neighbors (including itself if $w_{jj} > 0$):

$$u_j(x_j|\mathbf{x}_{-j}) = \sum_k w_{jk} u_k^I(x_k), \tag{3.10}$$

In the *global empathetic model*, we define $e_{jk}(x_k) = u_k(x_k|\mathbf{x}_{-k})$, so that k's total utility for x_k—which *may depend on k's neighbors*—influences j's utility for x_j, giving rise to

$$u_j(x_j|\mathbf{x}_{-j}) = w_{jj} u_j^I(x_j) + \sum_{k \neq j} w_{jk} u_k(x_k|\mathbf{x}_{-k}). \tag{3.11}$$

One can retrieve our empathetic model in Sec. 3.1 by posing the constraints $x_i = x_j$ for all $i, j \in \mathcal{N}$. Let $\mathbf{D}$ be the $n \times n$ diagonal matrix with $d_{jj} = w_{jj}$, We can write Eq. 3.11 as:

$$\mathbf{u}(\mathbf{x}) = (\mathbf{W} - \mathbf{D})\mathbf{u}(\mathbf{x}) + \mathbf{D}\mathbf{u}^I(\mathbf{x}). \tag{3.12}$$

We note that our proof for Prop. 3.1.1 directly apply for this equation and nonnegativity, normalizaton, and positive self-loop assumptions induce a unique fixed-point solution for utilities:

Proposition 3.3.1 (Generalized Fixed-point Utility)**.** *Assuming nonnegativity, normalizaton, and positive self-loop, Eq. 3.12 has a unique fixed-point solution* $\mathbf{u}(\mathbf{x}) = (\mathbf{I} - \mathbf{W} + \mathbf{D})^{-1}\mathbf{D}\mathbf{u}^I(\mathbf{x})$.

Social welfare for societal assignment $\mathbf{x}$ in the generalized local and global empathetic models, respectively, is given by

$$sw_l(\mathbf{x}, \mathbf{u}^I) = \boldsymbol{\omega}_l^\top \mathbf{u}^I(\mathbf{x}),$$

$$sw_g(\mathbf{x}, \mathbf{u}^I) = \boldsymbol{\omega}_g^\top \mathbf{u}^I(\mathbf{x}),$$

where $\boldsymbol{\omega}_l = \mathbf{e}^\top \mathbf{W}$ and $\boldsymbol{\omega}_g^\top = \mathbf{e}^\top (\mathbf{I} - \mathbf{W} + \mathbf{D})^{-1}\mathbf{D}$. This observation is the key to solve many social choice problems (e.g., assignments, matching, etc.) in our empathetic models. In this light, each social choice problem can be recast as its corresponding weighted version, with weights coming from the underlying social network. As an example, we show how the allocation (or assignment) problem with capacity constraints in the global empathetic model can be formulated. We first start with standard version of this problem:

$$\begin{aligned} \underset{\mathbf{x} \in D(\mathcal{A})^n}{\text{maximize}} \quad & sw(\mathbf{x}) \\ \text{subject to:} \quad & \sum_{i=1}^{n} I[x_i = a_j] \leq q_j, \forall \;\; 1 \leq j \leq m \end{aligned} \tag{3.13}$$

Here, the utilitarian social welfare is $sw(\mathbf{x}) = \sum_i u^I(x_i)$, $D(\mathcal{A}) = \mathcal{A} \cup \{\emptyset\}$, $I[.]$ is the indicator function, and $q_j \in \mathbb{N}$ is the quota (or capacity) for alternative a_j. $D(\mathcal{A}) = \mathcal{A} \cup \{\emptyset\}$ encodes all possible outcomes for each agent in our assignment problem (i.e, being assigned

to any of alternatives in $\mathcal{A}$ or none of them represented by $\emptyset$). Thus, $D(A)^n$ is the set of all possible assignments. We can obtain the the allocation (or assignment) problem with capacity constraints under global model simply by replacing social welfare to its corresponding weighted version:

$$\begin{aligned} \underset{\mathbf{x}\in D(\mathcal{A})^n}{\text{maximize}} \quad & \boldsymbol{\omega}_g^\top \mathbf{u}^I(\mathbf{x}) \\ \text{subject to:} \quad & \sum_{i=1}^{n} I[x_i = a_j] \leq q_j, \forall \;\; 1 \leq j \leq m \end{aligned} \tag{3.14}$$

Here, $D(A)^n$, again, is the set of all possible assignments (as discussed above). By replacing $\boldsymbol{\omega}_g$ with $\boldsymbol{\omega}_l$, we obtain the resource allocation under the local empathetic model. The standard non-empathetic setting is recovered by setting all ωs to 1 (or any positive constant value). This weighted maximization problem, in the general case, can be viewed as a *maximum weighted matching* problem on a bipartite graph, where individuals are on one side and alternatives—with repetitions determined by their quotas—are on the other side (dummy nodes can be added, if necessary). The weight on the edge between an individual i and alternative a_j is given by $\omega_i u_i^I(a_j)$.

As another example, we demonstrate how multi-winner elections can be accommodated in the empathetic model. We consider slates $S \subseteq \mathcal{A}$ and let k be the fixed size of allowable slates. We start with non-empathetic, standard problem:

$$\begin{aligned} \underset{S\subseteq\mathcal{A},\mathbf{x}\in S^n}{\text{maximize}} \quad & sw(\mathbf{x}) \\ \text{subject to:} \quad & u_i^I(x_i) \geq u_i^I(y), \forall \;\; y \in S, i \in \mathcal{N} \\ & |S| = k \end{aligned} \tag{3.15}$$

Here, the utilitarian social welfare is $sw(\mathbf{x}) = \sum_i u^I(x_i)$. The goal is to select slate $S \in \mathcal{A}$ with cardinality K such that it maximizes utilitarian social welfare defined based on the utilities that agents derive from their most preferred item on the slate. We can obtain the the multi-winner election problem under the global model simply by replacing social welfare with its corresponding weighted version:

$$\begin{aligned} \underset{S\subseteq\mathcal{A},\mathbf{x}\in S^n}{\text{maximize}} \quad & \boldsymbol{\omega}_g^\top \mathbf{u}^I(\mathbf{x}) \\ \text{subject to:} \quad & u_i^I(x_i) \geq u_i^I(y), \forall \;\; y \in S, i \in \mathcal{N} \\ & |S| = k \end{aligned} \tag{3.16}$$

We have initially changed $u_i^I()$ to $\omega_i u_i^I()$ in the constraints. However, ω_i on both side of the inequalities are cancelled out, thus making the inequality constraints only on intrinsic

utilities. By replacing $\boldsymbol{\omega}_g$ with $\boldsymbol{\omega}_l$, we obtain the multi-winner election problem under the local empathetic model. The standard non-empathetic setting is recovered by setting all ωs to 1 (or any positive constant value).

Some other social choice problems which focus on maximizing social welfare in presence of some constrains (e.g., segmentation, stable matching with utilities, etc.) can similarly be accommodated in our generalized empathetic model.

3.3.1 Iterative Methods for Societal Weight Computation

As noted, other social choice problems (e.g., assignment, multi-winner elections, matching, etc.) can be viewed as their corresponding weighted versions under our empathetic models when weights come from the underlying social network structure. We here discuss various methods for computation of these societal weights. The broad use of societal weights for various social choice problems motivate us to propose a scalable iterative method for estimating weights and their theoretical bounds. This iterative method is a building block for algorithms to solve other social choice problems in empathetic settings.

The problem of weight computation can be viewed as a linear system of equation:

Corollary 3.3.1. *$\boldsymbol{\omega}$ is the unique solution to the linear system of $\mathbf{A}\boldsymbol{\omega} = \mathbf{e}$ where $\mathbf{A} = (\mathbf{I} - \mathbf{W}^\top + \mathbf{D})\mathbf{D}^{-1}$.*

We now briefly describe the use of a standard Jacobi iterative method for estimating weights $\boldsymbol{\omega}$ in the global model.[4] Let $\boldsymbol{\omega}^{(t)}$ be the estimated weights after t iterations.

Theorem 3.3.2. *Consider the following update:*

$$\boldsymbol{\omega}^{(t+1)} = \mathbf{D}(\mathbf{W}^\top - \mathbf{D})\mathbf{D}^{-1}\boldsymbol{\omega}^{(t)} + \mathbf{De}$$

Assuming nonnegativity, normalizaton, and positive self-loop, this method converges to $\boldsymbol{\omega}$, the solution to linear system stated in Cor. 3.3.1.

For each $j \in \mathcal{N}$, this iterative method computes

$$\omega_j^{(t+1)} = w_{jj} + \sum_{k \neq j} \frac{w_{jj}}{w_{kk}} w_{kj} \omega_k^{(t)}(a). \tag{3.17}$$

where $\omega_j^{(t)}$ is the agent j's estimated societal weight after t iterations. One can readily bound the error of estimated weights after t iterations:

Theorem 3.3.3. *Assume $\boldsymbol{\omega}^{(0)} = (w_{11}, w_{22}, \ldots, w_{nn})^\top$. In the iterative scheme above,*

$$\left\| \boldsymbol{\omega} - \boldsymbol{\omega}^{(t)} \right\|_1 \leq n \frac{\hat{w}}{\tilde{w}} (1 - \tilde{w})^t (1 - \bar{w}),$$

[4] Recall that for the local model, $\boldsymbol{\omega}$ can be efficiently computed in $O(|E|)$ time as explained in Sec. 3.2.

where $\tilde{w} = \min_{1\leq j\leq n} w_{jj}$, $\hat{w} = \max_{1\leq j\leq n} w_{jj}$, *and* $\bar{w} = \frac{1}{n}\sum_j w_{jj}$.

Hence, this iterative method converges faster for societies in which individuals have self-loops with relatively larger weight (i.e., less empathy) compared to societies with greater empathy. The error bound on $\boldsymbol{\omega}$ allows one to bound the error in estimated social welfare

$$sw^{(t)}(\mathbf{x}) = \sum_j \omega_j^{(t)} u_j^I(x_j),$$

for the assignment vector $\mathbf{x}$.

Theorem 3.3.4. *Assume* $\boldsymbol{\omega}^{(0)} = (w_{11}, w_{22}, \ldots, w_{nn})^\top$. *Under normalization, nonnegativity, and self-positive loop, for any* t*:*

$$|sw(\mathbf{x}) - sw^{(t)}(\mathbf{x})| \leq n\frac{\hat{w}}{\tilde{w}}\left(1-\tilde{w}\right)^t\left(1-\bar{w}\right)\left\|\mathbf{u}^I(\mathbf{x})\right\|_2,$$

where $\tilde{w} = \min_{1\leq j\leq n} w_{jj}$, $\hat{w} = \max_{1\leq j\leq n} w_{jj}$, *and* $\bar{w} = \frac{1}{n}\sum_j w_{jj}$.

This bounds the error in our estimated social welfare at iteration t. As t grows, the error shrinks since $1-\tilde{w} < 1$. Due to having n and $\left\|\mathbf{u}^I(\mathbf{x})\right\|_2$ on the right side, for larger n, we require a greater number of iterations to obtain reasonable approximations. As a result of Theorem 3.3.4, we know that (under the same assumptions):

Proposition 3.3.5. *If* $sw^{(t)}(\mathbf{x}) - sw^{(t)}(\mathbf{y}) \geq n\frac{\hat{w}}{\tilde{w}}\left(1-\tilde{w}\right)^t\left(1-\bar{w}\right)\left(\left\|\mathbf{u}^I(\mathbf{x})\right\|_2 + \left\|\mathbf{u}^I(\mathbf{y})\right\|_2\right)$ *then* $sw(\mathbf{x}) > sw(\mathbf{y})$.

Using this proposition, by comparing the estimated social welfares of two options, one can assess the relative magnitude of their actual social welfares. We exploit this below.

To demonstrate the practical impact of these theoretical results, in our experiments below we show how these results can be deployed for solving a simplified allocation problem as an example. Moreover, we can use these weight approximation methods to develop another iterative algorithm for solving consensus decision making problems which are the main focus of this chapter. We call this new algorithm *weight-based iterated candidate elimination (WICE)*. For the ease of presentation, we present and explain WICE using our earlier notations in this chapter for the specific problem of consensus decision making. The intuition behind WICE is to iteratively update the estimated weights $\boldsymbol{\omega}^{(t)}$ and accordingly calculate the estimated social welfare of the subset $C \subset \mathcal{A}$ of candidates that are non-dominated, and gradually prune away any candidate that is dominated by another until only one, a^*, remains.

More precisely, WICE first initializes $C = \mathcal{A}$ and $\omega_j^{(0)} = w_{jj}$ for all $j \in \mathcal{N}$. An iteration of WICE consists of: (1) updating estimated weights using Eq. 3.17 for all $j \in \mathcal{N}$; (2) computing estimated social welfare of each $a \in C$; (3) determining the maximum estimated social welfare $\hat{sw}^{(t)}$ and its corresponding alternative $\hat{a}$; (4) testing each $a \in C$ for domination, i.e., $\hat{sw}^{(t)} - sw^{(t)}(a) \geq n\frac{\hat{w}}{\tilde{w}}\left(1-\tilde{w}\right)^t\left(1-\bar{w}\right)\left(\left\|\mathbf{u}^I(\hat{a})\right\|_2 + \left\|\mathbf{u}^I(a)\right\|_2\right)$; and (5) eliminating all dominated

Algorithm 2: Weight-based Iterated Candidate Elimination (WICE)

input : Social graph G, intrinsic utilities $u_i^I(a) \in [c,d]$, $\forall i \in \mathcal{N}$ and $\forall a \in \mathcal{A}$.
output: Consensus winner a^*.
Initialize $\omega_j^{(0)} \leftarrow w_{jj}$, $\forall j \in \mathcal{N}$;
// C is the possible winner candidate set
$C \leftarrow \mathcal{A}$;
$\tilde{w} = \min_{1 \le j \le n} w_{jj}$;
$\hat{w} = \max_{1 \le j \le n} w_{jj}$;
$\bar{w} = \frac{1}{n} \sum_j w_{jj}$;
$t \leftarrow 0$;
while *size*$(C) > 1$ **do**
  $t \leftarrow t+1$;
  foreach $j \in \mathcal{N}$ **do**
    $\omega_j^{(t)} \leftarrow w_{jj} + \sum_{k \neq j : e_{jk} \in E} \frac{w_{jj}}{w_{kk}} w_{kj} \omega_k^{(t-1)}$;
  foreach $a \in C$ **do**
    $sw^{(t)}(a) = \left(\boldsymbol{\omega}^{(t)}\right)^\top \mathbf{u}^I(a)$
  $\hat{a} = \arg\max_{a \in C} sw^{(t)}(a)$;
  $\hat{sw}^{(t)} \leftarrow sw^{(t)}(\hat{a})$;
  foreach $a \in C$ **do**
    if $\hat{sw}^{(t)} - sw^{(t)}(a) \geq n \frac{\hat{w}}{\tilde{w}} (1-\tilde{w})^t (1-\bar{w}) \left(\left\| \mathbf{u}^I(\hat{a}) \right\|_2 + \left\| \mathbf{u}^I(a) \right\|_2 \right)$ **then**
      $C \leftarrow C - \{a\}$
return $a^* \in C$

candidates from C. The algorithm terminates when only one candidate (i.e., a^*) remains in C (the pseudo-code is presented in Algorithm 2). The running time for each iteration of WICE is $O(|E| + mn)$, where $|E|$ is the number of edges.

Comparison of ICE and WICE algorithms. The running time comparison of ICE vs. that of WICE is not straightforward. Each iteration of WICE runs faster than ICE: $O(|E| + mn)$ vs. $O(|E|m)$. But, we expect ICE to converge much faster (in fewer iterations) than WICE by comparing their domination conditions (see Prop. 3.2.4 and Prop. 3.3.5). If so, ICE is more efficient for small number of alternatives and large scale networks while WICE is well-designed for large scale network and high number of alternatives. We study this in our experiments below.

3.3.2 Generalized Empathetic Model vs. Allocative Externalities Models

Empathetic utilities can be viewed as a form of network externality in an agent's utility function. But unlike typical models of *allocative externalities* (see Section 2.3.4 for a review), an agent's utility depends on the utility of her neighbors for the chosen alternative rather than the behavior of, or the (direct) allocation made to, her neighbors. More specifically, individual i's extrinsic utility in our empathetic model derives from i's neighbours' utilities, whereas in the allocative

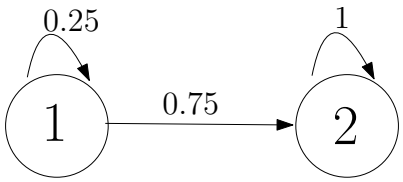


Figure 3.2: The social network of our counter example for demonstrating that the empathetic model is not subsumed by the metric labelling model.

externalities model, extrinsic utilities are determined by similarity/dissimilarity (or distance) between i's assignment and each of its neighbours' assignments.

To crystallize this distinction, we here show that the generalized empathetic model cannot be subsumed by the general allocative externalities model where we adopt the model of *metric labelling* [212] as a general model for allocative externalities.[5] The utility of individual i under the metric labelling model can be written as:

$$u_i(x_i|\mathbf{x}_{-i}) = w_{ii}u_i^I(x_i) + \sum_j w_{ij}u'(x_i, xj)$$

where x_i represents the assignment (or label) of individual i, $u_i^I(.)$ is i's intrinsic utility function, w_{ij} is the weight of the edge between i and j, and $u'(x_i, x_j)$ is extrinsic utility. Letting $\beta_0 \in \mathbb{R}$ and $\beta_1 \in \mathbb{R}^+$ be constants, we consider the extrinsic utility $u'(x_i, x_j) = \beta_0 - \beta_1 d(x_i, x_j)$, which is the linear function of the distance between x_i and x_j, denoted by $d(x_i, x_j)$. This implies that the closer x_i and x_j are according the distance function d, the corresponding extrinsic utility is higher and increase linearly with the distance. We assume that $d(.,.)$ is a valid distance function satisfying *non-negativity* $d(x, y) \geq 0$, *identity of indiscernibles* $d(x, y) = 0 \Leftrightarrow x = y$, and *symmetry* $d(x, y) = d(y, x)$. It is straightforward to see that

$$u_i(x_i|\mathbf{x}_{-i}) = w_{ii}U_i^I(x_i) + \beta_0 \sum_j w_{ij} - \beta_1 \sum_j w_{ij}d(x_i, xj).$$

We now present a counter example to show that the empathetic model is not subsumed by the metric labelling model. Consider the simple social network depicted in Fig. 3.2. We assume there are two options a and b where agents 1 and 2 have following intrinsic utilities over them: $u_1^I(a) = 1$, $u_2^I(a) = 0$, $u_1^I(b) = 0$, $u_2^I(b) = 1$. We now focus on agent 1's overall utility $u_1(.|x_2 = b)$ assuming agent 2 is assigned to item b (i.e., $x_2 = b$).

Let $u_1^L(.|x_2 = b)$ be the overall utility of individual 1 given $x_2 = b$ under the labelling framework. Then, we have $u_1^L(x_1|x_2 = b) = 0.25 \times u^I(x_1) + 0.75\beta_0 - 0.75 \times d(x_1, b)$. Given this, $u_1^L(a|x_2 = b) = 0.25 + 0.75\beta_0 - 0.75\beta_1 \times d(a, b)$. Similarly, $u_1^L(b|x_2 = b) = 0.75\beta_0 - 0.75\beta_1 \times d(b, b)$. As $d(b, b) = 0$ for a valid distance function due to identity of indiscernibles, we have $u_1^L(b|x_2 =$

[5]The metric labelling problem is typically formulated based on a node's costs and pairwise disagreement penalties/costs. However, we can represent this model equivalently using the concept of utility.

$b) = 0.75\beta_0$.

Let $u_1^E(.|x_2 = b)$ be the overall utility of individual 1 given $x_2 = b$ under empathetic framework. Then, we have $u_1^E(a|x_2 = b) = 0.25 \times u_1^I(a) + 0.75 \times u_2^E(b|x_2 = b)$. As $u_2^E(b|x_2 = b) = 1.0 \times u_2^I(b)$ and given intrinsic utilities above, we have $u_1^E(a|x_2 = b) = 0.25 \times 1 + 0.75 \times 1 = 1$. Similarly, $u_1^E(b|x_2 = b) = 0.75$.

Now we check whether there are valid values for β_0, β_1 and $d(.,.)$ which map the empathetic model to the labelling model. By letting $u_1^L(b|x_2 = b) = u_1^E(b|x_2 = b)$, we have $\beta_0 = 1$. Now we let $u_1^L(a|x_2 = b) = u_1^E(a|x_2 = b)$ and use $\beta_0 = 1$, we find that $1 - 0.75\beta_1 d(a, b) = 1$, thus $d(a, b) = 0$ or $\beta_1 = 0$. However, $\beta_1 > 0$ (as stated above) and also $d(a, b)$ can not be zero for a valid distance function (due to identity of indiscernibles). So, the empathetic model can not be subsumed by the labeling model.

3.4 Empirical Results

We describe experiments on randomly generated networks and intrinsic preferences to analyze our algorithms, and to contrast the decisions that result under standard non-empathetic, local empathetic, and global empathetic models.

Experimental Setup. We assume that individual intrinsic utilities arise from an underlying preference ordering over $\mathcal{A}$. In all experiments, we draw a random ordering for each agent j using either: the *impartial culture* model, in which all rankings are equally likely; or the Irish voting data set, which we explain in detail below. To draw connections to voting methods, j's utility for a is given by the Borda or plurality score of a in its ranking. As utilities, these embody very different assumptions: Borda treats utility differences as smooth and linear, whereas plurality utility is "all or nothing." We generate random social networks using a *preferential attachment* model for scale-free networks [25]:[6] starting with n_0 initial nodes, we add n nodes in turn, with a new node connected to $k \leq n_0$ existing nodes, where node i is selected as a neighbor with probability $\deg(i)/\sum_j \deg(j)$. We set $n_0 = 2$ and $k = 1$ in all experiments. We "direct" the graph by replacing each undirected edge with the two corresponding directed edges; add a self-loop to each node with weight α; then distribute weight $1 - \alpha$ equally to all other outgoing edges. Parameter $\alpha \in (0, 1]$ represents the degree of self-interest, and $1 - \alpha$ the degree of empathy. Unless noted, all experiments have $n = 1000$ agents (nodes), $\alpha = 0.25$, and are run over 50 random preference profiles on each of 50 random networks (2500 instances).

Performance Metrics. To examine the importance of modeling empathy in social choice, we distinguish actual user preferences—referred to as the *true model*—from how preferences are modeled in a group decision-support system—namely, the *assumed model*. Specifically, we let the true and assumed models be any of our intrinsic (non-empathetic), local or global models (9 possible combinations). We are interested in the extent to which these models *disagree* in their

[6] This is only one of many models that can be used. Results are similar for other types of networks.

decisions, and the loss in social welfare that results from such disgareement. If these measures are large, it indicates that, in situations where empathetic preferences exist, ignoring them by using classical preference aggregation techniques will lead to poor decisions. Specifically, we measure the percentage of *decision disgareement* (DD) (over 2500 instances for a fixed setting) in which the the true and assumed models propose different optimal decisions. We also measure the average loss in social welfare arising from making decisions using an assumed model that differs from the true model. Let $sw^t(\cdot)$ and $sw^a(\cdot)$ be social welfare under the true and assumed models, respectively, and a_t and a_a be the corresponding optimal options (or winners). Rather than directly comparing social welfare under various models, we define *relative social welfare loss (RSWL)* to be

$$RSWL = \frac{sw^t(a_t) - sw^t(a_a)}{sw^t(a_t)}.$$

We often report RSWL as a percentage. RSWL, by scaling differences in social welfare, helps calibrate the comparison between experiments. We can normalize RSWL by considering the range of possible social welfare values actually attainable. Let alternative a^- have *minimum social welfare* under the true model (so it is no better than the decision under the assumed model). *Normalized social welfare loss (NSWL)* is

$$NSWL = \frac{sw^t(a_t) - sw^t(a_a)}{sw^t(a_t) - sw^t(a^-)}.$$

This offers a more realistic picture of loss caused by using an inconsistent assumed utility model (by comparing it to the loss of the *worst possible decision* under the true model).

Impartial Culture. We first consider RSWL and NSWL for all nine combinations of assumed and true utility models. We fix $m = 5$ options and use Borda scoring. Average (maximum) losses are reported in Table 3.1 while the decision disagreement percentage is shown in Table 3.2. While RSWL is relatively small on average (though maximum losses are quite large), this is largely due to the uniformity of preferences generated by impartial culture (all options have the same *expected* score). By normalizing, we obtain a more accurate picture of the loss incurred by using non-empathetic voting: average normalized loss shows that the "controllable" error is quite large, especially when comparing the "standard" intrinsic model to either of the empathetic models. Moreover, the intrinsic model chooses the incorrect alternative in over half of all instances in both cases. Interestingly, assuming either the local model or global model when the true model is the other gives reasonable results: this means that the local model offers a good first-order approximation to the global model (see Sec. 3.2).

Irish Voting Data. Impartial culture is often viewed as an unrealistic model of real-world preferences. For this reason, we tested our methods using preferences drawn from the 2002 Irish General Election, using electoral data from the Dublin West constituency, which has 9

true model	assumed model		
	intrinsic	local	global
intrinsic	—	1.4(9.9)	1.1(8.0)
	—	28.4(100)	22.6(100)
local	2.9(19.3)	—	0.1(3.2)
	28.5(100)	—	1.2(86.9)
global	1.8(12.7)	0.1(2.7)	—
	22.3(100)	1.1(97.0)	—

Table 3.1: Avg. (max.) RSWL (1st rows) and NSWL (2nd rows): Borda, $m = 5$.

true model	assumed model		
	intrinsic	local	global
intrinsic	—	57.76	50.48
local	58.12	—	11.72
global	50.84	11.72	—

Table 3.2: Percentage decision disagreement: Borda, $m = 5$.

candidates and 29,989 ballots of top-t form, of which 3800 are complete rankings.[7] We assign full rankings, drawn randomly from the set of 3800 complete rankings to nodes in our network. Decision disagreement under both plurality and Borda scoring (Table 3.3) is quite high, ranging from 22-46%. Average NSWL (shown in Table 3.4) is not as high as with impartial culture (from 1-3%, with maximum loss around 40%).

The effect of m. Fig. 3.3 shows the average RSWL and decision disagreement (DD) for three "true vs. assumed" models as we increase the number of alternatives m using plurality scoring. We observe that average RSWL increases with m and approaches 70% when $m = 200$, while the optimal decision is rarely made. NSWL for the instrinsic model (shown in Fig. 3.4), even at $m = 5$, averages 20–30%. With Borda scoring, the effect of m is much less pronounced because of relatively small utility differences (or smoothing) between adjacent candidates (intrinsic loss ranges from 20-30% across all values of m), but the pattern decision disagreement is almost identical to plurality.

The effect of scoring rule (from Plurality to Borda). So far, we have experimented with both Borda (as a representative for smooth scoring rule) and Plurality (as a representative for sharp, one-or-nothing scoring rule). We now explore how RSWL changes when the scoring rule transitions from plurality to Borda. We consider the $\tau-$scoring rule

$$s_\tau(a_j, r) = \tau^{r(a_j)-1}(m - r(a_j)),$$

where $r(a_j)$ represents the rank of alternative a_j in ranking r and $\tau \in [0, 1]$. Note that when $\tau = 1$, the τ-scoring rule is equivalent to Borda whereas with $\tau = 0$, the τ-scoring rule is

[7] We have obtained the original data sets from www.dublincountyreturningofficer.com.

true model	assumed model		
	intrinsic	**local**	**global**
intrinsic	—	27.3 / 46.1	22.3 / 39.0
local	28.0 / 46.3	—	5.6 / 8.6
global	22.9 / 39.3	5.5 / 8.6	—

Table 3.3: Percentage decision disagreement, plurality/Borda: West Dublin, $m = 9$.

true model	assumed model		
	intrinsic	**local**	**global**
intrinsic	—	1.9(34.8)	1.3(19.9)
local	2.7(39.6)	—	0.1(7.1)
global	1.6(31.7)	0.1(8.8)	—

Table 3.4: Average (maximum) NSWL: 2500 runs, Plurality, West Dublin dataset, $m = 9$, $n = 1000$, $\alpha = 0.25$.

plurality.

We set $m = 10$ and vary τ over $\{0, 0.2, 0.4, 0.6, 0.8, 1\}$. Fig. 3.5 shows average (maximum, minimum) RSWL for three actual, assumed model combinations for various τ values. We observe that plurality is more susceptible to RSWL than Borda (compare $\tau = 0$ with $\tau = 1$). The change in RSWL is almost linear when moving from plurality to Borda. This implies that amongst the wide variety of scoring rules which exist between Borda and plurality, those scoring rules which are closer to plurality yield to higher RSWL when compared to those are closer to Borda. These results suggest that the sharpness or smoothness of scoring rule plays a role in RSWL; i.e., the sharper the scoring rule is, the higher RSWL is.

Self-loop weight α. Varying the self-loop weight α has a significant effect on NSWL and decision disagreement when true utility is global but intrinsic utility is assumed. Table 3.5 shows that, for both Borda and plurality, increasing α (i.e., decreasing overall degree of empathy) decreases both NWSL and DD, which is not surprising, as by increasing α the empathetic model gets closer to the intrinsic model. Similar trends hold for the local model. We also used a model in which nodes have different self-loop weights, drawing each node's α from a (truncated) Gaussian. As we vary the mean μ, we see a similar trend in Table 3.6.

The impact of directionality. The results above use networks with bi-directional edges (by replacing each undirected edge with two directed edges). To explore how directionality impacts NSWL, we consider networks with a *hierarchical orientation*, as often found in economic (e.g., supply chain), organizational (management/employee structure), and even some social networks (e.g., forms of status, following, etc.). We replace each undirected edge in the preferential attachment network with a directed edge from the "younger" node to the "older." The older node reciprocates with a directed edge to the younger with probability γ. If $\gamma = 1$, our standard bidirectional network results (as above); when $\gamma = 0$ we obtain a completely

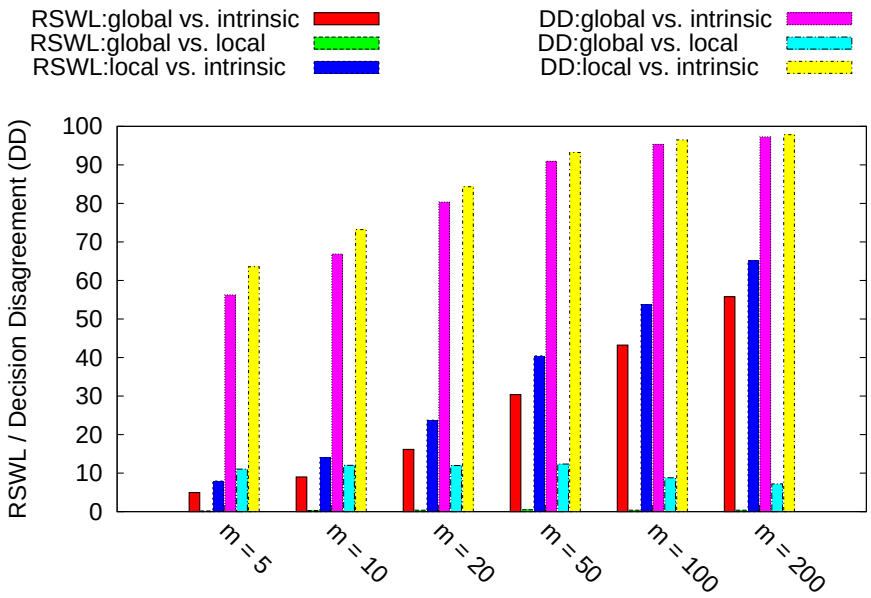


Figure 3.3: RSWL and decision disgreement (DD), plurality.

α	0.05	0.1	0.25	0.5	0.75
Borda	26.0 / 58.7	25.0 / 55.8	22.2 / 53.0	15.1 / 42.5	7.3 / 28.8
Plurality	28.8 / 59.8	26.7 / 58.1	22.7 / 53.8	16.9 / 46.9	7.8 / 31.3

Table 3.5: Average NSWL/decision disgreement: global vs. intrinsic, varying α.

hierarchical network.

Fixing $m = 10$, Fig. 3.6 depicts NSWL for both Borda and plurality as γ varies. Networks that are more hierarchical have higher NSWL for the global vs. intrinsic models, independent of the scoring rule, while NSWL for local vs. intrinsic is almost constant. However, plurality seems more susceptible to increasing loss due to hierarchical structure than Borda for all three combinations. Unlike earlier results, when the network is very hierarchical (e.g., $\gamma = 0$), the global and local models do not approximate each other well.

Number of Iterations of ICE. We examine how the self-loop weight α affects the expected number of iterations required by the ICE algorithm. We fix $m = 5$, and vary α. Fig. 3.7 illustrates estimated social welfare for each alternative in one representative run ($\alpha = 0.25$, Borda scoring): this instance of ICE converges in 24 iterations, with computation time under 2 ms., despite the large number of voters. Alternative a_4 is eliminated at iteration 16, a_5 at 17, a_1 at 20, and a_2 at 24, leaving a_3 as optimal. Note that the relative order of the alternatives is unchanged after 6 iterations, suggesting early termination as a robust means of approximation.

Table 3.7 shows the average number of iterations for various α, for Borda and plurality. In all cases, the number of iterations is small relative to network size. ICE is quite insensitive to

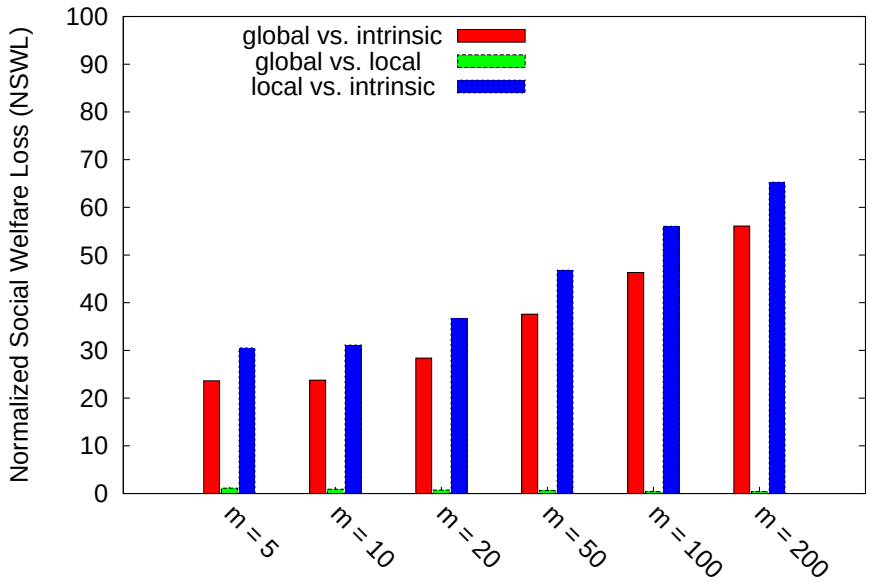


Figure 3.4: Average NSWL, impartial culture, and varying the number of alternatives m.

μ	0.05	0.1	0.25	0.5	0.75
Borda	27.6 / 59.2	24.3 / 54.9	21.3 / 52.3	15.6 / 42.5	8.1 / 31.1
Plurality	27.2 / 58.6	24.5 / 55.5	23.3 / 54.7	16.5 / 46.1	8.0 / 31.5

Table 3.6: Average NSWL/decision disgreement: global vs. intrinsic, α drawn from truncated Gaussian with mean μ and std. dev. 0.1.

the scoring rule, and termination time declines dramatically with increasing α, as is typical for iterative algorithms (e.g., for Markov chains).

Number of Iterations of WICE. We examine how the self-loop weight α affects the expected number of iterations required by the WICE algorithm. We fix $m = 5$, and vary α. Table 3.8 shows the average number of iterations for various α, for both Borda and plurality. The number of iterations is small relative to network size, but is almost twice that of ICE (see Table 3.7). WICE, similar to ICE, is quite insensitive to the scoring rule, and termination time declines dramatically with increasing α, as expected.

Performance Comparison of ICE vs. WICE. The experiment above shows that WICE requires a greater number of iterations on average when compared to ICE. We here examine how the running time of ICE compares to that of WICE. We fix $\alpha = 0.25$ and $n = 1000$ but vary m. Table 3.9 shows the ratio of average running time for ICE over the average running time of WICE (over 2500 common instances). WICE seems to be faster than ICE despite its greater number of iterations. The speed ratio generally increases with m. Given this performance, we suggest the use of WICE over ICE for applications with relatively large m.

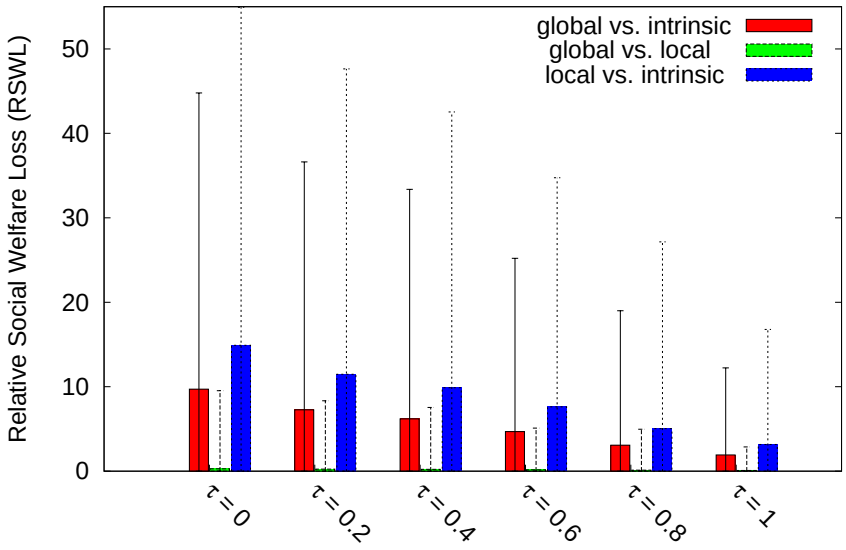


Figure 3.5: Avg. (maximum, minimum) RSWL (2500 runs): $n = 1000$, $m = 10$, τ-scoring rule.

α	0.05	0.1	0.25	0.5	0.75
Borda	104.1	51.4	19.5	8.7	4.7
Plurality	98.7	48.6	18.6	8.3	4.6

Table 3.7: Average number of iterations for ICE, varying α.

Empathetic Resource Allocation. We briefly demonstrate the value of accounting for empathetic utilities in a simple resource allocation problem. We use the plurality scoring rule, and fix $n = 1000$ agents and $m = 5$ alternatives. Each agent's ranking is drawn from a ϕ-Mallows model. For each alternative j, we let its quota be $q_j = \frac{n}{cm}$, where the constant c is set to 2, unless noted. Under plurality, the weighted constrained resource allocation (outlined in Eq 3.14) can be solved by greedily allocating each alternative a_j (given its quota q_j) to those agents who ranked a_j first and have high societal weight ω. In other words, for each alternative a_j, we first find those set of agents $\mathcal{N}_j = \{i \in \mathcal{N} | r_i(a_j) = 1\}$ who ranked a_j first. Then, we sort the agents in $\mathcal{N}_j$ based on their societal weights ω in the a descending order. We start allocating a_j from top of the list as far as the quota q_j allows us. After iterating over all alternatives, we can arbitrary assign remaining alternatives to unmatched agents (due to plurality, the agents are indifferent to items below their first-ranked items). Note that there might be many possible optimal solutions given the distribution of rankings, quotas, and societal weights.

We define relative social welfare loss (RSWL) in a similar fashion to voting above:

$$RSWL = \frac{sw^t(\mathbf{x}_t) - sw^t(\mathbf{x}_a)}{sw^t(\mathbf{x}_t)},$$

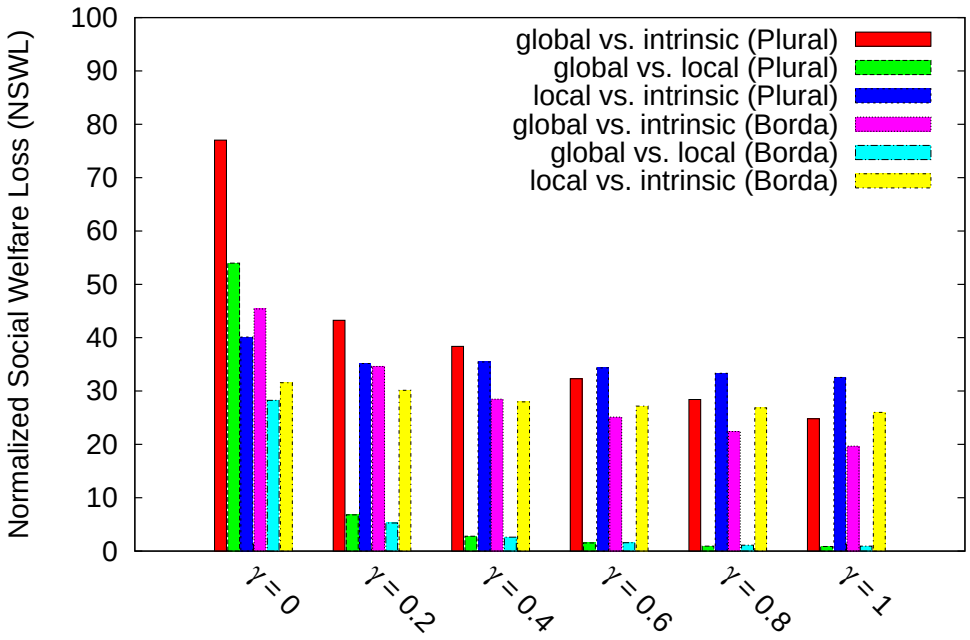


Figure 3.6: Average NSWL, $m = 10$, varying γ.

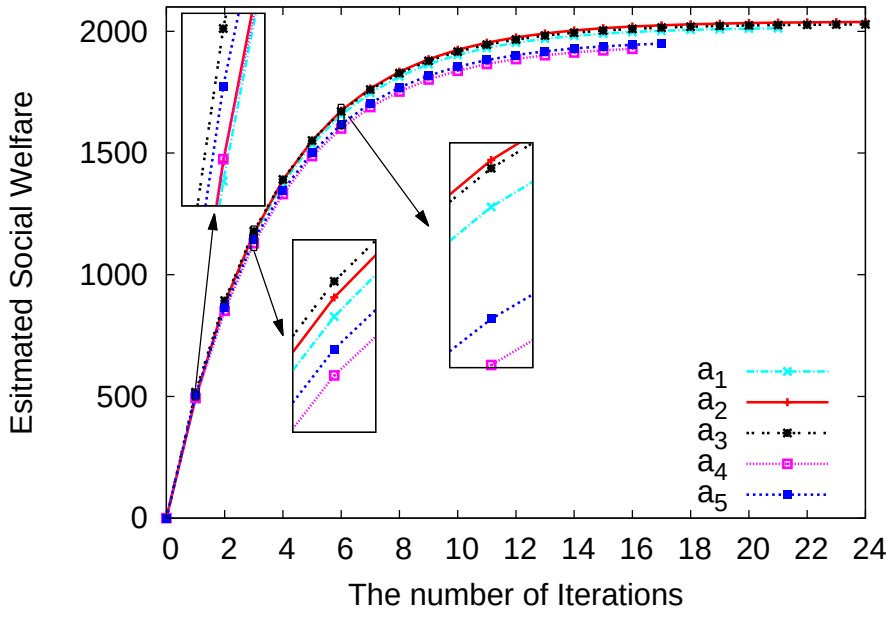


Figure 3.7: Estimated social welfare vs. iterations of ICE (one sample run).

α	0.05	0.1	0.25	0.5	0.75
Borda	219.4	101.1	33.6	12.9	6.1
Plurality	208.1	94.6	31.5	12.1	5.7

Table 3.8: Average number of iterations for WICE, varying α.

m	5	10	20	50	100	200	500	1000
Borda	1.592	1.737	1.882	1.817	1.829	2.316	2.715	2.574
Plurality	1.691	1.923	1.930	1.957	2.227	3.105	4.209	4.330

Table 3.9: The ratio of the average of running time for ICE over the average of running time for WICE (2500 common instances).

where $sw^t(\cdot)$ and $sw^a(\cdot)$ are social welfare under the true and assumed models, respectively, and $\mathbf{x}_t$ and $\mathbf{x}_a$ are the corresponding optimal options.

Fig. 3.8 demonstrates the average RSWL (over 2500 instances) for three "true vs. assumed" models as we increase ϕ. RSWL seems to be high for both "global vs. non-empathetic" and "local vs. non-empathetic" but it is very low for "local vs. global". RSWL decreases with ϕ, meaning that the more homogeneous the society is, the higher relative social welfare loss is. This is partially due to the constraints or quotas imposed: when there are more agents with the same first-rank alternative and limited capacity, the allocation mechanisms are required to consider societal weights (due to empathy) more seriously in their allocations.[8]

We also examine the effect of resource scarcity on RSWL. We set $\phi = 0.8$ and vary c. The higher c is, the higher the scarcity of resources. Fig. 3.9 shows the average RSWL (over 2500 instances) for various c. RSWL increases with c, suggesting that with higher scarcity, the allocation mechanism should be more cautious in assigning the resources to individuals. The general goal should be to first satisfy individuals with the higher empathetic influence (i.e., higher societal weights).

Iterative Weight Computation for Resource Allocation Finally, we examine the performance/accuracy of the weight updating scheme presented in Eq. 3.17. We assume that the true model is the global empathetic model. After each iteration, we consider the estimated weights as an assumed model and we compute RSWL for that specific iteration. This process allows us to monitor how RSWL evolves over iterations of our updating scheme. For this experiment, we set $c = 2$ and $\phi = 0.8$ while varying m. Table 3.10 demonstrates the average of RSWL (over 2500 instances) for various m over 10 iterations. The higher m is, the average RSWL is higher at iteration 0. However, only after 2 iterations, the average RSWL is close to zero for all m (with the maximum of almost 0.02). The average RSWL after 10 iterations (for all m) is

[8]This pattern is not observed for consensus decision making in our experiments. In contrast, our experiments suggest that homogeneity affects RSWL in the opposite way for consensus decision making: the higher homogeneity is, the lower RSWL is.

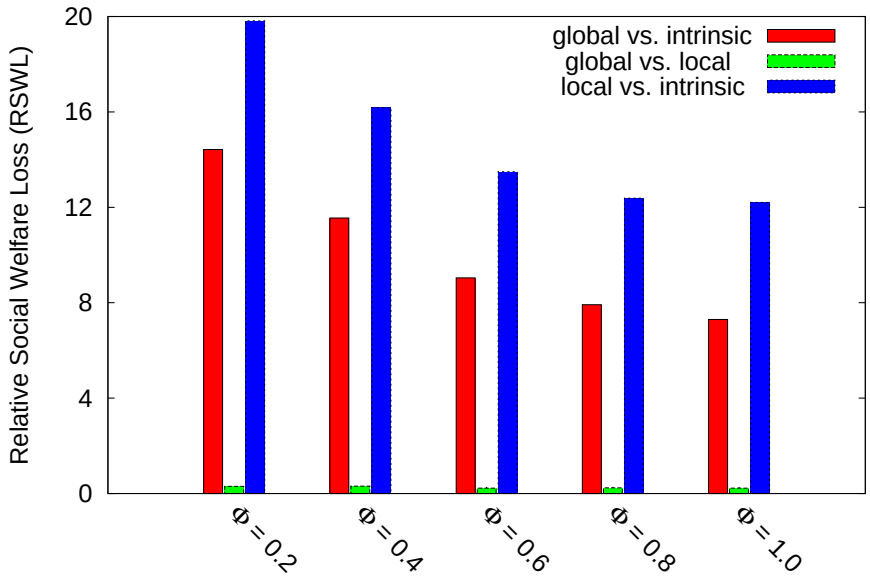


Figure 3.8: RSWL for allocation problem, plurality, n=1000, m=5, $q_j = 100$, varying ϕ

m/iter.	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
5	7.8503	0.2459	0.0135	0.0131	0.0035	0.0045	0.0010	0.0012	0.0003	0.0003	0.0001
10	8.8486	0.2033	0.0179	0.0129	0.0040	0.0034	0.0011	0.0009	0.0003	0.0002	0.0001
20	11.5266	0.2458	0.0193	0.0203	0.0042	0.0038	0.0011	0.0009	0.0003	0.0002	0.0001
40	15.1015	0.3120	0.0187	0.0344	0.0050	0.0068	0.0014	0.0016	0.0004	0.0004	0.0001
80	17.5271	0.3521	0.0171	0.0474	0.0045	0.0092	0.0013	0.0022	0.0004	0.0006	0.0001

Table 3.10: Average RSWL (estimated weights vs. global) for allocation problem over iterations, plurality, n=1000, $\phi = 0.8$, $q_j = \frac{n}{2m}$ for all $j \in \mathcal{A}$.

0.0001.

3.5 Summary and Future Work

This chapter presents a novel model for social choice, combining intrinsic and *empathetic* preferences, the latter reflecting one's desire to see others satisfied with a chosen alternative. Using a social network to measure degree of empathy, our proposed algorithms, for local and global empathetic settings, allow efficient computation of optimal decisions by weighting the contribution of each agent, and have a natural interpretation as empathetic voting when scoring rules are used. Critically, individuals need only specify their intrinsic preferences (and network weights): they need not reason explicitly about the preferences of others.

Our empathetic model is a starting point for the broader investigation of empathetic preferences in social choice and group decision making. One can explore more realistic processes for simultaneous generation of networks and preferences that better explain preference correlation (see, e.g., our work on *ranking network* framework [311] which is presented in the next chapter).

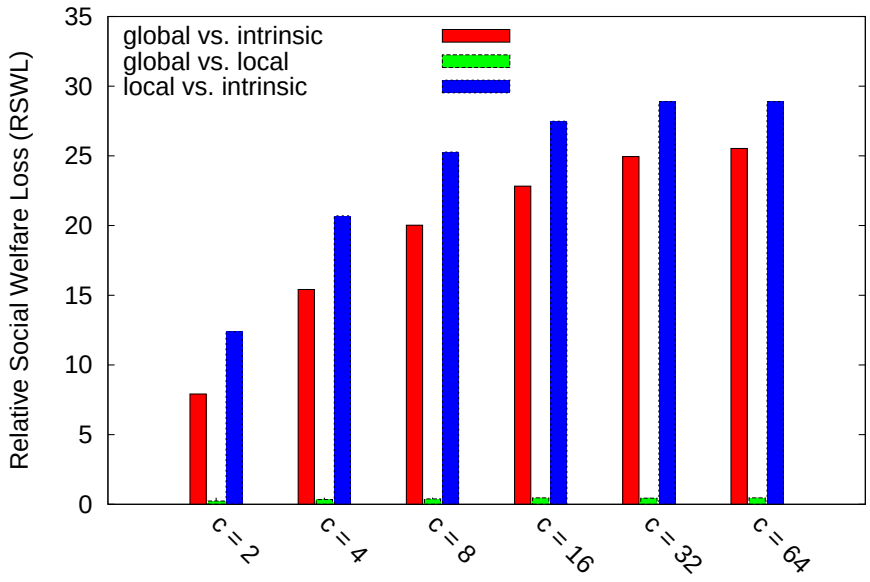


Figure 3.9: RSWL for allocation problem, plurality, $n = 1000$, $m = 5$, $\phi = 0.8$, $q_j = \frac{200}{c}$ for all $j \in \mathcal{A}$, varying c.

Methods to assess the prevalence of empathetic preferences, the extent to which social network structure reflects such preferences, and how they can be discovered effectively, are critical. Testing our model, and these extensions, on large data sets is of course critical to validating the existence of empathy of this form.

Although we show in Sec. 3.3 how to extend our empathetic social choice framework to accommodate other social choice problems (such as matching, assignment, and multi-winner election problems), each of these applications requires its own algorithmic developments. Other important directions include: voting schemes where agents specify tradeoffs between intrinsic and empathetic preference in a *qualitative* fashion; and analysis of manipulation in the context of such externalities in voting. We elaborate on future work in Sec. 6.3.

Chapter 4

Ranking Networks

Network formalisms offer a powerful tool for studying the interactions of entities in complex systems. Of special interest in network science is the development of *network formation models* that explain the emergence of common structural properties of real-world networks. Recent work has focused on modelling the role of both known and hidden attributes of nodes in forming networks—the intuition being that each node possesses attributes (e.g., geographical location [351] or social position [183, 49]) that determine the chance that any pair of nodes become connected. Examples of such models include spatial networks [27], random geometric graphs [295, 103], latent space models [183, 49], and multiplicative attribute graphs [203, 206]. We discussed latent and spatial networks in Sec. 2.3.3.

Most of these models take attributes to be binary, real-valued, or integral. Surprisingly little work has studied attribute-based network formation in which the underlying attributes correspond to *rankings* over some space of options. Such rankings may represent the preferences of agents over some set of alternatives (e.g., products or services, political candidates, jobs) or their subjective opinions about the quality or relevance of certain items (web pages, sports teams). Because of the increasing availability of ranked data brought about by e-commerce, recommender systems, web search, online/interactive polling and surveying, and numerous other applications, a deeper understanding of the possible interactions between ranked data and network formation seems imperative.

In Chapter 3, we analysed empathetic models for capturing the interdependency of preferences over social networks. We observed how empathy can induce correlation of preferences. However, correlation of preferences can also be due to homophily as discussed in Sec. 2.3.1. This chapter focuses on developing a generative model for ranking preferences (or, in general, rankings) based on homophily.

In this chapter, we introduce and formulate a model for the formation of *ranking networks* [311], in which each entity possesses a ranking over a set of m alternatives. The similarity of their rankings determines the probability with which two individuals will connect in the network. Such networks can be used for a variety of purposes, but we are especially interested in models where rankings reflect *agent preferences* over some space of options. We discuss the

extent to which we can leverage the inherent structure of ranking networks to address important computational and algorithmic questions that arise in the analysis of networks. For instance, what topological network properties emerge under such a model? How do model parameters—including the underlying ranking distribution and similarity metric—influence these properties? And can such structural properties be predicted efficiently?

Ranking network models can be deployed in a variety of contexts. It is widely acknowledged that the behaviours of individuals are correlated over the social networks [88, 118, 187]; and correlated preferences offer one possible explanation of this phenomenon. With suitable models such as ranking networks, one can more readily predict or elicit the preferences of specific individuals to develop more efficient recommender, social choice and advertising mechanisms (e.g., to make better decisions with less data or less elicitation). We will focus on this application of ranking networks in Chapter 5. Ranking networks may also have application in information retrieval and topic modelling over the web (or other information networks): given some common set of attributes (e.g., topics, tags), each node (e.g., web page, photo) may have a "relevance ranking" over those attributes. A ranking network can capture the correlation of such relevance rankings given the structure of underlying information network (and may help improve the efficiency of recently developed comparison-based interactive search methods [194, 195]). The predictive power of ranking networks might be also harnessed for link prediction [244, 237]—to help predict future interactions between entities—or *link mining*—to help infer unobserved links between entities [150].

We analyse some general topological properties of ranking networks (e.g., connectivity conditions, graph diameter, degree distribution, edge density, clustering coefficient). We show that ranking networks exhibit the *small world* effect—a commonly observed real-world social networks' property—by possessing short diameter and average path length. We demonstrate how the diameter can be approximated using random graph theory. Ranking networks also possess the diameter-shrinking property of real-world networks [232], where diameter (and average shortest path) shrinks as the network grows.

We observe that the structural properties of ranking networks (e.g., edge density and degree distribution) can be computed readily if model parameters are given. However, computation can be extremely intensive even for relatively small m due to the combinatorial size of ranking space. This motivates us to develop easy-to-compute approximations for the special class of *distance-based ranking models*, while studying properties that emerge among networks in this class.

Distance-based ranking networks are built upon distance-based distributions of rankings in which ranking probabilities decrease with increasing distance from some "modal" or "reference" ranking (see Sec. 2.2.2 for review). We bound the minimum and maximum probability that a node with a given ranking is connected to another randomly chosen node. Our theoretical bound has a natural interpretation: individuals that possess more likely or more "popular" rankings have greater odds of connecting with others (hence have higher expected degree) than

those with less likely or less "popular" rankings. This is not surprising given the nature of distance-based ranking models. In the context of social networks, this is one possible contributing factor to the intuition that people with more "popular" preferences tend to have more friends and social interactions than those with uncommon preferences, i.e., the intuition that *preference popularity governs social popularity*. From a different perspective, one can use observed node degree to make inferences about that node's ranking: higher degree is predictive of more "common" preferences (i.e., closer to the modal ranking). This can be exploited to support efficient estimation/learning of both the reference ranking, the ranking distribution, and the preferences of specific individuals. Using these theoretical bounds, we develop some approximation methods for predicting edge density and degree distribution under ranking models. We also show how these approximation methods can be used for learning model parameters when one is interested in maximum likelihood estimation of the parameters and dealing with missing data.

We perform some empirical experiments using the *Mallows' ϕ-model* [258]. We compare various structural properties (e.g., degree distribution, edge density, diameter) of the randomly-generated networks under a ranking network model with the predictions of our approximation methods. Our results confirm the efficiency and predictive accuracy of our computational approximations.

The remainder of this chapter is organized as follows. After reviewing the related work in Sec. 4.1, we first provide background on latent space models for network formation 4.2. We then define *ranking networks* by building upon these models and discuss some general theoretical properties in Sec. 4.3. In Sec. 4.4, we examine the special case of *distance-based ranking networks*, analysing some of its theoretical properties and developing computational approximations to estimate these properties. In Sec. 4.5, we show how our approximation methods can be used for efficient model learning. The results of some empirical experiments conducted on ranking network model are reported in Sec. 4.6. Finally, we highlight some future work in Sec. 4.7.

4.1 Related Work and Concepts

Network formation models have drawn considerable attention, with a variety of models proposed to capture structural properties of real-world networks [279, 26, 280].One can roughly divide these model into two categories. *Static models* (e.g., Erdős-Rènyi, small world) specify the network using simple static properties (e.g., probability of edge occurrence), whereas *dynamic models* (e.g., preferential attachment) involve some dynamic "growth" process by which nodes and edges are added to a network (see the review of network formation in Sec. 2.3.3).

Our ranking network model is a random, static model. It is also falls within the class of *spatial (or latent space) networks* [27, 183] in which nodes have a set of real-valued, binary, or integer-valued *latent variables*, with the probability of an edge forming between two nodes

determined by their attributes (see Sec. 2.3.3 for a review of latent and spatial models). Within spatial network formation models, recent research has addressed *inference and learning* of latent attributes and social network structure (for example, see[183, 204, 205] and the review in Sec. 2.3.4).

4.2 Latent Space Network Models Properties

Latent space or hidden variable network models [48] generally assume a set n nodes, where each node i is associated with a latent variable h_i. This represents some feature (or feature vector) of the node or individual in question. The parameters of such a model are given by a parameter vector $\boldsymbol{\theta} = (\boldsymbol{\lambda}, \boldsymbol{\eta})$ (we define model parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\theta}$ below). A random undirected graph is generated by assuming: (1) a parameterized distribution $\rho(h|\boldsymbol{\eta})$ specifying the probability that an arbitrary node takes on value h (vector $\boldsymbol{\eta}$ is parameter vector of distribution ρ); and (2) a symmetric *connection probability function*, $c(h, h^{'}|\boldsymbol{\lambda})$, where $c(h_i, h_j\boldsymbol{\lambda})$ denotes the probability that edge e_{ij} forms between nodes i and j given their variable values. The vector $\boldsymbol{\lambda}$ denotes those model parameters governing the connection probability function $c(\cdot|\boldsymbol{\lambda})$. A common variant of this model defines the connection probability using the relative distance d_{ij} between the values h_i and h_j in some metric space. In this case, we assume a function $c(d) : [0, \infty) \rightarrow [0, 1]$ that maps distances into connection probabilities. We generally assume distance-based connection functions below. A canonical example of such a connection probability is Waxman's [351] function $c(d) = \beta e^{-\frac{d}{d_0}}$, widely used for modeling computer network topologies, where β controls edge density and d_0 represents a "typical" distance between nodes. Another is $c(d) = (1+\frac{d}{\beta})^{-\alpha}$, used for modeling social networks [49]. Here β controls the average node degree while $\alpha > 1$ determines the degree of homophily (higher value for α, higher amount of homophily).

We now describe certain structural properties that emerge (e.g., edge density, clustering coefficient, etc.) under latent space models. Because we will deal with discrete ranking spaces below, we assume latent variables are discrete (e.g., marginalizations are given by summations rather than integrals). First, note that *edge density*, i.e., the probability of an edge occurring between two arbitrary nodes, is given by:

$$\mathcal{E}(\boldsymbol{\theta}) = \sum_{h,h'} \rho(h|\boldsymbol{\eta})\rho(h^{'}|\boldsymbol{\eta})c\left(d(h^{'}, h)|\boldsymbol{\lambda}\right). \tag{4.1}$$

As a result, the expected number of edges of a randomly chosen node is given by $\binom{n}{2}\mathcal{E}(\boldsymbol{\theta})$. The probability of an edge occurring between a node i with observed value h_i and a random node j is:

$$\mathcal{D}(h_i, \boldsymbol{\theta}) = \sum_{h'} \rho(h^{'}|\boldsymbol{\eta})c\left(d(h^{'}, h_i)|\boldsymbol{\lambda}\right), \tag{4.2}$$

where $\boldsymbol{\theta}$ is the model parameter vector, $\rho(.)$ is latent variable distribution, and $c(.)$ is the connection probability function. Thus, the expected degree of a node with value h_i is given by

$(n-1)\mathcal{D}(h_i, \boldsymbol{\theta})$. Note that $\mathcal{E}(\boldsymbol{\theta})$ can be rewritten as:

$$\mathcal{E}(\boldsymbol{\theta}) = \sum_h \rho(h|\boldsymbol{\eta})\mathcal{D}(h, \boldsymbol{\theta}). \tag{4.3}$$

The exact *degree distribution* $P(k, \boldsymbol{\theta})$ is:

$$P(k, \boldsymbol{\theta}) = \sum_h \rho(h|\boldsymbol{\eta})\mathcal{G}(h, k|\boldsymbol{\theta}), \tag{4.4}$$

where $\mathcal{G}(h, k|\boldsymbol{\theta})$ is the probability that a node with value h has k neighbours and is given by the binomial distribution of

$$\mathcal{G}(h, k|\boldsymbol{\theta}) = \binom{n-1}{k} \mathcal{D}(h, \boldsymbol{\theta})^k \left(1 - \mathcal{D}(h, \boldsymbol{\theta})\right)^{n-1-k}. \tag{4.5}$$

The *clustering coefficient* of a node i is the fraction of the *pairs* of neighboring nodes that are themselves connected. The expected clustering coefficient for a node i with value h_i is:

$$\mathcal{C}(h_i, \boldsymbol{\theta}) = \frac{1}{\mathcal{D}(h_i, \boldsymbol{\theta})^2} \sum_{h', h''} \rho(h')\rho(h'')c(d(h_i, h'))c(d(h_i, h''))c(d(h', h'')) \tag{4.6}$$

One can compute the average clustering coefficient by

$$\langle \mathcal{C}(\boldsymbol{\theta}) \rangle = \sum_h \rho(h, \boldsymbol{\eta})\mathcal{C}(h, \boldsymbol{\theta}).$$

4.3 Ranking Networks: A General Model

We adapt the general latent space network model to the setting where the latent attributes are *rankings* over some set of alternatives. For instance, nodes might represent individuals with attributes reflecting their preferences over some set of products, services, political candidates, etc. We assume a finite set of alternatives (or options) $\mathcal{A} = \{a_1, \ldots, a_m\}$ and a set of nodes $\mathcal{N} = \{1, \ldots, n\}$. Each node i has a ranking (or strict total order) over $\mathcal{A}$, denoted by $\succ_i$ (weaker notions, e.g., preorders, partial orders, can be accommodated, though some details of our model require modification). Let $\Omega(\mathcal{A})$ denote the set of all $m!$ rankings over $\mathcal{A}$. Our latent variables are rankings r drawn from $\Omega(\mathcal{A})$. We assume that each node i's ranking r_i is drawn independently from some (parameterized) distribution $\rho(r|\boldsymbol{\eta})$ over $\Omega(\mathcal{A})$. We also assume a *ranking distance metric* $d : \Omega(\mathcal{A})^2 \to \mathbb{R}$ which measures similarity between rankings. Finally, a *connection probability function* $c : [0, \infty) \to [0, 1]$ determines the probability that two nodes i, j are connected given the distance $d(r_i, r_j)$ between their rankings. We now detail each of these three components of our model.

Distance Metric on Rankings. We use the "similarity" of two rankings to determine their distance, which will be used below to determine connection probabilities. A variety of well-

known distance metrics for rankings can be used [111] (see Sec. 2.2.1 for more details). We again briefly describe several common distance metrics. A natural set of distances are "d_p distances", where $d_p(r, r') = \sum_{i=1}^{m} |r(a_i) - r'(a_i)|^p$ for $p \in [1, \infty)$. The well-known *footrule* ($p = 1$) and *Spearman* ($p = 2$) distances are instances of this. *Hamming* distance is another natural model, while *Kendall's* τ distance is used, where

$$d_\tau(r, r') = \sum_{k \neq l} I[r(a_k) > r(a_l) \text{ and } r'(a_k) < r'(a_l)]. \tag{4.7}$$

Here d_τ measures the number of pairwise swaps needed to transform r to r'. As the ranking space $\Omega(\mathcal{A})$ is discrete with finite size $m!$, there are finitely many realizable distances. It is easy to see that, for any $r, r' : d_\tau(r, r') \in \{0, 1, \cdots, \binom{m}{2}\}$.

Ranking Distributions. The $\rho(r|\boldsymbol{\eta})$ component of our ranking network model accommodates arbitrary ranking distributions. Distributional models of rankings developed in psychometrics and statistics, and now widely used in machine learning and information retrieval [242], include Mallows, Plackett-Luce, Bradley-Terry, and many others (see Sec. 2.2.2 for an overview). We use the Mallows' ϕ-model in our empirical experiments. It is characterized by a "modal" *reference ranking* σ and a *dispersion parameter* $\phi \in [0, 1)$, with the probability of a ranking r given by $\rho(r|\sigma, \phi) \propto \phi^{d_\tau(r,\sigma)}$.

Connection Probability Function. With ranking-based distance metrics in hand, we adopt standard connection functions for latent-space models (see above). We assume c is integrable and strictly decreasing.[1]

We now study several topological properties of general ranking networks. We can derive sufficient conditions for the connectivity of a ranking network:

Theorem 4.3.1. *Assume a ranking model $(\rho(r|\boldsymbol{\eta}), c(d|\boldsymbol{\lambda}))$ over m alternatives. The induced ranking network is connected with high probability (i.e., with probability $1 - o(1)$ where $o(1) \to 0$ as $n \to \infty$) if*

$$d_M(m) < c^{-1}\left(\frac{\log n}{n}\Big|\boldsymbol{\lambda}\right),$$

where $d_M(m)$ is the maximum possible distance under d given m alternatives.[2]

All proofs and theoretical details of this chapter can be found in Appendix B. This result can be used to derive suitable conditions for the connectivity of specific models. For instance, using the τ distance metric and Waxman connection function $c(x|\alpha, d_0) = \alpha e^{\frac{x}{d_0}}$, the emerging ranking network will be connected with high probability if $\binom{m}{2} < -d_0 \ln \frac{\log n}{\alpha n}$.

[1] The decreasing assumption results in homophilous network in which similar nodes are more likely to be acquainted than dissimilar nodes; when hetrophily is in effect in the network, some other set of assumptions (e.g., increasing) must be taken into consideration for the connection probability function.

[2] We assume that m is fixed and independent of the number n of individuals. This is consistent with the applications we have in mind (e.g., social choice, recommender systems, interactive search, etc.).

The *small world* effect is a commonly observed property of real-world networks: the diameter (or longest shortest path between any pair of nodes) is small, as is the average shortest path length [7]. Ranking networks exhibit these properties: the diameter $D(\boldsymbol{\theta})$ and the average shortest path $\langle l(\boldsymbol{\theta})\rangle$ can be approximated and bounded by:

$$D(\boldsymbol{\theta}) \leq \left\lceil \frac{\log(n)}{\log(n-1) + \log c(d_M(m)|\boldsymbol{\lambda})} \right\rceil; \quad \langle l(\boldsymbol{\theta})\rangle \leq \frac{\log(n)}{\log(n-1) + \log c(d_M(m)|\boldsymbol{\lambda})}. \tag{4.8}$$

The derivation of these approximations are detailed in Appendix B. Our empirical results in Sec. 4.6 confirm these properties and the tightness of these bounds. Ranking networks also possess the diameter-shrinking property of real-world networks [232], where diameter (and average shortest path) shrinks as the network grows:

Theorem 4.3.2. *Fix m and assume $\rho(r|\boldsymbol{\eta})$ distributes probability mass on more than one ranking. The asymptotic diameter of any ranking network over m options is 2 (as $n \to \infty$).*

These and other structural properties of ranking networks can be computed readily if model parameters are given. However, computation can be extremely intensive, even for relatively small m, due to the combinatorial size of ranking space (recall, the number of possible rankings is $m!$). Consequently, it would be computationally prohibitive for one to benefit from the predictive power of these mathematical expressions (stated above) to predict the topological properties of the emerging network. This motivates the development of easy-to-compute approximations for the special class of *distance-based ranking models*, while studying properties that emerge among networks in this class. We will also show how our approximation methods can be applied for efficient learning of model parameters.

4.4 Distance-Based Ranking Models

Distance-based ranking distributions [131, 259] have ranking probabilities that decrease exponentially with increasing distance from some modal or reference ranking $\sigma \in \Omega(\mathcal{A})$:

$$\rho(r|\sigma,\omega) = \frac{1}{\psi(\omega)} \exp(-\omega d(r,\sigma)), \tag{4.9}$$

where $\omega \in [0,\infty)$ is a *dispersion parameter* and $\psi(\omega)$ is a normalizing constant. As $\omega \to \infty$, ρ becomes concentrated at the reference ranking σ, whereas for $\omega = 0$, ρ is the uniform distribution. The Mallows ϕ-model above is an example of such a model (with dispersion $\phi = e^{-\omega}$ and distance d_τ). While we focus on unimodal models, mixtures of such models offer additional modeling flexibility [246].

In this section, we assume that ρ (our distance-based ranking distribution) and c (our distance-based connection function) use the same distance metric to measure similarity of rankings. For instance, when using d_τ as our distance measure, the ranking distribution is

the Mallows ϕ-model and the probability of connection between two nodes is determined based on this same distance.

We first observe that, as $\omega \to \infty$, all nodes' rankings will become more similar and converge to σ. Consequently, the ranking networks model converges to the well-studied Erdős-Rènyi random graph model $G(n,p)$ with $p = c(0|\boldsymbol{\lambda})$. We can also bound the probability $\mathcal{D}(r|\boldsymbol{\theta})$ that a node with a ranking r is connected to a randomly chosen node:

Theorem 4.4.1. *Given reference ranking σ and a distance-based ranking model, for any fixed $\boldsymbol{\theta}$ and any $r \in \Omega(\mathcal{A})$:*

$$\mathcal{D}(\sigma_M, \boldsymbol{\theta}) \leq \mathcal{D}(r, \boldsymbol{\theta}) \leq \mathcal{D}(\sigma, \boldsymbol{\theta}),$$

where σ_M is some ranking at maximum distance from σ.[3]

This theoretical observation has a natural interpretation: individuals that possess more probable or more "popular" rankings have greater odds of connecting with others (hence have higher expected degree). This is not surprising given the nature of distance-based ranking models. In the context of social networks, this is one possible contributing factor to the intuition that people with more "popular" preferences tend to have more friends and social interactions than those with uncommon preferences, i.e., *preference popularity governs social popularity.* From a different perspective, one can use observed node degree to draw inferences about its ranking: higher degree is predictive of more common preferences. (i.e., close to the modal ranking). This can be exploited to support efficient estimation/learning of the reference ranking, the ranking distribution, and preferences of specific individuals. Using Thm. 4.4.1, it is straightforward to bound the edge probability $\mathcal{E}(\boldsymbol{\theta})$.

Proposition 4.4.2. *Given a distance-based ranking network, $\mathcal{E}(\boldsymbol{\theta})$ is bounded by*

$$\mathcal{D}(\sigma_M, \boldsymbol{\theta}) \leq \mathcal{E}(\boldsymbol{\theta}) \leq \mathcal{D}(\sigma, \boldsymbol{\theta}),$$

where σ_M is some ranking at maximum distance from σ.

The upper bound $\mathcal{D}(\sigma, \boldsymbol{\theta})$ can be computed efficiently by:

$$\mathcal{D}(\sigma, \boldsymbol{\theta}) = \hat{\mathcal{D}}(m, \omega, \boldsymbol{\eta}) = \frac{1}{\psi(\omega)} \sum_{k=0}^{d_M(m)} n_k e^{-\omega k} c(k|\boldsymbol{\lambda}), \tag{4.10}$$

where $d_M(m)$ is the maximum possible distance for given d when there are m alternatives (e.g., for Kendall-τ, $d_M(m) = \binom{m}{2}$) and n_i is the number of rankings of distance i from an arbitrary fixed ranking (we present an algorithm for computing n_i below). When n_i is computed for all $i \in \{0, 1, \cdots d_M(m)\}$, Eq. 4.10 can be computed in $O(d_M(m))$ time. For Kendall-τ, the running time is $O(m^2)$. If d is symmetric in the sense that $n_i = n_{d_M(m)-i}$ for all $i \leq d_M(m)$

[3] If more than one ranking has maximum distance, one such ranking minimizes $\mathcal{D}$.

(e.g., Kendall-τ has this symmetric property but Hamming distance lacks this), a lower bound $\mathcal{D}(\sigma_M, \boldsymbol{\theta})$ can be computed by

$$\mathcal{D}(\sigma_M, \boldsymbol{\theta}) = \check{\mathcal{D}}(m, \omega, \boldsymbol{\lambda}) = \frac{1}{\psi(\omega)} \sum_{k=0}^{d_M(m)} n_k e^{-\omega k} c(d_M(m) - k | \boldsymbol{\lambda}) \tag{4.11}$$

For precomputed n_i, Eq. 4.11 can be computed in $O(d_M(m))$ time (e.g., for Kendall-τ, the running time is $O(m^2)$). In general, we can efficiently compute n_i for any d either in closed form or using dynamic programming.

Dynamic Programming for Computing The Number of Rankings

We here present a dynamic programming algorithm for computing n_i for the Kendall-τ distance, where n_i is the number of rankings of distance i from an arbitrary fixed ranking.

We let $T(j, k)$ denote the number of rankings that have distance k from some reference ranking with j alternatives. We note that $T(j, 0) = 1$ for any $j \geq 1$ and $T(j, k) = 0$ for any $k > \binom{j}{2}$. We focus on the fact that distance between two permutations measures the minimum number of swaps or inversions required to transform one permutation to the other. We observe that when there are $j-1$ alternatives, the j^{th} alternative can be inserted in j possible positions, thus causing between 0 and $j-1$ inversions or swaps (i.e., increasing distance by at most $j-1$). This implies that when $k \geq j-1$, the j^{th} alternative can cause $0, 1, \ldots$,or $j-1$ inversions while the other preceding $j-1$ alternatives involve in $k-0, k-1, \ldots$,or $k-j+1$ inversions. Similarly, when $k < j-1$, the j^{th} alternative can be part of either $0, 1, \ldots,$ *or* k inversions while the other preceding $j-1$ alternatives contribute to $k, k-1, \ldots$,or 0 inversions. Hence, we can write this recursive equation:

$$T(j, k) = \sum_{i=0}^{\min(k, j-1)} T(j-1, k-i) \tag{4.12}$$

Since $k \in [0, \binom{m}{2}]$ and $j \in [1, m]$, T can be viewed as a $m \times \frac{m(m-1)}{2}$ matrix. For calculating each element of this matrix, at most $m-1$ summation operations are required. Thus, the running time of this dynamic programming method is $O(m^4)$. However, this time complexity can be improved by slight change of the recursive formula in Eq. 4.12. From Eq. 4.12, we note that for $k \leq j-1$, $T(j, k) - T(j, k-1) = T(j-1, k)$ and for $k > j-1$, $T(j, k) - T(j, k-1) = T(j-1, k) - T(j-1, j-k)$. Hence, the recursive formula in Eq. 4.12 is equivalent to:

$$T(j, k) = \begin{cases} T(j, k-1) + T(j-1, k), & k \leq j-1 \\ T(j, k-1) + T(j-1, k) - T(j-1, k-j), & k > j-1 \end{cases} \tag{4.13}$$

Using this recursion, each element in matrix T can be calculated in $O(1)$, thus yielding in overall time complexity of $O(m^3)$. Note that, as we are interested in those values located in the m^{th} row of matrix T (i.e., $T(m, k)$ for $0 \leq k \leq \binom{m}{2}$), thus using the notation n_k instead of

$T(m,k)$.

Approximations

In this section, we present efficient methods for approximating $\mathcal{D}(r,\boldsymbol{\theta})$ and $\mathcal{E}(\boldsymbol{\theta})$ which exploit Eq. 4.10, Eq. 4.11, and the dynamic programming algorithm introduced above. We have empirically observed that $\mathcal{D}(r,\boldsymbol{\theta})$ (usually) decreases as $d(r,\sigma)$ increases. In other words, the distance of an individual's ranking to the reference ranking is negatively correlated with its degree. Define the linear function

$$\tilde{\mathcal{D}}(d,\boldsymbol{\theta}) = \left(1 - \frac{d}{d_M(m)}\right)\mathcal{D}(\sigma,\boldsymbol{\theta}) + \left(\frac{d}{d_M(m)}\right)\mathcal{D}(\sigma_M,\boldsymbol{\theta}). \tag{4.14}$$

We can approximate $\mathcal{D}(r,\boldsymbol{\theta})$ by $\tilde{\mathcal{D}}(d(r,\sigma),\boldsymbol{\theta})$, which can be used to effectively approximate other structural network properties. For example, the edge density $\mathcal{E}$ and degree distribution P, respectively, can be approximated by:

$$\tilde{\mathcal{E}}(\boldsymbol{\theta}) = \frac{1}{\psi(\omega)}\sum_{k=0}^{d_M(m)} n_k e^{-\omega k}\tilde{\mathcal{D}}(k,\boldsymbol{\theta}), \quad \text{and} \tag{4.15}$$

$$\tilde{P}(k,\boldsymbol{\theta}) = \frac{\binom{n-1}{k}}{\psi(\omega)}\sum_{i=0}^{d_M(m)} n_i e^{-\omega i}\tilde{\mathcal{D}}(i,\boldsymbol{\theta})^k\left(1-\tilde{\mathcal{D}}(i,\boldsymbol{\theta})\right)^{n-1-k}. \tag{4.16}$$

By pre-computing n_i values, these functions can be computed in $O(m^2)$ time for Kendall-τ as $d_M(m) = \binom{m}{2}$ (cf. the $O(m!)$ time required for naive exact computation of $\mathcal{E}$, which needs two nested loops each of which iterates over a ranking space with size of $m!$). For large m, even this might be problematic; but if $c(\cdot|\boldsymbol{\lambda})$ is convex, one can (loosely) approximate $\mathcal{D}$ and edge density $\mathcal{E}$ in $O(1)$ time by:

$$\tilde{\tilde{\mathcal{D}}}(x,\boldsymbol{\theta}) = \left(1-\frac{x}{d_M(m)}\right)c\left(\frac{d_M(m)e^{-\omega}}{1+e^{-\omega}}\right) + \frac{x}{d_M(m)}c\left(\frac{d_M(m)}{1+e^{-\omega}}\right), \text{ and} \tag{4.17}$$

$$\tilde{\tilde{\mathcal{E}}}(\boldsymbol{\theta}) = \frac{1}{1+e^{-\omega}}c\left(\frac{d_M(m)e^{-\omega}}{1+e^{-\omega}}\right) + \frac{e^{-\omega}}{1+e^{-\omega}}c\left(\frac{d_M(m)}{1+e^{-\omega}}\right). \tag{4.18}$$

(The derivations of these loose approximations can be found in Appendix B.) Our approximations $\tilde{\mathcal{E}}(\boldsymbol{\theta})$ and $\tilde{\mathcal{D}}(d(r,\sigma),\boldsymbol{\theta})$ can also be exploited for efficient model learning as we discuss in Section 4.5. Exact evaluation of the likelihood function of a ranking network when dealing with missing data (i.e., when some node rankings are unobserved) has computation of edge density $\mathcal{E}(\boldsymbol{\theta})$ and average connection probability $\mathcal{D}(r,\boldsymbol{\theta})$ as its main bottlenecks. Of course, one can learn model parameters using iterative methods such as EM, but direct evaluation of (an approximation of) the likelihood function using these approximations may be support more efficient learning and inference.

4.5 Model Learning for Distance-Based Ranking Models

We now show how our approximations can be exploited for efficient parameter learning of distance-based ranking models when dealing with missing rankings. Learning model parameters is necessary for predicting missing rankings (we discuss this in Chapter 5) or missing edges in ranking networks.

The basic learning problem is as follows. We first assume that individuals are partitioned into two sets: $O \subseteq \mathcal{N}$, whose complete preference rankings are observed (e.g., elicited or otherwise revealed); and $U = \mathcal{N} \setminus O$, whose preferences are unknown or "missing." Let $R^O = \{r_i | i \in O\}$ be the set of observed rankings and $R^U = \{r_i | i \in U\}$ be the set of random variables associated with unknown preferences. The likelihood function for the observed social network G and observed individual preferences R^O is given by

$$\mathbb{P}(R^O, G|\boldsymbol{\theta}) = \mathbb{P}(G|R^O, \boldsymbol{\theta})\mathbb{P}(R^O|\boldsymbol{\theta}), \tag{4.19}$$

where

$$\begin{aligned}\mathbb{P}(G|R^O, \boldsymbol{\theta}) = &\prod_{i,j \in U; i<j} \mathcal{E}(\boldsymbol{\theta})^{e_{ij}} \left(1 - \mathcal{E}(\boldsymbol{\theta})\right)^{1-e_{ij}} \times \\ &\prod_{i \in O, j \in U} \mathcal{D}(r_i, \boldsymbol{\theta})^{e_{ij}} \left(1 - \mathcal{D}(r_i, \boldsymbol{\theta})\right)^{1-e_{ij}} \times \\ &\prod_{i,j \in O; i<j} c(r_i, r_j|\boldsymbol{\lambda})^{e_{ij}} \left(1 - c(r_i, r_j|\boldsymbol{\lambda})\right)^{1-e_{ij}}, \end{aligned} \tag{4.20}$$

and

$$\mathbb{P}(R^O|\boldsymbol{\theta}) = \prod_{r \in O} \rho(r|\sigma, \omega). \tag{4.21}$$

To learn a maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ of the ranking network parameters $\boldsymbol{\theta}$, one can maximize the log likelihood function

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}) &= \log \mathbb{P}(R^O, G|\boldsymbol{\theta}) \\ &= \sum_{i,j \in U; i<j} e_{ij} \log \mathcal{E}(\boldsymbol{\theta}) + (1 - e_{ij}) \log \left(1 - \mathcal{E}(\boldsymbol{\theta})\right) \\ &+ \sum_{i \in O, j \in U} e_{ij} \log \mathcal{D}(r_i, \boldsymbol{\theta}) + (1 - e_{ij}) \left(1 - \mathcal{D}(r_i, \boldsymbol{\theta})\right) \\ &+ \sum_{i,j \in O; i<j} e_{ij} \log c(r_i, r_j|\boldsymbol{\lambda}) + (1 - e_{ij}) \log \left(1 - c(r_i, r_j|\boldsymbol{\lambda})\right) \\ &+ \sum_{r \in O} \rho(r|\sigma, \omega). \end{aligned} \tag{4.22}$$

which can be computed in $O((m!)^2 + n_o n_u + n_o^2)$ where n_o and n_u represent the number of observed and unobserved preferences respectively (assuming $n_u \geq 1$). Here, the $(m!)^2$ term

arises from exact computation of $\mathcal{E}$ and $n_o n_u$ arises from the second summation in Eq. 4.22, assuming that $\mathcal{D}(r, \theta)$ values are already stored when $\mathcal{E}$ is computed and can be accessed in $O(1)$. Finally, n_o^2 corresponds to the last summation in Eq. 4.22. The exact computation of the function in Eq. 4.22, and thus its direct maximization, is computationally very expensive even for relatively small m. So, we resort to approximating log likelihood $\mathcal{L}$ by

$$\begin{aligned}\tilde{\mathcal{L}}(\boldsymbol{\theta}) = &\sum_{i,j \in U;\, i<j} e_{ij} \log \tilde{\mathcal{E}}(\boldsymbol{\theta}) + (1 - e_{ij}) \log \left(1 - \tilde{\mathcal{E}}(\boldsymbol{\theta})\right) \\ &+ \sum_{i \in O, j \in U} e_{ij} \log \tilde{\mathcal{D}}(r_i, \boldsymbol{\theta}) + (1 - e_{ij}) \left(1 - \tilde{\mathcal{D}}(r_i, \boldsymbol{\theta})\right) \\ &+ \sum_{i,j \in O;\, i<j} e_{ij} \log c(r_i, r_j | \boldsymbol{\lambda}) + (1 - e_{ij}) \log \left(1 - c(r_i, r_j | \boldsymbol{\lambda})\right) \\ &+ \sum_{r \in O} \rho(r | \sigma, \omega), \end{aligned} \tag{4.23}$$

which can be computed in $\mathrm{O}(m^3 + n_o n_u + n_o^2)$ time, where $O(m^3)$ is the running time of the dynamic programming discussed in Sec 4.4. The terms $n_o n_u$ and n_o^2 correspond to second and third summations, respectively, in Eq. 4.23. This approximation is obtained by replacing the $\mathcal{E}$ and $\mathcal{D}$ terms in $\mathcal{L}$ with their corresponding approximations $\tilde{\mathcal{E}}$ and $\tilde{\mathcal{D}}$ that we developed in Sec. 4.4. We can find *approximate maximum likelihood estimates* $\tilde{\boldsymbol{\theta}}$ by maximizing $\tilde{L}$. In our experiments below, we show that these approximate estimates can not only be computed efficiently in practice but also explain the observed data well by maximizing the actual log likelihood function reasonably accurately.

4.6 Empirical Analyses

We describe experiments on a version of our ranking network model, using the Mallows ϕ-model under various parameter settings. We compare various structural properties (e.g., degree distribution, edge density, diameter) of the resulting networks with the predictions of our approximation methods.

Experimental Setup. For the distance-based ranking model with Kendall-τ distance, the normalizing constant is

$$\psi(\omega) = 1 \cdot (1 + e^{-\omega}) \cdot (1 + e^{-\omega} + e^{-2\omega}) \cdots (1 + \cdots + e^{-(m-1)\omega}). \tag{4.24}$$

We use a variant of the connection probability of Serrano *et al.* [322]:

$$c(d | \boldsymbol{\lambda}) = \gamma \left(1 + \frac{d}{\beta}\right)^{-\alpha}. \tag{4.25}$$

Here β controls average degree and $\alpha > 1$ determines the extent of homophily (higher values

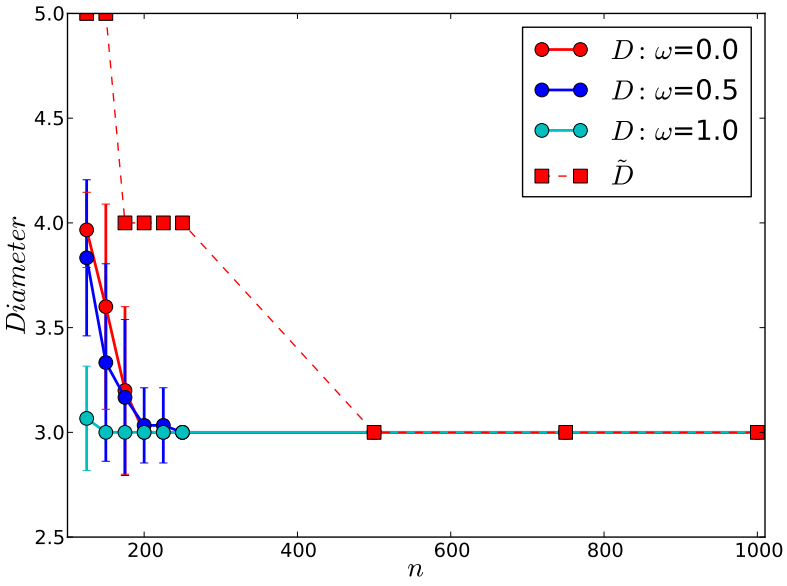


Figure 4.1: Mean and std. dev. of observed diameters (30 instances) for various ω and n, and its approximation $\tilde{D}$ ($m = 3$, $\alpha = 2$, $\gamma = 0.8$, and $\beta = 2$).

of α give greater homophily). We introduce $\gamma \in (0, 1]$ to control the probability of connecting nodes with the same ranking (to account for the discrete nature of ranking space). Unless noted, all experiments are run on networks with $n = 1000$ nodes, $\alpha = 2$, $\beta = 2$ and $\gamma = 0.8$, while varying m. For each parameter setting, we report results over 30 random ranking networks. Error bars in our figures represent one sample standard deviation.

Diameter. We first examine the effect of n and ω on the diameter of the emergent ranking networks, and compare the true observed diameter to our approximation $\tilde{D}$ (see Eq. 4.8). We fix $m = 5$. Fig. 4.1 confirms that diameter shrinks as n increases. Unsurprisingly, diameter decreases with increasing ω (i.e., more uniform distributions give larger diameter), largely due to the increased edge density caused by increasing ω (see below). $\tilde{D}$ provides a reasonable upper bound of diameter for any value of ω, with bounds that are very tight when n reaches 500. This is encouraging since for smaller networks, exact computation of diameter is viable, while the approximation is need for larger networks.

Edge Density. The effect of m and ω on edge density is illustrated in Fig. 4.2 and Fig. 4.3. Fig. 4.2 demonstrates that edge density increases with ω but decreases with m (compare boxes on the solid lines). Error bars are very tight (and barely observable, with a maximum standard deviation 0.0105 for $\omega = 2$ and $m = 5$). Our approximation $\tilde{\mathcal{E}}$ (Eq. 4.15) is relatively close to the observed edge density, especially when ω is relatively small ($\omega \leq 0.5$) or large ($\omega \geq 4$).

To better quantify the accuracy of our approximation, we compute the mean squared error

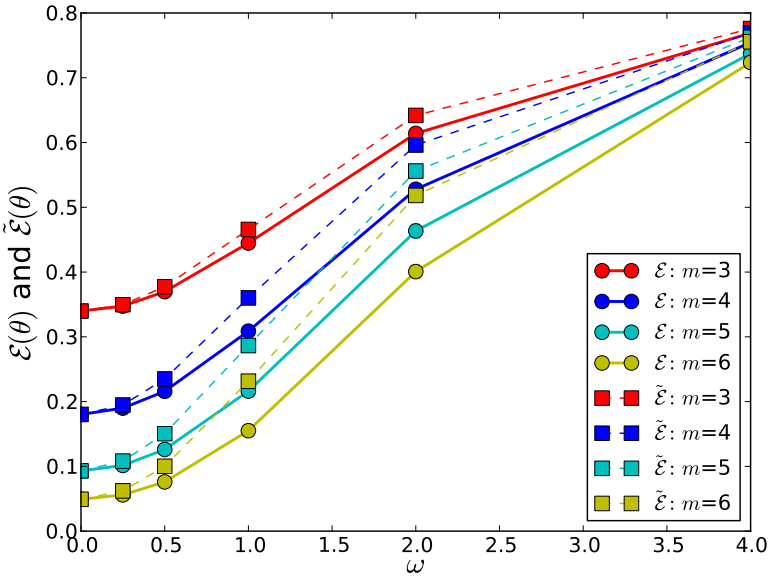


Figure 4.2: Mean observed edge density $\mathcal{E}(\boldsymbol{\theta})$ (30 instances) and its approximation $\tilde{\mathcal{E}}(\boldsymbol{\theta})$; varying number of options m, dispersion ω ($n = 1000$, $\alpha = 2$, $\beta = 2$, $\gamma = 0.8$).

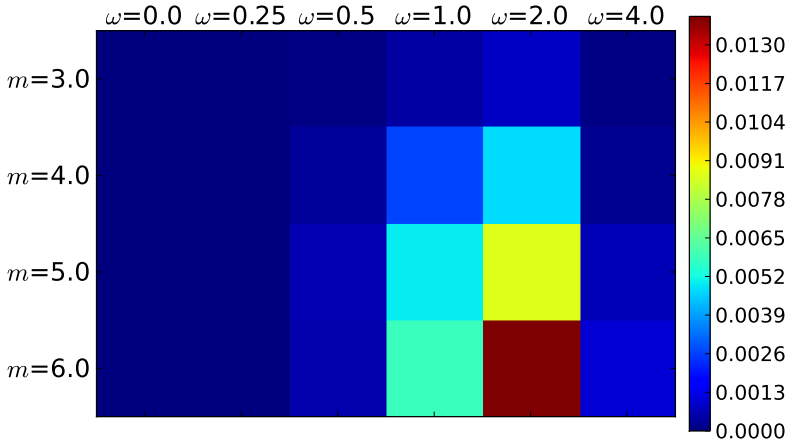


Figure 4.3: Mean squared error of approximate edge density; varying number of options m, dispersion ω ($n = 1000$, $\alpha = 2$, $\beta = 2$, $\gamma = 0.8$).

(MSE) between the actual edge density and our approximation over 30 randomly generated ranking networks. Fig. 4.3 shows that MSE is relatively low with a maximum of roughly 0.014 for $m = 6$, $\omega = 2$. MSE first increases and then decreases with ω. Moreover, MSE increases with m especially for $\omega = 1, 2$. This suggests that, for large m and small ω, our edge density approximation should be used cautiously. Refined approximations that account for the value of ω should be explored.

Degree Distribution. Fig. 4.4 shows the impact of ω on the degree distribution of ranking networks under the Mallows model, as well as the effectiveness of our approximation $\tilde{P}$ of the degree distribution (Eq. 4.16). The figure shows *cumulative degree distribution* to reduce noise in the plots: given degree distribution $P(k, n)$, the cumulative degree distribution is $P_c(k, n) = \sum_{l=k}^{n} P(l, n)$ (i.e., probability that a node n has the degree $d \geq k$). The mean cumulative degree distribution is shown for various values of ω with $m = 5$ and $n = 1000$. The expected degree increases with ω; e.g., with $\omega = 0$, most of nodes have degree around 100, whereas for $\omega = 4$, most nodes have degree around 800 (since mean degrees is $n\mathcal{E}$). However, variance in the degree distribution initially increases as ω increases from 0 but then decreases. For example, variance is low at $\omega = 0$ and $\omega = 4$, but high for $\omega = 1$ and $\omega = 2$. Interestingly, with $\omega = 1$ and $\omega = 2$, the the cumulative degree distribution shows several distinct modes. The relation between these modes and community structure in ranking networks is something that remains to be explored.

Fig. 4.4 also shows that the approximate degree distribution $\tilde{P}$ is reasonably close to actual observed degree distribution when $\omega \leq 0.5$ and $\omega \geq 4$, though it fails to account for the distinct modes, especially for $\omega = 1, 2$. Once again, more accurate approximations (sensitive to ω) may be possible.

Model Learning. We examine the efficiency of our proposed approximated likelihood function for learning model parameters. We compare the learned model parameters under exact and approximated log likelihood functions. To avoid invariant issues in comparison of likelihood estimates—rather than direct comparison of approximated MLE $\tilde{\boldsymbol{\theta}}$ and exact MLE $\hat{\boldsymbol{\theta}}$—we compare $\mathcal{L}(\tilde{\boldsymbol{\theta}})$ with $\mathcal{L}(\hat{\boldsymbol{\theta}})$, where $\mathcal{L}$ is the exact likelihood function. The closer these two values are, the better the approximated likelihood function is at preserving the structure of actual likelihood function. We specially measure the *log likelihood approximation ratio* (LLAR) defined by

$$LLAR = \frac{\mathcal{L}(\hat{\boldsymbol{\theta}})}{\mathcal{L}(\tilde{\boldsymbol{\theta}})},$$

where $\hat{\boldsymbol{\theta}} = \min_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta})$ and $\tilde{\boldsymbol{\theta}} = \min_{\boldsymbol{\theta}} \tilde{\mathcal{L}}(\boldsymbol{\theta})$. We note that $LLAR \in [0, 1]$. The higher LLAR is, the more accurate our approximated parameters are.

We set $(\alpha, \beta, \gamma) = (2, 2, 0.8)$ and $n = 200$ while varying ω and m. After each network generation, each node's ranking is observed or unobserved with probability 0.5 will be considered observed or unobserved. We assume that σ is known but intend to learn $\alpha, \beta, \gamma, \omega$. For

optimization, we deploy gradient ascent where the step length in each iteration is determined by the inexact backtracking line search (BLS) algorithm [58]. We terminate each optimization procedure after 30 iterations. We set the initial parameters to $(\phi, \alpha, \beta, \gamma, \omega) = (0.7, 2, 1, 0.5)$.

Fig. 4.5 shows the LLAR for various m and ω. LLAR is relatively very high with a minimum of roughly 0.97 for $m = 6$, $\omega = 2$. This suggests that our approximate likelihood function can be a reliable surrogate for exact likelihood function. LLAR first decreases and then increases with ω. Moreover, LLAR decreases with m especially for $\omega = 1, 2$. This might indicate that for large m and medium-range ω, our likelihood approximation should be used cautiously. These results are very consistent with those of Fig. 4.3, as LLAR is lower when the edge density approximation error is higher. Fig. 4.6 shows the running time ratio of learning with exact likelihood function over learning with approximation likelihood function. For $m = 3$, the running time for both methods are very close. But, for large m, the approximation method is much faster (e.g., for $m = 6$ approximation methods 200-280 times faster). Not surprisingly, the running time ratio increases with m but is less sensitive to ω.

4.7 Summary and Future Work

We have introduced *ranking networks*, a class of attribute-based (or latent-space) network formation models in which node attributes are rankings over a set of options, and connections are formed between nodes based on the similarity of their underlying rankings. We studied some structural properties (e.g., diameter, connectivity, edge density) of these networks, and showed that our model possesses some characteristics of real-world networks (e.g., shrinking diameter). We developed easy-to-compute approximations for the special class of *distance-based ranking models*, while studying properties that emerge among networks in this class. We also showed how these approximation methods can be used for learning model parameters when one is interested in maximum likelihood estimation with missing data. Empirically, we compared various structural properties (e.g., degree distribution, edge density, diameter) of the randomly-generated networks under a ranking network model with the predictions of our approximation methods. Our results confirm the efficiency and predictive accuracy of our computational approximations.

The ranking network model is a starting point for the broader investigation of the impact of rankings, and preferences in particular, on network formation. Future directions will be discussed in detail in Section 6.3, but briefly include: the analysis of more realistic ranking distributions (e.g., mixture models); generalization to partial rankings or pairwise comparisons; extensions of our model that account for heterogeneity/heterophily; modeling the dynamics and mutability of the underlying rankings themselves in response to network connections; and a simpler descriptive model which lacks a generative process but still can capture ranking correlations over a ranking network. Another practical extension is to incorporate the other nodal attributes in addition to rankings in the process of network formation. This may prove

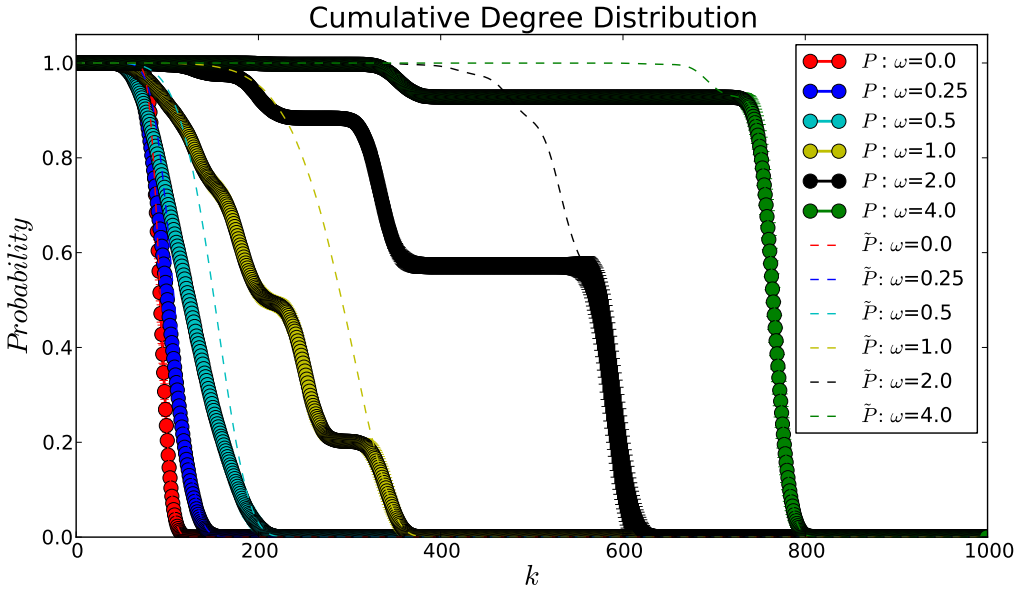


Figure 4.4: The mean cumulative degree distribution (30 instances) and its approximation, varying ω ($m = 5$, $n = 1000$, $\alpha = 2$, $\beta = 2$, $\gamma = 0.8$).

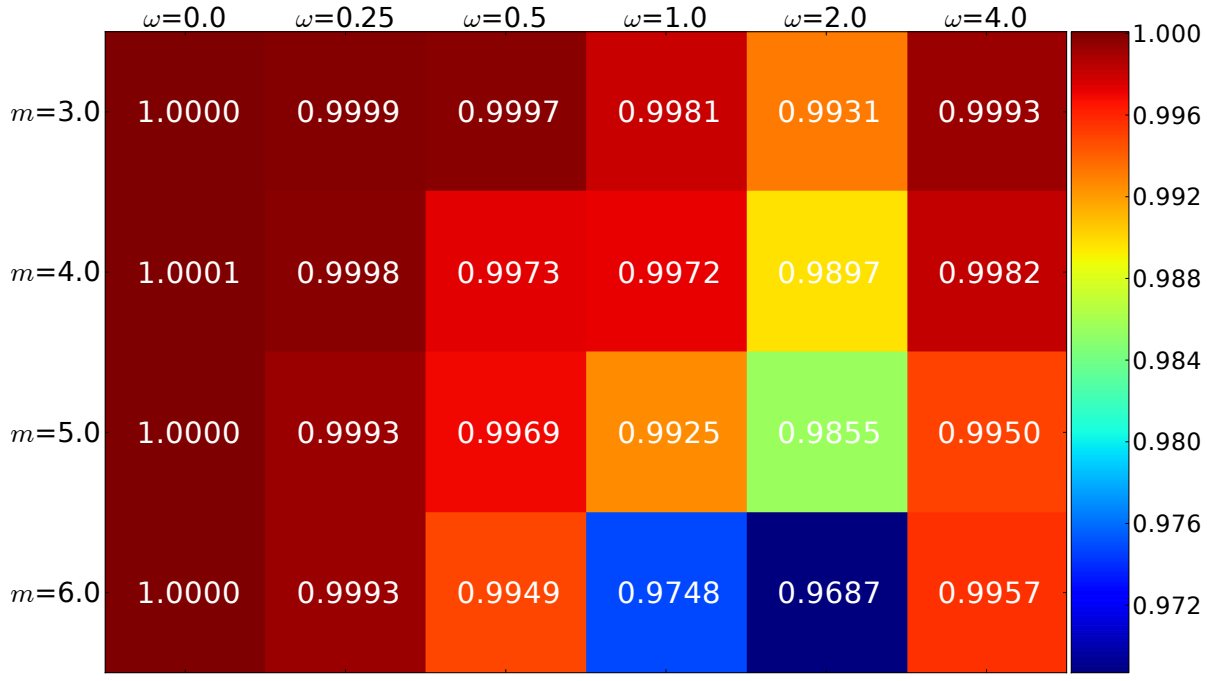


Figure 4.5: The log likelihood approximation ratio (LLAR), varying ω and m, but $n = 200$, $\alpha = 2$, $\beta = 2$, $\gamma = 0.8$).

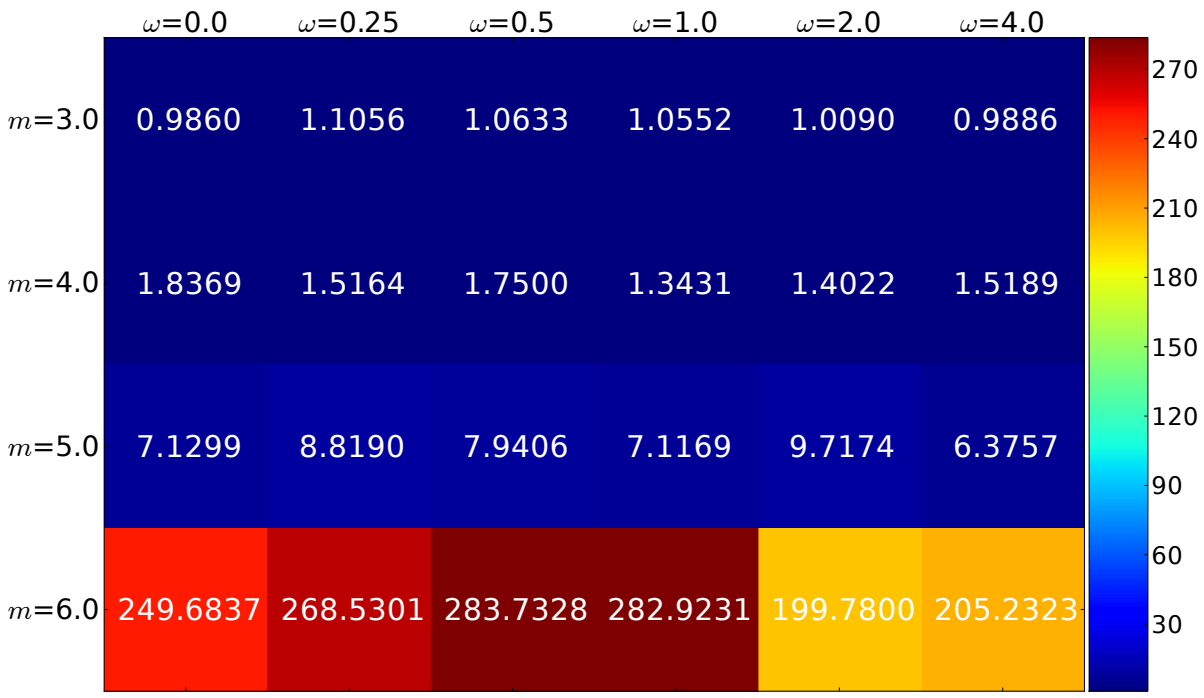


Figure 4.6: The running time ratio of the exact method over the approximation method for model learning, varying ω and m, but $n = 200$, $\alpha = 2$, $\beta = 2$, $\gamma = 0.8$).

to be useful in fitting model to real-world data when the rankings over a set of alternatives are not the only factors in the formation of the network.

Of practical importance is studying the extent to which rankings, such as preferences, play a role in shaping connections in real-world networks, and how the induced correlations can be best exploited in applications such as recommender systems, advertising, social choice and voting, web search and information retrieval; this last direction is partially addressed and discussed in the next chapter.

Chapter 5

Group Recommendation on Preference-Oriented Social Networks

It is widely recognized that individuals' behaviors and preferences are correlated [88, 118, 187], with those of their friends or connections (e.g., correlations of music tastes [234]). This has lead to research focused on inferring individual attributes and behaviour using social connections, e.g., inference of ratings over items [252, 253, 191], latent group membership [205], categorical attributes [204], or latent "social positions" [183]. Yet surprisingly, using social networks to infer individual preferences—in the form of rankings of alternatives—has received little attention. Methods for inference and learning of preference rankings have been extensively studied in econometrics, psychometrics, statistics, and machine learning and data mining; in the latter case, they find application to recommender systems, collaborative filtering, information retrieval and group decision/recommendation problems (i.e., social choice), especially when faced with partial information. In contrast to cardinal utilities, preference rankings (or ordinal preferences) are of special interest in social choice and group recommendation, since they help circumvent, to some extent, the problem of interpersonal comparisons of utilities [14, 320].

In this chapter, we address how to use *social network structure* to support more accurate inference of preference rankings and to make group decisions when some individual preferences are unknown [313]. Specifically, we exploit the fact that *homophily* or *social selection*—association with similar individuals—and *social influence*—adoption of properties and attitudes of those to whom one is connected—can be used to infer individual preferences more efficiently and with less data. This can in turn support more accurate group decision making with partial preferences. To capture correlations of preference rankings over social networks, we introduce *preference-oriented social networks (POSNs)*, a particular form of the ranking network model (introduced in Chapter 4), in which the similarity of the preference rankings of two individuals determines the odds with which they are connected. We exploit this model to infer unobserved individual preferences given observed preferences of others in the network. Intuitively, if we know something about the preferences of an individual's friends, family or colleagues—or their

friends, etc.—we should be able to more accurately predict their preference ranking if homophily or social influence shapes network dynamics. Moreover, we demonstrate how network structure, by allowing such predictions, can be used to support effective group recommendations/decisions with incomplete preferences.

This chapter is organized as follows. We introduce the POSN model in Sec. 5.1. In Sec. 5.2, we describe the inference and social choice problems on POSNs and discuss related work. In Secs. 5.3 and 5.4 we characterize probabilistic properties of the model and develop an MCMC method for inferring unobserved preferences. In Sec.5.6, we evaluate the ability of our methods to predict unobserved preferences and make accurate group decisions with partial preferences. Finally, we conclude and outline some future directions in Sec. 5.7

5.1 Preference-oriented Social Networks

We start by outlining our basic model (we contrast it with existing network generation models in Sec. 5.2.3). A *preference-oriented social network (POSN)* includes: (i) a social network, where nodes represent individuals, and edges represent some social relationship; and (ii) a finite set of options over which individuals have preferences, where these preferences take the form of an ordering or ranking. The model also includes a probabilistic generative process used to generate individual preferences and connections. This process captures the correlation of individuals' preferences across the social network.

The network in a POSN is an undirected graph $G = (\mathcal{N}, E)$ over individuals $\mathcal{N} = \{1, \ldots, n\}$. We can specify G using a binary adjacency matrix $[e_{ij}]$ where $e_{ij} = 1$ iff $(i, j) \in E$. We generally think of these edges as corresponding directly to some relationship in a social network. We assume a finite set of *alternatives (or options)* $\mathcal{A} = \{a_1, \ldots, a_m\}$, e.g., a set of products, political candidates, policies, genre of movies, etc. over which individuals have preferences. The preference of node (or individual) i is a *ranking* (or strict total order) r_i over $\mathcal{A}$. Let $\Omega(\mathcal{A})$ denote the set of all $m!$ rankings over $\mathcal{A}$.

The generative process for POSNs has two stages as in a standard ranking network: first, individual preferences are drawn from a ranking distribution; then individuals form connections with a probability increasing with the *similarity* of their preferences. Each node i's preference ranking r_i is drawn independently from some (parametrized) distribution $\rho(r|\boldsymbol{\eta})$ over $\Omega(\mathcal{A})$ with parameters $\boldsymbol{\eta}$. Many ranking distributions can be used, e.g., Plackett-Luce, Bradley-Terry, etc. (see Section 2.2.2 for an overview). Here, we focus on the *Mallows ϕ-model*, characterized by a "modal" *reference ranking* σ and a *dispersion parameter* $\phi \in [0, 1)$, with the probability of a ranking r decreasing exponentially with its *τ-distance* from σ (see Section 2.2.2 for the formal definition and details).

To compute connection probabilities, we define the similarity of two rankings using the τ distance metric (see Section 2.2.1). Intuitively, $d_\tau(r_i, r_j)$ measures the number of pairwise

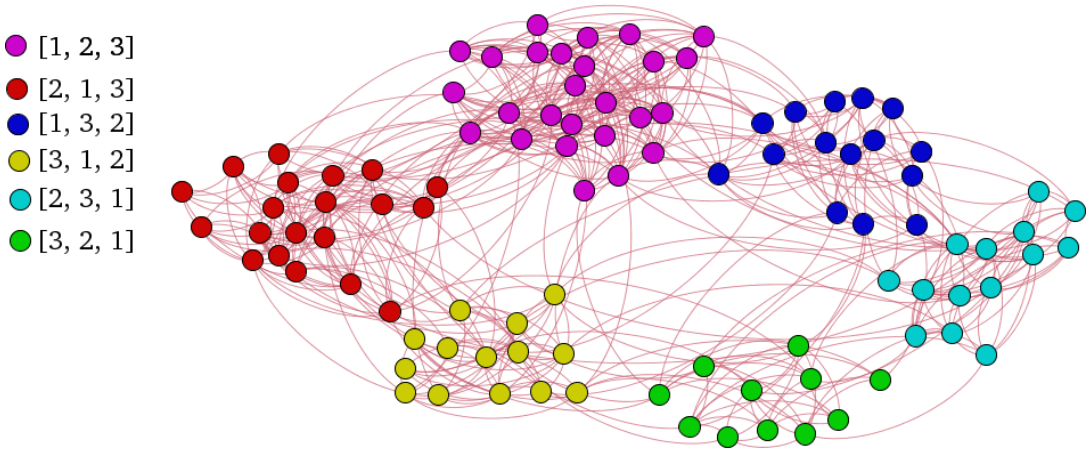


Figure 5.1: A randomly generated POSN for $n = 100$ individuals and $m = 3$ alternatives. The individuals' rankings are drawn from a ϕ-Mallows model with $\phi = 0.7$ and reference ranking $\sigma = (1, 2, 3)$. The connection probability function in Eq 5.1 is used with $(\alpha = 2; \gamma = 0.5; \beta = 0.5)$.

swaps needed to transform r_i to r_j.[1] A strictly decreasing *connection probability function* $c(d) : [0, \infty) \to [0, 1]$ specifies the probability that two nodes i, j are connected given the distance $d_\tau(r_i, r_j)$ between their corresponding rankings. We use the following connection function [322]:

$$c(d|\boldsymbol{\lambda}) = \gamma \left(1 + \frac{d}{\beta}\right)^{-\alpha}. \tag{5.1}$$

Here β controls average node degree and $\alpha > 1$ determines the extent of homophily (greater α implies more homophily). We use $\gamma \in (0, 1]$ to control the odds of connecting nodes with the same ranking (accounting for the discrete nature of ranking space). We sometimes write the connection probability as $c(r_i, r_j)$. Denote the parameters of c by $\boldsymbol{\lambda} = (\alpha, \beta, \gamma)$; the parameters of the ranking distribution by $\boldsymbol{\eta} = (\sigma, \phi)$; and all POSN parameters by $\boldsymbol{\theta} = (\boldsymbol{\lambda}, \boldsymbol{\eta})$. Fig.5.1 illustrates a small POSN, where individuals have preferences over three options. Nodes with similar preferences are more densely connected.

Our POSN model is an instance of a more general notion of a *ranking network*, introduced in Chapter 4, in which latent attributes are generic rankings over options. We analyzed the general topological properties of this model in Chapter 4. In this chapter, we focus on its application on preference inference and group recommendation when faced with missing preferences.

[1] Other distance metrics for rankings can be used as well, e.g., Spearman's rho or footrule, or Hamming distance.

5.2 Inference and Social Choice

We now address two tightly connected problems, *preference inference* and *single-option group recommendation* (or consensus decision making). While preference inference is interesting in its own right, it plays a vital role in group recommendation when preferences of some group members are unobserved.

5.2.1 Preference Inference

We assume that individuals are partitioned into two sets: $O \subseteq \mathcal{N}$ is the set of individuals, whose complete preference rankings are observed (e.g., elicited or otherwise revealed); and $U = \mathcal{N} \setminus O$ are those whose preferences are unobserved or "missing." Let $R^O = \{r_i | i \in O\}$ be the set of observed rankings and $R^U = \{r_i | i \in U\}$ be the set of random variables associated with unknown preferences. Our inference goal, at the high level, is to infer unobserved preferences R^U from the observed preferences R^O and social network. This setting captures scenarios in which a decision maker has access to the complete preferences of a part of some society/group but can't gain access to the preference of the other part. For example, a company may have access to the preferences of existing customers, and want to market to new prospects without knowing their preferences.[2]

In this chapter, we assume that the network G and model parameters $\boldsymbol{\theta}$ are known. Learning model parameters given observed preferences is an important problem which is partially addressed in Sec. 4.5; but other learning methods can exploit our solution to the inference problem (e.g., when using EM).

Our goal in preference inference is to compute the posterior distribution over unobserved preferences $\mathbb{P}(R^U | G, R^O, \boldsymbol{\theta})$ given observed preferences R^O. We discuss sampling methods for approximating the posterior distribution in Sec. 5.4. Other inference problems include the *most probable explanation (MPE)*, i.e., finding the instantiation of R^U which maximizes the posterior:

$$R^{MPE} = \arg\max_{R^U} \mathbb{P}(R^U | G, R^O, \boldsymbol{\theta}). \tag{5.2}$$

We may also be interested in the posterior over the preferences of a single individual $i \in U$:

$$\mathbb{P}(r_i | G, R^O, \boldsymbol{\theta}) = \sum_{R^U \setminus \{r_i\}} \mathbb{P}(R^U | G, R^O, \boldsymbol{\theta}) \tag{5.3}$$

[2]Our model extends to observations of partial preferences. To accommodate this, one can adopt paired-compared distributions (see Section 2.2.2) rather than ranking distributions. The same τ-distance metric and connection probability function can be used. Our inference and decision making algorithms can accommodate such changes with some minor modifications (e.g., an appropriate paired-compared proposal distribution should be adopted in our inference algorithm).

and as well as the "individual MPE:"

$$r_i^{MPE} = \arg \max_{r \in \Omega(\mathcal{A})} \mathbb{P}(r_i = r | G, R^O, \boldsymbol{\theta}). \tag{5.4}$$

5.2.2 Group Recommendation

A key goal in this work is to exploit social network structure to make higher quality group decisions with incomplete preference information. Suppose we need to select an option from $\mathcal{A}$ for a group or "subpopulation" $\mathcal{S} \subseteq \mathcal{N}$ using some preference aggregation method (i.e., a *social choice function*, for example, a voting rule). We distinguish the subsegment $\mathcal{S}$ (e.g., friends planning an activity, the electorate in a small district) from the larger society $\mathcal{N}$ (e.g., users of an online social network, eligible voters in a country): while many group decisions are local, they can be supported by knowledge of the preferences of individuals outside that group. We focus on the choice of a single option using the notion of "social welfare maximization" relative to a *scoring rule* $g : (\mathbb{N}, \mathbb{N}) \to \mathbb{R}^+$. Here, $g(k, m)$ is the positional score of an option ranked k^{th} relative to m options (the Borda and plurality score are common examples). Define the *social welfare* of $a \in \mathcal{A}$:

$$sw(a, \mathcal{S}) = \sum_{i \in \mathcal{S}} g(r_i(a), m), \tag{5.5}$$

with the goal of selecting $a^* \in \mathcal{A}$ that maximizes $sw(., \mathcal{S})$.

In general, we will not know the preferences of all individuals in $\mathcal{S}$, requiring that we *infer* the social welfare of an option a given the observations at hand. Define $sw(a, \mathcal{S} | R^O, G, \boldsymbol{\theta})$ to be this *inferred social welfare*, which varies depending on the method of inference (we sometimes omit mention of G and $\boldsymbol{\theta}$). Assuming each individual's contribution to social welfare is independent, it can be decomposed into *revealed social welfare* $sw_{rev}(a, R^O_{\mathcal{S}})$ (corresponding to observed preferences) and *inferred social welfare* $sw_{inf}(a, R^U_{\mathcal{S}} | R^O)$ (for unobserved preferences):

$$sw(a, \mathcal{S} | R^O) = sw_{rev}(a, R^O_{\mathcal{S}}) + sw_{inf}(a, R^U_{\mathcal{S}} | R^O). \tag{5.6}$$

Revealed social welfare is straightforward:

$$sw_{rev}(a, R^O_{\mathcal{S}}) = \sum_{r \in R^O_{\mathcal{S}}} g(r(a), m). \tag{5.7}$$

But there are various ways to define inferred social welfare. We define several forms of inferred social welfare. We briefly discuss the computational demands of each such form, and in Sec. 5.6 study the extent to which the final decisions of each form coincide.

Expected Score (ES). One might be interested in expected social welfare under the posterior distribution $\mathbb{P}(R^U_{\mathcal{S}} | G, R^O)$. For this purpose, a natural way to define inferred social welfare is

the *expected score*:

$$sw^{E}_{inf}\left(a, R^{U}_{\mathcal{S}}|G, R^{O}\right) = \sum_{R^{U}_{\mathcal{S}}} \mathbb{P}(R^{U}_{\mathcal{S}}|G, R^{O}) \sum_{i \in U_{\mathcal{S}}} g(r_i(a), m), \tag{5.8}$$

which can be computed in $O(m!^{|U_{\mathcal{S}}|}|U_{\mathcal{S}}|)$ time (where $U_{\mathcal{S}}$ are those individuals with unobserved preferences). Since there are $m!$ possible rankings for each unobserved preference, the cardinality of $R^{U}_{\mathcal{S}}$ is $m!^{|U_{\mathcal{S}}|}$ where $|U_{\mathcal{S}}|$ is the number of unobserved preferences in group S. If $\mathbb{P}(r_i|G, R^{O})$ is pre-computed for each $i \in U_S$, we can write

$$sw^{E}_{inf}\left(a, R^{U}_{\mathcal{S}}|G, R^{O}\right) = \sum_{i \in U_{\mathcal{S}}} \sum_{r_i} \mathbb{P}(r_i|G, R^{O}) g(r_i(a), m), \tag{5.9}$$

which can be computed in $O(m!|U_{\mathcal{S}}|)$ time. The term $m!$ again arises because of the size of ranking space.

Joint Most Probable Explanation Score (JMPES). One might consider the most probable rankings for R^{MPE} as the inferred rankings for the missing preferences. JMPES uses the instantiation of unobserved preferences, R^{MPE}, which maximizes the joint posterior $\mathbb{P}(R^{U}|G, R^{O}, \boldsymbol{\theta})$:

$$sw^{JM}_{inf}(a, \mathcal{S}, R^{MPE}) = \sum_{i\mathcal{S}} g(r^{MPE}_{i}(a), m). \tag{5.10}$$

If R^{MPE} is given, this can be computed in $O(|U_{\mathcal{S}}|)$ time , which is the number of iterations needed for the summation in Eq. 5.10.

Individual Most Probable Explanation Score (IMPES). One might consider an inferred preference for individual i as the post probable ranking under i's posterior distribution $\mathbb{P}(r_i|R^{O}, G)$. We can define and compute inferred social welfare using these instantiations. More specifically, we use the instantiation r^{MPE}_{j} for each $j \in U$ that maximizes the posterior $\mathbb{P}(r_j|G, R^{O}, \boldsymbol{\theta})$, and define $IMPES$ as

$$sw^{IM}_{inf}(a, \{r^{MPE}_{j}\}) = \sum_{j \in U_{\mathcal{S}}} g(r^{MPE}_{j}(a), m). \tag{5.11}$$

If the r^{MPE}_{j} for all $j \in \mathcal{S}$ are given, this is computable in $O(|U_{\mathcal{S}}|)$ time, which is the number of iterations that required in the summation of Eq. 5.11.

5.2.3 Related Work and Models

We review the related work on group recommendation in data mining and recommender systems. Then, we highlight the related work in network formation models, nodal attribute inference, preference ranking learning, collaborative filtering methods using social networks, and decision making on social networks.

Group Recommendation. Group recommendation can be broadly categorized as follows: (i) *Virtual/artificial profile methods* (see, e.g., [264]), where joint artificial user profiles for each group of users are created to keep track of their joint revealed/elicited preferences; (ii) *Profile-merging methods* (see, e.g., [368, 33]), which merge group member profiles to form a group profile, based on which recommendations are made; (iii) *Recommendation/scoring aggregation* methods (see, e.g., [261, 24, 12, 318, 148]), which aggregate the recommendations (or inferred preferences) for each group member into single group recommendation list (or recommended option). This aggregation is usually conducted by a *group consensus function* (or social choice function). Our method falls into this third category.

Group Recommendation using Social Factors. Group recommendations based on social factors or interaction patterns have recently drawn a fair amount of attention. Masthoff and Gatt [262] analyse the effect of group member relationship types on their emotional conformity and contagion in a group recommendation task. Social relationship strength has been considered in a group collaborative filtering context [301].

Network Formation Models. Our POSN model lies in the class of random, static network formation models [279]. It is also a *spatial (or latent space) networks* [183, 27], where nodes possess *latent attributes* and are connected (see Section 2.3.3).

Nodal Attribute Inference. Inference of nodal attributes, given social network structure, has also received attention (See Section 2.3.4). However, the inference of rankings as nodal attribute has largely been unaddressed in this work.

Learning Preference Rankings. Distributional models of rankings are widely studied in statistics, psychometrics and machine learning (see Section 2.2 for details), though accounting for social network structure has been unaddressed. Our model is distinct from those above as it models preference correlations induced by social ties, requiring new sampling and inference methods.

Collaborative Filtering and Social Networks. Collaborative Filtering (CF) methods which exploit social networks for rating prediction have recently become popular (see for example, [252, 191, 253, 361, 156, 260, 219] and [359] for a recent survey). See Section 2.3.4. Our model differs in that it considers preference ranking correlations rather than ratings correlations over social networks, and in its focus on group rather than individual recommendation.

5.3 Target Joint Posterior Distributions

In this section, we explain how the joint posterior distribution $\mathbb{P}(R^U|R^O, G)$ can be computed in the POSN model. The computation of this posterior distribution is necessary for our inference and group decision making tasks.

We observe that, using Bayes rule, the posterior over R^U given observed preferences R^O

and network G is given by:

$$\mathbb{P}(R^U|G, R^O, \boldsymbol{\theta}) = \frac{\mathbb{P}(G|R^U, R^O)\mathbb{P}(R^U|\boldsymbol{\eta})}{\mathbb{P}(G|R^O, \boldsymbol{\theta})}, \tag{5.12}$$

We now describe each of the components of Eq. 5.12, $\mathbb{P}(R^U|\boldsymbol{\eta})$, $\mathbb{P}(G|R^O, \boldsymbol{\theta})$, and $\mathbb{P}(G|R^U, R^O)$ in turn.

Assuming that the preference distribution parameters $\boldsymbol{\eta}$ are given, the joint prior distribution over R^U (due to being independent and identical distribution) is:

$$\mathbb{P}(R^U|\boldsymbol{\eta}) = \prod_{r \in R^U} \rho(r|\boldsymbol{\eta}), \tag{5.13}$$

where $\rho(r|\boldsymbol{\eta})$ is the preference distribution.

We now describe the form and structure of the joint distribution $\mathbb{P}(G|R^O, \boldsymbol{\theta})$ induced by POSNs. We first focus on the probability $\mathbb{P}(e_{ij} = 1)$ with which an edge occurs between two nodes i and j in G. We define it under three conditions: (1) the preferences of both i and j are unobserved (and drawn independently from $\rho(r|\boldsymbol{\eta})$); (2) one is observed and the other unobserved; and (3) both are observed.

Unobserved preferences for both nodes. In this case, $\mathbb{P}(e_{ij} = 1|\boldsymbol{\theta})$ is the chance of an edge between two nodes whose preferences are drawn independently from $\rho(r|\boldsymbol{\eta})$:

$$\mathcal{E}(\boldsymbol{\theta}) = \sum_{r \in \Omega(\mathcal{A})} \sum_{r' \in \Omega(\mathcal{A})} \rho(r|\boldsymbol{\eta})\rho(r'|\boldsymbol{\eta}) c\left(d_\tau(r', r)|\boldsymbol{\lambda}\right), \tag{5.14}$$

(The expected number of edges in a POSN is $\binom{n}{2}\mathcal{E}(\boldsymbol{\theta})$.)

Unobserved preference for one node. When only one node's preference is observed (say i), one can compute $\mathbb{P}(e_{ij} = 1|r_i, \boldsymbol{\theta})$ by

$$\mathcal{D}(r_i, \boldsymbol{\theta}) = \sum_{r' \in \Omega(\mathcal{A})} \rho(r'|\boldsymbol{\eta}) c\left(d_\tau(r', r_i)|\boldsymbol{\lambda}\right). \tag{5.15}$$

$\mathcal{D}(r_i, \boldsymbol{\theta})$ also determines the expected degree of a node with ranking r_i, which is simply $(n-1)\mathcal{D}(r_i, \boldsymbol{\theta})$.

Observed preferences for both nodes. The edge probability between i and j when both r_i and r_j are observed is $\mathbb{P}(e_{ij} = 1|r_i, r_j, \boldsymbol{\theta}) = c(d_\tau(r_i, r_j)|\boldsymbol{\lambda})$.

Using these edge probabilities, the probability $\mathbb{P}(G|R^O, \boldsymbol{\theta})$ of graph structure G given ob-

served preferences R^O is:

$$\mathbb{P}(G|R^O,\boldsymbol{\theta}) = \prod_{i,j\in U;\, i<j} \mathcal{E}(\boldsymbol{\theta})^{e_{ij}} \left(1-\mathcal{E}(\boldsymbol{\theta})\right)^{1-e_{ij}} \times \prod_{i\in O, j\in U} \mathcal{D}(r_i,\boldsymbol{\theta})^{e_{ij}} \left(1-\mathcal{D}(r_i,\boldsymbol{\theta})\right)^{1-e_{ij}} \times \prod_{i,j\in O;\, i<j} c(r_i,r_j|\boldsymbol{\lambda})^{e_{ij}} \left(1-c(r_i,r_j|\boldsymbol{\lambda})\right)^{1-e_{ij}}. \tag{5.16}$$

We formulate $\mathbb{P}(G|R^U,R^O)$ by focusing on the probability $\mathbb{P}(e_{ij}|r_i,r_j)$ of an edge between i and j. Since $\mathbb{P}(e_{ij}|r_i,r_j) = c(r_i,r_j)^{e_{ij}}(1-c(r_i,r_j)^{1-e_{ij}}$, we have:

$$\mathbb{P}(G|R^O R^U) = \prod_{\substack{i,j\in\mathcal{N}\\ i<j}} c(r_i,r_j)^{e_{ij}} \left(1-c\left(r_i,r_j\right)\right)^{1-e_{ij}} \tag{5.17}$$

As noted earlier, using Bayes rule, the posterior over R^U given observed preferences R^O and network G is given by:

$$\mathbb{P}(R^U|G,R^O,\boldsymbol{\theta}) = \frac{\mathbb{P}(G|R^U,R^O)\mathbb{P}(R^U|\boldsymbol{\eta})}{\mathbb{P}(G|R^O,\boldsymbol{\theta})}, \tag{5.18}$$

where one can use Eq. 5.13, Eq. 5.16, and Eq. 5.17 for computation of $\mathbb{P}(R^U|\boldsymbol{\eta})$, $\mathbb{P}(G|R^O,\boldsymbol{\theta})$, and $\mathbb{P}(G|R^U,R^O)$, respectively. We exploit this joint posterior distribution in our inference and group decision making tasks.

5.4 Sampling Methods

For both preference inference and group recommendation, we must compute the joint posterior distribution $\mathbb{P}(R^U|G,R^O,\boldsymbol{\theta})$. Exact computation is, not surprisingly, computationally expensive. The main computational bottle neck comes from the *partition function* (i.e., normalizing constant) $\mathbb{P}(G|R^o,\boldsymbol{\theta})$ (refer to Eq. 5.12). The exact computation of $\mathbb{P}(G|R^o,\boldsymbol{\theta})$ (see Eq. 5.16) requires evaluation of $\mathcal{E}(\boldsymbol{\theta})$ and $\mathcal{D}(r_i,\boldsymbol{\theta})$ which need $O(m!^2)$ and $O(m!)$ time, respectively. So we develop sampling methods to approximate the posterior. At a high level, we sample L *preference profiles* $R^{(1)},\ldots,R^{(L)}$ from the posterior, where each profile consists of a preference ranking for each individual $i\in U$, whose preferences are unobserved. Let $R_i^{(t)}$ denote the sampled preference ranking of individual $i\in U$ in the t^{th} profile. We approximate the posterior preference of any individual $i\in U$, $\mathbb{P}(r_i=l|G,R^O,\boldsymbol{\theta})$, $\forall\ l\in\Omega(\mathcal{A})$, and the expected score of

the inferred component of social welfare $sw^E_{inf}\left(a, G, R^O, O\right)$ as follows:

$$\mathbb{P}(r_i = l | G, R^O, \boldsymbol{\theta}) \approx \frac{1}{L} \sum_{t=1}^{L} I[R_i^{(t)} = l], \text{and} \tag{5.19}$$

$$sw^E_{inf}\left(a, G, R^O\right) \approx \frac{1}{L} \sum_{t=1}^{L} \sum_{i \in U} g(R_i^{(t)}(a), m) \tag{5.20}$$

Here $g(R_i^{(t)}(a), m)$ is the positional score of option a in i's preference ranking for the t^{th} sample. L must be sufficiently large to ensure a good approximation. More critically, we need the ability to draw independent samples from the (unknown) posterior. To do this, we use an Markov Chain Monte Carlo (MCMC) algorithm, specifically, *Gibbs sampling*, where individual variables are in turn sampled using *Metropolis sampling* [43].

We use iterative Gibbs sampling to sample unobserved preferences R^U. The process begins with an initial preference profile $R^{(0)}$, completing the rankings for all unobserved preferences by sampling independently from a ϕ-Mallows model. At each iteration l, we sample $R_i^{(l)}$ for each $i \in U$ from the conditional distribution

$$\mathbb{P}(r_i | R_1^{(l)}, \ldots, R_{i-1}^{(l)}, R_{i+1}^{(l-1)}, \ldots, R_{|U|}^{(l-1)}, R^O). \tag{5.21}$$

The order in which preferences are sampled can impact the efficiency of the method. The order can be deterministic or stochastic, and may be based on node degree or the number of observed preferences of their neighbors. In our experiments, we use a fixed arbitrary ordering.

Let R_{-i} be the set of all rankings except i's ranking. To sample r_i from the distribution $\mathbb{P}(r_i | R_{-i})$ we use *Metropolis sampling*. By Eqs. 5.16–5.12 and 5.13, the probability of r_i given all other individual preferences is $\mathbb{P}(r_i | R_{-i}) \propto \tilde{p}(r_i)$, where

$$\tilde{p}(r_i) = \prod_{j \in U} \mathbb{P}(e_{ij} | r_i, r_j) \prod_{j \in O} \mathbb{P}(e_{ij} | r_i, r_j) \phi^{d_\tau(\sigma, r_i)} \tag{5.22}$$

which can be computed in $O(n)$ time. To sample r_i at iteration l of Gibbs, we sample r^* from a *conditional proposal distribution*

$$q(r^* | R_i^{(l-1)}) = \frac{1}{z(\hat{\phi}_i)} \hat{\phi}_i^{\, d_\tau(r^*, R_i^{(l-1)})}, \tag{5.23}$$

which is a Mallows distribution that uses the previous sample of i's preference $R_i^{(l-1)}$ as its reference ranking and a dispersion parameter $\hat{\phi}_i$ (in our experiments, we fix $\hat{\phi}_i = \phi$; i.e., fixed to be identical to the POSN model preference dispersion parameter). In this proposal distribution, the new proposed sample will be relatively close the previous sample. The dispersion parameter $\hat{\phi}_i$ can control the extent of exploration. A small ϕ_i results in less exploration as the proposed sample will be closer to the previous sample. We accept proposal r^* as $R_i^{(l)}$ (i.e., set $R_i^{(l)} = r^*$)

with probability

$$A(r^*, R_i^{(l-1)}) = \min\left(1, \frac{\tilde{p}(r^*)}{\tilde{p}(R_i^{(l-1)})}\right); \tag{5.24}$$

otherwise, we set $R_i^{(l)} = R_i^{(l-1)}$. To sample from the Mallows model $q(.)$, we use the *repeated insertion model* [112]. One can sample L preference profiles given $|R^U|$ unobserved preferences in $O(L|R^U|nm)$ time using our proposed method. Assuming $|R^U|$ is a constant fraction of n, our sampling methods runs in $O(Ln^2m)$ time, which may prove intractable for very large networks. Designing more scalable sampling methods is an important future direction, but we partially address this in next section.

5.5 Speeding Up Sampling Methods

We here relax some dependency assumptions in our original POSN models to speed up our sampling methods. The intent is to (possibly) make our predictions more accurate by removing unrealistic dependence assumptions and retaining more realistic dependence. We first observe that $\mathbb{P}(G|R^O, \boldsymbol{\theta})$ in Eq. 5.16 and $\mathbb{P}(G|R^O, R^U, \boldsymbol{\theta})$ in Eq. 5.17 are defined over all pairs of nodes. This not only renders learning model parameters and inference computationally costly, but also imposes a strong assumption on negative examples as we treat lack of an edge as a negative example. To address both of these issues, we allows only a subset of all possible node pairs to contribute the definition of the network probabilities, thus simultaneously reducing number of negative examples and speeding up our sampling methods. By pruning negative examples, we make our model more realistic in the sense that not all absent edges should be treated as negative examples as some people/nodes might not have a chance to interact with each other.

We define the *opportunity network* $H = (\mathcal{N}, E_H)$, where the set of edges E_H determines which pairs of nodes should participate in the probability computation of the observed network G. The opportunity network H captures the idea that not all pairs have opportunities to make connection to each other. For the computation of network probability, one should account for such opportunities. For example, when two nodes have common neighbours, they are more likely to have opportunity to connect to each other than when they don't have common neighbours.

We usually select E_H in a way that covers all observed edges in G (as the existence of a social tie is an evidence for presence of interaction opportunity) but it includes small subset of non-edge pairs (i.e., negative examples). For example, in our experiments below, we stochastically create the opportunity network H by setting the edge $e_{ij}^H = 1$ with the probability $d_G(i,j)^{-\xi}$ where e_{ij}^H is a binary variable for edge between i and j in network H, $d_G(i,j)$ is the geodesic distance between i and j in the observed network G, and $\xi \in (0, \infty)$ is a parameter (described below).[3] Doing so, the edges in G surely appear in H (i.e., G is a sub-graph of H). Moreover, the closer two nodes are in G, they are are more likely to have opportunity in interacting with

[3]More complicated ways for creating H can be built using other network properties such as node degree, clustering coefficient, common number of neighbours, etc.

each other (i.e., to have an edge in opportunity network H). The parameter ξ controls the effect of distances on the edge probabilities in H. For $\xi = 0$, one can retrieve the original POSN models when the opportunity network H is a fully-connected network. The larger ξ is, the more sparse network H becomes. As ξ goes to infinity, H converges to G.

Once the opportunity network H is given (or created), we can compute the probability $\mathbb{P}(G|R^O, H, \boldsymbol{\theta})$ of graph structure G, given observed preferences R^O and the opportunity network H by

$$\begin{aligned}\mathbb{P}(G|H, R^O, \boldsymbol{\theta}) = &\prod_{i,j\in U;\, i<j} \Big(\mathcal{E}(\boldsymbol{\theta})^{e_{ij}} \left(1-\mathcal{E}(\boldsymbol{\theta})\right)^{1-e_{ij}}\Big)^{e^H_{ij}} \times \\ &\prod_{i\in O;\, j\in U} \Big(\mathcal{D}(r_i, \boldsymbol{\theta})^{e_{ij}} \left(1-\mathcal{D}(r_i, \boldsymbol{\theta})\right)^{1-e_{ij}}\Big)^{e^H_{ij}} \times \\ &\prod_{i,j\in O;\, i<j} \Big(c(r_i, r_j|\boldsymbol{\lambda})^{e_{ij}} \left(1-c(r_i, r_j|\boldsymbol{\lambda})\right)^{1-e_{ij}}\Big)^{e^H_{ij}}, \end{aligned} \tag{5.25}$$

where $e^H_{ij} = 1$ if and only if there is an edge between i and j in opportunity network H. Similarly, we formulate $\mathbb{P}(G|R^U, R^O, H)$ by

$$\mathbb{P}(G|R^O, R^U, H) = \prod_{\substack{i,j\in\mathcal{N}\\ i<j}} \Big(c(r_i, r_j)^{e_{ij}} \left(1-c\left(r_i, r_j\right)\right)^{1-e_{ij}}\Big)^{e^H_{ij}} \tag{5.26}$$

Using Bayes rule, the posterior over R^U given observed preferences R^O and network G is given by:

$$\mathbb{P}(R^U|G, R^O, H, \boldsymbol{\theta}) = \frac{\mathbb{P}(G|H, R^U, R^O)\mathbb{P}(R^U|\boldsymbol{\eta})}{\mathbb{P}(G|H, R^O, \boldsymbol{\theta})}, \tag{5.27}$$

Though this probabilistic framework is well-defined, it is not a generative probabilistic model as opposed to the original POSN. This is mainly because we have derived H from G.[4] However, we can still apply the sampling machinery that we developed in Sec. 5.4 by replacing Eq. 5.22 with

$$\tilde{p}(r_i) = \prod_{j\in U} \mathbb{P}(e_{ij}|r_i, r_j)^{e^H_{ij}} \prod_{j\in O} \mathbb{P}(e_{ij}|r_i, r_j)^{e^H_{ij}} \phi^{d_\tau(\sigma, r_i)}. \tag{5.28}$$

We study the performance and accuracy of this sampling method with its original form in our experiments below.

[4] One can extend this model to a generative process by assuming that first the opportunity network H is formed based on some latent variables (excluding rankings), then G is formed based on H and the nodal rankings in a similar fashion to POSN.

5.6 Empirical Analysis

We conduct experiments to assess the effectiveness of our inference and group recommendation algorithms. We measure the accuracy of preference inference, and more importantly, assess the quality of the group decisions reached when exploiting network structure to better deal with missing preferences of certain group members.

Experimental Setup. We experiment on three types of data sets: *two-sided synthetic data* in which both preferences and networks are randomly generated; *one-sided real-world data* in which preferences are derived from Irish electoral data[5], but networks are synthetically generated; and *two-sided real-world data* in which both preferences and network structure are derived from Flixster [191]. We assume the model parameters $\boldsymbol{\theta}$ are known (e.g., learned from a similar population). Unless otherwise noted, we set $(\alpha, \gamma, \beta) = (2, 0.7, 1)$ and $n = 200$. We use Borda as our scoring rule where $g(r(a), m) = m - r(a)$. While other rules can be used, Borda is a useful surrogate for random utility models [57] and serves to illustrate the value of the POSN model.

We vary the degree to which preferences are observed with parameter $\psi \in [0, 1]$, the probability that any node's ranking is observed.[6] By varying ψ, we can assess the impact of preference observability on the efficiency of our methods. We select the *decision making group* $\mathcal{S} \subseteq \mathcal{N}$ (with n_s members), for whom a group recommendation is to be made, using one of three methods. *RSA (Random Selection from All)* selects n_s individuals uniformly at random from $\mathcal{N}$. *RSU (Random Selection from Unobserved)* select n_s individuals uniformly at random from U (e.g., reflecting a company with access to a social network and the preferences of existing customers, and wanting to market to new prospects without knowing their preferences). *RSC (Random Selection from Community)* selects a connected community: it first selects a "seed" individual at random, then extends the group by selecting $n_s - 1$ friends of this seed at random; if this set is smaller than $n_s - 1$, friends of these friends are selected at random to complete the group.

Performance Metrics. To measure prediction accuracy made by our sampling algorithm, we determine how close inferred preferences are to the *true unobserved preferences* from a held out test set. We measure closeness using *mean scaled expected Kendall-τ (MSEK)*:

$$MSEK = \frac{1}{\binom{m}{2}|U|} \sum_{i \in U} \sum_{r \in \Omega(\mathcal{A})} \mathbb{P}(r|G, R^O, \boldsymbol{\theta}) d_\tau(r, \hat{r}_i), \tag{5.29}$$

where $\hat{r}_i$ is the true preference of i, $|U|$ is number of individuals with unobserved preferences, and $\binom{m}{2}$ is maximum τ-distance between two rankings over m options. MSEK lies in $[0, 1]$, with $MSEK = 0$ if all preferences are inferred correctly. In contrast, $MSEK = 1$ implies maximum

[5] We have obtained the original preference data sets from www.dublincountyreturningofficer.com.

[6] More sophisticated mechanisms for preference observability can be taken into account. For instance, one can consider the probability of observing an individual's preference depends on the structural properties of that individual in the underlying social network (e.g., degree or some other centrality measures)

"inaccuracy," meaning that all predicted rankings have the maximum possible distance to their true counterparts. By multiplying MSEK by the number of alternatives m, one can retrieve the expected number of "swaps" required to transform an inferred ranking to a *true* ranking. MSEK measures inference error in ranking space and is analogous to expected mean squared error (MSE) in Euclidean space.

To examine the decision/recommendation quality using inferred preferences, we compare its social welfare with that of *the decision that would be made had actual preferences been observed*. We measure the loss in social welfare that results from using inferred preferences. Let $sw(\cdot)$ denote social welfare with true preferences, and a^* and a^*_{inf} be the optimal options under given actual and inferred preferences. Rather than directly comparing social welfare, we define *relative social welfare loss (RSWL)* to be

$$[sw(a^*) - sw(a^*_{inf})]/sw(a^*)$$

(we report it as a percentage). RSWL, by scaling the difference in social welfare, makes the comparison across results of experiments easier especially when two experiments with various parameters impacting social welfare (e.g., m, or n) are compared.

Benchmarks. We consider several other ways of dealing with missing preferences in decision making, and use these as benchmarks. In *ϕ-Mallows inference (PM)*, we assume that all unobserved preferences are independent and are drawn from a ϕ-Mallows model (with parameters identical to those in the POSN model). We calculate the same inferred social welfare functions as in our model, namely, ES, JMPES, IMPES. Note that ES will be the same for all unobserved preferences and can be computed once by iterating over all permutations of alternatives. Moreover, JMPES and IMPES must be the same as the reference ranking σ. Another approach to missing preferences, dubbed *Discard Unobserved (DU)*, is to ignore them and make a decision using only observed preferences within decision making group as if some individuals "refused to vote."

For each fixed setting, we generate 10 partially observed POSNs. For each network, we burn-in 1000 samples, then collect 1000 samples using our Gibbs-Metropolis method (see Sec. 5.4). We report MSEK averaged over the 10 instances. For each instance, we also randomly select 40 decision making groups of fixed sizes $\{3, 5, 10, 15, 20\}$ using RSA, RSU, or RSC, giving 400 social choice instances per an experimental setting. RSWL is reported as the average over these 400 instances.

Two-sided synthetic data. For this set of experiments, we draw rankings independently from a ϕ-Mallows model and form the social network based on POSN model. Though this may not be an especially realistic approach to assess the ability of our prediction and decision making methods in real world settings, it plays important rule in checking the validity of our implemented algorithms. We set $\phi = 0.85$, $\sigma = (1, \ldots, m)$, and $\boldsymbol{\lambda}$ as stated above. Table 5.1 shows average MSEK for various ψ and m. Unsurprisingly, MSEK increases with m and

m / ψ	$\psi = 0.5$	$\psi = 0.6$	$\psi = 0.7$	$\psi = 0.8$
$m = 3$	0.000	0.000	0.000	0.000
$m = 4$	0.009	0.008	0.007	0.006
$m = 5$	0.168	0.158	0.152	0.148
$m = 6$	0.378	0.367	0.349	0.335

Table 5.1: Avg. MSEK (10 instances) for two-sided synthetic data, various m, ψ, but fixed $n = 200$, $(\alpha, \beta, \gamma) = (2, 0.7, 1)$.

n	200	400	600	800	1000
MSEK	0.335	0.2474	0.185	0.141	0.106

Table 5.2: Avg. MSEK (10 instances), $\psi = 0.8$, $m = 6$.

decreases with ψ. As m increases, the number of rankings increases factorially, as does the support of the ranking distribution. In such cases, lower MSEK requires more information for accurate prediction. When $m = 3$, $n = 200$ is sufficient to push MSEK to almost 0. With $m = 4$, it remains very low. To examine the effect of n on MSEK, we fix $m = 6$ and $\psi = 0.8$ but vary n: Table 5.2 shows that MSEK decreases with n as expected.

We now discuss decision quality of our methods. Fig. 5.2 shows average RSWL (over 400 instances) for various m, group selection methods, and inference methods. We set $\psi = 0.5$, which implies that, in expectation, half of all individual preferences are unobserved. Our POSN-ES and POSN-IMPES methods outperform the other benchmarks in most settings, including all situations in which no group member preferences are observed at all (see Fig. 5.2(b), 5.2(e), 5.2(h)), or even for $m = 6$ (see Fig. 5.2(g)-5.2(i)) with relatively high MSEK (see Table 5.1). RSWL in all of the benchmark methods, PM-ES, PM-JMPES, and DU, is very sensitive to group size, dramatically increasing as group size decreases (see Fig. 5.2(a)-(i)). However, our POSN-ES and POSN-IMPES methods are more robust to changes in group sizes, especially for $m = 4, 5$ (see Fig. 5.2(a)-(f)) with relatively low MSEK (ref. Table 5.1). POSN-IMPES approximates POSN-ES reasonably well, while POSN-JMPES is a reasonable approximation for $m = 4, 5$, but fails for $m = 6$; once again this might be due to its relatively high MSEK in our setting. In short, our ES and IMPES methods outperform the other benchmark methods in most settings, including over all group sizes, group selection methods, and various m (even for $m = 6$ with relatively high MSEK).

Irish data. We test our methods using real-world preferences from the 2002 Irish Election, Dublin West Constituency, with 9 candidates and 29,989 ballots of the top-t form, of which 3800 are complete rankings. We created preference data sets with various values m from these complete preferences, by choosing m candidates with highest aggregate Borda score, and limiting each individual's preferences to these m options.

For each m, we learn ϕ and σ from its corresponding data set and used those parameters in our methods (e.g. for $m = 4$, $\phi = 0.94$ and $\sigma = [4, 1, 2, 3]$; for $m = 5$, $\phi = 0.96$ and

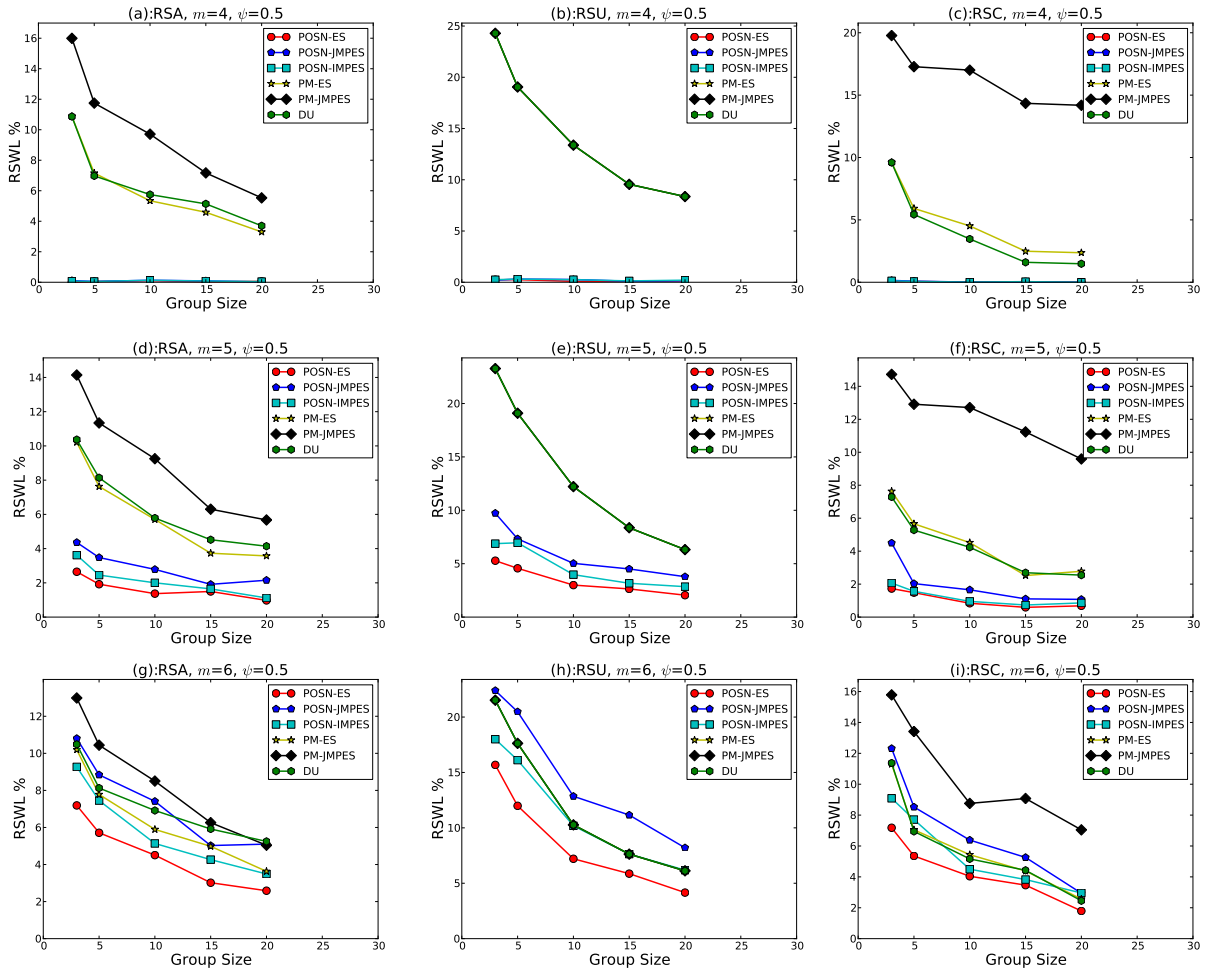


Figure 5.2: Avg. RSWL (over 400 instances) for various group sizes n_s, group selection methods, and m. POSN-ES, POSN-JMPES, and POSN-IMPES represent ES, JMPES, IMPES inferred social welfare under the POSN model, respectively. PM-ES and PM-JMPES correspond to ES and JMPES inferred social welfare under the ϕ-Mallows model. DU corresponds to our Discard Unobserved benchmark.

$\sigma = [5, 1, 2, 3, 4]$). Hence, we have a rough prior over preferences, but not a precise prior for specific group. For each experimental setting, we generate 10 partially observed POSNs with $\psi = 0.5$ and 200 individuals with preferences drawn from the filtered Irish data set. We then generate the POSN using our model, but with additional noise: we randomly change the parity of each e_{ij} (i.e., delete or add an edge) with probability ε. Though we create a synthetic social network using our POSN model, adding noise in this fashion reflects scenarios in which the social network is not generated using our specific model, or when learned model parameters provide a less-than-ideal fit to the underlying data. This tests the robustness of our methods when POSN parameters are inaccurate.

Table 5.3 reports average MSEK when ε varies ($m = 4, 5$). Unsurprisingly, MSEK increases with both m and ε when n and ψ are fixed. MSEK is very low when $m = 4$, even with high

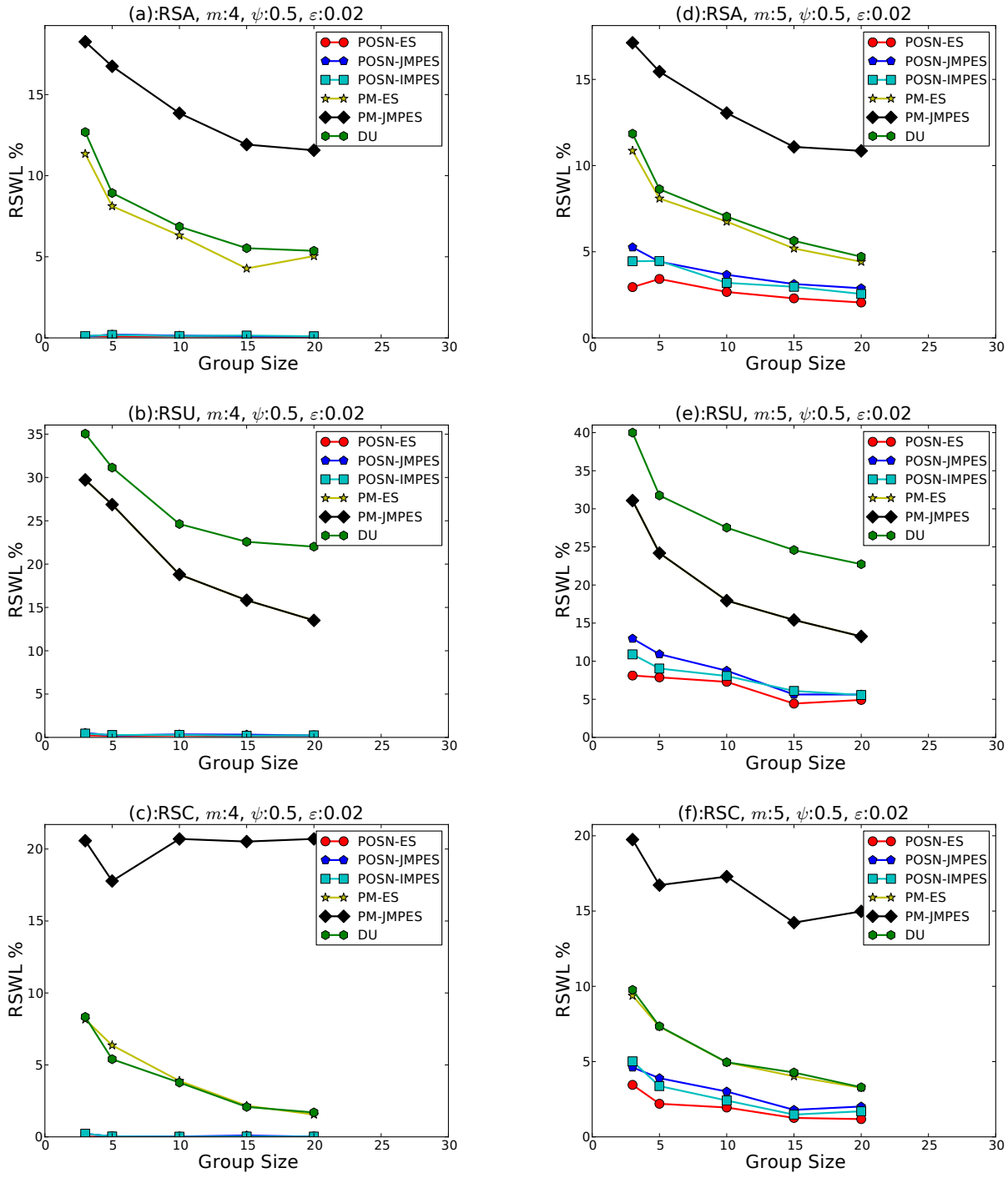


Figure 5.3: Avg. RSWL (over 400 instances) for one-sided synthetic data, various group sizes n_s, group selection methods, and m but fixed $\psi = 0.5$, $\varepsilon = 0.2$. POSN-ES, POSN-JMPES, and POSN-IMPES represent ES, JMPES, IMPES inferred social welfare under the POSN model, respectively. PM-ES and PM-JMPES correspond to ES and JMPES inferred social welfare methods under the ϕ-Mallows model. DU corresponds to our Discard Unobserved benchmark.

m / ε	$\varepsilon = 0.0$	$\varepsilon = 0.01$	$\varepsilon = 0.02$	$\varepsilon = 0.05$
$m = 4$	0.007	0.009	0.009	0.022
$m = 5$	0.172	0.198	0.214	0.274

Table 5.3: Avg. MSEK, $\psi = 0.5$, $n = 200$, Irish data set.

$\varepsilon = 0.05$ (10 edge flips per node in expectation). Tables 5.1 and 5.3 show comparable MSEK values for $m = 4, 5$, suggesting that that even in scenarios where the preference distribution $\rho(.)$ is not known *a priori*, but is a learned ϕ-Mallows model, POSNs support effective inference.

Fig. 5.3 shows average RSWL (over 400 instances) with $\psi = 0.5$ and $\varepsilon = 0.02$ (400 expected edge flips in the network). We vary m, the group selection method and the inference method. Our POSN-ES and POSN-IMPES approaches outperform the other benchmarks in most settings, including: all situations in which no group preferences are observed (see Fig. 5.3(b) and 5.3(e)); and even with $m = 5$ (see Fig. 5.3(d)-5.3(f)) despite its relatively high MSEK (see Table 5.3). RSWL in all benchmark methods (PM-ES, PM-JMPES, DU) is very sensitive to group size, increasing dramatically as group size decreases (see Fig. 5.3(a)-(f)). However, POSN-ES and POSN-IMPES are more robust to group size (see Fig. 5.3(a)-(f)). POSN-IMPES approximates POSN-ES reasonably well, while POSN-JMPES also performs well. These results suggest that social network structure can serve as vital evidence in social choice and group decision problems with partial preferences.

Flixster data. The Flixster dataset [191] consists of a social network of movie watchers and their ratings of movies, and allows a test of our methods using both real-world network and preference data. Because movie ratings are sparse, we aggregate them into preferences over movie genres (genres were determined automatically using the Rotten Tomatoes and IMDB web sites). Let $\tilde{r}_{um}$ be the rating of user u for movie m where $\tilde{r}_{um} \in \{0.5, 1, \cdots, 5\}$ if u has rated m, and is otherwise 0 (for missing ratings). For each user u and genre g, we define a *user-genre* score

$$S^C_{ug} = \frac{1}{\hat{I}_u} \sum_m \mathbb{1}(\tilde{r}_{um} > 0) A_{mg}, \tag{5.30}$$

where $\hat{I}_u = \sum_m \mathbb{1}(\tilde{r}_{um} > 0)$ is the number of movies rated by u, and $A_{mg} = 1$ if movie m has genre g (and $A_{mg} = 0$ otherwise). This score reflects the relative number of movies of each genre watched by a specific user. This is converted into a ranking of genres for each user u by ordering genres according to their scores S^C_{ug}. We also consider another method for extracting genre ranking by taking into account the actual values of movie ratings. For each user u and genre g, we define a *user-genre-rating* score

$$S^R_{ug} = \frac{1}{\bar{I}_u} \sum_m \tilde{r}_{um} A_{mg}, \tag{5.31}$$

where $\bar{I}_u = \sum_m \mathbb{1}(\tilde{r}_{um} > 0) A_{mg}$, and $A_{mg} = 1$ if movie m has genre g (and $A_{mg} = 0$ otherwise).

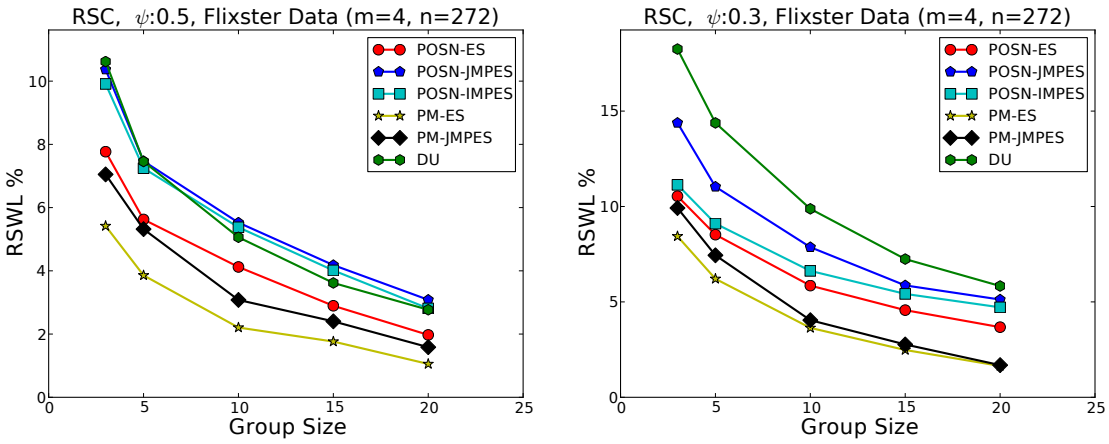


Figure 5.4: Avg. RSWL (400 instances), Flixster, RSC, user-genre-rating scores, ψ: $0.5, 0.3$. POSN-ES, POSN-JMPES, and POSN-IMPES represent ES, JMPES, IMPES inferred social welfare under the POSN model, respectively. PM-ES and PM-JMPES correspond to ES and JMPES inferred social welfare under the ϕ-Mallows model. DU corresponds to our Discard Unobserved benchmark.

This score reflects the average ratings of each genre watched by a specific user. This is converted into a ranking of genres for each user u by ordering genres according to their scores S^R_{ug}.

We limit our focus to four diverse genres—Kids/Family, Mystery/Suspense, Comedy, and Drama.[7] We run our methods on a 272-node subgraph of the Flixster data set, with 924 edges. We estimate a ϕ-Mallows model and POSN model parameters using maximum likelihood methods on this sub-network; the learned parameters are $(\alpha, \beta, \gamma, \phi) = (2.05, 1.06, 0.07, 0.33)$ for user-genre scores and $(\alpha, \beta, \gamma, \phi) = (1.30, 1.55, 0.07, 0.61)$ for user-genre-rating scores. For each run, we test our methods on 10 instances of a partially observed network, censoring each individual's genre preference with probability $\psi = 0.5$ or $\psi = 0.3$.

For user-genre-rating scores, the average MSEK is 0.416 and 0.415 for $\psi = 0.5$ and $\psi = 0.3$ (resp.). This suggests that genre-rating preferences are not reasonably predictable using the POSN model. Fig. 5.4 shows decision making performance, i.e., average RSWL, for the various methods described above using RSC to select groups for user-genre-rating scores and $\psi = 0.5$ or $\psi = 0.3$. Each of our POSN-sensitive methods—ES, IMPES, and JMPS—fail to outperform the ϕ-Mallows benchmark for all group sizes, but outperform DU.

The results for user-genre scores are by far more promising. The average MSEK is 0.242 and 0.256 for $\psi = 0.5$ and $\psi = 0.3$ (resp.). This suggests that genre preferences are reasonably predictable using the POSN model. Fig. 5.5 shows decision making performance, i.e., average RSWL, for the various methods described above using RSC to select groups. Each of our POSN-sensitive methods—ES, IMPES, and JMPS—outperform the ϕ-Mallows benchmark for all group sizes, and outperform DU significantly for small groups. DU performs comparably

[7] We focus on these four genres in part to increase data "density." Our choice of these genres may impact the results below; future investigation is needed to assess this impact.

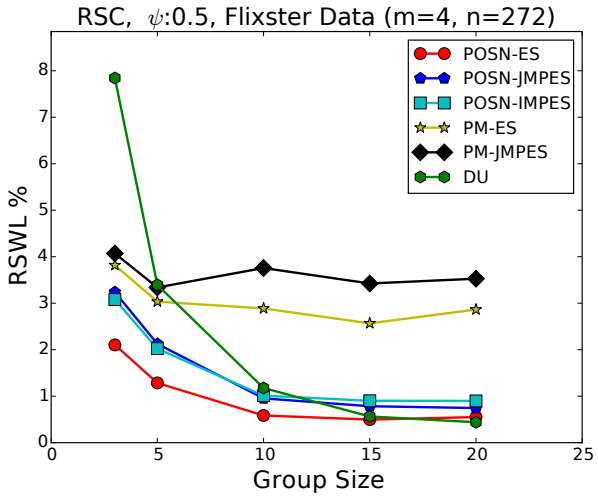


Figure 5.5: Avg. RSWL (400 instances), Flixster, RSC, user-genre scores, $\psi = 0.5$. POSN-ES, POSN-JMPES, and POSN-IMPES represent ES, JMPES, IMPES inferred social welfare under the POSN model, respectively. PM-ES and PM-JMPES correspond to ES and JMPES inferred social welfare under the ϕ-Mallows model. DU corresponds to our Discard Unobserved benchmark.

	n_s=3	n_s=5	n_s=10	n_s=15	n_s=20
POSN-ES	**7.34**	**5.04**	**2.86**	**2.57**	2.66
POSN-JMPES	9.89	7.67	4.33	3.69	3.29
POSN-IMPES	9.59	7.09	4.34	3.89	3.77
PM-ES	9.36	8.07	7.56	6.76	6.73
PM-JMPES	9.72	8.54	8.63	7.84	7.48
DU	20.61	12.07	5.31	2.92	**2.12**

Table 5.4: Std. RSWL in percentage (Flixster, ψ=0.5, user-genre scores)

to methods that account for network structure when groups are larger (15 or 20 individuals) since, in expectation, the preferences of 7–10 group members are observed: this is sufficient to make a good decision without estimating missing preferences explicitly due to normal sampling bounds from the underlying Mallows model. This, in addition to the fact that homophily across a large group makes it likely that the missing preferences are similar to those observed, means that making a group decision based only on observed preferences usually results in near-optimal decisions. Table 5.4 reports the std. dev. for these results when $\psi = 0.5$. ES has the smallest variance in RSWL in general, implying more robustness in the decisions made. Overall, ES is the most reliable method of those analyzed here. (Results for $\psi = 0.3$ are qualitatively similar as illustrated in Fig. 5.6.)

By comparing the results of user-genre-rating scores with those of user-genre scores, there are some interesting observations. Surprisingly, the POSN methods outperform the benchmarks for

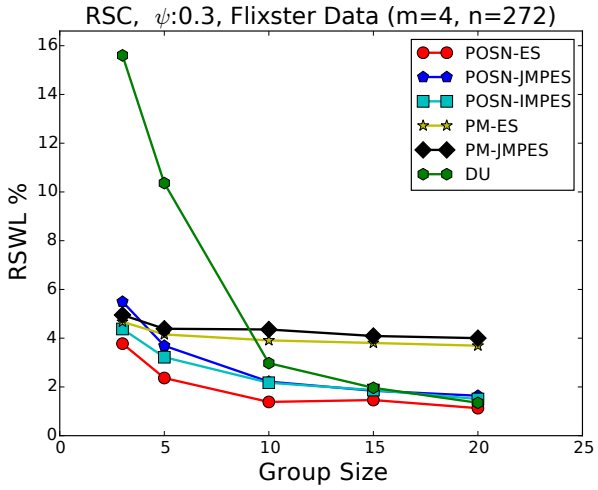


Figure 5.6: Avg. RSWL (400 instances), Flixster, RSC, user-genre scores, $\psi = 0.3$. POSN-ES, POSN-JMPES, and POSN-IMPES represent ES, JMPES, IMPES inferred social welfare under the POSN model, respectively. PM-ES and PM-JMPES correspond to ES and JMPES inferred social welfare under the ϕ-Mallows model. DU corresponds to our Discard Unobserved benchmark.

user-genre scores but not for user-genre-rating scores. This might suggest the user preferences are more correlated when the frequencies of genres are considered rather than ratings associated with those genres. This, in turn, might be indicative of behaviors in watching movies being more correlated than actual preferences are. In other words, the frequency with which a genre is watched by users might be more correlated over a social network than the extent to which the genre is preferred. Also, the failure of POSN for user-genre-rating scores suggests that further research is required to modify POSN model such that it can perform as well as benchmarks in the case of not ideal-fit to the data.

Speeding up Inference by Opportunity Networks. We examine the efficacy and performance of our group decision making using our sampling methods with opportunity networks discussed in Sec. 5.5. We use Flixster data with user-genre scores for our experiments and set preference observability $\psi = 0.5$. We vary opportunity network parameter ξ to study how the density of network opportunity affects the running time and the quality of group decisions. For $\xi = 0$, we recover our original POSN, sampling and decision making methods.

To compare the speed of our sampling methods for various ξ, we define the *relative running time improvement (RRTI)* for ξ to be the ratio of the average running time of original POSN model to the average running time using network opportunity with ξ. RRTI shows the magnitude that specific ξ make sampling and group decision making faster than that of the original model. Table 5.5 shows RRTI for various ξ. Unsurprisingly, $RRTI$ increases with ξ. Now, the important question is to understand the effect of ξ on the quality of group decisions. Fig. 5.7 shows the average relative social welfare loss for various ξ. We can observe that by increasing

ξ	0	0.5	1	2	3	5
RRTI	1.000	1.744	2.620	6.348	9.549	15.981

Table 5.5: Running time ratio (10 instances) with opportunity network, $\psi = 0.5$, Flixster Data, user-genre score, $n = 272$, and $m = 4$.

ξ, RSWL (specifically for POSN-ES and POSN-IMPES and large group sizes) first decreases (e.g. compare $\xi = 0$–2) and then increases (e.g., compare $\xi = 2$–5). RSWL for $\xi = 2$ are lowest. Not only does $\xi = 2$ result in better group decisions, but also it is 6.5 times faster than the original model for this network.[8] These results have an important message: the quality of group decisions can be improved by sub-sampling negative examples but the number of sub-samples should not be too large or too small to maximize prediction accuracy.

Through these experiments, we observe that our proposed opportunity network methods not only have made our inference tasks very faster but also help improving the quality of our group decisions. This approach seems to be a win-win approach but further explorations are required for a definitive conclusion.

5.7 Summary and Future Work

This chapter introduced preference-oriented social networks (POSNs) to capture the correlation of preference rankings between individuals who interact in social networks. We developed effective inference methods to predict an individual's preferences by exploiting these correlations. We also developed methods for group recommendation when the preferences of some (or even all) group members are unobserved. Our experiments showed the value of accounting for social ties in inference and group recommendation when faced with missing preferences.

This work is a starting point for the deeper modeling of preferences in a social network context. Interesting future directions include: more efficient sampling methods based on network topology; studying other aggregation functions (e.g., other social choice functions, voting rules, bargaining solution concepts, etc.), and extensions to other social choice problems (e.g., matchings, assignments). We elaborate more about these future directions in Sec. 6.3.

Of practical importance is investigating the extent to which preference rankings are correlated and play a role in shaping connections in real-world social networks. Developing scalable methods for learning model parameters is essential; such learning techniques can exploit our inference methods as important building block (e.g., in EM-based algorithms). Our model can provide the basis for more effective preference elicitation.

As decision making using MPE seems to provide a reasonable approximation to optimal decisions, studying how MPE can be computed or approximated without the use of sampling remains of interest. Similar to active learning methods [40], the tighter integration of inference

[8]The speed gain would be much higher for larger social networks.

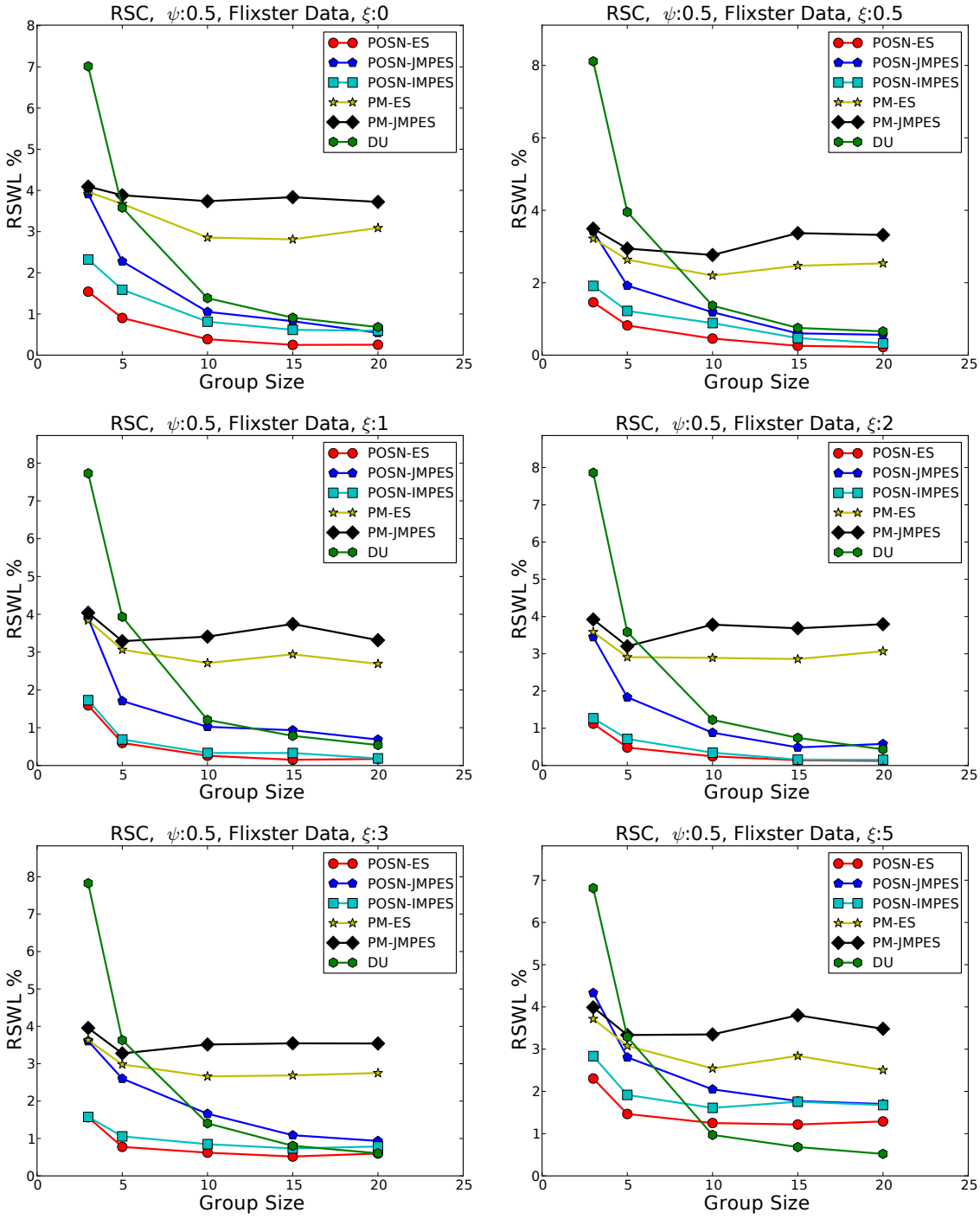


Figure 5.7: Avg. RSWL (over 400 instances) for various group sizes n_s, various ξ, the RSC group selection method, Flixster data $\psi = 0.5$. POSN-ES, POSN-JMPES, and POSN-IMPES represent ES, JMPES, IMPES inferred social welfare under the POSN model, respectively. PM-ES and PM-JMPES correspond to ES and JMPES inferred social welfare under the ϕ-Mallows model. DU corresponds to our Discard Unobserved benchmark.

and decision making methods would also be of value.

There are a number of potential extension to our POSN model. This includes accommodating partial information about the preferences of specific users (e.g., a small set of pairwise comparisons); and incorporating both the strength and types of relationships between individuals. Such generalizations may offer greater performance in certain preference inference and group recommendation settings. We discuss in detail some of these future directions in Sec. 6.3.

Chapter 6

Conclusion and Future Work

Group decision making (social choice) problems are prevalent in our day-to-day lives. Many applications that deal with user preferences (e.g., recommender systems, advertisement and marketing mechanisms, etc.) have some sort of social choice problem at their core. One of the main endeavours in any group decision making problem is learning the individual preferences upon which decisions or recommendations are made.

This thesis hypothesised that social networks–by capturing preference correlations over individuals induced by social interactions–provide a natural and informative platform for preference learning. Some specific questions addressed through this thesis: i) What dynamics dictate how individual preference become correlated in a social network over time (see Chapter 3)? (ii) How should one mathematically model such correlations and dynamics (see Chapter 3 and Chapter 4)? (iii) Can such correlations and dynamics be harnessed for more efficient preference learning and elicitation, and consequently more effective group decision making? (see Chapter 3 and Chapter 5).

Our theoretical and empirical analyses confirmed our hypothesis (discussed above) and suggest that social networks can play important role in group decision making by capturing the interdependency of individual preferences. The thesis advanced the understanding and mathematical modelling of preference dynamics and correlations over social networks and to exploit the computational and predictive power of these models to develop efficient algorithms for decision making and recommendations, while requiring less user data, and imposing lower cognitive and communication burden on users. We review the summary of main results, specific contributions of each chapter, and the future directions below.

6.1 Summary of Main Results

We summarize the main results of each chapter.

Chapter 3: Empathetic Social Choice on Social Networks. We have introduced empathetic social choice frameworks in which individuals derive utility based on both their own

intrinsic preferences and empathetic preferences, determined by the satisfaction of their acquaintances (e.g., friends voting for a vacation spot or a movie, while considering both their own and others' satisfaction). We first studied the theoretical conditions under which empathetic preferences are well-defined (i.e., converge to a fixed-point). We described (mild) conditions—*normalization*, *non-negativity*, and *positive self-loop*—under which such fixed points exist. We demonstrated that group decision making in the empathetic framework can be recast as a form of weighted voting. We then developed two scalable algorithms for consensus decision making (or group recommendation). We also generalized our empathetic framework to accommodate other social choice problems (e.g., assignment, matching, etc.) and showed how some of our theoretical results (e.g., fixed-point convergence) still hold. We also demonstrated that, in our general empathetic framework, certain other social choice problems can be viewed as their corresponding weighted versions, when weights are determined from social network structure. We developed a scalable iterative method for estimating those societal weights, which can serve as a building block for solving other social choice problems. We empirically demonstrated the value of accounting for empathetic preferences and the performance of our algorithms. Our theoretical and empirical results shed light on how individual preferences become correlated due to the presence of empathy. The results also confirmed that neglecting empathic preferences yields sub-optimal group decisions.

Chapter 4: Ranking Networks. To capture the correlation of preference rankings (or in general rank data) on social networks, we have introduced a network formation model called ranking networks in which the similarity of two individuals' rankings determines the chance they are connected to each other. We theoretically analysed general topological properties of this model, demonstrating that it exhibits some observed properties of real-work social networks such as a small diameter and the existence of a giant connected component. We also derived some closed-form formulas for estimating degree distribution, edge density, and clustering coefficients under this model, and showed that their computations are expensive. Thus, we developed easy-to-compute approximations for the special class of distance-based ranking models, while studying properties that emerge among networks in this class. We also demonstrated how these approximations can be exploited for efficient model learning. Through empirical experiments, we demonstrated the effectiveness of our approximation and learning methods.

Chapter 5: Group Recommendation on Preference-Oriented Social Networks. We introduced a special instance of ranking networks, that we call preference-oriented social networks (POSNs), for capturing the correlation of preferences over social networks. We characterized important probabilistic properties of the model such as, the joint posterior distribution of unobserved preferences given observed networks structure and some individuals' observed preferences. We then formulated the social choice/group decision making problem with missing preferences in a social network. We assumed that we need to select an option that implements a social choice function for a group or "subpopulation". We distinguished a subpopulation (e.g., a group of friends planning an activity, or electorate in a small district) from the larger

society (e.g., the users of an online social network in a university, or the eligible voters in a country) to reflect the fact that many group decisions are local, but can be supported by knowledge of the preferences of individuals outside that group. We developed a Markov Chain Monte Carlo (MCMC) method for inferring unobserved preferences and making social choices for sub-populations. We also proposed some enhancements to our inference and group decision making methods which improve both the speed and efficacy of our original methods. These enhancements build upon on a simple and seemingly natural assumption that not everyone has the opportunity to be acquainted with everyone else in a society. We studied various types of group selection strategies ranging from uniformly randomly selected subsets of individuals to those which choose a subset of individuals using network structure. In our experiments, we evaluated the ability of our methods to predict unobserved preferences and to support effective group decision making with partial preferences (or even in cases where preferences of group members are completely unknown). Using various datasets (e.g., Flixster, Irish election, etc.), we compared our group recommendation methods to different benchmarks which neglect the information contained in the social network. Our empirical results demonstrate that accounting for social ties can significantly improve predictions and group recommendation when faced with missing preferences.

6.2 Thesis Contributions

We summarize this thesis's contributions in point form below.

Chapter 3:

- Introduce and mathematically model the *empathetic framework*.
- Theoretically analyse the empathetic framework:
  - Prove conditions under which utilities are well-defined.
  - Prove that social choice in the empathetic framework can be viewed as a weighted voting where weights come from underlying social network.
- Propose two scalable iterative algorithms (ICE and WICE) for consensus decision making:
  - Prove their correctness, termination, and convergence rates.
- Generalize our empathetic framework to accommodate other social choice problems.
- Develop an iterative method for approximating societal weights, which can serve as building blocks of algorithms for solving other social choice problems.
- Prove the distinction between empathetic models and allocative externalities models.
- Run empirical experiments:

  - Demonstrate the importance of considering empathy in group decision making problems.
  - Evaluate the performance of the proposed algorithms.

Chapter 4:

- Introduce and mathematically model the generative *ranking network* model.
- Theoretically analyze general topological properties of this model, demonstrating that it exhibits some observed properties of real-work networks:
  - Prove the existence of the small-world phenomena.
  - Prove the existence of a giant connected component.
  - Prove the shrinking diameters property.
- Derive closed-form formulae for estimating degree distribution, edge density, and clustering coefficients.
  - Analyze the running time of these exact estimations and shown that their computations are expensive.
- Analyze and studied the special class of *distance-based ranking models*.
  - Develop easy-to-compute approximations for edge density and degree distribution.
  - Study specific properties that emerge among networks in this class; e.g., the relation of social popularity and preference popularity.
- Formulate maximum likelihood estimation in distance-based ranking models and propose efficient approximation methods for its computation.
- Empirically demonstrate the effectiveness of our approximation methods.

Chapter 5:

- Introduce *Preference-Oriented Social Network* (POSN) model a special instance of ranking networks.
- Exploit POSN to infer unobserved individual preferences given observed preferences of others in the social network.
- Formulate the target posterior distributions.
- Develop a Markov Chain Monte Carlo (MCMC) method for inferring unobserved preferences.
- Develop various group decision making processes under inferred preferences.

- Introduce the concept of "opportunity network" for speeding up the inference and group decision making.
- Run empirical experiments on various datasets:
  - Evaluate the ability of our methods to predict unobserved preferences.
  - Demonstrate that accounting for social ties can significantly improve predictions and group decision making/recommendation.

6.3 Future Directions

We highlight the possible directions of future research, mostly built on or continuation of this thesis with the focus of exploiting social ties for more efficient and effective group decision making, recommendation, and preference learning.

Empathetic Framework and Other Social Choice Problems One can apply our empathetic social choice framework to other social choice problems, such as matching, assignment, and multi-winner election problems. We have shown in Chapter 3 that social welfare can be written as weighted intrinsic utilities under both global and local empathetic models; this observation can be starting point for applying our framework to other social choice problems. Of special interest is exploiting empathetic framework for multi-winner elections problems. Examples of multi-winner elections with empathetic preferences are prevalent in real-world: when a city council decides to implement, say, two of a number of proposals for the use of vacant land, or when a social networking site decides to implement, say, three new functionalities from many possible additions.

Qualitative Preference Formation and Aggregation on Social Networks. One can generalize the empathetic social choice framework in Chapter 3 by considering scenarios in which individuals repeatedly update their own preference rankings by aggregating their own and their neighbours' preferences (any aggregation mechanism is applicable as long as it minimizes some notion of distance between aggregated preference and local preference profiles). This *local aggregation process* of preferences might capture various psychological dynamics on social networks including empathy, confrontation, influence, imitation, etc. Nonetheless, we believe that local aggregation is in heart of all these phenomena, since all require individual preferences become more similar to that of (a subset of) their friends over time. Similarly, a local aggregation process that iteratively minimizes the distance of an individual's ranking to those of her friends makes an individual's ranking more similar to those of her friends. As such, it is essential to study local aggregation by abstracting away other minor differences. Under this framework, one can study ranking preference formation and its interplay with social structure on social networks. Of special interest is how social structure impacts the formation of correlated preference rankings on social networks, when preferences converge, and how community structures diversify the resulting preferences.

Extensions to Ranking Networks. There are a number of potential extensions to our ranking network model of Chapter 4. This includes accommodating partial information about the preferences of specific users (e.g., a small set of pairwise comparisons); incorporating both the strength and types of relationships between individuals; considering heterogeneity/heterophily; and incorporating the other nodal attributes in addition to rankings in the process of network formation. Such generalizations may offer greater performance in certain preference inference and group recommendation settings.

Correlations of Rankings in Social Networks. In Chapter 5, we evaluated the performance of POSN using real-world data from Flixster, where movie ratings were mapped to some rankings over genres. However, it is interesting to explore the performance of POSN on preference rankings collected/elicited directly from users over a social network. (Unfortunately, a social network data set which includes elicited preference rankings from users was not been available when this thesis research conducted.) There are some scientifically interesting questions that one can explore in this regard. For example, to what extent and for which contexts are ranking preferences correlated in social networks? Which ranking distance metrics are most suitable to be used in POSNs? Is it possible that ranking preferences demonstrate various levels of correlation, determined by the strength of social ties (e.g., preferences are more correlated between couples or siblings than between classmates)?

Scalable Learning Methods for POSN. Of practical importance are scalable methods for learning POSN parameters. We developed efficient methods for model learning by optimizing the approximate likelihood function, drawing upon the approximation methods that we developed for estimating topological properties of ranking networks in Chapter 4. One can take another direction for learning model parameter by deploying our MCMC sampling methods developed in Chapter 5 within an EM algorithm. However, this approach requires some modifications to our sampling methods to make it more efficient and scalable.

Efficient Sampling Methods for POSN. One interesting direction is to design more efficient sampling methods by relaxing (or neglecting) some dependencies in the joint posterior distribution of POSNs. In Chapter 5, the original POSN model considers the absence and presence of all possible edges when computing the joint posterior distribution, thus yielding a sampling method that is very intensive for large social networks. To make computation more efficient, we proposed the concept "opportunity network" which takes into account the presence and absence of a subset of possible edges. However, we empirically studied only a very specific instance of opportunity networks which relies on geodesic distances in observed networks. There are more interesting variations of opportunity networks can be studied in future work in which other network properties (e.g., the degree distribution, clustering coefficient, etc.) are taken into account for its creation. The goal should be to produce sparser opportunity networks which can scale well for large-scale social networks but still result in high quality predictions and group decisions.

Appendices

Appendix A

Theoretical Background and Proofs of Chapter 3

This appendix first reviews the relevant definitions and results used for our theoretical analyses in Chapter 3. We then discuss the relevant lemmas and proofs.

A.1 Linear Algebra Background

We first provide relevant definitions and results used in our proofs.

Definition (**Spectrum $\sigma(\mathbf{A})$**)**.** The set of eigenvalues of an $n \times n$ matrix $\mathbf{A}$ is called its *spectrum* $\sigma(\mathbf{A})$.

Definition (**Spectral Radius $\rho(\mathbf{A})$**)**.** Let $\mathbf{A}$ be an $n \times n$ matrix with real or complex eigenvalues $\sigma(\mathbf{A})$. Then the *spectral radius* of $\mathbf{A}$ is

$$\rho(\mathbf{A}) = \max_{\lambda \in \sigma(\mathbf{A})} |\lambda|$$

.

Definition (**M-matrix**). A matrix $\mathbf{A}$ in the form of $\mathbf{A} = s\mathbf{I} - \mathbf{B}$ is an *M-matrix* if $s \geq \rho(\mathbf{B})$ and $\mathbf{B} \geq \mathbf{0}$.

Proposition A.1.1 (**Nonsingular M-matrix** [268])**.** *If $s > \rho(\mathbf{B})$ in an M-matrix $\mathbf{A} = s\mathbf{I} - \mathbf{B}$, then $\mathbf{A}$ is nonsingular and $\mathbf{A}^{-1} \geq \mathbf{0}$*

Note that an M-matrix can be either singular or non-singular. Therefore, the condition $s > \rho(\mathbf{B})$ in Prop. A.1.1 is necessary to guarantee the nonsingularity of an M-matrix.[1]

[1] In some references (e.g.,[268]), an M-matrix is defined with $s > \rho(\mathbf{B})$. By this definition, an M-matrix is non-singular.

Theorem A.1.2 (Gerschgorin Circles [268]). *The eigenvalues of matrix* $\mathbf{A} \in \mathcal{C}^{n\times n}$ *are contained in* $\cup_{i=1}^{n}\mathcal{G}_i$*, where* $\mathcal{G}_i$ *is the Gerschgorin circle defined by:*

$$\mathcal{G}_i = \left\{c \in \mathcal{C} \middle| |c - a_{ii}| \leq R_i\right\} \text{ where } R_i = \sum_{\substack{0\leq j\leq n \\ j\neq i}} |a_{ij}|$$

We exploit induced matrix norms in our analysis of convergence rate of our iterative method for fixed-point utilities. For a given vector norm $\|.\|$, the induced norm for $n \times m$ matrix $\mathbf{A} \in \mathcal{C}^{n\times m}$ is:

$$\begin{aligned}\|\mathbf{A}\| &= \max\{\|\mathbf{A}\mathbf{x}\| : \mathbf{x} \in \mathcal{C}^m \text{ and } \|\mathbf{x}\| = 1\} \\ &= \max\left\{\frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|} : \mathbf{x} \in \mathcal{C}^m \text{ and } \mathbf{x} \neq \mathbf{0}\right\}\end{aligned}$$

We here focus on the p-norm $\|.\|_p$ which is induced by the p-norm in vector spaces. More precisely, the p-norm of matrix $\mathbf{A}$ is

$$\|\mathbf{A}\|_p = \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|_p}{\|\mathbf{x}\|_p}$$

where the p-norm $\|\mathbf{x}\|_p$ of vector $\mathbf{x} \in \mathcal{C}^n$ is:

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}$$

p-norms have several important properties: (1) they are *submultiplicative*: $\|\mathbf{A}\mathbf{B}\|_p \leq \|\mathbf{A}\|_p\|\mathbf{B}\|_p$. A consequence of this consistency property is that, for any square matrix $\mathbf{A}$, $\|\mathbf{A}^k\|_p \leq \|\mathbf{A}\|_p^k$. (2) By definition, they are *compatible*: $\|\mathbf{A}\mathbf{x}\|_p \leq \|\mathbf{A}\|_p\|\mathbf{x}\|_p$ where $\mathbf{A} \in \mathcal{C}^{n\times m}$ and $\mathbf{x} \in \mathcal{C}^m$.

For the cases where $p = 1$ or $p = \infty$, the matrix p-norm can be computed easily. The 1-norm for matrix $\mathbf{A}$ is simply the maximum absolute column sum of $\mathbf{A}$:

$$\|\mathbf{A}\|_1 = \max_{1\leq j\leq n} \sum_{i=1}^{m} |a_{ij}| \tag{A.1}$$

The ∞-norm for matrix $\mathbf{A}$ is simply the maximum absolute row sum of $\mathbf{A}$:

$$\|\mathbf{A}\|_\infty = \max_{1\leq i\leq m} \sum_{j=1}^{n} |a_{ij}|. \tag{A.2}$$

We review the Jacobi iterative method and its convergence criteria and rate. Iterative methods offer practical advantages for solving linear systems [310]. A linear system is formally defined as follows: Given an $n \times n$ real-valued matrix $\mathbf{A}$ and a real n-vector $\mathbf{b}$, the problem is to find n-vector $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{A}\mathbf{x} = \mathbf{b}$.

The Jacobi method [310] is an iterative method for solving linear systems. Consider this decomposition $\mathbf{A} = \mathbf{\Lambda} - \mathbf{E} - \mathbf{F}$ where $\mathbf{\Lambda}$ is the diagonal matrix of $\mathbf{A}$, $\mathbf{E}$ is the strictly lower triangular matrix of $-\mathbf{A}$, and $\mathbf{F}$ is the strictly upper triangular matrix of $-\mathbf{A}$. Note that we assume that the diagonal entries of $\mathbf{A}$ are all non-zero (this corresponds to our positive self-loop assumption below). Each iteration of the Jacobi method takes the form of:

$$\mathbf{x}^{(t+1)} = \mathbf{\Lambda}^{-1}(\mathbf{E} + \mathbf{F})\mathbf{x}^{(t)} + \mathbf{\Lambda}^{-1}\mathbf{b} \tag{A.3}$$

Theorem A.1.3 (Convergence of Iterative Methods [310])**.** *Let an iterative method take the form of* $\mathbf{x}_{t+1} = \mathbf{G}\mathbf{x}_t + \mathbf{f}$ *where* $\mathbf{G}$ *is an* $n \times n$ *iteration matrix and* $\mathbf{f}$ *is an n-vector. It converges if and only if* $\rho(\mathbf{G}) < 1$*.*

Corollary A.1.1 (Jacobi Convergence)**.** *The Jacobi iterative method converges to the solution of linear system* $\mathbf{A}\mathbf{x} = \mathbf{b}$ *if* $\rho(\mathbf{G}) < 1$ *where* $\mathbf{G} = \mathbf{\Lambda}^{-1}(\mathbf{E} + \mathbf{F})$*.*

Proof The proof of convergence is trivial and immediately follows form the Theorem A.1.3 by letting $\mathbf{G} = \mathbf{\Lambda}^{-1}(\mathbf{E} + \mathbf{F})$ and $\mathbf{f} = \mathbf{\Lambda}\mathbf{b}$. Now, we prove that the Jacobi method converges to the solution of the linear system. Since it converges, let $\mathbf{x}^* = \lim_{t\to\infty} \mathbf{x}^{(t)}$. From Equation A.3, we have:

$$\begin{aligned}
&\lim_{t\to\infty} \mathbf{x}^{(t+1)} = \lim_{t\to\infty} \mathbf{\Lambda}^{-1}(\mathbf{E} + \mathbf{F})\mathbf{x}^{(t)} + \mathbf{\Lambda}^{-1}\mathbf{b} \\
\implies &\lim_{t\to\infty} \mathbf{x}^{(t+1)} = \mathbf{\Lambda}^{-1}(\mathbf{E} + \mathbf{F}) \left(\lim_{t\to\infty} \mathbf{x}^{(t)}\right) + \mathbf{\Lambda}^{-1}\mathbf{b} \\
\implies &\mathbf{x}^* = \mathbf{\Lambda}^{-1}(\mathbf{E} + \mathbf{F})\mathbf{x}^* + \mathbf{\Lambda}^{-1}\mathbf{b} \\
\implies &\mathbf{\Lambda}\mathbf{x}^* = (\mathbf{E} + \mathbf{F})\mathbf{x}^* + \mathbf{b} \\
\implies &(\mathbf{\Lambda} - \mathbf{E} - \mathbf{F})\mathbf{x}^* = \mathbf{b} \\
\implies &\mathbf{A}\mathbf{x}^* = \mathbf{b}
\end{aligned}$$

So $\mathbf{x}^*$ is the solution of the linear system. ∎

A.2 Proofs

Recall that the adjacency matrix is denoted by $\mathbf{W} = [w_{ij}]$ and $\mathbf{D}$ is $n \times n$ diagonal matrix with $d_{jj} = w_{jj}$. To prove Prop. 3.1.1, we first show

Lemma A.2.1. *Assuming nonnegativity, normalizaton, and positive self-loop,* $\rho(\mathbf{B}) < 1$ *where* $\mathbf{B} = \mathbf{W} - \mathbf{D}$*.*

Proof By the definition of $\mathbf{W}$ and $\mathbf{D}$, it can be seen that $\mathbf{B} = \mathbf{W} - \mathbf{D}$ is a matrix with $b_{ii} = 0$ and $b_{ij} = w_{ij}$ for all $i, j \in \mathcal{N}$ and $i \neq j$. Using the Gerschgorin Circle Theorem (Thm. A.1.2), we have $\sigma(\mathbf{B}) \subset \cup_{i=1}^{n} \mathcal{G}_i$ where

$$\mathcal{G}_i = \big\{c \in \mathcal{C} \big| |c - b_{ii}| \leq R_i\big\} \text{ and } R_i = \sum_{\substack{0 \leq j \leq n \\ j \neq i}} |b_{ij}|.$$

As $b_{ii} = 0$ and $b_{ij} = w_{ij}$ for $i \neq j$, we have:

$$\mathcal{G}_i = \big\{c \in \mathcal{C} \big| |c| \leq R_i\big\} \text{ where } R_i = \sum_{\substack{0 \leq j \leq n \\ j \neq i}} |w_{ij}|.$$

Note that each $\mathcal{G}_i$ is a closed disk in $\mathcal{C}$ which is centered at 0. So $\cup_{i=1}^{n} \mathcal{G}_i$ is the union of closed disks of various radii but the same center 0. Since the number of these disks is finite, we can cover all these closed disks with a closed covering disk defined by $\big\{c \in \mathcal{C} \big| |c| \leq R_{max}\big\}$ where $R_{max} = \max_{i=1}^{n} R_i$. Without loss of generality, let $l = \arg\max_i R_i$. So, we have

$$\sigma(\mathbf{B}) \subset \cup_{i=1}^{n} \big\{c \in \mathcal{C} \big| |c| \leq R_i\big\} \subseteq \big\{c \in \mathcal{C} \big| |c| \leq R_l\big\}$$

From this, it follows that:

$$|\lambda| \leq R_l, \forall \lambda \in \sigma(\mathbf{B}) \implies \max_{\lambda \in \sigma(\mathbf{B})} |\lambda| \leq R_l \implies \rho(\mathbf{B}) \leq R_l.$$

Using $R_l = \sum_{j \neq l} |w_{lj}|$ and the normalization assumption $\sum_j w_{lj} = 1$, we have $\rho(\mathbf{B}) \leq 1 - w_{ll}$. Since $w_{ll} > 0$ by self-loop positivity, we have $\rho(\mathbf{B}) \leq 1 - w_{ll} < 1$. ∎

Proposition 3.1.1. *Assuming nonnegativity, normalizaton, and positive self-loop, Eq. 3.6 has a unique fixed-point solution* $\mathbf{u}(a) = (\mathbf{I} - \mathbf{W} + \mathbf{D})^{-1}\mathbf{D}\mathbf{u}^I(a)$.

Proof of Prop. 3.1.1 Using Eq. 3.4, we can write:

$$\begin{aligned}
&\mathbf{u}(a) = (\mathbf{W} - \mathbf{D})\mathbf{u}(a) + \mathbf{D}\mathbf{u}^I(a) \\
&\implies \mathbf{u}(a) - (\mathbf{W} - \mathbf{D})\mathbf{u}(a) = \mathbf{D}\mathbf{u}^I(a) \\
&\implies (\mathbf{I} - (\mathbf{W} - \mathbf{D}))\mathbf{u}(a) = \mathbf{D}\mathbf{u}^I(a)
\end{aligned}$$

So it is sufficient to show that $(\mathbf{I} - (\mathbf{W} - \mathbf{D}))^{-1}$ exists to prove that $\mathbf{u}(a) = (\mathbf{I} - \mathbf{W} + \mathbf{D})^{-1}\mathbf{D}\mathbf{u}(a)$ exists and is unique. We need to show that $\mathbf{I} - (\mathbf{W} - \mathbf{D})$ is nonsingular to guarantee the existence of $(\mathbf{I} - (\mathbf{W} - \mathbf{D}))^{-1}$.

Let $\mathbf{B} = \mathbf{W} - \mathbf{D}$. By definitions of $\mathbf{W}$ and $\mathbf{D}$, the matrix $\mathbf{B}$ has $b_{ii} = 0$ and $b_{ij} = w_{ij}$ for all $i, j \in \mathcal{N}$ and $i \neq j$. By nonnegativity assumption, we have $w_{ij} \geq 0$, so $\mathbf{B} \geq \mathbf{0}$. By setting $s = 1$, $(\mathbf{I} - (\mathbf{W} - \mathbf{D})) = (s\mathbf{I} - \mathbf{B})$ which is an M-matrix (See Definition A.1). Using Lemma A.2.1, we have $\rho(\mathbf{B}) < 1$. Since $s = 1$, then $\rho(\mathbf{B}) < s$. By Proposition A.1.1, it follows that $(\mathbf{I} - (\mathbf{W} - \mathbf{D}))$ is nonsingular and $(\mathbf{I} - (\mathbf{W} - \mathbf{D}))^{-1} \geq \mathbf{0}$. ∎

Corollary 3.1.1. *In the global empathetic model, (global) social welfare of alternative a is given by* $sw_g(a, \mathbf{u}^I) = \boldsymbol{\omega}^\top \mathbf{u}^I$ *where* $\boldsymbol{\omega}^\top = \mathbf{e}^\top(\mathbf{I} - \mathbf{W} + \mathbf{D})^{-1}\mathbf{D}$.

Proof of Cor. 3.1.1 The proof is straightforward. From the definition of social welfare and Prop. 3.1.1, it follows

$$sw_g(a, \mathbf{u}^I) = \mathbf{e}^\top (\mathbf{I} - \mathbf{W} + \mathbf{D})^{-1} \mathbf{D} \mathbf{u}^I(a).$$

By setting $\boldsymbol{\omega}^\top = \mathbf{e}^\top (\mathbf{I} - \mathbf{W} + \mathbf{D})^{-1} \mathbf{D}$, we have $sw_g(a, \mathbf{u}^I) = \boldsymbol{\omega}^\top \mathbf{u}^I$. ∎

Theorem 3.2.1. *Consider the following iteration:*

$$\mathbf{u}^{(t+1)}(a) = (\mathbf{W} - \mathbf{D})\mathbf{u}^{(t)}(a) + \mathbf{D}\mathbf{u}^I(a). \tag{A.4}$$

Assuming nonnegativity, normalizaton, and positive self-loop, this method converges to $\mathbf{u}(a)$*, the solution to Eq. 3.6*

Proof of Theorem 3.2.1 From Eq. 3.4, we observe that $\mathbf{u}(a)$ is the solution of the linear system $\mathbf{A}\mathbf{u}(a) = \mathbf{b}$ with $\mathbf{A} = \mathbf{I} - (\mathbf{W} - \mathbf{D})$ and $\mathbf{b} = \mathbf{D}\mathbf{u}^I(a)$. The Jacobi method is:

$$\mathbf{u}(a)^{(t+1)} = \boldsymbol{\Lambda}^{-1}(\mathbf{E} + \mathbf{F})\mathbf{u}(a)^{(t)} + \boldsymbol{\Lambda}^{-1}\mathbf{b}$$

Since $\mathbf{A} = \mathbf{I} - (\mathbf{W} - \mathbf{D})$, we have that $\boldsymbol{\Lambda} = \mathbf{I}$ and $\mathbf{E} + \mathbf{F} = \mathbf{W} - \mathbf{D}$. As $\mathbf{b} = \mathbf{D}\mathbf{u}^I(a)$, we have:

$$\begin{aligned} \mathbf{u}(a)^{(t+1)} &= \mathbf{I}^{-1}(\mathbf{W} - \mathbf{D})\mathbf{u}(a)^{(t)} + \mathbf{I}^{-1}\mathbf{D}\mathbf{u}^I(a) \\ \implies \mathbf{u}(a)^{(t+1)} &= (\mathbf{W} - \mathbf{D})\mathbf{u}(a)^{(t)} + \mathbf{D}\mathbf{u}^I(a) \end{aligned}$$

From Lemma A.2.1, we have $\rho(\mathbf{W} - \mathbf{D}) < 1$. Then, using Corollary A.1.1, we know that $\mathbf{u}^{(t+1)}(a) = (\mathbf{W} - \mathbf{D})\mathbf{u}^{(t)}(a) + \mathbf{D}\mathbf{u}^I(a)$ converges to $\mathbf{u}(a)$ which is the solution to the linear system. ∎

Theorem 3.2.2. *In the iterative scheme above,*

$$\left\| \mathbf{u}(a) - \mathbf{u}^{(t)}(a) \right\|_\infty \leq (1 - \tilde{w})^t \left\| \mathbf{u}(a) - \mathbf{u}^{(0)}(a) \right\|_\infty,$$

where $\tilde{w} = \min_{1 \leq i \leq n} w_{ii}$.

Proof of Theorem 3.2.2 Using Eq. 3.4 and $\mathbf{u}^{(t)}(a) = (\mathbf{W} - \mathbf{D})\mathbf{u}^{(t-1)}(a) + \mathbf{D}\mathbf{u}^I(a)$, we can write $\mathbf{u}(a) - \mathbf{u}^{(t)}(a) = (\mathbf{W} - \mathbf{D})(\mathbf{u}(a) - \mathbf{u}^{(t-1)}(a))$. By induction on t, we have $\mathbf{u}(a) - \mathbf{u}^{(t)}(a) =$

$(\mathbf{W}-\mathbf{D})^t\left(\mathbf{u}(a)-\mathbf{u}^{(0)}(a)\right)$. Thus, we have

$$\begin{aligned}
\left\|\mathbf{u}(a)-\mathbf{u}^{(t)}(a)\right\|_\infty &= \left\|(\mathbf{W}-\mathbf{D})^t\left(\mathbf{u}(a)-\mathbf{u}^{(0)}(a)\right)\right\|_\infty \\
&\leq \left\|(\mathbf{W}-\mathbf{D})^t\right\|_\infty \left\|\mathbf{u}(a)-\mathbf{u}^{(0)}(a)\right\|_\infty \text{ (by compatibility)} \\
&\leq \|\mathbf{W}-\mathbf{D}\|_\infty^t \left\|\mathbf{u}(a)-\mathbf{u}^{(0)}(a)\right\|_\infty \text{ (by consistency)} \\
&= \left(\max_{1\leq i\leq n}\sum_{j=1}^{n}|w_{ij}-d_{ij}|\right)^t \left\|\mathbf{u}(a)-\mathbf{u}^{(0)}(a)\right\|_\infty (\infty\text{-norm}) \\
&= \left(\max_{1\leq i\leq n}\sum_{\substack{j=1\\ j\neq i}}^{n}|w_{ij}|\right)^t \left\|\mathbf{u}(a)-\mathbf{u}^{(0)}(a)\right\|_\infty \text{ (by defn. of } \mathbf{D}) \\
&= \left(\max_{1\leq i\leq n}\sum_{\substack{j=1\\ j\neq i}}^{n}w_{ij}\right)^t \left\|\mathbf{u}(a)-\mathbf{u}^{(0)}(a)\right\|_\infty \text{ (by nonnegativity)} \\
&= \left(1-\min_{1\leq i\leq n} w_{ii}\right)^t \left\|\mathbf{u}(a)-\mathbf{u}^{(0)}(a)\right\|_\infty \text{ (by normalization)}
\end{aligned}$$

Letting $\tilde{w}=\min_{1\leq i\leq n} w_{ii}$, we have shown that

$$\left\|\mathbf{u}(a)-\mathbf{u}^{(t)}(a)\right\|_\infty \leq (1-\tilde{w})^t \left\|\mathbf{u}(a)-\mathbf{u}^{(0)}(a)\right\|_\infty . \quad \blacksquare$$

Lemma A.2.2. *Assume nonnegativity and normalization, consider the iterative updating scheme:* $\mathbf{u}^{(t)}(a)=(\mathbf{W}-\mathbf{D})\mathbf{u}^{(t-1)}(a)+\mathbf{D}\mathbf{u}^I(a)$. *If* $\forall i\in\mathcal{N}$, $u_i^I(a)\in[c,d]$ *and* $u_i^{(0)}(a)\in[c,d]$, *then* $u_i(a)^{(t)}\in[c,d]$, $\forall i\in\mathcal{N}$ *and* $\forall t\in\mathbb{N}$. *Moreover, we have* $u_i(a)\in[c,d]$, $\forall i\in\mathcal{N}$.

Proof We first prove the first part of the lemma by induction on t. The base case is $t=0$ for which it is given that $u_i^{(0)}(a)\in[c,d]$, $\forall i\in\mathcal{N}$. The induction hypothesis is that $u_i^{(t)}(a)\in[c,d]$ for all $\forall i\in\mathcal{N}$. There are two useful inequalities which follow immediately from the induction hypothesis: $\max_{i\in\mathcal{N}} u_i^{(t)}(a)\leq d$ and $\min_{i\in\mathcal{N}} u_i^{(t)}(a)\geq c$. We can write the updating scheme for each individual $i\in\mathcal{N}$ and the alternative $a\in\mathcal{A}$ as follows:

$$u_i^{(t+1)}(a)=w_{ii}u_i^I(a)+\sum_{k\neq i}w_{ik}u_k^{(t)}(a) \tag{A.5}$$

where $u_i^{(t)}(a)$ denotes the utility of individual i for alternative a after t iterations. Using this equation and the two inequalities mentioned above, we will first show the upper bound on d and then the lower bound on c for $u_i^{(t+1)}(a)$, $\forall i\in\mathcal{N}$, and fixed a. For the upper bound, we can

write the following:

$$\begin{aligned}
u_i^{(t+1)}(a) \leq & \max_{i\in\mathcal{N}} \left\{ u_i^{(t+1)}(a) \right\} \\
= & \max_{i\in\mathcal{N}} \left\{ w_{ii} u_i^I(a) + \sum_{k\neq i} w_{ik} u_k^{(t)}(a) \right\} \\
\leq & \max_{i\in\mathcal{N}} \left\{ w_{ii} u_i^I(a) \right\} + \sum_{k\neq i} \max_{i\in\mathcal{N}} \left\{ w_{ik} u_k^{(t)}(a) \right\} \\
\leq & w_{ii} \max_{i\in\mathcal{N}} \left\{ u_i^I(a) \right\} + \sum_{k\neq i} w_{ik} \max_{i\in\mathcal{N}} \left\{ u_k^{(t)}(a) \right\} \\
\leq & w_{ii} d + \sum_{k\neq i} w_{ik} \max_{i\in\mathcal{N}} \left\{ u_k^{(t)}(a) \right\} \\
\leq & w_{ii} d + \sum_{k\neq i} w_{ik} d \\
= & d \sum_k w_{ik} = d.
\end{aligned}$$

Similarly, for the lower bound:

$$\begin{aligned}
u_i^{(t+1)}(a) \geq & \min_{i\in\mathcal{N}} \left\{ u_i^{(t+1)}(a) \right\} \\
= & \min_{i\in\mathcal{N}} \left\{ w_{ii} u_i^I(a) + \sum_{k\neq i} w_{ik} u_k^{(t)}(a) \right\} \\
\geq & \min_{i\in\mathcal{N}} \left\{ w_{ii} u_i^I(a) \right\} + \sum_{k\neq i} \min_{i\in\mathcal{N}} \left\{ w_{ik} u_k^{(t)}(a) \right\} \\
\geq & w_{ii} \min_{i\in\mathcal{N}} \left\{ u_i^I(a) \right\} + \sum_{k\neq i} w_{ik} \min_{i\in\mathcal{N}} \left\{ u_k^{(t)}(a) \right\} \\
\geq & w_{ii} c + \sum_{k\neq i} w_{ik} \left\{ u_k^{(t)}(a) \right\} \\
\geq & w_{ii} c + \sum_{k\neq i} w_{ik} c \\
= & c \sum_k w_{ik} = c.
\end{aligned}$$

So we have shown that $c \leq u_i^{(t+1)}(a) \leq d$, $\forall i \in \mathcal{N}$ and $a \in \mathcal{A}$, so proving the first part of the lemma.

Now, we will prove the second part of lemma by showing that $u_i(a) \in [c, d]$, $\forall i \in \mathcal{N}$ and $a \in \mathcal{A}$. Fix an arbitrary $i \in \mathcal{N}$. The sequence $u_i^{(t)}(a)$ with $t = 0, 1, 2, \ldots$ is a convergent sequence which converges to $u_i(a) = lim_{t\to\infty} u_i^{(t)}(a)$ (based on Theorem 3.2.1). Note that, from the first part of this lemma, we have $u_i^{(t)}(a) \in [c, d]$ for any $t \in \mathbb{N} \cup \{0\}$. So we can see $u_i^{(t)}(a)$ is a

convergent sequence on the closed set $[c,d]$. As $[c,d]$ is closed, the limit point of $u_i^{(t)}(a)$ sequence which is $u_i(a)$ must belong to $[c,d]$. As i and a are chosen arbitrarily, we have $u_i(a) \in [c,d]$. ∎

Theorem 3.2.3. *Assume $u_j^I(a), u_j^{(0)}(a) \in [c,d]$, for all j. Then $\left|sw(a) - sw^{(t)}(a)\right| \leq n(d-c)\,(1-\tilde{w})^t$, for all t, under the conditions above, where $\tilde{w} = \min_{1\leq i\leq n} w_{ii}$.*

Proof of Theorem 3.2.3 Let $\tilde{w} = \min_{1\leq i\leq n} w_{ii}$. Using Theorem 3.2.2, we can write

$$\begin{aligned}
&\left\|\mathbf{u}(a) - \mathbf{u}^{(t)}(a)\right\|_\infty \leq (1-\tilde{w})^t \left\|\mathbf{u}(a) - \mathbf{u}^{(0)}(a)\right\|_\infty \implies \\
&\left|u_i(a) - u_i^{(t)}(a)\right| \leq \max_i \left|u_i(a) - u_i^{(t)}(a)\right| \\
&\leq (1-\tilde{w})^t \left\|\mathbf{u}(a) - \mathbf{u}^{(0)}(a)\right\|_\infty \implies \\
&\sum_{i=1}^n \left|u_i(a) - u_i^{(t)}(a)\right| \leq n\,(1-\tilde{w})^t \left\|\mathbf{u}(a) - \mathbf{u}^{(0)}(a)\right\|_\infty . \qquad \text{(A.6)}
\end{aligned}$$

By Lemma A.2.2, we know that $u_i(a) \in [c,d]$. Based on this and the assumption that $u_i^{(0)}(a) \in [c,d]$, it follows that $|u_i(a) - u_i^{(0)}(a)| \leq d-c$. So we can continue Inequality (A.6) as follows:

$$\begin{aligned}
\sum_{i=1}^n \left|u_i(a) - u^{(t)}(a)\right| &\leq n\,(1-\tilde{w})^t \left\|\mathbf{u}(a) - \mathbf{u}^{(0)}(a)\right\|_\infty \\
&\leq n(d-c)\,(1-\tilde{w})^t . \qquad \text{(A.7)}
\end{aligned}$$

By Lemma A.2.2, we know that $u_i(a) \in [c,d]$ and $u_i^{(t)}(a) \in [c,d]$, $\forall t \in \mathbb{N}\cup\{0\}$. By the triangle inequality, we have:

$$\begin{aligned}
\sum_{i=1}^n \left|u_i(a) - u_i^{(t)}(a)\right| &\geq \left|\sum_{i=1}^n \left(u_i(a) - u_i^{(t)}(a)\right)\right| \\
&= \left|\sum_{i=1}^n u_i(a) - \sum_{i=1}^n u_i^{(t)}(a)\right| \\
&= \left|sw(a) - sw^{(t)}(a)\right| . \qquad \text{(A.8)}
\end{aligned}$$

By Inequalities (A.7) and (A.8), we conclude that

$$\left|sw(a) - sw^{(t)}(a)\right| \leq n(d-c)\,(1-\tilde{w})^t ,$$

where $\tilde{w} = \min_{1\leq i\leq n} w_{ii}$. ∎

Proposition 3.2.4. *If $sw^{(t)}(b) - sw^{(t)}(a) \geq 2n(d-c)\,(1-\tilde{w})^t$ then $sw(b) > sw(a)$.*

Proof of Proposition 3.2.4 Using the triangle inequality and the inequality presented in

Theorem 3.2.3, we can write:

$$\begin{aligned}
&sw^{(t)}(b) - sw^{(t)}(a) \\
&= sw^{(t)}(b) - sw(b) + sw(b) - sw^{(t)}(a) + sw(a) - sw(a) \\
&\leq |sw^{(t)}(b) - sw(b)| + sw(b) + |sw(a) - sw^{(t)}(a)| - sw(a) \\
&\leq n(d-c)(1-\tilde{w})^t + sw(b) + n(d-c)(1-\tilde{w})^t - sw(a) \\
&= 2n(d-c)(1-\tilde{w})^t + sw(b) - sw(a).
\end{aligned}$$

Using this and $sw^{(t)}(b) - sw^{(t)}(a) \geq 2n(d-c)(1-\tilde{w})^t$, we have $2n(d-c)(1-\tilde{w})^t \leq 2n(d-c)(1-\tilde{w})^t + sw(b) - sw(a)$. This implies $sw(b) \geq sw(a)$. ∎

Corollary 3.3.1. *$\boldsymbol{\omega}$ is the unique solution to the linear system of $\mathbf{A}\boldsymbol{\omega} = \mathbf{e}$ where $\mathbf{A} = (\mathbf{I} - \mathbf{W}^\top + \mathbf{D})\mathbf{D}^{-1}$.*

Proof of Corollary 3.3.1 The proof is trivial. From Corollary 3.1.1, we have

$$\begin{aligned}
&\boldsymbol{\omega}^\top = \mathbf{e}^\top(\mathbf{I} - \mathbf{W} + \mathbf{D})^{-1}\mathbf{D} \\
\implies &\boldsymbol{\omega} = \mathbf{D}^\top \left((\mathbf{I} - \mathbf{W} + \mathbf{D})^{-1}\right)^\top \mathbf{e} \\
\implies &\boldsymbol{\omega} = \mathbf{D} \left((\mathbf{I} - \mathbf{W} + \mathbf{D})^\top\right)^{-1} \mathbf{e} \\
\implies &\boldsymbol{\omega} = \mathbf{D}(\mathbf{I}^\top - \mathbf{W}^\top + \mathbf{D}^\top)^{-1}\mathbf{e} \\
\implies &\boldsymbol{\omega} = \mathbf{D}(\mathbf{I} - \mathbf{W}^\top + \mathbf{D})^{-1}\mathbf{e} \\
\implies &(\mathbf{I} - \mathbf{W}^\top + \mathbf{D})\mathbf{D}^{-1}\boldsymbol{\omega} = \mathbf{e}
\end{aligned}$$

Let $\mathbf{A} = (\mathbf{I} - \mathbf{W}^\top + \mathbf{D})\mathbf{D}^{-1}$, so we have the linear system of $\mathbf{A}\boldsymbol{\omega} = \mathbf{e}$ with the solution of $\boldsymbol{\omega}$. For uniqueness of $\boldsymbol{\omega}$, we need to show that $(\mathbf{I} - \mathbf{W}^\top + \mathbf{D})\mathbf{D}^{-1}$ is a non-singular matrix. Based on positive self-loop assumption, $\mathbf{D}$ and its inverse $\mathbf{D}^{-1}$ must be non-singular. In the proof of Proposition 3.1.1, we showed that $(\mathbf{I} - \mathbf{W} + \mathbf{D})$ is non-singular. As the transpose of any non-singular matrix is non-singular, $(\mathbf{I} - \mathbf{W} + \mathbf{D})^\top = (\mathbf{I} - \mathbf{W}^\top + \mathbf{D})$ is non-singular. ∎

Lemma A.2.3. *Assuming $\mathbf{B} = \mathbf{W} - \mathbf{D}$ and $\mathbf{G} = \mathbf{D}(\mathbf{W}^\top - \mathbf{D})\mathbf{D}^{-1}$, then $\rho(\mathbf{G}) = \rho(\mathbf{B})$*

Proof of Lemma A.2.3 By showing that both $\mathbf{G}$ and $\mathbf{B}$ have the same characteristic polynomial (i.e., $p_B(\lambda) = p_G(\lambda)$), we demonstrate that $\sigma(\mathbf{G}) = \sigma(\mathbf{B})$, thus yielding to $\rho(\mathbf{B}) = \rho(\mathbf{G})$ based on the definition. We first note that $\mathbf{G} = \mathbf{D}\mathbf{B}^\top\mathbf{D}^{-1}$ and $det(\mathbf{D}) \neq 0$ (due to positive self-loop assumption). Then, using the definition of $p_B(\lambda)$ and transpose, multiplication and

inverse properties of determinants, we have:

$$\begin{aligned} p_B(\lambda) =& det(\mathbf{B}-\lambda\mathbf{I}) = det\left((\mathbf{B}-\lambda\mathbf{I})^\top\right) = det(\mathbf{B}^\top-\lambda\mathbf{I}) = \frac{det(\mathbf{D})}{det(\mathbf{D})}det(\mathbf{B}^\top-\lambda\mathbf{I}) \\ =& det(\mathbf{D})det(\mathbf{B}^\top-\lambda\mathbf{I})det(\mathbf{D}^{-1}) = det(\mathbf{D}(\mathbf{B}^\top-\lambda\mathbf{I})\mathbf{D}^{-1}) = det(\mathbf{D}\mathbf{B}^\top\mathbf{D}^{-1}-\mathbf{D}\lambda\mathbf{D}^{-1}) \\ =& det(\mathbf{G}-\lambda\mathbf{I}) = p_G(\lambda) \end{aligned}$$

Thus, $\sigma(\mathbf{G}) = \sigma(\mathbf{B})$ and consequently $\rho(\mathbf{G}) = \rho(\mathbf{B})$. ∎

Theorem 3.3.2. *Consider the following update:*

$$\boldsymbol{\omega}^{(t+1)} = \mathbf{D}(\mathbf{W}^\top-\mathbf{D})\mathbf{D}^{-1}\boldsymbol{\omega}^{(t)} + \mathbf{D}\mathbf{e}$$

Assuming nonnegativity, normalizaton, and positive self-loop, this method converges to $\boldsymbol{\omega}$, the solution to linear system stated in Cor. 3.3.1.

Proof of Theorem 3.3.2 From Corollary 3.3.1, we observe that $\boldsymbol{\omega}$ is the unique solution of the linear system of $\mathbf{A}\boldsymbol{\omega} = \mathbf{e}$ where $\mathbf{A} = (\mathbf{I}-\mathbf{W}^\top+\mathbf{D})\mathbf{D}^{-1}$. The Jacobi method (as presented in Eq. A.3) for solving this linear system is

$$\boldsymbol{\omega}^{(t+1)} = \boldsymbol{\Lambda}^{-1}(\mathbf{E}+\mathbf{F})\boldsymbol{\omega}^{(t)} + \boldsymbol{\Lambda}^{-1}\mathbf{e}.$$

Since $\mathbf{A} = (\mathbf{I}-\mathbf{W}^\top+\mathbf{D})\mathbf{D}^{-1} = \mathbf{D}^{-1} - (\mathbf{W}^\top-\mathbf{D})\mathbf{D}^{-1}$, we have $\boldsymbol{\Lambda} = \mathbf{D}^{-1}$ and $\mathbf{E}+\mathbf{F} = (\mathbf{W}^\top-\mathbf{D})\mathbf{D}^{-1}$ based on the definitions, thus yielding the iteration:

$$\boldsymbol{\omega}^{(t+1)} = \mathbf{D}(\mathbf{W}^\top-\mathbf{D})\mathbf{D}^{-1}\boldsymbol{\omega}^{(t)} + \mathbf{D}\mathbf{e}.$$

From Lemma A.2.1 and Lemma A.2.3, we have $\rho(\mathbf{G}) < 1$ where $\mathbf{G} = \mathbf{D}(\mathbf{W}^\top-\mathbf{D})\mathbf{D}^{-1}$. Then, using Corollary A.1.1, we have shown that $\boldsymbol{\omega}^{(t+1)} = \mathbf{D}(\mathbf{W}^\top-\mathbf{D})\mathbf{D}^{-1}\boldsymbol{\omega}^{(t)} + \mathbf{D}\mathbf{e}$ converges to $\boldsymbol{\omega}$ which is the solution to the linear system of $\mathbf{A}\boldsymbol{\omega} = \mathbf{e}$ where $\mathbf{A} = (\mathbf{I}-\mathbf{W}^\top+\mathbf{D})\mathbf{D}^{-1}$. ∎

Lemma A.2.4. *If $\boldsymbol{\omega}$ is the solution to the linear system in Cor. 3.3.1 and $\boldsymbol{\omega}' = (w_{11}, w_{22}, \ldots, w_{nn})^\top$ then $\boldsymbol{\omega} \geq \boldsymbol{\omega}'$.*

Proof of Lemma A.2.4 We first show that $\boldsymbol{\omega} \geq \mathbf{0}$. As Corollary 3.1.1 demonstrates that $\boldsymbol{\omega}^\top = \mathbf{e}^\top(\mathbf{I}-\mathbf{W}+\mathbf{D})^{-1}\mathbf{D}$ and $\mathbf{D}$ is a nonnegative matrix (due to the nonnegativity assumption). It is sufficient to show that $(\mathbf{I}-\mathbf{W}+\mathbf{D})^{-1} \geq \mathbf{0}$. We can see that $\mathbf{I}-\mathbf{W}+\mathbf{D} = \mathbf{I}-(\mathbf{W}-\mathbf{D})$ is in the form of M-matrix. As $\rho(\mathbf{W}-\mathbf{D}) < 1$ (See Lemma A.2.1), by applying Proposition A.1.1, we have $(\mathbf{I}-\mathbf{W}+\mathbf{D})^{-1} \geq \mathbf{0}$ and consequently $\boldsymbol{\omega} \geq \mathbf{0}$.

Theorem 3.3.2 implies that the $\boldsymbol{\omega}$ is the fixed-point of the iterative process of $\boldsymbol{\omega}^{(t+1)} = \mathbf{D}(\mathbf{W}^\top-\mathbf{D})\mathbf{D}^{-1}\boldsymbol{\omega}^{(t)} + \mathbf{D}\mathbf{e}$. So using this and $\mathbf{D}\mathbf{e} = \boldsymbol{\omega}'$, we have $\boldsymbol{\omega} = \mathbf{D}(\mathbf{W}^\top-\mathbf{D})\mathbf{D}^{-1}\boldsymbol{\omega} + \boldsymbol{\omega}'$.

As $\mathbf{W}^\top$, $\mathbf{D}$, $\mathbf{D}^{-1}$ and $\boldsymbol{\omega}^\top$ are nonnegative,

$$\boldsymbol{\omega} \geq \mathbf{0} \implies \mathbf{D}(\mathbf{W}^\top - \mathbf{D})\mathbf{D}^{-1}\boldsymbol{\omega} \geq \mathbf{0} \implies \mathbf{D}(\mathbf{W}^\top - \mathbf{D})\mathbf{D}^{-1}\boldsymbol{\omega} + \boldsymbol{\omega}' \geq \boldsymbol{\omega}' \implies \boldsymbol{\omega} \geq \boldsymbol{\omega}'. \quad \blacksquare$$

Lemma A.2.5. *Assuming non-negativity, normalization and positive self-loop, $\boldsymbol{\omega}$ in the global model always satisfies $\mathbf{e}^\top\boldsymbol{\omega} = n$ or equivalently $\sum_i \omega_i = n$.*

Proof of Lemma A.2.5 We first note that $\mathbf{e}^\top\mathbf{W}^\top = n$ as a consequence of normalization assumption. Then, from Corollary 3.3.1, we can write

$$\begin{aligned}
&(\mathbf{I} - \mathbf{W}^\top + \mathbf{D})\mathbf{D}^{-1}\boldsymbol{\omega} = \mathbf{e} \\
\implies &\mathbf{e}^\top(\mathbf{I} - \mathbf{W}^\top + \mathbf{D})\mathbf{D}^{-1}\boldsymbol{\omega} = \mathbf{e}^\top\mathbf{e} \\
\implies &(\mathbf{e}^\top\mathbf{I} - \mathbf{e}^\top\mathbf{W}^\top + \mathbf{e}^\top\mathbf{D})\mathbf{D}^{-1}\boldsymbol{\omega} = n \\
\implies &(n - n + \mathbf{e}^\top\mathbf{D})\mathbf{D}^{-1}\boldsymbol{\omega} = n \\
\implies &\mathbf{e}^\top\mathbf{D}\mathbf{D}^{-1}\boldsymbol{\omega} = n \\
\implies &\mathbf{e}^\top\mathbf{I}\boldsymbol{\omega} = n \\
\implies &\mathbf{e}^\top\boldsymbol{\omega} = n \quad \blacksquare
\end{aligned}$$

Theorem 3.3.3. *Assume $\boldsymbol{\omega}^{(0)} = (w_{11}, w_{22}, \ldots, w_{nn})^\top$. In the iterative scheme above,*

$$\left\|\boldsymbol{\omega} - \boldsymbol{\omega}^{(t)}\right\|_1 \leq n\frac{\hat{w}}{\tilde{w}}\,(1 - \tilde{w})^t\,(1 - \bar{w}),$$

where $\tilde{w} = \min_{1\leq j\leq n} w_{jj}$, $\hat{w} = \max_{1\leq j\leq n} w_{jj}$, and $\bar{w} = \frac{1}{n}\sum_j w_{jj}$.

Proof of Theorem 3.3.3 From $\boldsymbol{\omega}^{(t)} = \mathbf{D}(\mathbf{W}^\top - \mathbf{D})\mathbf{D}^{-1}\boldsymbol{\omega}^{(t-1)} + \mathbf{De}$, we can write $\boldsymbol{\omega} - \boldsymbol{\omega}^{(t)} = \mathbf{D}(\mathbf{W}^\top - \mathbf{D})\mathbf{D}^{-1}(\boldsymbol{\omega} - \boldsymbol{\omega}^{(t-1)})$. By induction on t, we can show that $\boldsymbol{\omega} - \boldsymbol{\omega}^{(t)} = \mathbf{D}\left(\mathbf{W}^\top - \mathbf{D}\right)^t\mathbf{D}^{-1}(\boldsymbol{\omega} - \boldsymbol{\omega}^{(0)})$. Using this, we can write

$$\begin{aligned}
\left\|\boldsymbol{\omega} - \boldsymbol{\omega}^{(t)}\right\|_1 &= \left\|\mathbf{D}\left(\mathbf{W}^\top - \mathbf{D}\right)^t\mathbf{D}^{-1}(\boldsymbol{\omega} - \boldsymbol{\omega}^{(0)})\right\|_1 \\
&\leq \|\mathbf{D}\|_1\left\|\left(\mathbf{W}^\top - \mathbf{D}\right)^t\right\|_1\left\|\mathbf{D}^{-1}\right\|_1\left\|\boldsymbol{\omega} - \boldsymbol{\omega}^{(0)}\right\|_1 \qquad \text{p-norm compatibility property} \\
&\leq \|\mathbf{D}\|_1\left\|\mathbf{W}^\top - \mathbf{D}\right\|_1^t\left\|\mathbf{D}^{-1}\right\|_1\left\|\boldsymbol{\omega} - \boldsymbol{\omega}^{(0)}\right\|_1 \qquad \text{p-norm consistency property} \\
&= \left(\max_j \sum_i |d_{ij}|\right)\left(\max_j \sum_i |w_{ji} - d_{ij}|\right)^t\left(\max_j \sum_i \left|\frac{1}{d_{ij}}\right|\right)\left\|\boldsymbol{\omega} - \boldsymbol{\omega}^{(0)}\right\|_1 \qquad \text{1-norm Def.} \\
&= \left(\max_j |w_{jj}|\right)\left(\max_j \sum_{i\neq j} |w_{ji}|\right)^t\left(\max_j \left|\frac{1}{w_{jj}}\right|\right)\left\|\boldsymbol{\omega} - \boldsymbol{\omega}^{(0)}\right\|_1 \qquad \text{def. of matrix } \mathbf{D}
\end{aligned}$$

$$
\begin{aligned}
&= \left(\max_j w_{jj}\right)\left(\max_j \sum_{i\neq j} w_{ji}\right)^t \left(\frac{1}{\min_j w_{jj}}\right)\left\|\boldsymbol{\omega}-\boldsymbol{\omega}^{(0)}\right\|_1 && \text{nonnegativity assumption}\\
&= \left(\max_j w_{jj}\right)\left(1-\min_j w_{jj}\right)^t \left(\frac{1}{\min_j w_{jj}}\right)\left\|\boldsymbol{\omega}-\boldsymbol{\omega}^{(0)}\right\|_1 && \text{normalization assumption}\\
&= \left(\frac{\max_j w_{jj}}{\min_j w_{jj}}\right)\left(1-\min_j w_{jj}\right)^t \left(\sum_j |\omega_j - w_{jj}|\right) && \text{1-norm for vectors}\\
&= \left(\frac{\max_j w_{jj}}{\min_j w_{jj}}\right)\left(1-\min_j w_{jj}\right)^t \left(\sum_j (\omega_j - w_{jj})\right) && \text{Lemma A.2.4}\\
&= \left(\frac{\max_j w_{jj}}{\min_j w_{jj}}\right)\left(1-\min_j w_{jj}\right)^t \left(n - \sum_j w_{jj}\right) && \text{Lemma A.2.5}\\
&= n\left(\frac{\max_j w_{jj}}{\min_j w_{jj}}\right)\left(1-\min_j w_{jj}\right)^t \left(1 - \frac{1}{n}\sum_j w_{jj}\right)
\end{aligned}
$$

Let $\tilde{w} = \min_{1\leq j\leq n} w_{jj}$, $\hat{w} = \max_{1\leq j\leq n} w_{jj}$, and $\bar{w} = \frac{1}{n}\sum_j w_{jj}$. Thus,

$$\left\|\boldsymbol{\omega}-\boldsymbol{\omega}^{(t)}\right\|_1 \leq n\frac{\hat{w}}{\tilde{w}}\left(1-\tilde{w}\right)^t\left(1-\bar{w}\right).$$

∎

Theorem 3.3.4. *Assume $\boldsymbol{\omega}^{(0)} = (w_{11}, w_{22}, \ldots, w_{nn})^\top$. Under normalization, nonnegativity, and self-positive loop, for any t:*

$$|sw(\mathbf{x}) - sw^{(t)}(\mathbf{x})| \leq n\frac{\hat{w}}{\tilde{w}}\left(1-\tilde{w}\right)^t\left(1-\bar{w}\right)\left\|\mathbf{u}^I(\mathbf{x})\right\|_2,$$

where $\tilde{w} = \min_{1\leq j\leq n} w_{jj}$, $\hat{w} = \max_{1\leq j\leq n} w_{jj}$, and $\bar{w} = \frac{1}{n}\sum_j w_{jj}$.

Proof of Theorem 3.3.4 For $sw^{(t)}(\mathbf{x})$ and $sw(\mathbf{x})$, using the Cauchy-Schwarz inequality, we can write:

$$|sw(\mathbf{x}) - sw^{(t)}(\mathbf{x})| = |\boldsymbol{\omega}^\top\mathbf{u}^I(\mathbf{x}) - (\boldsymbol{\omega}^{(t)})^\top\mathbf{u}^I(\mathbf{x})| = |(\boldsymbol{\omega}-\boldsymbol{\omega}^{(t)})^\top\mathbf{u}^I(\mathbf{x})| \leq \left\|\boldsymbol{\omega}-\boldsymbol{\omega}^{(t)}\right\|_2\left\|\mathbf{u}^I(\mathbf{x})\right\|_2$$

In general, for a given vector $\mathbf{x}$, $\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1$. Thus, we here have $|sw(\mathbf{x}) - sw^{(t)}(\mathbf{x})| \leq \left\|\boldsymbol{\omega}-\boldsymbol{\omega}^{(t)}\right\|_1\left\|\mathbf{u}^I(\mathbf{x})\right\|_2$. By applying Theorem 3.3.3, we have:

$$|sw(\mathbf{x}) - sw^{(t)}(\mathbf{x})| \leq \left\|\boldsymbol{\omega}-\boldsymbol{\omega}^{(t)}\right\|_1\left\|\mathbf{u}^I(\mathbf{x})\right\|_2 \leq n\frac{\hat{w}}{\tilde{w}}\left(1-\tilde{w}\right)^t\left(1-\bar{w}\right)\left\|\mathbf{u}^I(\mathbf{x})\right\|_2$$

where $\tilde{w} = \min_{1\leq j\leq n} w_{jj}$, $\hat{w} = \max_{1\leq j\leq n} w_{jj}$, and $\bar{w} = \frac{1}{n}\sum_j w_{jj}$. Thus, we have shown

$$|sw(\mathbf{x}) - sw^{(t)}(\mathbf{x})| \leq n\frac{\hat{w}}{\tilde{w}}\left(1-\tilde{w}\right)^t\left(1-\bar{w}\right)\left\|\mathbf{u}^I(a)\right\|_2. \quad \blacksquare$$

Proposition 3.3.5. *If* $sw^{(t)}(\mathbf{x}) - sw^{(t)}(\mathbf{y}) \geq n\frac{\hat{w}}{\tilde{w}}\left(1-\tilde{w}\right)^t\left(1-\bar{w}\right)\left(\left\|\mathbf{u}^I(\mathbf{x})\right\|_2 + \left\|\mathbf{u}^I(\mathbf{y})\right\|_2\right)$ *then* $sw(\mathbf{x}) > sw(\mathbf{y})$.

Proof of Proposition 3.3.5 Using the inequality presented in Theorem 3.3.4, we can write:

$$\begin{aligned}
sw^{(t)}(\mathbf{x}) - sw^{(t)}(\mathbf{y}) &= sw^{(t)}(\mathbf{x}) - sw(\mathbf{x}) + sw(\mathbf{x}) - sw^{(t)}(\mathbf{y}) + sw(\mathbf{y}) - sw(\mathbf{y}) \\
&\leq |sw^{(t)}(\mathbf{x}) - sw(\mathbf{x})| + sw(\mathbf{x}) + |sw(\mathbf{y}) - sw^{(t)}(\mathbf{y})| - sw(\mathbf{y}) \\
&\leq n\frac{\hat{w}}{\tilde{w}}\left(1-\tilde{w}\right)^t\left(1-\bar{w}\right)\left\|\mathbf{u}^I(\mathbf{x})\right\|_2 + sw(\mathbf{x}) + n\frac{\hat{w}}{\tilde{w}}\left(1-\tilde{w}\right)^t\left(1-\bar{w}\right)\left\|\mathbf{u}^I(\mathbf{y})\right\|_2 - sw(\mathbf{y}) \\
&= n\frac{\hat{w}}{\tilde{w}}\left(1-\tilde{w}\right)^t\left(1-\bar{w}\right)\left(\left\|\mathbf{u}^I(\mathbf{x})\right\|_2 + \left\|\mathbf{u}^I(\mathbf{y})\right\|_2\right) + sw(\mathbf{x}) - sw(\mathbf{y})
\end{aligned}$$

Using this and $sw^{(t)}(\mathbf{x}) - sw^{(t)}(\mathbf{y}) \geq n\frac{\hat{w}}{\tilde{w}}\left(1-\tilde{w}\right)^t\left(1-\bar{w}\right)\left(\left\|\mathbf{u}^I(\mathbf{x})\right\|_2 + \left\|\mathbf{u}^I(\mathbf{y})\right\|_2\right)$, we have $sw(\mathbf{x}) \geq sw(\mathbf{y})$. $\blacksquare$

Appendix B

Theoretical Details and Proofs of Chapter 4

This Appendix first presents proofs of theoretical analyses of Chapter 4 in Section B.1. Then, Section B.2 discusses the rational behind the proposed approximations.

B.1 Proofs

Theorem 4.3.1. *Assume a ranking model* $(\rho(r|\boldsymbol{\eta}), c(d|\boldsymbol{\lambda}))$ *over* m *alternatives. The induced ranking network is connected with high probability (i.e., with probability* $1-o(1)$ *where* $o(1) \to 0$ *as* $n \to \infty$*) if*

$$d_M(m) < c^{-1}\left(\frac{\log n}{n}\Big|\boldsymbol{\lambda}\right),$$

where $d_M(m)$ *is the maximum possible distance under* d *given* m *alternatives.*

Proof of Theorem 4.3.1 We note that the connection probability between two nodes with rankings r and $r^{'}$ can be written as $c(d(r,r^{'}),\boldsymbol{\lambda}) = c(d_M(m),\boldsymbol{\lambda}) + (c(d(r,r^{'}),\boldsymbol{\lambda}) - c(d_M(m),\boldsymbol{\lambda}))$. So we can consider that the network formation in the ranking network model occurs in two phases. First any two nodes get connected to each other with fixed probability $c(d_M(m),\boldsymbol{\lambda})$ (similar to the random graph model $G(n,p)$ with $p = c(d_M(m),\boldsymbol{\lambda})$). Then, any pair of nodes with ranking r and $r^{'}$ get connected with the probability $c(d(r,r^{'}),\boldsymbol{\lambda}) - c(d_M(m),\boldsymbol{\lambda})$. (Note that $c(d(r,r^{'}),\boldsymbol{\lambda}) - c(d_M(m),\boldsymbol{\lambda}) \geq 0$ as $c(.)$ is strictly decreasing and $d(r,r') \leq d_M(m)$ for any r and r'.) So if we find conditions under which the network generated after phase 1 is almost surely connected, after the second phase, the connectivity property must hold since other edges can only be added to the network without removal of any existing edge.

It is well-known that the random graph $G(n,p)$ is almost surely connected when $p > \frac{\log n}{n}$ [52]. The first phase is actually the random graph model $G(n,p)$ with $p = c(d_M(m),\boldsymbol{\lambda})$. So we

require

$$c(d_M(m)|\boldsymbol{\lambda}) > \frac{\log n}{n}$$

Since $c(.|\boldsymbol{\lambda})$ is a strictly decreasing function, its inverse $c^{-1}(.|\boldsymbol{\lambda})$ is well-defined and is a strictly increasing function. So, we have

$$\begin{aligned} & c(d_M(m)|\boldsymbol{\lambda}) > \frac{\log n}{n} \\ \implies & c^{-1}(c(d_M(m)|\boldsymbol{\lambda})|\boldsymbol{\lambda}) < c^{-1}(\frac{\log n}{n}) \\ \implies & d_M(m) < c^{-1}\left(\frac{\log n}{n}\Big|\boldsymbol{\lambda}\right) \quad \blacksquare \end{aligned}$$

Theorem 4.3.2. *Fix m and assume $\rho(r|\boldsymbol{\eta})$ distributes probability mass on more than one ranking. The asymptotic diameter of any ranking network over m options is 2 (as $n \to \infty$).*

Proof of Theorem 4.3.2 Our proof has the following steps: first we will show that asymptotic diameter is at most 2 by demonstrating that any two arbitrary nodes are connected through at least one other node in the network. Second, we show that each node cannot be connected to all other nodes in the network, thus resulting in asymptotic diameter of more than 1. Then, the asymptotic diameter has to be 2.

For the first step, we start observing that the connection probability of any two nodes regardless of their rankings is at least $p = c(d_M(m)|\boldsymbol{\lambda})$. We note that $c(d_M(m)|\boldsymbol{\lambda}) > 0$ for any distance metric d and $m \geq 2$; so $p > 0$. Denote the probability that i and j have no common neighbour k by $P(\textit{not } i \sim k \sim j)$. Similarly let $P(\textit{not } i \sim S \sim j)$ denote the probability that i and j have no element of set S as a common neighbour. It is straightforward to see that $P(\textit{not } i \sim S \sim j) \leq 1 - p^2$. Similarly, due to independence of connections, one can observe that $P(\textit{not } i \sim \mathcal{N} \setminus \{i,j\} \sim j) \leq (1-p^2)^{n-2}$. Since $p > 0$, $\lim_{n\to\infty} P(\textit{not } i \sim \mathcal{N} \setminus \{i,j\} \sim j) = 0$. This implies that i and j have to have at least one common neighbour when $n \to \infty$. So the shortest path between these two nodes has length at most of 2. As i and j are two arbitrary nodes, any two pair of nodes at most have the shortest path of length 2. So the asymptotic diameter of the network is at most 2.

We now show that asymptotic diameter is greater than 1. We prove this by contradiction. We assume that the asymptotic diameter is 1. As the probability mass is not on a single ranking, there are at least two rankings r_m and r_M such that $0 < \rho(r_m) \leq \rho(r_M)$ and also $\rho(r_m) \leq \rho(r) \leq \rho(r_M)$ for any r with $\rho(r) > 0$. Let n_{r_m} and n_{r_M} denote the number of nodes with rankings r_m and r_M when there are n nodes. We consider a sufficiently large n such that $n_{r_m}, n_{r_M} > 0$. As r_m and r_M are distinct, $d(r_m, r_M) > 0$. So, $c(d(r_m, r_M)) < c(0)$. Let $P(r_m, n_{r_M})$ denote the probability that one node with ranking r_m gets connected to all nodes with ranking r_M. It is straightforward to see that $P(r_m, n_{r_M}) = c(d(r_m, r_M))^{n_{r_M}}$. We note when $n \to \infty$, $n_{r_M} \to \infty$ when m is fixed. As $c(d(r_m, r_M)) < c(0) \leq 1$, $\lim_{n\to\infty} P(r_m, n_{r_M}) = 0$.

So there must be some nodes with r_M that some node with r_m is not directly connected to. Therefore, the asymptotic diameter can not be 1. Based on this and the first step proof that asymptotic diameter is ≤ 2; the asymptotic diameter must be 2. ∎

Theorem 4.4.1. *Given reference ranking σ and a distance-based ranking model, for any fixed $\boldsymbol{\theta}$ and any $r \in \Omega(\mathcal{A})$:*

$$\mathcal{D}(\sigma_M, \boldsymbol{\theta}) \leq \mathcal{D}(r, \boldsymbol{\theta}) \leq \mathcal{D}(\sigma, \boldsymbol{\theta}),$$

where σ_M is some ranking at maximum distance from σ.[1]

Proof of Theorem 4.4.1. We first note that for $\omega = 0$, the proof is straightforward: when $\omega = 0$, for any $r \in \Omega(\mathcal{A})$, $\rho(r|\boldsymbol{\theta}) = \frac{1}{m!}$. Hence,

$$\begin{aligned} D(r, \boldsymbol{\theta}) &= \sum_{r' \in \Omega(\mathcal{A})} \rho(r'|\boldsymbol{\theta}) c\left(d_\tau(r', r)|\boldsymbol{\theta}\right) = \frac{1}{m!} \sum_{r' \in \Omega(\mathcal{A})} c\left(d_\tau(r', r)|\boldsymbol{\theta}\right) \\ &= \frac{1}{m!} \sum_{i=0}^{d_M(m)} n_i c\left(i|\boldsymbol{\theta}\right), \end{aligned}$$

where n_i represents the number of rankings at distance i from a specific ranking and $d_M(m)$ is the maximum possible distance under the metric d. We note that the value of $D(r, \boldsymbol{\theta})$ is same for any $r \in \Omega(\mathcal{A})$. So, for this case, for any $r \in \Omega(\mathcal{A})$, we have

$$\mathcal{D}(\sigma_M, \boldsymbol{\theta}) = \mathcal{D}(r, \boldsymbol{\theta}) = \mathcal{D}(\sigma, \boldsymbol{\theta}),$$

where σ_M is a ranking with maximum possible distance to σ.

To prove the inequalities for the case $\omega \in (0, \infty)$, we first let $k_{ij}^{(r)}$ denote the number of rankings at distance i from the reference ranking σ and distance j from an arbitrary ranking r. Once again, let n_i represent the number of rankings with distance i from an arbitrary ranking, where $i \in \{0, \cdots, d_M(m)\}$. So, we can write

$$\begin{aligned} \mathcal{D}(r, \boldsymbol{\theta}) &= \sum_{r' \in \Omega(\mathcal{A})} \rho(r'|\boldsymbol{\theta}) c\left(d(r', r)|\boldsymbol{\theta}\right) \\ &= \frac{1}{\psi(\omega)} \sum_{i=0}^{d_M} \sum_{j=0}^{d_M} k_{ij}^{(r)} e^{-\omega i} c(j|\boldsymbol{\theta}). \end{aligned} \tag{B.1}$$

We also observe that for any fixed $r \in \Omega(\mathcal{A})$,

$$\forall j \in \{0, \ldots, d_M\}, \sum_{i=0}^{d_M} k_{ij}^{(r)} = n_j \quad \text{and} \quad \forall i \in \{0, \ldots, d_M\}, \sum_{j=0}^{d_M} k_{ij}^{(r)} = n_i. \tag{B.2}$$

[1] If more than one ranking has maximum distance, one such ranking minimizes $\mathcal{D}$.

We first focus on proving $\mathcal{D}(r,\boldsymbol{\theta}) \leq \mathcal{D}(\sigma,\boldsymbol{\theta})$. By using Eq. B.1, we can view $\mathcal{D}(r,\boldsymbol{\theta})$ as a function over matrix $\mathbf{K}^{(r)}$, denoted by $f\left(\mathbf{K}^{(r)}\right)$, with the constraints stated in Eq. B.2. Now, we explore for which feasible instance of $\mathbf{K}^{(r)}$, $f(\mathbf{K}^{(r)})$ is maximized. Using Lemma B.1.1 below, it follows that $f\left(\mathbf{K}^{(r)}\right)$ is maximized with a diagonal matrix $\mathbf{K}^{(r^*)}$ with $k_{ii}^{(r^*)} = n_i$ for all $i \in \{0,\ldots,d_M\}$. This implies that $k_{00}^{r^*} = n_0 = 1$; so there is one ranking, say $r^{'}$, which has distance 0 to both r^* and σ. Obviously, this can be only true when $r^* = r^{'} = \sigma$ (based on the identity of indiscernibles property of any metric distance). It follows that σ maximizes the function $\mathcal{D}(r,\boldsymbol{\theta})$ when $\boldsymbol{\theta}$ is fixed. Hence, for any $r \in \Omega(\mathcal{A})$: $\mathcal{D}(r,\boldsymbol{\theta}) \leq \mathcal{D}(\sigma,\boldsymbol{\theta})$.

We now focus on $\mathcal{D}(\sigma_M,\boldsymbol{\theta}) \leq \mathcal{D}(r,\boldsymbol{\theta})$. By using Eq. B.1, we can view $\mathcal{D}(r,\boldsymbol{\theta})$ as a function over matrix $\mathbf{K}^{(r)}$, denoted by $f\left(\mathbf{K}^{(r)}\right)$, with the constraints stated in Eq. B.2. Using Lemma B.1.3 below, it follows that the r^* which minimizes $f\left(\mathbf{K}^{(r)}\right)$ must have distance $d_M(m)$ to the reference ranking. So $r^* = \sigma_M$ must be one of those rankings with maximum distance to σ. ∎

Lemma B.1.1. *Assume* $f(\mathbf{X}) = \sum_{i=0}^{d}\sum_{j=0}^{d} a_{ij}x_{ij}$ *and* $a_{ij} = \frac{1}{\psi(\omega)}e^{-\omega i}c(j)$ *where* $\omega \in (0,\infty)$ *and* $c(x|\boldsymbol{\lambda})$ *is a decreasing connection probability function. Consider this constrained maximization problem:*

$$\begin{aligned} \max_{\mathbf{X}\in\mathbb{R}^{(d+1)\times(d+1)}} \quad & f(\mathbf{X}) \\ \textit{subject to:} \quad & \forall i \in \{0,\ldots,d\},\ \sum_{j=0}^{d} x_{ij} = c_i \\ & \forall j \in \{0,\ldots,d\},\ \sum_{i=0}^{d} x_{ij} = c_j \\ & \forall i,j \in \{0,\ldots,d\},\ x_{ij} \geq 0 \end{aligned} \tag{B.3}$$

where $\forall i$, c_i *is a positive constant and given. The solution to this maximization problem is the diagonal matrix* $\mathbf{X}^*$ *such that* $\forall i$, $x_{ii}^* = c_i$.

Proof of Lemma B.1.1. We prove this by contradiction. The proof strategy is as follows. We assume that $\mathbf{X}^*$ maximizes the objective function $f(.)$ but is not diagonal. Then, by a slight modification of some elements in $\mathbf{X}^*$, we create matrix $\mathbf{Y}^*$ and show that it is indeed a feasible solution. Afterwards, we show that $f(\mathbf{Y}^*) > f(\mathbf{X}^*)$ which contradicts the maximality of $\mathbf{X}^*$.

Let $\mathbf{X}^*$ maximizes the objective function $f(.)$ and assume it is not diagonal. Consider this decomposition $\mathbf{X}^* = \mathbf{L}^* + \mathbf{D}^* + \mathbf{U}^*$ where $\mathbf{L}^*$, $\mathbf{D}^*$, and $\mathbf{U}^*$ are strictly lower triangular, diagonal, and strictly upper triangular matrices respectively. Since $\mathbf{X}^*$ is not diagonal, at least one of $\mathbf{L}^*$ and $\mathbf{U}^*$ must have none-zero elements. Moreover, from Lemma B.1.2 below, we can conclude that both $\mathbf{L}^* \neq \mathbf{0}$ and $\mathbf{U}^* \neq \mathbf{0}$. Define non-empty set $S = \{(i,j)|u_{ij}^* > 0 \textit{ and } j > i\}$. We consider the lexicographical ordering on S such that

$$(i,j) < (k,l) \Leftrightarrow i < k \textit{ or } (i = k \textit{ and } j < k).$$

There must be some (i_m, j_m) which is the minimum element in S based on our lexicographical ordering. Based on the definition of S and the minimality of (i_m, j_m) in S, we observe that $u^*_{i_m j_m} > 0$ and $u^*_{k i_m} = 0$ for any $k < i_m$. Hence,

$$\sum_{k<i_m} x^*_{k i_m} = 0 \;\; and \;\; x^*_{i_m j_m} > 0. \tag{B.4}$$

From those two constraints which correspond to the row i_m and column i_m (see Eq. B.3), we can write:

$$\sum_j x^*_{i_m j} = \sum_k x^*_{k i_m} = c_{i_m} \implies \sum_j x^*_{i_m j} - \sum_k x^*_{k i_m} = 0 \implies$$

$$\sum_{j\neq i_m} x^*_{i_m j} - \sum_{k\neq i_m} x^*_{k i_m} = 0 \implies \sum_{j\neq i_m} x^*_{i_m j} - \left(\sum_{k<i_m} x^*_{k i_m} + \sum_{k>i_m} x^*_{k i_m}\right) = 0$$

$$\implies x^*_{i_m j_m} - \left(\sum_{k<j_m} x^*_{k i_m} + \sum_{k>j_m} x^*_{k i_m}\right) \leq 0$$

Using this and Eq. B.7, it follows that

$$0 < x^*_{i_m j_m} \leq \sum_{k>i_m} x^*_{k i_m} \implies \sum_{k>i_m} x^*_{k i_m} > 0.$$

Hence, there is (at least) one $k > i_m$ such that $x^*_{k i_m} > 0$. We can define

$$\delta = \min(x^*_{k i_m}, x^*_{i_m, j_m}).$$

Note that $\delta > 0$ since both $x^*_{k i_m}$ and $x^*_{i_m, j_m}$ are positive. Let

$$B = \{(i_m, i_m), (i_m, j_m), (k, i_m), (k, j_m)\}.$$

Using δ, we can define matrix $\mathbf{Y}^*$ out of $\mathbf{X}^*$ as follows:

$$y^*_{ij} = \begin{cases} x^*_{ij}, & (i,j) \notin B \\ x^*_{i_m i_m} + \delta, & (i,j) = (i_m, i_m) \\ x^*_{i_m j_m} - \delta, & (i,j) = (i_m, j_m) \\ x^*_{k i_m} - \delta, & (i,j) = (k, i_m) \\ x^*_{k j_m} + \delta, & (i,j) = (k, j_m) \end{cases} \tag{B.5}$$

It is straightforward to see that $\mathbf{Y}^*$ satisfies all constraints stated in Eq. B.3. We can first observe that $y^*_{ij} \geq 0$ for all $(i,j) \in B$ mostly because of the way that δ is defined and $\delta > 0$. Moreover, $y^*_{ij} = x^*_{ij} \geq 0$ for all $(i,j) \notin B$. Since $\mathbf{Y}^*$ is different than $\mathbf{X}^*$ only in four corners

of a box determined by coordinates in B, to show that $\mathbf{Y}^*$ satisfies all other constraints, it is sufficient to demonstrate that the constraints related to rows i_m and k, and columns i_m and j_m are satisfied (note that as the other rows and columns are unchanged compared to $\mathbf{X}^*$, their corresponding constraints are already satisfied.). By definition of $\mathbf{Y}^*$ in Eq. B.5, we observe that whenever δ is added to the original X_{ij}, it is subtracted from another element in the same row and another element in the same column; this ensures that the constraints are still met.

As $\mathbf{Y}^*$ is a feasible solution, we can write:

$$\begin{aligned} f(\mathbf{Y}^*) - f(\mathbf{X}^*) &= \sum_i \sum_j y^*_{ij} a_{ij} - \sum_i \sum_j x^*_{ij} a_{ij} = \sum_{(i,j)\in B} a_{ij}(y^*_{ij} - x^*_{ij}) \\ &= a_{i_m i_m}\delta - a_{i_m j_m}\delta - a_{k i_m}\delta + a_{k j_m}\delta \\ &= \frac{\delta}{\psi(\omega)}\left[\left(e^{-\omega i_m}c(i_m) - e^{-\omega i_m}c(j_m)\right) - \left(e^{-\omega k}c(i_m) - e^{-\omega k}c(j_m)\right)\right] \\ &= \frac{\delta}{\psi(\omega)}\left(e^{-\omega i_m} - e^{-\omega k}\right)\left(c(i_m) - c(j_m)\right) \end{aligned}$$

Since $j_m > i_m$ and $c(.)$ is decreasing function, $c(i_m) - c(j_m) > 0$. Also, since $i_m < k$ and $\omega \in (0, \infty)$, we have $e^{-\omega i_m} - e^{-\omega k} > 0$. Hence,

$$\frac{\delta}{\psi(\omega)}\left(e^{-\omega i_m} - e^{-\omega k}\right)\left(c(i_m) - c(j_m)\right) > 0 \implies f(\mathbf{Y}^*) > f(\mathbf{X}^*),$$

which contradicts the optimality of $\mathbf{X}^*$. So $\mathbf{X}^*$ has to be diagonal. Using the diagonal property of $\mathbf{X}^*$ and constraints stated in Eq. B.3, we have $x^*_{ii} = c_i$ and for each $i \neq j \in \{1, \cdots, d\}$, $x^*_{ij} = 0$. ∎

Lemma B.1.2. *Assume matrix* $\mathbf{X}$ *satisfies the constraints in Eq. B.3 and decomposes into diagonal matrix* $\mathbf{D}$*, strictly upper triangular matrix* $\mathbf{U}$ *and strictly lower triangular matrix* $\mathbf{L}$ *(i.e.,* $\mathbf{X} = \mathbf{L} + \mathbf{D} + \mathbf{U}$*). Then,*

$$\mathbf{L} \neq \mathbf{0} \iff \mathbf{U} \neq \mathbf{0}$$

Proof of Lemma B.1.2 We first prove that $\mathbf{L} \neq \mathbf{0} \implies \mathbf{U} \neq \mathbf{0}$. If $\mathbf{L} \neq \mathbf{0}$ then there exist (at least) one (i, j) such that $l_{ij} > 0$ and $i > j$ (note that $l_{ij} = 0$ for all $j \geq i$ based on the strictly lower triangular property of $\mathbf{L}$). Hence, we can define the non-empty set $S = \{(j, i) | l_{ij} > 0 \text{ and } i > j\}$. As S has a finite number of elements, we can order elements of S lexicographically such that $(j, i) > (k, l)$ if and only if $j > k$ or $(j = k \text{ and } i > l)$. Let (j_m, i_m) be the minimum element in S based on our lexicographical ordering. Based on the definition of S and the minimality of (j_m, i_m) in S, we observe that $l_{i_m j_m} > 0$ and $l_{j_m k} = 0$ for any $k < j_m$. Hence,

$$\sum_{k<j_m} x_{j_m k} = 0 \quad \text{and} \quad x_{i_m j_m} > 0. \tag{B.6}$$

From constraints stated in Eq. B.3, we have:

$$\sum_i x_{ij_m} = \sum_k x_{j_m k} = c_{j_m} \implies \sum_i x_{ij_m} - \sum_k x_{j_m k} = 0 \implies$$

$$\sum_{i \neq j_m} x_{ij_m} - \sum_{k \neq j_m} x_{j_m k} = 0 \implies \sum_{i \neq j_m} x_{ij_m} - \left(\sum_{k<j_m} x_{j_m k} + \sum_{k>j_m} x_{j_m k} \right) = 0$$

$$\implies x_{i_m j_m} - \left(\sum_{k<j_m} x_{j_m k} + \sum_{k>j_m} x_{j_m k} \right) \leq 0$$

Using this and Eq. B.6, it follows that

$$0 < x_{i_m j_m} \leq \sum_{k>j_m} x_{j_m k} \implies \sum_{k>j_m} x_{j_m k} > 0.$$

Hence, there is (at least) one $l > j_m$ such that $x_{j_m l} > 0$. So, $u_{j_m l} > 0 \implies \mathbf{U} \neq \mathbf{0}$.

A similar type of argument can be made for the other direction: $\mathbf{U} \neq \mathbf{0} \implies \mathbf{L} \neq \mathbf{0}$. If $\mathbf{U} \neq \mathbf{0}$ then there exists (at least) one (i, j) such that $u_{ij} > 0$ and $i < j$ (note that $u_{ij} = 0$ for all $j \leq i$ based on the strictly upper triangular property of $\mathbf{U}$). Hence, we can define the non-empty set $T = \{(i, j) | u_{ij} > 0 \; and \; i < j\}$. As T has finite number of elements, we can order elements of T lexicographically such that $(i, j) > (k, l) \; if \; and \; only \; if \; i > k \; or \; (i = k \; and \; j > l)$. Let (i_m, j_m) be the minimum element in T based on our lexicographical ordering. Based on the definition of T and the minimality of (i_m, j_m) in T, we observe that $u_{i_m j_m} > 0$ and $u_{k i_m} = 0$ for any $k < i_m$. Hence,

$$\sum_{k<i_m} x_{k i_m} = 0 \;\; and \;\; x_{i_m j_m} > 0. \tag{B.7}$$

From constraints stated in Eq. B.3, we have:

$$\sum_j x_{i_m j} = \sum_k x_{k i_m} = c_{i_m} \implies \sum_j x_{i_m j} - \sum_k x_{k i_m} = 0 \implies$$

$$\sum_{j \neq i_m} x_{i_m j} - \sum_{k \neq i_m} x_{k i_m} = 0 \implies \sum_{j \neq i_m} x_{i_m j} - \left(\sum_{k<i_m} x_{k i_m} + \sum_{k>i_m} x_{k i_m} \right) = 0$$

$$\implies x_{i_m j_m} - \left(\sum_{k<j_m} x_{k i_m} + \sum_{k>j_m} x_{k i_m} \right) \leq 0$$

Using this and Eq. B.7, it follows that

$$0 < x_{i_m j_m} \leq \sum_{k>i_m} x_{k i_m} \implies \sum_{k>i_m} x_{k i_m} > 0.$$

Hence, there is (at least) one $k' > i_m$ such that $x_{k' i_m} > 0$. So, $l_{k' i_m} > 0 \implies \mathbf{L} \neq \mathbf{0}$. ∎

Lemma B.1.3. *Assume* $f(\mathbf{X}) = \sum_{i=0}^{d}\sum_{j=0}^{d} a_{ij}x_{ij}$ *and* $a_{ij} = \frac{1}{\psi(\omega)}e^{-\omega i}c(j)$ *where* $\omega \in (0,\infty)$ *and* $c(x|\boldsymbol{\lambda})$ *is a decreasing connection probability function. Consider this constrained integer programming minimization problem:*

$$\begin{aligned} \min_{\mathbf{X}\in(\mathbb{N}\cup\{0\})^{(d+1)\times(d+1)}} \quad & f(\mathbf{X}) \\ \textit{subject to:} \quad & \forall i \in \{0,\dots,d\},\ \sum_{j=0}^{d} x_{ij} = c_i \\ & \forall j \in \{0,\dots,d\},\ \sum_{i=0}^{d} x_{ij} = c_j \\ & \forall i,j \in \{0,\dots,d\},\ x_{ij} \geq 0 \end{aligned} \tag{B.8}$$

where $\forall i$, c_i *is a positive integer constant. Specifically, it is given that* $c_0 = 1$ *and* $c_d \geq 1$. *Matrix* $\mathbf{X}^*$ *is the solution to this minimization problem with* $x^*_{0d} = c_0 = 1$ *and* $x^*_{id} = 0$ *for* $i < d$.

Proof of Lemma B.1.3 The proof is by contradiction and its general idea is as follows. We assume that $\mathbf{X}^*$ is a feasible solution and minimizes the objective function $f(.)$ but does not satisfy $x^*_{0d} = c_0 = 1$ and $x^*_{id} = 0$ for $i < d$. Then, by slight modification of some elements in $\mathbf{X}^*$, we create matrix $\mathbf{Y}^*$ and show that it is indeed feasible solution. Afterwards, we show that $f(\mathbf{Y}^*) < f(\mathbf{X}^*)$ which contacts the minimality of $\mathbf{X}^*$.

Assume that $\mathbf{X}^*$ minimizes the objective function $f(.)$ but does not satisfy $x^*_{0d} = c_0 = 1$ and $x^*_{id} = 0$ for $i < d$. So, based on the row constraint of $\sum_{j=0}^{d} x^*_{0j} = c_0 = 1$, there should be $l < d$ such that $x^*_{0l} = 1$. Similarly, as $x^*_{0d} = 0$ and $\sum_{i=0}^{d} x^*_{id} = c_d \geq 1$, there should be $k > 0$ such that $x^*_{kd} \geq 1$. Let

$$B = \{(0,l),(0,d),(k,l),(k,d)\},$$

, we can define matrix $\mathbf{Y}^*$ as follows:

$$y^*_{ij} = \begin{cases} x^*_{ij}, & (i,j) \notin B \\ 0, & (i,j) = (0,l) \\ 1, & (i,j) = (0,d) \\ x^*_{kd} - 1, & (i,j) = (k,d) \\ x^*_{kl} + 1, & (i,j) = (k,l) \end{cases} \tag{B.9}$$

It is straightforward to see that $\mathbf{Y}^*$ satisfies all constraints stated in Eq. B.8. We can first observe that $y^*_{ij} \geq 0$ and $y^*_{ij} \in \mathbb{N}\cup\{0\}$ for all $(i,j) \in B$. Moreover, $y^*_{ij} = x^*_{ij} \geq 0$ for all $(i,j) \notin B$. Since $\mathbf{Y}^*$ is different than $\mathbf{X}^*$ only in four corners of a box determined by coordinates in B, to show that $\mathbf{Y}^*$ satisfy all other constraints, it is sufficient to demonstrate the constraints related to rows 0 and k, and columns l and d are satisfied (note that as the other rows and columns are unchanged compared to $\mathbf{X}^*$, their corresponding constraints are already satisfied.). From

the definition of $\mathbf{Y}^*$ in Eq. B.9, we observe that whenever 1 is added to the original x^*_{ij}, it is subtracted from another element in the same row and another element in the same column; this ensures that the constraints are still met.

As $\mathbf{Y}^*$ is a feasible solution, we can write:

$$\begin{aligned} f(\mathbf{Y}^*) - f(\mathbf{X}^*) &= \sum_i \sum_j y^*_{ij} a_{ij} - \sum_i \sum_j x^*_{ij} a_{ij} = \sum_{(i,j)\in B} a_{ij}(y^*_{ij} - x^*_{ij}) \\ &= a_{0d} - a_{0l} - a_{kd} + a_{kl} \\ &= \frac{1}{\psi(\omega)} \left[(c(d) - c(l)) - \left(e^{-\omega k} c(d) - e^{-\omega k} c(l) \right) \right] \\ &= \frac{\delta}{\psi(\omega)} \left(1 - e^{-\omega k} \right) (c(d) - c(l)) \end{aligned}$$

Since $d > l$ and $c(.)$ is decreasing function, $c(d) - c(l) < 0$. Also, since $0 < k$ and $\omega \in (0, \infty)$, we have $1 - e^{-\omega k} > 0$. Hence,

$$\frac{1}{\psi(\omega)} \left(1 - e^{-\omega k} \right) (c(d) - c(l)) < 0 \implies f(\mathbf{Y}^*) > f(\mathbf{X}^*),$$

which contradicts the optimality of $\mathbf{X}^*$. So matrix $\mathbf{X}^*$ must satisfy $x^*_{0d} = c_0 = 1$ and $x^*_{id} = 0$ for $i < d$. ∎

Proposition 4.4.2. *Given a distance-based ranking network, $\mathcal{E}(\boldsymbol{\theta})$ is bounded by*

$$\mathcal{D}(\sigma_M, \boldsymbol{\theta}) \leq \mathcal{E}(\boldsymbol{\theta}) \leq \mathcal{D}(\sigma, \boldsymbol{\theta}),$$

where σ_M is some ranking at maximum distance from σ.

Proof of Proposition 4.4.2 The proof is straightforward and follows from Thm. 4.4.1 and Eq. 4.3:

$$\begin{aligned} &\mathcal{D}(\sigma_M, \boldsymbol{\theta}) \leq \mathcal{D}(r, \boldsymbol{\theta}) \leq \mathcal{D}(\sigma, \boldsymbol{\theta}) \\ &\implies \sum_{r\in\Omega(\mathcal{A})} \rho(r|\boldsymbol{\theta}) \mathcal{D}(\sigma_M, \boldsymbol{\theta}) \leq \sum_{r\in\Omega(\mathcal{A})} \rho(r|\boldsymbol{\theta}) \mathcal{D}(r, \boldsymbol{\theta}) \leq \sum_{r\in\Omega(\mathcal{A})} \rho(r|\boldsymbol{\theta}) \mathcal{D}(\sigma, \boldsymbol{\theta}) \\ &\implies \mathcal{D}(\sigma_M, \boldsymbol{\theta}) \sum_{r\in\Omega(\mathcal{A})} \rho(r|\boldsymbol{\theta}) \leq \mathcal{E}(\boldsymbol{\theta}) \leq \mathcal{D}(\sigma, \boldsymbol{\theta}) \sum_{r\in\Omega(\mathcal{A})} \rho(r|\boldsymbol{\theta}) \\ &\implies \mathcal{D}(\sigma_M, \boldsymbol{\theta}) \leq \mathcal{E}(\boldsymbol{\theta}) \leq \mathcal{D}(\sigma, \boldsymbol{\theta}) \quad ∎ \end{aligned}$$

B.2 Rational Behind Approximations

We here explain the main idea and rational behind our approximation methods.

Approximation to Network Diameter. We know that two nodes have the lowest chance of connectivity when their rankings have the maximum possible distance $d_M(m)$. In other

words, the probability that any pair of nodes get connected is at least $c(d_M(m), \boldsymbol{\lambda})$. We also note that by increasing the edge probability, the diameter (or the average shortest path length) does not increase. So one can bound the diameter $D(\theta)$ (or the average shortest path length $\langle l(\theta)\rangle$) by the diameter $D^{'}(n,p)$ (or the average shortest path length $\langle l^{'}(n,p)\rangle$) of the random graph $G(n,p)$ with fixed $p = c(d_M(m), \boldsymbol{\lambda})$. So $D(\boldsymbol{\theta}) \leq D^{'}(n,p)$ and $\langle l(\boldsymbol{\theta})\rangle \leq \langle l^{'}(n,p)\rangle$ with $p = c(d_M(m), \boldsymbol{\lambda})$.

To give more clear picture, one can assume that the network formation under ranking network model runs in two phases. First any two nodes get connected to each other with fixed probability of $c(d_M(m), \boldsymbol{\lambda})$ (similar to random graph model) which results in diameter, say, $D^{'}(n,p)$ and average shortest path length, say, $\langle l^{'}(n,p)\rangle$. Then, any pair of nodes with ranking r and $r^{'}$ get connected with the probability $c(d(r,r^{'}), \boldsymbol{\lambda}) - c(d_M(m), \boldsymbol{\lambda})$. (Note that $c(d(r,r^{'}), \boldsymbol{\lambda}) - c(d_M(m), \boldsymbol{\lambda}) \geq 0$ as $c(.)$ is strictly decreasing and $d(r,r') \leq d_M(m)$ for any r and r'.) The latter phase results in final diameter $D(\theta)$ and average shortest path length $\langle l(\theta)\rangle$. Obviously adding more edge in the second phase can not increase the diameter and the average shortest path. So $D(\boldsymbol{\theta}) \leq D^{'}(n,p)$ and $\langle l(\theta)\rangle \leq \langle l^{'}(n,p)\rangle$ with $p = c(d_M(m), \boldsymbol{\lambda})$.

One can approximate the diameter $D^{'}(n,p)$ and the average shortest path length $\langle l^{'}(n,p)\rangle$ in $G(n,p)$ model by

$$D^{'}(n,p) \approx \left\lceil \frac{\log(n)}{\log(n-1)p} \right\rceil \ \textit{and}\ \langle l^{'}(n,p)\rangle \approx \frac{\log(n)}{\log(n-1)p}.$$

Setting $p = c(d_M(m), \boldsymbol{\lambda})$ and using the inequalities $D(\theta) \leq D^{'}(n,p)$ and $\langle l(\theta)\rangle \leq \langle l^{'}(n,p)\rangle$, we have:

$$D(\boldsymbol{\theta}) \leq \left\lceil \frac{\log(n)}{\log(n-1) + \log c(d_M(m)|\boldsymbol{\lambda})} \right\rceil \ \textit{and}\ \langle l(\boldsymbol{\theta})\rangle \leq \frac{\log(n)}{\log(n-1) + \log c(d_M(m)|\boldsymbol{\lambda})}.$$

$\tilde{\tilde{E}}$ **Approximation.** Let $p(d(r,\sigma) = i|\omega,\sigma)$ denote the probability that the ranking r drawn from the distance-based ranking distribution $\rho(r|\omega,\sigma)$ has distance i to the reference ranking σ. We note that for $i \in \{0, \cdots, d_M(m)\}$,

$$p(d(r,\sigma) = i|\omega,\sigma) = n_i \frac{e^{-i\omega}}{\psi(\omega)},$$

where n_i represents the number of rankings with distance of i to the reference ranking σ. One can approximate $p(d(r,\sigma) = i|\omega,\sigma)$ by a binomial distribution

$$B\left(i|d_M(m), \frac{e^{-\omega}}{1+e^{-\omega}}\right) = \binom{d_M(m)}{i} \frac{e^{-i\omega}}{(1+e^{-\omega})^{d_M(m)}}$$

where n_i, and $\psi(\omega)$ are approximated by $\binom{d_M(m)}{i}$ and $(1+e^{-\omega})^{d_M(m)}$ respectively. Using this approximation, the convexity of function of $c(.|\boldsymbol{\lambda})$, and Equations 4.14, 4.10, and 4.11, we can write

$$\begin{aligned}
\tilde{\mathcal{D}}(x,\boldsymbol{\theta}) &= \left(1-\frac{x}{d_M(m)}\right)\mathcal{D}(\sigma,\boldsymbol{\theta}) + \left(\frac{x}{d_M(m)}\right)\mathcal{D}(\sigma_M,\boldsymbol{\theta}) \\
&\approx \left(1-\frac{x}{d_M(m)}\right)\sum_{i=0}^{d_M(m)}\binom{d_M(m)}{i}\frac{e^{-i\omega}}{(1+e^{-\omega})^{d_M(m)}}c(i|\boldsymbol{\lambda}) \\
&+ \left(\frac{x}{d_M(m)}\right)\sum_{i=0}^{d_M(m)}\binom{d_M(m)}{i}\frac{e^{-i\omega}}{(1+e^{-\omega})^{d_M(m)}}c(d_M(m)-i|\boldsymbol{\lambda}) \\
&= \left(1-\frac{x}{d_M(m)}\right)\underset{x\sim B\left(d_M(m),\frac{e^{-\omega}}{1+e^{-\omega}}\right)}{\mathbb{E}\left[c(x,\boldsymbol{\lambda})\right]} + \frac{x}{d_M(m)}\underset{x\sim B\left(d_M(m),\frac{e^{-\omega}}{1+e^{-\omega}}\right)}{[c(d_M(m)-x,\boldsymbol{\lambda})]} \\
&\leq \left(1-\frac{x}{d_M(m)}\right)c(\mathbb{E}[x],\boldsymbol{\lambda}) + \left(\frac{x}{d_M(m)}\right)c(d_M(m)-\mathbb{E}[x],\boldsymbol{\lambda})
\end{aligned}$$

We let $\tilde{\tilde{\mathcal{D}}} = \left(1-\frac{x}{d_M(m)}\right)c(\mathbb{E}[x],\boldsymbol{\lambda}) + \left(\frac{x}{d_M(m)}\right)c(d_M(m)-\mathbb{E}[x],\boldsymbol{\lambda})$ Now, as $\underset{x\sim B\left(d_M(m),\frac{e^{-\omega}}{1+e^{-\omega}}\right)}{\mathbb{E}\left[x\right]} = \frac{d_M(m)e^{-\omega}}{1+e^{-\omega}}$, we have

$$\tilde{\tilde{\mathcal{D}}}(x,\boldsymbol{\theta}) = \left(1-\frac{x}{d_M(m)}\right)c\left(\frac{d_M(m)e^{-\omega}}{1+e^{-\omega}}\right) + \frac{x}{d_M(m)}c\left(\frac{d_M(m)}{1+e^{-\omega}}\right).$$

Using this approximation, one can approximate the edge density by

$$\begin{aligned}
\mathcal{E}\tilde{\tilde{(}}\boldsymbol{\theta}) &= \underset{x\sim B\left(d_M(m),\frac{e^{-\omega}}{1+e^{-\omega}}\right)}{\mathbb{E}\left[\tilde{\tilde{\mathcal{D}}}(x,\boldsymbol{\theta})\right]} \\
&= \frac{1}{1+e^{-\omega}}c\left(\frac{d_M(m)e^{-\omega}}{1+e^{-\omega}}\right) + \frac{e^{-\omega}}{1+e^{-\omega}}c\left(\frac{d_M(m)}{1+e^{-\omega}}\right)
\end{aligned}$$

Bibliography

[1] David J. Abraham, Avrim Blum, and Tuomas Sandholm. Clearing algorithms for barter exchange markets: enabling nationwide kidney exchanges. In *Proceedings of the 8th ACM conference on Electronic commerce*, EC '07, pages 295–304, New York, NY, USA, 2007. ACM.

[2] Daron Acemoglu, Munther A. Dahleh, Ilan Lobel, and Asuman Ozdaglar. Bayesian learning in social networks. *The Review of Economic Studies*, 78(4):1201–1236, 2011.

[3] Daron Acemoglu and Asuman Ozdaglar. Opinion dynamics and learning in social networks. *Dynamic Games and Applications*, 1(1):3–49, 2011.

[4] Lada A. Adamic, Thomas M. Lento, and Andrew T. Fiore. How you met me. In *Proceedings of the Sixth International Conference on Weblogs and Social Media*, 2012.

[5] Eytan Adar and Lada A. Adamic. Tracking information epidemics in blogspace. In *Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence*, WI '05, pages 207–214, Washington, DC, USA, 2005. IEEE Computer Society.

[6] Gagan Aggarwal, Jon Feldman, S. Muthukrishnan, and Martin Pál. Sponsored search auctions with Markovian users. In *Proceedings of the 4th International Workshop on Internet and Network Economics*, WINE '08, pages 621–628, Berlin, Heidelberg, 2008. Springer-Verlag.

[7] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.

[8] Noga Alon, Moshe Babaioff, Ron Karidi, Ron Lavi, and Moshe Tennenholtz. Sequential voting with externalities: Herding in social networks. In *Proceedings of the 13th ACM Conference on Electronic Commerce (EC-12)*, pages 36–36, 2012.

[9] Noga Alon, Felix Fischer, Ariel Procaccia, and Moshe Tennenholtz. Sum of us: Strategyproof selection from the selectors. In *Proceedings of the 13th Conference on Theoretical Aspects of Rationality and Knowledge*, pages 101–110. ACM, 2011.

[10] Luis A. Nunes Amaral, Antonio Scala, Marc Barthélémy, and H. Eugene Stanley. Classes of small-world networks. *Proceedings of the National Academy of Sciences*, 97(21):11149–11152, 2000.

[11] Luis A Nunes Amaral, Antonio Scala, Marc Barthélémy, and H Eugene Stanley. Classes of small-world networks. *Proceedings of the National Academy of Sciences*, 97(21):11149–11152, 2000.

[12] Sihem Amer-Yahia, Senjuti Basu Roy, Ashish Chawlat, Gautam Das, and Cong Yu. Group recommendation: Semantics and efficiency. *Proc. VLDB Endow.*, 2(1):754–765, 2009.

[13] Kenneth J Arrow. A difficulty in the concept of social welfare. *The Journal of Political Economy*, 58(4):328–346, 1950.

[14] Kenneth J. Arrow. *Social Choice and Individual Values*. Yale University Press, 1951.

[15] Kenneth J. Arrow. Extended sympathy and the possibility of social choice. *Philosophia*, 7(2):223–237, 1978.

[16] David Arthur, Rajeev Motwani, Aneesh Sharma, and Ying Xu. Pricing strategies for viral marketing on social networks. In *Proceedings of the 5th International Workshop on Internet and Network Economics*, WINE '09, pages 101–112. Springer, 2009.

[17] Pranjal Awasthi and Tuomas Sandholm. Online stochastic optimization in the large: application to kidney exchange. In *Proceedings of the 21st international jont conference on Artifical intelligence*, IJCAI'09, pages 405–411, San Francisco, CA, USA, 2009. Morgan Kaufmann Publishers Inc.

[18] Hossein Azari, David Parkes, and Lirong Xia. Random utility theory for social choice. In *Advances in Neural Information Processing Systems (NIPS-12)*, pages 126–134, 2012.

[19] Lars Backstrom, Paolo Boldi, Marco Rosa, Johan Ugander, and Sebastiano Vigna. Four degrees of separation. In *Proceedings of the 3rd Annual ACM Web Science Conference*, pages 33–42. ACM, 2012.

[20] Lars Backstrom and Jure Leskovec. Supervised random walks: Predicting and recommending links in social networks. In *Proceedings of the fourth ACM International Conference on Web Search and Data Mining*, WSDM '11, pages 635–644, New York, NY, USA, 2011. ACM.

[21] Kyle Bagwell. The economic analysis of advertising. *Handbook of Industrial Organization*, 3:1701–1844, 2007.

[22] Venkatesh Bala and Sanjeev Goyal. Learning from neighbours. *The Review of Economic Studies*, 65(3):595–621, 1998.

[23] Suhrid Balakrishnan and Sumit Chopra. Collaborative ranking. In *Proceedings of the Fifth ACM international Conference on Web Search and Data Mining*, WSDM '12, pages 143–152, New York, NY, USA, 2012. ACM.

[24] Linas Baltrunas, Tadas Makcinskas, and Francesco Ricci. Group recommendations with

rank aggregation and collaborative filtering. In *Proceedings of the 4th ACM Conference on Recommender Systems*, pages 119–126, 2010.

[25] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[26] Alain Barrat, Marc Barthelemy, and Alessandro Vespignani. *Dynamical Processes on Complex Networks*. Cambridge University Press, 2008.

[27] Marc Barthélemy. Spatial networks. *Physics Reports*, 499(1):1–101, 2011.

[28] John Bartholdi III, Craig A Tovey, and Michael A Trick. Voting schemes for which it can be difficult to tell who won the election. *Social Choice and welfare*, 6(2):157–165, 1989.

[29] John J Bartholdi III and James B Orlin. Single transferable vote resists strategic voting. *Social Choice and Welfare*, 8(4):341–354, 1991.

[30] John J Bartholdi III, Craig A Tovey, and Michael A Trick. The computational difficulty of manipulating an election. *Social Choice and Welfare*, 6(3):227–241, 1989.

[31] Joseph F. Bennett and William L. Hays. Multidimensional unfolding: Determining the dimensionality of ranked preference data. *Psychometrika*, 25(1):27–43, 1960.

[32] W.F. Benter. Computer-based horse race handicapping and wagering systems: A report. In Donald B. Hausch, Victor S.Y. Lo, and William T. Ziemba, editors, *Efficiency of Race Track Betting Markets*. Academic Press, January 1994.

[33] Shlomo Berkovsky and Jill Freyne. Group-based recipe recommendations: Analysis of data aggregation strategies. In *Proceedings of the 4th ACM Conference on Recommender Systems*, pages 111–118, 2010.

[34] Nadja Betzler, Jiong Guo, and Rolf Niedermeier. Parameterized computational complexity of dodgson and young elections. *Information and Computation*, 208(2):165–177, 2010.

[35] Nadja Betzler, Rolf Niedermeier, and Gerhard J Woeginger. Unweighted coalitional manipulation under the borda rule is np-hard. In *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence*, pages 55–60. AAAI Press, 2011.

[36] Anand Bhalgat, Sreenivas Gollapudi, and Kamesh Munagala. Mechanisms and allocations with positive network externalities. In *Proceedings of the 13th ACM Conference on Electronic Commerce (EC-12)*, pages 179–196, 2012.

[37] Shishir Bharathi, David Kempe, and Mahyar Salek. Competitive influence maximization in social networks. In *Internet and Network Economics*, pages 306–311. Springer, 2007.

[38] Sayan Bhattacharya, Dmytro Korzhyk, and Vincent Conitzer. Computing a profit-maximizing sequence of offers to agents in a social network. In *Proceedings of the 8th International Workshop on Internet and Network Economics*, pages 482–488. Springer,

2012.

[39] Ginestra Bianconi and Albert-László Barabási. Competition and multiscaling in evolving networks. *EPL (Europhysics Letters)*, 54(4):436, 2001.

[40] Mustafa Bilgic, Lilyana Mihalkova, and Lise Getoor. Active learning for networked data. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 79–86, 2010.

[41] Ken Binmore. *Natural Justice*. Oxford University Press, 2005.

[42] Ken Binmore. Interpersonal comparison of utility. 2007.

[43] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[44] Lawrence E. Blume. The statistical mechanics of strategic interaction. *Games and Economic Behavior*, 5(3):387–424, 1993.

[45] Stefano Boccaletti, Vito Latora, Yamir Moreno, Martin Chavez, and D-U Hwang. Complex networks: Structure and dynamics. *Physics reports*, 424(4):175–308, 2006.

[46] Elizabeth Bodine-Baron, Christina Lee, Anthony Chong, Babak Hassibi, and Adam Wierman. Peer effects and stability in matching markets. In *Proceedings of the Fourth Symposium on Algorithmic Game Theory*, pages 117–129, 2011.

[47] Anna Bogomolnaia and Jean-François Laslier. Euclidean preferences. *Journal of Mathematical Economics*, 43(2):87–98, 2007.

[48] Marián Boguñá and Romualdo Pastor-Satorras. Class of correlated random networks with hidden variables. *Physical Review E*, 68:036112, Sep 2003.

[49] Marián Boguñá, Romualdo Pastor-Satorras, Albert Díaz-Guilera, and Alex Arenas. Models of social networks based on social distance attachment. *Physical Review E*, 70:056122, Nov 2004.

[50] Paolo Boldi, Francesco Bonchi, Carlos Castillo, and Sebastiano Vigna. Voting in social networks. In *Proceedings of the Eighteenth Conference on Information and Knowledge Management (CIKM-09)*, pages 777–786, 2009.

[51] Paolo Boldi, Francesco Bonchi, Carlos Castillo, and Sebastiano Vigna. Viscous democracy for social networks. *Communications of the ACM*, 54(6):129–137, 2011.

[52] Béla Bollobás. *Random graphs*. Cambridge University Press, 2001.

[53] Philip Bonacich. Power and centrality: A family of mesasures. *American Journal of Socialogy*, pages 1170–1182, 1987.

[54] Robert M. Bond, Christopher J. Fariss, Jason J. Jones, Adam D.I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. A 61-million-person experiment in social

influence and political mobilization. *Nature*, 489(7415):295–298, 09 2012.

[55] Allan Borodin, Mark Braverman, Brendan Lucier, and Joel Oren. Strategyproof mechanisms for competitive influence in networks. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13, pages 141–150, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee.

[56] Allan Borodin, Yuval Filmus, and Joel Oren. Threshold models for competitive influence in social networks. In *Internet and Network Economics*, pages 539–550. Springer, 2010.

[57] Craig Boutilier, Ioannis Caragiannis, Simi Haber, Tyler Lu, Ariel D. Procaccia, and Or Sheffet. Optimal social choice functions: A utilitarian view. In *Proceedings of the 13th ACM Conference on Electronic Commerce (EC-12)*, pages 197–214, 2012.

[58] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[59] Ralph Allen Bradley and Milton E. Terry. Rank analysis of incomplete block designs. *Biometrika*, 39:324–345, 1952.

[60] Yann Bramoullé, Sergio Currarini, Matthew O. Jackson, Paolo Pin, and Brian W. Rogers. Homophily and long-run integration in social networks. *Journal of Economic Theory*, 2012.

[61] Steven Brams and Peter Fishburn. Approval voting. *American Political Science Review*, 72(3):831–847, 1978.

[62] Steven J Brams and Peter C Fishburn. *Approval voting*. Springer, 2007.

[63] Felix Brandt, Vincent Conitzer, and Ulle Endriss. Computational social choice. *Multiagent systems*, pages 213–283, 2012.

[64] Felix Brandt, Tuomas Sandholm, and Yoav Shoham. Spiteful bidding in sealed-bid auctions. In *Proceedings of the 20th International Joint Conference on Artifical intelligence*, IJCAI'07, pages 1207–1214, San Francisco, CA, USA, 2007. Morgan Kaufmann Publishers Inc.

[65] Simina Brânzei and Kate Larson. Coalitional affinity games and the stability gap. In *Proceedings of the 21st International Joint Conference on Artifical Intelligence*, IJCAI'09, pages 79–84, 2009.

[66] Simina Brânzei and Kate Larson. Social distance games. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI-11)*, pages 91–96, 2011.

[67] Daniel J. Brass. Being in the right place: A structural analysis of individual influence in an organization. *Administrative Science Quarterly*, pages 518–539, 1984.

[68] John S. Breese, David Heckerman, and Carl Kadie. Empirical analysis of predictive

algorithms for collaborative filtering. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, UAI'98, pages 43–52, 1998.

[69] Chris Burges, Tal Shaked, Erin Renshaw, Ari Lazier, Matt Deeds, Nicole Hamilton, and Greg Hullender. Learning to rank using gradient descent. In *Proceedings of the 22nd International Conference on Machine Learning*, ICML '05, pages 89–96, New York, NY, USA, 2005. ACM.

[70] Rainer E Burkard, Mauro Dell'Amico, and Silvano Martello. *Assignment problems*. Society for Industrial and Applied Mathematics (SIAM), 2009.

[71] Ronald S. Burt. *Structural Holes: The social Structure of Competition*. Harvard University Press, 1992.

[72] Robert Burton. *The Anatomy of Melancholy*. JW Moore, 1621.

[73] Ludwig M. Busse, Peter Orbanz, and Joachim M. Buhmann. Cluster analysis of heterogeneous rank data. In *Proceedings of the 24th International Conference on Machine Learning (ICML-07)*, pages 113–120, 2007.

[74] Antoni Calvo-Armengol and Matthew O. Jackson. The effects of social networks on employment and inequality. *The American Economic Review*, 94(3):426–454, 2004.

[75] Antoni Calvó-Armengol and Matthew O Jackson. Peer pressure. *Journal of the European Economic Association*, 8(1):62–89, 2010.

[76] Ozan Candogan, Kostas Bimpikis, and Asuman Ozdaglar. Optimal pricing in the presence of local network effects. In *Proceedings of the 6th International Workshop on Internet and Network Economics*, pages 118–132. Springer, 2010.

[77] Ozan Candogan, Kostas Bimpikis, and Asuman Ozdaglar. Optimal pricing in networks with externalities. *Operations Research*, 60(4):883–905, 2012.

[78] Ioannis Caragiannis, Jason A Covey, Michal Feldman, Christopher M Homan, Christos Kaklamanis, Nikos Karanikolas, Ariel D Procaccia, and Jeffrey S Rosenschein. On the approximability of dodgson and young elections. *Artificial Intelligence*, 2012.

[79] Tim Carnes, Chandrashekhar Nagarajan, Stefan M. Wild, and Anke Van Zuylen. Maximizing influence in a competitive social network: A follower's perspective. In *Proceedings of the Ninth International Conference on Electronic Commerce*, pages 351–360. ACM, 2007.

[80] John R. Chamberlin and Paul N. Courant. Representative deliberations and representative decisions: Proportional representation and the Borda rule. *The American Political Science Review*, 77(3):718–733, 1983.

[81] Weiwei Cheng, Jens Hühn, and Eyke Hüllermeier. Decision tree and instance-based learning for label ranking. In *Proceedings of the 26th Annual International Conference on*

Machine Learning, pages 161–168. ACM, 2009.

[82] Weiwei Cheng, Eyke Hüllermeier, and Krzysztof J Dembczynski. Label ranking methods based on the plackett-luce model. In *Proceedings of the 27th International Conference on Machine Learning*, pages 215–222, 2010.

[83] Yann Chevaleyre, Paul E Dunne, Ulle Endriss, Jérôme Lang, Michel Lemaitre, Nicolas Maudet, Julian Padget, Steve Phelps, Juan A Rodriguez-Aguilar, and Paulo Sousa. Issues in multiagent resource allocation. *Informatica*, 30(1):3–31, 2006.

[84] Yann Chevaleyre, Ulle Endriss, Sylvia Estivie, and Nicolas Maudet. Multiagent resource allocation in k-additive domains: Preference representation and complexity. *Annals of Operations Research*, 163(1):49–62, 2008.

[85] Flavio Chierichetti, Jon Kleinberg, and David Liben-Nowell. Reconstructing patterns of information diffusion from incomplete observations. In *Proceedings of the 25th Annual Conference on Neural Information Processing Systems (NIPS'11)*, pages 792–800, December 2011.

[86] Flavio Chierichetti, Jon Kleinberg, and Alessandro Panconesi. How to schedule a cascade in an arbitrary graph. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 355–368. ACM, 2012.

[87] Nicholas A. Christakis and James H. Fowler. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357(4):370–379, 2007.

[88] Nicholas A. Christakis and James H. Fowler. *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives – How Your Friends' Friends' Friends Affect Everything You Feel, Think, and Do*. Back Bay Books, 2011.

[89] Nicholas A. Christakis and James H. Fowler. Social contagion theory: examining dynamic social networks and human behavior. *Statistics in Medicine*, 2012.

[90] Aaron Clauset, Cristopher Moore, and Mark E.J. Newman. Hierarchical structure and the prediction of missing links in networks. *Nature*, 453(7191):98–101, 2008.

[91] William W. Cohen, Robert E. Schapire, and Yoram Singer. Learning to order things. *Journal of Artificial Intelligence Research*, 10(1):243–270, 1999.

[92] Vincent Conitzer and Tuomas Sandholm. Vote elicitation: Complexity and strategy-proofness. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI-02)*, pages 392–397, Edmonton, 2002.

[93] Vincent Conitzer and Tuomas Sandholm. Universal voting protocol tweaks to make manipulation hard. In *Proceedings of the 18th international joint conference on Artificial intelligence*, IJCAI'03, pages 781–788, 2003.

[94] Vincent Conitzer and Tuomas Sandholm. Communication complexity of common voting

rules. In *Proceedings of the Sixth ACM Conference on Electronic Commerce (EC-05)*, pages 78–87, Vancouver, 2005.

[95] Vincent Conitzer and Tuomas Sandholm. Nonexistence of voting rules that are usually hard to manipulate. In *Proceedings of the Twenty-first National Conference on Artificial Intelligence (AAAI-06)*, pages 627–634, 2006.

[96] Vincent Conitzer and Tuomas Sandholm. Computing optimal outcomes under an expressive representation of settings with externalities. *Journal of Computer and System Sciences*, 78(1):2–14, 2012.

[97] Vincent Conitzer, Tuomas Sandholm, and Jérôme Lang. When are elections with few candidates hard to manipulate? *Journal of the ACM (JACM)*, 54(3):14, 2007.

[98] Clyde H. Coombs. Psychological scaling without a unit of measurement. *Psychological Review*, 57(3):145–158, 1950.

[99] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms*. The MIT Press, 3rd edition, 2009.

[100] Koby Crammer and Yoram Singer. Pranking with ranking. In *Advances in Neural Information Processing Systems*, pages 641–647, 2001.

[101] Daryl J. Daley and David G. Kendall. Epidemics and rumours. *Nature*, 204, 1964.

[102] DJ Daley and David G Kendall. Stochastic rumours. *IMA Journal of Applied Mathematics*, 1(1):42–55, 1965.

[103] Jesper Dall and Michael Christensen. Random geometric graphs. *Physical Review E*, 66:016121, July 2002.

[104] H. E. Daniels. Rank correlation and population models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 12(2):171–191, 1950.

[105] Jessica Davies, George Katsirelos, Nina Narodytska, and Toby Walsh. Complexity of and algorithms for borda manipulation. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence*, pages 657–662, 2011.

[106] Jean-Charles de Borda. Mémoire sur les élections au scrutin. *Histoire de l'Académie Royale des Sciences*, 1781.

[107] Jean Antoine Nicolas de Caritat marquis de Condorcet. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. L'imprimerie royale, 1785.

[108] Morris H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.

[109] Peter M DeMarzo, Dimitri Vayanos, and Jeffrey Zwiebel. Persuasion bias, social influence, and unidimensional opinions. *The Quarterly Journal of Economics*, 118(3):909–968, 2003.

[110] Changrong Deng and Saša Pekec. Money for nothing: Exploiting negative externalities. In *Proceedings of the ACM Conference on Electronic commerce (EC)*, pages 361–370, 2011.

[111] Persi Diaconis. *Group Representations in Probability and Statistics*. Institute of Mathematical Statistics, Hayward, California, 1988.

[112] Jean-Paul Doignon, Aleksandar Pekec, and Michel Regenwetter. The repeated insertion model for rankings: Missing link between two subset choice models. *Psychometrika*, 69(1):33–54, 2004.

[113] Pedro Domingos and Matt Richardson. Mining the network value of customers. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 57–66. ACM, 2001.

[114] P. Alex Dow, Lada A. Adamic, and Adrien Friggeri. The anatomy of large facebook cascades. In *Proceedings of the Seventh International Conference on Weblogs and Social Media*, 2013.

[115] Joanna Drummond and Craig Boutilier. Elicitation and approximately stable matching with partial preferences. In *Proceedings of the Twenty-third International Joint Conference on Artificial Intelligence (IJCAI-13)*, Beijing, 2013.

[116] Richard Durrett. *Random Graph Dynamics*, volume 20. Cambridge University Press, 2007.

[117] Cynthia Dwork, Ravi Kumar, Moni Naor, and D. Sivakumar. Rank aggregation methods for the Web. In *Proceedings of the 10th international conference on World Wide Web*, WWW '01, pages 613–622, New York, NY, USA, 2001. ACM.

[118] David A. Easley and Jon M. Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, Cambridge, 2010.

[119] Holger Ebel, Lutz-Ingo Mielsch, and Stefan Bornholdt. Scale-free topology of e-mail networks. *Physical Review E*, 66:035103, Sep 2002.

[120] Federico Echenique and M. Bumin Yenmez. A solution to matching with preferences over colleagues. *Games and Economic Behavior*, 59(1):46–71, April 2007.

[121] Nicholas Economides. The economics of networks. *International Journal of Industrial Organization*, 14(6):673–699, 1996.

[122] Milad Eftekhar, Yashar Ganjali, and Nick Koudas. Information cascade at group scale. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '13, pages 401–409, New York, NY, USA, 2013. ACM.

[123] Jon X Eguia. Foundations of spatial preferences. *Journal of Mathematical Economics*, 47(2):200–205, 2011.

[124] Edith Elkind and Helger Lipmaa. Hybrid voting protocols and hardness of manipulation. In *Proceedings of the 16th international conference on Algorithms and Computation*, ISAAC'05, pages 206–215, 2005.

[125] Glenn Ellison. Learning, local interaction, and coordination. *Econometrica: Journal of the Econometric Society*, pages 1047–1071, 1993.

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

[127] Stephen Eubank, Hasan Guclu, V.S. Anil Kumar, Madhav V. Marathe, Aravind Srinivasan, Zoltan Toroczkai, and Nan Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 429(6988):180–184, 2004.

[128] Piotr Faliszewski, Edith Hemaspaandra, and Henning Schnoor. Copeland voting: Ties matter. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems*, pages 983–990. International Foundation for Autonomous Agents and Multiagent Systems, 2008.

[129] Piotr Faliszewski and Ariel D Procaccia. Ai's war on manipulation: Are we winning? *AI Magazine*, 31(4):53, 2010.

[130] Peter C Fishburn. *The theory of social choice*, volume 264. Princeton University Press, 1973.

[131] Michael A Fligner and Joseph S Verducci. Distance based ranking models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 359–369, 1986.

[132] Michael A. Fligner and Joseph S. Verducci. Multistage ranking models. *Journal of the American Statistical Association*, 83(403):892–901, 1988.

[133] Philippe Fontaine. Identification and economic behavior: Sympathy and empathy in historical perspective. *Economics and Philosophy*, 13(2):261–280, 1997.

[134] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3):75–174, 2010.

[135] James H. Fowler and Nicholas A. Christakis. Dynamic spread of happiness in a large social network: Longitudinal analysis over 20 years in the Framingham Heart Study. *British Medical Journal*, 337:a2338, 2008.

[136] James H. Fowler and Nicholas A. Christakis. Cooperative behavior cascades in human social networks. *Proceedings of the National Academy of Sciences*, 107(12):5334–5338, 2010.

[137] Ove Frank and David Strauss. Markov graphs. *Journal of the American Statistical Association*, 81(395):832–842, 1986.

[138] Linton C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, pages 35–41, 1977.

[139] John R.P. French, Jr. A formal theory of social power. *Psychological Review*, 63(3):181–194, 1956.

[140] Yoav Freund, Raj Iyer, Robert E. Schapire, and Yoram Singer. An efficient boosting algorithm for combining preferences. *Jouranl of Machine Learning Research*, 4:933–969, 2003.

[141] Ehud Friedgut, Gil Kalai, and Noam Nisan. Elections can be manipulated often. In *Proceedings of the 49th Annual IEEE Symposium on the Foundations of Computer Science (FOCS'08)*, pages 243–249, Philadelphia, 2008.

[142] Noah E. Friedkin and Eugene C. Johnsen. Social influence and opinions. *Journal of Mathematical Sociology*, 15(3-4):193–206, 1990.

[143] Noah E Friedkin and Eugene C Johnsen. Social influence networks and opinion change. *Advances in Group Processes*, 16(1):1–29, 1999.

[144] Nir Friedman and Daphne Koller. Being Bayesian about network structure. a bayesian approach to structure discovery in Bayesian networks. *Machine learning*, 50(1-2):95–125, 2003.

[145] D. Gale and L.S. Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.

[146] Douglas Gale and Shachar Kariv. Bayesian learning in social networks. *Games and Economic Behavior*, 45(2):329–346, 2003.

[147] Andrea Galeotti, Sanjeev Goyal, Matthew O. Jackson, Fernando Vega-Redondo, and Leeat Yariv. Network games. *The Review of Economic Studies*, 77(1):218–244, 2010.

[148] Mike Gartrell, Xinyu Xing, Qin Lv, Aaron Beach, Richard Han, Shivakant Mishra, and Karim Seada. Enhancing group recommendation by incorporating social relationship interactions. In *Proceedings of the 16th ACM International Conference on Supporting Group Work*, pages 97–106, 2010.

[149] Lisa Getoor, Nir Friedman, Daphne Koller, and Benjamin Taskar. Learning probabilistic models of link structure. *The Journal of Machine Learning Research*, 3:679–707, 2003.

[150] Lise Getoor and Christopher P. Diehl. Link mining: A survey. *ACM SIGKDD Explorations Newsletter*, 7(2):3–12, December 2005.

[151] Arpita Ghosh and Mohammad Mahdian. Externalities in online advertising. In *Proceedings of the 17th International Conference on World Wide Web*, WWW '08, pages 161–168, New York, NY, USA, 2008. ACM.

[152] Arpita Ghosh and Amin Sayedi. Expressive auctions for externalities in online advertising. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 371–380, New York, NY, USA, 2010. ACM.

[153] Allan Gibbard. Manipulation of voting schemes: a general result. *Econometrica: Journal of the Econometric Society*, pages 587–601, 1973.

[154] Malcolm Gladwell. *The Tipping Point: How Little Things can Make a Big Difference*. Little Brown, 2000.

[155] Sharad Goel, Duncan J. Watts, and Daniel G. Goldstein. The structure of online diffusion networks. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 623–638. ACM, 2012.

[156] Jennifer Golbeck and James Hendler. Filmtrust: Movie recommendations using trust in web-based social networks. In *Proceedings of the IEEE Consumer Communications and Networking Conference*, volume 96, pages 282–286, 2006.

[157] Benjamin Golub and Matthew O. Jackson. Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, 2(1):112–149, 2010.

[158] Benjamin Golub and Matthew O. Jackson. Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, 2(1):112–149, 2010.

[159] Benjamin Golub and Matthew O. Jackson. Using selection bias to explain the observed structure of internet diffusions. *Proceedings of the National Academy of Sciences*, 107(24):10833–10836, 2010.

[160] Benjamin Golub and Matthew O. Jackson. How homophily affects the speed of learning and best-response dynamics. *The Quarterly Journal of Economics*, 127(3):1287–1338, 2012.

[161] Benjamin Golub and Matthew O. Jackson. Network structure and the speed of learning: Measuring homophily based on its consequences. *Annals of Economics and Statistics*, 2012.

[162] Renato Gomes, Nicole Immorlica, and Evangelos Markakis. Externalities in keyword auctions: An empirical and theoretical assessment. In *Proceedings of the 5th International Workshop on Internet and Network Economics*, WINE '09, pages 172–183, Berlin, Heidelberg, 2009. Springer-Verlag.

[163] Manuel Gomez-Rodriguez, David Balduzzi, and Bernhard Schölkopf. Uncovering the temporal dynamics of diffusion networks. In *Proceedings of the 28th International Conference on Machine Learning*, pages 561–568, 2011.

[164] Manuel Gomez Rodriguez, Jure Leskovec, and Andreas Krause. Inferring networks of diffusion and influence. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '10, pages 1019–1028, New York, NY, USA, 2010. ACM.

[165] Steven M. Goodreau, James A. Kitts, and Martina Morris. Birds of a feather, or friend of a friend? using exponential random graph models to investigate adolescent social

networks*. *Demography*, 46(1):103–125, 2009.

[166] Isobel Claire Gormley and Thomas Brendan Murphy. A latent space model for rank data. In Edoardo M. Airoldi, David M. Blei, Stephen E. Fienberg, Anna Goldenberg, Eric P. Xing, and Alice X. Zheng, editors, *Statistical Network Analysis: Models, Issues, and New Directions*, pages 90–102. Springer, 2007.

[167] Isobel Claire Gormley and Thomas Brendan Murphy. Exploring voting blocs within the irish electorate. *Journal of the American Statistical Association*, 103(483):1014–1027, 2008.

[168] Sanjeev Goyal. *Connections : An introduction to the economics of networks*. Princeton University Press, Princeton, NJ, 2007.

[169] Sanjeev Goyal and Michael Kearns. Competitive contagion in networks. In *Proceedings of the 44th symposium on Theory of Computing*, STOC '12, pages 759–774, New York, NY, USA, 2012. ACM.

[170] Sanjeev Goyal, Marco J Van Der Leij, and José Luis Moraga-González. Economics: An emerging small world. *Journal of Political Economy*, 114(2):403–412, 2006.

[171] Mark S. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78(6):1360–1380, 1973.

[172] Jerrold W. Grossman. The evolution of the mathematical research collaboration graph. *Congressus Numerantium*, pages 201–212, 2002.

[173] John Guiver and Edward Snelson. Bayesian inference for plackett-luce ranking models. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 377–384. ACM, 2009.

[174] I.E. Hafalir. Stability of marriage with externalities. *International Journal of Game Theory*, 37(3):353–369, 2008.

[175] Nima Haghpanah, Nicole Immorlica, Vahab Mirrokni, and Kamesh Munagala. Optimal auctions with positive network externalities. In *Proceedings of the 12th ACM Conference on Electronic Commerce*, EC '11, pages 11–20, New York, NY, USA, 2011. ACM.

[176] Frank Harary. A criterion for unanimity in french's theory of social power. *In Dorwin Cartwright (Ed), Studies in Social Power*, pages 168–182, 1959.

[177] John C. Harsanyi. *Rational Behaviour and Bargaining Equilibrium in Games and Social Situations*. Cambridge University Press, 1977.

[178] Jason Hartline, Vahab Mirrokni, and Mukund Sundararajan. Optimal marketing strategies over social networks. In *Proceedings of the 17th International Conference on World Wide Web*, pages 189–198. ACM, 2008.

[179] Elaine Hatfield, John T. Cacioppo, and Richard L. Rapson. Emotional contagion. *Current*

Directions in Psychological Science, 2(3):96–99, 1993.

[180] Edith Hemaspaandra, Lane A Hemaspaandra, and Jörg Rothe. Exact analysis of dodgson elections: Lewis carroll's 1876 voting system is complete for parallel access to np. *Journal of the ACM (JACM)*, 44(6):806–825, 1997.

[181] Robert J. Henery. Permutation probabilities for gamma random variables. *Journal of Applied Probability*, pages 822–834, 1983.

[182] Marc Henry and Ismael Mourifié. Euclidean revealed preferences: Testing the spatial voting model. *Journal of Applied Econometrics*, 2011.

[183] Peter D. Hoff, Adrian E. Raftery, and Mark S. Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098, 2002.

[184] Paul W. Holland and Samuel Leinhardt. An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, 76(373):33–50, 1981.

[185] Petter Holme and Beom Jun Kim. Growing scale-free networks with tunable clustering. *Phys. Rev. E*, 65:026107, Jan 2002.

[186] Matthew O. Jackson. A survey of network formation models: Stability and efficiency. *Group Formation in Economics: Networks, Clubs and Coalitions*, pages 11–57, 2005.

[187] Matthew O. Jackson. *Social and Economic Networks*. Princeton University Press, 2008.

[188] Matthew O. Jackson and Dunia Lopez-Pintado. Diffusion and contagion in networks with heterogeneous agents and homophily. *Network Science*, 1(01):49–67, 2013.

[189] Matthew O. Jackson and Leeat Yariv. Social networks and the diffusion of behavior. *Yale Economic Review*, 3(2), 2006.

[190] Matthew O. Jackson and Leeat Yariv. Diffusion of behavior and equilibrium properties in network games. *The American Economic Review*, 97(2):92–98, 2007.

[191] Mohsen Jamali and Martin Ester. A matrix factorization technique with trust propagation for recommendation in social networks. In *Proceedings of the 4th ACM Conference on Recommender Systems (Recsys-10)*, pages 135–142. ACM, 2010.

[192] S Jespersen and Alexander Blumen. Small-world networks: Links with long-tailed distributions. *Physical Review E*, 62(5):6270, 2000.

[193] Meir Kalech, Sarit Kraus, Gal A. Kaminka, and Claudia V. Goldman. Practical voting rules with partial information. *Journal of Autonomous Agents and Multi-Agent Systems*, 22(1):151–182, 2011.

[194] Amin Karbasi, Stratis Ioannidis, and Laurent Massoulié. Content search through comparisons. In *Automata, Languages and Programming*, pages 601–612. Springer, 2011.

[195] Amin Karbasi, Stratis Ioannidis, and Laurent Massoulié. Hot or not: Interactive content search using comparisons. In *Information Theory and Applications Workshop (ITA), 2012*, pages 291–297. IEEE, 2012.

[196] E. Katz and P.F. Lazarsfeld. *Personal Influence: The Part Played by People in the Flow of Mass Communications*. Free Press, 1955.

[197] Leo Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, 1953.

[198] Michael Kearns, Michael L. Littman, and Satinder Singh. Graphical models for game theory. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, pages 253–260, 2001.

[199] John G. Kemeny and James L. Snell. *Finite Markov Chains*, volume 210. Springer-Verlag New York, 1976.

[200] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 137–146. ACM, 2003.

[201] David Kempe, Jon Kleinberg, and Éva Tardos. Influential nodes in a diffusion model for social networks. In *Automata, languages and programming*, pages 1127–1138. Springer, 2005.

[202] David Kempe and Mohammad Mahdian. A cascade model for externalities in sponsored search. In *Proceedings of the 4th International Workshop on Internet and Network Economics*, WINE '08, pages 585–596, Berlin, Heidelberg, 2008. Springer-Verlag.

[203] Myunghwan Kim and Jure Leskovec. Multiplicative attribute graph model of real-world networks. In *Algorithms and Models for the Web-Graph*, Lecture Notes in Computer Science, pages 62–73. Springer Berlin Heidelberg, 2010.

[204] Myunghwan Kim and Jure Leskovec. Modeling social networks with node attributes using the multiplicative attribute graph model. In *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence (UAI-11)*, pages 400–409, 2011.

[205] Myunghwan Kim and Jure Leskovec. Latent multi-group membership graph model. In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, pages 1719–1726, 2012.

[206] Myunghwan Kim and Jure Leskovec. Multiplicative attribute graph model of real-world networks. *Internet Mathematics*, 8(1-2):113–160, 2012.

[207] Alan Kirman. Ants, rationality, and recruitment. *The Quarterly Journal of Economics*, 108(1):137–156, 1993.

[208] Alan Kirman and Miriam Teschl. Selfish or selfless? The role of empathy in economics.

Philosophical Transactions of the Royal Society B: Biological Sciences, 365(1538):303–317, 2010.

[209] Bettina Klaus and Flip Klijn. Paths to stability for matching markets with couples. *Games and Economic Behavior*, 58(1):154–171, January 2007.

[210] Bettina Klaus, Flip Klijn, and Toshifumi Nakamura. Corrigendum to "stable matchings and preferences of couples" [j. econ. theory 121 (1) (2005) 75-106]. *Journal of Economic Theory*, 144(5):2227–2233, September 2009.

[211] Jon Kleinberg, Christos Papadimitriou, and Prabhakar Raghavan. Segmentation problems. *Journal of the ACM*, 51:263–280, 2004.

[212] Jon Kleinberg and Eva Tardos. Approximation algorithms for classification problems with pairwise relationships: Metric labeling and Markov random fields. *Journal of the ACM (JACM)*, 49(5):616–639, 2002.

[213] Jon M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, 1999.

[214] Jon M. Kleinberg. Navigation in a small world. *Nature*, 406(6798):845–845, 2000.

[215] Fuhito Kojima and Parag A. Pathak. Incentives and stability in large two-sided matching markets. *American Economic Review*, 99(3):608–27, June 2009.

[216] Daphne Kollar and Nir Friedman. *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press, 2009.

[217] Kathrin Konczak and Jérôme Lang. Voting procedures with incomplete preferences. In *IJCAI-05 Workshop on Advances in Preference Handling*, pages 124–129, Edinburgh, 2005.

[218] Sara H. Konrath, Edward H. O'Brien, and Courtney Hsing. Changes in dispositional empathy in american college students over time: A meta-analysis. *Personality and Social Psychology Review*, 15(2):180–198, 2011.

[219] Ioannis Konstas, Vassilios Stathopoulos, and Joemon M. Jose. On social networks and collaborative recommendation. In *Proceedings of the 32Nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 195–202, 2009.

[220] Yehuda Koren and Robert Bell. Advances in collaborative filtering. In Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor, editors, *Recommender Systems Handbook*, pages 145–186. Springer US, 2011.

[221] Mirjam Kretzschmar and Martina Morris. Measures of concurrency in networks and the spread of infectious disease. *Mathematical biosciences*, 133(2):165–195, 1996.

[222] Dmitri Krioukov, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat, and Marián Boguñá. Hyperbolic geometry of complex networks. *Physical Review E*, 82(3):036106,

2010.

[223] Piotr Krysta, Tomasz Michalak, Tuomas Sandholm, and Michael Wooldridge. Combinatorial auctions with externalities. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems*, AAMAS '10, pages 1471–1472, Richland, SC, 2010. International Foundation for Autonomous Agents and Multiagent Systems.

[224] Jérôme Lang. Vote and aggregation in combinatorial domains with structured preferences. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI-07)*, pages 1366–1371, Hyderabad, India, 2007.

[225] Amy N. Langville and Carl D. Meyer. *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton University Press, 2009.

[226] P.F. Lazarsfeld, B. Berelson, and H. Gaudet. *The People's Choice: How the Voter Makes Up His Mind in a Presidential Campaign*. Columbia University Press, 1948.

[227] Paul H. Lee and Philip L.H. Yu. Mixtures of weighted distance-based models for ranking data with applications in political studies. *Computational Statistics & Data Analysis*, 56(8):2486 – 2500, 2012.

[228] Harvey Leibenstein. *Beyond Economic Man: A New Foundation for Microeconomics*. Harvard University Press, 1976.

[229] Jure Leskovec, Lada A. Adamic, and Bernardo A. Huberman. The dynamics of viral marketing. *ACM Transactions on the Web (TWEB)*, 1(1):5, 2007.

[230] Jure Leskovec, Deepayan Chakrabarti, Jon Kleinberg, Christos Faloutsos, and Zoubin Ghahramani. Kronecker graphs: An approach to modeling networks. *The Journal of Machine Learning Research*, 11:985–1042, 2010.

[231] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. Predicting positive and negative links in online social networks. In *Proceedings of the 19th International Conference on World Wide Web*, pages 641–650. ACM, 2010.

[232] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graphs over time: Densification laws, shrinking diameters and possible explanations. In *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, KDD '05, pages 177–187, New York, NY, USA, 2005. ACM.

[233] Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, and Natalie Glance. Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 420–429. ACM, 2007.

[234] Kevin Lewis, Marco Gonzalez, and Jason Kaufman. Social selection and peer influence in an online social network. *Proceedings of the National Academy of Sciences*, 109(1):68–72, 2012.

[235] Hang Li. Learning to rank for information retrieval and natural language processing. *Synthesis Lectures on Human Language Technologies*, 4(1):1–113, 2011.

[236] Hang Li. A short introduction to learning to rank. *IEEE TRANSACTIONS on Information and Systems*, 94(10):1854–1862, 2011.

[237] David Liben-Nowell and Jon Kleinberg. The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology*, 58(7):1019–1031, 2007.

[238] David Liben-Nowell and Jon Kleinberg. Tracing information flow on a global scale using internet chain-letter data. *Proceedings of the National Academy of Sciences*, 105(12):4633–4638, 2008.

[239] Greg Linden, Brent Smith, and Jeremy York. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1):76–80, 2003.

[240] Nick Littlestone and Manfred K. Warmuth. The weighted majority algorithm. In *The 30th Annual Symposium on Foundations of Computer Science*, pages 256–261. IEEE, 1989.

[241] Nathan N. Liu and Qiang Yang. Eigenrank: A ranking-oriented approach to collaborative filtering. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, pages 83–90, New York, NY, USA, 2008. ACM.

[242] Tie-Yan Liu. Learning to rank for information retrieval. *Foundations and Trends in Information Retrieval*, 3(3):225–331, 2009.

[243] Ilan Lobel and Evan Sadler. Preferences, homophily, and social learning. *working paper*, 2013.

[244] Linyuan Lü and Tao Zhou. Link prediction in complex networks: A survey. *Physica A: Statistical Mechanics and its Applications*, 390(6):1150–1170, 2011.

[245] Tyler Lu and Craig Boutilier. Budgeted social choice: From consensus to personalized decision making. In *Proceedings of the Twenty-second International Joint Conference on Artificial Intelligence (IJCAI-11)*, pages 280–286, Barcelona, 2011.

[246] Tyler Lu and Craig Boutilier. Learning mallows models with pairwise preferences. In *Proceedings of the 28th International Conference on Machine Learning (ICML-12)*, pages 145–152, 2011.

[247] Tyler Lu and Craig Boutilier. Robust approximation and incremental elicitation in voting protocols. In *Proceedings of the Twenty-second International Joint Conference on Artificial Intelligence (IJCAI-11)*, pages 287–293, Barcelona, 2011.

[248] Tyler Lu and Craig Boutilier. Vote elicitation with probabilistic preference models: Empirical estimation and cost tradeoffs. In *Proceedings of the Second International Confer-*

ence on Algorithmic Decision Theory (ADT-11), pages 135–149, Piscataway, NJ, 2011.

[249] Tyler Lu and Craig Boutilier. Multi-winner social choice with incomplete preferences. In *Proceedings of the Twenty-third International Joint Conference on Artificial Intelligence (IJCAI-13)*, Beijing, 2013.

[250] Tyler Lu, Pingzhong Tang, Ariel D. Procaccia, and Craig Boutilier. Bayesian vote manipulation: Optimal strategies and impact on welfare. In *Proceedings of the Twenty-eighth Conference on Uncertainty in Artificial Intelligence (UAI-12)*, Catalina, CA, 2012.

[251] Robert Duncan Luce. *Individual Choice Behavior: A Theoretical Analysis*. Wiley, 1959.

[252] Hao Ma, Irwin King, and Michael R. Lyu. Learning to recommend with social trust ensemble. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 203–210, 2009.

[253] Hao Ma, Haixuan Yang, Michael R. Lyu, and Irwin King. Sorec: Social recommendation using probabilistic matrix factorization. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, pages 931–940, 2008.

[254] Fabio Maccheronia, Massimo Marinaccia, and Aldo Rustichini. Social decision theory: Choosing within and between groups. *The Review of Economic Studies*, 79(4):1591–1636, 2012.

[255] Sofus A. Macskassy. Using graph-based metrics with empirical risk minimization to speed up active learning on networked data. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 597–606. ACM, 2009.

[256] Mohammad Mahdian and Ying Xu. Stochastic kronecker graphs. In *Algorithms and Models for the Web-Graph*, Lecture Notes in Computer Science, pages 179–186. Springer Berlin Heidelberg, 2007.

[257] D.P. Maki and M. Thompson. *Mathematical Models and Applications: with Emphasis on the Social, Life, and Management Sciences*. Prentice-Hall, 1973.

[258] Colin L. Mallows. Non-null ranking models. In *Biometrika*, volume 44, pages 114–130, 1957.

[259] John I. Marden. *Analyzing and Modeling Rank Data*. Chapman and Hall, London, 1995.

[260] Paolo Massa and Paolo Avesani. Controversial users demand local trust metrics: An experimental study on epinions.com community. In *Proceedings of the 20th National Conference on Artificial Intelligence*, AAAI'05, pages 121–126, 2005.

[261] Judith Masthoff. Group modeling: Selecting a sequence of television items to suit a group of viewers. *User Modeling and User-Adapted Interaction*, 14(1):37–85, 2004.

[262] Judith Masthoff and Albert Gatt. In pursuit of satisfaction and the prevention of em-

barrassment: Affective state in group recommender systems. *User Modeling and User-Adapted Interaction*, 16(3-4):281–319, 2006.

[263] Kenneth O May. A set of independent necessary and sufficient conditions for simple majority decision. *Econometrica: Journal of the Econometric Society*, pages 680–684, 1952.

[264] Joseph F McCarthy and Theodore D Anagnost. Musicfx: an arbiter of group preferences for computer supported collaborative workouts. In *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work*, pages 363–372, 1998.

[265] Daniel McFadden. Econometric models for probabilistic choice among products. *Journal of Business*, pages S13–S29, 1980.

[266] Miller McPherson, Lynn Smith-Lovin, and James M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, pages 415–444, 2001.

[267] Marina Meila and Harr Chen. Dirichlet process mixtures of generalized mallows models. In *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, pages 358–367, 2010.

[268] Carl D. Meyer. *Matrix Analysis and Applied Linear Algebra*, volume 2. SIAM: Society for Industrial and Applied Mathematics, 2000.

[269] Stanley Milgram. The small world problem. *Psychology Today*, 2(1):60–67, 1967.

[270] Andriy Mnih and Ruslan Salakhutdinov. Probabilistic matrix factorization. In *Advances in Neural Information Processing Systems*, pages 1257–1264, 2007.

[271] Yamir Moreno, Maziar Nekovee, and Amalio F. Pacheco. Dynamics of rumor spreading in complex networks. *Physical Review E*, 69:066130, June 2004.

[272] Martina Morris. Epidemiology and social networks: Modeling structured diffusion. *Sociological Methods & Research*, 22(1):99–126, 1993.

[273] Stephen Morris. Contagion. *The Review of Economic Studies*, 67(1):57–78, 2000.

[274] Elchanan Mossel and Sebastien Roch. On the submodularity of influence in social networks. In *Proceedings of the Thirty-ninth Annual ACM Symposium on Theory of Computing*, STOC '07, pages 128–134, New York, NY, USA, 2007. ACM.

[275] Hervé Moulin. On strategy-proofness and single peakedness. *Public Choice*, 35(4):437–455, 1980.

[276] Thomas Brendan Murphy and Donal Martin. Mixtures of distance-based models for ranking data. *Computational Statistics & Data Analysis*, 41(3):645–655, 2003.

[277] Diana C. Mutz. Cross-cutting social networks: Testing democratic theory in practice. *American Political Science Review*, 96(1):111–126, 2002.

[278] Seth Myers and Jure Leskovec. On the convexity of latent social network inference. *Advances in Neural Information Processing Systems*, pages 1741–1749, 2010.

[279] Mark Newman. *Networks: An Introduction*. Oxford University Press, 2010.

[280] Mark Newman, Albert-László Barabási, and Duncan J. Watts. *The Structure and Dynamics of Networks*. Princeton University Press, 2006.

[281] Mark E.J. Newman. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2):404–409, 2001.

[282] Mark E.J. Newman. Assortative mixing in networks. *Physical Review Letters*, 89(20):208701, 2002.

[283] Mark E.J. Newman. Spread of epidemic disease on networks. *Physical review E*, 66(1):016128, 2002.

[284] Mark E.J. Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.

[285] Mark E.J. Newman. Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National academy of Sciences of the United States of America*, 101:5200–5205, 2004.

[286] Mark E.J. Newman. A measure of betweenness centrality based on random walks. *Social Networks*, 27(1):39–54, 2005.

[287] Mark E.J. Newman and Duncan J. Watts. Scaling and percolation in the small-world network model. *Physical Review E*, 60(6):7332, 1999.

[288] Joel Oren, Yuval Filmus, and Craig Boutilier. Efficient vote elicitation under candidate uncertainty. In *Proceedings of the Twenty-third International Joint Conference on Artificial Intelligence (IJCAI-13)*, Beijing, 2013.

[289] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, November 1999.

[290] Fragkiskos Papadopoulos, Maksim Kitsak, M Ángeles Serrano, Marián Boguñá, and Dmitri Krioukov. Popularity versus similarity in growing networks. *Nature*, 489(7417):537–540, 2012.

[291] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic dynamics and endemic states in complex networks. *Physical Review E*, 63(6):066117, 2001.

[292] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. *Physical Review Letters*, 86(14):3200–3203, 2001.

[293] Romualdo Pastor-Satorras and Alessandro Vespignani. Immunization of complex networks. *Physical Review E*, 65(3):036104, 2002.

[294] Prasanta K Pattanaik. On the stability of sincere voting situations. *Journal of Economic Theory*, 6(6):558–574, 1973.

[295] Mathew Penrose. *Random Geometric Graphs*. Oxford University Press, USA, 2003.

[296] Robin L. Plackett. The analysis of permutations. *Applied Statistics*, pages 193–202, 1975.

[297] Keith T. Poole. *Spatial Models of Parliamentary Voting*. Cambridge University Press, 2005.

[298] Keith T. Poole and Howard Rosenthal. A spatial model for legislative roll call analysis. *American Journal of Political Science*, pages 357–384, 1985.

[299] Richard F. Potthoff and Steven J. Brams. Proportional representation: Broadening the options. *Journal of Theoretical Politics*, 10(2):147–178, 1998.

[300] Areil D. Procaccia and Jeffrey S. Rosenschein. Junta distributions and the average-case complexity of manipulating elections. *Journal of Artificial Intelligence Research*, 28:157–181, 2007.

[301] Lara Quijano-Sanchez, Juan A Recio-Garcia, Belen Diaz-Agudo, and Guillermo Jimenez-Diaz. Social factors in group recommender systems. *ACM Transactions on Intelligent Systems and Technology*, 4(1):8, 2013.

[302] Anatol Rapoport. Spread of information through a population with socio-structural bias: I. assumption of transitivity. *The Bulletin of Mathematical Biophysics*, 15(4):523–533, 1953.

[303] Sidney Redner. How popular is your paper? an empirical study of the citation distribution. *The European Physical Journal B-Condensed Matter and Complex Systems*, 4(2):131–134, 1998.

[304] Matthew Richardson and Pedro Domingos. Mining knowledge-sharing sites for viral marketing. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 61–70. ACM, 2002.

[305] Garry Robins, Tom Snijders, Peng Wang, Mark Handcock, and Philippa Pattison. Recent developments in exponential random graph (p*) models for social networks. *Social networks*, 29(2):192–215, 2007.

[306] Everett M. Rogers. *Diffusion of Innovations, 5th Edition*. Free Press, 2003.

[307] A.E. Roth and M.A.O. Sotomayor. *Two-sided matching: A study in game-theoretic modeling and analysis*, volume 18. Cambridge University Press, 1990.

[308] Alvin E. Roth. The evolution of the labor market for medical interns and residents: A case study in game theory. *Journal of Political Economy*, 92:991–1016, 1984.

[309] Alvin E. Roth and Elliott Peranson. The redesign of the matching market for american physicians: Some engineering aspects of economic design. *American Economic Review*,

89(4):748–780, September 1999.

[310] Yousef Saad. *Iterative Methods for Sparse Linear Systems*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2nd edition, 2003.

[311] Amirali Salehi-Abari and Craig Boutilier. Ranking networks. In *NIPS-13 Workshop on Frontiers of Network Analysis: Methods, Models, and Applications*, 2013.

[312] Amirali Salehi-Abari and Craig Boutilier. Empathetic social choice on social networks. In *Proceedings of The 13th International Conference on Autonomous Agents and Multiagent Systems*, pages 693–700, 2014.

[313] Amirali Salehi-Abari and Craig Boutilier. Preference-oriented social networks: Group recommendation and inference. In *Proceedings of the 9th ACM Conference on Recommender Systems (RecSys-15)*, 2015.

[314] Hiroo Sasaki and Manabu Toda. Two-sided matching problems with externalities. *Journal of Economic Theory*, 70(1):93 – 108, 1996.

[315] Mark Allen Satterthwaite. Strategy-proofness and arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of economic theory*, 10(2):187–217, 1975.

[316] Barry Schwartz. *The Paradox of Choice: Why More is Less*. Harper Collins, 2004.

[317] John P. Scott. *Social Network Analysis: A Handbook*. SAGE Publications, 2000.

[318] Shunichi Seko, Takashi Yagi, Manabu Motegi, and Shinyo Muto. Group recommendation using feature space representing behavioral tendency and power balance among members. In *Proceedings of the Fifth ACM Conference on Recommender Systems*, pages 101–108, 2011.

[319] Amartya K. Sen. *Collective Choice and Social Welfare*. North-Holland Publishing Co., Amsterdam, 1970.

[320] Amartya Kumar Sen. *Collective Choice and Social Welfare*. North-Holand, 1970.

[321] Prithviraj Sen, Galileo Mark Namata, Mustafa Bilgic, Lise Getoor, Brian Gallagher, and Tina Eliassi-Rad. Collective classification in network data. *AI Magazine*, 29(3):93–106, 2008.

[322] M. Ángeles Serrano, Dmitri Krioukov, and Marián Boguñá. Self-similarity of complex networks and hidden metric spaces. *Physical Review Letters*, 100:078701, Feb 2008.

[323] Ma Angeles Serrano and Marián Boguná. Topology of the world trade web. *Physical Review E*, 68(1):015101, 2003.

[324] Burr Settles. Active learning literature survey. Technical Report Technical Report 1648, Department of Computer Science, University of Wisconsin-Madison, 2010.

[325] Bracha Shapira. *Recommender Systems Handbook*. Springer, 2011.

[326] George Silverman. *The Secrets of Word-of-mouth Marketing: How to Trigger Exponential Sales Through Runaway Word of Mouth*. AMACOM, 2001.

[327] Tania Singer. The neuronal basis and ontogeny of empathy and mind reading: Review of literature and implications for future research. *Neuroscience and Biobehavioral Reviews*, 30(6):855 – 863, 2006.

[328] B. Babington Smith. Discussion of professor ross's paper. *Journal of the Royal Statistical Society*, 12(1):41–59, 1950.

[329] Anand E. Sokhey and Scott D. McClurg. Social networks and correct voting. Preprint, 2008.

[330] Hal Stern. Models for distributions on permutations. *Journal of the American Statistical Association*, 85(410):558–564, 1990.

[331] David Strang and Sarah A. Soule. Diffusion in organizations and social movements: From hybrid corn to poison pills. *Annual Review of Sociology*, pages 265–290, 1998.

[332] Pingzhong Tang and Tuomas Sandholm. Optimal auctions for spiteful bidders. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, AAAI'12, 2012.

[333] Ben Taskar, Ming-Fai Wong, Pieter Abbeel, and Daphne Koller. Link prediction in relational data. In *Advances in Neural Information Processing Systems 16*. MIT Press, Cambridge, MA, 2004.

[334] Michael Taylor, John Guiver, Stephen Robertson, and Tom Minka. Softrank: Optimizing non-smooth rank metrics. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, WSDM '08, pages 77–86, New York, NY, USA, 2008. ACM.

[335] Louis L. Thurstone. A law of comparative judgment. *Psychological Review*, 34(4):273–286, 1927.

[336] Craig A Tovey. Some foundations for empirical study in the Euclidean spatial model of social choice. In *Political economy: institutions, competition, and representation: proceedings of the Seventh International Symposium in Economic Theory and Econometrics*, page 175. Cambridge University Press, 1993.

[337] Alan Tsang and Kate Larson. The echo chamber: Strategic voting and homophily in social networks. In *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-16)*, 2016.

[338] Johan Ugander, Brian Karrer, Lars Backstrom, and Cameron Marlow. The anatomy of the facebook social graph. *arXiv preprint arXiv:1111.4503*, 2011.

[339] Rian A.W. van Blokland-Vogelesang. Unfolding and consensus ranking: A prestige ladder for technical occupations. *New Developments in Psychological Choice Modeling*, 60:237–

258, 1989.

[340] Rian A.W. van Blokland-Vogelesang. A nonparametric distance model for unidimensional unfolding. In *Probability Models and Statistical Analyses for Ranking Data*, pages 241–276. Springer, 1993.

[341] Shankar Vembu and Thomas Gärtner. Label ranking algorithms: A survey. In *Preference Learning*, pages 45–64. Springer, 2011.

[342] Maksims Volkovs and Richard S. Zemel. Collaborative ranking with 17 parameters. In *Advances in Neural Information Processing Systems*, pages 2303–2311, 2012.

[343] Maksims N. Volkovs and Richard S. Zemel. Boltzrank: Learning to maximize expected ranking gain. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, pages 1089–1096, New York, NY, USA, 2009. ACM.

[344] Maksims N. Volkovs and Richard S. Zemel. A flexible generative model for preference aggregation. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12, pages 479–488, New York, NY, USA, 2012. ACM.

[345] Joan Walker and Moshe Ben-Akiva. Generalized random utility model. *Mathematical Social Sciences*, 43(3):303–343, 2002.

[346] Toby Walsh. Where are the really hard manipulation problems? the phase transition in manipulating the veto rule. In *Proceedings of the Twenty-first International Joint Conference on Artificial Intelligence (IJCAI-09)*, pages 324–329, Pasadena, California, 2009.

[347] Stanley Wasserman. *Social Network Analysis: Methods and Applications*, volume 8. Cambridge university press, 1994.

[348] Stanley Wasserman and Philippa Pattison. Logit models and logistic regressions for social networks: I. an introduction to Markov graphs and p*. *Psychometrika*, 61(3):401–425, 1996.

[349] Duncan J. Watts. *Small Worlds: The Dynamics of Networks between Order and Randomness*. Princeton University Press, 1999.

[350] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393(6684):440–442, 1998.

[351] Bernard M. Waxman. Routing of multipoint connections. *Selected Areas in Communications, IEEE Journal on*, 6(9):1617–1622, December 1988.

[352] Markus Weimer, Alexandros Karatzoglou, Quoc V. Le, and Alex J. Smola. Cofi rank - maximum margin matrix factorization for collaborative ranking. In *Proceedings of the Twenty-First Annual Conference on Neural Information Processing Systems*, 2007.

[353] Alistair Wilson, Mariagiovanna Baccara, Ayse Imrohoroglu, and Leeat Yariv. A field

study on matching with network externalities. In *Proceedings of the Behavioral and Quantitative Game Theory: Conference on Future Directions*, BQGT '10, pages 96:1–96:1, New York, NY, USA, 2010. ACM.

[354] Felix Ming Fai Wong and Peter Marbach. Who are your friends?—a simple mechanism that achieves perfect network formation. In *Proceedings IEEE INFOCOM*, pages 566–570. IEEE, 2011.

[355] Felix Ming Fai Wong and Peter Marbach. Cosine-neighbourhood-refinement: Towards a robust network formation mechanism. In *Proceedings IEEE INFOCOM*, pages 1710–1718. IEEE, 2012.

[356] Felix Ming Fai Wong and Peter Marbach. Similarity-based network formation. In *Proceedings 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 394–399. IEEE, 2012.

[357] Lirong Xia and Vincent Conitzer. Determining possible and necessary winners under common voting rules given partial orders. In *Proceedings of the Twenty-third AAAI Conference on Artificial Intelligence (AAAI-08)*, pages 202–207, Chicago, 2008.

[358] Lirong Xia, Michael Zuckerman, Ariel D. Procaccia, Vincent Conitzer, and Jeffrey S. Rosenschein. Complexity of unweighted coalitional manipulation under some common voting rules. In *Proceedings of the 21st international jont conference on Artifical intelligence*, IJCAI'09, pages 348–353, San Francisco, CA, USA, 2009.

[359] Yang Xiwang, Yang Guo, Yong Liu, and Harald Steck. A survey of collaborative filtering based social recommender systems. *Computer Communications*, 41:1–10, 2014.

[360] Jun Xu and Hang Li. Adarank: A boosting algorithm for information retrieval. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '07, pages 391–398, New York, NY, USA, 2007.

[361] Xiwang Yang, Harald Steck, and Yong Liu. Circle-based recommendation in online social networks. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1267–1275, 2012.

[362] Leeat Yariv and Matthew O. Jackson. Diffusion, strategic interaction, and social structure. *Handbook of Social Economics*, 2008.

[363] H Peyton Young. An axiomatization of borda's rule. *Journal of Economic Theory*, 9(1):43–52, 1974.

[364] H Peyton Young. Condorcet's theory of voting. *The American Political Science Review*, pages 1231–1244, 1988.

[365] H Peyton Young. The dynamics of social innovation. *Proceedings of the National Academy of Sciences*, 108(Supplement 4):21285–21291, 2011.

[366] Peyton Young. Optimal voting rules. *The Journal of Economic Perspectives*, 9(1):51–64, 1995.

[367] Stephen J. Young and Edward R. Scheinerman. Random dot product graph models for social networks. In *Proceedings of the 5th International Conference on Algorithms and Models for the Web-graph*, WAW'07, pages 138–149. Springer-Verlag, 2007.

[368] Zhiwen Yu, Xingshe Zhou, Yanbin Hao, and Jianhua Gu. Tv program recommendation for multiple viewers based on user profile merging. *User Modeling and User-Adapted Interaction*, 16(1):63–82, 2006.

[369] Lilin Zhang and Peter Marbach. "two is a crowd"–optimal trend adoption in social networks. In *Proceedings of International Conference on Game Theory for Networks*, pages 608–627. Springer, 2012.

[370] Tao Zhou, Linyuan Lü, and Yi-Cheng Zhang. Predicting missing links via local information. *The European Physical Journal B*, 71(4):623–630, 2009.

[371] Xiaojin Zhu, John Lafferty, and Zoubin Ghahramani. Combining active learning and semi-supervised learning using gaussian fields and harmonic functions. In *Workshop on the Continuum from Labeled to Unlabeled Data in Machine Learning and Data Mining*, pages 58–65, 2003.