# Empathetic Social Choice on Social Networks

Amirali Salehi-Abari
Department of Computer Science
University of Toronto
Ontario, Canada
abari@cs.toronto.edu

Craig Boutilier
Department of Computer Science
University of Toronto
Ontario, Canada
cebly@cs.toronto.edu

## ABSTRACT

Social networks play a central role in individual interactions and decision making. While it is recognized that networks can correlate behaviors and preferences among connected agents, relatively little work has considered mechanisms for *social choice* on such networks. We introduce a model for social choice—specifically, consensus decision making—on social networks that reflects *dependence* among the utilities of connected agents. We define an *empathetic social choice framework* in which agents derive utility based on both their own intrinsic preferences and the satisfaction of their neighbors. We translate this problem into a weighted form of classical preference aggregation (e.g., social welfare maximization or voting), and develop scalable optimization algorithms for this task. Empirical results validate the effectiveness of our methods and the value of empathetic preferences.

## Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence,Multiagent Systems; J.4 [**Social and Behavioral Sciences**]: Economics and Sociology

## General Terms

Algorithms, Economics, Human Factors, Theory

## Keywords

Social Choice, Social and Economic Networks, Voting, Preference Aggregation, Empathy, Decision Making.

## 1. INTRODUCTION

Social and economic networks play a fundamental role in facilitating interactions and behaviours between individuals, businesses, and organizations. It is widely acknowledged that the behaviors, and to a lesser extent preferences, of individuals connected in a social network are correlated in ways that can be explained, in part, by network structure [13, 16, 24]. Because of this, and the increasing availability of data that allows one to infer such relationships, the study of *social choice* (group decision) problems on social networks is one of tremendous practical import. Arguably most group decision problems, whether social, corporate, or policy-oriented, involve people at least some of whom are linked via myriad social ties. These ties may provide strong clues as to the preferences of individuals, which can then be used to facilitate preference aggregation and implement a social choice function (or make a group decision).

Despite these natural connections, social choice within social networks has received, until recently, relatively little attention. Recent work has examined, for example, the formation of (hedonic) coalitions on social networks [12, 11], social network games [38], coalition structure generation [40] and stable matching on social networks [7, 4]. The influence of social networks on voting behavior has received attention in the social sciences [2, 33, 10], and the emergence of online social networks has spawned research on mechanisms for vote delegation [8].

This paper considers the problem of *consensus decision making (or group recommendation) on social networks*, such as voting over some option space. Specifically, we wish to select a single option from a set of alternatives for some group of individuals 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 neighbors. 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 [18]. This inherent interdependency of agent utilities is captured in the economic literature on empathy, envy, and other forms of "other regarding" preferences [30, 26, 6]; our models also have ties to work on opinion spread and social learning [24, 1] (see Sec. 2.4).

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 neighbors for $a$, where the weight given to $j$'s preference depends on the strength of the relationship of $i$ with $j$. For instance, $i$ may be willing to trade off some of her intrinsic preference for a restaurant if her colleagues are happier with the cuisine (as her happiness depends to those of her friends); and she defers more to her closer friends. In our *global empathetic model*, $i$'s utility for each $a$ depends on her intrinsic preference and the *total utility* of her neighbors for $a$ (not just their intrinsic prefer-

**Appears in:** *Alessio Lomuscio, Paul Scerri, Ana Bazzan, and Michael Huhns (eds.), Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), May 5-9, 2014, Paris, France.*

ence): she wants her neighbors not only to be satisfied with $a$, but to have high utility, which depends on the utility of *their* neighbors, 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 neighbors, their neighbors, and others in the community, she may prefer to see $b$ elected rather than have grumpy neighbors for the next few years.

Our main contribution is a model for preference aggregation that selects consensus alternatives in a way that is sensitive to both intrinsic and empathetic preferences. We do not require agents to *compute* such combined preferences; indeed, they need not even have knowledge of their neighbors' preferences. Instead, agents specify only their preferences for options and the extent to which they care about their neighbors' satisfaction (the latter potentially estimated from social network structure). We describe methods for computing optimal options under the local and global models. 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 co-dependent—indeed, utility spreads much like PageRank values [34]—requires the solution of a linear system to determine the optimal (fixed-point) option. We describe (mild) conditions under which such fixed points exist, and show that it too results in a form of weighted voting. Experiments demonstrate the effectiveness of our algorithms and show that, in some settings, ignoring empathetic preferences results in suboptimal decisions and high social welfare loss.[1]

# 2. SOCIAL EMPATHETIC MODEL

We outline our basic social choice model, describe our empathetic models, and discuss related work.

## 2.1 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 $\mathcal{A} = \{a_1, \dots, a_m\}$ and a set of agents $\mathcal{N} = \{1, \dots, n\}$. 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 $u_j^I$. 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 preference and $f$ is (utilitarian) *social welfare maximization (SWM)*, we select $a^* = \arg\max \sum_j u_j^I(a)$. We ignore ties in the SWM option for ease of exposition. If preferences are given by preference rankings, $f$ might correspond to some voting rule.[2]

## 2.2 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.[3]

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 preference, 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 $\sum_k w_{jk} = 1$ for any $j$. 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. 1(a) for an illustration.

We consider pure consensus social choice scenarios in which a single option $a$ is selected. 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 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

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

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 $w_{jj} > 0$):

$$u_j(a) = \sum_k w_{jk} u_k^I(a). \tag{1}$$

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

$$u_j(a) = w_{jj} u_j^I(a) + \sum_{k \neq j} w_{jk} u_k(a). \tag{2}$$

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

[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 [28].

[2] Our model applies directly to more general social choice problems, such as assignment problems with network externalities, matching, etc., without difficulty. Our algorithms, however, are specific to the "single-choice" assumption.

[3] Suitable *qualititative* expression of such tradeoffs is an important ongoing research direction.

global model, the circular dependence of utilities requires a fixed point solution to the linear system Eq. 2 (see below).

Correlations of behavior and/or preferences among agents connected in social network is widely accepted, and can be explained by a variety of mechanisms [16, 24]. Among these are: *information diffusion*, in which agents become aware of opportunities or innovations from connections to their neighbors; *network externalities*, in which the benefits of adopting some behavior increase when more neighbors do the same; or *homophily*, in which people with similar characteristics (say, preferences) more readily form social ties. 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.

## 2.3 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. 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).

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. 1, $\mathbf{u}(a) = \mathbf{W}\mathbf{u}^I(a)$. Letting $\boldsymbol{\omega} = \mathbf{e}^\top \mathbf{W}$ (where $\mathbf{e}$ is a vector of ones), the *(local) social welfare* of $a$ is:

$$sw_l(a, \mathbf{u}^I) = \boldsymbol{\omega}^\top \mathbf{u}^I(a). \tag{3}$$

Thus SWM under the local model is simply weighted maximization of intrinsic preferences, where the weight of $j$'s intrinsic utility $\omega_j$ is the sum of its incoming edge weights. Fig. 1(b) shows the weights derived for each agent under the local model. Using preference rankings and any scoring rule (e.g., Borda, plurality, k-approval, etc.) to determine intrinsic utilities, the decision may be different in the local model than when only intrinsic preferences are used (e.g., for both Borda and plurality, $a$ wins in the intrinsic model while $b$ wins in the local model; see Fig. 1(a) and (b)). 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.

Things are more subtle in the global empathetic model. Computing the utility vector $\mathbf{u}(a)$ for alternative $a$ requires solving a linear system to find the fixed point of Eq. 2. A unique solution is not guaranteed to exist;[4] 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 $j$. Let $\mathbf{D}$ be the $n \times n$ diagonal matrix with $d_{jj} = w_{jj}$. We can write Eq. 2 as:

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

As a consequence,

PROPOSITION 1 (FIXED-POINT UTILITY). *Assuming non-negativity, normalizaton, and positive self-loop, Eq. 4 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 results are included in an online appendix.) As in the local model, SWM in the global model can be seen as weighted maximization of intrinsic preference:

COROLLARY 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. 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, $a$ wins in the intrinsic model, and $b$ wins in the local model. We discuss weight computation in Sec. 3.

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 process 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) [25, 32], 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 [21]. However, we believe our assumptions—and in particular, positive self-loop over all individuals—are appropriate in most social choice settings.

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

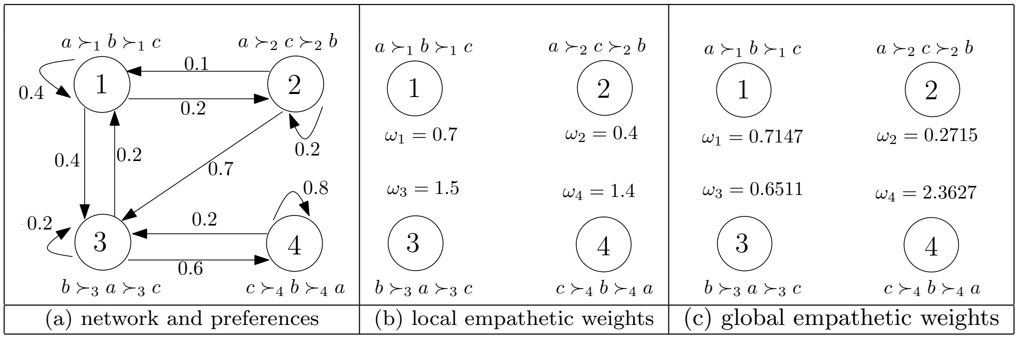


**Figure 1: A social network with ranked preferences, with weights under the local and global empathetic model. Using Borda or plurality-based utility, the consensus winner is different in each model: $a$ under intrinsic; $b$ under local empathetic; $c$ under global empathetic.**

## 2.4 Related Models and Concepts

The term empathy is used in several different ways in the literature [17]. Sometimes it refers to "seeing the world through the eyes of others" without being affected by this view, and such preferences [6] or "extended sympathy" [37, 3] is used to frame interpersonal comparison of utilities [23, 6]. However, our model is more consistent with an affective understanding of another, and having concern for that person's welfare [29], or having "other-regarding" preferences [26]. Empathy has recently drawn attention in neuroeconomics and social neurosience [39] 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 [30, 26, 6]). Most closely related to our work is the model of Maccheroni *et al.* [30], 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 [24; 1, Ch.8]. However, that work focuses on convergence of consensus (and correct) opinion among individuals with different initial opinions as individuals learn from one another. Our empathetic model can be viewed mathematically as a special case of a general model due to Friedkin and Johnson [20] (other special cases include [15, 19]). 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. Bodine-Baron *et al.* [7] study stable matchings (e.g., of students to residences) with *peer effects* that induce local network externalities. Brânzei and Larson address coalition formation on social networks where agent utility for a coalition depends on either her affinity weights [11] or distance to others in a network [12]. Maran *et al.* [31] study preference aggregation in combinatorial domains given the presence of social influence. Finally, auction design in social networks with externalities is studied in [22].

Boldi *et al.* [8] study *delegative democracy*, where an individual can either express her preferences directly, or to delegate her vote to a neighbor. In our model, individuals do not delegate their votes: we simply consider the dependency of their preferences on those of others.

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* [9], *hubs and authorities* [27] and *PageRank* [34]. 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. 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. 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, $\boldsymbol{\omega}^\top = \mathbf{e}^\top \mathbf{W}$, in time $O(n^2)$. However, social networks are generally extremely sparse, with the number of incoming edges to any node $j$ bounded by some small constant $c$. In such sparse networks, $\boldsymbol{\omega}$ can be computed in $O(n)$ time since $\omega_j$ 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. 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 [14, 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) which have $O(n)$ complexity (in sparse systems) per iteration and tend to converge very quickly in practice. We now describe one 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 1. *Consider the following iteration:*

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

*Assuming nonnegativity, normalizaton, and positive self-loop, this method converges to $\mathbf{u}(a)$, the solution to Eq. 4.*

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

$$u_j^{(t+1)}(a) = w_{jj} u_j^I(a) + \sum_{k \neq j} w_{jk} u_k^{(t)}(a), \qquad (5)$$

where $u_j^{(t)}(a)$ is agent $j$'s estimated utility for $a$ after $t$ iterations. This scheme has a natural interpretation: 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—simply let $u_k^{(0)}(a) = u_k^I(a)$. Critically, the error in the estimated utilities at the $t^{\text{th}}$ iteration can also be bounded:

THEOREM 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.

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

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

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

We can exploit Prop. 2 in a simple algorithm called *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. 5 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 $C$. 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, ICE converges quickly in practice.

# 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 non-empathetic (standard), 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*, 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 [5]:[5] 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 $\alpha$; then distribute weight $1 - \alpha$ equally to all other out-going 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 decisions, and the loss in social welfare that results from such disagreement. 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 disagreement* (DD) (over 2500 instances for a fixed setting) in which the the true and

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

| 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 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 2: Percentage decision disagreement: Borda, $m = 5$.**

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 $[sw^t(a_t) - sw^t(a_a)]/sw^t(a_t)$ (we often report it 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. *Normalized social welfare loss (NSWL)* is $[sw^t(a_t) - sw^t(a_a)]/[sw^t(a_t) - sw^t(a^-)]$. 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 1 while the decision disagreement percentage is shown in Table 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).

**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 2002 Irish General Election, using electoral data from the Dublin

| 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: Percentage decision disagreement, plurality/Borda: West Dublin, $m = 9$.**

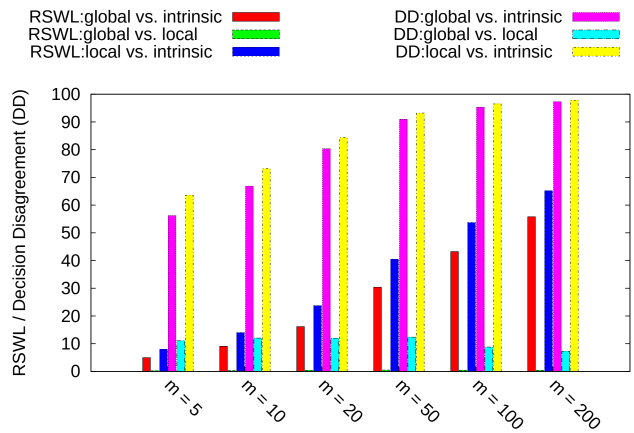


**Figure 2: RSWL and decision disgreement (DD), plurality.**

West constituency, which has 9 candidates and 29,989 ballots of top-$t$ form, of which 3800 are complete rankings.[6] 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) is quite high, ranging from 22-46%. Average NSWL (not shown) is not as high as with impartial culture (from 1-3%, with maximum loss around 40%).

**The effect of $m$.** Fig. 2 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 (not shown), 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.

**Self-loop weight $\alpha$.** Varying the self-loop weight $\alpha$ has a significant effect on NSWL and decision disagreement when true utility is global but intrinsic utility is assumed. Table 4 shows that, for both Borda and plurality, increasing $\alpha$ (i.e., decreasing overall degree of empathy) decreases both NWSL and DD. Similar trends hold for the local model. We also used a model in which nodes have different self-loop weights, drawing each node's $\alpha$ from a (truncated) Gaussian. As we vary the mean $\mu$, we see a similar trend in Table 5.

**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

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

| $\alpha$ | 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 4: Average NSWL/decision disgreement: global vs. intrinsic, varying $\alpha$.**

| $\mu$ | 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 5: Average NSWL/decision disgreement: global vs. intrinsic, $\alpha$ drawn from truncated Gaussian with mean $\mu$ and std. dev. 0.1.**

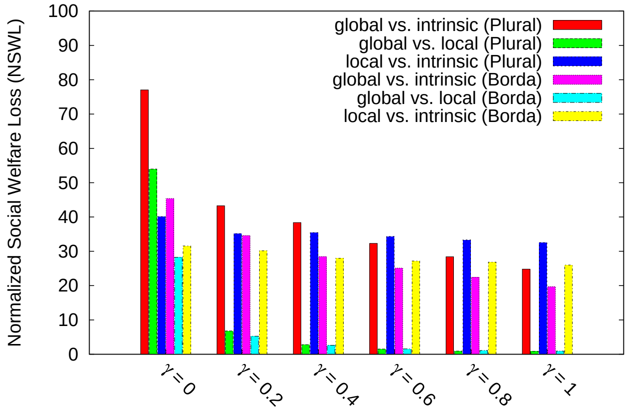


**Figure 3: Average NSWL, $m = 10$, varying $\gamma$.**

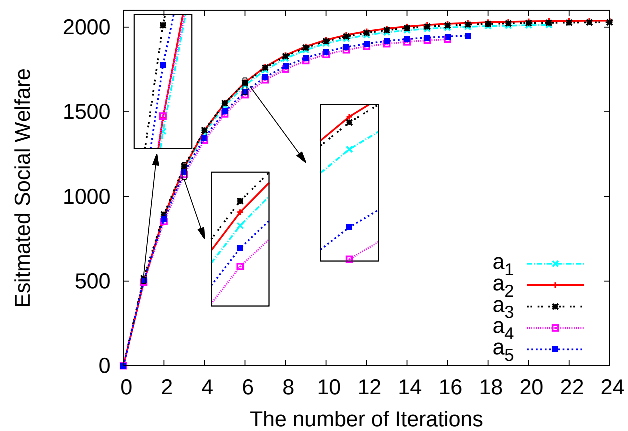


**Figure 4: Estimated social welfare vs. iterations of ICE (one sample run).**

| $\alpha$ | 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 6: Average number of iterations, varying $\alpha$.**

*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 $\gamma$. If $\gamma = 1$, our standard bidirectional network results (as above); when $\gamma = 0$ we obtain a completely hierarchical network.

Fixing $m = 10$, Fig. 3 depicts NSWL for both Borda and plurality as $\gamma$ 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.** Finally, we examine how the self-loop weight $\alpha$ affects the expected number of iterations required by the ICE algorithm. We fix $m = 5$, and vary $\alpha$. Fig. 4 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 6 shows the average number of iterations for various $\alpha$, for Borda and plurality. In all cases, the number of iterations is small relative to network size. ICE is quite insensitive to the scoring rule, and termination time declines dramatically with increasing $\alpha$.

# 5. CONCLUDING REMARKS

We have presented 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 algorithms 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.

This model is a starting point for the broader investigation of empathetic preferences in social choice. We are exploring more realistic processes for simultaneous generation of networks and preferences that better explain preference correlation (see, e.g., our *ranking network* framework [36]). 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. Our model can be applied to matching, assignment, and other group decision problems, each requiring 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.

## 6. ACKNOWLEDGMENTS

This research is supported by Natural Sciences and Engineering Research Council of Canada (NSERC).

## 7. REFERENCES


[1] D. Acemoglu and A. Ozdaglar. Opinion dynamics and learning in social networks. *Dynamic Games and Applications*, 1(1):3–49, 2011.

[2] S. L. Althaus. Information effects in collective preferences. *American Political Science Review*, 92(3):545–558, 1998.

[3] K. J. Arrow. Extended sympathy and the possibility of social choice. *Philosophia*, 7(2):223–237, 1978.

[4] M. Baccara, A. Imrohorolu, A. J. Wilson, and L. Yariv. A field study on matching with network externalities. *The American Economic Review*, 102(5):1773–1804, 2012.

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

[6] K. Binmore. *Natural Justice*. Oxford, 2005.

[7] E. Bodine-Baron, C. Lee, A. Chong, B. Hassibi, and A. Wierman. Peer effects and stability in matching markets. *SAGT-11*, pp.117–129, 2011.

[8] P. Boldi, F. Bonchi, C. Castillo, and S. Vigna. Voting in social networks. *CIKM-09*, pp.777–786, 2009.

[9] P. Bonacich. Power and centrality: A family of mesasures. *Amer. J. Sociology*, 1170–1182, 1987.

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

[11] S. Brânzei, K. Larson. Coalitional affinity games and the stability gap. *IJCAI-09*, pp.79–84, 2009.

[12] S. Brânzei and K. Larson. Social distance games. *IJCAI-11*, pp.91–96, 2011.

[13] N. A. Christakis and J. H. Fowler. *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives*. Back Bay Books, 2011.

[14] T. Cormen, C. Leiserson, R. Rivest, C. Stein. *Introduction to Algorithms*. MIT Press, 2009.

[15] M. H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.

[16] D. A. Easley and J. M. Kleinberg. *Networks, Crowds, and Markets*. Cambridge, 2010.

[17] P. Fontaine. Identification and economic behavior: Sympathy and empathy in historical perspective. *Economics and Philosophy*, 13(2):261–280, 1997.

[18] J. H. Fowler and N. 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.

[19] J. R. French. A formal theory of social power. *Psychological Review*, 63(3):181–194, 1956.

[20] N. Friedkin and E. Johnsen. Social influence and opinions. *J. Math. Soc.*, 15:193–206, 1990.

[21] B. Golub and M. O. Jackson. Naive learning in social networks and the wisdom of crowds. *Amer. Econ. J.: Microeconomics*, 2:112–149, 2010.

[22] N. Haghpanah, N. Immorlica, V. Mirrokni, and K. Munagala. Optimal auctions with positive network externalities. *ACM EC-11*, pp.11–20, 2011.

[23] J. C. Harsanyi. *Rational Behaviour and Bargaining Equilibrium in Games and Social Situations*. Cambridge, 1977.

[24] M. O. Jackson. *Social and Economic Networks*. Princeton, 2008.

[25] J. G. Kemeny and J. L. Snell. *Finite Markov Chains*. Springer-Verlag, New York, 1976.

[26] A. Kirman and M. Teschl. Selfish or selfless? The role of empathy in economics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1538):303–317, 2010.

[27] J. Kleinberg. Authoritative sources in a hyperlinked environment. *JACM*, 46:604–632, 1999.

[28] S. Konrath, E. O'Brien, C. Hsing. Changes in dispositional empathy in american college students over time: A meta-analysis. *Personality and Social Psych. Rev.*, 15:180–198, 2011.

[29] H. Leibenstein. *Beyond Economic Man: A New Foundation for Microeconomics*. Harvard, 1976.

[30] F. Maccheroni, M. Marinacci, and A. Rustichini. Social decision theory: Choosing within and between groups. *Review of Economic Studies*, 79(4):1591–1636, 2012.

[31] A. Maran, N. Maudet, M. S. Pini, F. Rossi, and K. B. Venable. A framework for aggregating influenced CP-nets and its resistance to bribery. *AAAI-13*, pp.668–674, 2013.

[32] C. D. Meyer. *Matrix Analysis and Applied Linear Algebra*, volume 2. SIAM, 2000.

[33] D. C. Mutz. Cross-cutting social networks: Testing democratic theory in practice. *American Political Science Review*, 96(1):111–126, 2002.

[34] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, November 1999.

[35] Y. Saad. *Iterative Methods for Sparse Linear Systems*. SIAM, 2003.

[36] A. Salehi-Abari and C. Boutilier. Ranking networks. *NIPS-13 Workshop on Frontiers of Network Analysis: Methods, Models, and Applications*, 2013.

[37] A. K. Sen. *Collective Choice and Social Welfare*. North-Holland Publishing Co., Amsterdam, 1970.

[38] S. Simon and K. R. Apt. Social network games. *Journal of Logic and Computation*, 2013.

[39] T. 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.

[40] T. Voice, S. D. Ramchurn, and N. R. Jennings. On coalition formation with sparse synergies. *AAMAS-12*,

pp.223–230, 2012

# APPENDIX

## A. LINEAR ALGEBRA BACKGROUND

We here first provides relevant definitions and results used in our proofs.

*Definition 1.* **Spectrum $\sigma(\mathbf{A})$.** The set of eigenvalues of $n \times n$ matrix $\mathbf{A}$ is called its spectrum $\sigma(\mathbf{A})$.

*Definition 2.* **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 3.* **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 3 (**Nonsingular M-matrix** [32]). *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. 3 is necessary to guarantee the nonsingularity of an M-matrix.[7]

THEOREM 4 (GERSCHGORIN CIRCLES [32]). *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 for 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:

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

We here focus on the $p$-norm matrix norm $\|.\|_p$ which are induced by $p$-norms 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:

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

$p$-norm matrix norms have several important properties: (1) they are *submultiplicative*: $\|\mathbf{AB}\|_p \leq \|\mathbf{A}\|_p \|\mathbf{B}\|_p$. 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 $\mathbf{A}$:

$$\|\mathbf{A}\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{m} |a_{ij}| \tag{6}$$

[7] In some references (e.g.,[32]), an M-matrix is defined with $s > \rho(\mathbf{B})$. By this definition, an M-matrix is non-singular.

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

$$\|\mathbf{A}\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^{n} |a_{ij}|. \tag{7}$$

We review the Jacobi iterative method and its convergence criteria and rate. Iterative methods offer practical advantages for solving linear systems [35]. 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 [35] 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. Each iteration of the Jacobi method takes the form of:

$$\mathbf{x}^{(t+1)} = \mathbf{\Lambda}^{-1}(\mathbf{E} + \mathbf{F})\mathbf{x}^{(t)} + \mathbf{\Lambda}^{-1}\mathbf{b} \tag{8}$$

THEOREM 5 (CONVERGENCE OF ITERATIVE METHODS [35]). *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 2 (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 5 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 8, we have:

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

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

## B. PROOFS

To prove Prop. 1, we first show

LEMMA 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. 4), we have $\sigma(\mathbf{B}) \subset \cup_{i=1}^{n} \mathcal{G}_i$ where

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

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

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

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 of 0. Since the number of these disks is finite, we can cover all these closed disks with a closed covering disk defined by $\{c \in \mathcal{C} \big| |c| \le R_{max}\}$ where $R_{max} = \max_{i=1}^n R_i$. Without loss of generality, let $l = \arg\max_i R_i$. So, we have

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

From this, it follows that:

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

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

Proof of Prop. 1. Using Eq. 2, 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 $i \ne j$. By nonnegativity assumption, we have $w_{ij} \ge 0$, so $\mathbf{B} \ge \mathbf{0}$. By setting $s = 1$, $(\mathbf{I} - (\mathbf{W} - \mathbf{D})) = (s\mathbf{I} - \mathbf{B})$ which is an M-matrix (See Definition 3). Using Lemma 1, we have $\rho(\mathbf{B}) < 1$. Since $s = 1$, then $\rho(\mathbf{B}) < s$. By Proposition 3, it follows that $(\mathbf{I} - (\mathbf{W} - \mathbf{D}))$ is nonsingular and $(\mathbf{I} - (\mathbf{W} - \mathbf{D}))^{-1} \ge \mathbf{0}$. $\square$

Proof Proof of Theorem 1. From Eq. 2, 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:

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

Since $\mathbf{A} = \mathbf{I} - (\mathbf{W} - \mathbf{D})$, we have that $\mathbf{\Lambda} = \mathbf{I}$ 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 1, we have $\rho(\mathbf{W} - \mathbf{D}) < 1$. Then, using Corollary 2, we have shown 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. $\square$

Proof of Theorem 2. Using Eq. 2 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

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

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

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

Lemma 2. *Assume nonnegativity and normalization, and 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: $\max_{i \in \mathcal{N}} u_i^{(t)}(a) \le d$ and $\min_{i \in \mathcal{N}} u_i^{(t)}(a) \ge c$. We can write the updating scheme for each individual $i \in \mathcal{N}$ and the alternative $a \in \mathcal{A}$ as follows:

$$u_i^{(t+1)}(a) = w_{ii} u_i^I(a) + \sum_{k \ne i} w_{ik} u_k^{(t)}(a) \tag{9}$$

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 of $d$ and then the lower found of $c$ for $u_i^{(t+1)}(a)$, $\forall i \in \mathcal{N}$, and fixed $a$. For the upper bound, we can write the following:

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

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 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 $u_i(a) \in [c, d]$. $\square$

Proof of Theorem 3. Let $\tilde{w} = \min_{1\leq i\leq n} w_{ii}$. Using Theorem 2, we can write

$$
\begin{aligned}
&\left\|\mathbf{u}(a)-\mathbf{u}^{(t)}(a)\right\|_\infty \leq (1-\tilde{w})^t\left\|\mathbf{u}(a)-\mathbf{u}^{(0)}(a)\right\|_\infty \implies\\
&\left|u_i(a)-u_i^{(t)}(a)\right| \leq \max_i\left|u_i(a)-u_i^{(t)}(a)\right|\\
&\leq (1-\tilde{w})^t\left\|\mathbf{u}(a)-\mathbf{u}^{(0)}(a)\right\|_\infty \implies\\
&\sum_{i=1}^n\left|u_i(a)-u_i^{(t)}(a)\right| \leq n\,(1-\tilde{w})^t\left\|\mathbf{u}(a)-\mathbf{u}^{(0)}(a)\right\|_\infty.
\end{aligned} \tag{10}
$$

By Lemma 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 (10) as follows:

$$
\begin{aligned}
\sum_{i=1}^n\left|u_i(a)-u^{(t)}(a)\right| &\leq n\,(1-\tilde{w})^t\left\|\mathbf{u}(a)-\mathbf{u}^{(0)}(a)\right\|_\infty\\
&\leq n(d-c)\,(1-\tilde{w})^t\,.
\end{aligned} \tag{11}
$$

By Lemma 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:

$$
\begin{aligned}
\sum_{i=1}^n\left|u_i(a)-u_i^{(t)}(a)\right| &\geq \left|\sum_{i=1}^n\left(u_i(a)-u_i^{(t)}(a)\right)\right|\\
&=\left|\sum_{i=1}^n u_i(a)-\sum_{i=1}^n u_i^{(t)}(a)\right|\\
&=\left|sw(a)-sw^{(t)}(a)\right|.
\end{aligned} \tag{12}
$$

By Inequalities (11) and (12), we conclude that

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

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

Proof of Proposition 2. Using the triangle inequality and the inequality presented in Theorem 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 $sw(b) \geq sw(a)$. $\square$

## C. THE ICE ALGORITHM

The pseudo code for our ICE algorithm is presented below.

**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`
$C \leftarrow \mathcal{A}$;
$\tilde{w} = \min_{1\leq i\leq n} w_{ii}$;
$t \leftarrow 0$;
**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** $\hat{sw}^{(t)} - sw^{(t)}(a) \geq 2n(d-c)\,(1-\tilde{w})^t$ **then**
      $C \leftarrow C - \{a\}$
**return** $a^* \in C$