

An Axiomatic Approach to Community Detection

Christian Borgs
Microsoft Research
borgs@microsoft.com

Jennifer Chayes
Microsoft Research
jchayes@microsoft.com

Adrian Marple
Stanford
adrian.marple@gmail.com

Shang-Hua Teng^{*}
USC
shanghua@usc.edu

ABSTRACT

Inspired by social choice theory in voting and other contexts [2], we provide the first axiomatic approach to community identification in social and information networks. We start from an abstract framework, called *preference networks* [3], which, for each member, gives their ranking of all the other members of the network. This preference model enables us to focus on the fundamental conceptual question:

What constitutes a community in a social network?

Within this framework, we axiomatically study the formation and structures of communities in two different ways. First, we apply social choice theory and define communities indirectly by postulating that they are fixed points of a preference aggregation function obeying certain desirable axioms. Second, we directly postulate six desirable axioms for communities to satisfy, without reference to preference aggregation. For the second approach, we prove a taxonomy theorem that provides a *structural characterization* of the family of *axiom-conforming* community rules as a lattice. We complement this structural theorem with a *complexity result*, showing that, while for some rules in the lattice, community characterization is straightforward, it is coNP-complete to characterize subsets according to others. Our study also sheds light on the limitations of defining community rules solely based on preference aggregation, namely that many aggregation functions lead to communities which violate at least one of our community axioms. These include any aggregation function satisfying Arrow's "independence of irrelevant alternatives" axiom, as well as commonly used aggregation schemes like the Borda count or generalizations thereof. Finally, we give a polynomial-time rule consistent with five axioms and weakly satisfying the sixth axiom.

^{*}Supported in part by a Simons Investigator Award from the Simons Foundation and by NSF grant CCF-1111270.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
ITCS'16, January 14–16, 2016, Cambridge, MA, USA.

Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM 978-1-4503-4057-1/16/01 ...\$15.00.

DOI: <http://dx.doi.org/10.1145/2840728.2840748>.

Categories and Subject Descriptors

Theory of computation [**Algorithmic game theory and mechanism design**]: Social networks; Information systems [**World Wide Web**]: Web applications—*Social networks*; Information systems [**Information systems applications**]: Data mining—*Clustering*; Human-centered computing [**Collaborative and social computing**]: Social network analysis

Keywords

Network Analysis; social choice theory; community identification; social and information networks; axiomatic framework; taxonomy of community rules; stable communities

1. INTRODUCTION

A fundamental problem in network analysis is the characterization and identification of subsets of nodes in a network that have significant structural coherence. This problem is usually studied in the context of *community identification* and *network clustering*. Like other inverse problems in data mining and machine learning, this one is conceptually challenging: There are many possible ways to measure the degree of coherence of a subset and many possible interpretations of preferences and affinities to model network data. As a result, various seemingly reasonable/desirable conditions to qualify a subset as a community have been studied in the literature [17, 14, 13, 19, 25, 9]. However, direct comparison of different community characterizations is quite difficult.

A community is formed by a group of individuals, while an information/social network is usually specified by data describing each individual's direct neighbors, or the individuals' affinities or preferences for other members in the network. Thus, in order to answer the fundamental question, "*What constitutes a community in a social network?*," it is desirable to first answer

How do individual preferences (affinities or connectivities) result in group preferences?

In this paper, we take what we believe is a novel and principled approach to the problem of community identification. Inspired by work on clustering [12], and, more conceptually, by classic work in social choice theory [2], we propose an axiomatic approach towards understanding network communities, both providing a framework for comparison of different community characterizations, and relating community identification to well-studied problems in social choice theory.

1.1 Preference Networks and Community Identification Functions

Our axiomatic approach uses an abstract social network framework. Our framework is analogous to the clustering approach of Kleinberg [12] in which network nodes are clustered according to similarity, specified by a similarity matrix derived from a metric over nodes. In Kleinberg’s view, a clustering algorithm inputs a similarity matrix, and outputs a set of *non-overlapping* subsets called *clusters*. Analogously, we view community detection as an algorithmic problem which inputs a *preference network*, and outputs a set of (overlapping) subsets identified as *communities*.

To formalize this, consider a non-empty finite set V , let $L(V)$ denote the set of all linear orders π on V , with $\pi(u)$ denoting the rank of member u , from 1 for the highest rank to $|V|$ for the lowest. For $\pi \in L(V)$ and $i, j \in V$, we use the notation $i \succ_{\pi} j$ if i is ranked higher than j , i.e., $\pi(i) < \pi(j)$.

DEFINITION 1.1 (PREFERENCE NETWORKS). *A preference network is a pair $A = (V, \Pi)$, where V is a non-empty finite set and Π is a preference profile on V , defined as an element $\Pi = \{\pi_i\}_{i \in V} \in L(V)^V$. Here π_i specifies the ranking of V in the order of i ’s preference. We denote the collection of all preference networks¹ by \mathcal{A} .*

The preference network framework is inspired by social choice theory [2]. This framework is already used extensively — for voting, college admissions, medical residency assignments [8, 21, 10], for studying coalition formation in collaborative games [5, 22], for specifying routing preferences in the Border Gateway Protocol between autonomous systems of the Internet [20, 6], to name just a few contexts.

Community Identification Rules

The preference framework provides a complete-information network model that enables us to focus on the fundamental conceptual question regarding how to transform individual preferences into community identification rules. Like the work of Arrow [2] and Kleinberg [12], our axiomatic theory identifies basic desirable axioms which *consistent* community characterizations should satisfy, and also specifies the structures of consistent rules and the communities that they identify. Within the framework of preference networks, we can define the community-identification rules, the main subject of our study, as set-theoretical functions.

DEFINITION 1.2 (COMMUNITY FUNCTIONS). *A community function or rule is a function \mathcal{C} that maps a preference network $A = (V, \Pi)$ onto a collection of non-empty sets $S \subseteq V$, $A \mapsto \mathcal{C}(A) \in 2^{2^V - \{\emptyset\}}$. We say a subset $S \subseteq V$ is a community in a preference network $A = (V, \Pi)$ according to a community function \mathcal{C} if and only if $S \in \mathcal{C}(A)$.*

The preference network framework offers a natural, although highly selective, community rule, which we call the *Clique Rule* (\mathcal{C}_{clique}). We say $S \subseteq V$ is a *clique* in preference network $A = (V, \Pi)$ if each member of S prefers every member of S over every non-member.

RULE 1. (CLIQUE RULE) *Cliques and only cliques are communities.*

¹To avoid paradoxes, we assume that V is a subset of some reference set \mathcal{V} , say the set of natural numbers.

Another example of natural rules for community identification is the rule defined in Balcan *et al.* [3] — which we will denote by $\mathcal{C}_{democratic}$ — based on “democratic voting.” For a preference network $A = (V, \Pi)$, let $\phi_S^{\Pi}(i)$, for $S \subseteq V$ and $i \in V$, denote the number of votes that i would receive when each member $s \in S$ casts one vote for each of its $|S|$ most preferred members according to its preference π_s .

RULE 2 (DEMOCRATIC RULE). *S is a democratically-certified community in a preference network $A = (V, \Pi)$, if every member in S receives more votes from S than every non-member, i.e., $\min_{u \in S} \phi_S^{\Pi}(u) - \max_{v \notin S} \phi_S^{\Pi}(v) > 0$.*

The set-theoretical formulation of community rules of Definition 1.2 implies some direct comparison of community rules. For example, we say a community function \mathcal{C}_1 is *more selective* than another community function \mathcal{C}_2 if, for all preference networks $A = (V, E)$, $\mathcal{C}_1(A) \subseteq \mathcal{C}_2(A)$. Clearly, \mathcal{C}_{clique} is more selective than $\mathcal{C}_{democratic}$.

1.2 Our Work

Within the framework of preference networks, we axiomatically study the formation and structures of communities in two related approaches:

1. **Axiomatization and Characterization Community Rules:** We postulate six desirable axioms for community-identification rules to satisfy as well as provide a *structural characterization* of the family of *axiom-conforming* community rules in terms of a lattice. We also complement this structural theorem with a *complexity result*.
2. **Understanding Social-Choice Aggregation-Based Community Rules:** We examine a family of natural community-identification rules inspired by the classic social-choice theory from the lens of our axiomatic framework. We present two impossibility results that illustrate the limitation of defining community rules solely based on social-choice preference aggregation.

We now present the highlight of these results.

Axiomatization

We use *axioms* to state properties, such as fairness and consistency, that a *desirable* community function should have when applied to *all* preference networks. Our framework characterizes community-identification rules using a natural set of six axioms which will be defined formally in Section 2. Below, we summarize these axioms, which can be organized into three sets:

Social-Choice Fundamentals: The first three axioms are inspired by social choice theory. They reflect the intuition that community identification in preference networks is a form of social choice within each subset in the network. The two most fundamental axioms are **Anonymity** (Axiom 1) and **Monotonicity** (Axiom 2). The former states that a community rule should be isomorphism-invariant, i.e., they should not use the individuals’ labels. The latter captures that the community characterization of a subset should be monotonically consistent — if community members’ preferences change in favor of its members, then the community should remain a community. The third axiom, **Embedding** (Axiom 3), states that, if newcomers join the population of a preference network in such way that members of the existing

population all retain their original preferences for each other, and prefer them over any new member, then the community characterization regarding members of the original network should remain unchanged in this bigger network, independent of the preferences towards and of the newcomers.

Cliques and the Entire Population: The fourth axiom, World Community (Axiom 4), is a basic one, stating that the entire population V of any preference network $A = (V, \Pi)$ is a community of A . Note that V is a clique of $A = (V, \Pi)$. In fact, if we replace Axiom World Community by the axiom that “all cliques are communities,” we obtain an equivalent axiomatic system.

Baseline Stability and Self-Respect: The next two axioms are inspired by the classical game-theoretical studies of stable marriage and coalition formation [8, 21, 10, 5, 22]. The first, Self-Approval (Axiom 5), states that a community should have the necessary “self-respect.” It is not the case that everyone in the community “unanimously prefers” an outside group of the same size over the community itself. The last, Group Stability (Axiom 6), states that a community should have the necessary stability: No subgroup in the community is replaceable by an equal-sized group of non-members who are “unanimously” preferred by the rest of community members.

Structural Characterization Community Rules

The set-theoretical formulation of community rules also provides the following natural definition.

DEFINITION 1.3. (OPERATIONS OVER COMMUNITY RULES) For two community functions \mathcal{C}_1 and \mathcal{C}_2 , we define their intersection and union, $\mathcal{C}_1 \cap \mathcal{C}_2$ and $\mathcal{C}_1 \cup \mathcal{C}_2$, respectively, to be the community functions which, for all preference networks $A = (V, \Pi)$ characterize subsets of V according to

$$(\mathcal{C}_1 \cap \mathcal{C}_2)(A) := \mathcal{C}_1(A) \cap \mathcal{C}_2(A) \quad (1)$$

$$(\mathcal{C}_1 \cup \mathcal{C}_2)(A) := \mathcal{C}_1(A) \cup \mathcal{C}_2(A). \quad (2)$$

Note that $\mathcal{C}_1 \cap \mathcal{C}_2$ is more *selective* than both \mathcal{C}_1 and \mathcal{C}_2 , and $\mathcal{C}_1 \cup \mathcal{C}_2$ is more *inclusive* than both \mathcal{C}_1 and \mathcal{C}_2 . Our main structural result is the following taxonomy theorem and the complete characterization of the most comprehensive and the most selective community rules consistent with all our community axioms.

The set of axiom-conforming community rules is not empty, and forms a lattice under the operations of union and intersection defined above.

This result provides an interesting contrast to the classic axiomatization result of Arrow [2] and the more recent result of Kleinberg on clustering [12] that inspired our work. Unlike voting or clustering where the basic axioms lead to impossibility theorems, the preference network framework has a natural community rule, the *Clique Rule*. Indeed, the Clique Rule satisfies all our axioms. One may ask:

Is the Clique rule the only axiom-conforming community identification rule?

Our initial attempt to prove the “impossibility beyond the Clique rule” conjecture in fact led us to another community rule consistent with all axioms.² We call this rule the *Comprehensive Rule* ($\mathcal{C}_{comprehensive}$) because our proof shows the

²“If we knew what we were doing it wouldn’t be research”, Einstein once said.

following: For any community rule \mathcal{C} satisfying all axioms, for all preference network A :

$$\mathcal{C}_{clique}(A) \subseteq \mathcal{C}(A) \subseteq \mathcal{C}_{comprehensive}(A). \quad (3)$$

Thus, $\mathcal{C}_{clique}(A)$ and $\mathcal{C}_{comprehensive}(A)$ form a lower and upper bound, respectively, for the lattice of axiom-comforming community rules. We complement this structural theorem with a complexity result: We show that while identifying a community by the Clique Rule is straightforward, it is coNP-complete to determine if a subset satisfies the Comprehensive Rule.

Communities as Fixed Points of Social Choice: Schema and Limitation

Our approach of starting from preference networks to study community identification connects community formation to social choice theory [2], thus providing a theoretical framework for understanding the problem of combining individual preferences into a community preference. One way to define communities is to generalize the principle of the Clique rule by classifying a set $S \subseteq V$ to be a community if, collectively, the members of S prefer every member in S to every element outside of S — community members collectively “certify themselves.” To formalize what “collectively prefer” means, we use the notion of preference aggregation functions from social choice theory [2].

A *preference aggregation function* is a function which generates a single aggregate preference from a set of individual preferences. In this context, it is useful to allow for ties in the aggregate preference. To formalize this, we introduce the set $\overline{L(V)}$ of rankings with ties (see Section 4 for the precise definition), and then define preference aggregation as a function $F : L(V)^* \rightarrow \overline{L(V)}$, where $L(V)^*$ is the union of $L(V)^S$ over all non-empty, finite subsets of a countable reference set, which here we take as the union \mathcal{V} of all possible groundsets V . Given a non-empty finite set $S \subset \mathcal{V}$ and a preference profile $\Pi_S = \{\pi_s : s \in S\} \in L(V)^S$, we say the image $F(\Pi_S)$ is the *aggregated preference* of S according to Π_S .

DEFINITION 1.4. (COMMUNITIES AS FIXED POINTS OF SOCIAL CHOICE) Let $F : L(V)^* \rightarrow \overline{L(V)}$ be a preference aggregation function. For $A = (V, \Pi)$, define f_F as the map on non-empty subsets $S \subset V$ which maps S to the subset $f_F(S) := \{v \in V \mid F(\Pi_S)(v) \leq |S|\}$. A non-empty subset $S \subset V$ is called a *community* of A with respect to F if and only if $f_F(S) = S$. The function \mathcal{C}_F mapping A into the set of communities defined above is called the *fixed-point rule* with respect to F .

The fixed-point rule captures the strongest notion of “collective self-preference” based on social-choice aggregation.

While Definition 1.2 is convenient for studying the mathematical structures of our theory, community identification is a computational problem as much as a mathematical problem. Thus, it is desirable that communities can be characterized by a constructive community function \mathcal{C} that is:

- **Consistent:** \mathcal{C} satisfies all (or nearly all) axioms;
- **Constructive:** Given a preference network $A = (V, \Pi)$, and a subset $S \subseteq V$, one can determine in polynomial-time (in $n = |V|$) if $S \in \mathcal{C}(A)$.

The above mentioned co-NP completeness of the Comprehensive Rule highlights the computational difficulty of community identification based on axiom-conforming rules, raising the question of whether there are natural community rules that are indeed constructive. In this context, communities defined in terms of fixed point rules provide a rich source of candidates in terms of various aggregation functions studied in social choice; indeed, any aggregation function that can be calculated in polynomial time gives rise to a constructive community rule via the fixed point Definition 1.4. Defining community rules in terms of aggregation functions also gives us another perspective on community rules, since we can examine aggregation functions themselves through the lens of social choice theory, and connect the properties of the aggregation functions with the properties of their fixed rules from the lens of our axiomatic framework.

Our studies shed light on the limitations of formulating community rules solely based on preference aggregation: We show that, although the fixed-point community rules systematically generalizes the Clique rule, many aggregation functions lead to rules which violate at least one of our community axioms. We prove two impossibility theorems.

1. For any aggregation function satisfying Arrow’s Independence of Irrelevant Alternatives axiom, its fixed-point rule must violate one of our axioms.
2. Any fixed-point rule based on commonly-used weighted aggregation schemes like Borda count or generalizations thereof is inconsistent with (at least) one of our axioms.

The second impossibility result was more surprising to us than the first one: While weighted fixed-point rules are natural from a social choice viewpoint, it turns out that the fixed-point rules of several weighted aggregation functions, including $\mathcal{C}_{democratic}$, are inconsistent with the Axiom Monotonicity! We believe this violation is illustrative of the fundamental subtlety of community rules. This leads us to consider fixed-point rules, $\mathcal{C}_{harmonious}$, which are not given in terms of weighted voting schemes, such as the following.

RULE 3 (HARMONIOUS COMMUNITIES). *Let $A = (V, \Pi)$ be a preference network. A non-empty subset $S \subseteq V$ is a harmonious community of A if for all $u \in S$ and $v \in V - S$, the majority of $\{\pi_s : s \in S\}$ prefer u over v .*

This rule can be formulated as a fixed-point rule of a topologically defined aggregation function — see Proposition 4.9 — and satisfies Axioms 1-5, as well as a weaker form of Axiom Group Stability.

2. COHERENT COMMUNITIES: AXIOMS

We now define our six core axioms. Below, we fix a ground set V and a community function \mathcal{C} .

AXIOM 1 (Anonymity (A)). *Let $S, S' \subseteq V$ and Π, Π' be such $S' = \sigma(S)$ and $\Pi' = \sigma(\Pi)$ for some permutation $\sigma : V \rightarrow V$. Then $S \in \mathcal{C}(V, \Pi) \iff S' \in \mathcal{C}(V, \Pi')$.*

This is a standard axiom: labels should have no effect on a community function.

AXIOM 2 (Monotonicity (Mon)). *Let $S \subseteq V$. If Π and Π' are such that $\forall s \in S, \forall u \in S$ and $\forall v \in V, u \succ_{\pi'_s} v \implies u \succ_{\pi_s} v$, then $S \in \mathcal{C}(V, \Pi') \implies S \in \mathcal{C}(V, \Pi)$.*

The Axiom Monotonicity states that, if the profile changes so that the rank of a community member increases without decreasing the rank of other members, then this remains a community in the new ranking. Mon also allows non-members to change arbitrarily, as long as their positions relative to any members remains the same or worse. To state the next axiom, we define the projection $A|_{V'}$ of a preference network $A = (V, \Pi)$ onto a subset $V' \subseteq V$ as the preference network $A|_{V'} = (V', \Pi|_{V'})$ where $\Pi|_{V'} = \{\pi'_s\}_{s \in V'}$ is defined by setting π'_s to be the ranking on V' which maintains the relative ordering of all members of V' , i.e., for all $s, u, v \in V', u \succ_{\pi'_s} v \iff u \succ_{\pi_s} v$. We say that A' is EMBEDDED into A if $A' = A|_{V'}$ for some $V' \subseteq V$.

AXIOM 3 (Embedding (Emb)). *If $A' = (V', \Pi')$ is embedded into $A = (V, \Pi)$ and $\pi_i(j) = \pi'_i(j)$ for all $i, j \in V'$ then $\mathcal{C}(A') = \mathcal{C}(A) \cap 2^{V'}$.*

In other words, if a network is embedded into a larger network in such a way that, with respect to the preferences in the larger network, the members of the smaller network prefer each other over everyone else, then the community classification regarding members of the smaller network remains unchanged. The next axiom is self explanatory.

AXIOM 4 (World Community (WC)). *For all preference profiles $\Pi, V \in \mathcal{C}(V, \Pi)$.*

To state the last two axioms, we start with a few definitions to formalize the notion that a member prefers a group over another of the same size. Given $(V, \Pi) \in \mathcal{A}$ and non-empty disjoint sets $G, G' \subseteq V$ of equal size, we say that $s \in V$ prefers G' over G if, after reordering the elements $g_1, \dots, g_{|G|}$ and $g'_1, \dots, g'_{|G|}$ of G and G' according to her preferences, s prefers g'_i to $g_i \forall i = 1, \dots, |G|$. We sometimes also refer to this preference as *lexicographical preference*. Let $(V, \Pi) \in \mathcal{A}$. A set $S \subseteq V$ is called *group stable* with respect to Π if for all non-empty $G \subseteq S$ there exists no $G' \subseteq V - S$ that has the same size as G and is preferred to G by all $s \in S - G$. S is called *self-approving* with respect to Π if there exists set no $G' \subseteq V - S$ that has the same size as S and is preferred to S by all $s \in S$.

AXIOM 5 (Self-Approval (SA)). *If Π is a preference profile over V and $S \in \mathcal{C}(V, \Pi)$, then S is SELF-APPROVING with respect to Π .*

AXIOM 6 (Group Stability (GS)). *If Π is a preference profile over V and $S \in \mathcal{C}(V, \Pi)$, then S is GROUP STABLE with respect to Π .*

Axiom Group Stability provides a type of game-theoretic stability [16, 15, 4, 23, 24], and states that no subgroup in a community can be replaced by an equal-size group of non-members who are lexicographically preferred by the remainder of the community members, while Axiom Self-Approval provides a stability notion of minimum *self-respect*, and requires that there is no outside group of the same size as S which is lexicographically preferred to S by everyone in S . Note that the set V is trivially group stable for all Π , and that any set S with $|S| > |V|/2$ is self-approving for all Π .

It is easy to check that the Clique Rule satisfies all six axioms. However, the clique rule has a structural feature which essentially rules out any non-trivial overlap of communities, while “Real-world” communities typically have non-trivial overlaps among themselves.

PROPOSITION 2.1. $\forall A = (V, \Pi)$, if $S_1, S_2 \in \mathcal{C}_{clique}(A)$, then either $S_1 \cap S_2 = \emptyset$ or $S_1 \subset S_2$, or $S_2 \subset S_1$.

Given this property, it is desirable to answer the question whether there are other community functions that satisfy all axioms. The answer is yes, and in fact, the set of all axiom conforming rules has interesting structural properties, which we study of our next section.

3. LATTICE OF AXIOM-CONFORMING COMMUNITY RULES

Let \mathcal{C} denote the family of all axiom-conforming community rules. Let \mathcal{C}_B be a superset of \mathcal{C} consisting of all community rules satisfying Axioms 1-4. The main result of this section is that both \mathcal{C} and \mathcal{C}_B are not empty, and in fact, they each form a lattice under the natural union and intersection operations given by Definition 1.3. Two community rules are special for the lattice of \mathcal{C} . The first one is the Clique Rule (\mathcal{C}_{clique}) defined in Section 1. The second one is the following Comprehensive Rule:

RULE 4 (COMPREHENSIVE RULE). *For any preference network A , let $\mathcal{C}_{SA}(A)$ and $\mathcal{C}_{GS}(A)$ denote the subsets $S \subset V$ which are self-approving and group stable, respectively. Then, $\mathcal{C}_{comprehensive} := \mathcal{C}_{SA} \cap \mathcal{C}_{GS}$.*

Finally, let \mathcal{C}_{all} be the rule declaring every non-empty set S to be a community.

3.1 Taxonomy of Community Rules

THEOREM 3.1 (LATTICE OF COMMUNITY RULES).

1. *The algebraic structure,*

$$\mathcal{T} = (\mathcal{C}, \cup, \cap, \mathcal{C}_{clique}, \mathcal{C}_{comprehensive})$$

is a bounded lattice, with $\mathcal{C}_{comprehensive}$ and \mathcal{C}_{clique} as the lattice's top and bottom.

2. *The algebraic structure,*

$$\mathcal{T}_B = (\mathcal{C}_B, \cup, \cap, \mathcal{C}_{clique}, \mathcal{C}_{all})$$

is a bounded lattice, with \mathcal{C}_{all} and \mathcal{C}_{clique} as the lattice's top and bottom.

Part 1 of the theorem implies that, for any axiom-conforming community function \mathcal{C} , it must be the case that, for every preference network A ,

$$\mathcal{C}_{clique}(A) \subseteq \mathcal{C}(A) \subseteq \mathcal{C}_{comprehensive}(A).$$

The following basic lemma is key for establishing our Taxonomy Theorem.

LEMMA 3.2. (INTERSECTION LEMMA) *For all community rule \mathcal{C} , if $\mathcal{C} \in \mathcal{C}_B$ then $\mathcal{C} \cap \mathcal{C}_{GS} \cap \mathcal{C}_{SA} \in \mathcal{C}$.*

PROOF. Let $\tilde{\mathcal{C}} := \mathcal{C} \cap \mathcal{C}_{GS} \cap \mathcal{C}_{SA}$, where $\mathcal{C} \in \mathcal{C}_B$. Because \mathcal{C}_{GS} and \mathcal{C}_{SA} are both consistent with Axioms A, WC, and Emb, $\tilde{\mathcal{C}}$ remains consistent with these three axioms. To see $\tilde{\mathcal{C}}$ satisfies Axiom Mon, choose Π, Π' such that, for all $u, s \in S$ and $v \in V$, $u \succ_{\pi'_s} v \implies u \succ_{\pi_s} v$. We need to show that if $S \in \tilde{\mathcal{C}}((V, \Pi'))$ then $S \in \tilde{\mathcal{C}}((V, \Pi))$. Suppose this is not the case, then either (1) $S \notin \mathcal{C}_{GS}((V, \Pi))$ or (2) $S \notin \mathcal{C}_{SA}((V, \Pi))$. In Case (1), there exists $G \subset S$, $G' \subset V - S$, $|G| = |G'|$,

and bijections ($f_s : S \rightarrow G' | s \in S - G$) such that $\forall s \in S - G, \forall u \in G, u \prec_{\pi_s} f_s(u)$. By Mon, $u \prec_{\pi'_s} f_s(u)$, which shows $S \notin \mathcal{C}_{GS}(A')$. In Case (2), $\exists G' \subset V - S$, bijections ($f_s : S \rightarrow G'$) such that $\forall s, u \in S, u \prec_{\pi_s} f_s(u)$. By Mon, $u \prec_{\pi'_s} f_s(u)$, which implies that $S \notin \mathcal{C}_{SA}(A')$.

Finally, by definition, $\mathcal{C} \cap \mathcal{C}_{GS} \cap \mathcal{C}_{SA}$ satisfies GS and SA. Thus, $\mathcal{C} \cap \mathcal{C}_{GS} \cap \mathcal{C}_{SA} \in \mathcal{C}$. \square

PROOF. (Proof of Theorem 3.1) We first prove that \mathcal{C}_{clique} and $\mathcal{C}_{comprehensive}$ are, respectively, the most selective and inclusive axiom-conforming community rules. On one hand it is easy to see that for any rule that satisfies WC and Emb, all cliques must be communities, showing that $\mathcal{C}_{clique}(A) \subseteq \mathcal{C}(A)$ whenever \mathcal{C} satisfies WC and Emb; see also Proposition 4.2 below. On the other hand, $\mathcal{C}_{all}(A) = 2^V - \{\emptyset\}$ satisfies Axioms 1-4, and hence $\mathcal{C}_{all} \in \mathcal{C}_B$. Since $\mathcal{C}_{comprehensive} = \mathcal{C}_{all} \cap \mathcal{C}_{GS} \cap \mathcal{C}_{SA}$, by the Intersection Lemma, $\mathcal{C}_{comprehensive} \in \mathcal{C}$. Thus, for any \mathcal{C} that satisfies all axioms, $\mathcal{C}_{clique}(A) \subseteq \mathcal{C}(A) \subseteq \mathcal{C}_{comprehensive}(A)$.

The two operations \cap and \cup over the community functions are both commutative and associative. They also satisfy the absorption property. In other words, $\forall \mathcal{C}_1, \mathcal{C}_2 \in \mathcal{C}$:

$$\mathcal{C}_1 \cup (\mathcal{C}_1 \cap \mathcal{C}_2) = \mathcal{C}_1 \quad \text{and} \quad \mathcal{C}_1 \cap (\mathcal{C}_1 \cup \mathcal{C}_2) = \mathcal{C}_1.$$

For example, to see the first one, for any affinity network A , we have:

$$\begin{aligned} (\mathcal{C}_1 \cup (\mathcal{C}_1 \cap \mathcal{C}_2))(A) &= \mathcal{C}_1(A) \cup (\mathcal{C}_1 \cap \mathcal{C}_2)(A) \\ &= \mathcal{C}_1(A) \cup (\mathcal{C}_1(A) \cap \mathcal{C}_2(A)) \\ &= \mathcal{C}_1(A). \end{aligned}$$

To complete the proof that \mathcal{T} and \mathcal{T}_B are lattices, we need to prove that \mathcal{T} and \mathcal{T}_B are closed under \cap and \cup . We organize the arguments as following:

- **A, WC:** it is obvious that if \mathcal{C}_1 and \mathcal{C}_2 satisfy Axioms A and WC, then both $\mathcal{C}_1 \cup \mathcal{C}_2$ and $\mathcal{C}_1 \cap \mathcal{C}_2$ also satisfy Axioms A and WC.
- **Mon:** Suppose $A = (V, \Pi)$ and $A' = (V, \Pi')$ are two preference networks considered in Axiom Mon, and $S \subset V$. Then if \mathcal{C}_1 and \mathcal{C}_2 satisfy Mon, we have $S \in \mathcal{C}_i(A') \implies S \in \mathcal{C}_i(A)$ for $i \in \{1, 2\}$. Thus, if $S \in \mathcal{C}_1(A') \cap \mathcal{C}_2(A')$ then $S \in \mathcal{C}_1(A) \cap \mathcal{C}_2(A)$, and if $S \in \mathcal{C}_1(A') \cup \mathcal{C}_2(A')$ then $S \in \mathcal{C}_1(A) \cup \mathcal{C}_2(A)$. Thus, both $\mathcal{C}_1 \cup \mathcal{C}_2$ and $\mathcal{C}_1 \cap \mathcal{C}_2$ also satisfy Axioms Mon.
- **Emb:** If both \mathcal{C}_1 and \mathcal{C}_2 satisfy Emb, then for any $A = (V, \Pi)$ and any "embedded world" $A' = (V', \Pi')$ such that Π, Π' satisfy the assumption of Axiom Emb, we have $\mathcal{C}_i(A') = \mathcal{C}_i(A) \cap 2^{V'}$ for $i \in \{1, 2\}$. So

$$\begin{aligned} \mathcal{C}_1(A') \cap \mathcal{C}_2(A') &= (\mathcal{C}_1(A) \cap 2^{V'}) \cap (\mathcal{C}_2(A) \cap 2^{V'}) \\ &= (\mathcal{C}_1(A) \cap \mathcal{C}_2(A)) \cap 2^{V'} \\ \mathcal{C}_1(A') \cup \mathcal{C}_2(A') &= (\mathcal{C}_1(A) \cap 2^{V'}) \cup (\mathcal{C}_2(A) \cap 2^{V'}) \\ &= (\mathcal{C}_1(A) \cup \mathcal{C}_2(A)) \cap 2^{V'}. \end{aligned}$$

Thus, both $\mathcal{C}_1 \cup \mathcal{C}_2$ and $\mathcal{C}_1 \cap \mathcal{C}_2$ satisfy Axioms Emb.

Together, this shows that $\forall \mathcal{C}_1, \mathcal{C}_2 \in \mathcal{C}_B, \mathcal{C}_1 \cap \mathcal{C}_2 \in \mathcal{C}_B$ and $\mathcal{C}_1 \cup \mathcal{C}_2 \in \mathcal{C}_B$. Thus, $\mathcal{T}_B = (\mathcal{C}_B, \cup, \cap, \mathcal{C}_{clique}, \mathcal{C}_{all})$ is a lattice with \mathcal{C}_{all} as the lattice's top and \mathcal{C}_{clique} as the lattice's bottom. (4) **GS, SA:** Assume $\mathcal{C}_1 \in \mathcal{C}$ and $\mathcal{C}_2 \in \mathcal{C}$ satisfy Axioms GS

and SA. We can then argue as for Axiom Mon above to show that both $\mathcal{C}_1 \cup \mathcal{C}_2$ and $\mathcal{C}_1 \cap \mathcal{C}_2$ satisfy Axioms GS, SA. Thus, $\mathcal{T} = (\mathcal{C}, \cup, \cap)$ is a lattice. \square

3.2 Complexity of Community Rules

The first part of Theorem 3.1 shows that $\mathcal{C}_{comprehensive}$ and \mathcal{C}_{clique} , respectively, are the most inclusive and selective axioms-conforming rules. While it is easy to determine whether, for $A = (V, \Pi)$, a subset of V lies in $\mathcal{C}_{clique}(A)$, the rule $\mathcal{C}_{comprehensive}$ turns out to be highly “non-constructive”.

THEOREM 3.3 (RICHNESS OF COMPREHENSIVE RULE). *It is coNP-complete to determine whether given $A = (V, \Pi)$, $S \subset V$ is a member of $\mathcal{C}_{comprehensive}(A)$.*

Before starting the proof, we introduce a notation which we will use throughout this section. Given a preference profile (V, Π) and a non-empty set $S \subset V$, we say that a set $G' \subset V - S$ is a *witness that S is not self-approving*, if S lexicographically prefers G' to S , and we say that a pair $(G, G') \subset S \times (V - S)$ is a *witness that S is not group-stable* if $S - G$ lexicographically prefers G' to G . We say that $G \subset S$ *threatens the stability of S* if there exists a $G' \subset V - S$ such that $S - G$ lexicographically prefers G' to G . Let's first characterize the complexity of Axiom Self-Approval.

THEOREM 3.4. (COMPLEXITY OF SA) *It is coNP-complete to determine whether a subset $S \subset V$ is self-approving in a given preference network $A = (V, \Pi)$.*

PROOF. We reduce 3-SAT to this decision problem: Suppose $\mathbf{c} = (c_1, \dots, c_m)$ is a 3-SAT instance with Boolean variables $\mathbf{x} = (x_1, \dots, x_n)$ (i.e., $c_j = \{u_j, v_j, w_j\} \subset \cup_{i=1}^n \{x_i, \bar{x}_i\}$). We define a preference network as follows:

- $V = A \cup B \cup D \cup X$ has $m + n + m + 2n$ members, where $A = \{a_1, \dots, a_m\}$, $B = \{b_1, \dots, b_n\}$, $D = \{d_1, \dots, d_m\}$, and $X = \{x_1, \dots, x_n, \bar{x}_1, \dots, \bar{x}_n\}$. The distinguished subset will be $S = A \cup B$, and for convenience we will denote its complement as $U = D \cup X$.
- Since we will focus on subset S , here we only define the preferences of members in S . The preferences of U can be chosen arbitrarily.
 - Member b_i has preference $D \succ A \succ \{x_i, \bar{x}_i\} \succ \{b_i\} \succ X - \{x_i, \bar{x}_i\} \succ B - \{b_i\}$, where preferences between elements of each set can be chosen arbitrarily.
 - Member a_j has preference $c_j \succ \{a_j\} \succ D \cup X - c_j \succ B \cup A - \{a_j\}$, where again preferences between elements of each set are arbitrary.

Intuitively, members of A are used to enforce clause consistency and members of B are used to enforce variable consistency (no variable can be set to both true and false at the same time). Subsets of X naturally constitute an assignment of the variables, and D provides necessary padding in order to apply Self-Approval. We now show that S is not self-approving if and only if the 3-SAT instance is satisfiable.

In one direction, suppose $Y = \{y_1, \dots, y_n\}$ where $y_i \in \{x_i, \bar{x}_i\}$ is a satisfying assignment for the 3-SAT instance. Let $G' = Y \cup D$. Consider the bijection, f , where $f(a_j) = d_j$ and $f(b_i) = y_i$. Then for all $s \in S$ and all i , $f(s) \succ_{\pi_{b_i}} s$. All that is left is to find similar bijections for each a_j . First,

note that for a_j all bijections f_j trivially satisfy $f_j(s) \succ_{\pi_{a_j}} s$ where $s \in B \cup A - \{a_j\}$, since this set is ranked at the bottom of π_{a_j} . Thus it is sufficient to show that there exists an element of G' that a_j prefers to itself. This happens so long as one of the literals from its clause is in G' , which must be true by the fact that Y is a satisfying assignment. In the other direction, suppose $G' \subset U = D \cup X$ is a witness that S is not self-approving. Note the following:

- $D \subset G'$ otherwise any b_i will have a member of A that cannot be mapped to a more preferred member of G' .
- Let $Y = X \cap G'$. Then $|Y| = n$ by the above fact and the fact that $|G'| = n + m$.
- $\{x_i, \bar{x}_i\} \cap G' \neq \emptyset$ by b_i 's preference, and by the pigeonhole principle the literals of Y are consistent (i.e. $\{x_i, \bar{x}_i\} \not\subseteq Y$).
- $c_j \cap Y \neq \emptyset$ by a_j 's preferences.

Thus, the variable assignment implied by Y is a satisfying assignment for the 3-SAT instance. \square

The following lemma allows us to reduce various complexity results concerning community axioms to Theorem 3.4.

LEMMA 3.5 (PADDING). *Let $\emptyset \neq S \subset V \subset V'$ be such that the size of $\tilde{S} = V' - V$ is at least $|S|$, and let $S' = S \cup \tilde{S}$. Then each preference profile Π on V can be mapped onto a preference profile Π' on V' such that*

- (i) $S' \in \mathcal{C}_{GS}(V', \Pi') \cap \mathcal{C}_{SA}(V', \Pi') \Leftrightarrow S' \in \mathcal{C}_{GS}(V', \Pi')$.
- (ii) $S' \in \mathcal{C}_{GS}(V', \Pi') \Leftrightarrow S \in \mathcal{C}_{SA}(V, \Pi)$.

PROOF. Since $|\tilde{S}| \geq |S|$, we can find a surjective map $g: \tilde{S} \rightarrow S$. Given such a map, define Π' arbitrarily, except for the following two constraints:

- If $s \in S$, then π'_s ranks all of $S' = \tilde{S} \cup S$ before anyone in $V - S = V' - S'$;
- If $\tilde{s} \in \tilde{S}$, then $\pi'_{\tilde{s}}$ ranks all of \tilde{S} first, and then gives the rank $\pi'_{\tilde{s}}(v) = |\tilde{S}| + \pi_{g(\tilde{s})}(v)$ to every $v \in V = V' - \tilde{S}$.

Since every $s \in S \subset S'$ ranks all of S' before $V' - S'$, no subset $G' \subset V' - S'$ can be lexicographically preferred by π'_s to a subset of S' . As a consequence, S' is trivially self-approving with respect to Π' , proving statement (i). Furthermore, G cannot threaten the stability of S' if $G \subset S'$ is such that $(S' - G) \cap S \neq \emptyset$. If $G \subset S'$ threatens the stability of S' , we therefore must have that $G \supset S$. On the other hand, if $G \not\supset S$, then G contains an element $\tilde{s} \in \tilde{S}$ which means that no set G' can be lexicographically preferred G , since all elements of S' prefer all of \tilde{S} to anyone in $V' - S'$.

Thus G can only threaten the stability of S' if $G = S$. In other words, $S \notin \mathcal{C}_{GS}(\Pi')$ if and only if there exists $G' \subset V' - S'$ such that for all $\tilde{s} \in \tilde{S} = S' - G$, G' is lexicographically preferred to S with respect to $\pi'_{\tilde{s}} = \pi_{g(\tilde{s})}$. Since by assumption, the image of \tilde{S} under g is all of S , this is equivalent to the statement that for all $s \in S$, G' is lexicographically preferred to S with respect to π_s , which is the condition that G' is a witness to $S \notin \mathcal{C}_{SA}(\Pi)$, proving statement (ii). \square

Given this lemma, both the next theorem and Theorem 3.3 are immediate consequences of Theorem 3.4.

THEOREM 3.6. (COMPLEXITY OF GS) *It is coNP-complete to determine whether a subset $S \subset V$ is group-stable in a given preference network $A = (V, \Pi)$.*

3.3 Paths up the Taxonomy Lattice

The Intersection Lemma provides us with a tool for exploring the taxonomy of community rules. Together with Theorem 3.1, it gives us the following scheme to map an arbitrary community rule, \mathcal{C} , to an axiom-conforming community rule: First take the unique smallest (with respect to the lattice \mathcal{T}_B) rule $\bar{\mathcal{C}}$ that contains \mathcal{C} and satisfies all axioms besides SA and GS. Then, remove all communities which are not both self-approving and group stable. In other words, the map is: $\mathcal{C} \mapsto \bar{\mathcal{C}} \cap \mathcal{C}_{SA} \cap \mathcal{C}_{GS}$. By moving up lattice \mathcal{T}_B , we can apply the Intersection Lemma to define more inclusive axiom-conforming community rules. We now give two example paths up taxonomy lattice, and state their algorithmic and complexity consequences.

RULE 5. (RELAXED CLIQUE RULE: $\mathcal{C}_{Clique(g)}$) For a non-negative function $g : \mathbb{N} \rightarrow \mathbb{N} \cup \{0\}$, a non-empty subset $S \subseteq V$ is a community in $A = (V, \Pi)$ if and only if $\forall u, s \in S$, $\pi_s(u) \in [1 : |S| + g(|S|)]$.

$\mathcal{C}_{Clique(g)} \in \mathcal{C}_B$ and hence $(\mathcal{C}_{Clique(g)} \cap \mathcal{C}_{GS} \cap \mathcal{C}_{SA})$ satisfies all axioms. As g varies from 0 to ∞ , $\mathcal{C}_{Clique(g)}$ moves up the lattice \mathcal{T}_B from \mathcal{C}_{clique} to \mathcal{C}_{all} . The intersection with $\mathcal{C}_{SA} \cap \mathcal{C}_{GS}$ provides a “vertical” glimpse of the taxonomy lattice \mathcal{T} . In particular, as the community rules along this vertical path become more inclusive (when g increases), they become less constructive for community identification.

PROPOSITION 3.7. Given a preference network $A = (V, \Pi)$ and a subset $S \subseteq V$, we can determine in $O(2^g |S|^{g+3})$ time whether or not $S \in (\mathcal{C}_{Clique(g)} \cap \mathcal{C}_{GS} \cap \mathcal{C}_{SA})(A)$. Particularly, if $g = \Theta(1)$, then this decision problem is in P. However, the decision problem is co-NP complete for $g = |S|^\delta$ for any constant $\delta \in (0, 1]$.

RULE 6. (HARMONIOUS PATH: $\mathcal{C}_{harmonious(\lambda)}$) For $\lambda \in [0 : 1]$, a non-empty subset S is a λ -harmonious community in $A = (V, \Pi)$ if $\forall u \in S, v \in V - S$, at least λ -fraction of $\{\pi_s : s \in S\}$ prefer u over v .

$\mathcal{C}_{harmonious(\lambda)} \in \mathcal{C}_B$. Thus, $\mathcal{C}_{harmonious(\lambda)} \cap \mathcal{C}_{GS} \cap \mathcal{C}_{SA}$ satisfies all axioms. Therefore, as λ varies from 1 to 0, the community function $\mathcal{C}_{harmonious(\lambda)}$ moves up the lattice \mathcal{T}_B from $\mathcal{C}_{harmonious(1)} = \mathcal{C}_{clique}$ to $\mathcal{C}_{harmonious(0)} = \mathcal{C}_{all}$.

PROPOSITION 3.8. Given $A = (V, \Pi)$ and a subset $S \subseteq V$, we can determine whether $S \in (\mathcal{C}_{harmonious(\lambda)} \cap \mathcal{C}_{GS} \cap \mathcal{C}_{SA})(A)$ in polynomial time, if $(1 - \lambda)|S| < 2$, while it is co-NP complete to answer this question if $(1 - \lambda)|S| \geq 16$.

3.4 Number of Potential Communities

PROPOSITION 3.9. Assume that $n \geq 8$. There exists a preference network A such that $\mathcal{C}_{comprehensive}(A) \geq 2^{n/2}$.

PROOF. The preference profile, $\Pi_{H\&S}$, that is about to be described has been dubbed the “hero and sidekick” example as will soon become clear. Consider a world composed of $n/2$ hero-sidekick duos. Each member of a hero-sidekick duo first prefers the hero of that duo then the sidekick of the duo, then all other heroes, followed lastly by all other sidekicks (in some fixed but arbitrary order). Now consider a subset, S , that is composed of all heroes and an arbitrary set of sidekicks. Note that because there are $2^{n/2}$ different sets of sidekicks, it is sufficient to show that S is a community in

$\mathcal{C}_{comprehensive}([n], \Pi_{H\&S})$. First, note that S clearly satisfies SA. To show that S satisfies GS, consider two sets $G \subset S$ and $G' \subset V - S$ of equal size. We first note that it will be enough to consider the case where $(S - G) \times G$ contains no hero-sidekick pair (u, v) , since otherwise u would prefer v over everyone else, in particular over everyone in G' . Applying this to the sidekicks in G , we conclude that G must contain at least as many heroes as sidekicks. On the other hand, G' can't be lexicographically preferred to G if G contains at least two heroes, showing that only two cases are possible: G consisting of a hero-sidekick pair, or G made up of just a single hero. But neither one leads to a counterexample if $|S - G| > |G| = |G'|$, since then we can find an $s \in S - G$ which is not the partner of any sidekick in G' , which means that s prefers the hero in G to everyone in G' . Since S contains all heroes by assumption, we see that S is group stable as soon as $n \geq 8$. \square

4. FIXED-POINT RULES AND LIMITATION OF SOCIAL-CHOICE AGGREGATION

We now examine aggregation functions and their community rules through the lens of our axiomatic framework. We prove two *impossibility results* which may help to shed light on the limitations of defining community rules solely based on preference aggregation.

4.1 Social Choice Axioms and Implications for Fixed-Point Community Rules

We first review the basic axioms from the traditional social choice theory [2]. To this end, we begin by formally defining the set, $\overline{L(V)}$, of preferences with ties (or indifference) as the set of all maps $\pi : V \rightarrow \{1, \dots, |V|\}$ such that whenever k elements v_1, \dots, v_k have the same rank $\pi(v_1) = \dots = \pi(v_k)$, then we skip the ranks $\pi(v_k) + 1, \dots, \pi(v_k) + k - 1$. So for example, if 3 elements are tied at rank 2, the next rank in the image of π will be 5. $\overline{L(V)}$ is also known as the *ordered partition* of V . We also need the notion of an election, which will be defined as a triple (V, F, S) where V and S are finite sets (called the set of candidates and voters, respectively), and $F : L(V)^* \rightarrow \overline{L(V)}$ is a preference aggregation function.

SOCIAL CHOICE AXIOM 1 (Unanimity (U)). An election (V, F, S) satisfies Unanimity if, for all preference profiles, $\Pi_S = \{\pi_s : s \in S\} \in L(V)^S$ and all pairs of candidates, $\{i, j\} \subseteq V$, $\pi_s(i) > \pi_s(j), \forall s \in S \implies F(\Pi_S)(i) > F(\Pi_S)(j)$.

The question then is: what properties capture the intuition behind Unanimity and how do they relate to this social choice axiom? To answer this, we define the following two properties:

PROPERTY 1 (Pareto Efficiency (PE)). A community function, \mathcal{C} , is Pareto Efficient if, given $A \in \mathcal{A}$ and $S \in \mathcal{C}(A)$, for all $u \in S, v \notin S$, there is a $s \in S$ such that $u \succ_{\pi_s} v$.

PROPERTY 2 (Clique (Cq)). A community function \mathcal{C} satisfies the Clique Property if for all $A \in \mathcal{A}$, $u \succ_{\pi_s} v, \forall u, s \in S, \forall v \notin S \implies S \in \mathcal{C}(A)$.

Property Pareto Efficiency is a negative property that states that subsets in which a non-member is preferred to a member by everyone inside the subset should not be a community. In contrast, Clique is a positive property that states

that a completely self-loving group (i.e., a clique) must be a community. It turns out that both of these properties are implied by Unanimity, and that the second is implied by the Axioms World Community and Embedding.

PROPOSITION 4.1. *Fix V and a preference aggregation function F , and let \mathcal{C}_F be the fixed point rule with respect to F . If all elections (V, F, S) with $S \subseteq V$ satisfies Unanimity, then \mathcal{C}_F satisfies the properties Pareto Efficiency and Clique.*

PROPOSITION 4.2. *Let \mathcal{C} be a community rule that satisfies the World Community and Embedding Axioms. Then \mathcal{C} must also satisfy the Clique Property.*

SOCIAL CHOICE AXIOM 2 (Non-Dictatorship (ND)). *An election (V, S, F) is Non-Dictatorial if there exists no dictator, i.e., no voter $i \in S$ such that $F(\Pi_S) = \pi_i$ for all preference profiles $\Pi_S \in L(V)^S$.*

Instead of showing properties implied by ND as we did with Unanimity, we do the inverse, and show that a dictatorship violates some of our axioms.

PROPOSITION 4.3. *Fix V and a preference aggregation function F . If \mathcal{C}_F , the fixed point rule with respect to F , satisfies Group Stability or Anonymity, then all elections (V, F, S) with $S \subseteq V$ and $1 < |S| < |V|$ satisfy Non-Dictatorship.*

SOCIAL CHOICE AXIOM 3. (Independence of Irrelevant Alternatives) *An election (V, F, S) satisfies Independence of Irrelevant Alternatives (IIA) if, for all preference profiles, Π_S and $\Pi'_S \in L(V)^S$ and all candidates $a, b \in V$, we have that*

$$(\forall s \in S, a \succ_{\pi_s} b \Leftrightarrow a \succ_{\pi'_s} b) \implies (a \succ_{F(\Pi_S)} b \Leftrightarrow a \succ_{F(\Pi'_S)} b).$$

This axiom can reasonably be considered the strongest of the three, in that it says that the aggregate preference between two candidates does not depend on the preferences voters have between either of the two and some other candidate. Arrow's celebrated impossibility result immediately leads to impossibility results in our settings, showing in particular that Independence of Irrelevant Alternatives for the aggregation function is inconsistent with at least one of Axioms 3, 4, or 6.

THEOREM 4.4. (IMPOSSIBILITY OF FIXED-POINT RULE WITH IIA AGGREGATION) *Let F be an aggregation function such that the fixed point rule with respect to F satisfies the Clique Property and the Group Stability Axiom. Then no election (V, F, S) with $S \subseteq V$ and $1 < |S| < |V|$ satisfies IIA.*

PROOF. Let $S \subseteq V$ such that $1 < |S| < |V|$. Assume that the election (V, F, S) satisfies IIA, and the resulting fixed point rule \mathcal{C}_F satisfies Cq and GS. We will first show that the election (V, F, S) must satisfy Unanimity. In the following preference profiles, $\Pi, \Pi', \Pi'' \in L(V)^S$, we assume that every member of S has the same preference, π, π' , and π'' respectively. First, let π rank all members of S above non-members. By the Clique Property, $S \in \mathcal{C}_F(V, \Pi)$ and thus

$$\forall s \in S, v \notin S, s \succ_{F(\Pi)} v. \quad (4)$$

Thus, by IIA, if $s \in S$ is unanimously preferred to $v \notin S$, s must be strictly preferred to v in the aggregate preference. Now let π' be the same as π only with the least preferred member of S , s' , and the most preferred non-member, v' ,

switched in rank. By the partial Unanimity property (4), in the aggregate $F(\Pi')$, all members of $S - \{s'\}$ are preferred to all $v \notin S$, and all members of S are preferred to all $v \in V - S - \{s'\}$. But, by GS, $S \notin \mathcal{C}_F(\Pi')$, which is only possible is if $v' \succeq_{F(\Pi')} s'$. Applying the partial Unanimity property once more yields the following two statements:

$$\forall s \in S - \{s'\}, s \succ_{F(\Pi')} s' \quad \text{and} \quad \forall v \notin S \cup \{v'\}, v' \succ_{F(\Pi')} v,$$

and by IIA, this in turn implies

$$\forall s \in S - \{s'\}, s \succ_{F(\Pi)} s' \quad \text{and} \quad \forall v \notin S \cup \{v'\}, v' \succ_{F(\Pi)} v. \quad (5)$$

By IIA, this means that for any two members or two non-members, if one is unanimously preferred to the other, then it must be strictly preferred in aggregate preference. Indeed, consider, e.g., $s, s' \in S$ and a profile $\bar{\Pi}_S$ such that $s \succ_{\bar{\pi}_i} s'$ for all $i \in S$. Choose Π in such a way that every member has the same profile, s' has rank $|S|$ and $s \succ_{\pi_i} s'$ for all $i \in S$. By IIA, $s \succ_{F(\bar{\Pi})} s' \iff s \succ_{F(\Pi)} s'$, so by (5), s is preferred to s' in aggregate.

Finally, consider π'' where v' is switched with the second lowest ranked member, s'' . By the above additional partial Unanimity property, s' must be strictly preferred to s'' in the aggregation $F(\Pi'')$, and thus v' must be strictly rather than weakly preferred to s'' in the aggregate preference. Again by IIA, if a non-member, $v \notin S$, is unanimously preferred to a member $s \in S$, v must be strictly preferred to s in the aggregate preference. Taken together, these three partial Unanimity properties constitute Unanimity. Since the election (V, F, S) satisfies both IIA and Unanimity, by Arrow's Impossibility Theorem [2], it must be a dictatorship, contradicting Proposition 4.3. \square

4.2 Weighted-Aggregations and Limitations of Their Fixed-Point Rules

There are many preference aggregation functions satisfying the other two standard axioms of social choice theory, Unanimity and Non-Dictatorship, e.g., the well-known Borda count [27]. Moreover, there are many interesting fixed point rules generalizing Borda count, several of which can be cast as *weighted voting schemes* as follows: Let $W = (w^1, w^2, \dots)$ be a sequence of weight vectors $w^i \in \mathbb{R}^V$. For $\Pi_S \in L(V)^*$, we then define the aggregate preference $F_W(\Pi_S)$ on V by

$$i \succ_{F(\Pi_S)} j \iff \sum_{s \in S} w_{\pi_s(i)}^{|S|} > \sum_{s \in S} w_{\pi_s(j)}^{|S|}.$$

For Borda count, $w^k = (|V|, |V| - 1, \dots, 1) \forall k$, and for the $\mathcal{C}_{democratic}$ rule of [3], w^k consists of k ones followed by $(|V| - k)$ zeros, implying that every voter has to choose k candidates, with all votes counting equally.

DEFINITION 4.5 (WEIGHTED FIXED POINT RULE). *For a sequence of vectors $W = (w^1, w^2, \dots)$ in \mathbb{R}^n , \mathcal{C}_W is the fixed point rule with respect to F_W .*

PROPOSITION 4.6. *All weighted fixed-point rules satisfy Axiom Anonymity. They satisfy Property Clique iff $\forall k \in [n - 1]$ the vector w^k is such that $w_i^k > w_j^k$ for $i \leq k$ and $j > k$.*

While weighted fixed-pointed rules are natural from a social choice viewpoint, it turns out that they are again incompatible with at least one of Axioms 3, 4, or 6. We first note that $\mathcal{C}_{democratic}$ violates both Axioms Mon and GS.

THEOREM 4.7. $\mathcal{C}_{democratic}$ does not satisfy Axioms Monotonicity or Group Stability. It satisfies all other axioms, as well as Properties Pareto Efficiency and Clique.

PROOF. From its voting function ϕ_S^Π , $\mathcal{C}_{democratic}$ satisfies Axioms A, WC, Emb, and Properties PE and Cq. Suppose $\mathcal{C}_{democratic}$ does not satisfy SA. Then, there exists a preference network $A = (V, \Pi)$, $S \in \mathcal{C}_{democratic}(A)$, $T \subseteq V - S$, and a tuple of bijections $(f_s : S \rightarrow T)$ that for all $s, u \in S$, $u \prec_{\pi_s} f_s(u)$. It follows that $\forall s \in S$, the numbers of votes cast by s for S according to ϕ_S^Π is less than the numbers of votes that s casts for T . Summing up the votes from S , the average votes that members of T receive is larger than the average votes that members of S receive, contradicting the assumption that everyone in S receives more votes than everyone in T . Thus, $\mathcal{C}_{democratic}$ satisfies SA.

Let $V = [1 : 6]$, $S = [1 : 3]$, let $\Pi = (\pi_1, \dots, \pi_6)$ be the preference profile

$$\begin{aligned} \pi_1 &= [142356], & \pi_2 &= [253416], & \pi_3 &= [631425] \\ \pi_4 &= [456123], & \pi_5 &= [156423], & \pi_6 &= [165423] \end{aligned}$$

and let Π' be the preference profile

$$\begin{aligned} \pi'_1 &= [142356], & \pi'_2 &= [234516], & \pi'_3 &= [314625] \\ \pi'_4 &= \pi_4, & \pi'_5 &= \pi_5, & \pi'_6 &= \pi_6. \end{aligned}$$

Then $S = [1 : 3] \in \mathcal{C}_{democratic}(V, \Pi)$, as each members of S receives two votes while everyone in $[4 : 6]$ receives only one vote. However, in violation of Axiom Mon, S is no longer a $\mathcal{C}_{democratic}$ community w.r.t Π' , since 4 now receives three votes, one more than 1, 2 and 3.

Note also $T = (1, 5, 6) \in \mathcal{C}_{democratic}(V, \Pi)$. Let $G = \{5, 6\} \subset T$ and $G' = (2, 4) \subset V - T$. As member 1 prefers 2 to 5 and 4 to 6, T does not satisfy Group Stability. \square

The violation of the monotonicity axiom was initially surprising and rather counterintuitive to us, and indeed motivated some of the research in this paper. This violation is illustrative of the subtlety of community rules. It leads us to the following general impossibility result, which together with Theorem 4.4, illustrates some basic limitations of fixed-point community rules.

THEOREM 4.8. (IMPOSSIBILITY OF WEIGHTED AGGREGATION SCHEMA) *Weighted Fixed Point Rules are inconsistent with either the Group Stability or the Clique Property.*

PROOF. Let $A = (V, \Pi)$ be a preference network, $S \subset V$, and \mathcal{C}_W a weighted fixed point rule satisfying the the Clique Property. Throughout the the proof, we will take

$$V = \{a, b, c, d, e\} \quad \text{and} \quad S = \{a, b, c\},$$

and consider preference profiles such that S violates Group Stability. In order for \mathcal{C}_W to obey the Axiom GS, we would need the weight vector $w^3 \in \mathbb{R}^5$ to be such that $S \notin \mathcal{C}(V, \Pi)$ for all Π considered in this proof. Our goal is to show that this will lead to a contradiction. We start under the assumption that the weights are decreasing, i.e., in addition to the already established fact that $w_i^3 > w_j^3$ when $i = 1, 2, 3$ and $j = 4, 5$ (since \mathcal{C}_W satisfies the the Clique Property), we first assume that $w_1^3 \geq w_2^3 \geq w_3^3$ and $w_4^3 \geq w_5^3$.

Consider the following scenario: $\pi_a = [adebc]$, $\pi_b = \pi_c = [abcde]$. Since a prefers d and e over b and c , S is not group stable and cannot be a community. By our assumption that $w_1^3 \geq w_2^3 \geq w_3^3 > w_4^3 \geq w_5^3$, we have that $a \succ_{F_W(\Pi_S)}$

$b \succeq_{F_W(\Pi_S)} c \succ_{F_W(\Pi_S)} e$ and $b \succ_{F_W(\Pi_S)} d$. Thus, the only way S cannot be a community is that $d \succeq_{F_W(\Pi_S)} c$, i.e., $w_2^3 + 2w_4^3 \geq 2w_3^3 + w_5^3$. This implies that we cannot have both $w_2^3 = w_3^3$ and $w_4^3 = w_5^3$.

Now consider a modified preference profile: $\pi'_a = \pi'_b = [abdce]$, $\pi'_c = [caebd]$. In this profile a and b prefer d over c , so again S violates GS and hence cannot be a community. On the other hand, we now have $a \succ_{F_W(\Pi'_S)} b$, $b \succeq_{F_W(\Pi'_S)} d \succ_{F_W(\Pi'_S)} e$. Thus we must have either $b \sim_{F_W(\Pi')} d$ or $d \succeq_{F_W(\Pi'_S)} c$. The former, however, implies $w_2^3 = w_3^3$ and $w_4^3 = w_5^3$ and is hence a contradiction. Therefore the latter must be true which implies $2w_3^3 + w_5^3 \geq w_1^3 + 2w_4^3$.

This brings us to the final preference profile:

$$\pi''_a = [abdce], \quad \pi''_b = [dcabe], \quad \pi''_c = [cbaed].$$

Again a and b prefer d to c , so the profile violates GS, and hence again can't be a community. Now $a \succ_{F_W(\Pi'')} c \succeq_{F_W(\Pi'')} b$ and $d \succ_{F_W(\Pi'')} e$, showing that for S not to be a community, we must have $d \succeq_{F_W(\Pi'')} b$, which gives $w_1^3 + w_3^3 + w_5^3 \geq 2w_2^3 + w_4^3$.

Defining $d_i = w_i^3 - w_{i-1}^3$, we can write the bounds obtained so far as (1) $d_4 \leq d_3 + d_5$, (2) $d_2 + d_3 + d_5 \leq d_4$, (3) $d_3 + d_4 + d_5 \leq d_2$. Chaining up these three bounds, we get

$$\begin{aligned} d_3 + d_5 &\geq d_4 \geq d_2 + d_3 + d_5 \\ &\geq d_3 + d_4 + d_5 + d_3 + d_5 = 2(d_3 + d_5) + d_4, \end{aligned}$$

contradicting our assumption $d_i \geq 0$ and the fact that Cq implies $d_4 > 0$.

To relax the constraint that the weights are ordered, we observe that all three profiles in the proof are such that, under arbitrary permutations of the first three and the last two positions, S still violates GS: for any permutation σ of $[1 : 5]$ that leaves $[1 : 3]$ and $[4 : 5]$ invariant, S violates GS under the profiles $\{\sigma \circ \pi_s\}_{s \in S}$, $\{\sigma' \circ \pi_s\}_{s \in S}$, and $\{\sigma'' \circ \pi_s\}_{s \in S}$. Choosing the permutation such that $\tilde{w}_i^3 = w_{\sigma(i)}^3$ are ordered, we obtain the above three inequalities for the weights \tilde{w}_i^3 , leading again to a contradiction. \square

4.3 Harmonious Communities

We will close this section with a discussion of the harmonious community rule, $\mathcal{C}_{harmonious}$, of Section 1. First, we prove $\mathcal{C}_{harmonious}$ is a fixed-point rule associated with a topologically defined aggregation function.

PROPOSITION 4.9. *There exists a preference aggregation function $F_{\mathcal{H}} : L(V)^* \rightarrow \overline{L}(V)$ such that defines $\mathcal{C}_{harmonious}$.*

PROOF. Given V , a finite set S , and a preference profile $\Pi_S \in L(V)^S$, we consider the following directed graph $G_{\Pi_S} = (V, E_{\Pi_S})$ where $(i, j) \in E_{\Pi_S}$ if at least half of S prefers i to j . Note that if $|S|$ is an odd number, then G_{Π_S} is a *tournament graph*. If $|S|$ is an even number, then E_{Π_S} contains both (i, j) and (j, i) if exactly half of Π_S prefer i to j . G_{Π_S} is *total* since for all $i, j \in V$, either $(i, j) \in E_{\Pi_S}$ or $(j, i) \in E_{\Pi_S}$. Because G_{Π_S} is total, the graph \hat{G}_{Π_S} obtained from G_{Π_S} by contracting each strongly connected component into a single vertex is an *acyclic*, tournament graph. Thus, the graph \hat{G}_{Π_S} has exactly one Hamiltonian path that totally orders its vertices. Let (V_1, \dots, V_t) be the strongly connected components of G_{Π_S} , sorted by the order determined by the Hamiltonian path. The partition (V_1, \dots, V_t) of V then defines an ordered partition $F_{\mathcal{H}}(\Pi_S)$, with $V_i \succ_{F_{\mathcal{H}}(\Pi_S)} V_j$ iff $i \leq j$.

Next, we consider a subset $T \subset V$. It is then easy to check that if T is of the form $T = \cup_{j \leq i} V_j$ for some $i \in [1 : t]$, then for all $u \in T, v \in V - T$, a majority of S prefers u to v , and vice versa. Specializing to $S = T$, we see that $\mathcal{C}_{\text{harmonious}}$ is defined by the preference aggregation function $F_{\mathcal{H}}$. \square

THEOREM 4.10. *$\mathcal{C}_{\text{harmonious}}$ satisfies Axioms 1-5. but it does not satisfy GS.*

PROOF. One easily establish that $\mathcal{C}_{\text{harmonious}}$ satisfies Axioms A, Mon, Emb and WC. We will now prove that $\mathcal{C}_{\text{harmonious}}$ satisfies SA: If $S \in \mathcal{C}_{\text{harmonious}}(A)$ does not satisfy SA, then there exists a $T \subset V - S$ of the same size as S such that each $s \in S$ lexicographically prefers T over S . This implies that, for each $s \in S$, there are at least $(1+2+\dots+|S|)$ pairs $(u, v) \in S \times T$ such that s prefers v over u . Thus the number of triples (s, u, v) such that $s \in S$ prefers $v \in T$ over $u \in S$ is at least $|S|^2(|S|+1)/2$. However, $S \in \mathcal{C}_{\text{harmonious}}(A)$ implies that this number has to be strictly smaller than $|S|^3/2$. The set T in the proof of Theorem 4.7 is also an example that $\mathcal{C}_{\text{harmonious}}$ violates Axiom GS. \square

While $\mathcal{C}_{\text{harmonious}}$ does not satisfy the GS Axiom, it satisfies the following weaker property.

PROPERTY 3. Weak Group Stability *For all preference profiles Π on V and all $S \in \mathcal{C}(V, \Pi)$, S is weakly group stable. Here a set $S \subset V$ is called weakly group stable if for all $G \subset S, G' \subset V - S$ s.t. $0 < |G| = |G'| \leq |S|/2$, and all bijections $(f : G \rightarrow G', i \in S - G)$ there exists $s \in S - G, u \in G$ such that $u \succ_{\pi_s} f(u)$.*

Note that the property is weaker than the GS Axiom in two ways: we restrict ourselves to groups G of size at most $|S|/2$, and we only allow for a global bijection f , rather than individual bijections f_s .

PROPOSITION 4.11. *$\mathcal{C}_{\text{harmonious}}$ is weakly group stable, while $\mathcal{C}_{\text{democratic}}$ and $\mathcal{C}_{\text{Borda}}$ are not.*

PROOF. Consider a set $S \in \mathcal{C}_{\text{harmonious}}(V, \Pi)$, subsets $G \subset S$ and $G' \subset V - S$ such that $0 < |G| = |G'| \leq |S|/2$, and a bijection $f : G \rightarrow G'$. For each $u \in G$ the majority of S prefer u to $f(u)$ (who is not a member of S), and since $|G| \leq |S|/2$, this implies that there must be at least one $s \in S - G$ such that s prefers u to $f(u)$, as required.

To give a counterexample for both $\mathcal{C}_{\text{democratic}}$ and $\mathcal{C}_{\text{Borda}}$, consider $V = [1 : 6]$, $G = [3 : 4]$ and $G' = [5 : 6]$, with preference profiles

$$\pi_1 = [125463], \pi_2 = [126354], \pi_3 = [341256], \pi_4 = [341256].$$

Then 1 and 2 prefer 5 over 4, and 6 over 3, but S is a community both with respect to $\mathcal{C}_{\text{democratic}}$ (where 1 and 2 get four votes, 3 and 4 get three votes, and 5 and 6 get only one vote), and with respect to Borda count (with counts 20, 16, 18, 16, 10, 8 for 1, ..., 6, respectively). \square

We will now compare fixed-point community rules based on the three aggregation functions that we have discussed so far: Borda voting, democratic voting, and $F_{\mathcal{H}}$. While all three have their own appealing simplicity and intuition, and all satisfy Axioms A, SA, Emb, and WC, they have significant differences with respect to Axioms Mon and GS, and the Outsider Departure property. (1) **Outsider Departure:** A harmonious community S remains a harmonious community when any outsider $v \notin S$ leaves the system, since the

departure does not alter any pairwise preferences. However, for a $\mathcal{C}_{\text{democratic}}$ or $\mathcal{C}_{\text{Borda}}$ community S , the departure of an outsider can increase the votes for other outsiders to destabilize the community. (2) **Monotonicity:** The harmonious rule satisfies Axiom Mon. The other two only satisfy the weaker **Outsider Respecting Monotonicity** property³. (3) **Group Stability:** The subset T in the proof of Theorem 4.7 is a community according to all these three community rules. But T violates GS because 1 prefers outsiders over 5 and 6, even though 5 and 6 prefer 1 over everyone else: Element 1 is an “arrogant” member of its community. All aggregation functions satisfying **Unanimity** seem to be prone to existence of “arrogant” members. The harmonious rule satisfies the stability of majority subgroup under a global bijection f . $\mathcal{C}_{\text{democratic}}$ and $\mathcal{C}_{\text{Borda}}$ essentially have no guarantee on group stability. (4) **Small World:** We say a community function \mathcal{C} satisfies the **Small World** property if

$$S \in \mathcal{C}((V, \Pi)) \text{ if and only if } S \in \mathcal{C}(S \cup U, \Pi|_{S \cup U}), \\ \forall U \subseteq V - S, |U| < |S|.$$

This Helly-type property [7] localizes the identification of a community. Note that the **Small World** property includes some form of **Outsider Departure** together with the property that every community is “locally” verifiable. One can easily show that $\mathcal{C}_{\text{democratic}}$ and $\mathcal{C}_{\text{Borda}}$ do not have the **Small World** property, while $\mathcal{C}_{\text{harmonious}}$ enjoys the following stronger variant of the small world property

$$S \in \mathcal{C}_{\text{harmonious}}((V, \Pi)) \text{ if and only if } \forall v \in V - S, \\ S \in \mathcal{C}_{\text{harmonious}}(S \cup \{v\}, \Pi|_{S \cup \{v\}}).$$

4.4 Stability of Communities

In real-world social interactions, some communities are more stable or durable than others, when people’s interests and preferences evolve over time. For example, some music bands stay together longer than others. We consider the following stability model that is inspired by the work of Balcan *et al.* [3] and Mishra *et al.* [14].

DEFINITION 4.12. (PREFERENCE PERTURBATIONS) *Let \mathcal{C} be a community rule and $A = (V, \Pi)$ be a preference network. For $\delta \in (0, 1)$, we say a community $S \in \mathcal{C}(A)$ is stable under δ -perturbations if $S \in \mathcal{C}((V, \Pi'))$ for all preference profiles Π' such that $|\{i \in S : \pi_i(v) \neq \pi'_i(v)\}| \leq \delta|S|, \forall v \in V$.*

In other words, stable communities remain communities even after some changes to their members’ preferences.

Both the community rule \mathcal{C} and stability parameter δ can impact community structures. The main result of Balcan *et al.* [3] can be restated as:

THEOREM 4.13. ([3]) *For $\delta \in (0, 1)$, every preference network $A = (V, \Pi)$ has $n^{O(\log(1/\delta)/\delta)}$ democratic communities, that are stable under δ -perturbations of Π .*

Definition 4.12 can be directly applied to $\mathcal{C}_{\text{harmonious}}$, which is connected with the following natural notion of robust harmonious communities: For $\delta \in [0 : 1/2]$, a non-empty subset S is a $(\delta + 1/2)$ -harmonious community in $A = (V, \Pi)$ if $\forall u \in S, v \in V - S$, at least $(1/2 + \delta)$ -fraction of $\{\pi_s : s \in S\}$ prefer u over v .

³Again, we can use the profiles from the proof of Theorem 4.7 to show that the Borda count rule does not satisfy Mon.

PROPOSITION 4.14. For any $\delta \in [0 : 1/2]$, and preference network A , every $(1/2 + \delta)$ -harmonious community is stable under $\delta/2$ -perturbations. Conversely, any community in $\mathcal{C}_{\text{harmonious}}$ that is stable under δ perturbations is a $(1/2 + \delta)$ -harmonious community.

Using a simple probabilistic argument, we can establish the following bound:

THEOREM 4.15. For any $\delta \in [0 : 1/2]$, the number of δ -stable harmonious communities in any preference network is at most $n^{12 \log n / \delta^2}$.

PROOF. Let S be a δ -stable harmonious communities. For any multi-set $T \subseteq S$, we say T identifies S if for all $u \in S$ and $v \in V - S$, the majority of T prefer u to v . Note that such a T determines S once the size of S is set. To see this, note that the condition implies that $u \succ_{F(\Pi_T)} v$ for all $(u, v) \in S \times (V - S)$, which in turn implies that S is of the form $V_1 \cup \dots \cup V_i$ where (V_1, V_2, \dots) are the components of the ordered partition $F(\Pi_T)$, ordered in such a way that $V_1 \succ_{F(\Pi_T)} V_2, \dots$ (see Proposition 4.9 and its proof). Thus once $F(\Pi_T)$ and the size of S are fixed, S is uniquely determined.

We now show that $\exists T \subset V$ of size $12 \log n / \delta^2$ that identifies S . To this end, we consider a sample $T \subset S$ of $k = 12 \log n / \delta^2$ randomly chosen elements (with replacements). We analyze the probability that T identifies S . Let $T = \{t_1, \dots, t_k\}$, and for each $u \in S$ and $v \in V - S$, let $x_i^{(u,v)} = [u \succ_{\pi_{t_i}} v]$, where $[B]$ denotes the indicator variable of an event B . Then T identifies S iff $\sum_{i=1}^k x_i^{(u,v)} > k/2, \forall u \in S, v \in V - S$. We now focus on a particular (u, v) pair and bound $\Pr \left[\sum_{i=1}^k x_i^{(u,v)} \leq k/2 \right]$. We first note that

$$\mathbf{E} \left[\sum_{i=1}^k x_i^{(u,v)} \right] = \sum_{i=1}^k \mathbf{E} \left[x_i^{(u,v)} \right] \geq \left(\frac{1}{2} + \delta \right) \cdot k.$$

By a standard use of the Chernoff-Hoeffding bound

$$\begin{aligned} \Pr \left[\sum_{i=1}^k x_i^{(u,v)} \leq k/2 \right] &\leq \Pr \left[\sum_{i=1}^k x_i^{(u,v)} \leq (1 + 2\delta)^{-1} \mathbf{E} \left[\sum_{i=1}^k x_i^{(u,v)} \right] \right] \\ &\leq \Pr \left[\sum_{i=1}^k x_i^{(u,v)} \leq (1 - \delta) \mathbf{E} \left[\sum_{i=1}^k x_i^{(u,v)} \right] \right] \\ &\leq e^{-\frac{\delta^2}{2} (1/2 + \delta) k} \leq e^{-\frac{\delta^2}{4} k} \leq \frac{1}{n^3}, \end{aligned}$$

where we used that $(1 + 2\delta)^{-1} \leq 1 - \delta$ in the third step. If T fails to identify S , then there exists $(u \in S, v \in V - S)$ such that $\sum_{i=1}^k x_i^{(u,v)} \leq k/2$. As there are at most $|S||V - S| \leq n^2$ such (u, v) pairs to consider, by the union bound,

$$\begin{aligned} \Pr [T \text{ identifies } S] &\geq 1 - \sum_{u \in S, v \in V - S} \Pr \left[\sum_{i=1}^k x_i^{(u,v)} \leq k/2 \right] \\ &> 1 - 1/n > 0. \end{aligned}$$

Thus, if S is a δ -stable harmonious communities, then there exists a multi-set $T \subset V$ of size $12 \log n / \delta^2$ that identifies S . We can thus enumerate all δ -stable harmonious communities by enumerating all (T, t) pairs, where T ranges from

all multi-subsets of V of size $12 \log n / \delta^2$ and $t \in [1 : n]$ and check if T can identify a set of size t . \square

5. REMARKS

Our work can be partially applied to other network models. With simple modifications to our axioms, we can extend results of this paper to preference networks that allow *indifferences*. In other words, each preference network $A = (V, \Pi)$ is now given by n ordered partitions $\pi_1, \dots, \pi_n \in \overline{L(V)}$. By allowing indifference, we can partially apply our results to practical social networks, where community identification is posed as the problem of detecting communities in an observed social network, $G = (V, E)$, which is usually sparse. To apply our framework, we first define an affinity network $(V, [\mathbf{w}_1, \dots, \mathbf{w}_n])$, where \mathbf{w}_v is the personalized PageRank vector of vertex v [11]. We then obtain a preference network $A_G = (V, \Pi)$: For each $v \in V$, we extract a preference vector, $\pi_v \in \overline{L(V)}$ from \mathbf{w}_v , by sorting entries in \mathbf{w}_v — elements with the same weight are assigned to the same partition. In other words, $\pi_v \in \overline{L(V)}$ ranks vertices in V by v 's PageRank contributions [1] to them. Although this conversion may lose some valuable affinity information encoded in the numerical values, it offers a path to apply our community identification theory to network analysis.

To better model preferences in practical social networks, we can also use *multifaceted preference networks* [3], in which each node can have more than one ranking of other nodes. For example, one member may have three rankings — one based on “family/friends”, one based on “academic interests”, and one based on “business interests.” Meanwhile, another member may have two rankings — one based on “sports” and one based on “music”. Formally, in a *multifaceted preference network* $A = (V, \Pi)$, each $u \in V$ specifies d_u preferences in Π . We call d_u the *preference degree* of u . Our results extend to multifaceted preference networks.

A real-world (observed) social network may be viewed as sparse, observed social interactions induced by an underlying preference/affinity network. Thus, the conceptual question of community identification studied in this paper is a basic question in Network Sciences regarding both network formation and network structures. Our work suggests that simultaneous axiomatization of network formation and community characterization could be beneficial and essential. We expect that an axiomatization theory, for (1) personalized ranking in networks and (2) for community characterization in preference networks with partially ordered preferences, will offer us new insight for addressing the two fundamental mathematical questions, that are essential for understanding community formation in social/information networks:

- Inference of the underlying network model from observed networks.
- Community formulation from individual affinities and preferences (based on the underlying network model).

Extending our work to preference networks with partially ordered preferences will provide a concrete step to understand community formation in networks with incomplete or incomparable preferences. Like our current study, we believe that the existing literature in social choice — e.g., [18] — will be valuable to our understanding. For both fundamental problems, we can also consider other frameworks, such as

the game-theory based incentive networks [26], for formulating the underlying network model from observed networks. These network models offer richer structures for capturing interactions among members in networks. As both parts of theory become sufficiently well developed, well-designed experiments with real-world social networks will be necessary to further enhance this theoretical framework.

In summary, our taxonomy theorem provides the basic structure of communities in a preference network, while the complexity (coNP-Completeness) result illustrates the algorithmic challenges for community identification in addition to community enumeration. On the other hand, our analysis of the harmonious rule and the work of [3] seem to suggest some efficient notion of communities can be defined. However, it remains an open question if there exists a natural, constructive and axiom-conforming community rule that allows overlapping communities, and has stable communities which are polynomial-time samplable and enumerable. Finally, we hope that we can further develop our axiomatic system to better connect with practical networks.

Acknowledgements

We thank Nina Balcan of CMU, Mark Braverman of Princeton, and Madhu Sudan of Microsoft Research for the insightful discussions. We also thank Gang Zeng and Xingwu Liu of Chinese Academy of Sciences for pointing out a subtle error in an earlier version of this work.

6. REFERENCES

- [1] R. Andersen, C. Borgs, J. Chayes, J. Hopcraft, V. S. Mirrokni, and S.-H. Teng. Local computation of pagerank contributions. In *Proceedings of the 5th International Conference on Algorithms and Models for the Web-graph*, WAW'07, pages 150–165. Springer-Verlag, 2007.
- [2] K. J. Arrow. *Social Choice and Individual Values*. Wiley, New York, 2nd edition, 1963.
- [3] M. F. Balcan, C. Borgs, M. Braverman, J. T. Chayes, and S.-H. Teng. Finding endogenously formed communities. In *SODA*, pages 767–783, 2013.
- [4] O. N. Bondareva. Some applications of the methods of linear programming to the theory of cooperative games. *Problemy Kibernet.*, (10):119–139, 1963.
- [5] S. J. Brams, M. A. Jones, and D. M. Kilgour. Dynamic models of coalition formation: Fallback vs. build-up. In *Proceedings of the 9th Conference on Theoretical Aspects of Rationality and Knowledge*, TARK '03, pages 187–200, 2003.
- [6] M. Caesar and J. Rexford. Bgp routing policies in isp networks. *Netw. Mag. of Global Internetwkg.*, 19(6):5–11, Nov. 2005.
- [7] L. Danzer, Grünbaum, and V. Klee. Helly's theorem and its relatives. *Proc. Symp. Pure Math. 7, American Mathematical Society*, pages 101–179, 1963.
- [8] D. Gale and L. S. Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.
- [9] R. Ghosh, S.-H. Teng, K. Lerman, and X. Yan. The interplay between dynamics and networks: Centrality, communities, and Cheeger inequality. In *KDD 2014*, pages 1406–1415. ACM, Aug. 2014.
- [10] D. Gusfield and R. W. Irving. *The Stable Marriage Problem: Structure and Algorithms*. MIT Press, Cambridge, MA, USA, 1989.
- [11] T. Haveliwala. Topic-sensitive pagerank: A context-sensitive ranking algorithm for web search. In *Trans. Knowl. Data Eng.*, volume 15(4), pages 784 – 796, 2003.
- [12] J. Kleinberg. An impossibility theorem for clustering. In *NIPS*, pages 463–470, 2002.
- [13] J. Leskovec, K. Lang, A. Dasgupta, and M. Mahoney. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics*, pages 29–123, 2009.
- [14] N. Mishra, R. Schreiber, I. Stanton, and R. Tarjan. Finding strongly-knit clusters in social networks. *Internet Mathematics*, pages 155–174, 2009.
- [15] J. Nash. Equilibrium points in n-person games. *Proceedings of the National Academy of the USA*, 36(1):48–49, 1950.
- [16] J. Nash. Noncooperative games. *Annals of Mathematics*, 54:289–295, 1951.
- [17] M. E. J. Newman. Modularity and community structure in networks. *PNAS*, 2006.
- [18] M. S. Pini, F. Rossi, K. B. Venable, and T. Walsh. Aggregating partially ordered preferences: Impossibility and possibility results. *TARK*, 10, 2005.
- [19] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi. Defining and identifying communities in networks. *PNAS*, 2004.
- [20] Y. Rekhter and T. Li. A border gateway protocol4. *IETF RFC 1771*, 1995.
- [21] A. 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.
- [22] A. E. Roth. *Stable Coalition Formation: Aspects of a Dynamic Theory*, pages 228–234. Wuerzberg: Physica-Verlag, 1984.
- [23] H. E. Scarf. The core of an N person game. *Econometrica*, 69:35–50, 1967.
- [24] L. S. Shapley. On balanced sets and cores. *Naval Res. Logist. Quarter.*, 14, 1967.
- [25] D. A. Spielman and S.-H. Teng. A local clustering algorithm for massive graphs and its application to nearly-linear time graph partitioning. *SIAM J. Comput.*, 42(1):1–26, 2013.
- [26] S.-H. Teng. *Scalable Algorithms for Data and Network Analysis*. Foundations and Trends in Theoretical Computer Science, 2016.
- [27] H. P. Young. An axiomatization of Borda's rule. *Journal of Economic Theory*, 9(1):43–52, 1974.