Artificial Intelligence 245 (2017) 86–114

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

Graph aggregation ☆

Ulle Endriss [a], Umberto Grandi [b]

[a] ILLC, University of Amsterdam, The Netherlands
[b] IRIT, University of Toulouse, France

ARTICLE INFO

Article history:
Received 20 June 2016
Received in revised form 6 January 2017
Accepted 8 January 2017

Keywords:
Social choice theory
Collective rationality
Impossibility theorems
Graph theory
Modal logic
Preference aggregation
Belief merging
Consensus clustering
Argumentation theory

ABSTRACT

Graph aggregation is the process of computing a single output graph that constitutes a good compromise between several input graphs, each provided by a different source. One needs to perform graph aggregation in a wide variety of situations, e.g., when applying a voting rule (graphs as preference orders), when consolidating conflicting views regarding the relationships between arguments in a debate (graphs as abstract argumentation frameworks), or when computing a consensus between several alternative clusterings of a given dataset (graphs as equivalence relations). In this paper, we introduce a formal framework for graph aggregation grounded in social choice theory. Our focus is on understanding which properties shared by the individual input graphs will transfer to the output graph returned by a given aggregation rule. We consider both common properties of graphs, such as transitivity and reflexivity, and arbitrary properties expressible in certain fragments of modal logic. Our results establish several connections between the types of properties preserved under aggregation and the choice-theoretic axioms satisfied by the rules used. The most important of these results is a powerful impossibility theorem that generalises Arrow's seminal result for the aggregation of preference orders to a large collection of different types of graphs.

1. Introduction

Suppose each of the members of a group of autonomous agents provides us with a different directed graph that is defined on a common set of vertices. Graph aggregation is the task of computing a single graph over the same set of vertices that, in some sense, represents a good compromise between the various individual views expressed by the agents. Graphs are ubiquitous in computer science and artificial intelligence (AI). For example, in the context of decision support systems, an edge from vertex x to vertex y might indicate that alternative x is *preferred* to alternative y. In the context of modelling interactions taking place on an online debating platform, an edge from x to y might indicate that argument x

☆ This work refines and extends papers presented at COMSOC-2012 [1] and ECAI-2014 [2]. We are grateful for the extensive feedback received from Davide Grossi, Sylvie Doutre, Weiwei Chen, several anonymous reviewers, and the audiences at the SSEAC Workshop on Social Choice and Social Software held in Kiel in 2012, the Dagstuhl Seminar on Computation and Incentives in Social Choice in 2012, the KNAW Academy Colloquium on Dependence Logic held at the Royal Netherlands Academy of Arts and Sciences in Amsterdam in 2014, a course on logical frameworks for multiagent aggregation given at the 26th European Summer School in Logic, Language and Information (ESSLLI-2014) in Tübingen in 2014, the Lorentz Center Workshop on Clusters, Games and Axioms held in Leiden in 2015, the SEGA Workshop on Shared Evidence and Group Attitudes held in Prague in 2016, and lectures delivered at Sun Yat-Sen University in Guangzhou in 2014 as well as École Normale Supérieure and Pierre & Marie Curie University in Paris in 2016. This work was partly supported by COST Action IC1205 on Computational Social Choice. It was completed while the first author was hosted at the University of Toulouse in 2015 as well as Paris-Dauphine University, Pierre & Marie Curie University, and the London School of Economics in 2016.

E-mail addresses: ulle.endriss@uva.nl (U. Endriss), umberto.grandi@irit.fr (U. Grandi).

http://dx.doi.org/10.1016/j.artint.2017.01.001
0004-3702/

undercuts or otherwise *attacks* argument y. And in the context of social network analysis, an edge from x to y might express that person x is *influenced* by person y. How to best perform graph aggregation is a relevant question in these three domains, as well as in any other domain where graphs are used as a modelling tool and where particular graphs may be supplied by different agents or originate from different sources. For example, in an election, i.e., in a group decision making context, we have to aggregate the preferences of several voters. In a debate, we sometimes have to aggregate the views of the individual participants in the debate. And when trying to understand the dynamics within a community, we sometimes have to aggregate information coming from several different social networks.

In this paper, we introduce a formal framework for studying graph aggregation in general abstract terms and we discuss in detail how this general framework can be instantiated to specific application scenarios. We introduce a number of concrete methods for performing aggregation, but more importantly, our framework provides tools for evaluating what constitutes a "good" method of aggregation and it allows us to ask questions regarding the existence of methods that meet a certain set of requirements. Our approach is inspired by work in social choice theory [3], which offers a rich framework for the study of aggregation rules for preferences—a very specific class of graphs. In particular, we adopt the *axiomatic method* used in social choice theory, as well as other parts of economic theory, to identify intuitively desirable properties of aggregation methods, to define them in mathematically precise terms, and to systematically explore their logical consequences.

An aggregation rule maps any given *profile* of graphs, one for each agent, into a single graph, which we are often going to refer to as the *collective graph*. The central concept we focus on in this paper is the *collective rationality* of aggregation rules with respect to certain properties of graphs. Suppose we consider an agent rational only if the graph she provides has certain properties, such as being reflexive or transitive. Then we say that a given aggregation rule F is collectively rational with respect to that property of interest if and only if F can guarantee that that property is preserved during aggregation. For example, if we aggregate individual graphs by computing their *union* (i.e., if we include an edge from x to y in our collective graph if at least one of the individual graphs includes that edge), then it is easy to see that the property of *reflexivity* will always transfer. On the other hand, the property of *transitivity* will not always transfer. For example, if we aggregate two graphs over the set of vertices $V = \{x, y, z\}$, one consisting only of the edge (x, y) and one consisting only of the edge (y, z), then although each of these two graphs is (vacuously) transitive, their union is not, as it is missing the edge (x, z). Thus, the union rule is collectively rational with respect to reflexivity, but not with respect to transitivity.

We study collective rationality with respect to some such well-known and widely used properties of graphs, but also with respect to large families of graph properties that satisfy certain *meta-properties*. We explore both a semantic and a syntactic approach to defining such meta-properties. In our semantic approach, we identify three high-level features of graph properties that determine the kind of aggregation rules that are collectively rational with respect to them. For example, transitivity is what we call a "contagious" property: under certain circumstances, namely in the presence of edge (y, z), inclusion of (x, y) *spreads* to (x, z). Transitivity also satisfies a second meta-property, which we call "implicativeness": the inclusion of two specific edges, namely (x, y) and (y, z), *implies* the inclusion of a third edge, namely (x, z). The third meta-property we introduce, "disjunctiveness", expresses that, under certain circumstances, at least one of two specific edges has to be accepted. This is satisfied, for instance, by the property of *completeness*: every two vertices x and y need to be connected in at least one of the two possible directions. In our syntactic approach, we consider graph properties that can be expressed in particular syntactic fragments of a logical language. To this end, we make use of the language of modal logic [4]. This allows us to establish links between the syntactic properties of the language used to express the integrity constraints we would like to see preserved during aggregation and the axiomatic properties of the rules used.

We prove both *possibility* and *impossibility results*. A possibility result establishes that every aggregation rule belonging to a certain class of rules (typically defined in terms of certain axioms) is collectively rational with respect to all graph properties that satisfy a certain meta-property. An impossibility result, on the other hand, establishes that it is impossible to define an aggregation rule belonging to a certain class that would be collectively rational with respect to any graph property that meets a certain meta-property—or that the only such aggregation rules would be clearly very unattractive for other reasons. Our main result is such an impossibility theorem. It is a generalisation of Arrow's seminal result for preference aggregation [5], which we shall recall in Section 3.1. Our approach of working with meta-properties has two advantages. First, it permits us to give conceptually simple proofs for powerful results with a high degree of generality. Second, it makes it easy to instantiate our general results to obtain specific results for specific application scenarios. For example, Arrow's Theorem follows immediately from our more general result by checking that the properties of graphs that represent preference orders (namely transitivity and completeness) satisfy the meta-properties featuring in our theorem, yet our proof of the general theorem is arguably simpler than a direct proof of Arrow's Theorem. This is so, because the meta-properties we use very explicitly exhibit specific features required for the proof, while those features are somewhat hidden in the specific properties of transitivity and completeness. Similarly, we show how alternative instantiations of our general result easily generate both known and new results in other domains, such as the aggregation of plausibility orders (which has applications in nonmonotonic reasoning and belief merging) and the aggregation of equivalence relations (which has applications in clustering analysis).

Related work. Our work builds on and is related to contributions in the field of social choice theory, starting with the seminal contribution of Arrow [5]. This concerns, in particular, contributions to the theory of voting and preference aggregation [6–10,3], but also judgment aggregation [11–17]. In fact, in terms of levels of generality, graph aggregation may be regarded as occupying the middle ground between preference aggregation (most specific) and judgment aggregation (most general). In computer science, these frameworks are studied in the field of computational social choice [18]. As we shall discuss in some detail, graph aggregation is an abstraction of several more specific forms of aggregation taking place in a wide range of different domains. Preference aggregation is but one example. Aggregation of specific types of graphs has been studied, for instance, in nonmonotonic reasoning [19], belief merging [20], social network analysis [21], clustering [22], and argumentation in multiagent systems [23]. As we shall see, several of the results obtained in these earlier contributions are simple corollaries of our general results on graph aggregation.

Paper overview. The remainder of this paper is organised as follows. In Section 2, we introduce our framework for graph aggregation. This includes the discussion of several application scenarios, the definition of a number of concrete aggregation rules, and the formulation of various axioms identifying intuitively desirable properties of such rules. It also includes the definition of the concept of collective rationality. Finally, we prove a number of basic results in Section 2: characterisation results linking rules and axioms, as well as possibility results linking axioms and collective rationality requirements. In Section 3, we present our impossibility results for graph aggregation rules that are collectively rational with respect to graph properties meeting certain semantically defined meta-properties. There are two such results. One identifies conditions under which the only available rules are so-called *oligarchies*, under which the outcome is always the intersection of the graphs provided by a subset of the agents (the oligarchs). A second result shows that, under slightly stronger assumptions, the only available rules are the *dictatorships*, where a single agent completely determines the outcome for every possible profile. Much of Section 3 is devoted to the definition and illustration of the meta-properties featuring in these results. Once they are in place, the proofs are relatively simple. In Section 4, we introduce our approach to describing collective rationality requirements in syntactic terms, using the language of modal logic. Our results in Section 4 establish simple conditions on the syntax of the specification of a graph property that are sufficient for guaranteeing that the property in question will be preserved under aggregation. The grounding of our approach in modal logic also allows us to provide a deeper analysis of the concept of collective rationality by considering the preservation of properties at three different levels, corresponding to the three levels naturally defined by the notions of Kripke frame, Kripke model, and possible world, respectively. In Section 5, we discuss four of our application scenarios in more detail, focusing on application scenarios previously discussed in the AI literature. We show how our general results allow us to derive new simple proofs of known results, how they clarify the status of some of these results, and how they allow us to obtain new results in these domains of application. Section 6, finally, concludes with a brief summary of our results and pointers to possible directions for future work.

2. Graph aggregation

In this section, we introduce a simple framework for graph aggregation. The basic definitions are given in Section 2.1. While this is a general framework that is independent of specific application scenarios and specific choices regarding the aggregation rule used, we briefly discuss several such specific scenarios in Section 2.2 and suggest definitions for several specific aggregation rules in Section 2.3. We then approach the analysis of aggregation rules from two different but complementary angles. First, in Section 2.4, we define several *axiomatic properties* of aggregation rules that a user may wish to impose as requirements when looking for a "fair" or "well-behaved" aggregation rule for a specific application. We also prove a number of simple results that show how some of these axioms relate to each other and to some of the aggregation rules defined earlier. Second, in Section 2.5, we introduce the central concept of *collective rationality* and we prove a number of simple positive results that show how enforcing certain axioms allows us to guarantee collective rationality with respect to certain graph properties.

2.1. Basic notation and terminology

Fix a finite set of *vertices* V. A (directed) *graph* $G = \langle V, E\rangle$ based on V is defined by a set of *edges* $E \subseteq V \times V$. We write xEy for $(x, y) \in E$. As V is fixed, G is in fact fully determined by E. We therefore identify sets of edges $E \subseteq V \times V$ with the graphs $G = \langle V, E\rangle$ they define. For any kind of set S, we use 2^S to denote the powerset of S. So $2^{V\times V}$ is the set of all graphs. We use $E(x) := \{y \in V \mid (x, y) \in E\}$ to denote the set of *successors* of a vertex x in a set of edges E and $E^{-1}(y) := \{x \in V \mid (x, y) \in E\}$ to denote the set of *predecessors* of y in E.

A given graph may or may not satisfy a specific *property*, such as transitivity or reflexivity. Table 1 recalls the definitions of several such properties.[1] We are often going to be interested in families of graphs that all satisfy several of these properties. For instance, a *weak order* is a directed graph that is reflexive, transitive, and complete. It will often be useful to think of a graph property P, such as transitivity, as a subset of $2^{V\times V}$ (the set of all graphs over the set of vertices V). For two

[1] Some of these may be less well known than others, so let us briefly review the less familiar definitions. The two Euclidean properties encode Euclid's idea that "things which equal the same thing also equal one another". Negative transitivity, a property commonly assumed in the economics literature on preferences, may equivalently be expressed as $\forall xyz.[(\neg xEy \wedge \neg yEz) \rightarrow \neg xEz]$, which explains the name of the property. Completeness requires any two distinct vertices to be related one way or the other. Connectedness only requires two (not necessarily distinct) vertices to be related one way or the other if they are both reachable from some common predecessor (the term "connectedness" is commonly used in the modal logic literature [4]). Nontriviality excludes the empty graph, while seriality (also a term used in the modal logic literature) requires every vertex to have at least one successor.

Table 1
Common properties of directed graphs.

Property	First-order condition
Reflexivity	$\forall x.xEx$
Irreflexivity	$\neg\exists x.xEx$
Symmetry	$\forall xy.(xEy \to yEx)$
Antisymmetry	$\forall xy.(xEy \wedge yEx \to x = y)$
Right Euclidean	$\forall xyz.[(xEy \wedge xEz) \to yEz]$
Left Euclidean	$\forall xyz.[(xEy \wedge zEy) \to zEx]$
Transitivity	$\forall xyz.[(xEy \wedge yEz) \to xEz]$
Negative Transitivity	$\forall xyz.[xEy \to (xEz \vee zEy)]$
Connectedness	$\forall xyz.[(xEy \wedge xEz) \to (yEz \vee zEy)]$
Completeness	$\forall xy.[x \neq y \to (xEy \vee yEx)]$
Nontriviality	$\exists xy.xEy$
Seriality	$\forall x.\exists y.xEy$

disjoint sets of edges S^+ and S^- and a graph property $P \subseteq 2^{V\times V}$, let $P[S^+, S^-] = \{E \in P \mid S^+ \subseteq E \text{ and } S^- \cap E = \emptyset\}$ denote the set of graphs in P that include all of the edges in S^+ and none of those in S^-.

Let $\mathcal{N} = \{1, \dots, n\}$ be a finite set of (two or more) *individuals* (or *agents*). We are often going to refer to subsets of $\mathcal{N}$ as *coalitions* of individuals. Suppose every individual $i \in \mathcal{N}$ specifies a graph $E_i \subseteq V \times V$. This gives rise to a *profile* $\boldsymbol{E} = (E_1, \dots, E_n)$. We use $N_e^{\boldsymbol{E}} := \{i \in \mathcal{N} \mid e \in E_i\}$ to denote the coalition of individuals accepting edge e under profile $\boldsymbol{E}$.

Definition 1. An **aggregation rule** is a function $F : (2^{V\times V})^n \to 2^{V\times V}$, mapping any given profile of individual graphs into a single graph.

We are sometimes going to denote the outcome $F(\boldsymbol{E})$ obtained when applying an aggregation rule F to a profile $\boldsymbol{E}$ simply as E and refer to it as the *collective graph*. An example for an aggregation rule is the *majority rule*, accepting a given edge if and only if more than half of the individuals accept it. More examples are going to be provided in Section 2.3.

2.2. Examples of application scenarios

Directed graphs are ubiquitous in computer science and beyond. They have been used as modelling devices for a wide range of applications. We now sketch a number of different application scenarios for graph aggregation, each requiring different types of graphs (satisfying different properties) to model relevant objects of interest, and each requiring different types of aggregation rules.

Example 1 *(Preferences).* Our main example for a graph aggregation problem is going to be preference aggregation as classically studied in social choice theory [5]. In this context, vertices are interpreted as alternatives available in an election and the graphs considered are weak orders on these alternatives, interpreted as preference orders. Our aggregation rules then reduce to so-called social welfare functions. Social welfare functions, which return a preference order for every profile of individual preference orders, are similar objects as voting rules, which only return a winning alternative for every profile. While the types of preferences typically considered in classical social choice theory are required to be complete, recent work in AI has also addressed the aggregation of partial preference orders [10], corresponding to a larger family of graphs than the weak orders. In the context of aggregating complex preferences defined over combinatorial domains, graph aggregation can also be used to decide which preferential dependencies between different variables one should try to respect, based on the dependencies reported by the individual decision makers [24].

Example 2 *(Knowledge).* If we think of V as a set of possible worlds, then a graph on V that is reflexive and transitive (and possibly also symmetric) can be used to model an agent's knowledge: (x, y) being an edge means that, if x is the actual world, then our agent will consider y a possible world [25]. If we aggregate the graphs of several agents by taking their intersection, then the resulting collective graph represents the distributed knowledge of the group, i.e., the knowledge the members of the group can infer by pooling all their individual resources. If, on the other hand, we aggregate by taking the union of the individual graphs, then we obtain what is sometimes called the shared or mutual knowledge of the individual agents, i.e., the part of the knowledge available to each and every individual on their own. Finally, if we aggregate by computing the transitive closure of the union of the individual graphs, then we obtain a model of the group's common knowledge [26, p. 512]. These concepts play a role in disciplines as diverse as epistemology [27], game theory [28], and distributed systems [29].

Example 3 *(Nonmonotonic reasoning).* When an intelligent agent attempts to update her beliefs or to decide what action to take, she may resort to several patterns of common-sense inference that will sometimes be in conflict with each other. To take a famous example, we may wish to infer that Nixon is a pacifist, because he is a Quaker and Quakers by default are pacifists, and we may at the same time wish to infer that Nixon is not a pacifist, because he is a Republican and Republicans by default are not pacifists. In a popular approach to nonmonotonic reasoning in AI, such default inference rules are modelled as graphs that encode the relative plausibility of different conclusions [30]. Thus, here the possible conclusions are the vertices and we obtain a graph by linking one vertex with another, if the former is considered at least as plausible as the latter. Conflict resolution between different rules of inference then requires us to aggregate such plausibility orders, to be able to determine what the ultimately most plausible state of the world might be [19].

Example 4 *(Social networks).* We may also think of each of the graphs in a profile as a different social network relating members of the same population. One of these networks might describe work relations, another might model family relations, and a third might have been induced from similarities in online purchasing behaviour. Social networks are often modelled using undirected graphs, which we can simulate in our framework by requiring all graphs to be symmetric. Aggregating individual graphs then amounts to finding a single meta-network that describes relationships at a global level. Alternatively, we may wish to aggregate several graphs representing snapshots of the same social network at different points in time. The meta-network obtained can be helpful when studying the social structures within the population under scrutiny [21].

Example 5 *(Clustering).* Clustering is the attempt of partitioning a given set of data points into several clusters. The intention is that the data points in the same cluster should be more similar to each other than each of them is to data points belonging to one of the other clusters. This is useful in many disciplines, including information retrieval and molecular biology, to name but two examples. However, the field is lacking a precise definition of what constitutes a “correct” partitioning of the data and there are many different clustering algorithms, such as k-means or single-linkage clustering, and even more parameterisations of those basic algorithms [31]. Observe that every partitioning that might get returned by a clustering algorithm induces an equivalence relation (i.e., a graph that is reflexive, symmetric, and transitive): two data points are equivalent if and only if they belong to the same cluster. Finding a compromise between the solutions suggested by several clustering algorithms is what is known as consensus clustering [32]. This thus amounts to aggregating several graphs that are equivalence relations.

Example 6 *(Argumentation).* In a so-called abstract argumentation framework, arguments are taken to be vertices in a graph and attacks between arguments are modelled as directed edges between them [33]. A graph property of interest in this context is acyclicity, as that makes it easier to decide which arguments to ultimately accept. If we think of V as the collection of arguments proposed in a debate, a profile $\boldsymbol{E} = (E_1, \dots, E_n)$ specifies an attack relation for each of a number of agents that we may wish to aggregate into a collective attack relation before attempting to determine which of the arguments might be acceptable to the group. Recent work has addressed the challenge of aggregating several abstract argumentation frameworks from a number of angles, e.g., by proposing concrete aggregation methods grounded in work on belief merging [34], by investigating the computational complexity of aggregation [35], and by analysing what kinds of profiles we may reasonably expect to encounter in this context [36].

Example 7 *(Logic).* Graph aggregation is also at the core of recent work on the aggregation of different logics [37]. The central idea here is that every logic is defined by a consequence relation between formulas. Thus, given a set of formulas, we can think of a logic $\mathcal{L}$ as the graph corresponding to the consequence relation defining $\mathcal{L}$. Aggregating several such graphs then gives rise to a new logic. Thus, this is an instance of our graph aggregation problem, except that for the case of logic aggregation it is more natural to model the set of vertices as being infinite.[2]

Recall that we have assumed that every individual specifies a graph on *the same* set of vertices V. This is a natural assumption to make in all of our examples above, but in general we might also be interested in aggregating graphs defined on different sets of vertices. For instance, Coste-Marquis et al. [34] have argued that, in the context of merging argumentation frameworks, the case of agents who are not all aware of the exact same set of arguments is of great practical interest. Observe that also in this case our framework is applicable, as we may think of V as the union of all the individual sets of vertices (with each individual only providing edges involving “her” vertices).

We are going to return to several of these application scenarios in greater detail in Section 5.

2.3. Aggregation rules

Next, we define a number of concrete aggregation rules. We begin with three that are particularly simple, the first of which we have already introduced informally.

Definition 2. The (strict) **majority rule** is the aggregation F_{maj} with $F_{maj} : \boldsymbol{E} \mapsto \{e \in V \times V : |N_e^{\boldsymbol{E}}| > \frac{n}{2}\}$.

[2] All results reported in this paper remain true if we permit graphs with infinite sets V of vertices. However, for ease of exposition and as most applications are more naturally modelled using finite graphs, we do not explore this generalisation here. The finiteness of the set $\mathcal{N}$ of agents, however, is crucial. It is going to be exploited in the proofs of Lemmas 9 and 10 below, on which all of our theorems rely.

Definition 3. The **intersection rule** is the aggregation rule $F_\cap$ with $F_\cap : \boldsymbol{E} \mapsto E_1 \cap \cdots \cap E_n$.

Definition 4. The **union rule** is the aggregation rule $F_\cup$ with $F_\cup : \boldsymbol{E} \mapsto E_1 \cup \cdots \cup E_n$.

In related contexts, the intersection rule is also known as the *unanimity rule*, as it requires unanimous approval from all individuals for an edge to be accepted. Similarly, the union rule is a *nomination rule*, as nomination by just one individual is enough for an edge to get accepted.

Under a *quota rule*, an edge will be included in the collective graph if the number of individuals accepting it meets a certain quota. A *uniform* quota rule uses the same quota for every edge.

Definition 5. A **quota rule** is an aggregation rule F_q defined via a function $q : V \times V \to \{0, 1, \ldots, n+1\}$, associating each edge with a quota, by stipulating $F_q : \boldsymbol{E} \mapsto \{e \in V \times V : |N_e^{\boldsymbol{E}}| \geq q(e)\}$. F_q is called **uniform** in case q is a constant function.

The class of uniform quota rules includes the three simple rules we have seen earlier as special cases: the (strict) *majority rule* F_{maj} is the uniform quota rule with $q = \lceil \frac{n+1}{2} \rceil$, the *intersection rule* $F_\cap$ is the uniform quota rule with $q = n$, and the *union rule* $F_\cup$ is the uniform quota rule with $q = 1$. We call the uniform quota rules with $q = 0$ and $q = n+1$ the *trivial* quota rules; $q = 0$ means that *all* edges will be included in the collective graph and $q = n+1$ means that *no* edge will be included (independently of the profile encountered). The idea of using quota rules is natural and widespread. For example, quota rules have also been studied in judgment aggregation [13].

We now introduce a new class of aggregation rules specifically designed for graphs that is inspired by approval voting [38]. Imagine we associate each vertex with an election in which all the possible successors of that vertex are the candidates (and in which there may be more than one winner). Each agent votes by stating which vertices she considers acceptable successors and, based on this information, a choice function selects which edges to include in the collective graph.

Definition 6. Let $v : (2^V)^n \to 2^V$ be a function associating any given vector of sets of vertices with a single set of vertices. Then the **successor-approval rule** based on v is the aggregation rule F_v defined by stipulating $F_v : \boldsymbol{E} \mapsto \{(x, y) \in V \times V \mid y \in v(E_1(x), \ldots, E_n(x))\}$.

Example 8 *(Successor-approval based on classical approval voting).* Consider a graph with four vertices: $V = \{x, y, z, w\}$. Suppose two individuals report the graphs $E_1 = \{(x, y), (x, z)\}$ and $E_2 = \{(x, z)\}$. When deciding which vertices to connect from x using a successor-approval rule, we look at $E_1(x) = \{y, z\}$ and $E_2(x) = \{z\}$ as approval ballots, and we use v to decide which vertex is the winner. If v is the classical approval voting rule, which selects the candidate with the most approvals, then z is the winner with a score of two approvals, followed by y with one approval, and x and w with none. Since all other vertices have no outgoing edges at all, we have that $F_v(\boldsymbol{E}) = \{(x, z)\}$.

We call v the *choice function* associated with F_v. It takes a vector of sets of vertices, one for each agent, and returns another such set. For example, the classical approval voting rule is formally defined as $v : (S_1, \ldots, S_n) \mapsto \text{argmax}_{x \in S_1 \cup \cdots \cup S_n} |\{i \in \mathcal{N} : x \in S_i\}|$. Note how the argmax-operator ranges over the union of all successors mentioned by any of the agents rather than the full set of vertices V. This ensures that, in case none of the agents approve any vertex as a successor, we do not end up accepting all vertices for all having the same "maximal" support. We are only going to be interested in choice functions v that are (i) *anonymous* and (ii) *neutral*, i.e., for which (i) $v(S_1, \ldots, S_n) = v(S_{\pi(1)}, \ldots, S_{\pi(n)})$ for any permutation $\pi : \mathcal{N} \to \mathcal{N}$ and for which (ii) $\{i \in \mathcal{N} \mid x \in S_i\} = \{i \in \mathcal{N} \mid y \in S_i\}$ entails $x \in v(S_1, \ldots, S_n) \Leftrightarrow y \in v(S_1, \ldots, S_n)$. There are a number of natural choices for v. Apart from the classical approval voting rule mentioned before, we might want to accept all edges receiving above-average support. While classical approval voting will typically result in very "sparse" output graphs, intuitively the latter rule will return graphs that have similar attributes as the input graphs. A third option is to use "even-and-equal" cumulative voting with $v : (S_1, \ldots, S_n) \mapsto \text{argmax}_{x \in S_1 \cup \cdots \cup S_n} \sum_{i | x \in S_i} \frac{1}{|S_i|}$, i.e., to let each individual distribute her weight evenly over the successors she approves of. This would be attractive, for instance, under an epistemic interpretation, where agents specifying fewer edges might be considered more certain about those edges. Finally, observe that the uniform quota rules (but not the general quota rules) are a special case of the successor-approval rules. We obtain F_q with the constant function $q : e \mapsto k$, mapping any given edge to the fixed quota k, by using $v : (S_1, \ldots, S_n) \mapsto \{x \in V : |\{i \in \mathcal{N} : x \in S_i\}| \geq k\}$.

While we are not going to do so in this paper, it is also possible to adapt the *distance-based rules*—familiar from preference aggregation, belief merging, and judgment aggregation [39–41]—to the case of graph aggregation. Such rules select a collective graph that satisfies certain properties and that minimises the distance to the individual graphs (for a suitable notion of distance and a suitable form of aggregating such distances). A downside of this approach is that distance-based rules are typically computationally intractable [42–45], while quota and successor-approval rules have very low complexity.

We can also adapt the *representative-voter rules* [46] to the case of graph aggregation. Here, the idea is to return one of the input graphs as the output, and for every profile to pick the input graph that in some sense is "most representative" of the views of the group.

Definition 7. A **representative-voter rule** is an aggregation rule F that is such that for every profile $\boldsymbol{E}$ there exists an individual $i^\star \in \mathcal{N}$ such that $F(\boldsymbol{E}) = E_{i^\star}$.

For instance, we might pick the input graph that is closest to the outcome of the majority rule. This majority-based representative-voter rule also has very low complexity. While we are not going to study any specific representative-voter rule in this paper, in Section 4.4 we are going to briefly discuss this class of rules as a whole.

We conclude our presentation of concrete (families of) aggregation rules with a number of rules that, intuitively speaking, are not very attractive.

Definition 8. The **dictatorship** of individual $i^\star \in \mathcal{N}$ is the aggregation rule $F_{i^\star}$ with $F_{i^\star} : \boldsymbol{E} \mapsto E_{i^\star}$.

Thus, for any given profile of input graphs, $F_{i^\star}$ always simply returns the graph submitted by the dictator $i^\star$. Note that every dictatorship is a representative-voter rule, but the converse is not true.

Definition 9. The **oligarchy** of coalition $C^\star \subseteq \mathcal{N}$, with $C^\star$ being nonempty, is the aggregation rule $F_{C^\star}$ with $F_{C^\star} : \boldsymbol{E} \mapsto \bigcap_{i \in C^\star} E_i$.

Thus, $F_{C^\star}$ always returns the intersection of the graphs submitted by the oligarchs in the coalition $C^\star$. So an individual in $C^\star$ can veto the acceptance of any given edge, but she cannot enforce its acceptance. In case $C^\star$ is a singleton, we obtain a dictatorship. In case $C^\star = \mathcal{N}$, we obtain the intersection rule.

2.4. Axiomatic properties and basic characterisation results

When choosing an aggregation rule, we need to consider its properties. In social choice theory, such properties are called *axioms* [9]. We now introduce several basic axioms for graph aggregation. The first such axiom is an independence condition that requires that the decision of whether or not a given edge e should be part of the collective graph should only depend on which of the individual graphs include e. This corresponds to well-known axioms in preference and judgment aggregation [5,17].

Definition 10. An aggregation rule F is called independent of irrelevant edges (**IIE**) if $N_e^{\boldsymbol{E}} = N_e^{\boldsymbol{E}'}$ implies $e \in F(\boldsymbol{E}) \Leftrightarrow e \in F(\boldsymbol{E}')$.

That is, if exactly the same individuals accept e under profiles $\boldsymbol{E}$ and $\boldsymbol{E}'$, then e should be part of either both or none of the corresponding collective graphs. The definition above applies to *all* edges $e \in V \times V$ and *all* pairs of profiles $\boldsymbol{E}, \boldsymbol{E}' \in (2^{V \times V})^n$. For the sake of readability, we shall leave this kind of universal quantification implicit also in later definitions.

IEE is a desirable property, because—if it can be satisfied—it greatly simplifies aggregation, in both computational and conceptual terms. As we shall see, some of the arguably most natural aggregation rules, the quota rules defined earlier, satisfy IIE. At the same time, as we shall also see, IEE is a very demanding property that is hard to satisfy if we are interested in richer forms of aggregation. Indeed, IIE will turn out to be at the very centre of our impossibility results.

While very much a standard axiom, we might be dissatisfied with IIE for not making reference to the fact that edges are defined in terms of vertices. Our next two axioms are much more graph-specific and do not have close analogues in preference or judgment aggregation. The first of them requires that the decision of whether or not to collectively accept a given edge $e = (x, y)$ should only depend on which edges with the same source x are accepted by the individuals. That is, acceptance of an edge may be influenced by what agents think about other edges, but not those edges that are sufficiently unrelated to the edge under consideration. Below we write $F(\boldsymbol{E})(x)$ for the set of successors of vertex x in the set of edges in the collective graph $F(\boldsymbol{E})$, and similarly $F(\boldsymbol{E})^{-1}(y)$ for the predecessors of y in $F(\boldsymbol{E})$.

Definition 11. An aggregation F is called independent of irrelevant sources (**IIS**) if $E_i(x) = E'_i(x)$ for all individuals $i \in \mathcal{N}$ implies $F(\boldsymbol{E})(x) = F(\boldsymbol{E}')(x)$.

Definition 12. An aggregation rule F is called independent of irrelevant targets (**IIT**) if $E_i^{-1}(y) = E_i'^{-1}(y)$ for all individuals $i \in \mathcal{N}$ implies $F(\boldsymbol{E})^{-1}(y) = F(\boldsymbol{E}')^{-1}(y)$.

Both IIS and IIT are strictly weaker than IIE. That is, we obtain the following result, which is easy to verify (simple counterexamples can be devised to show that the converse does not hold):

Proposition 1. *If an aggregation rule is IIE, then it is also both IIS and IIT.*

The fundamental economic principle of *unanimity* requires that an edge should be accepted by a group in case all individuals in that group accept it.

Definition 13. An aggregation rule F is called **unanimous** if it is always the case that $F(\boldsymbol{E}) \supseteq E_1 \cap \cdots \cap E_n$.

A requirement that, in some sense, is dual to unanimity is to ask that the collective graph should only include edges that are part of at least one of the individual graphs. In the context of ontology aggregation this axiom has been introduced under the name *groundedness* [47].

Definition 14. An aggregation F is called **grounded** if it is always the case that $F(\boldsymbol{E}) \subseteq E_1 \cup \cdots \cup E_n$.

The next axiom expresses a basic symmetry requirement, namely that the aggregation rule should treat all individuals the same. It is used in the exact same form in both preference and judgment aggregation [5,17].

Definition 15. An aggregation rule F is called **anonymous** if $F(E_1, \dots, E_n) = F(E_{\pi(1)}, \dots, E_{\pi(n)})$ for any permutation $\pi : \mathcal{N} \to \mathcal{N}$.

The axiom of *neutrality*, loosely speaking, postulates symmetry with respect to different parts of the graphs to be aggregated. We are going to mostly work with the following formalisation of this intuitive idea, which is inspired by the way in which neutrality is often defined in judgment aggregation [48].

Definition 16. An aggregation rule F is called **neutral** if $N_e^{\boldsymbol{E}} = N_{e'}^{\boldsymbol{E}}$ implies $e \in F(\boldsymbol{E}) \Leftrightarrow e' \in F(\boldsymbol{E})$.

Thus, this axiom says that, if two edges are accepted by the same coalition of individuals, then either both or neither should be included in the collective graph. (Observe that, while IIE speaks about one edge and two profiles, neutrality speaks about two edges within the same profile.) When we restrict attention to graphs that can be interpreted as preference orders, e.g., weak orders, this notion of neutrality, however, is different from how neutrality is usually defined in the preference aggregation literature [3], where it is taken to represent symmetry with respect to alternatives (i.e., vertices) rather than pairwise preferences (i.e., edges). The following alternative definition generalises this idea to arbitrary graphs. It is formulated in terms of a permutation $\pi : V \to V$ on vertices. Any such π naturally extends to edges $e = (x, y)$, graphs E, and profiles $\boldsymbol{E}$: $\pi((x, y)) = (\pi(x), \pi(y))$, $\pi(E) = \{\pi(e) \mid e \in E\}$, and $\pi(\boldsymbol{E}) = (\pi(E_1), \dots, \pi(E_n))$.

Definition 17. An aggregation rule F is called **permutation-neutral** if $F(\pi(\boldsymbol{E})) = \pi(F(\boldsymbol{E}))$ for any permutation $\pi : V \to V$.

The following two examples show that there are neutral aggregation rules that are not permutation-neutral and that there are permutation-neutral aggregation rules that are not neutral. However, as we shall see next, in the presence of IIE, the two definitions have the same logical strength.

Example 9 *(Neutral yet not permutation-neutral rule).* Let $V = \{x, y\}$ and consider the aggregation rule F that returns the empty graph $\emptyset$ in case agent 1 accepts edge (x, y) and that returns the complete graph $\{(x, x), (x, y), (y, x), (y, y)\}$ in all other cases. This rule is easily seen to be neutral, as the output graph always agrees on all edges. However, F is not permutation-neutral: if we swap x and y in a profile where agent 1 accepts only (x, y), then the output will change from the empty to the complete graph.

Example 10 *(Permutation-neutral yet not neutral rule).* Let $V = \{x, y, z\}$ and consider the aggregation rule F that first computes the intersection of all individual graphs and then, in certain special cases, removes one further edge: namely, if the intersection graph happens to be exactly $\pi(\{(x, y), (y, z)\})$, for some permutation $\pi : V \to V$, then the edge $\pi((y, z))$ is removed. In other words: if the intersection graph is a "line" of length 2, then the second half of that line is removed. This rule is permutation-neutral by definition. However, it is not neutral. For instance, if all agents accept both (x, y) and (y, z), and no other edges, then these two edges nevertheless are not treated symmetrically in the output.

Proposition 2. *Let F be an aggregation rule that is IIE. Then F is neutral if and only if it is permutation-neutral.*

Proof. It suffices to observe that both (i) aggregation rules that are IIE and neutral and (ii) aggregation rules that are IIE and permutation-neutral have the following property in common. Any such rule can be completely described by specifying which coalitions C of agents are such that it is the case that a given edge will get accepted by the rule if and only if exactly the agents in C accept it.[3] □

The following monotonicity axiom expresses that additional support for a collectively accepted edge should never cause that edge to be rejected. It applies in case profiles $\boldsymbol{E}$ and $\boldsymbol{E}'$ are identical, except that some individuals who do not accept

[3] We are going to explore this technique of describing aggregation rules in terms of so-called "winning coalitions" in depth in Section 3.

edge e in the former profile do accept it in the latter. Its definition is closely modelled on its counterpart in judgment aggregation [17].

Definition 18. An aggregation rule F is called **monotonic** if either $E'_i = E_i$ or $E'_i = E_i \cup \{e\}$ holding for all individuals $i \in \mathcal{N}$ implies $e \in F(\boldsymbol{E}) \Rightarrow e \in F(\boldsymbol{E'})$.

The link between aggregation rules and axiomatic properties is expressed in so-called characterisation results. For each rule (or class of rules), the aim is to find a set of axioms that uniquely define this rule (or class of rules, respectively). A simple adaptation of a result by Dietrich and List [13] yields the following characterisation of the class of quota rules:

Proposition 3. *An aggregation rule is a quota rule if and only if it is anonymous, monotonic, and IIE.*

Proof. To prove the left-to-right direction we simply have to verify that the quota rules all have these three properties. For the right-to-left direction, observe that, to accept a given edge (x, y) in the collective graph, an IIE aggregation rule will only look at the set of individuals i such that xE_iy. If the rule is also anonymous, then the acceptance decision is based only on the number of individuals accepting the edge. Finally, by monotonicity, there will be some minimal number of individual acceptances required to trigger collective acceptance. That number is the quota associated with the edge under consideration. □

If we add the axiom of neutrality, then we obtain the class of uniform quota rules. If we furthermore impose unanimity and groundedness, then this excludes the trivial quota rules. Similarly, it is easy to verify that IIS essentially characterises the class of successor-approval rules:

Proposition 4. *An aggregation rule is a successor-approval rule (with an anonymous and neutral choice function) if and only if it is anonymous, neutral, and IIS.*

An extreme form of violating anonymity is to use a *dictatorial* or an *oligarchic* aggregation rule, i.e., a rule that is either a dictatorship or an oligarchy (unless the oligarchy in question is the full set $\mathcal{N}$).

Sometimes we are only going to be interested in the properties of an aggregation rule as far as the *nonreflexive* edges $e = (x, y)$ with $x \neq y$ are concerned. Specifically, we call F *neutral on nonreflexive edges* (or just *NR-neutral*) if $N^{\boldsymbol{E}}_{(x,y)} = N^{\boldsymbol{E}}_{(x',y')}$ implies $(x, y) \in F(\boldsymbol{E}) \Leftrightarrow (x', y') \in F(\boldsymbol{E})$ for all $x \neq y$ and $x' \neq y'$. Analogously, we call F *dictatorial on nonreflexive edges* (or *NR-dictatorial*) if there exists an individual $i^\star \in \mathcal{N}$ such that $(x, y) \in F(\boldsymbol{E}) \Leftrightarrow (x, y) \in E_{i^\star}$ for all $x \neq y$. Finally, we call F *oligarchic on nonreflexive edges* (or *NR-oligarchic*) if there exists a nonempty coalition $C^\star \subseteq \mathcal{N}$ such that $(x, y) \in F(\boldsymbol{E}) \Leftrightarrow (x, y) \in \bigcap_{i \in C^\star} E_i$ for all $x \neq y$.

2.5. Collective rationality and basic possibility results

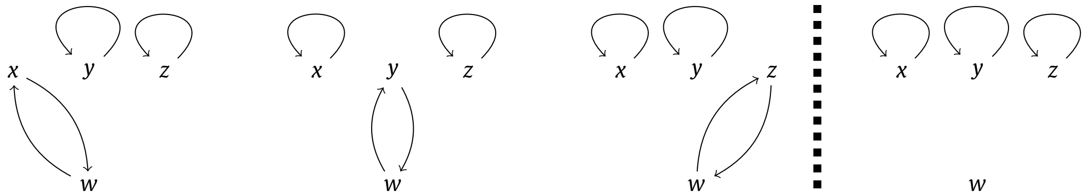
To what extent can a given aggregation rule ensure that a given property that is satisfied by each of the individual input graphs will be preserved during aggregation? This question relates to a well-studied concept in social choice theory, often referred to as *collective rationality* [5,11]. In the literature, collective rationality is usually defined with respect to a specific property that should be preserved (e.g., the transitivity of preferences or the logical consistency of judgments). Here, instead, we formulate a definition that is parametric with respect to a given graph property.[4]

Definition 19. An aggregation rule F is called **collectively rational** with respect to a graph property P if $F(\boldsymbol{E})$ satisfies P whenever all of the individual graphs in $\boldsymbol{E} = (E_1, \dots, E_n)$ do.

To illustrate the concept, let us consider two examples. Both concern the majority rule, but different graph properties. The first is a purely abstract example, while the second has a natural interpretation of graphs as preference relations.

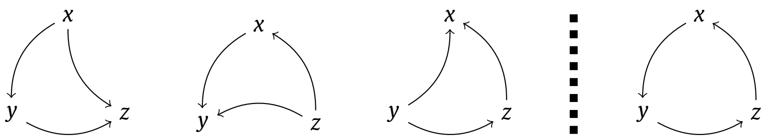
Example 11 *(Collective rationality).* Suppose three individuals provide us with three graphs over the same set $V = \{x, y, z, w\}$ of four vertices, as shown to the left of the dashed line below:

[4] In previous work on binary aggregation, a variant of judgment aggregation, we have used the term collective rationality in the same sense, with the property to be preserved under aggregation being encoded in the form of an integrity constraint [16].



If we apply the majority rule, then we obtain the graph to the right of the dashed line. Thus, the majority rule is not collectively rational with respect to seriality, as each individual graph is serial, but the collective graph is not. Symmetry, on the other hand, is preserved in this example.

Example 12 *(Condorcet paradox).* Now suppose three individuals provide us with the three graphs on the set of vertices $V = \{x, y, z\}$ shown on the lefthand side of the dashed line below:



The graph on the righthand side is once again the result of applying the majority rule. Observe that each of the three input graphs is transitive and complete. So we may interpret these graphs as (strict) preference orders on the candidates x, y, and z. For example, the preferences of the first agent would be $x \succ y \succ z$. The output graph, on the other hand, is not transitive (although it is complete). It does not correspond to a "rational" preference, as under that preference we should prefer x to y and y to z, but also z to x. This is the famous Condorcet paradox described by the Marquis de Condorcet in 1785 [49].

So the majority rule is not collectively rational with respect to either seriality or transitivity. On the other hand, we saw that both symmetry and completeness were preserved under the majority rule—at least for the specific examples considered here. In fact, it is not difficult to verify that this was no coincidence, and that the majority rule is collectively rational with respect to a number of properties of interest.

Fact 5. The majority rule is collectively rational with respect to reflexivity, irreflexivity, symmetry, and antisymmetry. In case n, the number of individuals, is odd, the majority rule furthermore is collectively rational with respect to completeness and connectedness.

Proof sketch. We give the proofs for symmetry and completeness. The other proofs are very similar. First, if the input graphs are symmetric, then the set of supporters of edge (x, y) is always identical to the set of supporters of the edge (y, x). Thus, either both or neither have a strict majority. Second, if the input graphs are complete, then each of them must include at least one of (x, y) and (y, x). Thus, by the pigeon hole principle, when n is odd, at least one of these two edges must have a strict majority. □

Rather than establishing further such results for specific aggregation rules, our main interest in this paper is the connection between the axioms satisfied by an aggregation rule and the range of graph properties preserved by the same rule. For some graph properties, collective rationality is easy to achieve, as the following simple *possibility results* demonstrate.

Proposition 6. *Any unanimous aggregation rule is collectively rational with respect to reflexivity.*

Proof. If every individual graph includes all edges of the form (x, x), then unanimity ensures the same for the collective graph. □

Proposition 7. *Any grounded aggregation rule is collectively rational with respect to irreflexivity.*

Proof. If no individual graph includes (x, x), then groundedness ensures the same for the collective graph. □

Proposition 8. *Any neutral aggregation rule is collectively rational with respect to symmetry.*

Proof. If edges (x, y) and (y, x) have the same support, then neutrality ensures that either both or neither will get accepted for the collective graph. □

Unfortunately, as we are going to see next, things do not always work out that harmoniously, and certain axiomatic requirements are in conflict with certain collective rationality requirements.

3. Impossibility results

In social choice theory, an *impossibility theorem* states that it is not possible to devise an aggregation rule that satisfies certain axioms and that is also collectively rational with respect to a certain combination of properties of the structures being aggregated (which in our case are graphs). In this section, we are going to prove two powerful impossibility theorems for graph aggregation, the *Oligarchy Theorem* and the *Dictatorship Theorem*. The latter identifies a set of requirements that are impossible to satisfy in the sense that the only aggregation rules that meet them are the dictatorships. The former drives on somewhat weaker requirements (specifically, regarding collective rationality) and permits a somewhat larger—but still decidedly unattractive—set of aggregation rules, namely the oligarchies.

Our results are inspired by—and significantly generalise—the seminal impossibility result for preference aggregation due to Arrow, first published in 1951 [5]. We recall Arrow's Theorem in Section 3.1. The following subsections are devoted to developing the framework in which to present and then prove our results. Section 3.2 introduces *winning coalitions*, i.e., sets of individuals who can force the acceptance or rejection of a given edge, discusses under what circumstances an aggregation rule can be described in terms of a family of winning coalitions, and what structural properties of such a family correspond to either dictatorial or oligarchic aggregation rules. Sections 3.3 and 3.4 introduce three so-called *meta-properties* for classifying graph properties and establish fundamental results for these meta-properties. Our impossibility theorems, which are formulated and proved in Section 3.5, apply to aggregation rules that are collectively rational with respect to graph properties that are covered by some of these meta-properties. Section 3.6, finally, discusses several variants of our theorems and provides a first illustration of their use.

3.1. Background: Arrow's Theorem for preference aggregation

The prime example of an impossibility result is Arrow's Theorem for preference aggregation, with preference relations being modelled as weak orders on some set of alternatives [5]. We can reformulate Arrow's Theorem in our framework for graph aggregation as follows:

> *For $|V| \geq 3$, every unanimous, grounded, and IIE aggregation rule that is collectively rational with respect to reflexivity, transitivity, and completeness must be a dictatorship.*

Thus, Arrow's Theorem applies to the following scenario. We wish to aggregate the preferences of several agents regarding a set of three or more alternatives. The agents are assumed to express their preferences by ranking the alternatives from best to worst (with indifferences being allowed), i.e., by each providing us with a weak order (a graph that is reflexive, transitive, and complete), and we want our aggregation rule to compute a single such weak order representing a suitable compromise. Furthermore, we want our aggregation rule to respect the basic axioms of unanimity (if all agents agree that x is at least as good as y, then the collective preference order should say so), groundedness (if no agent says that x is at least as good as y, then the collective preference order should not say so either), and IIE (it should be possible to compute the outcome on an edge-by-edge basis). Arrow's Theorem tells us that this is impossible—unless we are willing to use a dictatorship as our aggregation rule.

This result not only is surprising but also deeply troubling. It therefore is important to understand to what extent similar phenomena arise in other areas of graph aggregation. We are going to revisit Arrow's Theorem in Section 3.6, where we are also going to be in a position to explain why the standard formulation of the theorem, given in that section as Theorem 19, is indeed implied by the variant given here.

In the sequel, we are sometimes going to refer to aggregation rules that are unanimous, grounded, and IIE as *Arrovian* aggregation rules.

3.2. Winning coalitions, filters, and ultrafilters

As is well understood in social choice theory, impossibility theorems in preference aggregation heavily feed on independence axioms (in our case IIE). Observe that an aggregation rule F satisfies IIE if and only if for each edge $e \in V \times V$ there exists a set of *winning coalitions* $\mathcal{W}_e \subseteq 2^{\mathcal{N}}$ such that $e \in F(\boldsymbol{E}) \Leftrightarrow N_e^{\boldsymbol{E}} \in \mathcal{W}_e$. That is, F accepts e if and only if exactly the individuals in one of the winning coalitions for e do. Imposing additional axioms on F corresponds to restrictions on the associated family of winning coalitions $\{\mathcal{W}_e\}_{e \in V \times V}$:

- If F is *unanimous*, then $\mathcal{N} \in \mathcal{W}_e$ for any edge e (i.e., the grand coalition is always a winning coalition).
- If F is *grounded*, then $\emptyset \notin \mathcal{W}_e$ for any edge e (i.e., the empty set is not a winning coalition).
- If F is *monotonic*, then $C_1 \in \mathcal{W}_e$ implies $C_2 \in \mathcal{W}_e$ for any edge e and any set $C_2 \supset C_1$ (i.e., winning coalitions are closed under supersets).

- If F is *(NR-)neutral*, then $\mathcal{W}_e = \mathcal{W}_{e'}$ for any two (nonreflexive) edges e and e' (i.e., every edge must have exactly the same set of winning coalitions).

Thus, an aggregation rule that is both IIE and neutral can be fully described in terms of a single set $\mathcal{W}$ of winning coalitions. Any such $\mathcal{W}$ is a subset of the powerset of $\mathcal{N}$, the set of individuals. The proofs of our impossibility results are going to exploit the special structure of such subsets of the powerset of $\mathcal{N}$, enforced by both axioms and collective rationality requirements. Specifically, in our proofs we are going to encounter the concepts of *filters* and *ultrafilters* familiar from model theory [50].

Definition 20. A **filter** $\mathcal{W}$ on a set $\mathcal{N}$ is a collection of subsets of $\mathcal{N}$ satisfying the following three conditions:

(i) $\emptyset \notin \mathcal{W}$;
(ii) $C_1, C_2 \in \mathcal{W}$ implies $C_1 \cap C_2 \in \mathcal{W}$ for any two sets $C_1, C_2 \subseteq \mathcal{N}$ (closure under intersection);
(iii) $C_1 \in \mathcal{W}$ implies $C_2 \in \mathcal{W}$ for any set $C_2 \subseteq \mathcal{N}$ with $C_2 \supset C_1$ (closure under supersets).

Definition 21. An **ultrafilter** $\mathcal{W}$ on a set $\mathcal{N}$ is a collection of subsets of $\mathcal{N}$ satisfying the following three conditions:

(i) $\emptyset \notin \mathcal{W}$
(ii) $C_1, C_2 \in \mathcal{W}$ implies $C_1 \cap C_2 \in \mathcal{W}$ for any two sets $C_1, C_2 \subseteq \mathcal{N}$ (closure under intersection);
(iii) C or $\mathcal{N} \setminus C$ is in $\mathcal{W}$ for any set $C \subseteq \mathcal{N}$ (maximality).

Every ultrafilter is a filter; in particular, the ultrafilter conditions imply closure under supersets. Note that the condition $\emptyset \notin \mathcal{W}$ directly corresponds to groundedness, while closure under supersets corresponds to monotonicity.

The use of ultrafilters in social choice theory goes back to the work of Fishburn [6] and Kirman and Sondermann [7], who employed ultrafilters to prove Arrow's Theorem and its generalisation to an infinite number of individuals. The ultrafilter method also has found applications in judgment aggregation [15], and also filters have been used in both preference aggregation [8] and judgment aggregation [12]. The relevance of filters and ultrafilters to aggregation problems is due to the following simple results, which interpret well-known facts from model theory in our specific context.

Lemma 9 *(Filter Lemma). Let F be an IIE and NR-neutral aggregation rule and let $\mathcal{W}$ be the corresponding set of winning coalitions for nonreflexive edges, i.e., $(x, y) \in F(\boldsymbol{E}) \Leftrightarrow N^{\boldsymbol{E}}_{(x,y)} \in \mathcal{W}$ for all $x \neq y \in V$. Then F is NR-oligarchic if and only if $\mathcal{W}$ is a filter.*

Proof. ($\Rightarrow$) Recall that F being NR-oligarchic means that there exists a nonempty coalition $C^\star$ such that a given nonreflexive edge is accepted if and only if all the agents in $C^\star$ accept it. Thus, the winning coalitions are exactly $C^\star$ and its supersets. This family of sets does not include the empty set and is closed under both intersection and supersets.

($\Leftarrow$) Suppose F is determined by the filter $\mathcal{W}$ as far as nonreflexive edges are concerned. Let $C^\star := \bigcap_{C \in \mathcal{C}} C$, which is well-defined due to $\mathcal{N}$ being finite. Observe that $C^\star$ must be nonempty, due to the first two filter conditions. Now note that F is NR-oligarchic with respect to coalition $C^\star$. $\square$

Lemma 10 *(Ultrafilter Lemma). Let F be an IIE and NR-neutral aggregation rule and let $\mathcal{W}$ be the corresponding set of winning coalitions for nonreflexive edges, i.e., $(x, y) \in F(\boldsymbol{E}) \Leftrightarrow N^{\boldsymbol{E}}_{(x,y)} \in \mathcal{W}$ for all $x \neq y \in V$. Then F is NR-dictatorial if and only if $\mathcal{W}$ is an ultrafilter.*

Proof. ($\Rightarrow$) F being NR-dictatorial means that there exists an $i^\star \in \mathcal{N}$ such that the winning coalitions for nonreflexive edges are exactly $\{i^\star\}$ and its supersets. This family of sets does not include the empty set, is closed under intersection, and maximal.

($\Leftarrow$) Suppose F is determined by the ultrafilter $\mathcal{W}$ as far as nonreflexive edges are concerned. Take an arbitrary $C \in \mathcal{W}$ with $|C| \geq 2$ and consider any nonempty $C' \subsetneq C$. By maximality, one of C' and $\mathcal{N} \setminus C'$ must be in $\mathcal{W}$. Thus, by closure under intersection, one of $C \cap C' = C'$ and $C \cap (\mathcal{N} \setminus C') = C \setminus C'$ must be in $\mathcal{W}$ as well. Observe that both of these sets are nonempty and of lower cardinality than C. To summarise, we have just shown for any $C \in \mathcal{W}$ with $|C| \geq 2$ at least one nonempty proper subset of C is also in $\mathcal{W}$. By maximality, $\mathcal{W}$ is not empty. So take any $C \in \mathcal{W}$. Due to $\mathcal{N}$ being finite, we can apply our reduction rule a finite number of times to infer that $\mathcal{W}$ must include some singleton $\{i^\star\} \subsetneq \cdots \subsetneq C$. Hence, F is an NR-dictatorship with dictator $i^\star$. $\square$

3.3. The neutrality axiom and contagious graph properties

Recall that the neutrality axiom is required to be able to work with a single family of winning coalitions as outlined earlier, yet this axiom does not feature in Arrow's Theorem. As we shall see soon, the reason we do not need to assume neutrality is that, in Arrow's setting, the same restriction on winning coalitions is already enforced by collective rationality with respect to transitivity. This is an interesting link between a specific collective rationality requirement and a specific

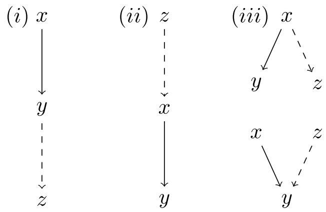


Fig. 1. Illustration of Definition 23, indicating given edges (solid) and implied edges (dashed).

axiom. In the literature, this fact is often called the *Contagion Lemma* [9], although the connection to neutrality is not usually made explicit. The same kind of result can also be obtained for other graph properties with a similar structure. Let us now develop a definition for a class of graph properties that are going to allow us to derive neutrality.

Recall that $P[S^+, S^-]$ denotes the set of graphs with property P that include all of the edges in S^+ and none of those in S^-. We start with a technical definition.

Definition 22. Let $x, y, z, w \in V$. A graph property $P \subseteq 2^{V \times V}$ is called ***xy*/*zw*-contagious** if there exist two disjoint sets $S^+, S^- \subseteq V \times V$ such that the following conditions hold:

(i) for every graph $E \in P[S^+, S^-]$ it is the case that $(x, y) \in E$ implies $(z, w) \in E$; and
(ii) there exist graphs $E_0, E_1 \in P[S^+, S^-]$ with $(z, w) \notin E_0$ and $(x, y) \in E_1$.

Part (i) of Definition 22 says that, if you accept edge (x, y), then you must also accept edge (z, w)—at least if the side condition of you also accepting all the edges in S^+ but none of those in S^- is met. That is, the property of xy/zw-contagiousness may be paraphrased as the formula $[\bigwedge S^+ \wedge \neg \bigvee S^-] \to [xEy \to zEw]$. Part (ii) is a richness condition that says that you have the option of accepting neither or both of (x, y) and (z, w). It requires the existence of a graph E_0 where neither (x, y) nor (z, w) are accepted, and the existence of a graph E_1 where both (x, y) and (z, w) are accepted.

Contagiousness with respect to two given edges will be useful for our purposes if those two edges stand in a specific relationship to each other. The following definition captures the relevant cases.

Definition 23. A graph property $P \subseteq 2^{V \times V}$ is called **contagious** if it satisfies at least one of the three conditions below:

(i) P is xy/yz-contagious for all triples of vertices $x, y, z \in V$.
(ii) P is xy/zx-contagious for all triples of vertices $x, y, z \in V$.
(iii) P is xy/xz-contagious and xy/zy-contagious for all triples of vertices $x, y, z \in V$.

That is, Definition 23 covers pairs of edges where (i) the second edge is a successor of the first edge, where (ii) the second edge is a predecessor of the first edge, and where (iii) the two edges share either a starting point or an end point. This covers all cases of two edges meeting in one point. The three cases are illustrated in Fig. 1. As will become clear in the proof of Lemma 12, case (iii) differs from the other two, as only one of these two types of connections would not be sufficient to "traverse" the full graph.

Fact 11. For $|V| \geq 3$, the two Euclidean properties, transitivity, negative transitivity, and connectedness are all contagious graph properties.

Proof. Let us first consider the property of being a right-Euclidean graph. It satisfies condition (i) of Definition 23. To prove this, we are going to show that the right-Euclidean property is xy/yz-contagious for all triples $x, y, z \in V$. Let $S^+ = \{(x, z)\}$ and $S^- = \emptyset$, i.e., $P[S^+, S^-]$ is the set of all right-Euclidean graphs containing (x, z). Condition (i) of Definition 22 is met: any graph in $P[S^+, S^-]$ contains (x, z); therefore, by the right-Euclidean property, (y, z) needs to be accepted whenever (x, y) is. Condition (ii) is also satisfied. Let E_0 be the graph only containing the single edge (x, z), and let E_1 be the graph containing exactly the three edges (x, y), (y, z), and (x, z). Both graphs are right-Euclidean and, since they include (x, z), they also belong to $P[S^+, S^-]$.

An alternative way of seeing that the right-Euclidean property is contagious is to observe that it is equivalent to the formula $[xEz] \to [xEy \to yEz]$, with all variables universally quantified. Similarly, the left-Euclidean property, which can be rewritten as $[zEy] \to [xEy \to zEx]$, is contagious by condition (ii). Transitivity satisfies condition (iii), as we can rewrite it either as $[yEz] \to [xEy \to xEz]$ or as $[zEx] \to [xEy \to zEy]$. Negative transitivity can be rewritten either as $[\neg(zEy)] \to [xEy \to xEz]$ or as $[\neg(xEz)] \to [xEy \to zEy]$, and this property thus also satisfies condition (iii). Connectedness, finally, can be rewritten as $[xEz \wedge \neg zEy] \to [xEy \to yEz]$ and thus satisfies condition (i). For all these properties, the richness conditions are easily verified to hold as well. $\square$

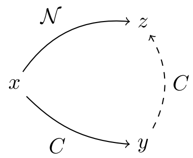


Fig. 2. Collective rationality with respect to the right-Euclidean property implies neutrality.

We are now ready to prove a powerful lemma, the *Neutrality Lemma*, showing that any Arrovian aggregation rule that is collectively rational with respect to a contagious graph property must be neutral (at least as far as nonreflexive edges are concerned). This generalises a result often referred to as the *Contagion Lemma* in the literature on preference aggregation [9], and our proof generalises the standard proof of that lemma.

Lemma 12 *(Neutrality Lemma). For $|V| \geq 3$, any unanimous, grounded, and IIE aggregation rule that is collectively rational with respect to a contagious graph property must be NR-neutral.*

Proof. We are first going to establish a generic result for collective rationality with respect to xy/zw-contagiousness. Let $x, y, z, w \in V$. Take any graph property P that is xy/zw-contagious and take any aggregation rule F that is unanimous, grounded, IIE, and collectively rational with respect to P. Let $\{\mathcal{W}_e\}_{e \in V \times V}$ be the family of winning coalitions associated with F. We want to show that $\mathcal{W}_{(x,y)} \subseteq \mathcal{W}_{(z,w)}$. So let C be a coalition in $\mathcal{W}_{(x,y)}$. Let $S^+, S^- \subseteq V \times V$ and $E_0, E_1 \in P[S^+, S^-]$ be defined as in Definition 22. Consider a profile $\boldsymbol{E}$ in which the individuals in C propose graph E_1 and all others propose E_0. That is, all individuals accept the edges in S^+, none accept any of those in S^-, exactly the individuals in C accept edge (x, y), and exactly those in C also accept (z, w). Now consider the collective graph $F(\boldsymbol{E})$. By unanimity $S^+ \subseteq F(\boldsymbol{E})$, by groundedness $S^- \cap F(\boldsymbol{E}) = \emptyset$, and finally $(x, y) \in F(\boldsymbol{E})$ due to C being a winning coalition for (x, y). By collective rationality, $F(\boldsymbol{E}) \in P$ and thus also $F(\boldsymbol{E}) \in P[S^+, S^-]$. But then, due to xy/zw-contagiousness of $F(\boldsymbol{E})$, we get $(z, w) \in F(\boldsymbol{E})$. As it was exactly the individuals in C who accepted (z, w), coalition C must be winning for (z, w), i.e., $C \in \mathcal{W}_{(z,w)}$, and we are done.

We are now ready to prove the lemma. Take any graph property P that is contagious and take any aggregation rule F that is unanimous, grounded, IIE, and collectively rational with respect to P. Let $\{\mathcal{W}_e\}_{e \in V \times V}$ be the family of winning coalitions associated with F. We need to show that there exists a unique $\mathcal{W} \subseteq 2^{\mathcal{N}}$ such that $\mathcal{W} = \mathcal{W}_e$ for every nonreflexive edge e. By unanimity, the sets $\mathcal{W}_e$ are not empty (because at least $\mathcal{N} \in \mathcal{W}_e$). Consider any three vertices $x, y, z \in V$ and any coalition $C \in \mathcal{W}_{(x,y)}$. We are going to show that C is also winning for both (y, z) and (y, x). If we can show this for any x, y, z, then we are done, as we can then repeat the same method several times until all nonreflexive edges are covered.

For each of the three possible ways in which P can be contagious (see Definition 23), we are going to use different instances of our generic result for xy/zw-contagiousness above:

- First, if P is contagious by virtue of condition (i), then we can use xy/yz-contagiousness to get $C \in \mathcal{W}_{(y,z)}$ and its instance xy/yx-contagiousness (with $z := x$) to obtain also $C \in \mathcal{W}_{(y,x)}$.
- Second, if P is contagious due to condition (ii), we use xy/yx-contagiousness to get $C \in \mathcal{W}_{(y,x)}$, and then yx/zy-contagiousness to get $C \in \mathcal{W}_{(z,y)}$ and zy/yz-contagiousness to get $C \in \mathcal{W}_{(y,z)}$.
- Third, suppose P is contagious by virtue of condition (iii). We first use xy/zy-contagiousness to obtain $C \in \mathcal{W}_{(z,y)}$ and then zy/zx-contagiousness to get $C \in \mathcal{W}_{(z,x)}$. From the latter, via zx/yx-contagiousness we get $C \in \mathcal{W}_{(y,x)}$. Finally, yx/yz-contagiousness then entails $C \in \mathcal{W}_{(y,z)}$.

Hence, we obtain the required transfer from one edge (x, y) to both its successor (y, z) and its inverse (y, x) in all three cases, and our proof is complete. □

Fig. 2 provides an illustration of a specific instance of the main argument in the proof of Lemma 12 when the right-Euclidean property is considered, which is xy/yz-contagious by Fact 11. We have $S^+ = \{(x, z)\}$ and $S^- = \emptyset$. E_1 is the graph that accepts all three edges (x, y), (y, z) and (x, z), and E_0 accepts only edge (x, z). Consider profile $\boldsymbol{E}$, in which the individuals in C choose E_1 and all others choose E_0. That is, the individuals in C accept (x, y) and (y, z), while (x, z) is accepted by all individuals in $\mathcal{N}$. By unanimity, (x, z) must be accepted, and due to $C \in \mathcal{W}_{(x,y)}$ also (x, y) should be accepted. We can now conclude, since F is collectively rational with respect to the right-Euclidean property, that (y, z) should also be accepted, and hence that $C \in \mathcal{W}_{(y,z)}$. It is then sufficient to consider all ordered triples to obtain neutrality over all (nonreflexive) edges.

3.4. Implicative and disjunctive graph properties

Let us briefly recapitulate where we are at this point. We now know that any Arrovian aggregation rule F that is collectively rational with respect to some contagious graph property P can be fully described in terms of a single family $\mathcal{W}$ of winning coalitions, at least as far as F's behaviour on nonreflexive edges is concerned. To prove our impossibility results, we need to derive structural properties of $\mathcal{W}$ that allow us to infer that $\mathcal{W}$ is either a filter or an ultrafilter (so we can use Lemma 9 or 10, respectively). These structural properties are going to be shown to follow from collective rationality requirements with respect to graph properties belonging to a certain class of such properties.

We are now going to introduce two such classes of graph properties, or "meta-properties" as we shall also call them. Recall that we have already seen one meta-property, namely contagiousness (which, however, is much more complex than the following meta-properties). First, a graph property is *implicative* if the inclusion of some edges can force the inclusion of a further edge, as is the case, for instance, for transitivity. The following definition makes this precise.

Definition 24. A graph property $P \subseteq 2^{V \times V}$ is called **implicative** if there exist two disjoint sets $S^+, S^- \subseteq V \times V$ and three distinct edges $e_1, e_2, e_3 \in V \times V \setminus (S^+ \cup S^-)$ such that the following conditions hold:

(i) for every graph $E \in P[S^+, S^-]$ it is the case that $e_1, e_2 \in E$ implies $e_3 \in E$; and
(ii) there exist graphs $E_0, E_1, E_2, E_{13}, E_{123} \in P[S^+, S^-]$ with $E_0 \cap \{e_1, e_2, e_3\} = \emptyset$, $E_1 \cap \{e_1, e_2, e_3\} = \{e_1\}$, $E_2 \cap \{e_1, e_2, e_3\} = \{e_2\}$, $E_{13} \cap \{e_1, e_2, e_3\} = \{e_1, e_3\}$, and $\{e_1, e_2, e_3\} \subseteq E_{123}$.

Part (i) expresses that all graphs with property P (that also include all edges in S^+ and none from S^-) must satisfy the formula $e_1 \wedge e_2 \rightarrow e_3$. Part (ii) is a richness condition saying that accepting/rejecting any combination of e_1 and e_2 is possible, that e_3 need not be accepted unless both e_1 and e_2 are, and that e_3 can be accepted even if only the first antecedent e_1 is.[5] Observe that Definition 24 has an existential form, i.e., we simply need to find two subsets S^+ and S^- for the precondition, and three edges e_1, e_2 and e_3 that satisfy the two requirements (i) and (ii). In this sense, implicativeness is much less demanding than contagiousness, which imposes conditions across the entire graph. Implicativeness may be paraphrased as the formula $[\bigwedge S^+ \wedge \neg \bigvee S^-] \rightarrow [e_1 \wedge e_2 \rightarrow e_3]$.

Fact 13. For $|V| \geq 3$, the two Euclidean properties, transitivity, and connectedness are all implicative graph properties.

Proof (sketch). Let $V = \{v_1, v_2, v_3, \ldots\}$. To see that transitivity satisfies Definition 24, choose $S^+ = S^- = \emptyset$, $e_1 = (v_1, v_2)$, $e_2 = (v_2, v_3)$, and $e_3 = (v_1, v_3)$. Transitivity implies that, if both e_1 and e_2 are accepted, then also e_3 should be accepted. All remaining acceptance/rejection patterns of e_1, e_2, and e_3 are possible, in accordance with condition (ii). The proofs for the Euclidean properties are similar. Rewriting connectedness as $[\neg yEz] \rightarrow [(xEy \wedge xEz) \rightarrow zEy]$ shows that it is implicative as well. □

Note that implicativeness is a very weak requirement: even transitivity restricted to a single triple of edges is sufficient to satisfy it. Next, we define *disjunctive* graph properties as properties that force us to include at least one of two given edges, as is the case, for instance, for completeness.

Definition 25. A graph property $P \subseteq 2^{V \times V}$ is called **disjunctive** if there exist two disjoint sets $S^+, S^- \subseteq V \times V$ and two distinct edges $e_1, e_2 \in V \times V \setminus (S^+ \cup S^-)$ such that the following conditions hold:

(i) for every graph $E \in P[S^+, S^-]$ we have $e_1 \in E$ or $e_2 \in E$; and
(ii) there exist two graphs $E_1, E_2 \in P[S^+, S^-]$ with $E_1 \cap \{e_1, e_2\} = \{e_1\}$ and $E_2 \cap \{e_1, e_2\} = \{e_2\}$.

Part (i) ensures that all graphs with property P (that meet the precondition of including all edges in S^+ and none from S^-) satisfy the formula $e_1 \vee e_2$. Part (ii) is a richness condition ensuring that there are at least two graphs that each include only one of e_1 and e_2. Definition 25 also has an existential form, and it may be paraphrased as the formula $[\bigwedge S^+ \wedge \neg \bigvee S^-] \rightarrow [e_1 \vee e_2]$.

Fact 14. For $|V| \geq 3$, negative transitivity, connectedness, completeness, nontriviality, and seriality are all disjunctive graph properties.

Proof. Let $V = \{v_1, \ldots, v_m\}$. For negative transitivity, choose $S^+ = \{v_1, v_2\}$, $S^- = \emptyset$, $e_1 = (v_1, v_3)$, and $e_2 = (v_3, v_2)$ to see that the conditions are satisfied. For connectedness, choose $S^+ = \{(v_1, v_2), (v_1, v_3)\}$, $S^- = \emptyset$, $e_1 = (v_2, v_3)$, and $e_2 = (v_3, v_2)$. For completeness, choose $S^+ = S^- = \emptyset$, $e_1 = (v_1, v_2)$, and $e_2 = (v_2, v_1)$. For nontriviality, choose $S^+ = \emptyset$, $S^- = \{(v_i, v_j) : \{i, j\} \neq \{1, 2\}\}$, $e_1 = (v_1, v_2)$, and $e_2 = (v_2, v_1)$. Finally, for seriality, choose $S^+ = \emptyset$, $S^- = \{(v_1, v_1), (v_1, v_2), \ldots, (v_1, v_{m-2})\}$, $e_1 = (v_1, v_{m-1})$, and $e_2 = (v_1, v_m)$. □

Note that some of these results could be strengthened to the case of $|V| = 2$, but doing so would not be useful for our purposes here.

[5] In our earlier work, we did not require the existence of E_{13} [2]. The slightly stronger formulation used here is necessary to prove one of our general impossibility theorems (Theorem 15), but not the other (Theorem 16).

3.5. Two general impossibility theorems for graph aggregation

We are now ready to present our impossibility results. We are going to prove two main theorems. What they have in common is that they talk about Arrovian aggregation rules F that are collectively rational with respect to a graph property P that is contagious and implicative. For the first theorem, we are going to show that under these assumptions F must be oligarchic (at least as far as nonreflexive edges are concerned). For the second theorem, we also assume that P is disjunctive, and show that then F must be dictatorial (at least on nonreflexive edges).

Theorem 15 *(Oligarchy Theorem). Let P be a graph property that is contagious and implicative. Then, for $|V| \geq 3$, any unanimous, grounded, and IIE aggregation rule F that is collectively rational with respect to P must be oligarchic on nonreflexive edges.*

Proof. Take any graph property P that is contagious and implicative, and any aggregation rule F that is unanimous, grounded, IIE, and collectively rational with respect to P. By Lemma 12, F must be NR-neutral. Hence, there exists a set of winning coalitions $\mathcal{W} \subseteq 2^{\mathcal{N}}$ determining F in the sense that $e \in F(\boldsymbol{E}) \Leftrightarrow N_e^{\boldsymbol{E}} \in \mathcal{W}$ for any nonreflexive edge e.

We shall prove that $\mathcal{W}$ is a filter (see Definition 20), from which the theorem then follows by Lemma 9. Condition (i) holds, as F is grounded. So we still need to show that $\mathcal{W}$ satisfies condition (ii), i.e., that it is closed under intersection, and condition (iii), i.e., that it is closed under supersets. To do so, we are going to make use of the assumption that P is implicative. Let $S^+, S^- \subseteq V \times V$ and $e_1, e_2, e_3 \in V \times V$; and let $E_0, E_1, E_2, E_{13}, E_{123} \in P[S^+, S^-]$ be defined as in Definition 24.

First, take any two winning coalitions $C_1, C_2 \in \mathcal{W}$. Consider a profile of graphs $\boldsymbol{E}$ satisfying P in which exactly the individuals in $C_1 \cap C_2$ propose E_{123}, those in $C_1 \setminus C_2$ propose E_1, those in $C_2 \setminus C_1$ propose E_2, and all others propose E_0. Thus, exactly the individuals in C_1 accept e_1, exactly those in C_2 accept e_2, and exactly those in $C_1 \cap C_2$ accept e_3. Furthermore, all individuals accept S^+ and all of them reject S^-. Hence, due to unanimity, all edges in S^+ must be part of the collective graph $F(\boldsymbol{E})$, while due to groundedness, none of the edges in S^- can be part of $F(\boldsymbol{E})$. As F is collectively rational with respect to P, we get $F(\boldsymbol{E}) \in P[S^+, S^-]$. Now, since C_1 and C_2 are winning coalitions, e_1 and e_2 must be part of $F(\boldsymbol{E})$. As P is implicative, this means that $e_3 \in F(\boldsymbol{E})$. Hence, we must have $C_1 \cap C_2 \in \mathcal{W}$, i.e., $\mathcal{W}$ is closed under intersection.

Now, take any winning coalition $C_1 \in \mathcal{W}$ and any other coalition C_2 with $C_1 \subseteq C_2$. Consider a profile of graphs $\boldsymbol{E}$ satisfying P in which the individuals in C_1 propose E_{123}, those in $C_2 \setminus C_1$ propose E_{13}, and those in $\mathcal{N} \setminus C_2$ propose E_1. In other words, the coalition of supporters of e_1 is $\mathcal{N}$, the coalition of supporters of e_2 is C_1, the coalition of supporters of e_3 is C_2, all individuals accept S^+, and all of them also reject S^-. Due to unanimity and as $C_1 \in \mathcal{W}$, e_1 and e_2 will be part of the collective graph $F(\boldsymbol{E})$. As F is collectively rational with respect to P, we thus also get $e_3 \in F(\boldsymbol{E})$. Hence, as e_3 was supported by C_2, it must be the case that $C_2 \in \mathcal{W}$, i.e., $\mathcal{W}$ is closed under supersets. □

If the graph property to be preserved under aggregation also is required to be disjunctive, we can further tighten this impossibility result and obtain a dictatorship. The proof is very similar to that of Theorem 15, the only added difficulty being that of proving maximality of the filter from collective rationality with respect to a disjunctive graph property.

Theorem 16 *(Dictatorship Theorem). Let P be a graph property that is contagious, implicative, and disjunctive. Then, for $|V| \geq 3$, any unanimous, grounded, and IIE aggregation rule F that is collectively rational with respect to P must be dictatorial on nonreflexive edges.*

Proof. Take any graph property P that is contagious, implicative, and disjunctive, and any aggregation rule F that is unanimous, grounded, and IIE, and collectively rational with respect to P. By Lemma 12, F must be NR-neutral, i.e., on nonreflexive edges, F must be determined by a single family $\mathcal{W}$ of winning coalitions. We shall prove that the $\mathcal{W}$ is an ultrafilter (see Definition 21), from which the theorem then follows by Lemma 10. Condition (i) holds, as F is grounded. Condition (ii) follows from P being implicative and can be proved exactly as for Theorem 15.

To derive condition (iii), we are going to make use of the assumption that P is disjunctive. Let $S^+, S^- \subseteq V \times V$ and $e_1, e_2 \in V \times V$; and let $E_1, E_2 \in P[S^+, S^-]$ be defined as in Definition 25. Now take any coalition $C \subseteq \mathcal{N}$. Consider a profile $\boldsymbol{E}$ satisfying P in which exactly the individuals in C propose E_1 and exactly those in $\mathcal{N} \setminus C$ propose E_2. Recall that $S^+ \subseteq E_1$ and $S^+ \subseteq E_2$, i.e., all individuals accept S^+. Thus, due to unanimity, all of the edges in S^+ must be part of the collective graph $F(\boldsymbol{E})$. Analogously, due to groundedness, none of the edges in S^- can be part of $F(\boldsymbol{E})$. Thus, as F is collectively rational with respect to P, we get $F(\boldsymbol{E}) \in P[S^+, S^-]$. As P is disjunctive, this means that one of e_1 and e_2 has to be part of $F(\boldsymbol{E})$. Hence, $C \in \mathcal{W}$ or $(\mathcal{N} \setminus C) \in \mathcal{W}$. □

It may be helpful to illustrate the main arguments in the proofs of Theorem 15 and 16 by instantiating them for specific graph properties rather than generic meta-properties. For instance, we can derive closure of intersection of $\mathcal{W}$ by using collective rationality with respect to transitivity, which by Fact 13 is an implicative property. Consider the profile depicted on the left in Fig. 3, in which exactly the individuals in C_1 accept edge $e_1 = (x, y)$, exactly those in C_2 accept $e_2 = (y, z)$, and exactly those in $C_1 \cap C_2$ accept $e_3 = (x, z)$. As both C_1 and C_2 are winning coalitions, we obtain that both (x, y) and (y, z) need to be collectively accepted. We can now conclude, since F is collectively rational with respect to transitivity, that the

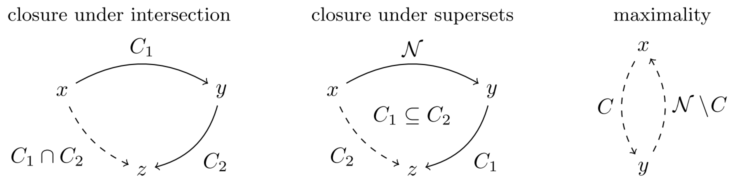


Fig. 3. Using collective rationality with respect to transitivity and completeness.

edge (x, z) should also be accepted. Hence, the coalition accepting (x, z), which is $C_1 \cap C_2$, must be a winning coalition as well. Similarly, we can obtain closure under supersets from collective rationality with respect to transitivity using the profile shown in the middle of Fig. 3. Here, all individuals accept $e_1 = (x, y)$, those in C_1 accept $e_2 = (y, z)$, and those in C_2, which is a superset of C_1, accept $e_3 = (x, z)$. As both $\mathcal{N}$ and C_1 are winning coalitions, both (x, y) and (y, z) get accepted. Thus, as F is collectively rational with respect to transitivity, so does (x, z). Hence, C_2, the coalition of supporters of (x, z), must also be winning. Finally, we can prove maximality of $\mathcal{W}$ by using collective rationality with respect to, say, completeness, which by Fact 14 is a disjunctive property. Consider the profile on the righthand side of Fig. 3, in which exactly the individuals in C accept $e_1 = (x, y)$ and exactly those in $\mathcal{N} \setminus C$ accept $e_2 = (y, x)$. As F is collectively rational with respect to completeness, one of the two edges has to get accepted in the outcome, i.e., one of the two coalitions accepting these two edges must be winning, meaning that either $C \in \mathcal{W}$ or $(\mathcal{N} \setminus C) \in \mathcal{W}$.

Observe that the converse of Theorem 16 holds as well: any dictatorship is unanimous, grounded, IIE, and collectively rational with respect to any graph property (and certainly with respect to those that are contagious, implicative, and disjunctive).[6] Thus, an alternative reading of Theorem 16 is as that of a family of characterisation theorems of the dictatorships (with one characterisation for every P that is contagious and implicative).

Our Theorem 16 is related to generalisations of Arrow's Theorem to judgment aggregation [51,14], particularly in the formulation due to Dokow and Holzman [14], who model sets of judgments (on m issues) as binary vectors in some subspace of $\{0, 1\}^m$. It is possible to embed graph aggregation into this form of judgment aggregation, by adapting the well-known approach for embedding preference aggregation into judgment aggregation [51,14,16]. This suggests that it should also be possible to derive Theorem 16 as a special case of the main result of Dokow and Holzman, which would involve showing that graph properties that are contagious, implicative, and disjunctive can be mapped into subspaces of $\{0, 1\}^m$ (with $m = |V \times V|$) that, in the terminology of Dokow and Holzman, are *totally blocked* and *not affine*.[7] While we conjecture this to be possible in principle, we also conjecture any such proof to be at least as technically involved as our proof given here and certainly much less valuable from a "didactic" point of view. Indeed, our proof arguably is easier and clearer than both the proofs for the corresponding result in the more specific domain of preference aggregation (i.e., Arrow's Theorem)[8] and the proofs for the corresponding results in the more general domain of judgment aggregation (i.e., the result of Dokow and Holzman [14] and its variant due to Dietrich and List [51]). The reason is that our meta-properties encode directly what we require in the proof steps where they are used.

3.6. Variants and instances of the general impossibility theorems

In the remainder of this section, we shall briefly discuss the implications of our general impossibility theorems for specific classes of graphs, particularly those that satisfy some of the properties of Table 1. We keep this discussion largely abstract; concrete applications are going to be discussed in Section 5. But first let us consider a number of variants of our theorems and mention additional assumptions that would allow us to remove the technical constraint on nonreflexive edges in Theorems 15 and 16, and to instead derive results on full dictatorships and full oligarchies, respectively.

First, note that if we remove the requirement of P being contagious but add the assumption of F being NR-neutral to Theorems 15 and 16, we can still derive the same conclusions (namely, F being NR-oligarchic or NR-dictatorial, respectively). If we impose full neutrality rather than just NR-neutrality, these conclusions can be strengthened to F being fully oligarchic or dictatorial, respectively. For ease of reference, we state these variants here explicitly:

[6] The same is not true for Theorem 15: it is not the case that every oligarchy is collectively rational with respect to every contagious and implicative graph property. The reason is that not every contagious and implicative graph property is closed under intersection, although many concrete such properties (e.g., transitivity) are. For example, the intersection rule does not preserve connectedness (which we have seen to be both contagious and implicative): if agent 1 provides the connected graph $\{(x, y), (x, z), (y, z)\}$ and agent 2 provides the connected graph $\{(x, y), (x, z), (z, y)\}$, then their intersection $\{(x, y), (x, z)\}$ nevertheless fails to be connected.

[7] Note that both Dokow and Holzman [14] and Dietrich and List [51] in fact prove *characterisation results* (in a different sense of that word than we have used in Section 2.4) that have both an impossibility and a possibility component. To use our terminology, they formulate meta-properties that are such that, whenever they are met, then nondictatorial aggregation is impossible, while whenever they are not met, nondictatorial aggregation is possible. We do not consider this second direction here. The reason is that, rather than proving theorems of maximal logical strength, we are interested in theorems that are easy to apply. That this is the case for our choice of meta-properties is going to be demonstrated in Section 5.

[8] This is true for proofs of Arrow's Theorem using the ultrafilter method, which is a refinement of the "decisive coalition method" going back to Arrow's original work [5]. There are, however, other proofs available that exploit the specific structure of preferences, and thus do not generalise to, e.g., judgment aggregation, which some readers will find more accessible [52].

Table 2
Meta-properties of common graph properties.

Property	Contagious?	Implicative?	Disjunctive?
Reflexivity	×	×	×
Irreflexivity	×	×	×
Symmetry	×	×	×
Antisymmetry	×	×	×
Right Euclidean	✓	✓	×
Left Euclidean	✓	✓	×
Transitivity	✓	✓	×
Negative Transitivity	✓	×	✓
Connectedness	✓	✓	✓
Completeness	×	×	✓
Nontriviality	×	×	✓
Seriality	×	×	✓

Theorem 17. *Let P be a graph property that is implicative. Then, for $|V| \geq 3$, any unanimous, grounded, IIE, and (NR-)neutral aggregation rule F that is collectively rational with respect to P must be (NR-)oligarchic.*

Theorem 18. *Let P be a graph property that is implicative and disjunctive. Then, for $|V| \geq 3$, any unanimous, grounded, IIE, and (NR-)neutral aggregation rule F that is collectively rational with respect to P must be (NR-)dictatorial.*

As implicativeness and disjunctiveness are much less demanding properties than contagiousness and as neutrality is often a reasonable axiom to impose, these variants of our main theorems are of some practical interest.

Next, recall that by Proposition 6, unanimity implies collective rationality with respect to reflexivity. Thus, our theorems remain true if we add reflexivity to the collective rationality requirements. In fact, they can be strengthened: for a unanimous rule and under the assumption that all input graphs are reflexive, every NR-dictatorial rule is in fact a full dictatorship and any NR-oligarchic rule is in fact a full oligarchy. Analogously, by Proposition 7 and in view of our assumption of groundedness, we can alternatively add irreflexivity to the collective rationality requirements and strengthen our theorems in the same manner. Thus, we obtain two further variants of Theorem 15 and two further variants of Theorem 16.

A simple instance of the first of these variants of Theorem 16 is Arrow's Theorem for weak orders (i.e., binary relations that are reflexive, transitive, and complete). An aggregation rule mapping profiles of weak orders to weak orders, i.e., a *social welfare function* [5], is simply a graph aggregation rule that is collectively rational with respect to reflexivity, transitivity, and completeness. Arrow uses two axioms, namely *independence* (which is the same as our IIE axiom), and the *weak Pareto condition*, according to which unanimously held strict preferences between two alternatives x and y should be respected by the aggregation rule.

Theorem 19 *(Arrow, 1963). Any weakly Paretian and independent preference aggregation rule, mapping profiles of weak orders over three or more alternatives to weak orders, must be a dictatorship.*

Proof. If we use the edges of a graph to represent weak preferences, then strict preference of x over y means that we accept edge (x, y) but reject edge (y, x). Thus, the weak Pareto condition together with IIE (independence) implies unanimity, while the weak Pareto condition together with collective rationality with respect to completeness implies groundedness.

Now the theorem follows immediately from Theorem 16, together with the insights that (i) transitivity is a graph property that is contagious (Fact 11) and implicative (Fact 13), that (ii) completeness is a graph property that is disjunctive (Fact 14), and that (iii) reflexivity allows us to conclude that the aggregation rule must be a full dictatorship rather than just an NR-dictatorial rule. □

Using the same approach, we can also easily derive a variant of Arrow's Theorem for strict linear preference orders (binary relations that are irreflexive, transitive, and complete) from Theorem 16. In this context, the weak Pareto condition is equivalent to the unanimity axiom, and groundedness is implied by the weak Pareto condition together with the collective rationality requirement for completeness.

But Arrow's Theorem now is just an example. We can immediately obtain any number of impossibility results such as this one, as long as the properties of the graphs we want to work with hit the appropriate meta-properties. Table 2 summarises which of our standard graph properties are contagious (see Fact 11), implicative (see Fact 13), and disjunctive (see Fact 14), respectively. Any combination of graph properties that together hit all three graph properties, by Theorem 16, gives rise to an impossibility theorem saying that all relevant aggregation rules are NR-dictatorial. Similarly, any combination of graph properties that together hit the first two meta-properties, by Theorem 15, gives rise to an impossibility theorem saying that the only relevant aggregation rules are NR-oligarchic. To be precise, when combining several graph properties, one needs to verify that the relevant richness conditions continue to be satisfied (which is trivially the case for all combinations of properties considered in Table 2). To exemplify the possibilities, we state two concrete instances of our general results

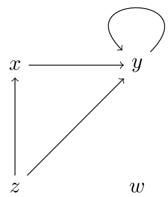


Fig. 4. Example for a modal logic frame with four possible worlds.

explicitly. They are particularly interesting, because they each require collective rationality with respect to just a single graph-property.

Corollary 20. *For $|V| \geq 3$, any unanimous, grounded, and IIE aggregation rule that is collectively rational with respect to transitivity must be oligarchic on nonreflexive edges.*

Corollary 21. *For $|V| \geq 3$, any unanimous, grounded, and IIE aggregation rule that is collectively rational with respect to connectedness must be dictatorial on nonreflexive edges.*

4. Integrity constraints in modal logic

So far we have worked with a definition of collective rationality that applies to every possible graph property (and we have specifically focused on common properties, such as transitivity). An alternative approach is to limit attention to properties that can be expressed in a restricted (logical) language. This is useful when we are interested in algorithmic aspects of collective rationality, e.g., the complexity of checking whether a given model satisfies the constraint (model checking). In our previous work on binary aggregation [16], we have focused on properties expressible in the language of propositional logic. Here, instead, we focus on fundamental properties of graphs that can be expressed using the language of *modal logic* [4].

As we shall see, this is interesting not only because modal logic is a widely used language for describing graphs, but also because the standard semantics of modal logic suggests a new distinction of different *levels* of collective rationality. After a brief review of relevant concepts from modal logic in Section 4.1, we introduce these three levels in Section 4.2. One of them operates at the level of *frames*, one at the level of *models*, and one at the levels of possible *worlds*. The first is equivalent to the basic notion of collective rationality used in the first part of this paper. Results for the other two are presented in Section 4.3 and 4.4, respectively.

4.1. Background: modal logic

In what follows, we briefly review the basic concepts of modal logic and introduce the relevant notation [4]. Fix a finite set Φ of propositional variables. The set of well-formed *formulas* φ is defined as follows (with p ranging over the elements of Φ):

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \varphi \vee \varphi \mid \varphi \rightarrow \varphi \mid \Box\varphi \mid \Diamond\varphi$$

A (Kripke) *model* $M = \langle G, Val \rangle$ consists of a graph $G = \langle V, E \rangle$ and a *valuation function* $Val : \Phi \rightarrow 2^V$. In line with standard terminology, we also refer to G as a (Kripke) *frame*, to V as the set of *possible worlds*, and to E as an *accessibility relation*. The valuation Val is mapping propositional variables p to sets of possible worlds—the worlds where the p in question is true. The *truth* of an arbitrary formula φ at a world $x \in V$ in a model $M = \langle G, Val \rangle$, denoted $M, x \models \varphi$, is defined recursively:

- $M, x \models p$ if $x \in Val(p)$ for any $p \in \Phi$
- $M, x \models \neg\varphi$ if $M, x \not\models \varphi$
- $M, x \models \varphi \wedge \psi$ if $M, x \models \varphi$ and $M, x \models \psi$
- $M, x \models \Diamond\varphi$ if $M, y \models \varphi$ for some $y \in E(x)$

Furthermore, $A \vee B$ is short for $\neg(\neg A \wedge \neg B)$, $A \rightarrow B$ is short for $\neg(A \wedge \neg B)$, and $\Box A$ is short for $\neg\Diamond\neg A$. Intuitively, $M, x \models \Diamond\varphi$ means that φ is true in at least one world accessible from x, and $M, x \models \Box\varphi$ means that formula φ is true in all worlds accessible from x.

Besides this notion of truth of φ at a given world, the semantics of modal logic provides two further ways of interpreting a formula φ on a graph G. First, a formula φ is *globally true* in model $M = \langle G, Val \rangle$, denoted $M \models \varphi$, if $M, x \models \varphi$ for every $x \in V$. Second, φ is *valid on frame* G, denoted $G \models \varphi$, if $\langle G, Val \rangle \models \varphi$ for every valuation Val. Two formulas φ and ψ are *equivalent* if $M, x \models \varphi$ implies $M, x \models \psi$ and *vice versa*, for every model M and every world x.

Table 3
Common frame properties and the corresponding modal formulas.

Property	Modal formula
Reflexivity	$p \to \Diamond p$
Symmetry	$p \to \Box\Diamond p$
Right Euclidean	$\Diamond p \to \Box\Diamond p$
Transitivity	$\Diamond\Diamond p \to \Diamond p$
Connectedness	$\Box(\Box p \to q) \vee \Box(\Box q \to p)$
Seriality	$\Diamond(p \vee \neg p)$

Example 13 *(Frame validity and global truth).* Consider the frame $G = \langle V, E\rangle$, with $V = \{x, y, z, w\}$, shown in Fig. 4. An example for a formula that is valid in this frame is $\Box q \to \Box\Box q$, because—whatever the model—in every world for which all accessible worlds satisfy q also all worlds accessible in exactly two steps satisfy q. The formula $p \to \Diamond p$, on the other hand, is not valid in G, because there exist models based on G, e.g., the model with $Val(p) = \{z\}$, in which it is not the case that from every world in which p is true we can access some world that also satisfies p. However, $p \to \Diamond p$ is globally true in some models based on G, e.g., in the model with $Val(p) = \emptyset$.

Recall that both truth at a world and global truth in a model are concepts that require the introduction of a valuation Val. Validity on a frame, on the other hand, is independent of the valuation and can be used to express global properties of frames, i.e., of graphs alone. For instance, it is well-known that $G = \langle V, E\rangle$ is reflexive (i.e., E is a reflexive relation on V) if and only if the formula $p \to \Diamond p$ is valid on G. To see this, consider a reflexive graph: by reflexivity we know that $x \in E(x)$ and, hence, whenever p is set to true in world x, x can "see" a world where p is true, namely itself, making $\Diamond p$ true. For the converse, if the accessibility relation E is not reflexive, then we can exhibit a valuation and a world at which the formula $p \to \Diamond p$ is false, namely the valuation that sets p to true only at the irreflexive vertex and false in the rest of the model. Results of this kind belong to the realm of *modal correspondence theory* [53]. Given these results, using the concept of validity on a frame, we are able to express a property of a graph by means of a formula in modal logic. Some of the most fundamental frame properties considered in correspondence theory are listed in Table 3. Such formulas can also be combined to characterise classes of graphs of interest. An equivalence relation, for instance, is a frame on which $p \to \Diamond p$, $p \to \Box\Diamond p$, and $\Diamond\Diamond p \to \Diamond p$ are valid. Note that not all graph properties have modal formulas defining them (e.g., irreflexivity, completeness, and negative transitivity do not).[9]

4.2. Three levels of collective rationality

Given a set of propositional variables Φ, we shall refer to modal formulas φ constructed from Φ as *modal integrity constraints*. We now introduce three definitions of collective rationality with respect to a modal integrity constraint. What distinguishes them is the level (frame, model, world) at which the modal integrity constraint is interpreted.

Definition 26. An aggregation rule F is **frame collectively rational** with respect to a modal integrity constraint φ if $\langle V, E_i\rangle \models \varphi$ for all $i \in \mathcal{N}$ implies $\langle V, F(\boldsymbol{E})\rangle \models \varphi$.

That is, F is frame collectively rational with respect to φ if validity of φ on all individual frames $\langle V, E_i\rangle$ implies validity of φ on the collective frame $\langle V, F(\boldsymbol{E})\rangle$. This is equivalent to our original Definition 19, with the only difference being that the property with respect to which we require collective rationality now has to be expressed by means of a modal formula.

Definition 27. An aggregation rule F is **model collectively rational** with respect to a modal integrity constraint φ if for every valuation $Val : \Phi \to 2^V$ we have $\langle\langle V, E_i\rangle, Val\rangle \models \varphi$ for all $i \in \mathcal{N}$ implying $\langle\langle V, F(\boldsymbol{E})\rangle, Val\rangle \models \varphi$.

That is, F is model collectively rational with respect to φ if—for any valuation Val—global truth of φ in all individual models $\langle\langle V, E_i\rangle, Val\rangle$ implies global truth of φ in the collective model $\langle\langle V, F(\boldsymbol{E})\rangle, Val\rangle$.

Definition 28. An aggregation rule F is **world collectively rational** with respect to a modal integrity constraint φ if for every valuation $Val : \Phi \to 2^V$ and every world $x \in V$ we have $\langle\langle V, E_i\rangle, Val\rangle, x \models \varphi$ for all $i \in \mathcal{N}$ implying $\langle\langle V, F(\boldsymbol{E})\rangle, Val\rangle, x \models \varphi$.

Thus, F is world collectively rational with respect to φ if—again, for any valuation—truth of φ at a given world in all individual models implies truth of φ at the same world in the collective model.

[9] The left-Euclidean property defined in Table 1 also cannot be expressed directly. It corresponds to the formula for the right-Euclidean property interpreted on the inverse relation E^{-1}.

Example 14 *(Levels of collective rationality).* Let us go back to our Example 11, in which aggregating three graphs that are serial by means of the majority rule yielded a fourth graph that fails to be serial. Specifically, in the majority graph the world w does not have a successor. In our discussion of Example 11, we concluded that the majority rule is not collectively rational with respect to seriality, which in the terminology of Definition 26 is expressed as the majority rule not being frame collectively rational with respect to $\Diamond(p \vee \neg p)$, a modal formula that corresponds to seriality. On the other hand, by Fact 5, the majority rule is frame collectively rational with respect to $p \to \Diamond p$, corresponding to reflexivity. But note that the majority rule is not model collectively rational with respect to the same formula $p \to \Diamond p$. To see this, consider a model with a valuation *Val* such that p is true at every world. Then $p \to \Diamond p$ is globally true in all individual models, but it is not globally true in the collective model, since the bottom world w is not connected to any of the other p-worlds (i.e., p is true at w, but $\Diamond p$ is not).

A straightforward analysis of Definitions 26–28 yields the following result:

Proposition 22. *Let F be an aggregation rule and let φ be a modal integrity constraint. Then the following implications hold:*

(*i*) *If F is world collectively rational with respect to φ, then F is also model collectively rational with respect to φ.*
(*ii*) *If F is model collectively rational with respect to φ, then F is also frame collectively rational with respect to φ.*

These inclusions are strict. For example, the aggregation rule F that returns the full graph in case all individual graphs satisfy $\Diamond(p \vee \neg p)$, and the empty graph otherwise, is model collectively rational but not world collectively rational. To see this, consider a profile of graphs with two worlds where $E_i = \{(x, y)\}$ for all $i \in \mathcal{N}$. The outcome returned by F is the empty graph, in violation of world collective rationality with respect to $\Diamond(p \vee \neg p)$ at world x. Moreover, Example 14 can be used to show the strict implication in item (ii), since it concerns an aggregation rule that is frame collectively rational with respect to modal formula $p \to \Diamond p$ but not model collectively rational with respect to the same formula.

Thus, frame collective rationality is the least demanding of our three notions of collective rationality and world collective rationality is the most demanding. Hence, negative results are strongest when formulated for frame collective rationality, while positive results are strongest when formulated for world collective rationality. Our (negative) impossibility results of Section 3 were indeed proved for frame collective rationality and these results thus immediately extend also to the other two levels (in those cases where the graph property in question has a corresponding modal formula). The (positive) possibility results for frame collective rationality of Section 2.5, however, do not automatically transfer. Indeed, as we are going to see next, they cannot be extended even to the next level, namely that of model collective rationality. Following this, we are going to complete the picture by establishing a number of positive results for world collective rationality, which immediately transfer to the other two levels as well.

4.3. Limitative results for collective rationality at the level of models

Recall that in Section 2.5 we have seen that every unanimous aggregation rule is collectively rational with respect to reflexivity (Proposition 6) and every neutral aggregation rule is collectively rational with respect to symmetry (Proposition 8). Given the well-known results in modal correspondence theory for these two properties, which we recall in Table 3, we can reformulate these results as follows[10]:

- Any unanimous aggregation rule is frame collectively rational with respect to $p \to \Diamond p$.
- Any neutral aggregation rule is frame collectively rational with respect to $p \to \Box\Diamond p$.

The following two examples show that these results are tight, in the sense that they cease to hold when we replace frame collective rationality by model collective rationality. Both examples use the intersection rule $F_\cap$, which is both unanimous and neutral.

Example 15 *(Counterexample for $p \to \Diamond p$).* Let $V = \{x, y\}$. Suppose two individuals provide the following two graphs: $E_1 = \{(x, y), (y, y)\}$ and $E_2 = \{(y, x), (x, x)\}$, i.e., $F_\cap$ will return the empty graph. Now consider the three models we obtain for these three graphs when we use the valuation $Val(p) = \{x, y\}$, which makes p true at every world. Then the formula $p \to \Diamond p$ is globally true in the two individual models, but it is not globally true in the model based on the collective (empty) graph. Hence, the intersection rule, despite being unanimous, is not model collectively rational with respect to $p \to \Diamond p$.

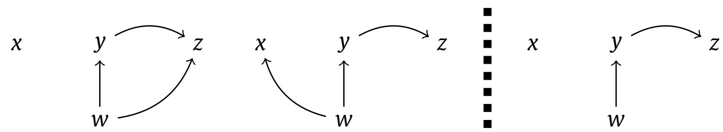
Example 16 *(Counterexample for $p \to \Box\Diamond p$).* Let $V = \{x, y, z\}$. Suppose two individuals report the graphs $E_1 = \{(x, y), (y, z)\}$ and $E_2 = \{(x, y), (y, x)\}$, respectively. If we aggregate using $F_\cap$, we obtain a collective graph with a single edge (x, y). Now consider the valuation $Val(p) = \{x, z\}$. While the formula $p \to \Box\Diamond p$ is globally true in both individual models, the same

10 Note that Proposition 7 cannot be reformulated in an analogous manner, because irreflexivity cannot be expressed in terms of a modal formula.

formula is not satisfied at x in the collective model (while p is true at x, x is connected in the collective graph to y at which $\Diamond p$ is not satisfied). Hence, despite being neutral, $F_\cap$ is not model collectively rational with respect to the modal formula corresponding to symmetry.

For more demanding modal integrity constraints, the situation is even more bleak. For example, we have already seen that transitivity is not preserved under the majority rule, which is an aggregation rule that meets essentially all axioms of interest. This is precisely what the Condorcet paradox shows (see Example 12). Thus, neither unanimity nor neutrality (nor any other basic axiom we have considered) could possibly guarantee an aggregation rule to be frame collectively rational, or indeed model collectively rational, with respect to $\Diamond\Diamond p \to \Diamond p$, the modal formula corresponding to transitivity. The best we can say is that all oligarchic rules are frame collectively rational with respect to $\Diamond\Diamond p \to \Diamond p$. This is so, because the intersection of several transitive graphs is always transitive itself. We conclude our discussion of limitative results with an example showing that even this basic result does not transfer to collective rationality at the level of models.

Example 17 *(Counterexample for $\Diamond\Diamond p \to \Diamond p$).* Let $V = \{x, y, z, w\}$ and suppose two individuals submit the two graphs depicted to the left of the dashed line below:



Under a valuation with $Val(p) = \{x, z, w\}$, the formula $\Diamond\Diamond p \to \Diamond p$ is globally true in both models: world w is the only world where $\Diamond\Diamond p$ is true, and in both models w also satisfies $\Diamond p$. Now, the intersection rule will return the graph shown to the right of the dashed line. In the corresponding model, the antecedent $\Diamond\Diamond p$ is still true at w, but $\Diamond p$ is not, since p is false at y. Hence, the intersection rule is not model collectively rational with respect to $\Diamond\Diamond p \to \Diamond p$.

4.4. Possibility results for collective rationality at the level of worlds

To complete the picture, we are now going to look for possibility results at the level of individual worlds. Recall that, by Proposition 22, any such result we are able to establish will immediately transfer to our other two notions of collective rationality as well. Unlike in Section 2.5, where we proved a number of simple possibility results for collective rationality at the level of frames for specific graph properties, the following results apply to all graph properties that can be expressed as modal integrity constraints meeting certain *syntactic* restrictions.

Recall that a formula is said to be in *negation normal form* (NNF) if it does not make use of the implication operator $\to$ and the negation operator $\neg$ only occurs immediately in front of propositional variables. As is well known, any modal formula can be translated into an equivalent formula in NNF. We call a formula in NNF that does not have any occurrences of $\Diamond$ a $\Box$-formula, and a formula in NNF without any occurrence of $\Box$ a $\Diamond$-formula.

The first straightforward observation to be made is that, if a formula φ does not involve any modal operators ($\Box$ and $\Diamond$), then *any* aggregation rule will be world collectively rational with respect to φ. This is immediate from Definition 28: the truth of such a φ only depends on the valuation *Val*, which is not subject to change during aggregation. For formulas involving only the universal modality $\Box$, we need to ensure that the frame resulting from the aggregation does not include "too many" edges:

Proposition 23. *If an aggregation rule F is such that for every profile $\boldsymbol{E}$ there exists an individual $i^\star \in \mathcal{N}$ such that $F(\boldsymbol{E}) \subseteq E_{i^\star}$, then F is world collectively rational with respect to all $\Box$-formulas.*

Proof. The proof hinges on a basic property of $\Box$-formulas, namely that of being preserved if the set of edges in a model gets reduced by deleting some of the edges. So let φ be a $\Box$-formula and let $\boldsymbol{E}$ be a profile. Fix a world $x \in V$ and a valuation *Val* such that $\langle\langle V, E_i\rangle, Val\rangle, x \models \varphi$ for all $i \in \mathcal{N}$. In particular, we have $\langle\langle V, E_{i^\star}\rangle, Val\rangle, x \models \varphi$. Since, by assumption, F is such that $F(\boldsymbol{E}) \subseteq E_{i^\star}$, all boxed formulas that are true in $\langle\langle V, E_i\rangle, Val\rangle$ at x are also true in the collective model $\langle\langle V, F(\boldsymbol{E})\rangle, Val\rangle$ at x; thus, $\langle\langle V, F(\boldsymbol{E})\rangle, Val\rangle, x \models \varphi$. $\Box$

Note that the individual $i^\star$ in Proposition 23 need not be the same in all profiles. But of course, it *can* be. This observation immediately leads to the following corollary:

Corollary 24. *Any oligarchic aggregation rule is world collectively rational with respect to all $\Box$-formulas.*

For formulas involving only the existential modality $\Diamond$, we have to ensure that the collective model includes "enough" edges:

Proposition 25. *If an aggregation rule F is such that for every profile $\boldsymbol{E}$ there exists an individual $i^\star \in \mathcal{N}$ such that $F(\boldsymbol{E}) \supseteq E_{i^\star}$, then F is world collectively rational with respect to all $\Diamond$-formulas.*

Proof. The proof is analogous to that of Proposition 23, this time using the property of $\Diamond$-formulas being preserved when the set of edges in a model is expanded by adding edges. Let φ be a $\Diamond$-formula and let $\boldsymbol{E}$ be a profile such that, for a given world $x \in V$ and valuation Val, we have that $\langle\langle V, E_i\rangle, Val\rangle, x \models \varphi$ for all $i \in \mathcal{N}$. By assumption, we know that $F(\boldsymbol{E}) \supseteq E_{i^\star}$. Hence, from the fact that $\langle\langle V, E_{i^\star}\rangle, Val\rangle, x \models \varphi$ and that φ is a $\Diamond$-formula we can conclude that $\langle\langle V, F(\boldsymbol{E})\rangle, Val\rangle, x \models \varphi$. $\square$

Examples for aggregation rules that satisfy the assumptions of Proposition 25 are the dictatorships and the union rule. Oligarchic rules (other than the dictatorships), however, do not. Instead, in analogy to Corollary 24, any aggregation rule that always returns the union of the graphs provided by some fixed coalition is world collectively rational with respect to all $\Diamond$-formulas.

Propositions 23 and 25 together suggest a sufficient condition for an aggregation rule to preserve truth for any kind of formula. Recall that a representative-voter rule is an aggregation rule F that is such that for every profile $\boldsymbol{E}$ there exists an individual $i^\star \in \mathcal{N}$ such that $F(\boldsymbol{E}) = E_{i^\star}$ (see Definition 7).

Proposition 26. *Any representative-voter rule is world collectively rational with respect to all modal integrity constraints.*

Proof. Immediate from Definition 28: If the collective graph is a copy of one of the individual graphs, then all formulas that are true at the individual level will remain true at the collective level. $\square$

Proposition 26 is related to a result for binary aggregation characterising the representative-voter rules as those binary aggregation rules that are collectively rational with respect to all *propositional* integrity constraints [16]. Interestingly, for graph aggregation and *modal* integrity constraints, we do *not* obtain such a result; the converse of Proposition 26 does not hold, as demonstrated by the following example.

Example 18 *(Beyond representative-voter rules).* Let $\mathcal{N} = \{1, 2\}$, $V = \{x, y, z\}$, and $\Phi = \{p\}$. Let F be the aggregation rule that is almost the dictatorship of agent 1, except that in case $E_1 = \{(x, y)\}$ and $E_2 = \{(y, z)\}$, rather than reproducing that graph, it returns the empty graph. Then F is not a representative-voter rule, but it nevertheless is world collectively rational with respect to any modal integrity constraint. To see this, first observe that we only need to check for the special profile where F returns the empty graph, as in all other cases the outcome will be equal to the graph of agent 1. Now start by considering $\Box$-formulas: As the outcome graph is empty, any such formula is true at any world in V under any valuation. F is thus world collectively rational with respect to any $\Box$-formula. Next consider $\Diamond$-formulas: In the special profile, for every world in V at least one of the two individual graphs does not have any outgoing edges. Hence, any such formula cannot be true at a given world in all individual models, making any requirement of world collective rationality vacuously satisfied. Finally, for propositional formulas, every aggregation rule is world collectively rational. Thus, we can conclude that F is world collectively rational with respect to any modal integrity constraint.

5. Applications in artificial intelligence

In Section 2.2, we have introduced several scenarios that together exemplify the range of applications in which graph aggregation can play a role. In this section, we are going to revisit some of these scenarios, particularly those featuring prominently in AI research, and show how our results, notably our general impossibility theorems, can be put to use in these domains. Some of the results we are going to present are new, but most of them instead highlight how our approach can be used to clarify known results and to obtain significantly simpler proofs for them. We are going to discuss applications of our approach to preference aggregation for agents that are not perfectly rational (Section 5.1), to nonmonotonic reasoning and belief merging (Section 5.2), to clustering analysis (Section 5.3), and to abstract argumentation in multiagent systems (Section 5.4).

Recall that an Arrovian aggregation rule is a rule that is unanimous, grounded, and IIE. We are going to use this terminology throughout this section. Also, to simplify the statements of theorems, when in this section we speak of "aggregation rules for X", with X being some family of graphs, we are referring to aggregation rules that are collectively rational with respect to the graph properties characterising X. For example, Arrow's Theorem speaks about aggregation rules for weak orders, i.e., aggregation rules that are collectively rational with respect to the three graph properties defining weak orders.

5.1. Bounded rationality: aggregation of incomplete preferences

In the economics literature, and thus in essentially all classical contributions to social choice theory, preferences are usually assumed to be *complete*. Thus, for any two alternatives, a decision maker is assumed to be able to decide which of them she prefers or whether she is indifferent between them. In AI, on the other hand, such an assumption would often be considered controversial. Rather, an agent may not always be able to provide a complete preference order. This kind of bounded rationality could be due to the agent lacking relevant information or due to her lacking the necessary computational resources to arrive at a complete ranking. This is particularly relevant in domains where agents are asked to express preferences over very large sets of alternatives. Indeed, many of the formal preference representation languages developed in AI, such as CP-nets [54], are not even able to express all complete preference orders [55].

It therefore is important to understand the options available to us for aggregating *incomplete preferences*, which are often modelled as *preorders*, i.e., binary relations that are reflexive and transitive.[11] First, observe that Arrow's Theorem does *not* apply to the aggregation of such incomplete preferences. A simple counterexample is the intersection rule, which is unanimous, grounded, IIE, and collectively rational with respect to both reflexivity and transitivity, i.e., it correctly maps profiles of preorders to single preorders—yet it is not a dictatorship. Of course, the intersection rule does not qualify as a very attractive rule either. It is an oligarchic rule, and in fact we can easily prove the following characterisation result:

Theorem 27. *Let F be an aggregation rule for preferences—modelled as preorders—over three or more alternatives. Then F is Arrovian if and only if it is oligarchic.*

Proof. The left-to-right direction follows from Theorem 15, as transitivity is contagious and implicative, and as reflexivity permits us to reduce NR-oligarchies to full oligarchies. The other direction is immediate. □

Let us say that a preorder E *has maxima* if there exists at least one element such that no other element is strictly preferred to it:

$$\exists x.\forall y.xEy \quad (E \text{ has maxima})$$

Thus, the preference order modelled by E may be incomplete, but there is at least one element that is at least as preferable as any other. Similarly, let us say that E *has minima* if there exists at least one element that is at least as bad as any other element:

$$\exists x.\forall y.yEx \quad (E \text{ has minima})$$

Pini et al. [10] study Arrovian impossibilities for incomplete preferences in detail. They call an incomplete preference order (i.e., a preorder) "restricted" if it has maxima or minima (or both). Their main result is a variant of Arrow's Theorem for such restricted incomplete preferences [10, Theorem 5]:

Theorem 28 *(Pini et al., 2009). Any Arrovian aggregation rule for preferences—when modelled as preorders that have maxima or minima—over three or more alternatives must be a dictatorship.*

Proof. The claim follows from Theorem 16, considering that transitivity is contagious and implicative, having maxima or minima is disjunctive, and reflexivity allows us to remove the restriction to nonreflexive edges. In other words, the proof is identical to that of Theorem 19, except that now the disjunctive property of having maxima or minima takes over the role of the disjunctive property of completeness. □

In fact, Theorem 28 is slightly stronger than the result stated by Pini et al., who only require preferences to be restricted in the output but admit arbitrary preorders in the input (note that by admitting a wider range of inputs, encountering an impossibility becomes more likely). Besides making available a much simpler proof than the one originally given by Pini et al., our approach shows that the focus on preorders that have maxima or minima is somewhat arbitrary. Any other property that is disjunctive, such as the strictly weaker nontriviality property (see Table 1), would have delivered the same result.

Pini et al. also prove variants of other classical theorems, notably the Muller–Satterthwaite Theorem and the Gibbard–Satterthwaite Theorem. Discussing these results is beyond the scope of this paper. Having said this, it is well known that in the classical setting they can be obtained as relatively simple corollaries to Arrow's Theorem [56], so our approach is likely to have fruitful applications also here.

5.2. Nonmonotonic reasoning and belief merging

Aggregation plays a role in several contributions to the literature on nonmonotonic reasoning in AI. This is the case both for models of commonsense reasoning for a single intelligent agent who has to aggregate the possibly conflicting views arising from several different inference rules [19], and for work on merging the beliefs of several agents in a multiagent system [20]. In some approaches to nonmonotonic reasoning, alternative states of belief that an agent or a multiagent system might adopt are structured in terms of plausibility orderings that indicate which states are preferred to which other states according to a given criterion or a given individual agent. Such plausibility orders (often referred to as preferences in

[11] Thus, a weak order, which we have used to model preferences up to this point, is a preorder that is complete.

the literature) of course are graphs, so this boils down to a question of graph aggregation.[12] Plausibility orders are reflexive and transitive, i.e., they are naturally modelled as preorders. In addition, different authors impose different additional requirements. We now review two contributions to nonmonotonic reasoning that involve graph aggregation.

The starting point of Doyle and Wellman [19] is the observation that prior attempts at integrating various specialised patterns of commonsense inference into a universal logic of nonmonotonic reasoning have failed, and they try to explain this observation in terms of an Arrovian impossibility result for plausibility orders. They recognise that Arrow's Theorem does not extend to the aggregation of preorders, but also do not consider adding a completeness requirement as being appropriate in this context. Instead, besides independence and the weak Pareto condition, they invoke one additional axiom. Doyle and Wellman call an aggregation rule F *conflict-resolving* if, for all $x, y \in V$, it is the case that if $(x, y) \in E_i$ holds for at least one $i \in \mathcal{N}$, then $(x, y) \in F(\boldsymbol{E})$ or $(y, x) \in F(\boldsymbol{E})$ must hold as well. That is, if at least one agent ranks x and y, then the output of F must rank x and y as well (but not necessarily in the same direction). The main theorem of Doyle and Wellman may be paraphrased as follows [19, Theorem 4.3]:

Theorem 29 *(Doyle and Wellman, 1991). Any aggregation rule for plausibility orders—when modelled as preorders—over three or more states of belief that is Arrovian and conflict-resolving must be a dictatorship.*

Proof. First, observe that every conflict-resolving aggregation rule is collectively rational with respect to nontriviality. This is straightforward from the definition of being conflict-resolving: Let $\boldsymbol{E}$ be profile of nontrivial graphs. Then there exist two vertices x and y such that $(x, y) \in E_1$, which implies that either (y, x) or (x, y) itself need to be in the collective graph. But then the claim is strictly weaker than claiming that, for $|V| \geq 3$, any Arrovian aggregation rule for nontrivial preorders must be dictatorial, which follows from Theorem 16 using the by now familiar approach. □

Doyle and Wellman prove their result by inspection of a published proof of Arrow's Theorem, noting that, in that proof, collective rationality with respect to completeness is only ever used when at least one individual expresses a preference between the relevant two alternatives. This is a valid approach, and indeed, the result of Doyle and Wellman is the theorem most similar to Arrow's original result amongst all the impossibility theorems discussed in this paper. Having said this, we believe that there is some added value in showing their result to be an immediate corollary to another theorem (as we have done here) rather than just showing how it follows *from the proof* of another theorem (as Doyle and Wellman have done), as this makes it considerably easier for others to verify the result and to prove similar new results themselves.

In work on belief merging, Maynard-Zhang and Lehmann [20] model plausibility orders as preorders that satisfy the property of negative transitivity (see Table 1), which they call *modularity*. They argue that assuming negative transitivity rather than completeness, together with a modification of the independence axiom, allows them to circumvent Arrow's Theorem and to make reasonable aggregation rules available for belief merging. In the discussion of their result, they stress the significance of both of these changes. However, our analysis clearly shows that replacing completeness by negative transitivity alone has no effect on Arrow's impossibility, as negative transitivity is also a disjunctive property (see Fact 14). Hence, the crucial source for the possibility result of Maynard-Zhang and Lehmann must be their modification of the independence axiom. Indeed, this modification is rather substantial, as it allows for independence to be violated whenever not doing so would lead to what they term a "conflict". Thus, our approach is helpful also in this context in pinpointing the precise sources of impossibilities, thereby providing guidance on how they can be avoided.

5.3. Consensus clustering

Given a set of data points, *clustering* is the task of partitioning that set into subsets, in a way that in some sense is meaningful or useful [31]. For example, someone designing an advertising campaign may wish to cluster a dataset about the past purchasing behaviour of a large group of people into a small number of groups of people with similar characteristics. Or someone designing a medical treatment may wish to cluster a medical dataset into subsets of patients with similar symptoms. Clustering has been exceptionally successful in practice, but is still lacking precise theoretical foundations. It is often difficult—and sometimes arguably impossible—to define what would constitute a "correct" clustering. The process of trying to find a compromise between the output of several different clustering algorithms is known as *consensus clustering*. Consensus clustering can be modelled as a problem of graph aggregation. To see this, observe that a specific clustering of a given set of data points can be modelled as an *equivalence relation* (i.e., a graph) on that set, by stipulating that two points are equivalent if and only of they belong to the same cluster.

Recall that an equivalence relation on a set V is a binary relation on V that is reflexive, symmetric, and transitive. To the best of our knowledge, Mirkin [58] was the first to analyse the aggregation of equivalence relations using the axiomatic method. Below we state a very similar result due to Fishburn and Rubinstein [22], who in their paper refer to oligarchies as "conjunctive operators".

[12] In other approaches to belief merging, belief bases themselves rather than the underlying plausibility orders are being aggregated [40]. These approaches are closely related to judgment aggregation [57], rather than graph aggregation, and we shall not discuss them here.

Theorem 30 *(Fishburn and Rubinstein, 1986). Any Arrovian aggregation rule for equivalence relations—which may represent alternative clusterings of a common dataset—over three or more data points must be an oligarchy.*

Proof. This follows from Theorem 15, together with the fact that transitivity is both a contagious and an implicative graph property, and the observation that collective rationality with respect to reflexivity eliminates the need to distinguish between NR-oligarchic and fully oligarchic rules. The additional requirement of collective rationality with respect to symmetry does not affect the result; in particular, it is easy to verify that the richness conditions in the definitions of contagiousness and implicativeness can still be met. □

In fact, not every possible clustering will be useful. In particular, the clustering that puts every single data point in its own little cluster might meet most of the required definitions (e.g., it vacuously ensures that similarity between data points of the same cluster is always greater than similarity between data points belonging to different clusters), but it hardly will be helpful in understanding the structure of the data or in using it. Note that this kind of trivial clustering corresponds to the empty graph. Thus, we may assume that all individual graphs are nontrivial (as defined in Table 1) and we may wish to impose the same constraint on the result of the aggregation rule, i.e., we may wish to impose collective rationality with respect to nontriviality. If we do so, we can further tighten the impossibility result of Fishburn and Rubinstein[13]:

Theorem 31. *Any Arrovian aggregation rule for nontrivial equivalence relations—which may represent alternative clusterings of a common dataset—over three or more data points must be a dictatorship.*

Proof. This follows from Theorem 16, in the same way as Theorem 30 follows from Theorem 15, together with the fact that nontriviality is a disjunctive graph property (see Fact 14). □

Thus, it is impossible to design useful algorithms for consensus clustering that operate on each pair of data points independently.

In a related line of research, an overview of which is available in the work of Barthélemy et al. [59], similar impossibility results have been obtained for the problem of consensus finding in the context of richer forms of classification that go beyond mere clustering. For example, Leclerc [60] has obtained such a result for valued quasi-orderings, which generalise both equivalence relations and weak orders. These results are similar in spirit to ours, in the sense that they also deal with the aggregation of information, but we use a different tool, namely graphs, to represent information.

While our approach applies to the problem of finding a consensus between the outputs produced by several clustering algorithms, we note that there also has been work on characterising those clustering algorithms themselves that is based on ideas originating in social choice theory [61,62].

5.4. Multiagent argumentation

The final application scenario introduced in Section 2.2 we are going to discuss in some more detail here is that of argumentation in multiagent systems. An *abstract argumentation framework* is a graph, the vertices of which are the *arguments* and the edges of which represent a so-called *attack-relation* between arguments. This model was introduced in the seminal work of Dung [33], who proposed several different semantics for abstract argumentation frameworks that specify principles according to which we may accept or reject arguments given the attacks between them. For example, if we accept argument x, and if x attacks y, then we should not also accept y.

In a multiagent system, each agent may be associated with a different abstract argumentation framework on the same set of arguments, i.e., each agent may have different views on what constitutes a valid attack. We may then wish to merge these different frameworks to arrive at a suitable representation of the views of the group as a whole. The aggregation of abstract argumentation frameworks has been studied by a number of authors [34,23,35,63].[14] Next, we review some of this work and demonstrate that there are several interesting connections to our own work on graph aggregation, which suggests that graph aggregation can be fruitfully applied also in this domain.

Coste-Marquis et al. [34] were the first to consider the problem of aggregating several argumentation frameworks. They propose a distance-based method for aggregation. While they formulate the unanimity axiom as a relevant property in the context of aggregation of argumentation frameworks, they do not explicitly link their work to social choice theory.

Tohmé et al. [23] were the first to make an explicit link to social choice theory. They formulate several choice-theoretic axioms for the aggregation of argumentation frameworks, e.g., an independence axiom and a (strong) monotonicity axiom (which is equivalent to the conjunction of our monotonicity axiom and IIE). They study collective rationality with respect to

[13] We are grateful to Shai Ben-David for alerting us to this connection between consensus clustering and our Dictatorship Theorem (personal communication, June 2015).

[14] In related work, other authors have studied the aggregation of alternative *extensions* of a given common abstract argumentation framework, i.e., alternative choices on which arguments to accept [64–66]. This line of work is more closely related to judgment aggregation and we shall not review it here. Bodanza and Auday [67] compare these two distinct approaches of combining abstract argumentation and social choice theory.

acyclicity. Acyclicity is an important graph property in the context of argumentation, because for an acyclic argumentation framework it is unambiguous which arguments to accept.[15] Acyclicity does not satisfy any of our three meta-properties (contagiousness, implicativeness, disjunctiveness), so our general impossibility theorems do not apply. Still, as Tohmé et al. argue, the options for designing an aggregation rule that is collectively rational with respect to acyclicity are very limited. Clearly, every oligarchic rule is collectively rational with respect to acyclicity, because acyclicity of graphs is preserved under intersection. In addition, as Tohmé et al. point out, also any aggregation rule based on a *collegium*, i.e., a coalition of agents who each can veto any given edge from being accepted, but who may not be able to jointly enforce the acceptance of an edge (as would be the case in an oligarchy), is also collectively rational with respect to acyclicity, besides being Arrovian.

Dunne et al. [35] introduced further choice-theoretic axioms into the study of the aggregation of abstract argumentation frameworks (also discussed by Delobelle et al. [63]). Arguably, some of their "axioms" are better characterised as collective rationality requirements. For example, their "nontriviality axiom" in fact is just collective rationality with respect to non-triviality of graphs (as defined in Table 1). Probably the most important innovation in the work of Dunne et al. [35] is the introduction of collective rationality requirements (albeit not under this name) with respect to graph properties that are specific to the context of abstract argumentation, such as the property of being "decisive" (in the sense of not permitting any ambiguity about which arguments are to be accepted). While, as explained above, acyclicity entails decisiveness, the converse is not true, i.e., studies of collective rationality with respect to acyclicity can only ever approximate the properties we should postulate for an aggregation rule for argumentation frameworks.

Modal logic can be used to define a semantics for argumentation frameworks by specifying rules for labelling arguments in a given argumentation framework as being either "in" or "out", or possibly "undecided" [68,69]. This provides yet another connection to our work on graph aggregation. Let $\Phi = \{\mathtt{in}, \mathtt{out}, \mathtt{undec}\}$. We can use the following formula to express that every argument must get labelled using exactly one of these three options:

$$(\mathtt{in} \wedge \neg\mathtt{out} \wedge \neg\mathtt{undec}) \vee (\neg\mathtt{in} \wedge \mathtt{out} \wedge \neg\mathtt{undec}) \vee (\neg\mathtt{in} \wedge \neg\mathtt{out} \wedge \mathtt{undec})$$

In addition, we can express constraints on the labelling of arguments that are linked to each other by means of the attack-relation. Let our graph describe the inverse of the attack-relation in an argumentation framework (rather than the attack-relation itself). Thus, the formula $\Diamond\mathtt{in}$, for example, will be true at a world, if that world represents an argument that is attacked by an argument that is "in", i.e., that is accepted. The formula $\Box\mathtt{out}$ is true if all attacking arguments are "out" (i.e., rejected). We now may wish to impose some of the following modal integrity constraints:

- $\mathtt{in} \to \Box\mathtt{out}$ (expressing that an argument can only be "in", if all of its attackers are "out")
- $\Box\mathtt{out} \to \mathtt{in}$ (expressing that, if all of an argument's attackers are "out", then it should be "in")
- $\mathtt{out} \to \Diamond\mathtt{in}$ (expressing that an argument should only be "out", if one of its attackers is "in")
- $\Diamond\mathtt{in} \to \mathtt{out}$ (expressing that an argument that has an attacker that is "in" must be "out")

A labelling that satisfies all of four of these constraints corresponds to what Dung calls a *complete extension* [33,68]. A labelling that furthermore does not label any argument as being undecided, i.e., that makes $\neg\mathtt{undec}$ true at every world, corresponds to a so-called *stable extension* [33,68].

Observe that each one of the four formulas above is equivalent to either a $\Box$-formula or a $\Diamond$-formula, although the conjunction of all four is not. Thus, in case we, for instance, are only interested in the first two of them, we can refer to Proposition 23 to identify aggregation rules that are collectively rational with respect to these modal integrity constraints. If, however, we require an aggregation rule that preserves the property of having a complete (or stable) extension, then the best we can say at this point is that, by Proposition 26, any representative-voter rule meets this kind of requirement.

6. Conclusion

We have introduced the problem of graph aggregation and analysed it in view of its possible use to combine information coming from different agents who each specify an alternative set of edges on the same set of vertices. Our focus has been on the concept of collective rationality, i.e., the preservation of certain properties of graphs under aggregation. Our results are formulated with respect to various meta-properties that may or may not be met by a specific property of graphs one may be interested in. We have explored two different approaches to the definition of such meta-properties. Using a semantic approach, we have defined certain templates (namely contagiousness, implicativeness, and disjunctiveness), which are easy to recognise in common graph properties and which make the features of graph properties required to carry through our proofs particularly salient. Using a syntactic approach, we have used formulas expressible in certain fragments of modal logic to describe properties of graphs.

Most of our technical results establish conditions under which it is either possible or impossible to guarantee collective rationality with respect to graph properties that meet certain meta-properties. Our main technical result is a generalisation

[15] First, accept all arguments that are not attacked by any argument. Then reject all arguments that are attacked by at least one accepted argument. Then accept all arguments that are only attacked by rejected arguments. Finally, repeat the last two steps until all arguments are either accepted or rejected. This process is well-defined in case of acyclicity.

of Arrow's Theorem for preference aggregation to aggregation problems for a large family of types of graphs that include the types of graphs used by Arrow to model preferences. To establish this theorem, as well as a closely related theorem identifying conditions that can only be satisfied by an oligarchic aggregation rule, we have refined the (ultra)filter method for proving impossibility theorems in social choice theory. Besides these technical contributions, we have also demonstrated how insights from the abstract setting of graph aggregation can be put to use in a variety of application domains.

While we have been able to demonstrate that our choice of meta-properties is particularly useful for quickly proving results in a wide variety of different domains, our impossibility theorems only establish sufficient conditions for impossibilities and there is room for future research on other such sufficient conditions and also for a complete characterisation of the family of types of graphs for which Arrovian aggregation is impossible. A good starting point for such an undertaking would be closely related work in judgment aggregation [14,51].

Much of this paper has focused on Arrovian aggregation rules, and more specifically on the consequences of accepting the axiom of independence. The prime direction to escape our impossibilities is therefore to relax this axiom, and to study rules that are not independent. In Section 2.3, we have already briefly mentioned the idea of adapting distance-based aggregation rules—familiar from preference aggregation, judgment aggregation, and belief merging—to our setting. Distance-based rules are collectively rational by definition, but unfortunately have the drawback of typically being computationally intractable. Investigating the trade-off between complexity and collective rationality when designing aggregation rules for specific classes of graph aggregation problems thus presents an important challenge for future research.

Besides such technical investigations, future work should continue to focus on applications of graph aggregation. Our discussion in Section 5 demonstrates the usefulness of adopting the general perspective of graph aggregation in the domains of preference aggregation, nonmonotonic reasoning and belief merging, cluster analysis, and argumentation. Future work should also address the other application scenarios identified in Section 2.2 and it should identify new ones. One promising direction concerns work on *theory change* in the philosophy of science, where one recent model has used the Arrovian framework of preference aggregation to analyse how scientists choose between rival scientific theories in terms of preferences induced by criteria such as simplicity or fit with available data [70]. The more general framework of graph aggregation opens up new possibilities for investigating the subtle differences that presumably exist between the preferences of an economic agent and the preferences induced by scientific criteria for accepting a novel theory. Another—entirely different but equally promising—direction for future research is in the area of the Semantic Web and concerns work on *XML data integration* [71]. The basic structure underlying documents encoded in XML (the *extensible markup language*) is that of a tree, i.e., a special kind of graph. Thus, if we want to combine information encoded using XML that has been obtained from different sources on the Semantic Web, we need to use some form of graph aggregation as well.[16] But also this extended list of potential applications is bound to be incomplete, given the ubiquity of graphs across so much of science and scholarship.

References

[1] U. Endriss, U. Grandi, Graph aggregation, in: Proceedings of the 4th International Workshop on Computational Social Choice, COMSOC-2012, AGH University of Science and Technology, Kraków, 2012.
[2] U. Endriss, U. Grandi, Collective rationality in graph aggregation, in: Proceedings of the 21st European Conference on Artificial Intelligence, ECAI-2014, 2014.
[3] K.J. Arrow, A.K. Sen, K. Suzumura (Eds.), Handbook of Social Choice and Welfare, Elsevier, 2002.
[4] P. Blackburn, M. de Rijke, Y. Venema, Modal Logic, Cambridge University Press, 2001.
[5] K.J. Arrow, Social Choice and Individual Values, 2nd ed., John Wiley and Sons, 1963, first edition published in 1951.
[6] P.C. Fishburn, Arrow's impossibility theorem: concise proof and infinite voters, J. Econ. Theory 2 (1970) 103–106.
[7] A. Kirman, D. Sondermann, Arrow's theorem, many agents, and invisible dictators, J. Econ. Theory 5 (1972) 267–277.
[8] B. Hansson, The existence of group preference functions, Public Choice 28 (1976) 89–98.
[9] A.K. Sen, Social choice theory, in: K.J. Arrow, M.D. Intriligator (Eds.), Handbook of Mathematical Economics, vol. 3, North-Holland, 1986, pp. 1073–1181.
[10] M.S. Pini, F. Rossi, K.B. Venable, T. Walsh, Aggregating partially ordered preferences, J. Log. Comput. 19 (2009) 475–502.
[11] C. List, P. Pettit, Aggregating sets of judgments: an impossibility result, Econ. Philos. 18 (2002) 89–110.
[12] P. Gärdenfors, A representation theorem for voting with logical consequences, Econ. Philos. 22 (2006) 181–190.
[13] F. Dietrich, C. List, Judgment aggregation by quota rules: majority voting generalized, J. Theor. Polit. 19 (2007) 391–424.
[14] E. Dokow, R. Holzman, Aggregation of binary evaluations, J. Econ. Theory 145 (2010) 495–511.
[15] F. Herzberg, D. Eckert, The model-theoretic approach to aggregation: impossibility results for finite and infinite electorates, Math. Soc. Sci. 64 (2012) 41–47.
[16] U. Grandi, U. Endriss, Lifting integrity constraints in binary aggregation, Artif. Intell. 199–200 (2013) 45–66.
[17] C. List, C. Puppe, Judgment aggregation: a survey, in: P. Anand, P. Pattanaik, C. Puppe (Eds.), Handbook of Rational and Social Choice, Oxford University Press, 2009, pp. 457–482.
[18] F. Brandt, V. Conitzer, U. Endriss, J. Lang, A.D. Procaccia (Eds.), Handbook of Computational Social Choice, Cambridge University Press, 2016.
[19] J. Doyle, M.P. Wellman, Impediments to universal preference-based default theories, Artif. Intell. 49 (1991) 97–128.
[20] P. Maynard-Zhang, D.J. Lehmann, Representing and aggregating conflicting beliefs, J. Artif. Intell. Res. 19 (2003) 155–203.
[21] H.C. White, S.A. Boorman, R.L. Breiger, Social structure from multiple networks, I: blockmodels of roles and positions, Am. J. Sociol. (1976) 730–780.
[22] P.C. Fishburn, A. Rubinstein, Aggregation of equivalence relations, J. Classif. 3 (1986) 61–65.
[23] F.A. Tohmé, G.A. Bodanza, G.R. Simari, Aggregation of attack relations: a social-choice theoretical analysis of defeasibility criteria, in: Proceedings of the 5th International Symposium on Foundations of Information and Knowledge Systems, FoIKS-2008, Springer-Verlag, 2008.

[16] This approach would be complementary to recently taken initial steps of another kind of use of the methodology of social choice theory in the area of the Semantic Web, namely the application of ideas originating in judgment aggregation to ontology merging [47].

[24] S. Airiau, U. Endriss, U. Grandi, D. Porello, J. Uckelman, Aggregating dependency graphs into voting agendas in multi-issue elections, in: Proceedings of the 22nd International Joint Conference on Artificial Intelligence, IJCAI-2011, 2011.
[25] J. Hintikka, Knowledge and Belief: An Introduction to the Logic of the Two Notions, Cornell University Press, 1962.
[26] P. Egré, Epistem logic, in: L. Horsten, R. Pettigrew (Eds.), The Bloomsbury Companion to Philosophical Logic, Bloomsbury Academic, 2011, pp. 503–542.
[27] D. Lewis, Elusive knowledge, Australas. J. Philos 74 (1996) 549–567.
[28] R.J. Aumann, Agreeing to disagree, Ann. Stat. (1976) 1236–1239.
[29] J.Y. Halpern, Y. Moses, Knowledge and common knowledge in a distributed environment, J. ACM 37 (1990) 549–587.
[30] Y. Shoham, Reasoning about Change: Time and Causation from the Standpoint of Artificial Intelligence, MIT Press, 1987.
[31] P.-N. Tan, M. Steinbach, V. Kumar, et al., Introduction to Data Mining, Pearson, 2005.
[32] A. Gionis, H. Mannila, P. Tsaparas, Clustering aggregation, ACM Trans. Knowl. Discov. Data 1 (2007) 4.
[33] P.M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, Artif. Intell. 77 (1995) 321–358.
[34] S. Coste-Marquis, C. Devred, S. Konieczny, M.-C. Lagasquie-Schiex, P. Marquis, On the merging of Dung's argumentation systems, Artif. Intell. 171 (2007) 730–753.
[35] P.E. Dunne, P. Marquis, M. Wooldridge, Argument aggregation: basic axioms and complexity results, in: Proceedings of the 4th International Conference on Computational Models of Argument, COMMA-2012, IOS Press, 2012.
[36] S. Airiau, E. Bonzon, U. Endriss, N. Maudet, J. Rossit, Rationalisation of profiles of abstract argumentation frameworks, in: Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems, AAMAS-2016, 2016.
[37] X. Wen, H. Liu, Logic aggregation, in: Proceedings of the 4th International Workshop on Logic, Rationality, and Interaction, Springer-Verlag, 2013.
[38] S.J. Brams, P.C. Fishburn, Approval Voting, 2nd ed., Springer, 2007.
[39] J. Kemeny, Mathematics without numbers, Daedalus 88 (1959) 577–591.
[40] S. Konieczny, R. Pino Pérez, Merging information under constraints: a logical framework, J. Log. Comput. 12 (2002) 773–808.
[41] M.K. Miller, D. Osherson, Methods for distance-based judgment aggregation, Soc. Choice Welf. 32 (2009) 575–601.
[42] E. Hemaspaandra, H. Spakowski, J. Vogel, The complexity of Kemeny elections, Theor. Comput. Sci. 349 (2005) 382–391.
[43] U. Endriss, U. Grandi, D. Porello, Complexity of judgment aggregation, J. Artif. Intell. Res. 45 (2012) 481–514.
[44] J. Lang, M. Slavkovik, How hard is it to compute majority-preserving judgment aggregation rules?, in: Proceedings of the 21st European Conference on Artificial Intelligence, ECAI-2014, 2014.
[45] U. Endriss, R. de Haan, Complexity of the winner determination problem in judgment aggregation: Kemeny, Slater, Tideman, Young, in: Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems, AAMAS-2015, 2015.
[46] U. Endriss, U. Grandi, Binary aggregation by selection of the most representative voter, in: Proceedings of the 28th AAAI Conference on Artificial Intelligence, AAAI-2014, 2014.
[47] D. Porello, U. Endriss, Ontology merging as social choice: judgment aggregation under the open world assumption, J. Log. Comput. 24 (2014) 1229–1249.
[48] U. Endriss, Judgment aggregation, in: F. Brandt, V. Conitzer, U. Endriss, J. Lang, A.D. Procaccia (Eds.), Handbook of Computational Social Choice, Cambridge University Press, 2016.
[49] I. McLean, A.B. Urken (Eds.), Classics of Social Choice, University of Michigan Press, Ann Arbor, 1995.
[50] B.A. Davey, H.A. Priestley, Introduction to Lattices and Order, 2nd ed., Cambridge University Press, 2002.
[51] F. Dietrich, C. List, Arrow's theorem in judgment aggregation, Soc. Choice Welf. 29 (2007) 19–33.
[52] J. Geanakoplos, Three brief proofs of Arrow's impossibility theorem, Econ. Theory 26 (2005) 211–215.
[53] J. van Benthem, Correspondence theory, in: D. Gabbay, F. Guenthner (Eds.), Handbook of Philosophical Logic, 2nd ed., Kluwer Academic Publishers, 2001, pp. 167–247.
[54] C. Boutilier, R.I. Brafman, C. Domshlak, H.H. Hoos, D. Pool, CP-nets: a tool for representing and reasoning with conditional ceteris paribus preference statements, J. Artif. Intell. Res. 21 (2004) 135–191.
[55] J. Lang, Logical preference representation and combinatorial vote, Ann. Math. Artif. Intell. 42 (2004) 37–71.
[56] U. Endriss, Logic and social choice theory, in: A. Gupta, J. van Benthem (Eds.), Logic and Philosophy Today, vol. 2, College Publications, 2011, pp. 333–377.
[57] P. Everaere, S. Konieczny, P. Marquis, Belief merging versus judgment aggregation, in: Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems, AAMAS-2015, 2015.
[58] B.G. Mirkin, On the problem of reconciling partitions, in: H.M. Blalock, et al. (Eds.), Quantitative Sociology: International Perspectives on Mathematical and Statistical Modelling, Academic Press, 1975, pp. 441–449.
[59] J.-P. Barthélemy, B. Leclerc, B. Monjardet, On the use of ordered sets in problems of comparison and consensus of classifications, J. Classif. 3 (1986) 187–224.
[60] B. Leclerc, Efficient and binary consensus functions on transitively valued relations, Math. Soc. Sci. 8 (1984) 45–61.
[61] M. Ackerman, S. Ben-David, Measures of clustering quality: a working set of axioms for clustering, in: Proceedings of the 22nd Annual Conference on Neural Information Processing, NIPS-2008, 2008.
[62] J. Kleinberg, An impossibility theorem for clustering, in: Proceedings of the 15th Annual Conference on Neural Information Processing Systems, NIPS-2002, 2002.
[63] J. Delobelle, S. Konieczny, S. Vesic, On the aggregation of argumentation frameworks, in: Proceedings of the 24th International Joint Conference on Artificial Intelligence, IJCAI-2015, 2015.
[64] M. Caminada, G. Pigozzi, On judgment aggregation in abstract argumentation, Auton. Agents Multi-Agent Syst. 22 (2011) 64–102.
[65] I. Rahwan, F. Tohmé, Collective argument evaluation as judgement aggregation, in: Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems, AAMAS-2010, 2010.
[66] R. Booth, E. Awad, I. Rahwan, Interval methods for judgment aggregation in argumentation, in: Proceedings of the 14th International Conference on Principles of Knowledge Representation and Reasoning, KR-2014, 2014.
[67] G.A. Bodanza, M.R. Auday, Social argument justification: some mechanisms and conditions for their coincidence, in: Proceedings of the 10th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU-2009, Springer-Verlag, 2009.
[68] M.W.A. Caminada, D.M. Gabbay, A logical account of formal argumentation, Stud. Log. 93 (2009) 109–145.
[69] D. Grossi, On the logic of argumentation theory, in: Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems, AAMAS-2010, 2010.
[70] S. Okasha, Theory choice and social choice: Kuhn versus Arrow, Mind 120 (2011) 83–115.
[71] A. Halevy, A. Rajaraman, J. Ordille, Data integration: the teenage years, in: Proceedings of the 32nd International Conference on Very Large Data Bases, VLDB-2006, 2006.