Opinion Dynamics with Median Aggregation

Petra Berenbrink
University of Hamburg
Hamburg, Germany
petra.berenbrink@uni-hamburg.de

Martin Hoefer
RWTH Aachen University
Aachen, Germany
mhoefer@cs.rwth-aachen.de

Dominik Kaaser
TU Hamburg
Hamburg, Germany
dominik.kaaser@tuhh.de

Marten Maack
University of Hamburg
Hamburg, Germany
marten.maack@uni-hamburg.de

Malin Rau
Chalmers University of Technology
Gothenburg, Sweden
malin.rau@chalmers.se

Lisa Wilhelmi
RWTH Aachen University
Aachen, Germany
wilhelmi@algo.rwth-aachen.de

ABSTRACT

Understanding the formation and evolution of opinions is of broad interdisciplinary interest. Many classical models for opinion formation focus on the impact of different notions of *locality*, e.g., locality due to network effects among agents or the role of the proximity of opinions. In practice, however, opinion formation is often governed by the interplay of *local* and *global* influences.

In this paper, we study an asynchronous opinion dynamics in a social network. Each agent has a static intrinsic opinion as well as a public opinion that is updated asynchronously over time. Moreover, agents have access to a global aggregate (e.g., the outcome of a vote) of all public opinions. We focus on the popular median voting rule and show that pure Nash equilibria always exist. For every initial state of the dynamics, a pure equilibrium can be reached. The set of reachable equilibria forms a complete lattice, and extremal equilibria can be computed in polynomial time. Indeed, there are instances and initial states from which the number of reachable equilibria is exponentially large. The global median in these equilibria can be any of the initial opinions.

We show that by uniformly increasing the influence of the aggregate median we can enforce that the median opinion is the same in every reachable equilibrium. We can compute the increase scheme that achieves this property in polynomial time. Furthermore, we show that finding the k most influential agents is NP-complete.

CCS CONCEPTS

• **Theory of computation → Algorithmic game theory and mechanism design**; *Distributed algorithms.*

KEYWORDS

Opinion Formation; Median Voting; Nash Equilibria

ACM Reference Format:
Petra Berenbrink, Martin Hoefer, Dominik Kaaser, Marten Maack, Malin Rau, and Lisa Wilhelmi. 2025. Opinion Dynamics with Median Aggregation. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025*, IFAAMAS, 9 pages.

Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Y. Vorobeychik, S. Das, A. Nowé (eds.), May 19 – 23, 2025, Detroit, Michigan, USA.

1 INTRODUCTION

Opinion formation processes [27, 28, 40] model how individuals develop, modify and express opinions in a social context. Opinion formation is highly influenced by various factors, including the social environment, domain experts, or professionals for PR and advertising. Such processes play a crucial role in public discourse, decision-making, and collective behavior [35, 47, 48]. They decide the outcome of elections and the fate of political parties, the success of new products, companies, and entire economies; they influence whether political movements are successful, and they lead to new trends and directions in research and development. Understanding how opinions are formed is essential for promoting constructive dialogue, managing conflicts, and designing effective communication strategies. Research on opinion formation processes, therefore, sheds light on the emergence of collective phenomena, such as public opinion shifts [47], social movements [38], and the spread of misinformation [23]. Understanding the formation and evolution of opinions is of broad interdisciplinary interest, including research in sociology, economics, mathematics, physics, and computer science.

Many classical models for opinion formation focus on the impact of social network effects (such as, e.g., Friedkin-Johnsen [34] or voter models [37]) or the role of the proximity of opinions (such as, e.g., Deffuant-Weisbuch [21] or Hegselmann-Krause [36]). These models introduce notions like *locality* between agents or *distance* between opinions. In practice, however, opinion formation is often governed by the interplay of *local* and *global* influences. Agents are exposed to the opinions of their friends, but they also have access to public opinion polls, media reports, research studies or other forms of aggregated information about the global opinion landscape.

In this paper our goal is to shed light on the interplay between global and local aspects of opinion dynamics. We study an agent-based opinion formation process on a static social network graph $G = (V, E)$. Each agent $i \in V$ has a (static) *intrinsic opinion* $s_i \in \mathbb{R}$ that they keep to themselves and a *public opinion* $z_i \in \mathbb{R}$ that they disclose to their neighbors in G. In addition to these local opinions, agents have access to a global aggregate opinion $f(\boldsymbol{z})$ of the *strategy profile* $\boldsymbol{z}$, i.e., the vector $\boldsymbol{z}$ of all agents' public opinions. We assume that agents update their public opinion in a sequential fashion. For the choice of the public opinion z_i, agent i strives to strike a balance between (1) their intrinsic opinion s_i, (2) the public opinions z_j in their local neighborhood $N(i) = \{j \mid (i, j) \in E\}$, and (3) the global aggregate opinion $f(\boldsymbol{z})$. Balance here means to minimize a *cost function* comprising the distances between the public opinions, the intrinsic opinion, and the global aggregate opinion $f(z)$. More precisely, the cost function of agent i is given by

$$\text{cost}_i(z) = \alpha_i|s_i - z_i| + \sum_{j\in N(i)} \beta_{ij}|z_j - z_i| + \gamma_i|f(z) - z_i|.$$

with weights α_i, β_{ij}, and γ_i for each $j \in N(i)$. We focus on the popular *median voting rule* $f(z) = \text{med}(z)$. Median voting has many favorable properties, e.g., optimality properties, as well as incentive compatibility for all agents [16, 41].

Filter bubbles [44] are a phenomenon known to occur in real-world social networks. They describe the fact that algorithmic curation in social networks disconnects users from information that disagrees with their viewpoints. The resulting *social media echo chambers* [17] are known to inhibit opinion exchange while amplifying extreme views. Models for opinion formation that are solely based on the *locality* of agents and the *proximity* of the opinions are known to model this behavior, see, e.g., the prominent model by Hegselmann and Krause [36]. Since the expressed opinions of individuals in a modern society are not only influenced by their local peer groups [42] but also subject to global influences, we assume in our model that agents have access to a global aggregate of all opinions. As our main contribution, we will show that the range and quality of reachable equilibria depends crucially on the weight each agent assigns to this global aggregate opinion. This implies that the median opinion in the society can be stabilized by giving agents access to the global aggregate opinion, shattering filter bubbles, enriching perspectives and promoting diverse discourse.

Overview of Results. Our first class of results shed light on the structural properties of such opinion dynamics in networks. We show that for any agent i the cost function is minimized by choosing the public opinion as a weighted median of their intrinsic opinion, the public opinions of their neighbors, and the aggregate opinion. We show that for any instance (defined by the network and the intrinsic opinions), there exists a *(pure Nash) equilibrium* where no agent can further reduce their costs. Moreover, for any instance and an initial vector of public opinions, an equilibrium can be reached by a sequence of best-response dynamics. The set of *all* such reachable equilibria forms a complete lattice, i.e., there is a pointwise maximal and a pointwise minimal reachable equilibrium. These equilibria can be computed in polynomial time.

Our results highlight the impact of the order in which the agents update their public opinion. We show that there exists a family of instances in which any of the intrinsic opinions can become the final median, depending on the order in which the agents are updated. Furthermore, in this family, the number of reachable equilibria is exponentially large in the number of opinions and agents, and the update sequence of the agents is crucial to steer the process towards one or another equilibrium. These results also show that the dynamics can potentially terminate in many different equilibria with fundamentally different aggregation outcomes.

Our second class of results shed light on the global influence of the median rule on possible equilibria. We observe that the range of reachable equilibria is decreasing as soon as the *impact of the median opinion* increases. Here, increasing the impact of the median opinion means that each component of the vector $\boldsymbol{\gamma}$ is increased by an additive value δ. We show that for every instance where the global median is not unique in all reachable equilibria, there is a threshold δ such that the following observation holds: if the median weight of all agents is increased by at least δ, then the global median is the same in all reachable equilibria of the modified instance. Additionally, we show that the threshold δ can be computed efficiently.

We contrast this result with a different perspective on influencing the global median. Instead of increasing $\boldsymbol{\gamma}$ for all agents by an additive term δ, suppose we select k agents to increase their median weights. The decision problem StableMedian is whether for a given k there exists a set of k agents such that each reachable equilibrium has the same global median. We show it is NP-complete.

2 RELATED WORK

The literature on opinion formation is vast. A comprehensive survey is beyond the scope of this paper. We give a brief overview of models that are commonly analyzed in the AI and multi-agent community. For a large number of additional references we refer to [14].

Early works on modeling opinion formation include Downs [24], Abelson and Bernstein [1], and DeGroot [22]. In their seminal paper [34], Friedkin and Johnsen introduce a model of opinion formation where each agent has an intrinsic opinion and a public opinion. Agents update their public opinion based on their own intrinsic opinion and the public opinions of their neighbors in a social network graph. Co-evolutionary and game-theoretic variants of this process are studied in [7, 8, 9, 26, 33], with a focus on the existence of equilibria and their *social quality*, measured by the price of anarchy.

Epitropou et al. [26] build upon the model by Friedkin and Johnsen and study continuous opinion formation games with aggregation aspects. They use the average public opinion to model the interplay of global properties of the opinions in the society and local influences and prove the existence of a unique equilibrium. As their main contribution the authors prove that even with outdated information about the average their model converges to the unique equilibrium within distance $\varepsilon > 0$ after $O(n^2 \log(n/\varepsilon))$ synchronous updates on the average. In addition, the authors bound the price of anarchy of average-oriented opinion formation games.

Fanelli and Fotakis [31] analyze preference games with local aggregation. They assume that agents have an innate preference and a public strategy which is affected by the agents' social neighbors. At each agent, an aggregation function "summarizes" the opinions of all agents. The authors provide a comprehensive set of general results on the existence and the structure of equilibria and on the price of anarchy of preference games with local aggregation.

Opinion formation processes are studied intensively in the AI and multi-agent systems community. Auletta et al. [3] present a co-evolutionary model, where the intrinsic opinion changes over time. A substantial body of work considers opinion diffusion, the spread of information and opinions among social networks [2, 10, 11, 13, 20, 30]. In particular, Bredereck et al. [12] study a process where agents in a network have preference rankings over a set of alternatives. Agents update their ranking sequentially by applying an aggregation method to the rankings in their neighborhood. The authors focus on single-peaked rankings and examine which aggregation rules preserve this property. Moreover, Zhan et al. [50]

analyze the impact of the network structure on the speed and extent of opinion diffusion and Coates et al. [18, 19] present a framework and a simulator for agent-based opinion formation processes.

Chazelle [15] shows that Hegselmann-Krause systems with synchronous updates always converge to a stable state where no further opinion changes are possible. A series of related works have analyzed bounds on the time it takes until such a stable state is reached [5, 29, 39, 46, 49]. Variants of the model restrict communication to a predefined social network [6, 43] or consider asynchronous activation of agents [4, 32].

3 MODEL AND PRELIMINARIES

A *median opinion game* is a tuple $\mathcal{G} = (G, \boldsymbol{s}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$. We are given a set V of n agents who are connected via a directed graph $G = (V, E)$. Each agent $i \in V$ has an intrinsic opinion $s_i \in \mathbb{R}$. Moreover, each agent can choose a public opinion $z_i \in \mathbb{R}$ as a strategy. An individual cost for agent i is given by the weighted sum of distances to their intrinsic opinion, the public opinions of their out-neighbors $N(i)$, and the median of public opinions $\operatorname{med}(\boldsymbol{z})$:

$$\operatorname{cost}_i(\boldsymbol{z}) = \alpha_i|s_i - z_i| + \sum_{j \in N(i)} \beta_{ij}|z_j - z_i| + \gamma_i|\operatorname{med}(\boldsymbol{z}) - z_i|.$$

All $\alpha_i, \gamma_i, \beta_{ij} \in \mathbb{N}_0$ are constant and non-negative integers. β_{ij} is defined as the weight of the edge (i, j). In that case that all $\alpha_1 = \alpha 2 = \ldots = \alpha_n$, $\gamma_1 = \gamma_2 = \ldots = \gamma_n$, and $\beta_{i,j} = \beta_{i',j'}$ for all $(i, j), (i', j') \in E$ we call the opinion game *uniform*. We assume that the median function $\operatorname{med}(\cdot)$ always refers to the lower median.

We consider opinion dynamics starting from any initial strategy profile $\boldsymbol{z}$. In each step, one agent i is chosen, either randomly or by an arbitrary deterministic rule. Agent i updates their public opinion to a best response $z'_i \in \arg\min_{y\in\mathbb{R}} \operatorname{cost}_i(y, \boldsymbol{z}_{-i})$, i.e., an opinion that minimizes their cost function. Note that this choice depends on $\operatorname{med}(y, \boldsymbol{z}_{-i})$ rather than $\operatorname{med}(\boldsymbol{z})$. Clearly, there might be several best responses (see example directly after Definition 3.1). We use $B_i(\boldsymbol{z}) = \arg\min_{y\in\mathbb{R}} \operatorname{cost}_i(y, \boldsymbol{z}_{-i})$ to denote the *set* of all best responses for i in $\boldsymbol{z}$. We define $\operatorname{best}_i^-(\boldsymbol{z}) = \min B_i(\boldsymbol{z})$ and $\operatorname{best}_i^+(\boldsymbol{z}) = \max B_i(\boldsymbol{z})$ as the smallest and largest best-response for agent i in profile $\boldsymbol{z}$, respectively. For some of our results, we assume that each agent chooses a best response that minimizes the distance to their latest public opinion. We call this the *lazy* best-response and denote it by $\operatorname{best}_i^l(\boldsymbol{z})$. Note that the intrinsic opinion of an agent remains static.

In the following we characterize best-response strategies. First we need the following definition.

Definition 3.1 (Weighted Median). Given a vector $\boldsymbol{x} = (x_i)_{i\in[n]}$ with n pairwise different values x_i, and a vector $\boldsymbol{w} = (w_i)_{i\in[n]}$ of weights $w_i \geq 0$, a value a is called *weighted median* of $(\boldsymbol{x};\boldsymbol{w})$ if

$$\sum_{\substack{j\in[n] \\ x_j<a}} w_j \leq \frac{\|\boldsymbol{w}\|_1}{2} \quad \text{and} \quad \sum_{\substack{j\in[n] \\ x_j>a}} w_j \leq \frac{\|\boldsymbol{w}\|_1}{2}.$$

Note that for each $(\boldsymbol{x};\boldsymbol{w})$ there is at least one $j \in [n]$ such that x_j is a weighted median. The weighted median is unique, except if we have $\sum_{j=1}^{i^*} x_j = \|\boldsymbol{w}\|_1/2$ for some value $i^* \in [n-1]$, because then each value $a \in [x_{i^*}, x_{i^*+1}]$ is a weighted median.

Proposition 3.2. *A value z_i^* is a best response for agent i against $\boldsymbol{z}_{-i}$ if and only if it is a weighted median of*

$$(s_i, \boldsymbol{z}^{N(i)}, \operatorname{med}(z_i^*, \boldsymbol{z}_{-i});\ \alpha_i, \boldsymbol{\beta}^{N(i)}, \gamma_i),$$

where $\boldsymbol{z}^{N(i)}$ is the vector of public opinions z_j of all neighbors $j \in N(i)$ of agent i and $\boldsymbol{\beta}^{N(i)}$ the corresponding vector of weights β_{ij} for all $j \in N(i)$.

Proof. Consider $\operatorname{cost}_i(z_i, \boldsymbol{z}_{-i})$ as a function of z_i. We show the function represents a piecewise linear function that decreases until the first weighted median is reached, stays constant until the last weighted median is exceeded, and increases afterwards.

Let $V' = \{i\} \cup N(i)$, $x_j = z_j$, $x_i = s_i$, $w_i = \alpha_i$, and $w_j = \beta_{ij}$ for each $j \in N(i)$.

First, assume for simplicity that the weight γ_i of the median of public opinions is zero and can therefore be ignored. If z_i is strictly smaller than the smallest weighted median, increasing z_i results in a linear cost reduction of the function, until the next value from $\{z_j \mid j \in N(i)\} \cup \{s_i\}$ is reached. The gradient of this linear function is $\sum_{j\in V':x_j\leq z_i} w_j - \sum_{j\in V':x_j>z_i} w_j$. Each time a value from $\{z_j \mid j \in N(i)\} \cup \{s_i\}$ is reached, the slope changes but remains negative until the first weighted median is reached and turns positive only after the last weighted median is passed.

Now if the median of public opinions has a weight strictly larger than zero, the argument is slightly more complicated. The contribution of the median of public opinions remains the same as long as the public opinion of agent i is strictly larger or strictly smaller then the median. If z_i is changed in the direction of the median, then the cost due to the median decreases linearly until the median is reached. At this point the public opinion of agent i itself is the median of public opinions. If z_i is changed further in the same direction, the cost contribution of the median remains at zero until the value of the next opinion is reached and that value takes on the role of the median. Afterwards, the cost contribution starts to grow linearly again. Hence, the cost of agent i depending on z_i remains a convex and piecewise linear function. The minima of this function are exactly the weighted medians. □

Proposition 3.2 has many consequences. Best responses represent a closed interval in $\mathbb{R}$, i.e., $B_i(\boldsymbol{z}) = [\operatorname{best}_i^-(\boldsymbol{z}), \operatorname{best}^+(\boldsymbol{z})]$. Moreover, the smallest, largest, and lazy best responses of every agent i are monotone in the strategy profile $\boldsymbol{z}$.

Corollary 3.3. *Suppose $\boldsymbol{z} \leq \boldsymbol{z}'$ pointwise. Then $\operatorname{best}_i^-(\boldsymbol{z}) \leq \operatorname{best}_i^-(\boldsymbol{z}')$, $\operatorname{best}_i^+(\boldsymbol{z}) \leq \operatorname{best}_i^+(\boldsymbol{z}')$, and $\operatorname{best}_i^l(\boldsymbol{z}) \leq \operatorname{best}_i^l(\boldsymbol{z}')$.*

Yet another consequence of Proposition 3.2 is a restriction of the strategy space. If agents always choose their best responses from $\operatorname{best}_i^+(\boldsymbol{z})$, $\operatorname{best}_i^-(\boldsymbol{z})$ or $\operatorname{best}_i^l(\boldsymbol{z})$, then at any point of time the opinions in the strategy profile remain a subset of the initial public and intrinsic opinions.

Observation 3.4. *During any dynamics starting from $\boldsymbol{z}$, in which every agent i chooses from the three best responses $\operatorname{best}_i^-(\boldsymbol{z})$, $\operatorname{best}_i^+(\boldsymbol{z})$, and $\operatorname{best}_i^l(\boldsymbol{z})$, all chosen strategies come from $\{s_i\} \cup \{z_j \mid j \in V\}$.*

Proof. Proposition 3.2 guarantees that all three best responses are a weighted median of

$$(s_i, \boldsymbol{z}^{N(i)}, \operatorname{med}(\operatorname{best}_i(\boldsymbol{z}), \boldsymbol{z}_{-i}); \alpha_i, \boldsymbol{\beta}^{N(i)}, \gamma_i). \tag{1}$$

If the weighted median is unique, then $\text{best}_i^-(z) = \text{best}_i^+(z) = \text{best}_i^l(z)$ and the median has to be one of the values from the vector in (1), which are always included in $\{s_i\} \cup \{z_j \mid j \in V\}$. If $B_i(z)$ is an interval, then each of the three best responses is one of the borders, which are again included in $\{s_i\} \cup \{z_j \mid j \in V\}$. □

Observe that no new opinions are introduced during such dynamics. For a given initial profile, let $I = \{s_j, z_j \mid j \in V\}$ be the set of *initial opinions* in z and s. When considering such dynamics, we can assume w.l.o.g. that the set of possible opinions is restricted to I with at most $2n$ opinions. Moreover, the lower weighted median is independent of the numeric opinion values and based only on (weights and) the total ordering of opinions. As such, we can assume that the available opinions are the integers $1, 2, \ldots, |I| \leq 2n$.

4 ANALYSIS OF EQUILIBRIA

We start by analyzing the properties of equilibria and introduce an efficient method to compute equilibria, the 2-Phase algorithm. We show that the set of equilibria has substantial structure. For any initial profile z, let $\Xi = \{z' \mid \min I \leq z'_i \leq \max I \text{ for all } i \in V\}$ be the subset of strategy profiles bounded by initial opinions.

Proposition 4.1. *For any median opinion game $\mathcal{G}$ with initial profile z, the set $\{z^* \mid z^* \in \Xi$ is an equilibrium$\}$ forms a complete lattice with respect to $\leq$ (componentwise).*

Proof. The set of strategy profiles Ξ forms a complete lattice with respect to $\leq$ (componentwise). By Corollary 3.3 the multi-function f that maps each component to $B_i(z)$ is monotone. By Observation 3.4 and monotonicity we see that if $z \in \Xi$, then $B_i(z) \subseteq \Xi$ for all $i \in V$. Hence, $f : \Xi \to 2^{\Xi}$, and the fixed-points of f (in which $z_i \in B_i(z)$ for all $i \in V$) are exactly the equilibria from Ξ. An extension of the Knaster-Tarski theorem to order-preserving multi-functions [25, 45, 51] shows that the set of fixed-points of f forms a complete lattice. □

Since every complete lattice consists of at least one element, Proposition 4.1 implies that every game has at least one equilibrium. Moreover, when we restrict to equilibria from Ξ, the componentwise maximal and minimal equilibria are both unique.

Beyond the mere existence of equilibria, we are interested in equilibria that can be *reached* by best-response dynamics. We call such an equilibrium *reachable* (from initial profile z). Indeed, for every game and initial strategy profile z, there is at least one reachable equilibrium, and it can be computed efficiently using the following 2-Phase algorithm.

2-Phase Algorithm: In the beginning of each round $k \geq 1$ in phase 1, consider the current state $z^{(k)}$ (where $z^{(1)} = z$). We consider the set $U(z_k) = \{i \mid z_i^{(k)} \notin B_i(z^{(k)}), \text{best}_i^+(z^{(k)} > z_i^{(k)}\}$ of agents that want to deviate to a larger opinion. We pick an agent $i \in \arg\max\{\text{best}_j^+(z^{(k)} \mid j \in U(z^{(k)})\}$ that wants to deviate to the largest best-response and allow them to deviate. We continue phase 1 until no agent wants to deviate to a higher opinion. Let $\acute{z}$ be the state that emerges after phase 1.

In the beginning of each round $k \geq 1$ in phase 2, consider the current state $z^{(k)}$ (where $z^{(1)} = \acute{z}$). We consider the set $D(z^{(k)}) = \{i \mid \text{best}_i^+(z^{(k)}) < z_i^{(k)}\}$) of agents that want to deviate to a lower opinion. We pick an agent $i \in \arg\min\{\text{best}_j^+(z^{(k)} \mid j \in D(z^{(k)})\}$ whose largest best-response is smallest and allow them to deviate. We continue phase 2 until no agent wants to deviate to a lower opinion. Let $\hat{z}$ be the state that emerges after phase 2.

There are two variants of the algorithm. We call the one described above the *ascending* 2-Phase algorithm. In the *descending* 2-Phase algorithm, the phases are reversed, we consider smallest best-responses best_i^- in both phases, and among the agents that want to deviate we choose one for which best_i^- is smallest in phase 1 and largest in phase 2. If not stated otherwise, we only consider the ascending 2-Phase algorithm in the following proofs, as the arguments for the two variants are very analogous. We first observe that our algorithms indeed converge to an equilibrium.

Lemma 4.2. *For any median opinion game $\mathcal{G}$ and initial profile z, both variants of the 2-Phase algorithm compute an equilibrium in $O(n)$ steps. The resulting equilibrium is unique over possible tie-breaking in choosing the deviating agent i in each step.*

Proof. We only consider the ascending 2-Phase algorithm. Consider any round k of phase 1 in which some agent i is chosen to deviate. Profile $z^{(k+1)}$ evolves after i deviated to $z_i^{(k+1)} = \text{best}_i^+(z^{(k)})$. Consider any other agent $j \neq i$. Due to monotonicity (Corollary 3.3), $\text{best}_j^+(z^{(k+1)}) \geq \text{best}_j^+(z^{(k)})$. The set of agents that want to deviate to smaller opinions can only shrink. Since best responses are weighted medians, $\text{best}_j^+(z^{(k+1)})$ can grow to at most $z_i^{(k+1)}$. Thus, if $\text{best}_j^+(z^{(k)}) > z_i^{(k+1)}$ previously, then after the deviation it remains $\text{best}_j^+(z^{(k+1)}) = \text{best}_j^+(z^{(k)})$. Overall, $\text{best}_j^+(z^{(k+1)}) \leq \max\{\text{best}_j^+(z^{(k)}), z_i^{(k+1)}\}$ for all $j \in V$. We choose agent i to deviate such that $\text{best}_i^+(z^{(k)})$ is maximal. Hence, all agents in $U(z^{(k+1)})$ want to deviate to at most $z_i^{(k+1)}$. Thus, $z_i^{(k+1)} = \text{best}_i^+(z^{(k+1)})$ remains the largest best-response throughout the rest of phase 1. Each agent i deviates at most once in phase 1.

If there are several agents $i \in U(z^{(k)})$ with maximal $z' = \text{best}_i^+(z^{(k)}) = \max\{\text{best}_j^+(z^{(k)}) \mid j \in U(z^{(k)})\}$, by the above arguments they will be chosen sequentially in the algorithm and all deviate to z' independent of their ordering. As such, the state $\acute{z}$ emerging at the end of phase 1 is unique.

The analysis of phase 2 follows symmetrically. By the same argument, we choose the agent from $D(z^{(k)})$ with smallest $\text{best}_i^+(z^{(k)})$ in phase 2. In the emerging profile $z^{(k+1)}$, by monotonicity, we only lower all $\text{best}_j^+(z^{(k+1)}) \leq \text{best}_j^+(z1(k))$. This maintains the invariant that no player wants to deviate to a higher opinion. Moreover, by the properties of weighted medians, $\text{best}_j^+(z^{(k+1)}) \geq \min\{\text{best}_j^+(z^{(k)}), z_i^{(k+1)}\}$. Thus, each agent deviates at most once, and if there are several agents $i \in D(z^{(k)})$ with smallest $\text{best}_i^+(z^{(k)})$, the order in which they are allowed to deviate does not affect their chosen opinion. □

For each initial profile z, the two versions of the 2-Phase algorithm yield two unique reachable equilibria. These equilibria bound all of the reachable equilibria from z.

Theorem 4.3. *For any median opinion game $\mathcal{G}$ and initial profile z, the ascending (descending) 2-Phase algorithm computes the unique* *maximum (minimum) reachable equilibrium in polynomial time. The maximum (minimum)* median opinion *among all reachable equilibria occurs at the maximum (minimum) equilibrium.*

Proof. Consider the state $\acute{z}$ after phase 1. We will first show inductively that $\acute{z} \geq \tilde{z}$ (componentwise) for any state $\tilde{z}$ reachable by best-response dynamics from $\mathbf{z}$. We number the agents based on the round they deviated in phase 1. The statement clearly holds for agent 1. $\text{best}_1^+(\mathbf{z})$ is the largest opinion any agent wants to deviate to in $\mathbf{z}$. No deviation of any agent j to $x \leq \text{best}_1^+(\mathbf{z})$ can make $\text{best}_k^+(\mathbf{z})$ grow to beyond x, for any $k \in V$. Hence, in any sequence of best responses starting from $\mathbf{z}$, we maintain the invariant that no agent ever wants to deviate to an opinion $x > \text{best}_1^+(\mathbf{z})$. This proves the statement for agent 1. Now, inductively, given that the statement holds for agents $1, \ldots, k-1$, consider agent k. We again use $\mathbf{z}_k$ to denote the state in the beginning of round k of the phase 1. Agents $j = 1, \ldots, k-1$ are at their maximal reachable opinions $\acute{z}_j \geq \text{best}_k^+(\mathbf{z}_k)$. By monotonicity, this only increases $\text{best}_j^+(\mathbf{z}_j)$ for $j = k, k+1, \ldots$. By the choice of k, all other agents $j \geq k$ want to deviate to opinions at most $\text{best}_k^+(\mathbf{z}_k)$, even after k deviates to $\text{best}_k^+(\mathbf{z}_k)$. Thus, for any sequence of best responses from $\mathbf{z}$, given that agents $1, \ldots, k-1$ never play higher opinions, agents $k, k+1, \ldots$ never want to deviate to any opinion $x > \text{best}_k^+(\mathbf{z}_k)$. This proves the statement for agent k, and thus all agents that deviate during phase 1 of the 2-Phase algorithm. Now finally, consider all agents i with $\acute{z}_i = z_i$ who did not deviate during phase 1. By monotonicity, these agents must also have $\tilde{z}_i \leq z_i$ in any reachable state.

Consider any reachable equilibrium $\bar{z}$. Since $\acute{z} \geq \tilde{z}$ for any reachable state $\tilde{z}$, phase 2 maintains the invariant that $\text{best}_i^+(z) \geq \text{best}_i^+(\bar{z})$ for all $i \in V$, by monotonicity. Thus, the equilibrium computed by 2-Phase fulfills $\hat{z} \geq \bar{z}$.

An analogous argument using the descending 2-Phase algorithm proves the statement regarding the minimum equilibrium. By monotonicity, the maximum and minimum median opinions among the reachable equilibria have to occur at the (componentwise) maximum and minimum equilibria, respectively. □

5 NUMBER OF EQUILIBRIA

By Proposition 3.2, best responses form a closed interval. Hence, for some initial state $\mathbf{z}$, there are trivial examples with infinitely many reachable equilibria. In this section we will show that even for *lazy* best-response dynamics (in which all best responses remain within the set I of initial opinions) the number of reachable equilibria can be very large, and the median opinions can be very diverse. We can show these properties even for instances with *undirected* graphs.

Let $k = 2^q$ with $q \in \mathbb{N}, q \geq 2$. In the following we define a uniform family of instances T_k (games and initial states) with k opinions and $n = \text{poly}(k)$ agents with the following properties:

- For each opinion j ($1 \leq j \leq k$), there exists a reachable equilibrium where the median of the public opinions equals the public opinion j.
- The number of reachable equilibria is exponentially large in k.

The family $\mathcal{G}_k = (T_k, \mathbf{s}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ with corresponding initial profile $\mathbf{z}$ is defined as follows. For a fixed k, $T_k(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ consists of $\log k$ layers with $\mathbf{s} \subset \{1, \ldots k\}$. See Figure 1 for an example. The first

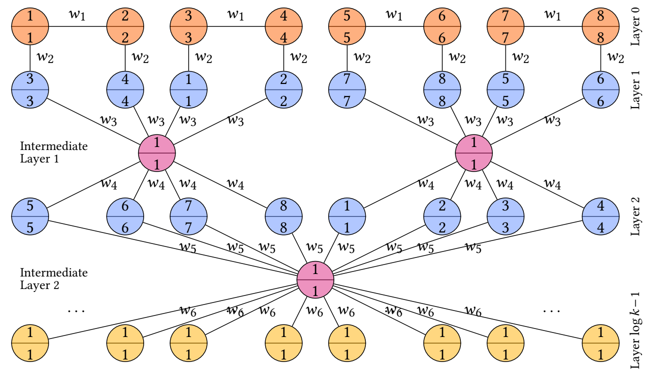

Figure 1: Tournament graph T_k for $k = 8$ with 8 player agents (orange), 16 transfer agents (blue), 3 elimination agents (purple), and 48 stabilizing agents (yellow).

$\log k - 1$ layers consist of k agents each, while the last layer has $2k \log k$ agents.

- The jth agent of layer 0 is denoted p_j (colored orange) and it is initialized with private opinion $z_{p_j} = s_{p_j} = j$. We call these agents *player* agents.
- For $1 \leq \ell \leq \log k - 2$ the jth agent of the ℓth layer is denoted $t_{j,\ell}$. All agents $t_{j,\ell}$ with $1 \leq j \leq k, 1 \leq \ell \leq \log k - 2$ (colored blue) are initialized with opinion $z_{t_{j,\ell}} = s_{t_{j,\ell}} = \left\lfloor \frac{j}{4\ell} \right\rfloor + ((j + 2\ell) \bmod 4\ell)$. They are called *transfer* agents.
- The jth agent from the last layer $\log k - 1$ is denoted m_j (colored yellow) and they are initialized with opinion $z_{m_j} = s_{m_j} = k/2$. These agents are called *median stabilizing* agents.

In addition to these layers we have $\log k - 2$ *intermediate layers* (colored pink), layer $1 \leq i < \log k - 2$ containing $2^{\log k - 1 - i}$ many agents. The jth agent of the ℓth layer is denoted $e_{j,\ell}$ and is initialized with opinion $z_{e_{j,\ell}} = s_{e_{j,\ell}} = 1$. We call these agents *elimination* agents. It remains to define the edges of T_k.

- Player agents: for each $i = 2j - 1$, $j \in \mathbb{N}$, we connect p_i and p_{i+1} with an edge of weight $w_1 = w_2 + \alpha + \beta + \gamma + 1$. Additionally, p_j will be connected to a transfer agent $t_{j,1}$ (blue) with an edge of weight w_2.
- Elimination agents: agents at level $1 \leq \ell < \log(k) - 2$ are connected to transfer agents at level ℓ and level $(\ell + 1)$ with edges of weight

$$w_{2\ell+1} = w_{2\ell} + \alpha + \beta + \gamma + 1$$
$$\text{and} \quad w_{2\ell+2} = \frac{1}{2}\left(\gamma + \beta + \alpha + 2^{\ell+1} \cdot w_{2\ell+3}\right) + 1,$$

respectively. More precisely, elimination agent $e_{j,\ell}$ is connected to transfer agents $t_{(j-1)2^{\ell+1}+1,\ell}$ to $t_{j \cdot 2^{\ell+1},\ell}$ and transfer agents $t_{(j-1)2^{\ell+1}+1,\ell+1}$ to $t_{j \cdot 2^{\ell+1},\ell+1}$ if $\ell < \log k - 1$. The single elimination agent at level $\log k - 2$ is connected at level $\log k - 2$ with edge weight $w_{2\log(k/2)-3}$. Furthermore, it is connected to all median stabilizing agents with edge weight $w_{2\log(k/2)-2} = \alpha + \beta + \gamma + 1$.

Lemma 5.1. *Suppose that $k = 2^q$ with $q \in \mathbb{N}, q \geq 2$. Let $T_k(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ be an uniform game with $\alpha, \beta, \gamma \in \mathbb{N}_0$ with k opinions and $O(k \log(k))$*

agents. For each $1 \leq j \leq k$ there exists an equilibrium where the global median takes on the value j.

Proof sketch. In the following we specify an activation order in which each agent is activated exactly once. We start with propagating opinion j from the jth player agent to the bottom of $T_k(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ via the elimination agents. As soon as opinion j reaches the elimination agent on level $\log k - 2$, *all* median stabilizing agents are updated. All these agents adopt opinion j. After these updates the global median becomes opinion j. After that the remaining nodes are activated one layer after the other, starting from layer 0. All agents who have been already updated cannot change their opinion anymore. This holds due to the corresponding edge weights. Hence, due to the sheer amount of median stabilizing agents, opinion j remains global median. □

For each player pair p_i, p_{i+1} for $i = 2j + 1$, $j \in \mathbb{N}$, in each equilibrium both agents have either opinion i or opinion $i + 1$. The following corollary is a consequence.

Corollary 5.2. *$T_k(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ has an exponential (in k) number of reachable equilibria.*

6 STABILIZING THE MEDIAN

Let $\mathcal{G} = (G, \boldsymbol{s}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ be a median opinion game with initial profile $\boldsymbol{z}$. We call the pair $\mathcal{G}, \boldsymbol{z}$ *stable* if all equilibria reachable from $\boldsymbol{z}$ share the same global median. In this section we consider the question of how to change the median weights γ_i of a given opinion game $\mathcal{G}$ with a fixed initial profile $\boldsymbol{z}$ such that the pair becomes stable. We say that these changes *stabilize* $\mathcal{G}$ and $\boldsymbol{z}$.

Observe that stabilizing $\mathcal{G}$ and $\boldsymbol{z}$ is always possible when all the median weights γ_i are increased by some value δ that is large enough. In particular, there is only one reachable equilibrium – and hence a unique global median – if γ_i for each agent i is larger than the weight of the intrinsic opinion, together with the overall weight of the neighbors of agent i. This is the case, e.g., if we set δ to $\xi = 1 + \max\left\{\alpha_i + \left(\sum_{j \in N(i)} \beta_{ij}\right) - \gamma_i \,\middle|\, i \in V\right\}$. However, it seems intuitive that in many cases, much lower values δ should suffice; the minimum value of δ should depend on $\mathcal{G}$ and $\boldsymbol{z}$. The main result of this section is that we can efficiently compute the minimum value of δ applying the 2-Phase algorithm described in Section 4.

Theorem 6.1. *There exists a polynomial time algorithm that, given a median opinion game $\mathcal{G} = (G, \boldsymbol{s}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ and an initial profile $\boldsymbol{z}$, finds the minimal integer value $\delta \in \mathbb{Z}_{\geq 0}$ such that the global median of $\boldsymbol{z}$ is stabilized when all the median weights γ_i are increased by δ.*

We prove Theorem 6.1 in several steps, taking a closer look at the two phases of (the two variants of) the 2-Phase algorithm. Let us first introduce some notation. Let $\acute{\boldsymbol{z}}$ and $\hat{\boldsymbol{z}}$ denote the strategy profiles that are obtained from $\boldsymbol{z}$ after phase 1 and after phase 2 of the ascending 2-Phase algorithm, respectively. Note that Lemma 4.2 and its proof imply that both $\acute{\boldsymbol{z}}$ and $\hat{\boldsymbol{z}}$ are well-defined and independent of tie-breaking in the ordering of the updating agents in the two phases. We define $\grave{\boldsymbol{z}}$ and $\check{\boldsymbol{z}}$ accordingly with regard to the descending 2-Phase algorithm. Furthermore, we will consider these four profiles with regard to different values δ that are added to the median weight and write $\acute{\boldsymbol{z}}(\delta)$, $\hat{\boldsymbol{z}}(\delta)$, $\grave{\boldsymbol{z}}(\delta)$ and $\check{\boldsymbol{z}}(\delta)$, respectively. The proof of Theorem 6.1 relies upon the following two lemmas.

Lemma 6.2. *For $\delta \in [0, \xi] \cap \mathbb{Z}$, the value $\mathrm{med}(\acute{\boldsymbol{z}}(\delta))$ is monotonically non-increasing in δ. In contrast, the value $\mathrm{med}(\grave{\boldsymbol{z}}(\delta))$ is monotonically non-decreasing in δ.*

Lemma 6.3. *Let $\lambda, \lambda', \rho, \rho' \in [0, \xi] \cap \mathbb{Z}$ with $\lambda < \rho$ and $\lambda' < \rho'$ such that $\mathrm{med}(\acute{\boldsymbol{z}}(\lambda)) = \mathrm{med}(\acute{\boldsymbol{z}}(\rho))$ and $\mathrm{med}(\grave{\boldsymbol{z}}(\lambda')) = \mathrm{med}(\grave{\boldsymbol{z}}(\rho'))$.*

- *For $\delta \in [\lambda, \rho] \cap \mathbb{Z}$ the median $\mathrm{med}(\hat{\boldsymbol{z}}(\delta))$ is monotonically non-decreasing in δ, and*
- *for $\delta \in [\lambda', \rho'] \cap \mathbb{Z}$ the median $\mathrm{med}(\check{\boldsymbol{z}}(\delta))$ is monotonically non-increasing in δ.*

With these lemmas we can show the main theorem of this section.

Proof of Theorem 6.1. Recall that the *ascending* 2-Phase algorithm computes the unique maximum reachable equilibrium (see Theorem 4.3). In this algorithm each activated agent i updates to $\mathrm{best}_i^+(\boldsymbol{z})$, the latter is either one of the private opinions from $\mathcal{G}$, or an opinion from the initial strategy profile $\boldsymbol{z}$ (see Observation 3.4). Hence, the median can adopt at most $2n$ different values during the whole run of the algorithm. This observation combined with Lemma 6.2 implies that $\mathrm{med}(\acute{\boldsymbol{z}}(\delta))$ is a monotonically non-increasing step function that can have only $\mathrm{O}(n)$ many steps. Moreover, since δ can only take on integer values, we can find the exact step borders of the step function using $\mathrm{O}(n)$ rounds of binary search over the value of δ. These step borders partition $[0, \xi] \cap \mathbb{Z}$ into $\mathrm{O}(n)$ many non-overlapping, discrete intervals such that $\mathrm{med}(\acute{\boldsymbol{z}}(\delta))$ is invariable for each such interval $[\lambda, \rho] \cap \mathbb{Z}$. Following this, we can employ the same approach for each such interval utilizing Lemma 6.3 to derive a partition of $[0, \xi] \cap \mathbb{Z}$ into $\mathrm{O}(n^2)$ many non-overlapping, discrete intervals such that $\mathrm{med}(\hat{\boldsymbol{z}}(\delta))$ is invariable for each such interval.

An analogous approach can be used for the *descending* 2-Phase algorithm. This shows that $[0, \xi] \cap \mathbb{Z}$ can be partitioned into $\mathrm{O}(n^2)$ many non-overlapping, discrete intervals such that $\mathrm{med}(\check{\boldsymbol{z}}(\delta))$ is invariable for each such interval.

At this point we have $\mathrm{O}(n^2)$ many interval border values (from the ascending and descending 2-Phase algorithm). We can consider these values in increasing order. For each border value δ, at least one of the values $\mathrm{med}(\hat{\boldsymbol{z}}(\bar{\delta}))$ or $\mathrm{med}(\check{\boldsymbol{z}}(\bar{\delta}))$ changes. We are interested in the minimum value $\bar{\delta}$ with $\mathrm{med}(\hat{\boldsymbol{z}}(\bar{\delta})) = \mathrm{med}(\check{\boldsymbol{z}}(\bar{\delta}))$. This value $\bar{\delta}$ is indeed the minimum non-negative integer value such that each reachable equilibrium has the same median if all the median weights γ_i are increased by $\bar{\delta}$ (see Theorem 4.3). □

It remains to prove Lemmas 6.2 and 6.3. For both lemmas, we will only prove the statements for the ascending 2-Phase algorithm, i.e., regarding $\acute{\boldsymbol{z}}$ and $\hat{\boldsymbol{z}}$. The statements for the descending 2-Phase algorithm work analogously. To show the lemmas, we consider two values $\delta_1, \delta_2 \in [0, \xi] \cap \mathbb{Z}$ with $\delta_1 < \delta_2$. To distinguish the runs of the 2-Phase algorithm w.r.t. δ_1 and δ_2, we will talk about process 1 and process 2, respectively. We abbreviate $\boldsymbol{z}^{(\ell)} = \boldsymbol{z}(\delta_\ell)$ and $\mu^{(\ell)} = \mathrm{med}(\boldsymbol{z}(\delta_\ell))$ for $\ell = 1, 2$. Moreover, for any opinion x and comparative operator $\gtreqless \in \{<, >, \leq, \geq, =\}$, we set $V^{(\ell)}_{\gtreqless x} = \{i \in V \mid z_i^{(\ell)} \gtreqless x\}$. This notation is further extended in an obvious fashion, e.g., $\hat{\mu}^{(\ell)} = \mathrm{med}(\hat{\boldsymbol{z}}^{(\ell)})$. Finally, we will examine combined states in the progress of both processes in more detail. Formally, such a (combined) state s is a pair $(s^{(1)}, s^{(2)})$, where $s^{(\ell)}$ denotes the state of process $\ell \in \{1, 2\}$. We will write $\boldsymbol{z}^{(\ell,s)}$ to denote the strategy profile of process ℓ in state s and extend this notation in an obvious fashion as above.

Proof of Lemma 6.2. Recall that $\acute{\mu}^{(2)}$ is defined as the median that is reached after the first phase of process 2. For slightly simpler notation in the following, we set $\bar{\mu} = \acute{\mu}^{(2)}$.

We define the combined state $s = (s^{(1)}, s^{(2)})$ as follows. State $s^{(1)}$ denotes the first time in phase 1 of process 1 in which no agent remains that can deviate to an opinion greater than or equal to $\bar{\mu}$. State $s^{(2)}$, on the other hand, denotes the first time in phase 2 of process 2 that the median takes the value $\bar{\mu}$. We claim that $V^{(1,s)}_{\geq\bar{\mu}} \supseteq V^{(2,s)}_{\geq\bar{\mu}}$. Note that by definition, we have $\mu^{(2,s)} = \bar{\mu}$. Hence, the claim implies $\mu^{(1,s)} \geq \bar{\mu}$ and therefore shows the lemma.

Assume that the claim is not true. Since the two processes start with the same strategy profile, this implies that there has to be a first agent i^* in process 2 with $i^* \in V^{(2,s)}_{\geq\bar{\mu}} \setminus V^{(1,s)}_{\geq\bar{\mu}}$. We consider all the influences supporting the deviation of i^* in process 2. The choice of i^* implies that any neighbor i' of i^* that supports the deviation of i^* to at least $\bar{\mu}$ in process 2, also does so in process 1 in state s. Furthermore, the definition of state $s^{(2)}$ implies that the median works against the deviation of i^* in process 2. This may or may not also be the case regarding process 1 and state s but in any case the median has a smaller influence in process 1 since $\delta_1 < \delta_2$. Lastly, the private opinion of i^* obviously has the same influence in both processes. Hence, i^* has an incentive to deviate to at least $\bar{\mu}$ in process 1 in state s, but no agent in process 1 has such an incentive in state s – a contradiction. □

In the proof of Lemma 6.3 we are interested in the case that the two processes reach the same median at the end of phase 1. Hence, we assume $\acute{\mu}^{(1)} = \acute{\mu}^{(2)}$ in the following analysis and again denote this value as $\bar{\mu}$. We can use a similar argument as we did above, but first have to take a closer look at phase 1 of the two processes. Intuitively, the fact that $\delta_1 < \delta_2$ should lead to the opinions being clustered more closely around $\bar{\mu}$ for $\acute{z}^{(2)}$ than for $\acute{z}^{(1)}$. This is indeed utilized in the following.

Proof of Lemma 6.3. Recall that $\hat{\mu}^{(2)}$ is defined as the median that is reached after phase 2 of process 2. Both processes have the median value $\bar{\mu}$ at the end of phase 1 and the median can only decrease in phase 2. Hence, there is nothing further to show if $\hat{\mu}^{(2)} = \bar{\mu}$ and we assume $\hat{\mu}^{(2)} < \bar{\mu}$ in the following.

We consider the combined state a, where state $a^{(\ell)}$ of process $\ell \in \{1, 2\}$ denotes the first time in phase 1 in which no agent remains that can deviate to an opinion larger or equal to $\bar{\mu}$. We claim that $V^{(1,a)}_{\geq\bar{\mu}} \subseteq V^{(2,a)}_{\geq\bar{\mu}}$.

Assume for the sake of contradiction that this is not true. Let $i^* \in V^{(1,a)}_{\geq\bar{\mu}} \setminus V^{(2,a)}_{\geq\bar{\mu}}$ and let i^* be the first such agent in process 1. The choice of i^* implies that any neighbor i' of i^* that supports the deviation of i^* to at least $\bar{\mu}$ in process 1, also does so in process 2 in state a. Moreover, the median may or may not support the deviation of i^* in process 1. But in process 2 the median $\bar{\mu}$ is reached in state a, and hence the median does support a deviation to at least $\bar{\mu}$ and with a stronger influence since $\delta_2 > \delta_1$. Hence, i^* has an incentive to deviate to at least $\bar{\mu}$ in process 2 in state a, but no agent has such an incentive in state a – a contradiction.

We continue with the combined state b in which both processes finished phase 1 of the 2-Phase algorithm. We claim $V^{(1,b)}_{<\bar{\mu}} \supseteq V^{(2,b)}_{<\bar{\mu}}$ and $z_i^{(1,b)} \leq z_i^{(2,b)}$ for each $i \in V^{(1,b)}_{<\bar{\mu}}$. Now, the first statement of the claim is directly implied by the claim for state a, and we assume, for the sake of contradiction, that the second statement is not true. For $i \in V^{(1,b)}_{<\bar{\mu}} \setminus V^{(2,b)}_{<\bar{\mu}}$, we already know that i deviates to a value smaller than $\bar{\mu}$ in phase 1 of process 1 but at least $\bar{\mu}$ in phase 1 of process 2. Hence, there has to be a first agent i^* in process 1 with $i^* \in V^{(1,b)}_{<\bar{\mu}} \cap V^{(2,b)}_{<\bar{\mu}}$ and $z_{i^*}^{(1,b)} > z_{i^*}^{(2,b)}$. Since the final median $\bar{\mu}$ for both processes in phase 1 is already reached in state a, the median supports a deviation of i^* to at least $z_{i^*}^{(1,b)}$ in process 2 in state b. Furthermore, this support is stronger than the one for the deviation of i^* in process 1 because $\delta_1 < \delta_2$. Next, we consider the neighbors of i^* that support their deviation in process 1. Any such neighbor i' also supports a deviation of i^* to $z_{i^*}^{(1,b)}$ in process 2 in state b. To see this, note that i' either did not deviate in phase 1 of process 1 before i^* or it did to either an opinion of at least $\bar{\mu}$ or at most $\bar{\mu}$. The support for a deviation of i^* to $z_{i^*}^{(1,b)}$ in process 2 in state b is guaranteed in the first case, since the two processes have the same initial strategy profile; in the second, due to the claim for state a; and in the third case, because of the choice of i^*. Hence, i^* has an incentive to deviate to a higher opinion in state b in process 2, but no agent has an incentive to do so in state b – a contradiction.

Finally, we define the combined state c as follows. State $c^{(1)}$ of process 1 denotes the first time in phase 2 in which no agent remains that can deviate to an opinion smaller or equal to $\hat{\mu}^{(2)}$. State $c^{(2)}$ of process 2, on the other hand, denotes the first time in phase 2 that the median takes the value $\hat{\mu}^{(2)}$. For clarity of notation, we set $\dot{\mu} = \hat{\mu}^{(2)}$. We claim that $V^{(1,c)}_{\leq\dot{\mu}} \supseteq V^{(2,c)}_{\leq\dot{\mu}}$. Note that by definition, we have $\mu^{(2,c)} = \dot{\mu}$. Hence, the claim implies $\mu^{(1,c)} \leq \dot{\mu}$ and therefore suffices to show the lemma.

Assume that the claim is not true. Remember that we have $\dot{\mu} < \bar{\mu}$, as discussed at the start of the proof. Moreover, regarding state b, we did show $V^{(1,b)}_{<\bar{\mu}} \supseteq V^{(2,b)}_{<\bar{\mu}}$ and $z_i^{(1,b)} \leq z_i^{(2,b)}$ for each $i \in V^{(1,b)}_{<\bar{\mu}}$. This implies that the corresponding claim is true at the end of phase 1, i.e., for the combined state b. Hence, there has to be a first agent i^* in process 2 with $i^* \in V^{(2,c)}_{\leq\dot{\mu}} \setminus V^{(1,c)}_{\leq\dot{\mu}}$. The definition of state $c^{(2)}$ implies that the median does not support the deviation of i^* in process 2. Moreover, a neighbor i' supporting the deviation of i^* to a value of at most $\dot{\mu}$ in process 2 also does so in process 1 in state c. To see this, note that i' either did deviate in phase 2 of process 2 prior to i^* or it did not. In the first case, the choice of i^* implies $i' \in V^{(1,c)}_{\leq\dot{\mu}}$ and in the second i' has opinion $z_{i'}^{(2,b)} \geq z_{i'}^{(1,b)}$ when i^* deviates in process 2. Hence, we have a contradiction. □

The proof technique of Theorem 6.1 can be used to show further results. For example, Lemma 6.2 and Lemma 6.3 work as well if the median weight is only increased for a *given subset* V' of agents. However, in this case, it can be impossible to stabilize the median. Rather, we aim to obtain the smallest possible difference between the minimum and maximum median of reachable equilibria. Let this value be $\tau_{V'}$. For a collection $\mathcal{V} \subseteq 2^V$ of subset of agents, we may be interested in the smallest $\tau_{\min} = \min\{\tau_{V'} \mid V' \in \mathcal{V}\}$. Moreover, from the subsets that guarantee $\tau_{\min}$, we want to choose one that achieves this optimal difference using the smallest (integer) increase in median weight, i.e., $\delta_{\min} = \min\{\delta_{V'} \mid \tau_{V'} = \tau_{\min}, V' \in \mathcal{V}\}$. We say that a pair $(V', \delta_{V'})$ with $\delta_{V'} = \delta_{\min}$ has an *optimal median gap with smallest increase* in $\mathcal{V}$. It is easy to show the following result:

Corollary 6.4. *Let $k \in \mathbb{Z}_{\geq 0}$ be a constant and $\mathcal{V} = \{S \subseteq V \mid |S| \leq k\}$. There is a polynomial-time algorithm to find a pair $(V', \delta_{V'})$ that has an optimal median gap with smallest increase in $\mathcal{V}$.*

Proof. First note that increasing any median weight by more than ξ will lead to the same result as increasing it by ξ. Hence, we can simply enumerate all $O(|V|^k)$ many possibilities for V'. By increasing the weight by ξ for each, we determine $\tau_{\min}$. Then, by restricting attention to sets V' with $\tau_{\min}$, we can use the approach from the proof of Theorem 6.1 to find $\delta_{V'}$ for each of them. In this way, we obtain a pair $(V', \delta_{V'})$ that has optimal median gap with smallest increase. □

This polynomial-time enumeration can be applied even when we have further restrictions on $\mathcal{V}$. One might wonder about the complexity of the problem if k is not part of the input (i.e., not necessarily constant). A special case of the problem with arbitrary k is studied in more detail in the next section.

7 HARDNESS RESULTS

In this section, we will first prove the following theorem, and then we discuss a variant with unit weights. Remember that we call a median opinion game $\mathcal{G}$ and an initial profile $\boldsymbol{z}$, *stable* (with respect to the median of reachable equilibria) if all equilibria reachable from $\boldsymbol{z}$ share the same global median. We define the decision problem StableMedian$(\mathcal{G}, \boldsymbol{z})$ as the problem to decide if the global median of the initial profile $\boldsymbol{z}$ in the median opinion game $\mathcal{G}$ can be stabilized by increasing the influence of the weight of the aggregation function γ_i for k agents $i \in V$ by any value $\delta \geq 0$.

Theorem 7.1. *StableMedian is NP-complete.*

Proof. Considering Theorem 4.3, we observe that StableMedian is indeed in NP: for a given guess of k agents, the problem can be decided by increasing their median weights by ξ and running the two variants of the 2-Phase algorithm (as was done in the proof of Corollary 6.4).

We show NP-hardness starting from the classical vertex cover problem. In this problem, a graph $G = (V, E)$ and an integer $k \in \{1, \ldots, |V|\}$ are given, and the goal is to decide whether a selection $C \subseteq V$ of (at most) k vertices exist such that each edge is connected to at least one vertex in C. We will assume w.l.o.g. that $|E| > k$.

The reduction uses a construction that can be divided into three gadgets. There is a vertex gadget, an edge gadget, and a pivot gadget. There are only two opinions: 0 and 1, and 1 is the initial median opinion. In the vertex gadget, there are vertex agents with public opinion 1 who may change their opinion. The edge gadget contains edge agents again with initial opinion 1 who are connected to their respective vertex agents and may change their opinion if both neighboring vertex agents do so as well. Finally, the pivot gadget contains many agents with public opinion 1 who may deviate to 0 as soon as one of the edge agents does. This can be realized in a way that the existence of a suitable vertex cover guarantees a suitable selection of (vertex) agents and vice versa. We proceed

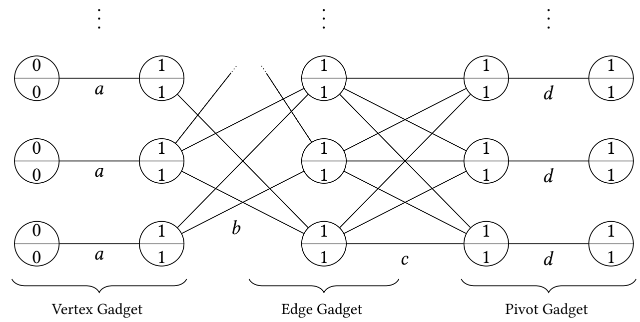

Figure 2: The gadgets of the reduction used in the proof of Theorem 7.1. The labels of the agents are the initial opinions. The vertex gadget includes $|V|$ many pairs of agents, the edge gadget $|E|$ many agents, and the pivot gadget X many pairs of agents. The Y many dummy agents are not shown.

with a detailed description of the gadgets. The median weight of each agent is 1, and the weights of the intrinsic opinions are 0. The construction uses parameters a, b, c, d, X, Y with $a = \Delta b + 2$ with Δ the maximum degree in G, $b = (Xc + 2)/2$, $c = 1$, $d = |E|c$, $X = 2|V| + 2|E|$, and $Y = 2|V|$. See Figure 2 for an overview.

- *Vertex gadget:* For each vertex $v \in V$, there is a *vertex agent* $\mathsf{VAgt}(v)$ with initial opinion 1 and a *vertex flip agent* $\mathsf{VFAgt}(v)$ with initial opinion 0. The agents $\mathsf{VAgt}(v)$ and $\mathsf{VFAgt}(v)$ are connected with weight a.
- *Edge gadget:* For each edge $e \in E$, there is an *edge agent* $\mathsf{EAgt}(e)$ with initial opinion 1. $\mathsf{EAgt}(\{u, v\})$ is connected to the vertex agents $\mathsf{VAgt}(u)$ and $\mathsf{VAgt}(v)$ with edges of weight b.
- *Pivot gadget:* For each $i \in [X]$ there is one *pivot agent* $\mathsf{PAgt}(i)$ with initial opinion 1 and one *pivot balance agent* $\mathsf{PBAgt}(i)$ with initial opinion 0. Each pivot agent is connected to each edge agent with weight c. Moreover, $\mathsf{PAgt}(i)$ is connected to $\mathsf{PBAgt}(i)$ with weight d.
- *Dummy agents:* For each $i \in [Y]$ there is one *dummy agent* $\mathsf{DAgt}(i)$ with initial opinion 1. Each dummy agent is connected to each other dummy agent with weight 1, i.e., the dummy agents form a clique, and the dummy agents are not connected to any other agent.

Given this construction, one can verify that the median can be stabilized by increasing the median weight of k agents if and only if there is a k vertex cover in G. □

Our final corollary shows that the large edge weights used in the reduction of Theorem 7.1 are not necessary.

Corollary 7.2. *StableMedian is NP-complete even if $\boldsymbol{\beta} = \mathbf{1}$.*

ACKNOWLEDGMENTS

Petra Berenbrink, Martin Hoefer, Malin Rau, and Lisa Wilhelmi gratefully acknowledge support by DFG Research Unit ADYN under grant DFG 411362735.

REFERENCES

[1] R. Abelson and A. Bernstein. 1963. A computer simulation model of community referendum controversies. *Public Opinion Quarterly*, 27, 1, 93–122.

[2] A. Anagnostopoulos, L. Becchetti, E. Cruciani, F. Pasquale, and S. Rizzo. 2020. Biased opinion dynamics: when the devil is in the details. In *Proc. 29th Int. Joint Conf. Artif. Intell. (IJCAI)*, 53–59.

[3] V. Auletta, A. Fanelli, and D. Ferraioli. 2019. Consensus in opinion formation processes in fully evolving environments. In *Proc. 33rd Conf. Artif. Intell. (AAAI)*, 6022–6029.

[4] P. Berenbrink, M. Hoefer, D. Kaaser, P. Lenzner, M. Rau, and D. Schmand. 2024. Asynchronous opinion dynamics in social networks. *Distributed Comput.*, 37, 3, 207–224.

[5] A. Bhattacharyya, M. Braverman, B. Chazelle, and H. L. Nguyen. 2013. On the convergence of the Hegselmann-Krause system. In *Proc. Symp. Innov. Theoret. Comput. Sci. (ITCS)*, 61–66.

[6] A. Bhattacharyya and K. Shiragur. 2015. How friends and non-determinism affect opinion dynamics. In *Proc. 54th IEEE Conf. Decision and Control (CDC)*, 6466–6471.

[7] K. Bhawalkar, S. Gollapudi, and K. Munagala. 2013. Coevolutionary opinion formation games. In *Proc. Symp. Theory Comput. (STOC)*, 41–50.

[8] V. Bilò, A. Fanelli, and L. Moscardelli. 2018. Opinion formation games with dynamic social influences. *Theor. Comput. Sci.*, 746, 73–87.

[9] D. Bindel, J. M. Kleinberg, and S. Oren. 2015. How bad is forming your own opinion? *Games Econ. Behav.*, 92, 248–265.

[10] S. Botan, U. Grandi, and L. Perrussel. 2019. Multi-issue opinion diffusion under constraints. In *Proc. 18th Conf. Auton. Agents and Multi-Agent Syst. (AAMAS)*, 828–836.

[11] R. Bredereck and E. Elkind. 2017. Manipulating opinion diffusion in social networks. In *Proc. 26th Int. Joint Conf. Artif. Intell. (IJCAI)*, 894–900.

[12] R. Bredereck, A. George, J. Israel, and L. Kellerhals. 2022. Single-peaked opinion updates. In *Proc. 31st Int. Joint Conf. Artif. Intell. (IJCAI)*, 137–143.

[13] R. Bredereck, L. Jacobs, and L. Kellerhals. 2020. Maximizing the spread of an opinion in few steps: opinion diffusion in non-binary networks. In *Proc. 29th Int. Joint Conf. Artif. Intell. (IJCAI)*, 1622–1628.

[14] C. Castellano, S. Fortunato, and V. Loreto. 2009. Statistical physics of social dynamics. *Reviews of Modern Physics*, 81, 2, 591–646.

[15] B. Chazelle. 2011. The total s-energy of a multiagent system. *SIAM J. Control. Optim.*, 49, 4, 1680–1706.

[16] S. Ching. 1997. Strategy-proofness and "median voters". *Int. J. Game Theory*, 26, 473–490.

[17] M. Cinelli, G. D. F. Morales, A. Galeazzi, W. Quattrociocchi, and M. Starnini. 2021. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118, 9.

[18] A. Coates, L. Han, and A. Kleerekoper. 2018. A unified framework for opinion dynamics. In *Proc. 17th Conf. Auton. Agents and Multi-Agent Syst. (AAMAS)*, 1079–1086.

[19] A. Coates, L. Han, and A. Kleerekoper. 2018. A unified opinion framework simulator. In *Proc. 17th Conf. Auton. Agents and Multi-Agent Syst. (AAMAS)*, 1803–1805.

[20] A. De, S. Bhattacharya, and N. Ganguly. 2018. Shaping opinion dynamics in social networks. In *Proc. 17th Conf. Auton. Agents and Multi-Agent Syst. (AAMAS)*, 1336–1344.

[21] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch. 2000. Mixing beliefs among interacting agents. *Adv. Complex Syst.*, 3, 1-4, 87–98.

[22] M. H. DeGroot. 1974. Reaching a consensus. *J. Amer. Stat. Assoc.*, 69, 345, 118–121.

[23] M. Del Vicario, A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, H. E. Stanley, and W. Quattrociocchi. 2016. The spreading of misinformation online. *Proc. Natl. Acad. Sci.*, 113, 3, 554–559.

[24] A. Downs. 1957. *An economic theory of democracy*. Harper & Row New York.

[25] F. Echenique. 2005. A short and constructive proof of Tarski's fixed-point theorem. 33, 2, 215–218.

[26] M. Epitropou, D. Fotakis, M. Hoefer, and S. Skoulakis. 2019. Opinion formation games with aggregation and negative influence. *Theory Comput. Syst.*, 63, 7, 1531–1553.

[27] J. M. Epstein. 2012. *Generative Social Science: Studies in Agent-Based Computational Modeling*. Princeton University Press.

[28] J. M. Epstein and R. Axtell. 1996. *Growing Artificial Societies: Social Science from the Bottom Up*. Brookings Institution Press.

[29] S. R. Etesami and T. Başar. 2015. Game-theoretic analysis of the hegselmann-krause model for opinion dynamics in finite dimensions. *IEEE Trans. Autom. Control.*, 60, 7, 1886–1897.

[30] P. Faliszewski, R. Gonen, M. Koutecký, and N. Talmon. 2018. Opinion diffusion and campaigning on society graphs. In *Proc. 27th Int. Joint Conf. Artif. Intell. (IJCAI)*, 219–225.

[31] A. Fanelli and D. Fotakis. 2020. Local aggregation in preference games. arXiv: 2002.01251 [cs.GT].

[32] S. Fortunato. 2005. On the consensus threshold for the opinion dynamics of krause–hegselmann. *Int. J. Modern Phys. C*, 16, 02, 259–270.

[33] D. Fotakis, D. Palyvos-Giannas, and S. Skoulakis. 2016. Opinion dynamics with local interactions. In *Proc. 25th Int. Joint Conf. Artif. Intell. (IJCAI)*, 279–285.

[34] N. E. Friedkin and E. C. Johnsen. 1990. Social influence and opinions. *J. Math. Soc.*, 15, 3-4, 193–206.

[35] M. Granovetter. 1978. Threshold models of collective behavior. *Amer. J. Soc.*, 83, 6, 1420–1443.

[36] R. Hegselmann and U. Krause. 2002. Opinion dynamics and bounded confidence models, analysis, and simulation. *J. Artif. Societ. Soc. Simul.*, 5, 3.

[37] R. A. Holley and T. M. Liggett. 1975. Ergodic theorems for weakly interacting infinite systems and the voter model. *Ann. Prob.*, 643–663.

[38] H. Johnston, (Ed.) 1995. *Social Movements and Culture*. Routledge.

[39] A. Martinsson. 2016. An improved energy argument for the hegselmann–krause model. *J. Diff. Equat. Appl.*, 22, 4, 513–518.

[40] J. H. Miller and S. E. Page. 2009. *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton University Press.

[41] H. Moulin. 1980. On strategy-proofness and single peakedness. *Public Choice*, 35, 4, 437–455.

[42] M. Moussaïd, J. E. Kämmer, P. P. Analytis, and H. Neth. 2013. Social influence and the collective dynamics of opinion formation. *PLoS One*, 8, 11, 1–8.

[43] R. Parasnis, M. Franceschetti, and B. Touri. 2019. On the convergence properties of social hegselmann-krause dynamics. arXiv: 1909.03485 [math.OC].

[44] E. Pariser. 2011. *The Filter Bubble: What the Internet Is Hiding from You*. Penguin UK.

[45] R. E. Smithson. 1971. Fixed points of order preserving multifunctions. *Proc. Amer. Math. Soc.*, 28, 304–310.

[46] B. Touri and A. Nedic. 2011. Discrete-time opinion dynamics. In *Conference Record of the 45th Asilomar Conference on Signals, Systems and Computers (ACSCC)*, 1172–1176.

[47] J. Van Ginneken. 2003. *Collective Behavior and Public Opinion: Rapid Shifts in Opinion and Communication*. Routledge.

[48] D. J. Watts and S. H. Strogatz. 1998. Collective dynamics of 'small-world' networks. *Nature*, 393, 6684, 440–442.

[49] E. Wedin and P. Hegarty. 2015. A quadratic lower bound for the convergence rate in the one-dimensional Hegselmann-Krause bounded confidence dynamics. *Discret. Comput. Geom.*, 53, 2, 478–486.

[50] M. Zhan, H. Liang, G. Kou, Y. Dong, and S. Yu. 2019. Impact of social network structures on uncertain opinion formation. *IEEE Trans. Comput. Soc. Syst.*, 6, 4, 670–679.

[51] L. Zhou. 1994. The set of nash equilibria of a supermodular game is a complete lattice. 7, 2, 295–300.