Decentralized Ranking Aggregation via Gossip: Convergence and Robustness

Kerrian Le Caillec* **Anna van Elst*** **Igor Colin** **Stephan Clémençon**
LTCI, Télécom Paris, Institut Polytechnique de Paris
name.surname@telecom-paris.fr

Abstract

The concept of ranking aggregation plays a central role in preference analysis, and numerous algorithms for calculating median rankings, often originating in social choice theory, have been documented in the literature, offering theoretical guarantees in a centralized setting, *i.e.*, when all the ranking data to be aggregated can be brought together in a single computing unit. For many technologies (*e.g.*, peer-to-peer networks, IoT, multi-agent systems), extending the ability to calculate consensus rankings with guarantees of convergence and resilience to potential contamination in a decentralized setting, when preference data is initially distributed across a communicating network, remains a major methodological challenge. Indeed, in recent years, the literature on decentralized computation has mainly focused on computing or optimizing statistics such as arithmetic means using gossip algorithms. The purpose of this article is precisely to study how to achieve reliable and resilient consensus on collective rankings in a decentralized setting, thereby raising new questions, such as robustness to corrupted nodes and scalability through reduced communication costs. The approach proposed and analyzed here relies on randomized gossip algorithms, which allow autonomous agents to compute a global ranking consensus using local communications only, without coordination or a central authority.

1 Introduction

Ranking aggregation aims to summarize several rankings relating to a common set of alternatives, obtained from different sources, into a single representative or *median ranking*. This problem, also known as *consensus ranking*, has its origin in social choice theory [5, 16, 17, 41] and voting systems [7]. More recently, rank aggregation has emerged as a core primitive in machine learning. In learning to rank [45], the goal is to train models from ranked supervision signals arising from multiple annotators or queries, and aggregating those signals consistently is a rank aggregation problem. In preference learning [26], models learn utility functions from pairwise comparisons, and consolidating preferences across distributed data sources requires principled aggregation. Most prominently, Reinforcement Learning from Human Feedback [14, 67], especially in Natural Language Processing tasks, relies on aggregating ranked preferences from multiple human annotators to shape reward models, yet standard pipelines collapse this aggregation into a single centralized step, ignoring the social choice structure of the problem and its vulnerabilities [15, 58]. Other applications include crowdsourcing platforms [11, 12], and recommender systems [3]. While the theoretical and empirical properties of centralized rank aggregation methods are now well established [22, 43], *decentralized* ranking aggregation has received much less attention in the literature.

To our knowledge, no prior work establishes convergence guarantees for fully decentralized rank aggregation; in particular, federated settings [52, 56] require a central aggregator. Indeed, existing

*Main contributors.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

work has mainly focused on configurations which are inherently centralized: they require a single trusted entity to collect and process all individual preferences, introducing a single point of failure and communication bottlenecks. In this paper, we consider a fully decentralized framework in which voters are distributed across a communication network, have limited computing resources, and can only communicate with their immediate neighbors, with no access to a central server [6, 42]. This model is standard in *e.g.*, peer-to-peer systems [2], and in wireless sensor networks and IoT networks [55]. Our goal is to study how a network of autonomous agents, each with its own preference ranking, can reach a consensus ranking without a central authority. Unlike real-valued quantities, rankings are elements of the symmetric group $\mathfrak{S}_n$, and the coordinate-wise arithmetic mean does not preserve their permutation structure. In this paper, we study a general class of score-based ranking consensus methods, subsuming many classical voting rules such as Borda [5] and Copeland [17] as canonical instances [8]. Our approach applies a single decentralized gossip algorithm [6] for score estimation across any rule in this class, followed by a local ranking projection onto the symmetric group $\mathfrak{S}_n$, yielding a unified and communication-efficient framework for decentralized consensus. A key consideration is robustness against corrupt nodes and vote manipulation, a topic explored through the notions of spam resistance [22] and breakdown points [31].

Contributions. Our main contributions are as follows. *(i)* We propose a unified gossip algorithm for decentralized score-based consensus applicable to any voting rule following a specific decomposition. We show that the average Kendall-τ error decays exponentially fast under any locally ranking-stable aggregation map, at a rate depending on a stability radius. *(ii)* We provide an empirical robustness analysis via the breakdown function and then we relate contamination levels to the gossip process. *(iii)* We present experiments on various network topologies and ranking datasets, showing how Borda and Copeland consensus exhibit robustness under data contamination (see Figure 1 for illustration of the protocols).

Other related work. Data ranking over gossip aggregation has recently attracted attention from a statistical perspective [61, 62], establishing connections between gossip-based averaging and robust estimation for rankings and trimmed means. Byzantine robustness in decentralized learning has been studied extensively [27, 34], with particular attention to characterizing the influence of corrupted nodes on convergence; the dual-approach framework in [28] provides tight clipping-based guarantees for average consensus that complement our breakdown function analysis in Section 4. Jiao et al. [39] and Jiao and Vert [38] approximate the Kemeny consensus using an embedding for permutations based on the usual voting rules.

2 Background and preliminaries

We briefly review the fundamental concepts of consensus ranking theory and gossip-based decentralized computing. Throughout this article, preferences over n items indexed by $i \in [\![n]\!] := \{1, \dots, n\}$ are represented as permutations $\sigma \in \mathfrak{S}_n$, where $\mathfrak{S}_n$ is the symmetric group and $\sigma(i)$ denotes the rank assigned to item i. We write $\mathbb{I}\{\mathcal{E}\}$ for the indicator function of any event $\mathcal{E}$. We denote scalars by lowercase letters $x \in \mathbb{R}$, vectors by boldface lowercase letters $\mathbf{x} \in \mathbb{R}^n$, and matrices by boldface uppercase letters $\mathbf{X} \in \mathbb{R}^{n \times m}$. For $d \in \mathbb{N}$, we write $\{\mathbf{e}_k \mid k \in [\![d]\!]\}$ for the canonical basis of $\mathbb{R}^d$, $\mathbf{M}^\top$ for the transpose of any matrix $\mathbf{M}$, $|F|$ for the cardinality of any finite set F, $\mathbf{I}_d$ for the identity matrix in $\mathbb{R}^{d \times d}$, $\mathbf{1}_d$ for the vector in $\mathbb{R}^d$ whose coordinates are all equal to 1, and $\|\cdot\|$ for the usual Euclidean (ℓ^2-)norm. For any probability distribution P on $\mathfrak{S}_n$ and any map $f : \mathfrak{S}_n \to \mathcal{X}$, where $\mathcal{X}$ is a Hilbert space, we write $\mathbb{E}_P f$ for $\mathbb{E}_{\Sigma \sim P} f(\Sigma)$ when there is no ambiguity.

Ranking aggregation/consensus. Let P be a probability distribution over $\mathfrak{S}_n$, written as $P = \sum_{\sigma \in \mathfrak{S}_n} P(\sigma)\, \delta_\sigma$, where δ_σ denotes the Dirac measure at σ. We denote by $\Delta_{\mathfrak{S}_n}$ the probability simplex over $\mathfrak{S}_n$. Given a measurable map $g : \mathfrak{S}_n \times \Delta_{\mathfrak{S}_n} \to \mathbb{R}$, the ranking aggregation problem [43] consists in finding a consensus permutation $\sigma_\star \in \mathfrak{S}_n$ such that

$$\sigma_\star \in \arg\min_{\sigma \in \mathfrak{S}_n} g(\sigma, P).$$

Let $(\sigma_v)_{v=1}^N$ be a dataset of N permutations in $\mathfrak{S}_n$ such that $N \geq n$ (in practical cases, $n \ll N$), representing one ranking per voter, with repetitions allowed. The underlying distribution P is then approximated by the empirical measure $\widehat{P}_N = N^{-1} \sum_{v=1}^N \delta_{\sigma_v}$. Since the order of the voters is

irrelevant, a dataset is a multiset of N permutations, and we denote by $\mathrm{Sym}^N(\mathfrak{S}_n)$ the set of all such multisets, and by $\mathcal{P}_N(\mathfrak{S}_n)$ the set of the corresponding empirical distributions. This notion of consensus ranking can be interpreted in different ways. The consensus can correspond to the mode of a probability distribution on $\mathfrak{S}_n$ (typically the Mallows distribution [49], see Appendix H for details), considering the rankings σ_v as i.i.d. realizations of it. Alternatively, the consensus can be defined as a barycenter of the dataset with respect to a metric on $\mathfrak{S}_n$, an approach widely used in practice. In particular, for some (pseudo-)metric $d : \mathfrak{S}_n \times \mathfrak{S}_n \to \mathbb{R}_+$, a ranking consensus can be defined as any solution of the (discrete) optimization problem

$$\min_{\sigma \in \mathfrak{S}_n} \mathbb{E}_{\Sigma \sim P}\, d(\sigma, \Sigma). \tag{1}$$

Among the many possible choices of d (see, *e.g.*, [19]), a widely used one is the Kendall-τ distance d_τ, defined as the number of discordant pairs: for any $\sigma, \sigma' \in \mathfrak{S}_n$,

$$d_\tau(\sigma, \sigma') = \sum_{1 \le i < j \le n} \mathbb{I}\left\{(\sigma(i) - \sigma(j))\,(\sigma'(i) - \sigma'(j)) < 0\right\}.$$

When $d = d_\tau$, any solution of Equation (1) is referred to as a *Kemeny consensus* [41] and, remarkably, the two approaches, probabilistic and metric, coincide: the Kemeny consensus is the maximum likelihood estimator of the central ranking of a Mallows model, under the assumption that the σ_v are i.i.d. realizations of it (see [66]). From a social choice perspective, the Kemeny consensus satisfies the Condorcet criterion: any candidate $i \in [\![n]\!]$ who would win against all others in head-to-head majority votes is guaranteed to be ranked first (see [8]). The Kemeny consensus further satisfies the extended Condorcet property, which ensures robustness to adversarial inputs (see [22]).

However, these properties are achieved at a significant computational cost, as computing a Kemeny consensus is NP-hard in general (see, *e.g.*, [37]). This motivates score-based ranking aggregation methods. The principle is to define a score function $S : \mathcal{P}_N(\mathfrak{S}_n) \to \mathbb{R}^n$ that maps a dataset to a vector of candidate scores $\mathbf{s} = (s_1, \ldots, s_n)^\top$. The score vectors are then ranked in ascending/descending order depending on the score counting method, and this yields a unique consensus in $\mathfrak{S}_n$ when there are no ties, *i.e.,* when all score coordinates are distinct (see Appendix E for details). For a given distribution P, we also denote the *pairwise probabilities* for a pair of alternatives (i, j) as $p_{ij} = \mathbb{P}(\sigma(i) < \sigma(j))$. We now present examples of score-based voting rules.

Example 1 (Borda method, [5])**.** *Each permutation $\sigma_v \in \mathfrak{S}_n$ is embedded as its rank vector $(\sigma_v(1), \ldots, \sigma_v(n))^\top \in \mathbb{R}^n$, the Borda score function S_B returns the Borda score by averaging the N vectors and taking the opposite:*

$$S_B\left(\widehat{P}_N\right) = -\frac{1}{N}\sum_{v=1}^{N}(\sigma_v(1), \ldots, \sigma_v(n))^\top.$$

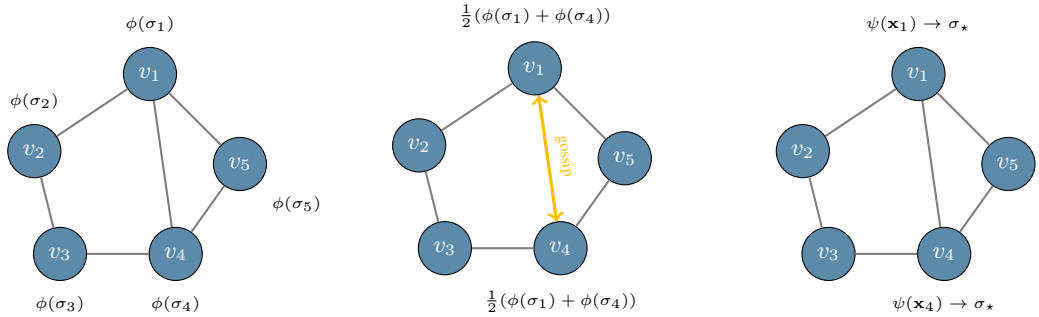
The Borda consensus $\sigma_\star^B \in \mathfrak{S}_n$ ranks candidates in decreasing order of $[S_B(\widehat{P}_N)]_i$, i.e., in ascending order of their average ranks. This method exploits the positional information in each ranking and, when Borda scores are distinct, yields the unique minimizer of the sum of Spearman-ρ distances [31] (for more details see Appendix F).

Example 2 (Copeland method, [17])**.** *Each permutation σ_v is embedded into its pairwise preference matrix, where each pair (i, j) with $i < j$ records $\mathbb{I}\{\sigma_v(i) < \sigma_v(j)\}$. Averaging over the N voters yields the pairwise preference matrix with entries $\widehat{p}_{ij} = N^{-1}\sum_{v=1}^{N}\mathbb{I}\{\sigma_v(i) < \sigma_v(j)\}$. The Copeland score function S_C returns a vector in $\mathbb{R}^n$ whose i-th coordinate is:*

$$\left[S_C\left(\widehat{P}_N\right)\right]_i = \sum_{j \neq i}\mathbb{I}\{\widehat{p}_{ij} > \tfrac{1}{2}\} - \mathbb{I}\{\widehat{p}_{ij} < \tfrac{1}{2}\}, \tag{2}$$

counting pairwise wins minus losses. The Copeland consensus $\sigma_\star^C$ ranks candidates in descending order of $[S_C(\widehat{P}_N)]_i$ (for more details, see Appendix G).

The distribution P is *weakly stochastically transitive* if $p_{ij} \ge 1/2$ and $p_{jk} \ge 1/2$ imply $p_{ik} \ge 1/2$, and *strongly stochastically transitive* if furthermore $p_{ik} \ge \max(p_{ij}, p_{jk})$, for all distinct $(i, j, k) \in [\![n]\!]^3$. Under these conditions, Borda's, Copeland's, and Kemeny's methods yield identical unique solutions (see [43]). Outside such favorable cases, Arrow's impossibility theorem establishes that no rank aggregation method can simultaneously satisfy all desirable criteria [1].



(a) At iteration $t = 0$, each node v holds a ranking $\sigma_v \in \mathfrak{S}_n$ and embeds it as $\phi(\sigma_v)$.

(b) At iteration $t = 1$, the two connected nodes update their estimates using the average.

(c) After enough iterations, each node applies ψ to its estimate and ranks the scores to obtain $\sigma_\star$.

Figure 1: Illustration of randomized (ϕ, ψ)-gossip on $N = 5$ nodes, each holding a ranking, where ϕ is an embedding function and ψ an aggregation map.

Decentralized estimation via gossip algorithms. Let $\mathcal{G} = (V, E)$ be a connected, non-bipartite graph with $V = [\![N]\!]$. Each agent $v \in V$ holds an initial local state $\mathbf{x}_v(0)$ lying in a Euclidean space, and the collective goal is to compute the global average $\bar{\mathbf{x}} = N^{-1} \sum_{v=1}^{N} \mathbf{x}_v(0)$ through purely local interactions. Two canonical protocols achieve this. In *synchronous gossip*, all nodes update simultaneously via a doubly stochastic *gossip matrix* $\mathbf{W}$ ($W_{uv} > 0$ only if $(u, v) \in E$ or $u = v$ and $\mathbf{W}\mathbf{1}_N = \mathbf{W}^\top \mathbf{1}_N = \mathbf{1}_N$), $\mathbf{X}(t+1) = \mathbf{W}\mathbf{X}(t)$ [42, 60] with $\mathbf{X}(t) := [\mathbf{x}_1(t)^\top, \dots, \mathbf{x}_N(t)^\top]^\top$, converging to $\bar{\mathbf{x}}\mathbf{1}_N^\top$. The rate of convergence depends on the spectral gap of $\mathbf{W}$ denoted $\lambda_{\mathbf{W}}$ (see Appendix A for more details). Larger spectral gaps, achieved by denser or expander-like graphs, yield faster convergence [42]. In other words, each node $v \in [\![N]\!]$ receives simultaneously the values of the nodes in its neighborhood denoted $\mathcal{N}_v$ and updates its state accordingly. In *randomized gossip* [6], a single edge $e = (u, v)$ is sampled from $p := (p_e)_{e \in E}$ at each step, and the two incident nodes average their states while all others remain unchanged. The update of state is thus given by the gossip matrix $\mathbf{W}_e = \mathbf{I}_N - (1/2)(\mathbf{e}_u - \mathbf{e}_v)(\mathbf{e}_u - \mathbf{e}_v)^\top$. The expected update matrix is then $\mathbb{E}\mathbf{W}_e = \mathbf{I}_N - (1/2)\mathbf{L}(p)$, where $\mathbf{L}(p) = \sum_{e \in E} p_e (\mathbf{e}_u - \mathbf{e}_v)(\mathbf{e}_u - \mathbf{e}_v)^\top$ is the weighted graph Laplacian. The spectral gap of $\mathbb{E}\mathbf{W}_e$ is denoted $\lambda(p)/2$ and corresponds to the second smallest eigenvalue of $(1/2)\,\mathbf{L}(p)$. We refer to [6, 55] for further details (see Appendix A for more details about graph constructions). In the next section, we connect score-based ranking aggregation to gossip estimation.

3 Decentralized algorithms for ranking aggregation

We consider a system of $N \geq 2$ agents on a graph $\mathcal{G}$, each holding a preference ranking over $n \geq 1$ alternatives, represented by a permutation $\sigma_v \in \mathfrak{S}_n$, for $v = 1, \dots, N$, and we denote by $\widehat{P}_N$ the associated empirical distribution. We consider a pair (ϕ, ψ) such that the score of the voting rule can be written as $S(\widehat{P}_N) = \psi\left(N^{-1} \sum_{v=1}^{N} \phi(\sigma_v)\right)$, where I is a finite index set with $|I|$ polynomial in n (see Table 1), $\phi : \mathfrak{S}_n \to \mathbb{R}^I$ is an embedding and $\psi : \mathbb{R}^I \to \mathbb{R}^n$ is an aggregation map. Our framework follows the generalized scoring-rule of [65], up to a normalization factor N^{-1}, which can be absorbed by ψ. For ϕ, we define the mixed norm as $|||\phi||| := \max_{\sigma \in \mathfrak{S}_n} \|\phi(\sigma)\|$. It can be shown that both Borda and Copeland scores follow this construction [65]. In fact, for Borda score, we have $\psi_B = -\mathrm{Id}$ and ϕ_B the application that maps a permutation σ onto its $\mathbb{R}^n$ representation, then we find that Borda score is just $S_B(\widehat{P}_N) = -N^{-1} \sum_{v=1}^{N} \phi_B(\sigma_v)$. For Copeland score, we consider ϕ_C the application that returns the empirical pairwise probabilities and $[\psi_C((x_{ij})_{i<j})]_i = \sum_{j \neq i} \mathbb{I}\{x_{ij} > 1/2\} - \mathbb{I}\{x_{ij} < 1/2\}$ (see Appendix G for more details). We refer to S as a (ϕ, ψ)-score function when it admits the decomposition following [65]. In what follows, we denote the *average embedding* as $\bar{\mathbf{x}} := \left(\bar{x}^{(1)}, \dots, \bar{x}^{(|I|)}\right)^\top := N^{-1} \sum_{v=1}^{N} \phi(\sigma_v)$ and such that $\bar{x}^{(i)} = N^{-1} \sum_{v=1}^{N} [\phi(\sigma_v)]_i$. Let $\mathbf{s} := S(\widehat{P}_N)$ be the score vector associated with S and $\widehat{P}_N$, where $s_i = [\psi(\bar{\mathbf{x}})]_i$. For any $\mathbf{s} \in \mathbb{R}^n$, we denote by $[\mathbf{s}]_{\mathfrak{S}_n} \in \mathfrak{S}_n$ the permutation induced by argsorting the coordinates of $\mathbf{s}$ in decreasing

order, ties being broken by a fixed arbitrary rule (*e.g.* by item index). We now introduce a notion of stability of the induced ranking, which will be key to deriving our convergence rate bounds.

Definition 1 (Locally ranking-stable)**.** *An aggregation map $\psi : \mathbb{R}^I \to \mathbb{R}^n$ is said to be* locally ranking-stable *at $\mathbf{x} \in \mathbb{R}^I$ if there exists $r > 0$ such that*

$$\forall \mathbf{y} \in \mathbb{R}^I, \quad \|\mathbf{y} - \mathbf{x}\| \leq r \implies [\psi(\mathbf{y})]_{\mathfrak{S}_n} = [\psi(\mathbf{x})]_{\mathfrak{S}_n}.$$

The stability radius *of ψ at $\mathbf{x}$ is defined as the supremum of the radii $r \geq 0$.*

A perturbation of $\mathbf{x}$ that stays within the ℓ_2-ball of radius $r_\psi(\mathbf{x})$ is sufficient, but not necessary, for the ranking to remain unchanged. We note that this quantity is only known asymptotically by each node (see Appendices F.2 and G.2 for direct applications on Borda's and Copeland's methods). Throughout this section, the dataset $\{\sigma_1, \ldots, \sigma_N\}$ is treated as fixed, so that $\widehat{P}_N$, $\bar{\mathbf{x}}$, the consensus $\sigma_\star$ and $r_\psi(\bar{\mathbf{x}})$ are deterministic quantities. We now describe our two algorithms (Algorithms 1 and 2) and refer to Figure 1 for an overview of Algorithm 2.

Algorithm 1 Synchronous (ϕ, ψ)-Score Gossip.

Init: Given a coefficient $\alpha \in [1, 2]$, a ranking σ_v, each voter v sets $\mathbf{x}_v = \phi(\sigma_v) \in \mathbb{R}^I$.
for $t \in [\![T]\!]$ **do**
  for $v \in [\![N]\!]$ in parallel **do**
    for u in neighborhood of v **do**
      Send $\mathbf{x}_v$, receive $\mathbf{x}_u$.
    end for
    if $t = 1$ **then** $\mathbf{x}_v, \mathbf{x}'_v \leftarrow \sum_u W_{vu}\mathbf{x}_u,\ \mathbf{x}_v$.
    else $\mathbf{x}_v, \mathbf{x}'_v \leftarrow (1-\alpha)\mathbf{x}'_v + \alpha \sum_u W_{vu}\mathbf{x}_u,\ \mathbf{x}_v$.
  end for
end for
Each voter v computes $\psi(\mathbf{x}_v)$ locally.
Output: Each voter v returns the ranking associated with $\psi(\mathbf{x}_v)$.

Algorithm 2 Randomized Asynchronous (ϕ, ψ)-Score Gossip.

Init: Given a probability distribution $p = (p_e)_{e \in E}$, a ranking σ_v, each voter v sets $\mathbf{x}_v = \phi(\sigma_v) \in \mathbb{R}^I$.

for $t \in [\![T]\!]$ **do**
  Select an edge $e = (u, v) \in E$ with probability $p_e > 0$.
  Update $\mathbf{x}_u, \mathbf{x}_v \leftarrow (\mathbf{x}_u + \mathbf{x}_v)/2$.
end for

Each voter v computes $\psi(\mathbf{x}_v)$ locally.
Output: Each voter v returns the ranking associated with $\psi(\mathbf{x}_v)$.

Synchronous score gossip. The synchronous gossip algorithm for (ϕ, ψ)-scores is given in Algorithm 1. At each round $t \geq 1$, every node $v \in [\![N]\!]$ simultaneously exchanges its current estimate $\mathbf{x}_v(t)$ with all its neighbors and performs a weighted average update governed by the gossip matrix $\mathbf{W}$. After T rounds, each node independently decodes its local estimate into a ranking via ψ.

The convergence rate is governed by the spectral gap $\lambda_{\mathbf{W}}$: larger values of $\lambda_{\mathbf{W}}$ yield faster concentration of the local estimates $\mathbf{x}_v(t)$ around the average embedding $\bar{\mathbf{x}}$. Since ψ is assumed to be locally ranking-stable at $\bar{\mathbf{x}}$, any estimate that is sufficiently accurate recovers the consensus ranking $\sigma_\star$. Introducing an auxiliary iterate $\mathbf{x}'_v(t)$ yields the accelerated second-order scheme of [10], whose rate scales with $\sqrt{\lambda_{\mathbf{W}}}$. This notably involves an expression of the update matrix as a Chebyshev polynomial (see [4, 44]), which gives the optimal memory coefficient $\alpha_\star = (2 - 2\sqrt{\lambda_{\mathbf{W}}(1 - \lambda_{\mathbf{W}}/4)})/(1 - \lambda_{\mathbf{W}}/2)^2$. When $\alpha = 1$, the second estimate $\mathbf{x}'_v$ is unused and the scheme reduces to the standard first-order synchronous gossip protocol, in which each node maintains only a single embedding vector $\mathbf{x}_v$. Our convergence guarantee is stated in Theorem 1, and its proof is given in Appendix D.

Theorem 1. *Consider Algorithm 1 with $\alpha = \alpha_\star$. Let $\sigma_\star$ be the consensus ranking for $\widehat{P}_N$, and $\mathbf{W}$ be the gossip matrix, with absolute spectral gap $\lambda_{\mathbf{W}} > 0$. Assume ψ satisfies local ranking-stability at the average embedding $\bar{\mathbf{x}}$. For each $i \in I$, we denote the estimate of the i-th coordinate across all voters at time t as $\mathbf{x}^{(i)}(t) = (x_{1i}(t), x_{2i}(t), \ldots, x_{Ni}(t))^\top \in \mathbb{R}^N$. Then, the average expected Kendall-τ error satisfies for $t \geq 0$:*

$$\frac{1}{N}\sum_{v=1}^{N} d_\tau\left(\sigma_\star, \widehat{\sigma}_v(t)\right) \leq \frac{2\binom{n}{2}}{N r_\psi(\bar{\mathbf{x}})^2} e^{-\sqrt{\lambda_{\mathbf{W}}}t} \underbrace{\sum_{i \in I} \|\mathbf{x}^{(i)}(0) - \bar{x}^{(i)}\mathbf{1}_N\|^2}_{=:\gamma}, \tag{3}$$

where $\widehat{\sigma}_v(t) := [\psi(\mathbf{x}_v(t))]_{\mathfrak{S}_n}$, i.e., the ranking induced by the scores computed by node v at time t.

We call γ the *initial dispersion* of the embeddings; it is deterministic since the dataset is fixed. The acceleration comes at a memory cost: each node $v \in [\![N]\!]$ must store an additional embedding vector $\mathbf{x}'_v$. Without this additional vector, we recover the standard synchronous gossip update. The convergence is thus in $\mathcal{O}(e^{-\lambda_{\mathbf{W}} t})$ when $t \to \infty$, which is slower as $\lambda_{\mathbf{W}} < 1$ due to the doubly stochastic nature of $\mathbf{W}$. Theorem 1 also yields a bound on the number of rounds needed for all nodes to recover the consensus ranking.

Remark 1 (δ-mixing time)**.** *We can upper bound the* δ-mixing time *defined for any* $\delta \geq 0$ *as* $T^{\mathrm{Sync}}(\delta) = \min\{t \geq 0 \mid N^{-1} \sum_{v=1}^{N} d_\tau(\widehat{\sigma}_v(t), \sigma_\star) \leq \delta\}$*. By Theorem 1, the mixing time needed to recover the consensus ranking (*$\delta = 0$*) satisfies* $T^{\mathrm{Sync}}(0) \leq \max\{0, \lfloor t^\star \rfloor + 1\}$*, where we define* $t^\star := \log\left(2\binom{n}{2}\gamma / r_\psi(\bar{\mathbf{x}})^2\right) / \sqrt{\lambda_{\mathbf{W}}}$*.*

Randomized score gossip (asynchronous). The randomized gossip algorithm for (ϕ, ψ)-scores is given in Algorithm 2. All expectations are taken with respect to the random edge sampling only. The asynchronous gossip process is a Markov chain with a fixed initial state. By relating it to standard randomized gossip averaging [6], we obtain the following convergence result (see Appendix B for further properties of randomized gossip processes). The proofs of Theorem 2 and of the remarks of this section are given in Appendix D.

Theorem 2. *Consider Algorithm 2 with edge probabilities* $p = (p_e)_{e \in E}$ *such that* $\lambda(p) > 0$*. Assume* ψ *satisfies local ranking-stability at* $\bar{\mathbf{x}}$*. Then, the average expected Kendall-*τ *error satisfies, for all* $t \geq 0$*,*

$$\frac{1}{N} \sum_{v=1}^{N} \mathbb{E}\, d_\tau\left(\sigma_\star, \widehat{\sigma}_v(t)\right) \leq \frac{\binom{n}{2}}{N r_\psi(\bar{\mathbf{x}})^2} e^{-\lambda(p) t/2} \gamma, \tag{4}$$

where $\widehat{\sigma}_v(t) := [\psi(\mathbf{x}_v(t))]_{\mathfrak{S}_n}$ *is the ranking induced by the scores computed by node* v *at time* t*, and* γ *and* $\lambda(p)$ *are defined in Equation* (3) *and Section 2, respectively.*

Theorem 2 shows that each node recovers the consensus ranking $\sigma_\star$ asymptotically, at an exponential rate in expectation. Note that this bound is not tight, and we observe faster convergence in practice (see Section 5). The convergence of both algorithms, as well as the stability radius, admits a simple geometric interpretation.

Remark 2 (Geometric interpretation)**.** *The map* $[\psi(\cdot)]_{\mathfrak{S}_n}$ *partitions the convex hull of the initial embeddings* $(\mathbf{x}_v(0))_{v=1}^N$ *into ranking cells* $C(\sigma) = \{\mathbf{x} \in \mathbb{R}^I \mid [\psi(\mathbf{x})]_{\mathfrak{S}_n} = \sigma\}$*, each consisting of the points* $\mathbf{x}$ *whose induced ranking is* $\sigma \in \mathfrak{S}_n$*. The ranking consensus is achieved once every node embedding belongs to* $C(\sigma_\star)$*, without requiring exact averaging in* $\mathbb{R}^I$*. Moreover, if a perturbation occurs during the gossip protocol but the average embedding vector remains inside the cell* $\mathcal{C}(\sigma_\star)$*, then the consensus ranking is unchanged. The stability radius* $r_\psi(\bar{\mathbf{x}})$ *is then the distance from* $\bar{\mathbf{x}}$ *to the boundary of the corresponding ranking cell* $C(\sigma_\star)$*.*

As in the synchronous case, Theorem 2 also yields a bound on the number of iterations needed to recover the consensus ranking.

Remark 3 (Expected δ-mixing time)**.** *Similarly to the synchronous case (Remark 1), we define the expected* δ*-mixing time, for any* $\delta > 0$*, as*

$$T^{\mathrm{rand}}(\delta) = \min\left\{ t \geq 0 \;\middle|\; \frac{1}{N} \sum_{v=1}^{N} \mathbb{E}\, d_\tau(\widehat{\sigma}_v(t), \sigma_\star) \leq \delta \right\}.$$

Theorem 2 yields the following upper bound: $T^{\mathrm{rand}}(\delta) \leq \max\{0, \lfloor t^\star_{\mathrm{rand}} \rfloor + 1\}$*, where we define* $t^\star_{\mathrm{rand}} := (2/\lambda(p)) \log\left(\binom{n}{2}\gamma / (\delta r_\psi(\bar{\mathbf{x}})^2)\right)$*. Hence, the mixing time of the randomized process scales as* $2/\lambda(p)$*, whereas it scales as* $1/\sqrt{\lambda_{\mathbf{W}}}$ *in the synchronous case. Moreover,* $T^{\mathrm{rand}}(\delta)$ *is lower bounded as* $T^{\mathrm{rand}}(\delta) \geq \max\{0, (\sum_{v=1}^{N} d_\tau(\widehat{\sigma}_v(0), \sigma_\star) - N\delta)/(n(n-1))\}$*.*

In practice, voters often provide partial rather than complete rankings of alternatives. Appendix E provides convergence guarantees for this more general setting. Appendices F and G specialize our results to the Borda and Copeland rules, and Appendix I extends the framework to metric-based consensus under the Kendall-τ and Spearman footrule distances. Note that consensus under the Spearman-ρ distance d_ρ reduces to the Borda case. We now study the robustness of our gossip algorithms.

4 Robustness guarantees using empirical breakdown functions

We consider a *weak adversarial model* in which a subset $B \subset [\![N]\!]$ of $b := |B|$ nodes is corrupted and may report arbitrary rankings. The remaining nodes $H := [\![N]\!] \setminus B$ constitute the *honest subgraph* $\mathcal{G}_H$, defined as the induced subgraph of $\mathcal{G}$ with vertex set H and edge set $E \cap H^2$. In particular, following vote manipulation theory [47], we consider that each corrupted node holds an adversarial initial ranking while still participating in the consensus process. We assume $\mathcal{G}_H$ remains connected, which ensures that honest nodes can still reach consensus among themselves through gossip averaging [27, 59]. An adversary controlling b nodes can shift the final consensus regardless of the communication graph. Although our adversarial model assumes corrupted nodes follow the averaging protocol, the geometric structure of the embedding provides resilience against stronger Byzantine attacks in which corrupted nodes inject arbitrary continuous values during the averaging phase. Since $\phi(\sigma) \in \mathcal{P}_\phi$ for all $\sigma \in \mathfrak{S}_n$, the polytope $\mathcal{P}_\phi$ is compact and bounded. Consequently, any Byzantine perturbation not in $\mathcal{P}_\phi$ injected during averaging is implicitly clipped by ψ. The contamination influence on the output ranking is bounded by the stability radius $r_\psi(\bar{\mathbf{x}})$ characterized in Remark 2, regardless of the magnitude of the injected values.

We define the minimal contamination fraction needed to modify the estimated score $S(\widehat{P}_N)$, a concept originally introduced in the robust statistics literature [32, 36], subsequently adapted to gossip learning in [62] and to ranking distributions in [30, 31], as follows.

Definition 2 (Empirical Breakdown Function)**.** *The* Empirical Breakdown Function *is the minimum fraction of corrupted votes required to shift the consensus by a distance of at least $\delta > 0$:*

$$\varepsilon^\star_{\widehat{P}_N,S}(\delta) = \inf\left\{ \frac{b}{N} \;\middle|\; b \in [\![N]\!], \sup_{\widehat{P}^b_N \in \mathcal{B}(\widehat{P}_N, b/N)} d_\tau\left(S\left(\widehat{P}_N\right), S\left(\widehat{P}^b_N\right)\right) \geq \delta \right\}, \tag{5}$$

where $\mathcal{B}(\widehat{P}_N, \varepsilon) \subset \Delta_{\mathfrak{S}_n}$ denotes the set of empirical distributions $\widehat{P}^b_N$ that can be obtained by modifying $b = \lceil \varepsilon N \rceil$ of the N votes. Formally,

$$\mathcal{B}(\widehat{P}_N, b) = \left\{ \widehat{P}^b_N \in \Delta_{\mathfrak{S}_n} \;\middle|\; d_{\mathrm{TV}}(\widehat{P}_N, \widehat{P}^b_N) \leq \frac{b}{N},\, \widehat{P}^b_N = \frac{1}{N}\sum_{v=1}^{N} \delta_{\tau_v}, \{\tau_1, \ldots, \tau_N\} \in \mathrm{Sym}^N(\mathfrak{S}_n) \right\},$$

where d_{TV} denotes the total variation distance: $d_{\mathrm{TV}}(P, Q) = (1/2)\sum_{\sigma \in \mathfrak{S}_n} |P(\sigma) - Q(\sigma)|$.

Remark 4. *The concept of Margin of Victory (MoV) [54] is related to the empirical breakdown. Indeed, the MoV is defined as the minimal number of votes required to change the top alternative, while the empirical breakdown measures how the full consensus ranking changes with respect to the Kendall-τ metric d_τ. Furthermore, computing the MoV is NP-complete for pair-based voting rules such as Copeland and maximin, while it is computable in polynomial time for methods such as Bucklin or Borda [64], which leads us to believe that computing the empirical breakdown function in the general case is also NP-complete.*

Proposition 1. *Let $\delta > 0$. For a given distribution $\widehat{P}_N$ and score function S with (ϕ, ψ)-decomposition, the breakdown function admits the following lower bound for d_τ:*

$$\varepsilon^- := \frac{r_\psi(\bar{\mathbf{x}})}{2\,|||\phi|||} \leq \varepsilon^\star_{\widehat{P}_N,S}(\delta). \tag{6}$$

Proposition 1 provides a general characterization of robustness for any score function admitting a (ϕ, ψ)-decomposition. The lower bound ε^- quantifies the minimum fraction of corrupted votes an adversary must inject to shift the consensus. In other words, as long as the contamination fraction remains below this threshold, the consensus ranking is unchanged. The limiting factor in Proposition 1 is the geometry of ψ. The lower bound ε^- is independent of δ, reflecting the fact that once the threshold is passed, then the contamination can corrupt with an error to consensus $\delta > 1$. We note that $\varepsilon^- = \mathcal{O}(n^{-\beta})$ where β is a constant depending on ϕ outputs (see Appendix H).

Time-dependent breakdown in gossip protocols. Consider a gossip process as described in Algorithm 2. Let $\sigma_\star \in \mathfrak{S}_n$ denote the target consensus reached by honest nodes when $B = \emptyset$, *i.e.*,

$\hat{\sigma}_v(t) \to \sigma_\star$ when $t \to \infty$ for any $v \in H$. Theorem 3 establishes that breakdown resilience improves exponentially with the number of iterations. We consider the *average* breakdown point over honest nodes:

$$\varepsilon_t^H(\delta) := \inf \left\{ \frac{b}{N} \;\middle|\; \sup_{\widehat{P}_N^b \in \mathcal{B}(\widehat{P}_N,\, b)} \mathbb{E}\, d_\tau \left(\psi \left(\frac{1}{N-b} \sum_{v \in H} \mathbf{x}_v(t) \right), \sigma_\star \right) \geq \delta \right\}.$$

If $B = \emptyset$, the Kendall-τ distance of the honest average with the true consensus $\sigma_\star$ is always null. The following theorem quantifies this convergence for $B \neq \emptyset$, as $t \to \infty$, the bound converges to the weaker bound $\varepsilon^- \delta / \binom{n}{2}$.

Theorem 3. *Let $\lambda_1(\bar{\mathbf{W}})$, $\lambda_{\bar{\mathbf{W}}}$ be the smallest eigenvalue and the spectral gap of $\bar{\mathbf{W}}$, respectively. Then, the empirical breakdown point at iteration $t \geq 0$ is lower bounded by:*

$$\varepsilon_t^H \geq \frac{\left(\sqrt{\xi_t^2 + 4\varepsilon^- \delta (1+\zeta_t) / \binom{n}{2}} - \xi_t \right)^2}{4(1+\zeta_t)^2}, \tag{7}$$

where $\xi_t = (1-\lambda(p))^{t/2}$ and $\zeta_t = 1 - \lambda_1^t(\bar{\mathbf{W}})$.

The bound in (7) reflects two competing spectral phenomena. The term ξ_t controls the residual disagreement among honest nodes (how far their average is from mixing), while ζ_t captures the accumulated contamination from adversarial nodes, growing as their influence percolates through the graph.

5 Numerical experiments

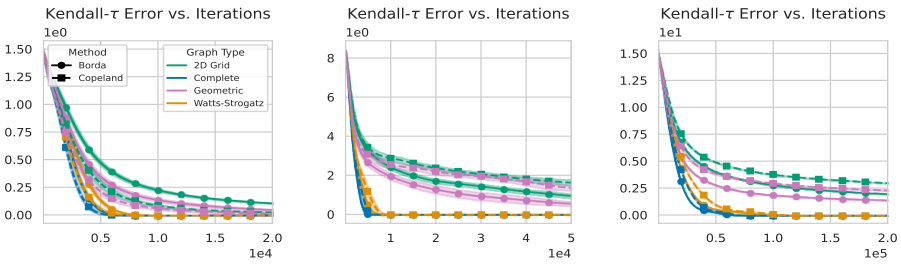
In this section, we evaluate the proposed decentralized consensus methods through a series of numerical experiments[2]. We consider two classes of datasets: synthetic data generated from the Mallows model, and real-world preference data [51], specifically the Sushi [40] and Debian's 2005 leader election datasets. In all settings, the gossip communication follows a uniform distribution over edges, *i.e.*, $p = (1/|E|)_{e \in E}$. Additional experimental setups and supplementary results are provided in Appendix J.

5.1 Gossip convergence experiments

We evaluate our methods on three datasets. Mixture of Mallows (see Appendix H for the definition), $N = 1000$ synthetic rankings over $n = 10$ alternatives drawn from a two-component mixture with weights $(0.7, 0.3)$, dispersion parameters $(\varphi = 0.4, \varphi = 0.6)$, and reference rankings $\sigma_\star$ and σ_{rev} which is $\sigma_\star$ in reversed order, respectively; we run 500 trials over 20000 iterations. Sushi: strict complete rankings of $n = 10$ alternatives by $N = 5000$ voters; 25 trials over 200000 iterations. Debian: partial rankings of $n = 7$ alternatives by $N = 504$ voters; 250 trials over 50000 iterations. Sushi and Debian are taken from PrefLib [51]. All experiments use four graph topologies (complete, 2D grid, geometric, Watts-Strogatz), with data reshuffled across nodes per trial. To handle partial rankings, we apply rank normalization to infer the Borda consensus and the corresponding weak pairwise preferences to deduce the Copeland consensus, as detailed in Appendix E.

In Figure 2, we present the convergence behavior of the decentralized Borda and Copeland methods. The plots show the Kendall-τ distance to the ground-truth consensus averaged over the trials with ± 2 standard error. For both the Sushi and Debian datasets, we operate in a favorable setting in which both methods (Borda and Copeland) yield identical consensus rankings. For both gossip-based Borda and Copeland, convergence is exponential and consistent with the underlying graph topology, slower on the 2D grid and geometric graphs, and faster on the Watts-Strogatz and complete graphs (see Appendix A for more details). For all experiments, the errors of the two consensus algorithms decrease at very similar rates and are consistent with Theorem 2. This highlights that convergence speed is governed by the spectral gap of the graph and the initial data distribution rather than the choice of voting rule.

[2]Code for all experiments is publicly available at https://github.com/kerrianlc/decentralized-ranking-aggregation.



(a) Mixture of Mallows (b) Debian dataset (c) Sushi dataset

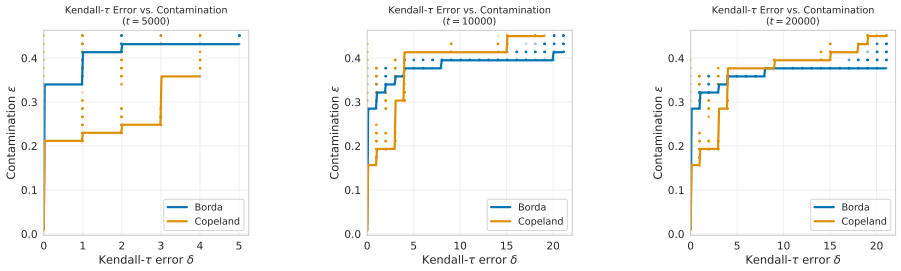
Figure 2: Average local Kendall-τ error to the respective consensus under randomized gossip.

5.2 Breakdown point study

We consider $N = 3000$ agents and $n = 7$ alternatives over a Watts-Strogatz communication graph. The honest fraction $1 - \varepsilon$ of agents holds rankings drawn i.i.d. from a Mallows model P_θ^M with parameter $\theta = (\sigma_\star, \varphi) \in \mathfrak{S}_n \times [0, 1]$, where $\varphi = 0.6$. The contaminated fraction ε of agents holds σ_{rev}, the maximally reversed ranking, which is the worst-case adversarial perturbation. We vary the contamination fraction $\varepsilon \in [0.01, 0.45]$ across 25 values and run 1000 independent trials per configuration. For each trial, we run the gossip algorithms for $t \in \{5000, 10000, 20000\}$ iterations and measure the Kendall-τ distance δ between the honest consensus and the true consensus $\sigma_\star$. We estimate the empirical breakdown curves by running 1000 trials with a different contaminated subset at each trial. This methodology is motivated by the computational complexity discussed in Remark 4.

Figure 3 reports the empirical breakdown curves. The solid step curve is a worst-case statistic. For each $\delta \geq 0$ on the x-axis, it marks the smallest contamination ε at which at least one of the 1000 trials produced an error greater than δ.

First, at low iteration counts ($t = 5000$), the breakdown curve exhibits a pronounced staircase structure with discrete jumps as contamination crosses critical thresholds, which smooths into a plateau by $t = 10000$ as the residual spectral term ξ_t in Theorem 3 decays. Second, Copeland exhibits an initial fragility, breaking at lower contamination levels than Borda in early iterations, but shows lower Kendall-τ error as contamination percolates through the network. Both methods achieve breakdown thresholds substantially higher than the worst-case bound ε^- from Proposition 1, reflecting the gap between theoretical guarantees and average-case robustness. Third, breakdown thresholds slightly decrease with t from $t = 10000$ to $t = 20000$, reflecting contamination propagation through the network as adversarial rankings diffuse into the honest average, though the degradation rate indicates that practical deployments reach near-stationary robustness within finite runtime.



(a) At iteration $t = 5000$ (b) At iteration $t = 10000$ (c) At iteration $t = 20000$

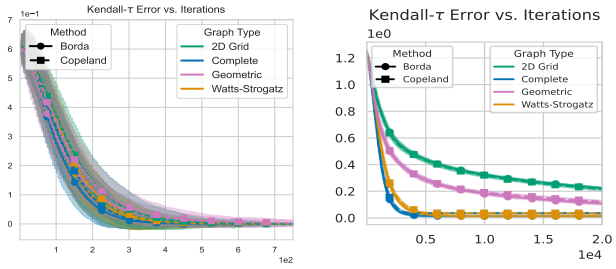
Figure 3: Average Kendall-τ error vs. contamination for the Mallows distribution.

5.3 Application to inference tasks

We consider the following setup. Each node $u \in [\![N]\!]$ on the communication graph represents a learner which holds a private local dataset D_u on which a model is trained. Then, during an

inference phase, each learner share their prediction on a common query and find a consensus on the shared prediction using a communication protocol. This framework is referred to as *Decentralized ensemble learning* [13], where the goal is to combine the predictions of multiple models to improve overall performance. In our case, we assume the output to be a ranking.

We consider two experiments for inference tasks. First, we consider a linear model $y = \langle \boldsymbol{\theta}, \mathbf{d} \rangle + \epsilon$ with $\boldsymbol{\theta} \in \mathbb{R}^d$ $(d = 4)$ and $\epsilon \sim \mathcal{N}(0, \varsigma_u^2)$, where the noise level ς_u specific to node $u \in [\![N]\!]$ $(N = 100, n = 6)$ sampled following a uniform distribution on $[0.1, 0.4]$. Each node u holds a private training set of i.i.d. features $\mathbf{d}$ with their noisy labels, and computes the least-squares estimate $\hat{\boldsymbol{\theta}}_u$. All nodes share a test set $\mathbf{d}_1, \ldots, \mathbf{d}_n \in \mathbb{R}^d$ which allow to make a test score vector $\mathbf{s}_u = (\langle \hat{\boldsymbol{\theta}}_u, \mathbf{d}_1 \rangle, \ldots, \langle \hat{\boldsymbol{\theta}}_u, \mathbf{d}_n \rangle)$ is then argsorted to obtain the ranking $\sigma_u \in \mathfrak{S}_n$. The true scores $\langle \boldsymbol{\theta}, \mathbf{d}_j \rangle$ are pairwise distinct and define similarly the ground-truth ranking $\sigma^\star \in \mathfrak{S}_n$. We consider a second setting motivated by RLHF-style pipelines, we consider a rating dataset, each node fits a local Plackett-Luce [48, 53] utility model from its preferences and returns a score which can be converted into a (partial) ranking. We use the 'MovieLens small' dataset [33] as a proof of concept of the pipeline, where we selected $N = 400$, $n = 25$ of the most contributing users, and most rated movies. Figure 4 illustrates the convergence of the previous tasks.



(a) Toy model (b) MovieLens

Figure 4: Convergence of inference tasks under randomized gossip.

6 Conclusion and perspectives

We have introduced a family of gossip-based algorithms for decentralized ranking aggregation. Our theoretical analysis establishes exponential convergence rates for score-based consensus.

Limitations. Our framework rests on several assumptions worth highlighting. Convergence requires $\mathcal{G}$ to remain connected, and per-round communication costs scale as $O(n^2)$ for Copeland, which may be prohibitive for large n. The adversarial model of Section 4 further assumes corrupted nodes follow the averaging protocol, ruling out more sophisticated attacks such as delayed, selective, or adaptive message manipulation. The upper bounds considered in the convergence analysis are often loose. In practice, the process may exhibit a faster decay as suggested by some empirical observations.

Perspectives. A natural extension is the design of provably robust gossip protocols under structured adversarial models [34]. Decentralized settings also require careful consideration of neighbor trustworthiness. Algorithms 1 and 2 provide no privacy protection, because ϕ is injective. Indeed, since the embedding ϕ maps $\mathfrak{S}_n$ injectively onto the finite set $\{\phi(\sigma) \mid \sigma \in \mathfrak{S}_n\}$, the first message from node v to any neighbor reveals σ_v exactly. A solution is to compose these algorithms with a differential-privacy mechanism [18] that injects noise into each exchanged embedding. The variance must be sufficiently small to ensure that the ranking is preserved with high probability.

Acknowledgments and Disclosure of Funding

This work benefited from Hi! PARIS and State funding managed by the French National Research Agency (ANR) under the France 2030 program, reference ANR-23-IACL-0005. It was also supported by the PEPR IA Foundry grant.

References

[1] Kenneth J. Arrow. Social choice and individual values. *Cowles Monograph No. 12*, 1951.

[2] Lyes Badis, Mourad Amad, Djamil Aïssani, and Sofiane Abbar. P2pcf: A collaborative filtering based recommender system for peer to peer social networks. *Journal of High Speed Networks*, 27(1):13–31, 2021.

[3] Linas Baltrunas, Tadas Makcinskas, and Francesco Ricci. Group recommendations with rank aggregation and collaborative filtering. In *Proceedings of the 4th ACM conference on Recommender systems*, 2010.

[4] Raphaël Berthier, Francis Bach, and Pierre Gaillard. Accelerated gossip in networks of given dimension using jacobi polynomial iterations. *SIAM Journal on Mathematics of Data Science*, 2(1):24–47, 2020.

[5] Jean-Charles de Borda. Mémoire sur les élections au scrutin. *Histoire de l'Académie Royale des Sciences*, 1781.

[6] Stephen Boyd, Arpita Ghosh, Balaji Prabhakar, and Devavrat Shah. Randomized gossip algorithms. *IEEE transactions on information theory*, 52(6):2508–2530, 2006.

[7] Felix Brandt, Vincent Conitzer, and Ulle Endriss. Computational social choice. *Multiagent systems*, 2:213–284, 2012.

[8] Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D Procaccia. Handbook of computational social choice. *Cambridge University Press*, 2016.

[9] Róbert Busa-Fekete, Eyke Hüllermeier, and Balázs Szörényi. Preference-based rank elicitation using statistical models: The case of mallows. In *International conference on machine learning*, pages 1071–1079. PMLR, 2014.

[10] Ming Cao, Daniel A. Spielman, and Edmund M. Yeh. Accelerated gossip algorithms for distributed computation. In *Proceedings of the 44th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, Illinois, USA, September 2006.

[11] Ioannis Caragiannis, Xenophon Chatzigeorgiou, George A Krimpas, and Alexandros A Voudouris. Optimizing positional scoring rules for rank aggregation. *Artificial Intelligence*, 267:58–77, 2019.

[12] Sujoy Chatterjee, Anirban Mukhopadhyay, and Malay Bhattacharyya. A weighted rank aggregation approach towards crowd opinion analysis. *Knowledge-Based Systems*, 149:47–60, 2018.

[13] Hong-You Chen and Wei-Lun Chao. FedBE: Making Bayesian Model Ensemble Applicable to Federated Learning. October 2020. URL `https://openreview.net/forum?id=dgtpE6gKjHn`.

[14] Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL `https://papers.nips.cc/paper_files/paper/2017/hash/d5e2c0adad503c91f91df240d0cd4e49-Abstract.html`.

[15] Pierre Colombo, Nathan Noiry, Ekhine Irurozki, and Stéphan Clémençon. What are the best systems? new perspectives on nlp benchmarking. *Advances in neural information processing systems*, 35:26915–26932, 2022.

[16] Nicolas de Condorcet. Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix. *L'imprimerie royale*, 1785.

[17] Arthur H Copeland. A reasonable social welfare function. *Seminar on Applications of Mathematics to Social Sciences*, 1951.

[18] Edwige Cyffers, Mathieu Even, Aurélien Bellet, and Laurent Massoulié. Muffliato: Peer-to-peer privacy amplification for decentralized optimization and averaging. *Advances in Neural Information Processing Systems*, 35:15889–15902, 2022.

[19] Michel Marie Deza and Elena Deza. Encyclopedia of Distances. *Springer*, 2009.

[20] Persi Diaconis. *Group representations in probability and statistics*, volume 11. Institute of mathematical statistics Hayward, CA, 1988.

[21] Jean-Paul Doignon, Aleksandar Pekeč, and Michel Regenwetter. The repeated insertion model for rankings: Missing link between two subset choice models. *Psychometrika*, 69(1):33–54, 2004.

[22] Cynthia Dwork, Ravi Kumar, Moni Naor, and Dandapani Sivakumar. Rank aggregation methods for the web. In *Proceedings of the 10th international conference on World Wide Web*, pages 613–622, 2001.

[23] Gábor Erdélyi, Michael R. Fellows, Jörg Rothe, and Lena Schend. Control complexity in Bucklin and fallback voting: A theoretical analysis. *Journal of Computer and System Sciences*, 81(4):632–660, 2015. ISSN 0022-0000. doi: 10.1016/j.jcss.2014.11.002. URL https://www.sciencedirect.com/science/article/pii/S0022000014001433.

[24] Ronald Fagin, Ravi Kumar, and Dandapani Sivakumar. Efficient similarity search and classification via rank aggregation. In *Proceedings of the 2003 ACM SIGMOD international conference on Management of data*, pages 301–312, 2003.

[25] Piotr Faliszewski, Edith Hemaspaandra, Lane A Hemaspaandra, and Jörg Rothe. Copeland voting fully resists constructive control. In *International Conference on Algorithmic Applications in Management*, pages 165–176. Springer, 2008.

[26] Johannes Fürnkranz and Eyke Hüllermeier, editors. *Preference Learning*. Springer, Berlin, Heidelberg, 2011. ISBN 978-3-642-14124-9 978-3-642-14125-6. doi: 10.1007/978-3-642-14125-6. URL http://link.springer.com/10.1007/978-3-642-14125-6.

[27] Renaud Gaucher, Aymeric Dieuleveut, and Hadrien Hendrikx. Unified breakdown analysis for byzantine robust gossip. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=OOYPeymDz2.

[28] Renaud Gaucher, Hadrien Hendrikx, and Aymeric Dieuleveut. Byzantine-robust gossip: Insights from a dual approach. *Transactions on Machine Learning Research*, 2026. ISSN 2835-8856. URL https://openreview.net/forum?id=wrLiUpfk4s.

[29] E. N. Gilbert. Random plane networks. *Journal of the Society for Industrial and Applied Mathematics*, 9(4):533–543, 1961. doi: 10.1137/0109045. URL https://doi.org/10.1137/0109045.

[30] Morgane Goibert, Stephan Clemencon, Ekhine Irurozki, and Pavlo Mozharovskyi. Statistical depth functions for ranking distributions: Definitions, statistical learning and applications. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pages 10376–10406. PMLR, 28–30 Mar 2022. URL https://proceedings.mlr.press/v151/goibert22a.html.

[31] Morgane Goibert, Clément Calauzènes, Ekhine Irurozki, and Stephan Clémençon. Robust consensus in ranking data analysis: Definitions, properties and computational issues. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, 2023.

[32] Frank R. Hampel. A general qualitative definition of robustness. *The Annals of Mathematical Statistics*, 42(6):1887–1896, 1971.

[33] F. Maxwell Harper and Joseph A. Konstan. The MovieLens Datasets: History and Context. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 5(4):19:1–19:19, December 2015. ISSN 2160-6455. doi: 10.1145/2827872. URL https://dl.acm.org/doi/10.1145/2827872.

[34] Lie He, Sai Praneeth Karimireddy, and Martin Jaggi. Byzantine-robust decentralized learning via clippedgossip, April 2023. URL http://arxiv.org/abs/2202.01545. arXiv:2202.01545 [cs].

[35] Gan Keng Hoon, Saravadee Sae Tan, Chan Huah Yong, and Tang Enya Kong. Rank aggregation model for meta search: an approach using text and rank analysis measures. In *International Conference on Intelligent Information Processing*, pages 325–339. Springer, 2004.

[36] Peter J. Huber and Elvezio M. Ronchetti. *Robust statistics*. Wiley series in probability and statistics. Wiley, Hoboken (N.J.), 2nd ed edition, 2009. ISBN 978-0-470-12990-6.

[37] Olivier Hudry. NP-hardness results for the aggregation of linear orders into median orders. *Annals of Operations Research*, 2008.

[38] Yunlong Jiao and Jean-Philippe Vert. The kendall and mallows kernels for permutations. In *International Conference on Machine Learning*, pages 1935–1944. PMLR, 2015.

[39] Yunlong Jiao, Anna Korba, and Eric Sibony. Controlling the distance to a Kemeny consensus without computing it. In *International Conference on Machine Learning*, pages 2971–2980. PMLR, 2016.

[40] Toshihiro Kamishima. Nantonac collaborative filtering: Recommendation based on order responses. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 583–588, 2003. doi: 10.1145/956750.956823.

[41] John G Kemeny. Mathematics without numbers. *Daedalus*, 88:571–591, 1959.

[42] D. Kempe, A. Dobra, and J. Gehrke. Gossip-based computation of aggregate information. In *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.*, pages 482–491, Cambridge, MA, USA, 2003. IEEE Computer. Soc. ISBN 978-0-7695-2040-7. doi: 10.1109/SFCS.2003.1238221. URL http://ieeexplore.ieee.org/document/1238221/.

[43] Anna Korba, Stephan Clémençon, and Eric Sibony. A learning theory of ranking aggregation. In *Artificial intelligence and statistics*, 2017.

[44] Ji Liu, Brian D. O. Anderson, Ming Cao, and A. Stephen Morse. Analysis of accelerated gossip algorithms. *Automatica*, 49(4):873–883, April 2013. ISSN 0005-1098. doi: 10.1016/j.automatica.2013.01.001. URL https://www.sciencedirect.com/science/article/pii/S0005109813000022.

[45] Tie-Yan Liu. Learning to rank for information retrieval. *Foundations and Trends® in Information Retrieval*, 3(3):225–331, 2009.

[46] Tyler Lu and Craig Boutilier. Effective sampling and learning for mallows models with pairwise-preference data. *The Journal of Machine Learning Research*, 15(1):3783–3829, 2014.

[47] Tyler Lu, Pingzhong Tang, Ariel D. Procaccia, and Craig Boutilier. Bayesian vote manipulation: Optimal strategies and impact on welfare. In *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence (UAI)*, 2012.

[48] R. Duncan Luce. *Individual choice behavior*. Individual choice behavior. John Wiley, Oxford, England, 1959. Pages: xii, 153.

[49] Colin L. Mallows. Non-null ranking models. *Biometrika*, 44(1-2):114–130, 1957.

[50] John C Mason and David C Handscomb. *Chebyshev polynomials*. Chapman and Hall/CRC, 2002.

[51] Nicholas Mattei and Toby Walsh. Preflib: A library of preference data HTTP://PREFLIB.ORG. In *Proceedings of the 3rd International Conference on Algorithmic Decision Theory*, 2013.

[52] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In Aarti Singh and Jerry Zhu, editors, *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pages 1273–1282. PMLR, 20–22 Apr 2017. URL https://proceedings.mlr.press/v54/mcmahan17a.html.

[53] R. L. Plackett. The analysis of permutations. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 24(2):193–202, 1975. ISSN 0035-9254. doi: 10.2307/2346567. URL https://doi.org/10.2307/2346567.

[54] Yannick Reisch, Jörg Rothe, and Lena Schend. The margin of victory in schulze, cup, and copeland elections: Complexity of the regular and exact variants. In *STAIRS*, pages 250–259, 2014.

[55] Devavrat Shah. Gossip algorithms. *Foundations and trends in networking*, 3(1):1–125, 2009.

[56] Jin Sima, Vishal Rana, and Olgica Milenkovic. Federated aggregation of mallows rankings: A comparative analysis of borda and lehmer coding. *arXiv preprint arXiv:2409.00848*, 2024.

[57] Paul B. Simpson. On defining areas of voter choice: Professor tullock on stable voting. *The Quarterly Journal of Economics*, 83(3):478–490, 1969.

[58] Anand Siththaranjan, Cassidy Laidlaw, and Dylan Hadfield-Menell. Distributional preference learning: Understanding and accounting for hidden context in RLHF. In *International Conference on Learning Representations*, volume 2024, pages 27448–27471, 2024.

[59] Lili Su and Nitin Vaidya. Byzantine multi-agent optimization: Part II. *arXiv preprint arXiv:1506.04681*, 2015.

[60] John N. Tsitsiklis. *Problems in decentralized decision making and computation*. Massuchusetts Institute of Technology, 1984.

[61] Anna van Elst, Igor Colin, and Stephan Clémençon. Fast and efficient asynchronous gossip algorithms for robust and non-smooth convex decentralized learning. *arXiv preprint arXiv:2601.20571*, 2026.

[62] Anna Van Elst, Igor Colin, and Stephan Clémençon. Robust distributed estimation: Extending gossip algorithms to ranking and trimmed means. *Advances in Neural Information Processing Systems*, 38:26493–26529, 2026.

[63] Duncan Watts and Steven Strogatz. Collective Dynamics of Small-World Networks. 1998.

[64] Lirong Xia. Computing the margin of victory for various voting rules. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, EC '12, pages 982–999, New York, NY, USA, June 2012. Association for Computing Machinery. ISBN 978-1-4503-1415-2. doi: 10.1145/2229012.2229086. URL https://dl.acm.org/doi/10.1145/2229012.2229086.

[65] Lirong Xia and Vincent Conitzer. Generalized scoring rules and the frequency of coalitional manipulability. In *Proceedings of the 9th ACM conference on Electronic commerce*, EC '08, pages 109–118, New York, NY, USA, July 2008. Association for Computing Machinery. ISBN 978-1-60558-169-9. doi: 10.1145/1386790.1386811. URL https://dl.acm.org/doi/10.1145/1386790.1386811.

[66] H. Peyton Young. Condorcet's theory of voting. *American Political science review*, 82(4): 1231–1244, 1988.

[67] Daniel M. Ziegler, Nisan Stiennon, Jeffrey Wu, Tom B. Brown, Alec Radford, Dario Amodei, Paul Christiano, and Geoffrey Irving. Fine-tuning language models from human preferences, January 2020. URL http://arxiv.org/abs/1909.08593. arXiv:1909.08593 [cs].

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract claims gossip-based algorithms for decentralized score-based ranking aggregation with convergence guarantees, robustness analysis via breakdown functions, and empirical validation across network topologies and datasets. These are fully reflected in the paper.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: A dedicated Limitations paragraph is included in Section 6, discussing the assumptions underlying our theoretical guarantees, the scope of the adversarial model, communication costs, and the asymptotic nature of our convergence results.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: All proofs are provided in Appendix D for the main theorems and results and Appendices F and G for the Borda and Copeland-specific results. Each theorem explicitly states its assumptions before the result.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: All information needed to reproduce the experiments is provided: dataset sources, graph topologies and construction parameters, gossip edge distribution, number of agents, alternatives, iterations, trials, and Mallows parameters.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: Experimental code is available on dedicated repository with comments and format necessary for main experimental runs.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Section 5 and Appendix J specify all experimental details: dataset sizes, graph topologies, gossip edge distribution, iteration counts, and Mallows parameters.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Figures display ± 2 standard error bands, Tables 4 and 5 report mean ± standard error over 100 trials, with variability from random data reshuffling, graph generation, and Mallows sampling.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Appendix K details both the hardware/workers used for the main experiments of the paper as well as an approximation of the computing time.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: No potential harms arising from the research process are identified; no human participants were involved and the datasets used raise no privacy (as data is anonymous), consent, or deprecation concerns. None of the societal impact categories outlined in the Code of Ethics are at risk. Reproducibility is ensured through public release of the code.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: This work is foundational research on decentralized ranking aggregation algorithms with no direct path to negative societal impacts. The robustness guarantees developed here are, if anything, protective against manipulation rather than enabling it.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper proposes gossip-based algorithms for ranking aggregation and releases no pretrained models or scraped datasets. The experimental data is sourced from public repositories.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: All non-synthetic datasets are sourced from PrefLib [51] and cited accordingly. The library used to load them is credited in the code. No proprietary assets are used.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: The code is released on a dedicated repository, no new data provided, no new model considered.

Guidelines:

- The answer [N/A] means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: The paper does not involve any crowdsourced datasets nor research with human subjects beyond existing datasets.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve any crowdsourced datasets nor research with human subjects beyond existing datasets.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: No LLMs are involved in the core methodology of the paper.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.

Outline of the Supplementary Materials

The supplementary material is organized as follows. Appendix A describes the graph topologies used throughout our experiments. Appendix B details the randomized gossip communication protocols and some properties of these protocols. Appendix C presents the two-layer aggregation framework for score-based aggregation. Appendix D contains all proofs of the main theoretical results stated in the paper. Appendix E discusses extensions of our framework to broader settings especially considering partial rankings. Appendices F and G provide further analysis of the Borda and Copeland voting rules under the developed framework. Appendix H formalizes and discusses the probabilistic model considerations for data. Appendix I describes additional baselines notably for metric-based consensus and their comparisons and robustness. Finally, Appendix J presents additional experiments and ablation studies, while Appendix K details the compute resources used for the different experiments of the paper.

A Graph requirements and topologies

A.1 Spectral properties of gossip matrices

The convergence of gossip-based averaging algorithms is tightly linked to the spectral properties of the underlying communication graph $\mathcal{G} = (\llbracket N \rrbracket, E)$. We briefly recall the relevant definitions and

describe some key spectral quantities. In gossip protocols, transitions from one state to another are given by a gossip matrix whose definition is recalled here.

Definition A.1 (Gossip Matrix)**.** *A matrix* $\mathbf{W}$ *is said to be a* gossip matrix *if and only if:*

- $\mathbf{W}$ *is symmetric, i.e.,* $W_{uv} = W_{vu}$ *for all* $u \neq v$.
- $W_{uv} > 0$ *when* $(u, v) \in E$ *and* $W_{uv} = 0$ *when* $(u, v) \notin E$.
- $W_{vv} \geq 0$ *for all* $v \in [\![N]\!]$.
- $\mathbf{W}$ *is doubly stochastic, i.e.,* $\mathbf{W}\mathbf{1}_N = \mathbf{W}^\top \mathbf{1}_N = \mathbf{1}_N$.

A necessary structural requirement is that $\mathcal{G}$ remains connected. For a symmetric real matrix $\mathbf{W}$, we denote its spectrum by $\mathrm{Sp}(\mathbf{W})$, with eigenvalues ordered as $\lambda_1(\mathbf{W}) \leq \cdots \leq \lambda_N(\mathbf{W})$. Since $\mathcal{G}$ is connected, the Perron-Frobenius theorem ensures that the eigenvalue 1 of $\mathbf{W}$ is simple and its associated eigenspace is $\mathrm{span}(\mathbf{1}_N)$. In the randomized gossip setting, the eigenvalues of the weighted graph Laplacian $\mathbf{L}(p)$ satisfy

$$0 = \lambda_1(p) \leq \lambda_2(p) \leq \cdots \leq \lambda_N(p).$$

Definition A.2 (Spectral Gap)**.** *For a given gossip matrix* $\mathbf{W}$*, the* spectral gap *is defined as*

$$\lambda_{\mathbf{W}} := \min_{\lambda \in \mathrm{Sp}\,(\mathbf{W}) \setminus \{1\}} 1 - |\lambda|.$$

In the randomized gossip setting, since all eigenvalues of $\mathbf{L}(p)$ are non-negative, the spectral gap reduces to $\lambda_2(\mathbf{L}(p))/2$. This quantity is strictly positive if and only if $\mathcal{G}$ is connected. When connectivity fails, $\lambda_{\mathbf{W}} = 0$ and the convergence guarantee breaks down. The spectral gap therefore serves as the key parameter controlling the rate at which nodes reach consensus.

A.2 Topology considerations

More generally, if $\mathcal{G}$ decomposes into $K \geq 2$ disjoint connected components, then $\mathbf{L}(p)$ has exactly K zero eigenvalues and the first non-zero eigenvalue is $\lambda_{K+1}(p)$. In this regime, a local consensus is reached within each component, but there is no guarantee that distinct components converge to the same value, so the global average $\bar{\mathbf{x}}$ is no longer recoverable. Whereas deterministic topologies such as rings or 2D grids are always connected by construction, random graph models may yield disconnected realizations, and care must be taken when using such constructions in experiments.

Small-world model (Watts–Strogatz) [63]. The graph is built using the following random procedure. The graph $\mathcal{G}$ contains N nodes and is initialized as a ring lattice where each node is linked to its k nearest neighbors. Then, for each node $v \in [\![N]\!]$ and each of its $k/2$ rightmost edges, the edge is rewired with probability $p > 0$ to a node $u \neq v$ chosen uniformly at random. For $k \gg \log(N)$, $\mathcal{G}$ is almost surely connected.

Random geometric graph [29]. N points are drawn uniformly at random in the unit square, and two points are connected if their Euclidean distance is below a threshold r.

B Preliminaries on randomized gossip protocol

Markov chain considerations. In this section, we first formalize the probabilistic constructions necessary to prove the main results. We consider the filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0})$, where the universe Ω contains the trajectories $\omega = (\omega_1, \omega_2, \dots)$ and $(\mathcal{F}_t)_{t\geq 0}$ is the filtration associated with the gossip stochastic process $(\mathbf{X}(t))_{t\geq 0}$. The randomness of a gossip process arises from the edge sampling process $e(\omega) = (e_t)_{t\geq 1}$. At each iteration $t > 0$, the edge e_t is drawn independently of the previous iterations, with $\mathbb{P}(e_t = e) = p_e > 0$. The gossip process is actually a Markov chain thanks to the construction of the edge sampling process. We denote $\mathbf{x}^{(i)}(t) := (x_{1,i}(t), \dots, x_{N,i}(t))^\top \in \mathbb{R}^N$ for $i \in I$ (component-wise estimation at iteration t across all voters), $\mathbf{x}_v(t) := (x_{v,1}(t), \dots, x_{v,|I|}(t))^\top \in \mathbb{R}^I$ (voter-wise embedding estimation at iteration t) for $v \in [\![N]\!]$ and finally $\mathbf{X}(t) := (\mathbf{x}_1(t), \dots, \mathbf{x}_N(t)) = (\mathbf{x}^{(1)}(t)^\top, \dots, \mathbf{x}^{(|I|)}(t)^\top)^\top$. For simplicity, we write L^2 for $L^2(\Omega, \mathcal{F}, \mathbb{P})$.

Main properties of randomized gossip processes. For completeness, we recall standard properties of gossip matrices. We define the normalized vector $\mathbf{z}_N = \mathbf{1}_N/\sqrt{N}$.

Lemma B.1 (Boyd et al. [6])**.** *Let I be a finite index set, $p = (p_e)_{e\in E}$. The vector $\mathbf{x}^{(i)}(t)$ evolves according to the gossip-based averaging procedure described in Section 2. Then, for any $i \in I$ and $t > 0$, we have*

$$\mathbb{E}\left\|\mathbf{x}^{(i)}(t) - \bar{x}^{(i)}\mathbf{1}_N\right\|^2 \leq \left(1 - \frac{\lambda(p)}{2}\right)^t \left\|\mathbf{x}^{(i)}(0) - \bar{x}^{(i)}\mathbf{1}_N\right\|^2,$$

and thus $\mathbf{x}^{(i)}(t)$ converges towards $\bar{x}^{(i)}\mathbf{1}_N$ in L^2 when $t \to \infty$.

Lemma B.2. *Let $t > 0$. If $e \in E$ is sampled with $e = (u, v)$, then the randomized gossip matrix at iteration t is given by $\mathbf{W}_t = \mathbf{I}_N - \frac{1}{2}\mathbf{L}_e$, where $\mathbf{L}_e := (\mathbf{e}_u - \mathbf{e}_v)(\mathbf{e}_u - \mathbf{e}_v)^\top$. The following properties hold:*

1. *The matrices $\mathbf{W}_t$ are gossip matrices. Moreover, the matrices are idempotent, i.e., $\mathbf{W}_t^2 = \mathbf{W}_t$.*
2. *Since $\sum_{e\in E} p_e = 1$, $\bar{\mathbf{W}} = \mathbb{E}[\mathbf{W}_t] = \mathbf{I}_N - \frac{1}{2}\sum_{e\in E} p_e\mathbf{L}_e$. The matrix $\bar{\mathbf{W}}$ is also doubly stochastic and it follows that $\mathbf{1}_N$ is an eigenvector with eigenvalue 1.*
3. *$\bar{\mathbf{W}}$ is positive semi-definite.*
4. *The matrix $\bar{\mathbf{W}} - \mathbf{z}_N\mathbf{z}_N^\top$ satisfies, by construction, $(\bar{\mathbf{W}} - \mathbf{z}_N\mathbf{z}_N^\top)\mathbf{1}_N = 0$, and it can be shown that $\|\bar{\mathbf{W}} - \mathbf{z}_N\mathbf{z}_N^\top\|_{\mathrm{op}} \leq \lambda_{N-1}(\bar{\mathbf{W}})$ and $\|\cdot\|_{\mathrm{op}}$ denotes the operator norm of a matrix. Moreover, the eigenvalue $\lambda_{N-1}(\bar{\mathbf{W}})$ satisfies $0 \leq \lambda_{N-1}(\bar{\mathbf{W}}) < 1$ and $\lambda_{N-1}(\bar{\mathbf{W}}) = 1 - \lambda(p)/2$ where $\lambda(p)$ is the spectral gap (or second smallest eigenvalue) of the Laplacian of the weighted graph defined as $\mathbf{L}(p) = \sum_{e\in E} p_e\mathbf{L}_e$.*

Proof. For points *1.*, *2.*, and *4.*, see Boyd et al. [6]. For point *3.*, observe that each individual gossip matrix $\mathbf{W}_e = \mathbf{I}_N - (\mathbf{e}_u - \mathbf{e}_v)(\mathbf{e}_u - \mathbf{e}_v)^\top/2$ is positive semi-definite. Thus, for any $\mathbf{y} \in \mathbb{R}^N$,

$$\langle \mathbf{y}, \mathbf{W}_e\mathbf{y}\rangle = \|\mathbf{y}\|^2 - \frac{(y_u - y_v)^2}{2} \geq 0,$$

where the inequality follows from $(y_u - y_v)^2 \leq 2(y_u^2 + y_v^2) \leq 2\|\mathbf{y}\|^2$. Since $\bar{\mathbf{W}} = \sum_{e\in E} p_e\mathbf{W}_e$ is a convex combination of positive semi-definite matrices with $p_e > 0$ and $\sum_{e\in E} p_e = 1$, it is itself PSD. □

C Examples of the two-layer framework

We now describe the general score-based framework underlying our algorithms. Let I be an index set and consider mappings $\phi : \mathfrak{S}_n \to \mathbb{R}^I$, $\psi : \mathbb{R}^I \to \mathbb{R}^n$, where ϕ encodes a ranking and ψ produces a score vector. The final ranking is $[\psi(\bar{\mathbf{x}})]_{\mathfrak{S}_n}$, where $\bar{\mathbf{x}} = N^{-1}\sum_{v=1}^N \phi(\sigma_v)$ and the notation $[\,\cdot\,]_{\mathfrak{S}_n}$ is defined in Section 3. We present several such constructions in Table 1 (see [65, Prop. 2]).

Method	Embedding $\phi(\sigma)$	Aggregation $\psi(\mathbf{x})$	Dim. $\lvert I\rvert$
Borda	$(\sigma(1), \dots, \sigma(n))^\top$	$-\mathrm{Id}$	n
Copeland	$(\mathbb{I}\{\sigma(i) < \sigma(j)\})_{i<j}$	$\left(\sum_{j\neq i}\left[\mathbb{I}\{\mathbf{x}_{ij} > \frac{1}{2}\} - \mathbb{I}\{\mathbf{x}_{ij} < \frac{1}{2}\}\right]\right)_{i=1}^n$	$\binom{n}{2}$
Bucklin [23]	$(\mathbb{I}\{\sigma(i) = k\})_{i,k\in[\![n]\!]}$	$\left(-\min\left\{k \in [\![n]\!] \mid \sum_{\ell=1}^k \mathbf{x}_{i\ell} > 1/2\right\}\right)_{i=1}^n$	n^2
Plurality [8]	$(\mathbb{I}\{\sigma(i) = 1\})_{i\in[\![n]\!]}$	Id	n
Maximin [57]	$(\mathbb{I}\{\sigma(i) < \sigma(j)\})_{i<j}$	$(\min_{j\neq i}\mathbf{x}_{ij})_{i=1}^n$	$\binom{n}{2}$

Table 1: Common ranking methods expressed in embedding-aggregation form. The map ϕ encodes a permutation $\sigma \in \mathfrak{S}_n$ into the embedding space, ψ aggregates information into scores, and $|I|$ denotes the dimensionality of the embedding.

Remark C.1 (Communication cost)**.** *At each gossip round, each node exchanges a vector of dimension $|I|$. For Borda and plurality, $|I| = n$, for Copeland and Maximin, $|I| = \binom{n}{2}$, for Bucklin, $|I| = n^2$ (see Table 1). In contrast, a centralized approach requires collecting all N rankings at a single node, incurring $O(Nn)$ total communication. The gossip approach replaces this with $O(T|I|)$ per-node cost, where T is the mixing time.*

Proposition C.2 establishes a useful regularity property of the stability radius, regardless of the score-based voting rule.

Proposition C.2. *For a given function ψ, the function $\mathbf{x} \mapsto r_\psi(\mathbf{x})$ is 1-Lipschitz on $\mathbb{R}^I$.*

Proof. We show that $\mathbf{x} \mapsto r_\psi(\mathbf{x})$ is 1-Lipschitz, which implies continuity. Let $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^I$. We can write the stability radius as $r_\psi(\mathbf{x}) = \inf_{\mathbf{z} \in T(\mathbf{x})} \|\mathbf{x} - \mathbf{z}\|$, where $T(\mathbf{x}) := \{\mathbf{y} \in \mathbb{R}^I \mid [\psi(\mathbf{y})]_{\mathfrak{S}_n} \neq [\psi(\mathbf{x})]_{\mathfrak{S}_n}\}$. If $[\psi(\mathbf{x})]_{\mathfrak{S}_n} = [\psi(\mathbf{x}')]_{\mathfrak{S}_n}$, then $T(\mathbf{x}) = T(\mathbf{x}')$. Since the distance to a fixed set is 1-Lipschitz:

$$|r_\psi(\mathbf{x}) - r_\psi(\mathbf{x}')| \leq \|\mathbf{x} - \mathbf{x}'\|.$$

Otherwise, we get $\mathbf{x}' \in T(\mathbf{x})$ and $\mathbf{x} \in T(\mathbf{x}')$. Following the definition of $T(\mathbf{x})$, $r_\psi(\mathbf{x}) \leq \|\mathbf{x} - \mathbf{x}'\|$ and $r_\psi(\mathbf{x}') \leq \|\mathbf{x} - \mathbf{x}'\|$. Therefore, $\max(r_\psi(\mathbf{x}), r_\psi(\mathbf{x}')) \leq \|\mathbf{x} - \mathbf{x}'\|$. Since $\max(r_\psi(\mathbf{x}), r_\psi(\mathbf{x}')) \geq |r_\psi(\mathbf{x}) - r_\psi(\mathbf{x}')|$, we conclude that r_ψ is 1-Lipschitz in both cases. □

Remark C.3. *Although $r_\psi(\bar{\mathbf{x}})$ depends on the unknown global average $\bar{\mathbf{x}}$, Proposition C.2 allows each node to estimate it consistently from its local gossip iterate. More precisely, the Lipschitz property gives $\mathbb{E}|r_\psi(\mathbf{x}_v(t)) - r_\psi(\bar{\mathbf{x}})|^2 \leq \mathbb{E}\|\mathbf{x}_v(t) - \bar{\mathbf{x}}\|^2$, which decays at the same rate as the gossip process itself. This enables practical estimation of the stability radius appearing in the mixing time bounds of Remarks 1 and 3, each node can compute $r_\psi(\mathbf{x}_v(t))$ locally to approximate the global threshold $r_\psi(\bar{\mathbf{x}})$ and implement adaptive stopping criteria without coordination.*

D Proofs of main results

D.1 Convergence of shift-register gossip via Chebyshev polynomials

Let $\mathbf{W}$ be a gossip matrix (see Appendix A), with eigenvalues $-1 \leq \lambda_1 \leq \cdots \leq \lambda_{N-1} \leq \lambda_N = 1$, where $\lambda_k := \lambda_k(\mathbf{W})$, and an orthonormal basis of eigenvectors $(\mathbf{z}_k)_{k=1}^N$ with $\mathbf{z}_N = \mathbf{1}_N/\sqrt{N}$.

Given initial values $\mathbf{a} \in \mathbb{R}^N$ with average $\bar{a} = (1/N)\sum_{v=1}^N a_v$, we consider the shift-register (second-order) gossip iteration with parameter α:

$$\mathbf{x}(0) = \mathbf{a}, \qquad \mathbf{x}(1) = \mathbf{W}\mathbf{a}, \qquad \mathbf{x}(t+1) = \alpha\mathbf{W}\mathbf{x}(t) + (1-\alpha)\mathbf{x}(t-1), \quad t \geq 1. \tag{8}$$

The first step is a plain gossip step. We use the optimal parameter of [44], written in terms of $\lambda_{\mathbf{W}}$:

$$\alpha_\star = 2\,\frac{1 - \sqrt{\lambda_{\mathbf{W}}(1 - \lambda_{\mathbf{W}}/4)}}{(1 - \lambda_{\mathbf{W}}/2)^2} = \frac{2}{1 + \sqrt{\lambda_{\mathbf{W}}(1 - \lambda_{\mathbf{W}}/4)}}. \tag{9}$$

It is worth noting that for $\alpha' := \sqrt{\lambda_{\mathbf{W}}(1 - \lambda_{\mathbf{W}}/4)}$ and $1 - \alpha'^2 = (1 - \lambda_{\mathbf{W}}/2)^2$. Moreover, $\alpha_\star = (1 - \alpha'^2)/(1 + \alpha')^2$ and thus

$$\frac{\alpha_\star - 1}{\alpha_\star^2} = \frac{1 - \alpha'^2}{4} = \frac{(1 - \lambda_{\mathbf{W}}/2)^2}{4}. \tag{10}$$

Proposition D.1 (Convergence rate)**.** *Assume $\lambda_{\mathbf{W}} \in (0, 1]$, and let $\mathbf{y}(t)$ be the iterates of* (8) *with $\alpha = \alpha_\star$ given by* (9)*. Then $\alpha_\star \in (1, 2)$ and, for all $t \geq 0$,*

$$\|\mathbf{x}(t) - \bar{a}\mathbf{1}_N\|^2 \leq 2(\alpha_\star - 1)^t\|\mathbf{a} - \bar{a}\mathbf{1}_N\|^2 \leq 2e^{-\sqrt{\lambda_{\mathbf{W}}}t}\|\mathbf{a} - \bar{a}\mathbf{1}_N\|^2.$$

The proof of Proposition D.1, presented below, is inspired by elements of [4, 44].

Lemma D.2 (Polynomial representation)**.** *Define polynomials by $P_0 = 1$, $P_1 = X$ and $P_{t+1} = \alpha_\star X P_t + (1 - \alpha_\star)P_{t-1}$ for $t \geq 1$. Then $\mathbf{y}(t) = P_t(\mathbf{W})\mathbf{a}$, $P_t(1) = 1$, and*

$$\|\mathbf{x}(t) - \bar{a}\mathbf{1}_N\|^2 = \sum_{k=1}^{N-1} |P_t(\lambda_k)|^2 \langle \mathbf{a}, \mathbf{z}_k \rangle^2.$$

Proof. The identity $\mathbf{x}(t) = P_t(\mathbf{W})\mathbf{a}$ holds for $t = 0, 1$, and it propagates through (8) by induction. We have $P_0(1) = P_1(1) = 1$, and if $P_t(1) = P_{t-1}(1) = 1$ then $P_{t+1}(1) = \alpha_\star + 1 - \alpha_\star = 1$. Writing $\mathbf{a} = \bar{a}\mathbf{1}_N + \sum_{k \leq N-1} \langle \mathbf{a}, \mathbf{z}_k \rangle \mathbf{z}_k$ and using $P_t(\mathbf{W})\mathbf{1}_N = \mathbf{1}_N$ and $P_t(\mathbf{W})\mathbf{z}_k = P_t(\lambda_k)\mathbf{z}_k$, we get $\mathbf{x}(t) - \bar{a}\mathbf{1}_N = \sum_{k \leq N-1} P_t(\lambda_k)\langle \mathbf{a}, \mathbf{z}_k \rangle \mathbf{z}_k$. The formula follows from Pythagora's theorem. □

Lemma D.3. *Let $\lambda \in \mathrm{Sp}(\mathbf{W}) \setminus \{1\}$ and $t \geq 0$, then $|P_t(\lambda)|^2 \leq 2(\alpha_\star - 1)^t$.*

Proof. Let $\lambda \in \mathrm{Sp}(\mathbf{W}) \setminus \{1\}$ and $t \geq 1$, the Chebyshev polynomials $P_t(\lambda)$ verify the second-order linear relation [50]:

$$P_{t+1}(\lambda) = \alpha_\star \lambda P_t(\lambda) + (1 - \alpha_\star) P_{t-1}(\lambda). \tag{11}$$

Since the discriminant $\alpha_\star^2 \lambda^2 - 4(\alpha_\star - 1)$ is negative for all eigenvalues, we know the closed form of the relation in Equation (11) is given for all $t \geq 0$ by $P_t(\lambda) = A_\lambda (r_\lambda^+)^t + B_\lambda (r_\lambda^-)^t$, where A_λ, B_λ are two constants and:

$$r_\lambda^\pm = \frac{\alpha_\star \lambda}{2} \pm i\sqrt{\alpha_\star - 1 - \alpha_\star^2 \lambda^2 / 4}.$$

We note that $|r_\lambda^\pm|^2 = \alpha_\star - 1$. We solve the linear system whose constants are given by $P_0(\lambda) = 1$ and $P_1(\lambda) = \lambda$ to find $A_\lambda = (r_\lambda^- - \lambda)/(r_\lambda^- - r_\lambda^+)$ and $B_\lambda = 1 - A_\lambda$. Given the previous expressions, it is obvious that $A_\lambda = \overline{B_\lambda}$ and $r_\lambda^+ = \overline{r_\lambda^-}$. Thus, $|A_\lambda| = |B_\lambda|$ and $|r_\lambda^+| = |r_\lambda^-|$, which leads to:

$$\begin{aligned} |P_t(\lambda)|^2 &\leq 4|A_\lambda|^2 |r_\lambda^+|^{2t} = \frac{4(\alpha_\star - 1)(1 - \lambda^2)}{4(\alpha_\star - 1) - \alpha_\star^2 \lambda^2} (\alpha_\star - 1)^t = \frac{(1 - \alpha'^2)(1 - \lambda^2)}{1 - \alpha'^2 - \lambda^2} (\alpha_\star - 1)^t, \\ &= \frac{(1 - \alpha'^2)(1 - \lambda^2 - \alpha'^2 + \alpha'^2)}{1 - \alpha'^2 - \lambda^2} (\alpha_\star - 1)^t = (1 - \alpha'^2) \left(1 + \frac{\alpha'^2}{1 - \alpha'^2 - \lambda^2}\right) (\alpha_\star - 1)^t. \end{aligned} \tag{12}$$

Since $\lambda^2 \leq (1 - \lambda_{\mathbf{W}})^2 \leq (1 - \alpha'^2)^2$, we can conclude using Equation (12):

$$|P_t(\lambda)|^2 \leq (2 - \alpha'^2)(\alpha_\star - 1)^t \leq 2(\alpha_\star - 1)^t.$$

□

Proof of Proposition D.1. Lemmas D.2 and D.3 give

$$\|\mathbf{x}(t) - \bar{a}\mathbf{1}_N\|^2 = \sum_{k=1}^{N-1} |P_t(\lambda_k)|^2 \langle \mathbf{a}, \mathbf{z}_k \rangle^2 \leq 2(\alpha_\star - 1)^t \sum_{k=1}^{N-1} \langle \mathbf{a}, \mathbf{z}_k \rangle^2 = 2(\alpha_\star - 1)^t \|\mathbf{a} - \bar{a}\mathbf{1}_N\|^2.$$

For the remaining inequalities, note that $\alpha_\star - 1 = \frac{1 - \alpha'}{1 + \alpha'} = \exp(-2 \operatorname{artanh} \alpha')$ and, since $\alpha' \geq 0$, $\operatorname{artanh} \alpha' \geq \alpha'$, so $\alpha_\star - 1 \leq e^{-2\alpha'}$. Since $\lambda_{\mathbf{W}} \leq 1$, we have $1 - \lambda_{\mathbf{W}}/4 \geq 3/4$, hence $\alpha' \geq \frac{\sqrt{3}}{2}\sqrt{\lambda_{\mathbf{W}}}$ and $2\alpha' \geq \sqrt{3}\sqrt{\lambda_{\mathbf{W}}} \geq \sqrt{\lambda_{\mathbf{W}}}$. Therefore $(\alpha_\star - 1)^t \leq e^{-\sqrt{3}\sqrt{\lambda_{\mathbf{W}}}\, t} \leq e^{-\sqrt{\lambda_{\mathbf{W}}}\, t}$. □

D.2 Proof of Theorem 1

We first establish the following auxiliary result, which is useful for the convergence guarantees for both the synchronous and randomized gossip protocols. Throughout, we write $\mathbb{E}_P f = \sum_{\sigma \in \mathfrak{S}_n} P(\sigma) f(\sigma)$ for any measurable map f from $\mathfrak{S}_n$ to a Euclidean space.

Lemma D.4. *For a graph $\mathcal{G} = (\llbracket N \rrbracket, E)$. Let $(\mathbf{x}_v(t))_{t \geq 0,\, v \in \llbracket N \rrbracket}$ be the distributed stochastic process generated by a (ϕ, ψ)-voting rule. Then, for all $t \geq 0$, the average expected Kendall-τ error satisfies*

$$\mathbb{E}\left[\frac{1}{N}\sum_{v=1}^{N} d_\tau\left(\sigma_\star,\, \widehat{\sigma}_v(t)\right)\right] \leq \frac{\binom{n}{2}}{N r_\psi(\bar{\mathbf{x}})^2} \sum_{i \in I} \mathbb{E}\left\|\mathbf{x}^{(i)}(t) - \bar{x}^{(i)}\, \mathbf{1}_N\right\|^2.$$

Proof. Let $t \geq 0$ and a (ϕ, ψ)-voting rule. For node $v \in \llbracket N \rrbracket$, we need to consider the set of per-node events $\Omega(\mathbf{x}_v(t))$ such that the running estimate $\mathbf{x}_v(t)$ is inside the stability radius of estimate $\bar{\mathbf{x}} = \mathbb{E}_{\widehat{P}_N} \phi$ which is assumed to be fixed and where

$$\Omega(\mathbf{x}) := \{\|\mathbf{x} - \bar{\mathbf{x}}\| \leq r_\psi(\bar{\mathbf{x}})\} = \{\|\mathbf{x} - \bar{\mathbf{x}}\|^2 \leq r_\psi^2(\bar{\mathbf{x}})\}.$$

Then, for a node $v \in [\![N]\!]$, the expected Kendall-τ error satisfies the conditional expectation relation:

$$\mathbb{E}\, d_\tau\left(\sigma_\star, \widehat{\sigma}_v(t)\right) = \mathbb{E}\left[d_\tau\left(\sigma_\star, \widehat{\sigma}_v(t)\right) \,|\, \Omega(\mathbf{x}_v(t))\right] \mathbb{P}(\Omega(\mathbf{x}_v(t))) + \mathbb{E}\left[d_\tau\left(\sigma_\star, \widehat{\sigma}_v(t)\right) \,|\, \Omega^c(\mathbf{x}_v(t))\right] \mathbb{P}(\Omega^c(\mathbf{x}_v(t))),$$

where $\Omega^c(\mathbf{x}_v(t))$ is the complementary of $\Omega(\mathbf{x}_v(t))$ in Ω. In $\Omega(\mathbf{x}_v(t))$, the L^2-error of gossip variables is lower than the stability radius for all nodes, then we can conclude that $\mathbb{E}\left[d_\tau\left(\sigma_\star, \widehat{\sigma}_v(t)\right) \,|\, \Omega(\mathbf{x}_v(t))\right] = 0$. Furthermore, because the Kendall-τ distance is upper bounded by $\binom{n}{2}$, then we can upper bound the error as follows:

$$\mathbb{E}\, d_\tau\left(\sigma_\star, \widehat{\sigma}_v(t)\right) \leq \binom{n}{2} \mathbb{P}(\Omega^c(\mathbf{x}_v(t))). \tag{13}$$

Applying Markov's inequality on event $\Omega^c(\mathbf{x}_v(t))$ yields:

$$\mathbb{P}(\Omega^c(\mathbf{x}_v(t))) \leq \frac{\mathbb{E}\left\|\mathbf{x}_v(t) - \bar{\mathbf{x}}\right\|^2}{r_\psi(\bar{\mathbf{x}})^2} = \sum_{i \in I} \frac{\mathbb{E}\left(x_{v,i}(t) - \bar{x}^{(i)}\right)^2}{r_\psi(\bar{\mathbf{x}})^2}. \tag{14}$$

Substituting (14) into (13) and averaging over all nodes $v \in [\![N]\!]$:

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{N}\sum_{v=1}^{N} d_\tau\left(\sigma_\star, \widehat{\sigma}_v(t)\right)\right] &\leq \frac{\binom{n}{2}}{N} \sum_{v=1}^{N} \frac{1}{r_\psi(\bar{\mathbf{x}})^2} \sum_{i \in I} \mathbb{E}\left(x_{v,i}(t) - \bar{x}^{(i)}\right)^2, \\
&= \frac{\binom{n}{2}}{r_\psi(\bar{\mathbf{x}})^2} \sum_{i \in I} \frac{1}{N} \sum_{v=1}^{N} \mathbb{E}\left(x_{v,i}(t) - \bar{x}^{(i)}\right)^2, \\
&= \frac{\binom{n}{2}}{N r_\psi(\bar{\mathbf{x}})^2} \sum_{i \in I} \mathbb{E}\left\|\mathbf{x}^{(i)}(t) - \bar{x}^{(i)} \mathbf{1}_N\right\|^2,
\end{aligned}$$

which completes the proof. □

Proof of Theorem 1. Let $t \geq 0$ and $i \in I$. In Algorithm 1, the auxiliary variable $\mathbf{x}'_v$ stores the previous estimate of node v, the first round is a plain gossip step, and the subsequent rounds use the memory coefficient $\alpha_\star$. Hence the vector $\mathbf{x}^{(i)}(t) \in \mathbb{R}^N$ of the i-th coordinates across all nodes follows the shift-register iteration (8) with $\mathbf{a} = \mathbf{x}^{(i)}(0)$ and $\bar{a} = \bar{x}^{(i)}$:

$$\mathbf{x}^{(i)}(1) = \mathbf{W}\mathbf{x}^{(i)}(0), \qquad \mathbf{x}^{(i)}(t+1) = \alpha_\star \mathbf{W}\mathbf{x}^{(i)}(t) + (1-\alpha_\star)\mathbf{x}^{(i)}(t-1), \quad t \geq 1.$$

Since $\lambda_{\mathbf{W}} \in (0, 1]$, Proposition D.1 gives

$$\|\mathbf{x}^{(i)}(t) - \bar{x}^{(i)}\mathbf{1}_N\|^2 \leq 2\, e^{-\sqrt{\lambda_{\mathbf{W}}}\, t}\, \|\mathbf{x}^{(i)}(0) - \bar{x}^{(i)}\mathbf{1}_N\|^2. \tag{15}$$

The synchronous process is deterministic, so the expectation in Lemma D.4 is trivial. Combining Lemma D.4 with (15) summed over $i \in I$, we obtain

$$\frac{1}{N}\sum_{v=1}^{N} d_\tau\left(\sigma_\star, \widehat{\sigma}_v(t)\right) \leq \frac{\binom{n}{2}}{N r_\psi(\bar{\mathbf{x}})^2} \sum_{i \in I} \|\mathbf{x}^{(i)}(t) - \bar{x}^{(i)}\mathbf{1}_N\|^2 \leq \frac{2\binom{n}{2}}{N r_\psi(\bar{\mathbf{x}})^2}\, e^{-\sqrt{\lambda_{\mathbf{W}}}\, t}\, \gamma. \qquad \square$$

D.3 Proof of Remark 1

Proof. Since d_τ is integer-based, we note the equivalence between $N^{-1}\sum_{v=1}^{N} d_\tau\left(\sigma_\star, \widehat{\sigma}_v(t)\right) = 0$ and $N^{-1}\sum_{v=1}^{N} d_\tau\left(\sigma_\star, \widehat{\sigma}_v(t)\right) < \frac{1}{N}$. Applying Theorem 1 yields a sufficient condition for the previous condition on the average of the Kendall-τ distances to happen:

$$\frac{2\binom{n}{2}}{N r_\psi(\bar{\mathbf{x}})^2} e^{-\sqrt{\lambda_{\mathbf{W}}}\, t} \gamma < \frac{1}{N}.$$

This condition holds for every integer $t > t^\star$, where $t^\star := \lambda_{\mathbf{W}}^{-1/2} \log\left(\frac{2\binom{n}{2}\gamma}{r_\psi^2(\bar{\mathbf{x}})}\right)$. If $t^\star < 0$, it already holds at $t = 0$. Otherwise, the smallest integer $t > t^\star$ is $\lfloor t^\star \rfloor + 1$. Since $T^{\mathrm{Sync}}(0)$ is the smallest time such that the average of the Kendall-τ distance is null, we obtain the desired bound $T^{\mathrm{Sync}}(0) \leq \max\{0, \lfloor t^\star \rfloor + 1\}$. □

D.4 Proof of Theorem 2

Proof. Let $t \geq 0$. Applying Lemma D.4 directly yields:

$$\mathbb{E}\left[\frac{1}{N}\sum_{v=1}^{N} d_\tau\left(\sigma_\star, \widehat{\sigma}_v(t)\right)\right] \leq \frac{\binom{n}{2}}{N r_\psi(\bar{\mathbf{x}})^2} \sum_{i \in I} \mathbb{E}\left\|\mathbf{x}^{(i)}(t) - \bar{x}^{(i)}\,\mathbf{1}_N\right\|^2. \tag{16}$$

We then apply Lemma B.1 to bound each term in the right-hand side sum:

$$\sum_{i \in I} \mathbb{E}\left\|\mathbf{x}^{(i)}(t) - \bar{x}^{(i)}\,\mathbf{1}_N\right\|^2 \leq \gamma\left(1 - \frac{\lambda(p)}{2}\right)^t. \tag{17}$$

Combining Equation (16) and Equation (17) gives:

$$\mathbb{E}\left[\frac{1}{N}\sum_{v=1}^{N} d_\tau\left(\sigma_\star, \widehat{\sigma}_v(t)\right)\right] \leq \frac{\binom{n}{2}\gamma}{N r_\psi(\bar{\mathbf{x}})^2}\left(1 - \frac{\lambda(p)}{2}\right)^t.$$

The conclusion follows from the convex inequality $1 - \lambda(p)/2 \leq \exp(-\lambda(p)/2)$. □

D.5 Proof of Remark 3

Proof. For the upper bound, we use similar arguments as the ones used for Remark 1. By Theorem 2 and since $N \geq 1$, the expected average Kendall-τ error is at most $\binom{n}{2}\gamma\, e^{-\lambda(p)t/2}/r_\psi^2(\bar{\mathbf{x}})$, which is at most δ for every integer $t > t^\star_{\mathrm{rand}}$, where

$$t^\star_{\mathrm{rand}} := \frac{2}{\lambda(p)} \log\left(\binom{n}{2}\frac{\gamma}{\delta r_\psi^2(\bar{\mathbf{x}})}\right).$$

If $t^\star_{\mathrm{rand}} < 0$, this holds at $t = 0$. Otherwise, the smallest such integer is $\lfloor t^\star_{\mathrm{rand}} \rfloor + 1$. Hence $T^{\mathrm{rand}}(\delta) \leq \max\{0, \lfloor t^\star_{\mathrm{rand}} \rfloor + 1\}$.

For the lower bound, we note that at each iteration $t+1 \geq 0$ of the randomized gossip protocol only two nodes share their embedding and we have:

$$\frac{1}{N}\sum_{v=1}^{N} d_\tau(\widehat{\sigma}_v(t+1), \sigma_\star) - \frac{1}{N}\sum_{v=1}^{N} d_\tau(\widehat{\sigma}_v(t), \sigma_\star) \geq -\frac{n(n-1)}{N}.$$

Telescoping the sums and taking the expectation yields:

$$\frac{1}{N}\sum_{v=1}^{N} \mathbb{E}\, d_\tau(\widehat{\sigma}_v(t+1), \sigma_\star) - \frac{1}{N}\sum_{v=1}^{N} d_\tau(\widehat{\sigma}_v(0), \sigma_\star) \geq -\frac{n(n-1)}{N}t.$$

At iteration $T^{\mathrm{rand}}(\delta)$, we can rewrite the previous inequality as $\delta \geq N^{-1}\sum_{v=1}^{N} d_\tau(\widehat{\sigma}_v(0), \sigma_\star) - n(n-1)T^{\mathrm{rand}}(\delta)/N$, which yields

$$T^{\mathrm{rand}}(\delta) \geq \max\left(\frac{\sum_{v=1}^{N} d_\tau(\widehat{\sigma}_v(0), \sigma_\star) - N\delta}{n(n-1)}, 0\right).$$

□

D.6 Proof of Proposition 1

Proof. We consider a score function S that admits a (ϕ, ψ)-decomposition. Let $1 \leq \delta \leq \binom{n}{2}$. We now restate the empirical breakdown function and define the associated set $\mathcal{E}$:

$$\varepsilon^\star_{\widehat{P}_N, S}(\delta) = \inf \mathcal{E} := \inf\left\{\frac{b}{N}\,\middle|\, b \in [\![N]\!],\ \sup_{\widehat{P}^\varepsilon_N \in \tilde{B}(\widehat{P}_N, b/N)} d_\tau\left(S\left(\widehat{P}_N\right), S\left(\widehat{P}^\varepsilon_N\right)\right) \geq \delta\right\}.$$

Let $\varepsilon \in \mathcal{E}$,

$$\begin{aligned}\varepsilon \in \mathcal{E} &\iff \exists\, \widehat{P}_N^{\varepsilon} \in \Delta_{\mathfrak{S}_n} \text{ such that } d_{\mathrm{TV}}(\widehat{P}_N, \widehat{P}_N^{\varepsilon}) \leq \varepsilon, d_\tau(\psi(\mathbb{E}_{\widehat{P}_N^{\varepsilon}}\phi), \psi(\mathbb{E}_{\widehat{P}_N}\phi)) \geq \delta, \\ &\implies \exists\, \widehat{P}_N^{\varepsilon} \in \Delta_{\mathfrak{S}_n} \text{ such that } d_{\mathrm{TV}}(\widehat{P}_N, \widehat{P}_N^{\varepsilon}) \leq \varepsilon, \|\mathbb{E}_{\widehat{P}_N^{\varepsilon}}\phi - \mathbb{E}_{\widehat{P}_N}\phi\| \geq r_\psi\left(\mathbb{E}_{\widehat{P}_N}\phi\right). \quad (18)\end{aligned}$$

Furthermore, we can bound the norm of the difference of expectations:

$$\begin{aligned}\left\|\mathbb{E}_{\widehat{P}_N^{\varepsilon}}\phi - \mathbb{E}_{\widehat{P}_N}\phi\right\| &\leq \left\|\sum_{\sigma\in\mathfrak{S}_n}\left[\widehat{P}_N^{\varepsilon}(\sigma) - \widehat{P}_N(\sigma)\right]\phi(\sigma)\right\| \leq \sum_{\sigma\in\mathfrak{S}_n}|\widehat{P}_N^{\varepsilon}(\sigma) - \widehat{P}_N(\sigma)|\|\phi(\sigma)\| \\ &\leq 2d_{\mathrm{TV}}(\widehat{P}_N, \widehat{P}_N^{\varepsilon}) \max_{\sigma\in\mathfrak{S}_n}\|\phi(\sigma)\| \leq 2\varepsilon\,|||\phi|||\,. \quad (19)\end{aligned}$$

Using both Equation (18) and Equation (19):

$$r_\psi\left(\mathbb{E}_{\widehat{P}_N}\phi\right) \leq \left\|\mathbb{E}_{\widehat{P}_N^{\varepsilon}}\phi - \mathbb{E}_{\widehat{P}_N}\phi\right\| \leq 2\varepsilon\,|||\phi|||\,,$$

which implies $\varepsilon \geq \frac{r_\psi\left(\mathbb{E}_{\widehat{P}_N}\phi\right)}{2|||\phi|||}$. Taking the infimum on E allows us to conclude on the lower bound. □

D.7 Proof of Theorem 3

Let $\delta > 0$ and a given set $B \subseteq [\![N]\!]$ of size $b = |B|$ and $H = [\![N]\!] \setminus B$. The gossip dynamics are governed by the cocycle $\mathbf{W}_{0:t} := \prod_{s=1}^{t}\mathbf{W}_s := \mathbf{W}_t \cdots \mathbf{W}_1$, where each $\mathbf{W}_s$ is a gossip matrix, drawn independently with $\mathbb{E}\mathbf{W}_s = \overline{\mathbf{W}}$ for all $s \geq 1$. The breakdown event is defined in terms of the consensus output $\bar{\mathbf{x}}_h(t) := \frac{1}{N-b}\sum_{v\in H}\mathbf{x}_v(t)$ compared to the initial honest consensus $\bar{\mathbf{x}}_h = \frac{1}{N-b}\sum_{v\in H}\mathbf{x}_v(0)$ is governed by the stability radius $r_\psi(\bar{\mathbf{x}}_h)$. Let $\mathbf{y}_u = \mathbf{x}_u(0) - \bar{\mathbf{x}}_h$ be the initial error at node $u \in [\![N]\!]$. The strategy of the proof is to make appear the fraction $\frac{b}{N}$ in an upper bound of the expectation term of Equation (29) using the decomposition:

$$\mathbb{E}\,\|\bar{\mathbf{x}}_h(t) - \bar{\mathbf{x}}\| \leq \mathbb{E}\,\|\bar{\mathbf{x}}_h(t) - \bar{\mathbf{x}}_h\| + \|\bar{\mathbf{x}}_h - \bar{\mathbf{x}}\|\,. \quad (20)$$

Furthermore, by writing $\bar{\mathbf{x}}$ as a barycenter of $\bar{\mathbf{x}}_h$ and $\bar{\mathbf{x}}_b$, one can find

$$\|\bar{\mathbf{x}}_h - \bar{\mathbf{x}}\| = \left\|\bar{\mathbf{x}}_h - \frac{b}{N}\bar{\mathbf{x}}_b - \frac{N-b}{N}\bar{\mathbf{x}}_h\right\| = \frac{b}{N}\|\bar{\mathbf{x}}_h - \bar{\mathbf{x}}_b\| \leq 2\frac{b}{N}\,|||\phi|||\,. \quad (21)$$

Thus, the only difficulty is to bound the term $\mathbb{E}\,\|\bar{\mathbf{x}}(t) - \bar{\mathbf{x}}_h\|$. In order to achieve so, we consider a second decomposition. Since $\mathbf{W}_{0:t}$ is doubly stochastic and initial values are fixed, we write:

$$\begin{aligned}\bar{\mathbf{x}}_h(t) - \bar{\mathbf{x}}_h &= \frac{1}{N-b}\sum_{v\in H}\mathbf{x}_v(t) - \bar{\mathbf{x}}_h = \frac{1}{N-b}\sum_{(u,v)\in[\![N]\!]\times H}[\mathbf{W}_{0:t}]_{vu}(\mathbf{x}_v(0) - \bar{\mathbf{x}}_h), \\ &= \underbrace{\frac{1}{N-b}\sum_{(u,v)\in H^2}[\mathbf{W}_{0:t}]_{vu}\mathbf{y}_u}_{=:\mathbf{y}_H} + \underbrace{\frac{1}{N-b}\sum_{(w,v)\in B\times H}[\mathbf{W}_{0:t}]_{vw}\mathbf{y}_w}_{=:\mathbf{y}_B}. \quad (22)\end{aligned}$$

The strategy is now to find a way to bound $\mathbf{y}_H$ which represent the uncontaminated dynamic of the process, while $\mathbf{y}_B$ denotes the effect of the added contamination over the rest of the nodes. We note that the non-nullity of $\mathbf{y}_H$ arises from the assumption that $\mathcal{G}_H$ is connected. Lemma D.5 bounds the first term while Lemma D.6 bounds the effect of the contamination on the error of the process.

Lemma D.5. *Given the definition of $\mathbf{y}_H$ in Equation* (22)*, then:*

$$\mathbb{E}\|\mathbf{y}_H\| \leq 2\,|||\phi|||\,\sqrt{\frac{b}{N}}\xi_t.$$

Proof. Let $\mathbf{h} := \frac{1}{N-b}\mathbf{1}_H$ be the normalized indicator of honest nodes. Swapping the order of summation gives:

$$\mathbf{y}_H = \sum_{u\in H}\left(\frac{1}{N-b}\sum_{v\in H}[\mathbf{W}_{0:t}]_{vu}\right)\mathbf{y}_u = \sum_{u\in H}[\mathbf{W}_{0:t}\mathbf{h}]_u\mathbf{y}_u. \quad (23)$$

The right hand-side term is the weighted sum of the vector family $(\mathbf{y}_u)_{u\in H}$, whose weights are $a_u := [\mathbf{W}_{0:t}\mathbf{h}]_u$ for $u \in H$. Given a scalar family $(a_u)_{u\in H}$ and a vector family $(\mathbf{y}_u)_{u\in H}$, we recall that using Cauchy-Schwarz inequality twice yields the following inequality:

$$\left\|\sum_{u\in H} a_u\mathbf{y}_u\right\|^2 = \sum_{(u,v)\in H^2} a_u a_v\langle \mathbf{y}_u, \mathbf{y}_v\rangle \leq \sum_{(u,v)\in H^2} a_u a_v\|\mathbf{y}_u\|\|\mathbf{y}_v\|$$
$$= \left(\sum_{u\in H} a_u\|\mathbf{y}_u\|\right)\left(\sum_{u\in H} a_u\|\mathbf{y}_u\|\right) \leq \left(\sum_{u\in H} a_u^2\right)\left(\sum_{u\in H}\|\mathbf{y}_u\|^2\right). \quad (24)$$

Since $\sum_{u\in H}\mathbf{y}_u = 0$ we may again remove $1/N$ to Equation (23), and Equation (24) gives:

$$\|\mathbf{y}_H\|^2 \leq \underbrace{\sum_{u\in H}\left((\mathbf{W}_{0:t}^\top\mathbf{h})_u - \tfrac{1}{N}\right)^2}_{=\left\|\left[\mathbf{W}_{0:t}^\top - \mathbf{z}_N\mathbf{z}_N^\top\right]\mathbf{h}\right\|^2}\left(\sum_{u\in H}\|\mathbf{y}_u\|^2\right).$$

We now can upper bound the norm on the subspace $\mathrm{span}(\mathbf{1}_N)^\perp$. Following Lemma B.2, $\bar{\mathbf{W}}$ is positive semi-definite and thus its spectral gap is given as $\xi_1^2 = 1 - \lambda_{\bar{\mathbf{W}}} = \lambda_{N-1}(\bar{\mathbf{W}})$. We consider the filtration introduced in Appendix B to upper bound the following expectation. We note that $(\mathbf{z}_N\mathbf{z}_N^\top)\mathbf{h} = \langle \mathbf{h}, \mathbf{z}_N\rangle\mathbf{z}_N = N^{-1}\mathbf{1}_N$, hence:

$$\mathbb{E}\left\|\left[\mathbf{W}_{0:t}^\top - \mathbf{z}_N\mathbf{z}_N^\top\right]\mathbf{h}\right\|^2 = \mathbb{E}\left\|\mathbf{W}_{0:t}^\top\left[\mathbf{h} - \frac{1}{N}\mathbf{1}_N\right]\right\|^2,$$
$$\text{(By [6])} \leq \lambda_{N-1}(\bar{\mathbf{W}})^t\left\|\mathbf{h} - \frac{1}{N}\mathbf{1}_N\right\|^2 = \xi_t^2\frac{b}{N(N-b)}.$$

Combined with $\sum_{u\in H}\|\mathbf{y}_u\|^2 \leq 4|H|\,|||\phi|||^2$, we find $\mathbb{E}\|\mathbf{y}_H\|^2 \leq 4\xi_t^2\frac{b}{N}|||\phi|||^2$, and Jensen's inequality on the square root function gives $\mathbb{E}\|\mathbf{y}_H\| \leq 2\xi_t\sqrt{\frac{b}{N}}\,|||\phi|||$ concluding the proof. □

Lemma D.6. *Given the definition of $\mathbf{y}_B$ in Equation* (22)*, then:*

$$\mathbb{E}\|\mathbf{y}_B\| \leq 2\,|||\phi|||\,\frac{b}{N}\zeta_t. \quad (25)$$

Proof. Let $B \subseteq [\![N]\!]$ such that $|B| = b$ and $H = [\![N]\!] \setminus B$. Let $(\mathbf{z}_k)_{k=1}^N$ be the eigenvectors associated with the eigenvalues $(\lambda_k(\bar{\mathbf{W}}))_{k=1}^N$. By the spectral theorem, $\bar{\mathbf{W}} = \sum_{k=1}^N \lambda_k(\bar{\mathbf{W}})\,\mathbf{z}_k\mathbf{z}_k^\top$. Let us denote $\mathbf{1}_H, \mathbf{1}_B$ respectively the binary indicator of nodes in H and in B. Setting $\hat{b}_k := \langle \mathbf{z}_k, \mathbf{1}_B\rangle$ and $\hat{h}_k := \langle \mathbf{z}_k, \mathbf{1}_H\rangle$, we can derive the following equation.

$$Q_t(H) := \frac{1}{N-b}\sum_{(v,w)\in H\times B}[\bar{\mathbf{W}}^t]_{vw},$$
$$= \frac{1}{N-b}\langle \mathbf{1}_H, \bar{\mathbf{W}}^t\mathbf{1}_B\rangle,$$
$$= \frac{1}{N-b}\sum_{k=1}^N \lambda_k(\bar{\mathbf{W}})^t\langle \mathbf{z}_k, \mathbf{1}_B\rangle\langle \mathbf{1}_H, \mathbf{z}_k\rangle,$$
$$= \frac{1}{N-b}\sum_{k=1}^N \lambda_k(\bar{\mathbf{W}})^t\,\hat{h}_k\hat{b}_k. \quad (26)$$

The $k = N$ terms can be computed as $\hat{h}_N = \langle \mathbf{z}_N, \mathbf{1}_H\rangle = (N-b)/\sqrt{N}$ and $\hat{b}_N = b/\sqrt{N}$. For $k \leq N-1$, the eigenbasis is orthogonal and $\langle \mathbf{z}_k, \mathbf{z}_N\rangle = \frac{1}{\sqrt{N}}\langle \mathbf{z}_k, \mathbf{1}_H\rangle + \frac{1}{\sqrt{N}}\langle \mathbf{z}_k, \mathbf{1}_B\rangle = 0$ gives $\hat{b}_k = -\hat{h}_k$, so:

$$Q_t(H) = \frac{b}{N} - \frac{1}{N-b}\sum_{k=1}^{N-1}\lambda_k^t(\bar{\mathbf{W}})\,\hat{h}_k^2.$$

Using Pythagora's theorem, one can show that $\sum_{k=1}^{N-1} \hat{h}_k^2 = \|\mathbf{1}_H\|^2 - \hat{h}_N^2 = (N-b) - \hat{h}_N^2 = \frac{b(N-b)}{N}$. Knowing that for all $k \in [\![N]\!]$, $\lambda_k(\bar{\mathbf{W}}) \geq \lambda_1(\bar{\mathbf{W}}) \geq 0$ under semi-definite positiveness of $\bar{\mathbf{W}}$, $Q_t(H)$ is upper bounded by $\frac{b}{N}\zeta_t$. Then, by definition of $\mathbf{y}_B$ and the triangle inequality:

$$\mathbb{E}\|\mathbf{y}_B\| \leq \frac{1}{N-b} \sum_{(v,w)\in H\times B} \mathbb{E}[\mathbf{W}_{0:t}]_{vw} \|\mathbf{y}_w\| \leq 2\,|||\phi|||\; Q_t(H) \leq 2\,|||\phi|||\; \frac{b}{N}\zeta_t.$$

□

Proof of Theorem 3. Let $\delta > 0$ and a given score with (ϕ, ψ)-decomposition. Lemma 1 gives that the event $\{d_\tau\left([\psi(\bar{\mathbf{x}}(t))]_{\mathfrak{S}_n}, \sigma_\star\right) \geq \delta\}$ is included in $\Omega^c(\bar{\mathbf{x}}(t))$. Taking the expectation yields the standard bound:

$$\begin{aligned}\mathbb{E}\,\|\bar{\mathbf{x}}(t) - \bar{\mathbf{x}}\| &\geq \mathbb{E}[\|\bar{\mathbf{x}}(t) - \bar{\mathbf{x}}\|\,\mathbb{I}\left\{d_\tau([\psi(\bar{\mathbf{x}}(t))]_{\mathfrak{S}_n}, \sigma_\star) \geq 1\right\}] \\ &\geq \mathbb{P}\left(d_\tau([\psi(\bar{\mathbf{x}}(t))]_{\mathfrak{S}_n}, \sigma_\star) \geq 1\right) r_\psi(\bar{\mathbf{x}}). \quad (27)\end{aligned}$$

Since the Kendall-τ distance is upper bounded by $\binom{n}{2}$, we can then derive the following:

$$\mathbb{E}\, d_\tau([\psi(\bar{\mathbf{x}}(t))]_{\mathfrak{S}_n}, \sigma_\star) = \sum_{k=1}^{\binom{n}{2}} k\,\mathbb{P}(d_\tau([\psi(\bar{\mathbf{x}}(t))]_{\mathfrak{S}_n}, \sigma_\star) = k) \leq \binom{n}{2} \mathbb{P}(d_\tau([\psi(\bar{\mathbf{x}}(t))]_{\mathfrak{S}_n}, \sigma_\star) \geq 1). \quad (28)$$

We consider Equation (27). By dividing both sides by $\binom{n}{2}$ in Equation (28) in order to isolate the probability term and using the assumption $\mathbb{E}\, d_\tau([\psi(\bar{\mathbf{x}}(t))]_{\mathfrak{S}_n}, \sigma_\star) \geq \delta$, we can derive:

$$\mathbb{E}\,\|\bar{\mathbf{x}}(t) - \bar{\mathbf{x}}\| \geq \frac{\delta}{\binom{n}{2}} r_\psi(\bar{\mathbf{x}}). \quad (29)$$

Thanks to the triangle inequality, we can upper bound Equation (22) and take the expectation $\mathbb{E}\,\|\bar{\mathbf{x}}(t) - \bar{\mathbf{x}}_h\| \leq \mathbb{E}\|\mathbf{y}_H\| + \mathbb{E}\|\mathbf{y}_B\|$. Applying Lemma D.5 and Lemma D.6 back to the previous decomposition yields:

$$r_\psi(\bar{\mathbf{x}}_h)\frac{\delta}{\binom{n}{2}} \leq \mathbb{E}\,\|\bar{\mathbf{x}}_h(t) - \bar{\mathbf{x}}_h\| \leq \mathbb{E}\|\mathbf{y}_H\| + \mathbb{E}\|\mathbf{y}_B\| + \|\bar{\mathbf{x}}_h - \bar{\mathbf{x}}\| \leq 2\,|||\phi|||\; \left(\frac{b}{N}(1+\zeta_t) + \sqrt{\frac{b}{N}\xi_t}\right).$$

Solving the quadratic inequality verified by $\sqrt{b/N}$ yields the following condition:

$$\varepsilon_t^H \geq \frac{\left(-\xi_t + \sqrt{\xi_t^2 + 4\varepsilon_- \delta(1+\zeta_t)/\binom{n}{2}}\right)^2}{4(1+\zeta_t)^2}.$$

□

E Extensions to partial rankings

In many practical settings, an agent $v \in [\![N]\!]$ ranks only a subset of $k_v < n$ alternatives, either because the full set is too large to evaluate, or because only a portion is relevant to the agent. Formally, a partial ranking by agent v is an injection $\sigma_v : [\![k_v]\!] \to [\![n]\!]$, where $\sigma_v(i) = a$ means that alternative a is assigned rank i by agent v. We denote by $A_v \subseteq [\![n]\!]$ the set of alternatives ranked by v, with $|A_v| = k_v$.

Completing partial rankings. An approach consists of imputing the average rank $(k_v + 1 + n)/2$ to all unranked items [35], where k_v is the number of ranked alternatives by node v. Alternatively, one can normalize ranked items as $\sigma(i)_{\text{scaled}} = (\sigma(i) - 1)/k$ and assign a normalized rank of 1 to unranked items [35]. Borda scores are then computed from these completed or normalized rankings in the usual way. For Copeland's method, we use *weak pairwise probabilities* $p_{ij} = \mathbb{I}\{\sigma(i) < \sigma(j)\} + (1/2)\,\mathbb{I}\{\sigma(i) = \sigma(j)\}$.

Modified embedding and stability radius. Under any of the above completion schemes, the embedding ϕ is applied to the completed ranking $\widetilde{\sigma}_v$ (not necessarily in $\mathfrak{S}_n$), yielding a local state $\mathbf{x}_v(0) := \phi(\widetilde{\sigma}_v) \in \mathbb{R}^I$. We denote by $J \subseteq I$ the index set of coordinates corresponding to pairs (i,j) such that at least one agent has ranked both i and j, and restrict attention to these. We define the *reduced initial dispersion* $\gamma^+ := \sum_{i\in J} \left\|\mathbf{x}^{(i)}(0) - \bar{x}^{(i)}\mathbf{1}_N\right\|^2$, where $\bar{x}^{(i)} = N^{-1}\sum_{v=1}^N [\phi(\widetilde{\sigma}_v)]_i$ is the population average over the completed embeddings for coordinate i. Similarly, since imputed coordinates are shared across agents and cancel in the averaging, the *reduced stability radius* is

$$r_\psi^+(\bar{\mathbf{x}}) := \inf\left\{r > 0 \,\middle|\, \exists\, \mathbf{y} \in \mathbb{R}^I, \|\mathbf{y}_J - \bar{\mathbf{x}}_J\| \le r, d_\tau\left(\psi(\mathbf{y}),\, \psi(\bar{\mathbf{x}})\right) \ge 1\right\},$$

i.e. the stability radius measured only over the informative coordinate directions. Since imputed coordinates are shared across agents they cancel in the averaging, and one has $r_\psi^+(\bar{\mathbf{x}}) \le r_\psi(\bar{\mathbf{x}})$, reflecting a degradation in ranking identifiability relative to the complete-ranking setting.

Corollary E.1 (Convergence under partial rankings)**.** *Under the same assumptions as Theorems 1 and 2, the average expected Kendall-τ error satisfies*

$$\frac{1}{N}\sum_{v=1}^N \mathbb{E}\, d_\tau(\sigma_\star, \widehat{\sigma}_v(t)) \le \frac{\binom{n}{2}}{N r_\psi^+(\bar{\mathbf{x}})^2}\gamma^+\rho(t), \tag{30}$$

where $\rho(t) = e^{-\sqrt{\lambda_{\mathbf{W}}}t}$ under Algorithm 1 and $\rho(t) = e^{-\lambda(p)t/2}$ under Algorithm 2. The bound degrades relatively to the complete ranking setting through the reduced stability radius $r_\psi^+(\bar{\mathbf{x}}) \le r_\psi(\bar{\mathbf{x}})$.

Proof sketch. The proof mirrors that of Theorems 1 and 2, with two modifications. First, in the application of Lemma D.4, the Markov-inequality step linking L^2-error to Kendall-τ error, the summation over I is restricted to J, and the stability radius $r_\psi(\bar{\mathbf{x}})$ is replaced by its reduced counterpart $r_\psi^+(\bar{\mathbf{x}})$, since ranking changes can only be triggered by perturbations in the informative coordinate directions. The remainder of the argument is identical, giving the stated bounds with γ^+ in place of γ. Second, the gossip averaging is applied to the completed embeddings $(\mathbf{x}_v(t))_{v\in[\![N]\!]}$, the gossip update apply verbatim in both synchronous and asynchronous cases, yielding the same exponential decay of $\|\mathbf{x}^{(i)}(t) - \bar{x}^{(i)}\mathbf{1}_N\|^2$ for every $i \in J$. □

In the extreme case $k_v = 1$ for all $v \in [\![N]\!]$ (each agent ranks a single alternative), the imputed Borda scores concentrate near $(n+1)/2$ and $r_\psi^+(\bar{\mathbf{x}}) = 0$, consistent with the loss of ranking identifiability.

F Further considerations for Borda consensus

F.1 Borda scoring as a median estimation

We identify each permutation $\sigma \in \mathfrak{S}_n$ with its vector representation in $\mathbb{R}^n$, namely $\phi_B(\sigma) := (\sigma(1), \ldots, \sigma(n))^\top$. We begin with the Spearman ρ distance:

$$d_\rho(\sigma, \sigma') = \sum_{j=1}^n \big(\sigma(j) - \sigma'(j)\big)^2 = \|\phi_B(\sigma) - \phi_B(\sigma')\|^2.$$

Let $\widehat{P}_N$ be an empirical distribution over $\mathfrak{S}_n$. Consider the Fréchet functional associated with d_ρ as stated in Equation (1), $\mathbf{s}_B = \operatorname{argmin}_{\mathbf{s}\in\mathbb{R}^n} N^{-1}\sum_{v=1}^N \|[\sigma_v] - \mathbf{s}\|^2$. Since the squared Euclidean norm is strictly convex, this minimization problem admits a unique solution. Taking the gradient with respect to $\mathbf{s}$ and setting it to zero yields the first-order condition:

$$\nabla_{\mathbf{s}}\left[\frac{1}{N}\sum_{v=1}^N \|\phi_B(\sigma_v) - \mathbf{s}\|^2\right] = -\frac{2}{N}\sum_{v=1}^N(\phi_B(\sigma_v) - \mathbf{s}) = 0,$$

which gives $\mathbf{s}_B := N^{-1}\sum_{v=1}^N \phi_B(\sigma_v)$. This procedure coincides with the first order estimator for Borda score.

F.2 Borda stability radius.

We now analyze the robustness of the induced ranking for the Borda method. Let $\mathbf{s} \in \mathbb{R}^n$ be the score vector associated with $\widehat{P}_N$. Assume that all coordinates of $\mathbf{s}$ are distinct. Let $(i,j) \in \arg\min_{i' \neq j'} |s_{i'} - s_{j'}|$, $\Delta := |s_i - s_j|$. We can use an alternative definition for the stability radius:

$$r_{\mathrm{Id}}(\mathbf{s}) = \inf \{\|\mathbf{y} - \mathbf{s}\| \mid \mathbf{y} \in \mathbb{R}^n,\, d_\tau(\mathbf{y}, \mathbf{s}) \geq 1\}\,.$$

A ranking changes if and only if there exists $i' \neq j'$ such that $(y_{i'} - y_{j'})(s_{i'} - s_{j'}) \leq 0$, *i.e.*, $\mathbf{y}$ crosses the hyperplane $H_{i'j'} := \{\mathbf{y} \in \mathbb{R}^n \mid y_{i'} = y_{j'}\}$. The ℓ_2-distance from $\mathbf{s}$ to $H_{i'j'}$ is

$$d_2(\mathbf{s}, H_{i'j'}) = \frac{|s_{i'} - s_{j'}|}{\|\mathbf{e}_{i'} - \mathbf{e}_{j'}\|} = \frac{|s_{i'} - s_{j'}|}{\sqrt{2}}.$$

The stability radius is therefore the minimum of this quantity over all pairs $i' \neq j'$, which is achieved at (i,j), and we conclude:

$$r_{\mathrm{Id}}(\mathbf{s}) = \frac{\Delta}{\sqrt{2}}. \tag{31}$$

G Further considerations for Copeland consensus

G.1 Copeland as a tournament-based voting rule.

Given a multiset $\{\sigma_1, \dots \sigma_n\} \in \mathrm{Sym}^N(\mathfrak{S}_n)$, with the associated pairwise probabilities $\widehat{p}_{ij} = N^{-1} \sum_{v=1}^N \mathbb{I}\{\sigma_v(i) < \sigma_v(j)\}$. Copeland scores stated in Equation (2) can be rewritten as:

$$s_i := [S_C(\widehat{P}_N)]_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} \mathrm{sgn}(\widehat{p}_{ij} - 1/2),$$

which is generally considered as a tournament-based score, each element i faces all the other contestants $j \in [\![n]\!] \setminus \{i\}$ and $+1$ is added to its score whenever i wins over j (and thus $\widehat{p}_{ij} > 1/2$) and -1 when it loses ($\widehat{p}_{ij} < 1/2$). The minimum score gap $\Delta = \min_{i \neq j} |s_i - s_j| \in \{0, 2\}$. Indeed, the difference between two scores is even. Let $i, j \in [\![n]\!]$ and $i \neq j$, and let us denote their respective number of wins as w_i, w_j where $w_k = \sum_{j \neq k} \mathbb{I}\{\widehat{p}_{kj} > 1/2\}$ for all $k \in [\![n]\!]$. We know that $s_i = 2w_i - (n-1)$ and $s_j = 2w_j - (n-1)$, if we do the difference of the two we find that 2 is a divisor of $|s_i - s_j|$ which allows us to conclude that 2 divides Δ. We can also see that the minimum score gap respects $\Delta < 4$. Indeed, if we assume $\Delta \geq 4$, then for a given pair (i,j) realizing Δ, *i.e.* $\Delta = |s_i - s_j|$, and without loss of generality, let us assume $s_i < s_j$, then j has two more wins than i. As $\Delta \geq 4$, this is true for all successive pairs of score components, and thus the total number of wins of the element with greatest score is greater than $2(n-1)$ which is absurd as they are only $n-1$ contestants. Thus, $\Delta < 4$ and $\Delta \in \{0, 2\}$.

Remark G.1. *Furthermore, if for all pairs (i,j) such that $i \neq j$, $\widehat{p}_{ij} \neq 1/2$ and $\widehat{P}_N$ is strict stochastic transitive (SST), then we can ensure that $\Delta = 2$. Indeed, for a pair (i,j), we can assume without loss of generality that $\widehat{p}_{ij} > 1/2$. Then for all $k \in [\![n]\!]$ such that $\widehat{p}_{ki} > 1/2$, then by SST, $\widehat{p}_{kj} > 1/2$. Thus, j has at least as many losses as i against other opponents and loses directly against i meaning that $s_j^C < s_i^C$.*

G.2 Copeland score stability radius

We now analyze the robustness of the induced ranking for the Copeland method. Let $\mathbf{s} \in \mathbb{R}^n$ be the score vector associated with P. Assume that all coordinates of $\mathbf{s}$ are distinct. Let $(i,j) \in \arg\min_{i' < j'} |s_{i'} - s_{j'}|$, $\Delta := |s_i - s_j|$. Without loss of generality, assume that $p_{ij} < 1/2$, and thus $s_i < s_j$. In order to swap scores, we need to add $|p_{ij} - 1/2|$ to p_{ij}. If we denote $\mathbf{s}'$ the corrupted score vector, then $s'_i = s_i + 2$ and $s'_j = s_j - 2$. A score swap occurs when $s'_i - s'_j > 0$ and:

$$s'_i - s'_j = (s'_i - s_i) - (s'_j - s_j) + (s_i - s_j) = -\Delta + 4,$$

combining the two expressions, we find that a score swap occurs when $\Delta < 4$. Using the result from previous subsection and the following Remark G.1 allows us to conclude that under strict stochastic transitivity:

$$r_{\psi_C}(\bar{\mathbf{x}}) = \min_{i<j} |p_{ij} - 1/2|. \tag{32}$$

We can of course extend the previous result to an empirical distribution $\widehat{P}_N$ as it is the case in the decentralized estimation.

Remark G.2. *Copeland scoring can be extended to a Copeland-α scoring function [25] for $\beta \in [0,1]$. While the embedding map is share with the original Copeland method, the aggregation map $\psi_{C,\beta}$ is given for all $i \in [\![n]\!]$ by*

$$[\psi_{C,\beta}(\mathbf{x})]_i = \sum_{j \neq i} \mathbb{I}\left\{\mathbf{x}_{ij} > \frac{1}{2}\right\} - \mathbb{I}\left\{\mathbf{x}_{ij} < \frac{1}{2}\right\} + (2\beta - 1)\mathbb{I}\left\{\mathbf{x}_{ij} = \frac{1}{2}\right\}.$$

In fact, for $\alpha = 1/2$, we retrieve the original Copeland voting rule described in Table 1. Considering $\mathbf{x}$ outside of the hyperplane $H_{ij} := \{\mathbf{x} \in [0,1]^{\binom{n}{2}} \mid \mathbf{x}_{ij} = 1/2\}$ for $i < j$, it is easy to see that $\psi_{C,\beta}(\mathbf{x}) = \psi_C(\mathbf{x})$ for all $\beta \in [0,1]$. Since the plane has an empty interior, it forces the cell boundaries $\partial C(\sigma_\star)$ to be equal when no pairwise ties are present (see Remark 2). When $\bar{\mathbf{x}} \in H_{ij}$ then the distance to the cell is null.

H Likelihood estimation on $\mathfrak{S}_n$

In this section, we introduce parametric probabilistic models and discuss how the previous constructions adapt to maximum likelihood estimation (MLE). Indeed, for the model shown below finding the consensus $\sigma_\star$ given the method of consensus often translates into finding the MLE. We discuss two common parametric models on $\mathfrak{S}_n$.

H.1 Mallows model

Setup. We consider a right-invariant metric d (*e.g.* Kendall-τ d_τ) on $\mathfrak{S}_n$, *i.e.* for $\sigma, \sigma' \in \mathfrak{S}_n$ $d(\sigma', \sigma) = d(\sigma' \circ \sigma^{-1}, \mathrm{Id})$, where Id denotes the identity permutation in $\mathfrak{S}_n$ (See Chap. 6 [20]). Mallows model P_θ^M [49] is a parametric distribution on $\mathfrak{S}_n$, with parameter $\theta = (\sigma_\star, \varphi)$, where $\varphi \in [0,1]$ is the dispersion parameter and $\sigma_\star \in \mathfrak{S}_n$ the central permutation, such that for $\sigma \in \mathfrak{S}_n$:

$$P_\theta^M(\sigma) = Z(\varphi)^{-1} \varphi^{d(\sigma, \sigma_\star)},$$

where $Z(\varphi)$ is the normalizing constant associated with the distribution, the constant depends only on the dispersion parameter thanks to the right-invariant property of d:

$$\sum_{\sigma \in \mathfrak{S}_n} \varphi^{d(\sigma, \sigma_\star)} = \sum_{\sigma \in \mathfrak{S}_n} \varphi^{d(\sigma_\star \circ \sigma^{-1}, \mathrm{Id})} = \sum_{\sigma' \in \mathfrak{S}_n} \varphi^{d(\sigma', \mathrm{Id})} =: Z(\varphi).$$

When $\varphi \to 0$, the distribution tends to the Dirac distribution $\delta_{\sigma_\star}$. When $\varphi \to 1$, the probability distribution is the Haar measure on $\mathfrak{S}_n$ *i.e.*, for all $\sigma \in \mathfrak{S}_n$, $P_{\mathrm{Haar}}(\sigma) = (n!)^{-1}$. The interest of such probability distribution is that it admits a mode given by $\sigma_\star$ as well as symmetry properties with respect to the distance d. A *mixture of Mallows distributions* is defined, for parameters $\theta_1, \ldots, \theta_k$ and a weight vector $(\pi_i)_{i=1}^k$ with $\sum_{i=1}^k \pi_i = 1$, as:

$$P^M(\sigma) = \sum_{i=1}^k \pi_i P_{\theta_i}^M(\sigma).$$

We quickly discuss the process to sample Mallows distributed samples for the synthetic dataset experiments. To sample permutations following P_θ^M, we used the RIM algorithm or variants (See [21, 46]). The idea is to switch the elements of the central parameter $\sigma_\star$, with acceptance φ^k with $k \in [\![n]\!]$ being the inversion length.

We aim to estimate the central parameter, the log-likelihood of P_θ^M is as follows:

$$\ell(\theta) = -\log(1/\varphi) \sum_{v=1}^N d(\sigma_\star, \sigma_v) - N \log(Z(\varphi)).$$

Thus, the MLE $\widehat{\sigma}$ for $\sigma_\star$ is

$$\widehat{\sigma} \in \operatorname*{argmin}_{\sigma \in \mathfrak{S}_n} \sum_{v=1}^N d(\sigma, \sigma_v).$$

We can see that finding the MLE is the same as solving the problem of Equation (1). Using the Corollary 3 from [9], we know that the pairwise probability for a Mallows distribution with dispersion parameter φ follows the conditions given in Lemma H.1.

Lemma H.1 (Busa-Fekete et al. [9]). *For Mallows model P_θ^M with parameter $\theta = (\sigma_\star, \varphi)$, and a given pair (i, j) with $i \neq j$, the following properties hold:*

1. $p_{ij} \geq \frac{1}{1+\varphi} > \frac{1}{2} \implies \sigma_\star(i) < \sigma_\star(j)$, *with equality when* $\sigma_\star(i) = \sigma_\star(j) - 1$.
2. $p_{ij} \leq \frac{\varphi}{1+\varphi} < \frac{1}{2} \implies \sigma_\star(i) > \sigma_\star(j)$, *with equality when* $\sigma_\star(j) = \sigma_\star(i) - 1$.
3. $p_{ij} > \frac{1}{2}$ *iff* $\sigma_\star(i) < \sigma_\star(j)$ *and* $p_{ij} < \frac{1}{2}$ *iff* $\sigma_\star(i) > \sigma_\star(j)$.

Derivation of the Borda stability radius. An analogous characterization holds for the Borda method. Under P_θ^M, the expected Borda score of item i is $s_i^B = \mathbb{E}_{P_\theta^M}[\sigma(i)]$, which depends on both the position $\sigma_\star(i)$ and the dispersion φ. The minimum score gap $\Delta_B(\varphi) := \min_{i \neq j} |s_i^B - s_j^B|$ governs the Borda stability radius via $r_{\mathrm{Id}}(\bar{x}) = \Delta_B(\varphi)/\sqrt{2}$ (see Appendix F.2). Two limiting regimes clarify its behavior. As $\varphi \to 0$, the distribution concentrates on $\sigma_\star$, so $s_i^B \to \sigma_\star(i)$ and consecutive score gaps approach 1, giving $r_{\mathrm{Id}} \to 1/\sqrt{2}$. As $\varphi \to 1$, all Borda scores collapse toward the common value $(n+1)/2$ under the Haar measure, so $\Delta_B(\varphi) \to 0$ and $r_{\mathrm{Id}} \to 0$. In summary:

$$r_{\mathrm{Id}}(\bar{\mathbf{x}}) = \frac{\Delta_B(\varphi)}{\sqrt{2}}, \qquad \text{with } \Delta_B(\varphi) \xrightarrow[\varphi \to 0]{} 1 \quad \text{and} \quad \Delta_B(\varphi) \xrightarrow[\varphi \to 1]{} 0.$$

Derivation of the Copeland stability radius. From Lemma H.1, for any pair (i, j) with $\sigma_\star(i) < \sigma_\star(j)$, we have $p_{ij} \geq \frac{1}{1+\varphi}$, so that

$$\left|p_{ij} - \tfrac{1}{2}\right| \;\geq\; \frac{1}{1+\varphi} - \frac{1}{2} = \frac{1-\varphi}{2(1+\varphi)},$$

with equality precisely when $\sigma_\star(i) = \sigma_\star(j) - 1$, *i.e.*, for adjacent pairs in $\sigma_\star$. Since Appendix G.2 establishes that $r_{\psi_C}(\bar{x}) = \min_{i<j} |p_{ij} - 1/2|$ under strict stochastic transitivity, taking the minimum over all pairs yields the claimed expression:

$$r_{\psi_C}\left(\mathbb{E}_{P_\theta^M} \phi_C\right) = \frac{1-\varphi}{2(1+\varphi)}.$$

This quantity is strictly positive for $\varphi \in [0, 1)$ and tends to zero as $\varphi \to 1$ (*i.e.*, as the distribution approaches the Haar measure), reflecting the fact that the consensus ranking becomes unidentifiable in the uniform regime.

Comparison of breakdown thresholds. Substituting these stability radii into the general lower bound of Proposition 1, and using the mixed norms $|||\phi_C||| = \sqrt{n(n-1)/2}$ and $|||\phi_B||| = \sqrt{n(n+1)(2n+1)/6}$, the method-specific breakdown thresholds under a Mallows model with dispersion factor φ read:

$$\varepsilon_C^- = \frac{1-\varphi}{2(1+\varphi)\sqrt{2n(n-1)}}, \qquad \varepsilon_B^- = \frac{\Delta_B(\varphi)}{2\sqrt{n(n+1)(2n+1)/3}}.$$

The ratio $r_\psi(\bar{x}) / |||\phi|||$, which controls the breakdown threshold, grows as $O(n^{-\beta})$ and $\beta = 1$ for Copeland and $\beta = 3/2$ for Borda as $n \to \infty$. We note however that both thresholds vanish as $\varphi \to 1$, consistently with the loss of identifiability of $\sigma_\star$ under near-uniform preferences.

I Approximated metric-based consensus.

We are interested in problems that follow a metric-based minimization problem as described in (1). A correspondence can be made between barycentric (metric-based) and score-based approaches. Indeed, minimizing the sum of Spearman-ρ correlations $d_\rho(\sigma, \sigma') = \sum_{i=1}^n (\sigma(i) - \sigma'(i))^2$ leads to the Borda consensus. Spearman's footrule $d_1(\sigma, \sigma') := \sum_{i=1}^n |\sigma(i) - \sigma'(i)|$, in turn, gives rise to the median aggregation rule [24]. If the coordinate-wise median ranks across voters form a valid

permutation, then this permutation minimizes the total footrule distance [24]. We first describe techniques to make the computations of specific medians more tractable, before giving specific details about a greedy approach for Kemeny consensus, we first describe heuristic methods to estimate the minimum of a median estimation problem for Spearman Footrule distance d_1 and the use of Local Kemenization to enforce robustness with respect to d_τ.

I.1 Spearman Footrule relaxation

In this section, we present a gossip-based method for the Footrule consensus (see Alg. 3) using GoalPD for decentralized median estimation [61]. We also introduce a decentralized implementation of Local Kemenization (see Alg. 4), which enables closer approximation to the Kemeny consensus while remaining computationally tractable.

Algorithm 3 Decentralized Footrule Consensus [24].

Init: Each voter v initializes its estimate as $\mathbf{x}_v = \sigma_v$.
Initialize median estimator: `goalpd.init()`.
for $t = 1, 2, \ldots, T$ **do**
  Select an edge $e = (u, v) \in E$ with probability p_e.
  $\mathbf{x}_u, \mathbf{x}_v \leftarrow$ `goalpd.update`$(\mathbf{x}_u, \mathbf{x}_v)$.
end for
Output: For all node $v \in [\![N]\!]$, return the Footrule (median rank) estimate $\widehat{\sigma}_v$ by sorting the components of $\mathbf{x}_v$ in ascending order.

I.2 Local Kemenization

Beyond global optimality, local Kemeny optimality [22] offers a computationally tractable alternative that satisfies the extended Condorcet property, ensuring that Condorcet winners are ranked highly while Condorcet losers are placed near the bottom. Every Kemeny optimal solution is locally Kemeny optimal, but the converse does not hold, making local optimality an attractive relaxation that approximates the Kemeny consensus. Any initial ranking can be transformed into a locally Kemeny optimal solution via the following iterative procedure, known as local Kemenization [22]: *(i)* initialize with an aggregation from a tractable method, such as Borda or Copeland, *(ii)* For each pair of adjacent candidates (i, j) in the current ranking, check whether a majority of voters prefer j over i (*i.e.*, $\hat{p}_{ij} < 1/2$). If so, swap i and j, *(iii)* Repeat step b until no further swaps occur. This procedure converges in polynomial time and inherits robustness properties against *spam* or noisy inputs.

Algorithm 4 Decentralized Local Kemenization [22].

Select aggregation method `consensus` (*e.g.*, Copeland, Borda).
Init: Each voter v initializes its estimate as $\mathbf{x}_v = \left(\widehat{p}^v_{ij}\right)_{i<j}$, where $\widehat{p}^v_{ij} = \mathbb{I}[\sigma_v(i) < \sigma_v(j)]$. Do `consensus.init()`.
for $t = 1, 2, \ldots, T$ **do**
  Select an edge $e = (u, v) \in E$ with probability p_e.
  Update pairwise scores: $\mathbf{x}_u, \mathbf{x}_v \leftarrow (\mathbf{x}_u + \mathbf{x}_v)/2$.
  Update consensus scores: `consensus.update()`.
end for
Compute preliminary ranking: $\widehat{\sigma}_v \leftarrow$ `consensus.sort()`. Apply Local Kemenization on $\widehat{\sigma}_v$ using $x_{v,(i,j)}$ in place of $\widehat{p}_{ij}$, following the procedure described above.
Output: For all nodes $v \in [\![N]\!]$, return the locally Kemeny optimal aggregation $\widehat{\sigma}_v$.

We denote by $L_P(\sigma) := \mathbb{E}_{\Sigma \sim P}\, d_\tau(\sigma, \Sigma)$ the Kemeny loss for distribution P. If $\sigma_\star$ is a solution of Equation (1) with $P = \widehat{P}_N$ and $d = d_\tau$, the optimal loss is denoted by $L^\star_{\widehat{P}_N} := L_{\widehat{P}_N}(\sigma_\star)$. The excess risk can be defined for a given consensus $\sigma \in \mathfrak{S}_n$ as:

$$\Delta L_{\widehat{P}_N}(\sigma) = L_{\widehat{P}_N}(\sigma) - L^\star_{\widehat{P}_N}. \tag{33}$$

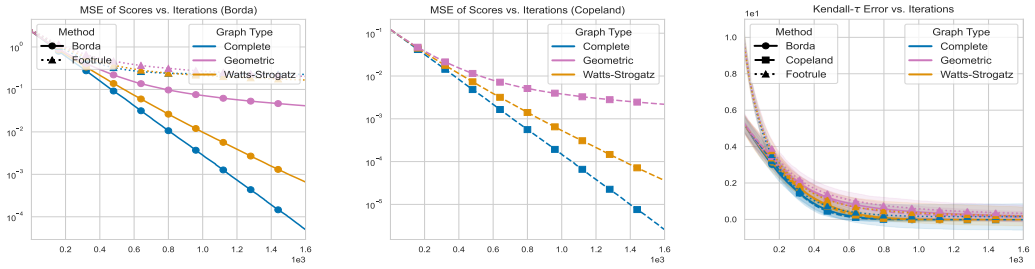
This method stems from the fact that the Kendall-τ distance is minimized when the minimal number of inversions occurs.

J Other experiments

In this section, we present several experiments not included in Section 5 notably on synthetic data models especially with probabilistic models as detail in Appendix H, as well as considerations for metric-based consensus as mentioned in Appendix I. As in Section 5, for all settings, the gossip communication follows a uniform distribution over edges, *i.e.*, $p = (1/|E|)_{e \in E}$

J.1 Asynchronous convergence study on synthetic data

Experimental setup. We generate synthetic ranking data according to the Mallows model P_θ^M with parameter $\theta = (\sigma_\star, \varphi)$. We set the number of agents to $N = 151$, the number of items to $n = 8$, and the dispersion parameter to $\varphi = 0.5$, with a central permutation $\sigma_\star$ drawn uniformly at random from $\mathfrak{S}_n$ and fixed across all trials. In each trial, the rankings are sampled independently from the Mallows distribution with parameters θ. Communication graphs are generated using three standard random graph models: complete graphs, Watts–Strogatz graphs, and random geometric graphs. For geometric graphs we set the connectivity radius $r = \sqrt{(\log(N) + 16)/(N\pi)}$, and for Watts–Strogatz graphs we set $k = \max(\lfloor N/500 \rfloor, 6)$. All generated graphs are verified to have a connectivity constant $\lambda(p) \in (0, 1)$, ensuring they are connected and non-bipartite, as required by standard convergence assumptions. We report results averaged over 100 independent trials, with each trial running for $2,000$ iterations. To apply the Goal-PD algorithm, we consider the step size $\rho = 1$ and refer to [61] for more details on their significance.



(a) MSE of Borda and Footrule scores

(b) MSE of pairwise probabilities (Copeland's mathod)

(c) Average d_τ error to respective consensus

Figure 5: Convergence of consensus methods (Borda, Copeland, Footrule) on rankings sampled from a Mallows model ($N = 151$ agents, $n = 8$ items, $\varphi = 0.5$) across different graph topologies. Results show mean $\pm$ standard deviation over 100 trials.

Results. Figure 5 presents the convergence behavior of the proposed gossip-based decentralized consensus methods under the Mallows model. The first plot shows the evolution of the mean squared error (MSE) of the Borda scores (*i.e.*, mean of ranks) and Footrule scores (*i.e.*, median of ranks), averaged over all node estimates relative to their true scores. Both methods converge rapidly across all graph topologies, achieving an MSE below 0.5 in fewer than 1000 iterations. The Footrule score converges slightly more slowly than the Borda score, which is expected given that estimating the median is inherently more challenging than estimating the mean. The second plot shows the convergence of the pairwise scores underlying the Copeland consensus, which similarly shows fast and stable convergence. The third plot reports the average Kendall-τ distance between the estimated consensus rankings (Borda, Copeland, and Footrule) and their respective ground-truth consensus rankings. Convergence is again rapid across all methods, with the complete graph converging fastest, followed by the Watts–Strogatz and geometric graphs, consistent with their respective degrees of connectivity.

J.2 Comparison of convergence rate between the synchronous and asynchronous versions

Experimental setup. We consider a setup with $N = 250$ voters and $m = 7$ alternatives. In this experiment, we compare the communication costs of Algorithm 1 and Algorithm 2. To do so, we compare the empirical estimations of $T^{\mathrm{Rand}}(0)$ and $T^{\mathrm{Sync}}(0)$ on 30 trials. To make the comparison more accurate, we also consider the number of communication *i.e.* the number of times a node calls its neighbors in the synchronous setting, in order to compare it with the randomized setting and the value of $T^{\mathrm{Rand}}(0)$.

Scoring rule	Graph	$T^{\mathrm{Rand}}(0)$	$T^{\mathrm{Sync}}(0)$	Num. of communications
Borda	2D Grid	8060.0	47	21855
	Geometric	8013	28	41244
	Watts-Strogatz	1880	5	6250
Copeland	2D Grid	6794	41	19065
	Geometric	6549	24	35352
	Watts-Strogatz	1677	4	5025

Table 2: Comparison of mixing times in the synchronous and randomized settings.

Results. In Table 2, we present the mixing times and total number of communications. While, the number of time steps required for the complete mixing of the rankings in the randomized setting is usually significantly higher that $T^{\mathrm{Sync}}(0)$, it is in fact comparable and usually lower than the total number of communication required in the synchronous case.

J.3 Convergence study on Mallows model with varying φ

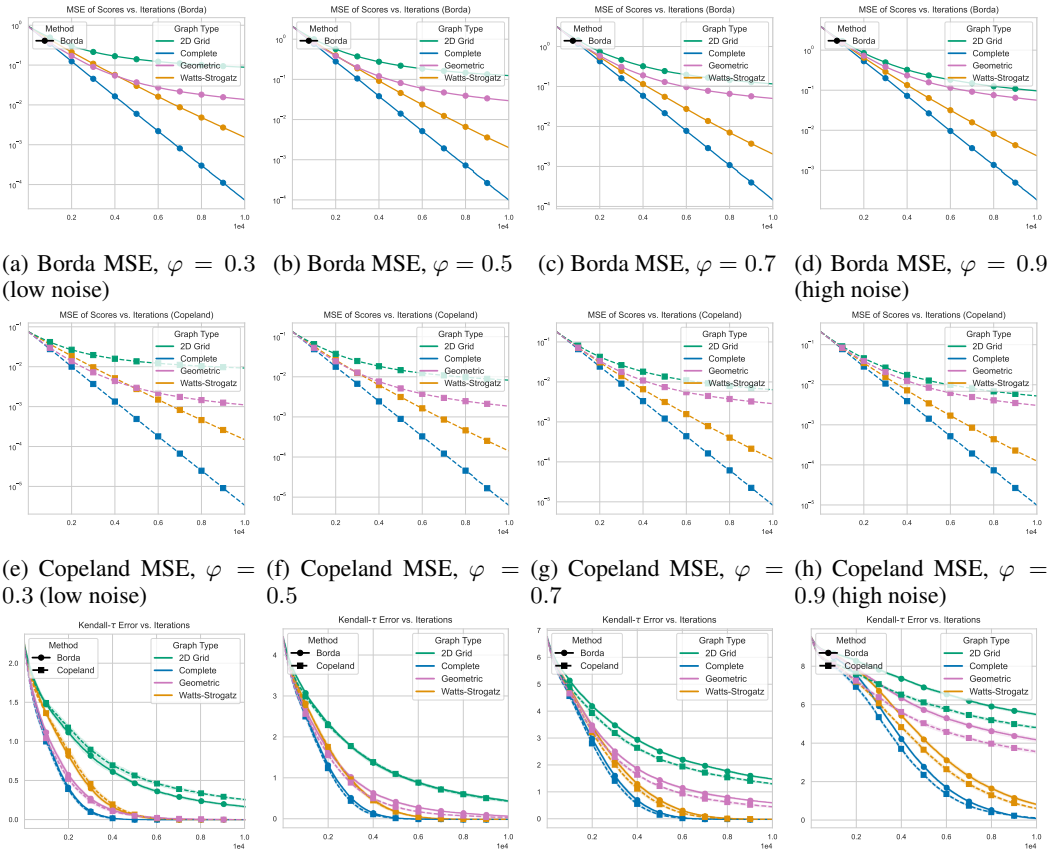
Experimental setup. We consider a setup with $N = 1000$ voters and $n = 7$ alternatives. Preferences are drawn from a Mallows model with a fixed central ranking $\sigma_\star = \mathrm{Id}$ and dispersion parameter $\varphi \in \{0.3, 0.5, 0.7, 0.9\}$, where small values of φ correspond to rankings concentrated near $\sigma_\star$ (low noise) and large values correspond to near-uniform preferences (high noise). We evaluate on the four graph topologies previously considered to govern the gossip communication structure, all graphs are configured to have $N = 1000$ nodes. We measure performance along two axes: the MSE between each agent's estimated Borda (resp. Copeland) score vector and the ground-truth score vector induced by $\sigma_\star$, and the Kendall-τ distance between the consensus ranking decoded from the aggregated scores and the central ranking $\sigma_\star$. The gossip process is run for up to $10{,}000$ iterations, with metrics recorded at logarithmically spaced checkpoints.

Results. The results are presented in Figure 6. Across all values of φ, graph topology is the dominant factor governing convergence: the Complete graph achieves the fastest decay in both MSE and Kendall-τ error, while the 2D Grid is consistently the slowest, with Watts-Strogatz and Geometric graphs occupying intermediate positions. As φ increases from 0.3 to 0.9, preferences become more dispersed and convergence degrades for all topologies, with the 2D Grid suffering most severely, at $\varphi = 0.9$, its Kendall-τ error remains well above zero after 10000 iterations while the Complete graph still converges to near zero. Borda and Copeland exhibit nearly identical ranking recovery across all settings, indicating that the convergence bottleneck lies in the gossip mixing time of the graph rather than the choice of aggregation rule.

J.4 Empirical breakdown function sweep for different n and N

Experimental setup. We follow the experimental setup of Section 5.2, and sweep the empirical breakdown function $\varepsilon(\delta)$ for different values of the number of alternatives $n \in \{5, 7, 10, 15\}$ and $N \in \{100, 250, 500\}$ for both Borda and Copeland methods. We fix $\delta = 1$ in the results to show the minimum contaminated fraction required to move the ranking of at least 1 in Kendall-τ distance with respect to the original consensus ranking.

Results. Following the results of Table 3, the empirical breakdown function increases with N, as expected from the topology of the graph. For smaller values of N, for which the honest signal



(a) Borda MSE, $\varphi = 0.3$ (low noise)
(b) Borda MSE, $\varphi = 0.5$
(c) Borda MSE, $\varphi = 0.7$
(d) Borda MSE, $\varphi = 0.9$ (high noise)
(e) Copeland MSE, $\varphi = 0.3$ (low noise)
(f) Copeland MSE, $\varphi = 0.5$
(g) Copeland MSE, $\varphi = 0.7$
(h) Copeland MSE, $\varphi = 0.9$ (high noise)
(i) Borda vs. Copeland consensus, $\varphi = 0.3$
(j) Borda vs. Copeland consensus, $\varphi = 0.5$
(k) Borda vs. Copeland consensus, $\varphi = 0.7$
(l) Borda vs. Copeland consensus, $\varphi = 0.9$

Figure 6: Score estimation error and consensus quality under the Mallows model ($N = 1000$ voters, $n = 7$ alternatives) across four dispersion levels $\varphi \in \{0.3, 0.5, 0.7, 0.9\}$ (columns, left to right). Small φ indicates low noise (rankings concentrated near the central order); large φ indicates high noise (near-uniform preferences). Rows 1-2 show the MSE of estimated Borda and Copeland scores, respectively, as a function of the number of sampled rankings. Row 3 compares the consensus rankings produced by the two rules, measured by their distance to the ground-truth central ranking.

is noisier, only a small contamination fraction is needed to break the consensus. However, the contamination required increases once N grows large relatively to n.

	$\varepsilon_B(1)$			$\varepsilon_C(1)$		
n/N	100	250	500	100	250	500
5	0.010	0.065	0.120	0.010	0.065	0.102
7	0.010	0.047	0.120	0.010	0.047	0.120
10	0.010	0.010	0.102	0.010	0.065	0.083
15	0.010	0.010	0.157	0.010	0.047	0.083

Table 3: Empirical breakdown function valued at $t = 7500$ and $\delta = 1$ for both Borda and Copeland methods.

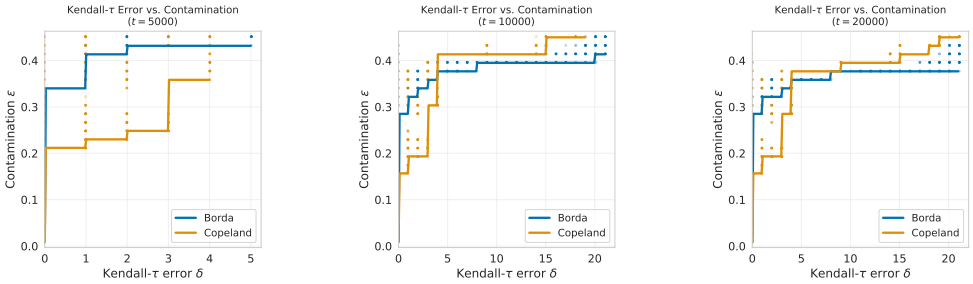
Consensus	$d_\tau(\cdot, \sigma_\star)$	$\Delta L_{\widehat{P}_N}$ w/o Local Kemenization	$\Delta L_{\widehat{P}_N}$ w/ Local Kemenization
Borda	0.25 ± 0.46	0.01 ± 0.03	0.00 ± 0.00
Copeland	0.07 ± 0.26	0.00 ± 0.00	0.00 ± 0.00
Footrule	1.04 ± 0.81	0.17 ± 0.17	0.00 ± 0.00
Kemeny	0.06 ± 0.24	N/A[3]	N/A[3]

Table 4: Robustness of consensus methods for recovering the central ranking $\sigma_\star$ under random contamination. Rankings are sampled from a Mallows model ($N = 100$, $n = 8$ items, $\varphi = 0.5$) where $\varepsilon = 0.3$ of samples are replaced with uniform random permutations. We report the Kendall-tau distance to the ground truth $d_\tau(\cdot, \sigma_\star)$ and the excess risk $\Delta L_{\widehat{P}_N}$ follows the definition in Equation (33), where $L^\star_{\widehat{P}_N}$ is the Kemeny optimal value for the empirical distribution $\widehat{P}_N$. Results show mean $\pm$ standard deviation over 100 trials.

J.5 Breakdown study with a stubborn model

Setup. We replicate the same setup as the one from Section 5.2, but we consider another methodology regarding the contaminated nodes. Indeed, we consider in this setup that the adversarial nodes contaminates the honest ones, without participating in the communication protocol, hence we refer to the adversarial nodes as 'stubborn'.

In Figure 7, the results are similar to the ones of Figure 3, as the trends of the empirical breakdown functions for both Borda and Copeland consensus are closely related.



(a) At iteration $t = 5000$ (b) At iteration $t = 10000$ (c) At iteration $t = 20000$

Figure 7: Average Kendall-τ error vs. contamination for the Mallows distribution.

J.6 Local Kemenization as a Kemeny consensus approximation

Experimental setup. We further evaluate the robustness of classical consensus methods under contamination, as reported in Tables 4 and 5. For simplicity, the analysis is conducted in a centralized setting, though it extends naturally to the decentralized case. For each table, we report for each method the mean Kendall-tau distance $d_\tau(\,\cdot\,, \sigma_\star)$ to the true reference ranking of the Mallows distribution (the target we aim to recover), the excess loss $\Delta L_{\widehat{P}_N}$, and the effect of local Kemenization on $\Delta L_{\widehat{P}_N}$ as a post-processing step to approximate the Kemeny consensus. We consider two contamination distributions for the corrupted rankings $(\sigma_w)_{w \in B}$. In the random contamination setting, corrupted rankings are sampled i.i.d. from the Haar measure P_{Haar} on $\mathfrak{S}_n$. In the adversarial contamination setting, they are sampled i.i.d. from a Mallows model P_θ^M with $\theta = (\sigma^{\text{rev}}, 0)$ centered at the reversed ranking σ^{rev}.

Results. Our results as illustrated in Table 4 and Table 5 show that Kemeny, followed by Copeland, is the most robust consensus method under both contamination models. Notably, applying local Kemenization to either the Borda or Copeland consensus consistently recovers the Kemeny consensus, suggesting that local Kemenization is a promising and lightweight strategy for improving the robustness of simpler consensus methods.

[3]For the global Kemeny consensus, the system is already locally Kemeny optimal.

Consensus	$d_\tau(\cdot, \sigma_\star)$	$\Delta L_{\widehat{P}_N}$ w/o Local Kemenization	$\Delta L_{\widehat{P}_N}$ w/ Local Kemenization
Borda	0.91 ± 0.85	0.03 ± 0.04	0.00 ± 0.00
Copeland	0.59 ± 0.71	0.00 ± 0.02	0.00 ± 0.01
Footrule	1.26 ± 0.99	0.13 ± 0.11	0.01 ± 0.02
Kemeny	0.58 ± 0.67	N/A[3]	N/A[3]

Table 5: Robustness of consensus methods under adversarial contamination. Setup is identical to Table 4, except contaminated samples are drawn from a Mallows model centered at the reversed ranking σ_{rev} rather than uniform random permutations.

Experiments	Time to compute
Debian / Inference model convergence	$<$ 5min
Sushi convergence	$\sim$ 1h
Mixture of Mallows / MovieLens convergence	$\sim$ 15min
Experiment 5.2 (Breakdown studies)	$\sim$ 3h
Experiment J.1 (Synthetic data with Spearman Footrule)	$\sim$ 1h
Experiment J.2 (Comparison of convergence rates)	$\sim$ 1h
Experiment J.3 (Varying dispersion parameter)	$\sim$ 1h05
Experiment J.4 (Sweep for different n and N)	$\sim$ 3h
Experiment J.6 (Local Kemenization study)	$\sim$ 30min

Table 6: Time to compute per simulation.

K Experiments compute resources

All experiments were run on a single compute node equipped with one NVIDIA P100 16GB GPU, 30GB RAM, and 20 CPU cores. Despite GPU availability, all computations are CPU-bound (Numpy/Numba), with parallelism limited to 8 workers across trial configurations. In Table 6, we describe approximately the time to execute the different numerical simulations.