A Tale of HodgeRank and Spectral Method: Target Attack Against Rank Aggregation is the Fixed Point of Adversarial Game

Ke Ma®, Member, IEEE, Qianqian Xu®, Senior Member, IEEE, Jinshan Zeng®, Guorong Li®, Xiaoichun Cao®, Senior Member, IEEE, and Qingming Huang®, Fellow, IEEE

Abstract—Rank aggregation with pairwise comparisons has shown promising results in elections, sports competitions, recommendations, and information retrieval. However, little attention has been paid to the security issue of such algorithms, in contrast to numerous research work on the computational and statistical characteristics. Driven by huge profit, the potential adversary has strong motivation and incentives to manipulate the ranking list. Meanwhile, the intrinsic vulnerability of the rank aggregation methods is not well studied in the literature. To fully understand the possible risks, we focus on the purposeful adversary who desires to designate the aggregated results by modifying the pairwise data in this paper. From the perspective of the dynamical system, the attack behavior with a target ranking list is a fixed point belonging to the composition of the adversary and the victim. To perform the targeted attack, we formulate the interaction between the adversary and the victim as a game-theoretic framework consisting of two continuous operators while Nash equilibrium is established. Then two procedures against HodgeRank and RankCentrality are constructed to produce the modification of the original data. Furthermore, we prove that the victims will produce the target ranking list once the adversary masters the complete information. It is noteworthy that the proposed methods allow the adversary only to hold incomplete information or imperfect feedback and perform the purposeful attack. The effectiveness of the suggested target attack strategies is demonstrated by a series of toy simulations and several real-world data experiments. These experimental results show that the proposed methods could achieve the attacker’s goal in the sense that the leading candidate of the perturbed ranking list is the designated one by the adversary.

1 INTRODUCTION

Given partially observed pairwise comparisons, we are interested in aggregating these partial orders into a full ranking of all candidates. Such a statistical estimation problem arises in various disciplines, including the social choice theory [1], statistics [2], machine learning [3], [4], sports [5], recommendation system [6], information retrieval [7], psychology [8], bioinformatics [9], etc. Roughly speaking, the rank aggregation approach treats pairwise comparisons as access to estimate the underlying “qualities” or “scores” of the items being compared (e.g., skill levels of tennis players, preference for political candidates, and interest degree of advertisements). These latent preference scores represent the voters’ choices.

Besides the statistical [10], [11] and computational issues [12], [13], security is a new perspective for understanding the ranking aggregation problem. In pursuit of substantial economic benefits, the potential attackers have strong motivations and incentives to manipulate the aggregated results when the pairwise ranking algorithms are utilized in high-stakes applications, e.g., elections (the so-called “Condorcet winner” [14] is a winner who will win every pairwise comparison), sports competitions (in “round-robin tournament”, each contestant meets all other contestants in turn), and recommendation (“click-through” data is always in paired form). A profit-oriented adversary could try his/her best to designate the ranking list to favor his/her demands. The attacker could place the particular object at the top of the recommendation list or help the specific candidate win an
election. If the attackers compromise the integrity of ranking results, these high-stakes applications’ fairness and rationality will be profoundly damaged.

However, the possible risk and potential threat of pairwise ranking algorithms have not been comprehensively examined. To the best of our knowledge, the existing adversarial arsenals do not cause damage to the ranking aggregation algorithms. Two significant disparities exist between the rank aggregation and the learning problems (e.g. classification and regression).

- Considering the process of learning problems, it separates into training and test phases where the used datasets are non-overlapping. Conversely, the aggregation problem needs to combine all collected comparisons into a single global ranking list in a one-shot way, and the test protocol does not exist.
- From the view of data, the typical learning data consists of features (always a continuous variable) and the desired output value (also called the supervisory information). The aggregation object is a pairwise comparison which involves two candidates and the partial order (a.k.a preference) between them. A binary variable always represents such an order.

Consequently, the attack methods against learning and aggregation have their characteristics. The Attack against learning could occur in the training phase (poisoning attack [15]) and test phase (evasion attack, a.k.a adversarial examples [16]) separately. Meanwhile, the attack against aggregation, which modifies the pairwise comparisons, should happen before the combination of pairwise comparisons. Furthermore, the attack against learning always conducts a continuous optimization problem which operates a single learning data point into the “weapon” of the adversary. Meanwhile, due to its discrete nature, modification of aggregation data naturally derives a discrete optimization problem, which decides to insert, delete or flip a pairwise comparison. To achieve the targeted attack against the ranking aggregation, the adversary must analyze the characteristics of the pairwise ranking problem, including the properties of both paired data and the victim algorithms. Consequently, the adversary would predict and induce the victim’s behavior with parsimonious actions to achieve his/her goals.

Given these challenges, we propose a principle framework for adversarial perturbations of pairwise comparisons, aiming to manipulate the results of rank aggregation algorithms. In particular, we focus on two well-known procedures: least squared approach [2] and spectral method [4]. These algorithms give the full ranking based on appropriate estimates of the latent preference scores. We make the following contributions of the paper:

- We identify the purposeful attack process as a fixed point problem in the dynamical system. Such a dynamical system is composed of two continuous operators characterized by the adversary and the victim. The core of the proposed target attack procedure lies in a perspective of the continuous operators characterized by the adversary and the victim. The core of the proposed target attack procedure lies in a perspective of the continuous operator with optimal conditions, which models the interaction between the adversary and the rank aggregation algorithms. We prove that the target ranking of the adversary is a fixed point of the dynamical system consisting of the complete information adversary and the specific ranker, e.g. HodgeRank and RankCentrality.
- The targeted attack strategy against HodgeRank is established with the help of the optimal condition of the least squared problem. Moreover, we show that the proposed method allows the attacker to successfully manipulate HodgeRank with incomplete information or imperfect feedback.
- Different target attacks against RankCentrality corresponding to the reversible and irreversible stochastic transition matrices are formulated into the inverse eigenvalue problem. The worst-case asymptotic variance analysis can help handle incomplete information or imperfect feedback with constructive solutions.
- A complexity model named “Incremental Pairwise Comparison Oracle” (ICPO) is proposed to evaluate the efficiency of purposeful attack strategies against rank aggregation methods.

The rest of the paper is structured as follows: in the next section, we introduce the basic concepts of rank aggregation and two representative algorithms as HodgeRank and RankCentrality. We then establish a general framework for manipulating different rank aggregation algorithms in Section 3. We identify the manipulation process as a fixed point problem in the dynamical system. Such a dynamical system is composed of two continuous operators characterized by the adversary and the victim. The core of the proposed target attack procedure lies in a perspective of the continuous operator with the victim’s optimal condition, which elicits the interaction between the adversary and the victim. We prove that the target ranking is a fixed point of the dynamical system consisting of the complete information adversary and the specific ranker. Based on the proposed framework, we derive the target attack strategies against HodgeRank and RankCentrality in Sections 4 and 5 correspondingly. Moreover, we establish the sufficient conditions that the target rankings lists are the fixed points of the two cases by Theorem 2 (HodgeRank), Theorem 4 (RankCentrality with reversible stochastic transition matrix), and Theorem 6 (RankCentrality with irreversible stochastic transition matrix). To the best of our knowledge, this is the first systematic study of attacking rank aggregation with specific purposes. The extensive evaluations are conducted on several datasets from different high-stake domains. Our experiments demonstrate that the proposed attacks could achieve the attacker’s goal in the sense that the leading candidate of the perturbed ranking list is the designated one by the adversary.

Notations. In the remainder of this paper, we will use positive integers to indicate alternatives and voters. Let $V$ be the set $|V| = \{1, \ldots, n\}$ which denotes a set of alternatives to be ranked. $U = \{u_1, \ldots, u_m\}$ denotes a set of voters. We will adopt the following notation from combinatorics:

$$\{V \atop l} := \text{set of all } l \text{ elements subset of } V.$$
In particular, \[
\begin{bmatrix}
V
\end{bmatrix} := \text{set of all unordered pairs of elements of } V
\]
\[= \left\{ (i, j) \mid \forall i, j \in V, i \neq j \right\}.\]

Moreover, for any \(i, j \in V, i \neq j\), we write \(i \succ j\) to mean that alternative \(i\) is preferred over alternative \(j\). Such a comparison could be converted into an ordered pair \((i, j)\). The set of ordered pairs will be denoted as
\[
V \times V := \left\{ (i, j) \mid i \succ j, \forall i, j \in V, i \neq j \right\}.
\]

Ordered and unordered pairs will be delimited by parentheses \((i, j)\) and braces \(\{i, j\}\) respectively. If we wish to emphasize the preference judgment from a particular voter \(u \in U\), we will write \(i \succ_u j\).

## 2 Ranking With Pairwise Comparisons

We begin with a formal description of the parametric model for binary comparisons, a.k.a Bradley-Terry-Luce (BTL) model [17]. Then we revisit the comparison graph and the Laplacian matrix which are essential for the ranking algorithms tailored to the BTL model. Two popular approaches which rank the items based on appropriate estimation of the latent preference scores, named HodgeRank [2] and Rank Centrality [4], are chosen as the victims to motivate our target attack strategies.

### 2.1 Parametric Model and Pairwise Comparisons

Given a collection \(V\) of \(n\) alternatives, the parametric model of pairwise comparisons assumes that each \(i \in V\) has a certain numeric quality score \(\theta_i^*\). Suppose that \(\theta^* \in \mathbb{R}^n\)
\[
\begin{bmatrix}
\theta^*
\end{bmatrix} = [\theta_1^*, \ldots, \theta_n^*]^T
\]
comprises the underlying preference scores assigned to each of the \(n\) items. Without loss of generality, \(\theta^*\) could be positive as
\[
\theta_i^* > 0, \forall i \in [n].
\]

Specifically, a comparison of any pair \(\{i, j\} \in \begin{bmatrix} V \end{bmatrix}^2\) is generated via the comparing between the corresponding scores \(\theta_i^*\) and \(\theta_j^*\) (in the presence of noise) by the BTL model. Let \(y_{ij}\) denote the outcome of the comparison of the pair \(i\) and \(j\) based on \(\theta^*\), such that \(y_{ij} = 1\) if \(i\) is preferred over \(j\) and \(y_{ij} = -1\) otherwise. Then, according to the BTL model,
\[
y_{ij} = \begin{cases} 
1, & \text{with probability } \frac{\theta_i^*}{\theta_i^* + \theta_j^*}, \\
-1, & \text{otherwise}.
\end{cases}
\]

Since the BTL model is invariant under the scaling of the scores, the latent preference score is not unique. Indeed, under the BTL model, a score vector \(\theta^* \in \mathbb{R}^n\) is the equivalence class
\[
\Theta^* = \left\{ \theta \mid \exists \alpha > 0 \text{ such that } \theta = \alpha \theta^* \right\}.
\]
The outcome of a comparison depends on the equivalence class \(\Theta^*\).

### 2.2 Comparison Graph and Combinatorial Laplacian

A graph structure, named comparison graph, arises naturally from pairwise comparisons as follows. Let \(G = (V, E)\) stand for a comparison graph, where the vertex set \(V = [n]\) represents the \(n\) candidates. In our problem setting, we pay attention to the complete graph setting: the directed edge set \(E = V \times V\) and \(N := |E| = n(n - 1)\). One can further associate weights \(w^*\) on \(E\) as voters \(U\) would have rated, i.e. assigned cardinal scores or given an ordinal ordering to, the complete set of the alternatives \(V\). But no matter how incomplete the rated portion is, one may always convert such judgments into pairwise rankings that have no missing values as follows. For each voter \(u \in U\), the pairwise ranking matrix is a skew-symmetric matrix \(\mathbf{Y}^u = (y_{ij}^u) \in \{-1, 0, 1\}^{n \times n}\) as
\[
y_{ij}^u = -y_{ji}^u, \forall (i, j) \in E, \forall u \in U,
\]
where
\[
y_{ij}^u = \begin{cases} 
1, & \text{if } i \succ_u j, \\
-1, & \text{if } j \succ_u i, \\
0, & \text{otherwise}.
\end{cases}
\]

Furthermore, we associate weight with each directed edge as \(w^* = [w_{12}^*, w_{13}^*, \ldots, w_{n,n-1}^*]^T \in \mathbb{Z}_+\)
\[
w_{ij}^* := \sum_{u \in U} I[y_{ij}^u > 0] + I[y_{ij}^u < 0],
\]
where \(I[\cdot]\) is the Iverson bracket. Consequently, we can represent any pairwise ranking data as a comparison graph \(G\) with edge weights \(w^*\).

Given a graph \(G\) and weights \(w^*\), it is common to consider the weight matrix \(\mathbf{W}^*\) with \(w^*_{ij}\) as matrix elements, as well as the diagonal degree matrix \(\mathbf{D}^* = \text{diag}(d_1^*, \ldots, d_n^*)\) given by \(d_i^* = \sum_{j=1}^n w_{ij}^*\), which represents the volume taken by each node in the graph \(G\). The combinatorial Laplacian \(\mathbf{L}_0\) is defined as
\[
\mathbf{L}_0 = \mathbf{D}^* - \mathbf{W}^*.
\]
In both solving process and the theoretical analysis, the combinatorial Laplacian \(\mathbf{L}_0\) plays a vital role for the popular approaches based on the parametric model.

### 2.3 HodgeRank

The HodgeRank method discussed in [2] consists in finding the relative ranking score by solving the following least-squares problem:
\[
\begin{aligned}
\min_{\theta \in \mathbb{R}^n} & \quad \frac{1}{2} \sum_{(i,j) \in E} w_{ij}^* (y_{ij} - \theta_j + \theta_i)^2 \\
\text{such that } & \quad \mathbf{Y}^u = \alpha \mathbf{Y}^u,
\end{aligned}
\]
where \(y = [y_{12}, y_{13}, \ldots, y_{n,n-1}]^T\) represents the directions of edges. As \(G\) is a complete graph, we set \(y_{ij} = 1\) which indicates a direct edge from node \(i\) to \(j\). Based on combinational Hodge theory [2], the minimal norm solution of (7) is simply given as
\[
\bar{\theta} = -\mathbf{L}_0^\dagger \cdot \text{div}(y),
\]
where \(\mathbf{L}_0^\dagger\) is the Moore-Penrose pseudo-inverse of \(\mathbf{L}_0\), and the divergence operator \(\text{div}\) is defined as...
depends on how often and accordingly. We adopt \( \theta_i = \frac{1}{\text{deg}_i} \), where \( \text{deg}_i \) is the out-degree of a node, noted as \( \text{deg}_i \). This re-scaling ensures that each row-sum is at most one. Finally, to ensure that each row-sum is exactly one, the spectral method adds a self-loop to each node of \( V \). Concretely, the transition matrix \( P^* \) is converted from the pairwise comparison data \( w^* \) in such a way that

\[
\begin{align*}
P^*_{i,j} &= \begin{cases} 
0, & \text{if } i \neq j, \ w^*_{ij} + w^*_{ji} = 0, \\
\frac{1}{\text{deg}_i} \frac{w^*_{ij}}{w^*_{ij} + w^*_{ji}}, & \text{if } i \neq j, \ w^*_{ij} + w^*_{ji} \neq 0, \\
1 - \frac{1}{\text{deg}_i} \sum_{k \neq i} w^*_{ik} + w^*_{ki}, & \text{otherwise}.
\end{cases}
\end{align*}
\]

(11)

Rank centrality estimates the probability distribution obtained by applying matrix \( P^* \) repeated starting from any initial condition. Precisely, let \( \theta_0(i) = \frac{1}{n} \) denote the distribution of the random walk at time \( t \) with \( \theta_t = \{ \theta_t(i) \} \subseteq R^n \) as an arbitrary starting distribution on \( V \). Then the random walk holds

\[
\theta_{i+1} = \theta_i P^*.
\]

(12)

One expects the stationary distribution of the sample version \( P^* \) to form a good estimate of true relative ranking score, provided the sample size is sufficiently large. When the transition matrix has a unique left eigenvector \( \theta^* \) related to the largest eigenvalue, then starting from any initial distribution \( \theta_0 \), the limiting distribution \( \theta_{i+1} \) is unique. This stationary distribution \( \lim_{t \to \infty} \theta_t = \theta \) is the top left eigenvector of \( P^* \) as

\[
\lim_{t \to \infty} \theta_t = \bar{\theta} \quad \text{and} \quad \bar{\theta}^T = \bar{\theta} P^*,
\]

which only involves a simple eigenvector computation.

3 General Framework

In this section, we systematically introduce the general game-theoretic framework for target attacks on pairwise ranking. Specifically, we firstly propose the threat model of the purposeful adversary, then analyze the adversarial game from the view of the fixed point theorem, and finally present the definition of the continuous operator game and show the existence of equilibrium.

3.1 Threat Model of Purposeful Adversary

We provide here a detailed framework for target attacks against pairwise ranking algorithms. The framework consists of goal of the adversary, knowledge of the attacked method, and capability of manipulating the pairwise data, to eventually define the target attack strategies.

The Goal of Adversary. To execute the target attack, an adversary desires to provide the ranking algorithm with the tempered data. This action by the adversary will induce its opponent to produce a designated ranking result. If the ranking result meets the demand of adversary, we will say that the adversary has executed a successful attack. Let \( A \) and \( R \) be the adversary and the ranking algorithm respectively. Suppose that \( \bar{w} \) is the substitute of the original pairwise comparison data \( w^* \). The relative ranking scores \( \bar{\theta} \) and \( \bar{\theta} \) are produced by \( R \) with the original data \( w^* \) and the poisoned data \( \bar{w} \) accordingly. We adopt \( \pi_\theta \) as the ranking list decided by the relative ranking score \( \theta \). An adversary may have multiple intentions to manipulate the ranking result \( \pi_\bar{\theta} \). However, designating a given candidate to be the winner, a.k.a the top-1 position of \( \pi_\bar{\theta} \) must be the most attractive achievement. Consequently, we are concerned about the following situation: A hopes that \( R \) would arrange \( i_0 \in [n] \) as \( \pi_\bar{\theta}(1) \). The goal of \( A \) could be described as a set of ranking lists

\[
C_A := \{ \pi \mid \pi(1) = i_0 \}.
\]

(14)

If it holds that \( \pi_\bar{\theta} \in C_A \), we will say that \( A \) has a successful target attack strategy by substituting \( w^* \) with \( \bar{w} \). However, (14) always leads to the integer optimization problems which is NP-complete [18]. To avoid solving such a challenging problem, we turn our attention to the latent preference score. As the latent preference score \( \hat{\theta} \) decides \( \pi_\hat{\theta} \), we can convert \( C_A \) into the constraints on the relative ranking scores as

\[
C_A := \{ \theta \in R^n \mid \max_{i \in [n], i \neq i_0} \theta_i \leq \theta_{i_0} \}.
\]

(15)

This set is more convenient to incorporate with the optimization process of \( R \) and establish the targeted attack strategies.

The Knowledge of Adversary. We assume four distinct target attack scenarios which are distinguished by the knowledge of \( A \). The knowledge of \( A \) is partitioned into two
dimensions: the completeness of information predominated by $A$ and the accuracy of the feedback given by $R$. As the adversarial mechanism among $A$ and $R$ can be naturally modeled as a game, we introduce the following definitions from game theory.

**Definition 1 (Complete/Incomplete Game).** In the adversary game between ranking algorithm $R$ and the attacker $A$, there exists the following decomposition of $w^*$

$$w^* = w^*_a + w^*_u,$$

(16)

where $w^*_a \in Z^N_+$ represents the available part for $A$, which can be modified and substituted; $w^*_u \in Z^N_+$ is the inaccessible part and it could create a disturbance for $A$.

- If $w^*_a = 0$, it is a Complete Information Game and $A$ has the full accessibility to manipulate the input data for $R$.
- If $w^*_a \neq 0$, it is an Incomplete Information Game and there exists unavailable data for $R$.

It is known that uncertainty is a common phenomenon in adversarial operation. The uncertainty comes from the limitation of the attack ability, the defense mechanism of the victim, the randomness of data generation process and so on.

**Definition 2 (Perfect/Imperfect Game).** Let $\tilde{\theta}$ be the relative ranking score obtained by the victim algorithm $R$ with $w^*$ and $\pi_R$ be the corresponding ranking list.

- If the attacker can get the relative ranking score $\tilde{\theta}$ obtained by $R$, it is a Perfect Information Game.
- If the attacker only knows the ranking results $\pi_R$, it is an Imperfect Information Game.

Even though the perfect feedback $\tilde{\theta}$ is inaccessible in the majority of cases, $A$ still is possible to simulate $\tilde{\theta}$ with $w^*$ or $w^*_u$. Such a simulation could be completed with the public algorithm of $R$. Due to the existence of equivalence class, the numerical accuracy of $\tilde{\theta}$ would not strongly influence the behavior of $A$. As a consequence, $A$ could construct an imprecise copy of $\tilde{\theta}$ by $\pi_R$ under the imperfect information situation. Of course, we can further limit the attacker’s knowledge to design the corresponding attack strategies, which would be more efficient in actual situations but more challenging. The target attack strategies with the least knowledge could be the future direction of our research.

The Capability of Adversary. To modify the accessible part of pairwise comparisons $w^*_a$, $A$ could inject an arbitrary pairwise comparison into $w^*_a$, delete the existing comparison $(i,j)$ in $w^*_u$ or just flip $(i,j)$ as $(j,i)$. We authorize the following three operations to $A$.

- The adding operation can be described as an assignment in the following two cases:
  - $\exists u \in U$, $i, j \in V$, $i \neq j$, such that $y^*_u = 0$,
  - $\forall u \notin U$, $i, j \in V$, $i \neq j$, $y^*_u$ does not exist,

then

$$y^*_u \leftarrow 1 \text{ or } -1 \text{ arbitrarily}.$$
which is a mapping function as
\[
\pi \mapsto T_\mathcal{R}(T_A\pi) := \bigcup_{w \in T_A\pi} T_\mathcal{R}w.
\] (20)

Then the complete interaction between \( \mathcal{A} \) and \( \mathcal{R} \) can be described by the composition of \( T_\mathcal{R} \) and \( T_A \). It is worth noting that the purpose of the adversary could be represented as \( \pi_\mathcal{R} \in \mathcal{C}_\mathcal{A} \) and the target ranking \( \pi_A \) also satisfies \( \pi_A \in \mathcal{C}_\mathcal{A} \). By the definition of \( \mathcal{C}_\mathcal{A} \) (14) and (15), it is obvious that the leading candidate of \( \pi_A \) is invariant under \( T_\mathcal{R} \circ T_A \). Accordingly, any element of \( \mathcal{C}_\mathcal{A} \) is, to some degree, a “fixed point” of the adversarial game between \( \mathcal{A} \) and \( \mathcal{R} \).

It is known [19], [20] that a Nash equilibrium of every finite zero-sum game is a fixed point of the game’s best response correspondence. A Nash equilibrium is a solution in which each player correctly predicts what the other players will do and responds optimally, so that no player can improve their utilities by choosing different strategies. But we could not apply the classic Kakutani fixed-point theorem [21] in this adversarial game. On one hand, the adversarial game between \( \mathcal{A} \) and \( \mathcal{R} \) is a nonzero-sum game. The purpose of \( \mathcal{A} \) is allocating the target candidate \( x_0 \) as the leading one in the aggregation result; \( \mathcal{R} \) pursues a best aggregated ranking list by the pairwise comparisons. On the other hand, (14) is not a convex subset of \( \mathbb{R}^n \).

But the situation will change with the choice of \( \mathcal{R} \). If \( \mathcal{R} \) executes the parametric-model-based ranking aggregation approaches, e.g., HodgeRank in Sec 2.3 and RankCentrality in Sec 2.4, \( \mathcal{A} \) would convert her/his manipulation purpose of ranking aggregation result into the constraints on the latent preference score like (15). Obviously, (15) is a convex subset of \( \mathbb{R}^n \). With the convex set, we can further discuss the fixed points of operators in the adversarial game.

**Definition 3 (Fixed Points of Operators).** The set of fixed points of an operator \( T : \mathcal{X} \to \mathcal{X} \) is denoted by \( \text{Fix}[T] \), i.e.
\[
\text{Fix}[T] = \{ x \in \mathcal{X} \mid Tx = x \}.
\] (21)

The following proposition shows the existence of fixed point for the continuous operators.

**Proposition 1.** Let \( (\mathcal{X}, d) \) be a complete metric space with distance (or metric) \( d \) and \( T : \mathcal{X} \to \mathcal{X} \) be a Lipschitz continuous operator with the Lipschitz constant \( L \in [0, 1) \). Given \( x_0 \in \mathcal{X} \), define the following sequence \( \{ x_n \} \)
\[
x_{n+1} := Tx_n, \quad \forall \ n \in \mathbb{N}.
\] (22)

Let \( d_n(x) := d(x_n, x) \). The sequence \( \{ d_n(x) \} \) is monotonically decreasing:
\[
d(x_{n+1}, x) \leq Ld(x_n, x), \quad \forall \ n \in \mathbb{N}.
\] (23)

Moreover, \( x \) is the unique fixed point of \( T \).

For \( x, y \in \mathcal{X} \), we denote the preference relation between \( x \) and \( y \) by \( x \succeq y \). If \( x \geq y \) but \( y \not\geq x \), we denote this by \( x \succ y \). If \( x \geq y \) and \( y \succeq x \), we denote this by \( x \sim y \). Since a preference relation is complete, we are able to establish a preference ordering between any two elements of \( \mathcal{X} \), and transitivity prevents cycles in preferences. Notice that there exist infinitely many choices of \( \theta_\mathcal{A} \) and \( \theta_\mathcal{R} \), we consider the adversarial games in which both \( \mathcal{A} \) and \( \mathcal{R} \) may have infinitely many pure strategies.

**Definition 4.** A continuous operator game is a tuple \( (\mathcal{P}, \mathcal{S}, \{ T_p \}) \), where
- \( \mathcal{P} \) is a finite (non-empty) set of players in the game,
- \( \mathcal{S} \) is the set of strategy profiles as
\[
\mathcal{S} = \bigcup_{p \in \mathcal{P}} \mathcal{S}_p,
\] (24)
where \( \mathcal{S}_p \) denotes non-empty compact metric space of each player \( p \in \mathcal{P} \),
- \( T_p : \mathcal{S}_p \to \mathbb{R}^n \), \( \forall \ p \in \mathcal{P} \) (25)
is the continuous operator acting as the payoff function in traditional game.

A compact metric space is a general mathematical structure for representing infinite sets that can be well approximated by large finite sets. One important fact is that, in a compact metric space, any infinite sequence has a convergent sub-sequence. We next state the analogue of Nash’s theorem for continuous operator games. For continuous utility function, this result is known as Glicksberg’s theorem [22], [23].

**Theorem 1.** Every continuous operator game has a mixed strategy Nash equilibrium.

According to this result, we know that the desired latent preference score \( \theta_\mathcal{A} \in \mathcal{C}_\mathcal{A} \) could be a Nash equilibrium of the adversarial game. It means that the purposeful attack behavior of \( \mathcal{A} \) has the opportunity of success. As discussed before, we pay attention to two famous rank aggregation methods, HodgeRank and RankCentrality, as the victim \( \mathcal{R} \). Next, we will discuss how to construct the payoff function of \( \mathcal{A} \) and form the corresponding adversarial game with the specific victim \( \mathcal{R} \).

## 4 Adversarial Game with HodgeRank

Now we are ready to propose the payoff operator and the concrete attack strategies of \( \mathcal{A} \) to execute the target attack against HodgeRank. Remember that HodgeRank aggregates the pairwise comparisons \( \{ G, y, w \} \) by solving a weighted least-squared problem as (7). When the victim \( \mathcal{R} \) is designated as HodgeRank, the corresponding adversarial game would satisfy the following claims:

- The set of players contains two members:
\[
\mathcal{P} := \{ \mathcal{A}, \mathcal{R} \},
\] (26)
where \( \mathcal{R} \) represents the HodgeRank.
- The strategy profile \( \mathcal{S}_\mathcal{R} \) is
\[
\mathcal{S}_\mathcal{R} := \mathbb{R}^n,
\] (27)
which is a non-empty compact metric space.
- The continuous operator of \( \mathcal{R} \) bases on the weighted least-squared problem (7).
By introducing the comparison matrix of $C$, we can change HodgeRank problem (7) into a matrix formulation:

$$
\ell_{R}(\theta) := \frac{1}{2} \sum_{(i,j) \in E} w_{ij}^* (y_{ij} - \theta_j + \theta_j)^2, 
$$

$$
= \frac{1}{2} \| y - C\theta \|^2_{L_2}, 
$$

$$
= \frac{1}{2} \| W^2 y - W^2 C\theta \|^2_{L_2}, 
$$

where the comparison matrix $C = \{c_{ml}\}_{m \in [N], l \in [n]} \in \{-1, 0, 1\}^{N \times N}$ indicates the edge flow of comparison graph $G$, $W^2 = \text{diag}(\sqrt{w^*})$ and

$$
\sqrt{w^*} = \left[ \sqrt{w_{12}^*}, \sqrt{w_{13}^*}, \ldots, \sqrt{w_{n,n-1}^*} \right]^T.
$$

We index each directed edge $(i, j)$ by $m$ (noted as $(i, j) \sim m$) in the following way:

$$
m = \begin{cases} 
(i - 1) \cdot (n - 1) + j - 1, & \text{if } 1 \leq i < j \leq n, \\
(i - 1) \cdot (n - 1) + j, & \text{if } 1 \leq j < i \leq n.
\end{cases}
$$

Then the elements of $C$ are organized as

$$
c_{ml} = \begin{cases} 
1, & \text{if } (i, j) \sim m, l = i, \\
1, & \text{if } (i, j) \sim m, l = j, \\
0, & \text{otherwise}.
\end{cases}
$$

With simple calculation, we will know that

$$
1 = [1, 1, \ldots, 1]^T \in \mathbb{R}^n
$$

is an element in the null spaces of $C$ and $W^2 C$

$$
C \cdot 1 = 0, \quad W^2 C \cdot 1 = 0.
$$

As a consequence, the classic Tikhonov regularization is introduced to generate a unique solution of HodgeRank

$$
\min_{\theta \in \mathbb{R}^n} \ell_{R}(\theta) + \| \Gamma \theta \|^2_{L_2},
$$

where $\Gamma$ is always chosen as $\lambda_0 I$. Then the corresponding optimization problem and the operator of $R$ becomes

$$
(T_R) \min_{\theta \in \mathbb{R}^n} \ell_{R}(\theta) + \lambda_0 \| \theta \|^2_{L_2}.
$$

When the regularization parameter $\lambda_0 > 0$ satisfies

$$
\text{Null}(C^T W^* C + \lambda_0 I) = \{0\},
$$

the objective function of regularized HodgeRank (34) will be strongly convex. Actually, the ranker $R$ always sets $\lambda_0$ to be a small constant as $C^T W^* C$ is a semi-definite positive matrix where $W^* = (W^2)^{-1} W^2$. Given the original pairwise comparisons $\{G, y, w^*\}$, we note $\bar{\theta}$ is the corresponding ranking score which is a solution of (34) and holds

$$
0 \in \partial \ell_{R}(\bar{\theta}) + 2\lambda_0 \bar{\theta}.
$$

Now we try to construct the continuous operator of $A$ and the corresponding strategy profile. It starts with complete information and perfect feedback, and then expands the analysis to the adversary with incomplete information or imperfect feedback.

**Complete and Perfect Game.** To archive the goal of manipulation, the adversary $A$ replaces the pairwise comparisons $w^*$ with $\hat{w}$ and hopes that the regularized HodgeRank (34) will aggregate $\hat{w}$ into a unique ranking score $\theta_A \in C_A$. With complete information and perfect feedback, there do not exist unknown pairwise comparisons $w^*_u$ in (16) and the desired relative ranking score $\theta_A$ is a permutation of $\bar{\theta}$ as

$$
\theta_A = p(\bar{\theta}),
$$

where $p : R^N_A \rightarrow C_A$ just redistributes the elements of $\bar{\theta}$ without modifying their values.

The optimal condition (36) is an enlightenment of $A$. When the perturbed pairwise comparisons $\hat{w}$ and desired relative ranking score $\theta_A$ correspond to the input and output of $R$, it implies that

$$
\theta_A = \arg\min_{\theta \in \mathbb{R}^n} \frac{1}{2} \| W^2 y - W^2 C\theta \|^2_{L_2} + \lambda_0 \| \theta \|^2_{L_2},
$$

where $W^2 = (W^2)^{-1} W^2$ and $\hat{w} = [\hat{w}_{12}, \hat{w}_{13}, \ldots, \hat{w}_{n,n-1}]^T$. Consequently, $\hat{w}$ and $\theta_A$ also satisfy the following condition like (36):

$$
(C^T W^* + 2\lambda_0 I) \theta_A = C^T W y,
$$

where $W = W^2 C^T W^2 = \text{diag}(\sqrt{\hat{w}})$. As a necessary condition of $\theta_A$, (39) illustrates the relationship between the pairwise comparisons $\hat{w}$ and the latent preference score considered by HodgeRank. On one hand, the ranker $R$ deduces the solution $\theta_A$ from (39) with given data $\hat{w}$. On the other hand, the (39) could instruct $A$ to modify the pairwise comparisons. Once $A$ has determined the target $\theta_A$, the corresponding attack strategy will be obtained by (39). In addition, in order to cover up his/her own behavior, $A$ wants to make as few changes as possible to the original data $w^*$:

$$
\hat{w} \in \arg\min_{w \in \mathbb{R}^N} \frac{1}{2} \| w - \hat{w} \|^2_{L_2},
$$

With some abuse of symbols, we treat $w^*$ and $\hat{w}$ as real number vectors in $\mathbb{R}^N$. Although $w^*$ represents the numbers of pairwise comparisons for each type, we can turn the integer into the real number by dividing the total amount as

$$
\hat{w} = \frac{w^*}{\sum_{(i,j) \in E} w^*_{ij}}.
$$

This operator will lead to a scale-different solution in (34). However, these two ranking results (integer version and real number version) will be consistent as the ranking operator is scale-invariant:

$$
\pi_{\theta} = \pi_{\hat{\theta}}.
$$
Furthermore, the real number \( \hat{\omega} \) can be converted to integer with the complete information. Avoiding the challenge integer programming and keeping the same ranking result, we adopt the real number version to construct the attack strategy.

Combining (39) and (40), we construct the continuous operator of \( \mathcal{A} \) as

\[
\begin{align*}
&\text{minimize} & & \frac{1}{2} \| w - w' \|^2_2 \\
&\text{subject to} & & C^T \text{diag}(C + 2\lambda_0 I)\theta_A - y)w = 0, & (43)
\end{align*}
\]

where \( \text{diag}(x) \) is a diagonal matrix with \( x \in \mathbb{R}^N \) being the diagonal components and \( \lambda_0 > 0 \) is a small constant. We can adopt the Alternating Direction Method of Multipliers (ADMM) [24], [25], [26], [27] method to solve \( \hat{\omega} \) from (43).

The detailed process is involved in the incomplete and imperfect case. Here we give the following theorem and show that \( \theta_A \) is a fixed point of the compositing operator \( T_R \circ T_A \) which is also the Nash equilibrium of the adversarial operator game between HodgeRank \( T_R \) (34) and \( T_A \) (43).

**Theorem 2.** Given the desired latent preference score \( \theta_A \), the corresponding attack strategy \( \hat{\omega} \) are the accumulation point of (43). If the solution set to the Karush–Kuhn–Tucker (KKT) conditions for (43) is not empty and the penalty parameter \( \rho \) of ADMM (Algorithm 1) satisfies \( \rho \in (0, (1 + \sqrt{7})/2) \), \( \theta_A \) is a global optimal solution of (34).

We provide the detailed proof process in the supplementary material, available online. This theorem indicates that the adversary will be able to successfully control the ranking results of HodgeRank once he/she has complete information. It reveals an imminent threat to the integrity of the ranking results. What’s more serious is that even with partial information, the adversary \( \mathcal{A} \) still has the chance to manipulate HodgeRank. Next, we will discuss the incomplete and imperfect game between HodgeRank and the adversary \( \mathcal{A} \).

### Incomplete and Imperfect Games

With the Definition 1, we know that there must exist some pairwise comparisons \( w_{i,j}^* \) that cannot be modified by attacker \( \mathcal{A} \) when he/she plays the incomplete game with the victim \( \mathcal{R} \). The obstruction \( w_{i,j}^* \) will lead \( \mathcal{R} \) to the other ranking result which is unable to achieve the goal of \( \mathcal{A} \). If \( \mathcal{A} \) constructs the \( \hat{\omega} \) by (43) which only replaces the known part of \( w_{i,j}^* \), the modified data given to HodgeRank will be

\[
\hat{\omega} = \hat{\omega} + w_{i,j}^*.
\]

By Theorem 2, if the adversary \( \mathcal{A} \) hopes \( \theta_A \) to be the aggregated result, it will holds that

\[
C^T (W + W_{i,j}^*) \theta_A + 2\lambda_0 \theta_A = C^T (W + W_{i,j}^*) y
\]

(44)

Notice that \( \hat{\omega} \) is the solution of (43), we have

\[
(C^T \mathcal{W}C + 2\lambda_0 I)\theta_A = C^T \mathcal{W} y
\]

(45)

Substituting (45) into (44), we have

\[
C^T \mathcal{W}^* \mathcal{C} \theta_A = C^T \mathcal{W}^* y.
\]

(46)

is an unrealistic requirement for the adversary \( \mathcal{A} \). Without any prior knowledge of \( w_{i,j}^* \), \( \mathcal{A} \) could not choose \( \theta_A \in C_A \) to satisfy (46). To eliminate the uncertainty of incomplete information, the attacker \( \mathcal{A} \) will introduce some constraints on \( w_{i,j}^* \), which involve the feedback \( \theta_R \) from the victim \( \mathcal{R} \). Different from the perfect case, \( \mathcal{A} \) only masters the ranking result \( \pi_R \). Let \( \theta_R \) be the feedback of \( \mathcal{R} \) utilized by \( \mathcal{A} \):

\[
\theta_R = \begin{cases}
\hat{\theta}, & \text{perfect case,} \\
\{ s: N^n \rightarrow R^n \}, & \text{imperfect case,}
\end{cases}
\]

where \( s: N^n \rightarrow R^n \) assigns scores to each item in a ranking list according to its position. Furthermore, \( \mathcal{A} \) constructs \( \theta_A \in C_A \) as the pursued latent preference score to construct \( \hat{\omega} \) and mislead \( \mathcal{R} \): \( \theta_A = p(\theta_R) \) where \( p: R^n \rightarrow R^n \) is a permutation of \( \theta_R \) to assure \( \theta_A \in C_A \). We will conduct a profound analysis and design the corresponding algorithm based on the above analysis of \( w_{i,j}^* \) and \( \theta_R \). Eliminating the interference of \( w_{i,j}^* \) is the most urgent issue for \( \mathcal{A} \) in the incomplete game. The adversary \( \mathcal{A} \) should forecast the effect of \( w_{i,j}^* \) and construct the corresponding attack strategy \( \hat{\omega} \). Here we split the attack strategy \( \hat{\omega} \) into two variables like (16) and each variable plays a different role in the following attack: \( \hat{\omega} = \omega_k + w_u \). Generally speaking, \( \omega_k \) substitutes the accessible part \( w_{i,j}^* \) as the actual attack strategy and \( w_u \) imitates the behavior of \( w_{i,j}^* \) in \( T_R \) (34). Although the adversary does not know \( w_{i,j}^* \) and \( \theta_R \) would be imperfect, \( \theta_A \) could be the solution of an optimization problem like (34) and the first-order optimal condition holds:

\[
C^T (W_k^* + W_u^*) \mathcal{C} \theta_R + 2\lambda_0 \theta_R = C^T (W_k^* + W_u^*) y
\]

(48)

If \( \mathcal{A} \) hopes \( w_u \) has the similar behavior like \( w_{i,j}^* \) in (34), \( w_u \) will satisfy

\[
C^T (W_k^* + W_u^*) \mathcal{C} \theta_R + 2\lambda_0 \theta_R = C^T (W_k^* + W_u^*) y
\]

(49)
With the help of $\mathbf{w}_k$, the adversary will come back to a situation like the complete game. By Theorem 2, $\mathbf{w}_k$ and $\mathbf{w}_k$ will play an ensemble to mislead $\mathcal{R}$ as

$$C^T(W_k + W_k^c)C\theta_A + 2\lambda_0\theta_A$$

$$= C^T(W_k + W_k^c)y.$$  (50)

Adopting the minimal modification like (40) as the objective function, the continuous operator of $\mathcal{A}$ in the incomplete game will be

$$\begin{align*}
\text{minimize} & \quad \frac{1}{2} \| \mathbf{w}_k - \mathbf{w}_k^c \|^2_2 \\
(T_\mathcal{A}) \quad \text{subjectto} & \quad B_\mathcal{R}(\mathbf{w}_k^c + \mathbf{w}_k) = 0, \\
& \quad B_\mathcal{A}(\mathbf{w}_k + \mathbf{w}_k) = 0.
\end{align*}$$  \hspace{1cm} (51)

where

$$\begin{align*}
B_\mathcal{R} &= C^T \text{diag}((C + 2\lambda_0I)\theta_\mathcal{R} - \mathbf{y}) \\
B_\mathcal{A} &= C^T \text{diag}((C + 2\lambda_0I)\theta_\mathcal{A} - \mathbf{y}).
\end{align*}$$  \hspace{1cm} (52)

Here we adopt the ADMM to solve the attack operation $\mathbf{w}_k$. Let the augmented Lagrangian function $L(\mathbf{w}_k, \mathbf{w}_k^c, \mu_1, \mu_2)$ of optimization problem (51) be

$$L(\mathbf{w}_k, \mathbf{w}_k^c, \mu_1, \mu_2)$$

$$= \frac{1}{2} \| \mathbf{w}_k^c - \mathbf{w}_k \|^2_2 + \left( \mu_1, (B_\mathcal{R}(\mathbf{w}_k^c + \mathbf{w}_k)) \right) + \mu_2, (B_\mathcal{A}(\mathbf{w}_k + \mathbf{w}_k)) \right)$$

$$+ \frac{\rho_1}{2} \| B_\mathcal{R}(\mathbf{w}_k^c + \mathbf{w}_k) \|^2_2 + \frac{\rho_2}{2} \| B_\mathcal{A}(\mathbf{w}_k + \mathbf{w}_k) \|^2_2,$$

where $\mu_1, \mu_2 \in \mathbb{R}^n$ are the Lagrangian multiplier variables, and $\rho_1, \rho_2 > 0$ are penalty parameters.

The ADMM consists of the following iterations:

$$\begin{align*}
\mathbf{w}_k^{t+1} &= \text{argmin}_{\mathbf{w}_k \geq 0}\ L(\mathbf{w}_k, \mathbf{w}_k^c, \mu_1, \mu_2), \\
\mathbf{w}_k^c^{t+1} &= \text{argmin}_{\mathbf{w}_k^c \geq 0}\ g_{\gamma t+1}^w, \mathbf{w}_k^c + \frac{1}{\gamma^t} \| \mathbf{w}_k^c - \mathbf{w}_k \|^2_2, \\
\mu_1^{t+1} &= \mu_1^t + \rho_1, B_\mathcal{R}(\mathbf{w}_k^{t+1} + \mathbf{w}_k^c), \\
\mu_2^{t+1} &= \mu_2^t + \rho_2, B_\mathcal{A}(\mathbf{w}_k^{t+1} + \mathbf{w}_k^c),
\end{align*}$$  \hspace{1cm} (53)

where $\gamma > 0$ is a proximal parameter and $g_{t+1}$ is defined

$$g_{\gamma t+1}^w := \nabla_{\mathbf{w}_k}L(\mathbf{w}_k^{t+1}, \mathbf{w}_k, \mu_1^t, \mu_2^t) \bigg|_{\mathbf{w}_k = \mathbf{w}_k^c}. \hspace{1cm} (54)$$

Specifically, the update rule of $\mathbf{w}_k$ is

$$\mathbf{w}_k^{t+1} = \text{Proj}_+ \left[ \left( I + \rho_2 B_\mathcal{R}^T B_\mathcal{R} \right)^{-1} \left( \mathbf{w}_k^c - \rho_1 B_\mathcal{R}^T \mathbf{w}_k^c \mathbf{w}_k^c \mathbf{w}_k^c \right) \right]. \hspace{1cm} (55)$$

and the update rule of $\mathbf{w}_k$ is

$$\mathbf{w}_k^{t+1} = \text{Proj}_+ \left[ \left( \frac{1}{\gamma} I + \rho_1 B_\mathcal{R}^T B_\mathcal{R} + \rho_2 B_\mathcal{A}^T B_\mathcal{A} \right)^{-1} \left( B_\mathcal{R}^T \mu_1^t + \rho_1 B_\mathcal{R}^T B_\mathcal{R} \mathbf{w}_k^c + B_\mathcal{A}^T \mu_2^t \right) \right.

\left. + \rho_2 B_\mathcal{A}^T B_\mathcal{A} \mathbf{w}_k^{t+1} - \frac{1}{\gamma} \mathbf{w}_k \right]. \hspace{1cm} (56)$$

where $I$ is an identity matrix. According to [28], the adopted ADMM has a linear convergence rate for solving the concerned optimization problem (51). The general framework for $\mathcal{A}$ to construct the attack strategies in complete game and incomplete game is illustrated as Algorithm 1.

Remark. As we introduce a variable $\mathbf{w}_k$, to estimate the unknown pairwise comparisons $\mathbf{w}_k$, there exist two penalty parameters $\rho_1$ and $\rho_2$ in Algorithm 1. The complete game discussed Theorem 1 (only involved single penalty $\rho$) is a special case of the incomplete game as $\mathbf{w}_k \equiv 0$ in (51). We provide the augmented Lagrangian function in the complete case for Algorithm 1 in the supplementary material, available online. It is noteworthy that we can obtain that $\theta_\mathcal{A}$ is a fixed point of $T_\mathcal{A} \circ T_\mathcal{R}$ when $\theta_\mathcal{R} = \overline{\theta}$ and $\mathcal{A}$ only masters incomplete information $\mathbf{w}_k$ through Theorem 1. Establishing and characterizing the fixed points of the incomplete and imperfect dynamic system $T_\mathcal{A} \circ T_\mathcal{R}$ are challenging but still open problems.

5 Adversarial Game With Spectral Method

In this section, we construct the operator of the adversary to execute the target attack against RankCentrality [4]. By solving the eigenvector problem (13), RankCentrality obtains the latent preference score $\overline{\theta}$, which is the leading eigenvector of the empirical transition matrix $\mathbf{P}$. When the victim $\mathcal{R}$ is designated as RankCentrality, the corresponding adversarial game would satisfy the following claims:

- The set of players contains two members $\mathcal{P} := \{\mathcal{A}, \mathcal{R}\}$. Here $\mathcal{R}$ represents the RankCentrality.
- The strategy profile $\mathbf{s}_\mathcal{R}$ is $\mathbb{R}^n$, which is a non-empty compact metric space.
- The continuous operator of $\mathcal{R}$ is the eigenvector problem as (12) and (13).

Different from attacking against HodgeRank, the operator of RankCentrality is not involved with the pairwise comparisons $\{\mathbf{G}, \mathbf{y}, \mathbf{w}\}$ directly. As a consequence, $\mathcal{A}$ could not produce the poisoned data $\tilde{\mathbf{w}}$ with the help of optimal conditions for the leading eigenvector problem (12) and (13). Instead of manipulating the weights of pairwise comparisons, the attacker $\mathcal{A}$ would like to counterfeit the independent and identically distributed (i.i.d.) pairwise samples from $\mathbf{V}$ by $\theta_\mathcal{A} = \{\theta_1, \theta_2, \ldots, \theta_n\} \in \mathcal{C}_\mathcal{A}$. However, such a kind of the manipulation is prohibited from the unknown distribution which gives the probability that $i > j$ turns out true. Moreover, verification of general linear model assumptions like (2) is challenging under the most practical scenarios. Sampling pairwise comparisons from general linear models with $\theta_\mathcal{A} \in \mathcal{C}_\mathcal{A}$ would be a sub-optimal solution.

To design distribution-free methods for generating paired data and conduct the tractable target attack strategies
on spectral ranking algorithm, the attacker $A$ will construct a new Markov chain with the target invariant distribution $\theta_A \in \mathcal{C}_A$. By (13), we know that $\theta_A$ must be the largest left eigenvector of an unknown transition matrix $P_A$:

$$\theta_A = \theta_A \cdot P_A. \quad (57)$$

Once the attacker $A$ obtains $P_A$, he/she could elicit the poisoned data $\tilde{w}$ from the corresponding transition matrix $P_A$ like (11). Here we adopt the Markov Chain Monte Carlo (MCMC) method to generate the desired Markov chain $P_A$ with the constraint (57), which can be understood as the other form of “optimal condition”. The basic idea underlying the MCMC is to construct a Markov chain with the desired distribution $\theta_A$ as its invariant distribution. As the procedure runs long enough, the samples generated from the Markov chain serve as a good approximation to the desired samples drawn from the unknown pairwise comparison distribution. Given the state space $V$ and the initial distribution, the desired Markov chain is characterized by its transition matrix $P_A$ whose leading left eigenvector is $\theta_A$. Constructing a (row) stochastic matrix with a prescribed left eigenvector $\theta_A$ associated with the Perron root 1 is a special type of the so-called “inverse eigenvalue problem” [29], [30]. To be more specific, since the main concern of $A$ is to maintain $\theta_A$ as the leading eigenvector of the matrix $P_A$, $A$ will solve an inverse eigenvalue problem as the continuous operator. Establishing the inverse eigenvalue problem and ensuring that $\theta_A$ becomes a fixed point of the continuous operator game need some constraints on $P_A$. The obvious structural constraint is that $P_A$ will be row stochastic. In addition, we introduce the worst-case performance measurement as the other constraint on $P_A$. Considering the possible incomplete information and imperfect feedback, the worst-case performance measurement tells how good the approximation of MCMC to the unknown pairwise comparison distribution is. In the literature, one of the commonly employed measures for evaluating the performance of an MCMC algorithm is the so-called asymptotic variance [31] which is the focus of this paper.

Suppose that $V = \{v_0, v_1, \ldots\}$ is a discrete time Markov chain on $V$ with some transition probability matrix $P_A$ and invariant distribution $\theta_A$. We then use the time average

$$\frac{1}{m} \sum_{i=0}^{m-1} f(v_i) \quad (58)$$

as an estimation for the space average

$$E_{\theta_A}[f] = \sum_{v \in V} f(v) \cdot \theta_v, \quad (59)$$

where $f$ is a real-valued function defined on $V$. By the strong law of large numbers, we should have

$$\frac{1}{m} \sum_{i=0}^{m-1} f(v_i) \xrightarrow{a.s.} E_{\theta_A}[f]. \quad (60)$$

Moreover, the asymptotic mean squared error

$$v(f, P_A) := \lim_{m \to \infty} m \cdot E_{\mu_0} \left[ \frac{1}{m} \sum_{i=0}^{m-1} f(v_i) - E_{\theta_A}[f] \right] \quad (61)$$

is equivalent to the asymptotic variance when the initial distribution $\mu_0$ on $V$ has the transition matrix $P_A$. Note that the mean square errors of time averages and their variance coincide up to an order $1/m$ for large times of sampling $m$. The difference is equal to the square of the bias, which is of order $1/n$. It is known in the literature [32] that for any irreducible transition matrix $P_A$, the limit in (61) exists no matter what initial distribution $\mu_0$ is. Obviously, a smaller asymptotic variance indicates a better approximation. The real challenge is that in the context of MCMC methods, we also look for a matrix $P_A$ of which the corresponding worst-case asymptotic variance is minimized. Here the worst-case asymptotic variance means that the game between the attacker and the ranker is the incomplete information case: it is possible to exist some unknown comparisons for the attacker but they are known to the ranker. When the adversary $A$ only accesses the incomplete information and the imperfect feedback, blindly pursuing the efficiencies of MCMC algorithms would deviate from the purpose of $A$. As a consequence, it is of great interest to analyze the asymptotic variance over all functions $f$ in the most pessimistic view. In the present paper, we take the maximum of asymptotic variance (61) over functions of interest $F$ as the criterion for evaluating the performance of MCMC algorithms. Moreover, we adopt the worst-case analysis to derive a lower bound of the asymptotic variance over general Markov chains with invariant probability $\theta_A$. The worst scenario of the asymptotic variance for each fixed $P_A$ amounts to solving

$$\inf_{P} \sup_{f \in F} v(f, P). \quad (62)$$

Assuming the generic condition that the Perron root of $P_A$ is unique and identifying the function $f$ in (58) and (59) on $V$ as a vector $f \in \mathbb{R}^n$, it is known [32] that the asymptotic variance $v(f, P_A)$ (61) can be expressed as

$$v(f, P_A) = 2 \cdot \left( \left( I - P_A - \Theta \right)^{-1} \Theta \right) \cdot \left( f - E_{\theta_A}[f] \right) \cdot \left( f - E_{\theta_A}[f] \right)_{\theta_A}, \quad (63)$$

where $E_{\theta_A}[f]$ is defined as

$$E_{\theta_A}[f] = \sum_{i \in [n]} f_i \cdot \hat{\theta}_i, \quad f = [f_1, \ldots, f_n]^T,$$

and $\langle \cdot, \cdot \rangle_{\theta}$ stands for the weighted inner product with respect to $\theta$

$$\langle x, y \rangle_{\theta} = y^T \Theta x = (\Theta^T x, \Theta^T y), \quad (64)$$

$\Theta = \text{diag}(\theta)$ is a diagonal matrix with $\theta \in \mathbb{R}^n$ being the diagonal components. Throughout the rest of this paper, we restrict $f$ in the space of mean zero vector $f \in F$

$$F := \left\{ f \in \mathbb{R}^n \mid E_{\theta_A}[f] = 0 \right\}, \quad (65)$$
and by re-centering $f$ if necessary, we may assume without loss of generality henceforth that $E_{\theta_A}[f] = 0$. Then we have
\[
    v(f, P_A) = 2 \cdot \left( (I - P_A)^{-1} f, f \right)_{\theta_A} - \left( f, f \right)_{\theta_A},
\]
where $I - P_A$ is the fundamental matrix of $P_A$. Since $v(f, P_A)$ in (66) is homogeneous in $f$, it suffices to assume further that
\[
    \|f\|_{\theta_A} := \left( f, f \right)_{\theta_A} = 1.
\]
Thus, for a fixed transition probability matrix $P_A$, the asymptotic variance $v(f, P_A)$ depends on the underlying vector $f$. The worst-case analysis aims to solving
\[
    \inf_{P_A} \sup_{f \in F, \|f\|_{\theta_A} = 1} \left( (I - P_A)^{-1} f, f \right)_{\theta_A}.
\]

### 5.1 Reversible Stochastic Transition Matrix

Many popular MCMC methods assume reversibility. One good reason is that a reversible matrix with minimum asymptotic variance is achievable in closed form. It is known that minimizing the asymptotic variance of reversible matrix is equivalent to finding the reversible $P_A$ whose second largest eigenvalue is minimized [33]. Here we give the definition of the reversible stochastic transition matrix.

**Definition 5 (Reversibility of Stochastic Transition Matrix).** A row stochastic matrix $P$ is said to be reversible relative to non-negative $\theta$ if and only if it satisfies the so-called detailed balance condition,
\[
    \Theta P = P^T \Theta,
\]
where $\Theta = \text{diag}(\theta)$ is a diagonal matrix with $\theta \in \mathbb{R}^n$ being the diagonal components.

The following theorem dissects the spectral structure of the reversible stochastic transition matrix $P_A$ with the given leading eigenvector $\theta_A$. This is a necessary condition of the desired matrix to achieve the purpose of $A$. Its proof can be found in the supplementary material, available online which is a special case of [34, Theorem 1].

**Theorem 3.** Given a reversible stochastic transition matrix $P_A$ and its leading eigenvector $\theta_A$, we have:

1) The second largest eigenvalue $\lambda_2$ of stochastic matrix $P_A = \{P_j\}$, reversible w.r.t $\theta_A = [\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n]^T$ satisfies
\[
    \lambda_2 \geq -\frac{\hat{\theta}_1}{1 - \hat{\theta}_1}, \text{ where } 0 < \hat{\theta}_1 \leq \hat{\theta}_2 \leq \ldots \leq \hat{\theta}_n.
\]

2) If the second largest eigenvalue of matrix $P_A$ attains its lower bound as
\[
    \lambda_2 = -\frac{\hat{\theta}_1}{1 - \hat{\theta}_1}, \quad \lambda_2 \leq -\frac{\hat{\theta}_1}{1 - \hat{\theta}_1}, \quad \lambda_2 \leq \ldots \leq \hat{\theta}_n.
\]

$P_A$ has the form:
\[
    P_A = \begin{pmatrix}
    0 & \hat{\theta}_2 & \cdots & \hat{\theta}_n \\
    1 - \hat{\theta}_1 & 0 & \cdots & \hat{\theta}_n \\
    \vdots & \vdots & \ddots & \vdots \\
    1 - \hat{\theta}_1 & 1 - \hat{\theta}_1 & \cdots & 0
    \end{pmatrix}
\]
(71)

where $P_{A_k}$ is in detailed balance with the vector
\[
    \theta_A/\theta_1 = [\hat{\theta}_2, \ldots, \hat{\theta}_n],
\]
and has constant row sums.

Based on the Theorem 3, we can recursively construct the desired reversible stochastic transition matrix as following. Begin with $\theta^{(k)} := \theta_A$, for $k = 1, \ldots, n - 1$, we define
\[
    P^{(k)} := \begin{pmatrix}
    0 & \theta^{(k)}_{k+1} & \cdots & \theta^{(k)}_{n} \\
    1 - \alpha_{k+1} & 0 & \cdots & \theta^{(k)}_{n} \\
    \vdots & \vdots & \ddots & \vdots \\
    1 - \alpha_{k+1} & 1 - \alpha_{k+1} & \cdots & 0
    \end{pmatrix},
\]
where
\[
    \beta_{k+1} := 1 - \theta^{(k)}_{k+1} \quad (73a)
\]
\[
    \alpha_{k+1} := 1 - \theta^{(k)}_{k+1} / \beta_{k+1} \quad (73b)
\]
\[
    \theta^{(k+1)} := \theta^{(k)} / \beta_{k+1}. \quad (73c)
\]

The following corollary shows that the recurrent construction of $P_A$ could attain the minimal value of the second largest eigenvalue.

**Corollary 1.** The following $P_A$ is a reversible stochastic matrix:
\[
    P_A := P^{(1)} + \left( 1 - \sum_{j=1}^{n-1} P^{(1)}_{nj} \right) e_n e_n^T,
\]
(74)

where $P^{(1)}$ denotes the $(n, j)$ entry of $P^{(1)}$ and $e_n$ is the $n$-th standard unit vector of $\mathbb{R}^n$. Furthermore, the second largest eigenvalue $\lambda_2$ of $P_A$ is minimal among all reversible matrices with the stationary distribution $\theta_A$.

We end the discussion of the reversible stochastic transition matrix case with the following theorem which establishes the lower bound of worst-case asymptotic variance [35] of (74). With Corollary 1, we know that $\theta_A$ would be a fixed point of the complete game between the perpetrator and the spectral ranking algorithm.

**Theorem 4.** Let $P_A$ be a stochastic, reversible with respect to $\theta_A$ and $v(f, P_A)$ be the asymptotic variance of
\[
    \frac{1}{m} \sum_{k=0}^{m-1} f(v_k)
\]
for norm 1 function $f : [n] \to \mathbb{R}$ like (67), where $V = \{v_0, v_2, \ldots\}$ is a discrete time Markov chain on $V$ with some transition probability matrix $P_A$ and invariant distribution $\theta_A$. Then
be the number of the candidates involved in the aggregate transition matrix.

and any stochastic matrix which attains the lower bound must satisfy

\[ \lambda_2 = -\frac{\hat{\theta}_1 - \hat{\theta}_i}{1 - \hat{\theta}_i}, \]

and hence be of the form (74).

5.2 Irreversible Stochastic Transition Matrix

However, the original RankCentrality approach does not require the reversibility of \( \mathbf{P}^* \) in (11). Next, we explore the worst-case analysis of the asymptotic variance for irreversible transition matrix. We first establish the lower bound of (66) by the following theorem which is a special case of [36, Theorem 2.1] with [37, Lemma 11]. The detailed proof is provided in the supplementary material, available online.

**Theorem 5.** Assume that \( \mathbf{P}_A \) is an irreducible transition matrix with invariant distribution \( \theta_A = [\theta_1, \ldots, \theta_n]^T \). It holds that

\[
\frac{\hat{\theta}_n}{\hat{\theta}_1 + \hat{\theta}_n} \leq v(\mathbf{P}_A),
\]

where

\[
v(\mathbf{P}_A) := \sup_{f \in \mathcal{F}} \sup_{\mathbf{f}_A = 1} v(f, \mathbf{P}_A)
= \sup_{f \in \mathcal{F}} \sup_{\mathbf{f}_A = 1} \left< (\mathbf{I} - \mathbf{P}_A)^{-1} f, f \right>_{\theta_A}
= 2\left< \mathbf{Zf}, f \right>_{\theta_A} - 1.
\]

Let \( \mathbf{Z} = (\mathbf{I} - \mathbf{P}_A)^{-1} \). From the lower bound of worst-case problem (68), we know that

\[
v(\mathbf{P}_A) \geq \max_{f \in \mathcal{F}, \mathbf{f}_A = 1} \left< \mathbf{Zf}, f \right>_{\theta_A} \geq \frac{\hat{\theta}_n}{\hat{\theta}_1 + \hat{\theta}_n}.
\]

Based on this result, our goal is to present an optimal transition matrix which achieves this lower bound (79). For simplicity, we construct the transition matrix such that

\[
\left< \mathbf{Zf}, f \right>_{\theta_A} = \frac{\hat{\theta}_n}{\hat{\theta}_1 + \hat{\theta}_n}, \quad f \in \mathcal{F}, \quad \mathbf{f}_A = 1.
\]

One possible construction of irreversible stochastic transition matrix \( \mathbf{P}_A = \{\hat{P}_{ij}\} \) is

\[
\hat{P}_{ij} = \begin{cases} 
0, & 1 \leq j < i < n, \\
\frac{\hat{\theta}_i - \hat{\theta}_k}{\hat{\theta}_1 + \hat{\theta}_n}, & 1 \leq j = i < n, \\
\frac{\hat{\theta}_j (\hat{\theta}_1 + \hat{\theta}_n)}{\hat{\theta}_1 (\hat{\theta}_1 + \hat{\theta}_n)} \sum_{k=j}^{i-1} \frac{\hat{\theta}_k - \hat{\theta}_1}{\hat{\theta}_k + \hat{\theta}_n}, & 1 \leq j < i = n, \\
\frac{2\hat{\theta}_i (\hat{\theta}_1 + \hat{\theta}_n)}{(\hat{\theta}_1 + \hat{\theta}_n)(\hat{\theta}_1 + \hat{\theta}_n)} \sum_{k=i}^{n-1} \frac{\hat{\theta}_k - \hat{\theta}_1}{\hat{\theta}_k + \hat{\theta}_n}, & 1 \leq i < j < n, \\
1 - \hat{P}_{ik}, & \text{otherwise.}
\end{cases}
\]

The following corollary states that \( \theta_A \) is the leading eigenvector of irreducible \( \mathbf{P}_A \) constructed by (81).

**Corollary 2.** For the irreversible stochastic transition matrix \( \mathbf{P}_A = \{\hat{P}_{ij}\} \) defined by (81), it holds that

1. \( \mathbf{P}_A \) is irreducible.
2. \( \hat{P}_{ij} \geq 0 \) for any \( i, j \in [n] \).
3. \( \theta_A \) is the largest left eigenvector of \( \mathbf{P}_A \).

At last, we show that the irreversible stochastic transition matrix \( \mathbf{P}_A \) in (81) will attain its lower bound of the worst-case asymptotic variance. It indicates that \( \theta_A \) will be the fixed point of the continuous operator games between \( \mathcal{A} \) and \( \mathcal{R} \) with irreversible stochastic transition matrix (81) and the minimal worst-case asymptotic variance.

**Theorem 6.** Given the stochastic transition matrix \( \mathbf{P}_A \) with its invariant distribution \( \theta_A = [\theta_1, \ldots, \theta_n]^T \) by (81), and the asymptotic variance \( v(f, \mathbf{P}) \) of the sample mean with respect to vector \( f \in \mathcal{F}, \mathbf{f}_A = 1 \) from the Markov chain with \( \mathbf{P} \) as defined in (66). Then we have \( \mathbf{P}_A \) is one of the optimal irreversible transition matrices which attains the following bound:

\[
\inf_{\mathbf{P}} \sup_{f \in \mathcal{F}, \mathbf{f}_A = 1} v(f, \mathbf{P}) = \frac{\hat{\theta}_n - \hat{\theta}_1}{\hat{\theta}_1 + \hat{\theta}_n}.
\]

Now we discuss the actions of \( \mathcal{A} \) to eliminate the unknown data \( w^*_v \) in detail. Although the above results have pursued the minimization of worst-case asymptotic variance to resist the effect of \( w^*_v \), the adversary \( \mathcal{A} \) could still seek the help of \( \mathcal{R} \)’s feedback \( \theta_R \). Regardless of \( \theta_R \) is perfect or not, \( \mathcal{A} \) reconstructs the stochastic transition matrix \( \mathbf{P}_R \) by \( \theta_R \), estimates the unknown pairwise comparisons and induces the actual attack strategy. The general framework for \( \mathcal{A} \) to construct the attack strategies is listed as Algorithm 2.

6 Experiments

In this section, three examples are exhibited with both simulated and real-world data to illustrate the validity of the proposed attack strategies on HodgeRank and RankCentrality. The first example is with simulated data while the latter two exploit real-world datasets involved crowdsourcing and election.

6.1 General Settings

Let \( n \) be the number of the candidates involved in the aggregation process of \( \mathcal{R} \). The original data is \( \{\mathbf{G}, y, w\} \). With the limitation of the space, we do not test all \( n! \) possible permutations of \([n]\) but construct some target ranking lists of \( \mathcal{A} \) as follows. Regardless of the completeness of information and the accuracy of feedback, the adversary \( \mathcal{A} \) always gets the ranking list \( \mathbf{r}_A \) aggregated by the victim \( \mathcal{R} \) with \( w^* \). \( \mathcal{A} \) chooses a candidate of \( \mathbf{r}_A \) and generates

\[
\mathbf{r}_A' = [\pi_0(t); \pi_0/\pi_0(t)], \quad t = 2, \ldots, n,
\]

where \( \pi(t) \) is the \( t \)-th candidate of \( \mathbf{r}_A \), and \( \pi/\pi(t) \) denotes an ordered sequence with \( n - 1 \) elements which deletes \( t \)-th candidates from \( \pi \). Due to the significance of top-1 candidate in rank aggregation, constructing the target ranking in this way can not only reflects the purpose of the adversary,
but also illustrates the applicability of the proposed attack methods. When attacking against HodgeRank, the original weight \( w' \) and the comparison matrix \( C \) are generated as (5) and (30). The empirical transition matrix \( P' \) is constructed as (11) when the victim is RankCentrality.

**Algorithm 2. Target Attack Against Spectral Method**

**Input:** \( w'_k, \theta_R, \theta_A, d_1, d_2 \).
1. \( A \) recovers the transition matrix \( P_R \) from the feedback \( \theta_R \) by (74) or (81)
   \[
   P_R = \text{InverseEigen}(\theta_R).
   \]
2. \( A \) constructs the transition matrix \( P_A \) with the desired score \( \theta_A \) by (74) or (81)
   \[
   P_A = \text{InverseEigen}(\theta_A).
   \]
3. \( A \) estimates the unavailable data \( w_s \) by \( P_R \) and \( w'_k \)
   \[
   w_s = \text{Reconstruction}(P_R, d_1) - w'_k, \quad w_s = \text{Reconstruction}(P_A, d_2) - w_s,
   \]
   where \( d_1, d_2 \) are two scale factors which acted as maximum out-degree \( d_{\text{max}} \) in (11), and \( \text{Reconstruction} \) is the inverse process of (11).
4. \( A \) transfers \( w_s \) into integer:
   \[
   \hat{w}_s = \text{Integer}(w_s),
   \]
   \( \mathcal{R} \) executes RankCentrality algorithm by solving (13) with the poisoned data \( \hat{w}_s \)
   \[
   \hat{\theta} = \text{RankCentrality}(\hat{w}_s)
   \]
   where
   \[
   \hat{w}_s = \hat{w}_s + w'_s,
   \]
   **Output:** The misguided ranking list \( \hat{\theta} \).

**Confrontation Scenarios.** To validate the vulnerability of ranking aggregation approaches like HodgeRank and RankCentrality, we establish different scenarios divided by the completeness of the adversary’s knowledge and the perfection of the victim’s feedback. We illustrate the attack results of the proposed methods in the three following cases:

- **Complete information and perfect feedback.** The adversary obtains the whole of the original data \( w' \) and the perfect feedback \( \hat{\theta} \) of the victim. \( \hat{\theta} \) is the output of HodgeRank or RankCentrality by solving (34) or (13) with \( \hat{w}_s \) respectively.
- **Incomplete information and perfect feedback.** The adversary obtains a part of the original data \( w' \) and the perfect feedback \( \hat{\theta} \) of the victim. The original data \( w' \) has two parts as (16): \( w'_1 \) and \( w'_2 \). Only the former part is accessible to the adversary.
- **Complete information and imperfect feedback.** The adversary obtains the whole of the original data \( w' \) and the imperfect feedback \( \pi_j \) of the victim. \( \pi_j \) is the ranking list obtained from \( w' \) by HodgeRank or RankCentrality.

**Competitors.** In the interest of fairness and evaluating the performance of different target attack strategies, we define the Incremental Pairwise Comparison Oracle (IPCO) complexity model. At each step \( s = 0, 1, 2, \ldots \), an attacker \( A \) picks a pairwise comparison \( (i, j) \in V \times V \) for \( i, j \in V \), \( i \neq j \), and the oracle \( O_A \) returns a variable \( z_{ij}^s \in \{-1, 0, 1\} \) which decides the operation of the adversary. Notice that different attack strategies correspond to different oracles. Let \( \hat{w}_s = [\hat{w}_{i1}^s, \hat{w}_{i2}^s, \ldots, \hat{w}_{in}^s] \) be the modified pairwise comparisons at \( s \) step and \( \hat{w}_0 = w' \). The strategy profile \( \hat{w}_s \) corresponds with \( z_{ij}^s \) like:

\[
\hat{w}_{ij}^s = \begin{cases} 
\hat{w}_{ij}^s - 1, & \text{if } z_{ij}^s = 1, \\
\hat{w}_{ij}^s + 1, & \text{if } z_{ij}^s = -1, \\
\hat{w}_{ij}^s, & \text{if } z_{ij}^s = 0.
\end{cases}
\]

Let \( S \) be the maximum step of accessing to the oracle \( O_A \), and the adversary \( A \) chooses one profile \( \hat{w}_s \) in \( W_S = \{\hat{w}_0, \ldots, \hat{w}_S\} \) as the attack strategy which will send to the victim \( \mathcal{R} \) as its input. When the attack strategy archives the purpose of the adversary, say that the perturbed latent preference score \( \hat{B} \) in \( \mathcal{C}_A \) is defined as (15) where \( i_0 = t \). IPCO complexity which is defined as the minimal number of interactions between \( A \) and \( O_A \) indicates the efficiency of the corresponding strategy/oracle. If \( A_1 \) has a lower IPCO complexity (smaller \( S \)) than that of \( A_2 \) against the same victim with the same purpose, we say that \( A_1 \) is more efficient than \( A_2 \). To the best of our knowledge, the proposed methods are the first overturer to attack strategies on pairwise ranking algorithms with specified purposes. We compare the proposed methods with the following three competitors: the random perturbation strategy (referred to as ’Rand.’), the naive purposeful strategy (referred to as ’Naive’) and the probabilistic model based target strategy (referred to as ’Prob.’). For the interest of fairness, we compare different attack profiles with the same number of IPCO interactions. The interaction of our proposed methods is defined as

\[
\Delta = \|w'_k - \hat{w}_s\|_1 = \sum_{(i,j) \in E} |w'_{ij} - \hat{w}_{ij}|,
\]

and the modified data with the same \( \Delta \) of each attack strategy is delivered to \( \mathcal{R} \).

- **Random** perturbation involves a completely random oracle which generates a random sequence as the operation indicators \( \{z_{ij}^s\} \) without any prior knowledge. This method does not rely on any information of the desired ranking \( \pi_A \) or latent preference score \( \theta_A \). We conjecture that the purposeless behavior of the random perturbation would not sculpt the desired results out of the poisoned data \( \hat{w}_s \). However, this strategy is still the evidence to prove the importance of \( \pi_A \) and \( \theta_A \) in the attack operations with special purpose.
- **Naive** strategy implements its oracle with the so-called “Matthew Principle”: increasing the number of the pairwise comparisons which are consistent with the desired ranking list \( \pi_A \) and decreasing the number of inconsistent comparisons. The operation variables of this oracle are produced as follows: for all \( 0 < s \leq S \), \( (i, j) \in V \times V \), \( \forall i, j \in V \), \( i \neq j \),
where $i \succ j$ represents that candidate $i$ has priority over candidate $j$ in the target ranking list $\pi_{t}^{\prime}$. Obviously, this strategy/oracle only focuses the partial orders of any paired candidates and ignores the global nature of the target ranking $\pi_{t}^{\prime}$. We speculate that the order of query would impress the efficiency of this strategy seriously.

- **Probabilistic** approach establishes its oracle with the BTL model (2) and the target preference score $\theta_{t} = \{\theta_{1}, \theta_{2}, \ldots, \theta_{n}\}$. If the probability that the pairwise comparison $(i,j)$ turns out true in $\pi_{t}^{\prime}$ is larger than a given threshold $\rho$, the oracle will return a positive variable of operation. Mathematically, for all $0 < s \leq T$, the operation variables $\{z_{ij}^{s}\}$ are given as

$$z_{ij}^{s} = \begin{cases} 1, & \text{if } \tilde{\theta}_{ij} > \rho, \\
0, & \text{if } \tilde{\theta}_{ij} = \rho, \\
-1, & \text{if } \tilde{\theta}_{ij} < \rho, \end{cases}$$

where the hyper-parameter $\rho \geq 0.5$ represents the belief in the BTL model. Moreover, the probabilistic feedbacks of this oracle reflect the global character of the desired ranking $\pi_{t}^{\prime}$.

It is worth noting that the above three attack strategies have some common issues: they fail to predict the behaviors of the victims $\mathcal{R}$ and ignore the completeness of the accessible pairwise comparisons $\mathbf{w}_{b}$. On the basis of the pairwise comparisons, the aggregated result depends on the algorithmic processes and statistical characters of the victims. It is known that the same votes would translate into different election outcomes in the social choice theory or voting theory [1]. Instead of proposing a targeted strategy toward the special victim, blindly modifying or perturbing the existing data would not archive the goal of the adversary. In addition, the completeness of observed pairwise comparisons also impacts on the success chance of the target attack operations. Once it exists, the inaccessible part $\mathbf{w}_{b}$ becomes the invertible obstruction.

**Evaluation Metrics.** Based on the IPCO complexity model, we need some metrics to measure whether the perturbed ranking $\pi_{b}$ achieves the attacker’s goal or not. More specifically, given a threshold of IPCO complexity, the larger the “similarity” between the target ranking $\pi_{t}^{\prime}$ and the perturbed ranking $\pi_{b}$ is, the more efficient the attack strategy is. Here we adopt Reciprocal Rank (R-Rank) and Kendall $\tau$ Coefficient (Kendall-$\tau$) for evaluating the correlation between two ranking lists. The first metric focuses on the top-1 candidates of the two ranking lists. Due to the significance of the leading position in ranking list, this “local” metric is the most concerned indicator for the attackers. The second metric considers the position consistency of every candidate pair in the two ranking lists. So it is a “global” metric.

**Reciprocal Rank (R-Rank).** The reciprocal rank is a statistical measure for evaluating any process that produces an order list of possible responses to a series of queries, ordered by the probability of correctness or the ranking scores. The reciprocal rank of a ranking list is the multiplicative inverse of the leading object’s position in the new order list. Let $\pi_{t}^{\prime}$ be the target ranking of the adversary and $\pi_{b}$ be the aggregation result of the victim with the manipulated data from $\mathcal{A}$. The $R – Rank$ is defined as

$$R – Rank(\pi_{t}^{\prime}, \pi_{b}) = \frac{1}{\pi_{b}^{-1}(\pi_{t}^{\prime}(1))}, \tag{87}$$

where $\pi(i)$ refers to the item index which lies in the $i$-th position of ranking list $\pi$, and $\pi^{-1}[i]$ indicates the position of ranking list $\pi$ which belongs to the item $i$. If it holds that $\pi_{b}(1) = \pi_{t}^{\prime}(1)$, we have $R – Rank(\pi_{t}^{\prime}, \pi_{b})$ archives its maximum value 1. The larger R-Rank value indicates a better purposeful attack result.

**Kendall $\tau$ Coefficient (Kendall-$\tau$).** The Kendall rank correlation coefficient evaluates the degree of similarity between two ranking lists given the same objects. This coefficient depends upon the number of inversions of pairs of objects which would be needed to transform one rank order into the other. Let $\mathbf{V} = [n]$ be a set of $n$ candidates and $\pi_{1}, \pi_{2}$ are two total orders or permutations on $\mathbf{V}$, the Kendall $\tau$ coefficient is defined to be

$$d_{\tau}(\pi_{1}, \pi_{2}) = \frac{2}{n(n-1)} \cdot \theta,$$ \tag{88}

where

$$\theta = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \theta(\pi_{1}^{-1}[i], \pi_{1}^{-1}[j], \pi_{2}^{-1}[i], \pi_{2}^{-1}[j])$$ \tag{89}

is the number of different pairs between these two ordered sets $\pi_{1}, \pi_{2}$ as

$$\theta(\pi_{1}^{-1}[i], \pi_{1}^{-1}[j], \pi_{2}^{-1}[i], \pi_{2}^{-1}[j]) = \begin{cases} 1, & \text{if } (\pi_{1}^{-1}[i] – \pi_{2}^{-1}[i])(\pi_{2}^{-1}[j] – \pi_{2}^{-1}[j]) > 0, \\
-1, & \text{if } (\pi_{1}^{-1}[i] – \pi_{2}^{-1}[i])(\pi_{2}^{-1}[j] – \pi_{2}^{-1}[j]) < 0, \\
0, & \text{otherwise} \end{cases}$$

Kendall $\tau$ coefficient counts the number of inconsistent comparisons between two rank orders. Larger Kendall-$\tau$ value indicates a better purposeful attack result. If $d_{\tau}(\pi_{1}, \pi_{2}) = 1$, we have $\pi_{1} = \pi_{2}$.

6.2 Simulated Study

We validate the proposed target attack strategies against Hodgerank and RankCentrality on simulated data. Here the pairwise comparisons are generated as follows. First we construct a complete graph $G = (\mathbf{V}, \mathbf{E})$ where $\mathbf{V} = [10]$. Then the latent preference score is assigned to each candidate/vertex of $\mathbf{V}$ and the true ranking is obtained by these scores. Next, we randomly sample the pairwise comparisons $\mathbf{w}'$ from $\mathbf{V} \times \mathbf{V}$ based on the true ranking. Notice that $\mathbf{w}'$ could contain the comparisons which are inconsistent with the true ranking. Even with the same true ranking, the pairwise comparisons could be different. It means that the same $t$ can lead to induce different $\pi_{b}(t)$ in (83). To ensure the statistical stability, the attack procedures are conducted 50 times for the same target $t$ with different comparisons.
Comparative Results. The reciprocal rank and Kendall-\(\tau\) coefficient between the perturbed ranking \(\pi_\theta\) and the target ranking \(\pi^*_t\) are illustrated in Fig. 1. Here the \(\pi_\theta\) is obtained by sorting the latent preference scores \(\hat{\theta}\) aggregated by HodgeRank or RankCentrality (Algorithm 1 or 2). The results indicate that all of our approaches exhibit higher success rate due to the parsimonious mechanism. Only the proposed methods stably lead \(R - \text{Rank}(\pi^*_t, \pi_\theta)\) to be 1 and get \(d_t(\pi^*_t, \pi_\theta)\) close to 1. However, the irreversible stochastic matrix attack method against RankCentrality would exhibit non-ideal performance with imperfect feedback as Fig. 1a and 1p. We speculate that this phenomenon comes from the complicated construction of the irreversible stochastic matrix as (81). If the adversary only predominates imperfect feedback, he/she would not reconstruct the empirical stochastic matrix through the first step in Algorithm 2. This phenomenon inspires us to develop the ranking aggregation algorithms that can resist target attacks in the future.

Visualization. We then visualize the attacked data distribution \(\hat{w}_i\) of the proposed method with different targets in Fig. 2. From the same original data \(w_i\), the proposed methods make efficient manipulation with specific purposes. We observe that the bins which represent the target winning the other candidates are higher after the attack procedure. Meanwhile, the number of the pairwise comparisons which indicate the target is defeated by the other candidates is decreased. Furthermore, we display these phenomenons with a “magnifier” in Fig. 3. Here the horizontal axis lists all possible pairwise comparisons. We index them as follows: \(\text{No.}(i - 1) \times 9 + 1\) to \(i \times 9\) are the comparisons \(\{(i, j) \mid j \in \{10, j \neq i\}\}\). The size of red dots stands for the number of pairwise comparisons after attack. The larger the radius of the red dot, the more the number of the perturbed pairwise comparisons. The vertical axis shows the changes in each comparison. The positive/negative vertical coordinate of red dot stands for the increment/decrement. The pairwise comparisons whose indices range from 73 to 81 are involved.
with the target candidate \( t = 9 \). We can see that the numbers of No. 77 – 81 comparisons are increased after the attack procedure; meanwhile, the number of No. 90 (10 > 9) is decreased by 40. Moreover, all comparisons which stand for the original top-1 candidates are defeated by the other candidates (No. 9, 18, ..., 72) which are also reduced by the proposed methods. The manipulation operation on the pairwise comparisons of the adversary naturally leads to the following questions: is such a modification of the data easy to detect? We answer this question by Fig. 4. We plot each pair of poisoned data by different methods on the left part (lower triangle part) and calculate the Pearson correlation on the right part (upper triangle part). The scatter plots in the lower triangle part show a pair of data whose source is labeled in the “gray” regions. The upper left points of each scatter plot belongs to the data labeled by the top “gray” regions; the lower right points come from the right gay box. In scatter plot, each point represents one type of pairwise comparison and these points are in ascending order of the weight. It is noteworthy that the horizontal axis of the scatter plot also shows the weight but not the index of pairwise comparisons. The Pearson correlations of every pair of data are calculated in the upper triangle part and the “***” stands for the \( p \)-value is smaller than 0.001. The diagonal blocks are the data distribution normalized by the total number of pairwise comparisons.

### 6.3 Human Age

**Description.** 30 images from human age dataset FGNET\(^2\) are annotated by a group of volunteer users on a crowdsourcing platform\(^3\). The ground-truth age ranking is known to us. The annotator is presented with two images and given a binary choice of which one is older. Totally, we obtain 8,017 pairwise comparisons from 94 annotators.

**Comparative Results.** It is worth mentioning that this real-world data has a high percentage of outliers (about 20% of all comparisons conflict with the correct age ranking). The proposed methods against HodgeRank and RankCentrality with reversible stochastic transition matrix still show promise attack effect as Table 2. An irreversible stochastic matrix would be failed except the complete information and perfect feedback scenario. These observations further inspires us that the spectral ranking with irreversible stochastic matrix has the potential to resist the target attack.

### 6.4 Dublin Election

**Description.** The Dublin election data set\(^4\) contains a complete record of votes for elections held in county Meath, Dublin, Ireland in 2002. This set contains 64,081 votes over 14 candidates. These votes could be a complete or partial list over the candidate set. The ground-truth ranking of 14 candidates is based on their obtained first preference votes\(^5\). The five candidates who receive the most first preference votes will be the winner of the election. We are interested in the top-5 performance of the pairwise rank aggregation method. Then these votes are converted into the pairwise comparisons. The total number of comparisons is 652,817.

**Comparative Results.** It is worth noting that the election result is not obtained by pairwise ranking aggregation. However, the ordered list aggregated from induced comparisons still shows positive correlations with the actual election result. Once \( A \) generates the successful strategy against the HodgeRank and RankCentrality, this attack plan could be adopted to manipulate the election in the real world. Different from the manipulation or strategic voting problem in social choice [38], the poisoning attack framework could break the barrier of [39] or the defense.
7 CONCLUSION

We establish the first study of target attacks in the context of ranking aggregation with pairwise comparisons to the best of our knowledge. The attack problem with specified purpose is formulated as the fixed point of the continuous operator. Moreover, a game-theoretic framework is proposed to model the interaction between the adversary and the victim. Two typical algorithms are selected as the victims. The attack strategy against HodgeRank is modeled as the least squared problems with a constraint which derived from the optimal condition of HodgeRank. We prove that the target ranking would be a fixed point in the complete game. When attacking against spectral ranking algorithm like RankCentrality, an inverse eigenvalue problem of the stochastic transition matrix is constructed. Moreover, we adopt minimizing the worst-case asymptotic variance as a constraint of the corresponding inverse eigenvalue problem. Our empirical studies show that the proposed attack methods could achieve the attacker’s goal in the sense that the leading candidate of the aggregated ranking list is the designated one by the adversary.

There are many avenues for further investigation – providing the finite-sample and asymptotic results, characterizing the theoretical performance of the adversarial game with the minimum knowledge and extending our attacks to more general ranking system like election and cryptocurrency. Vulnerability is an interesting question which can help to develop the reliable ranking algorithms in the future.

ACKNOWLEDGMENTS

The authors like to thank Angelina Cao for helpful discussion and carefully proof-reading this manuscript.

REFERENCES

[1] K. Arrow and E. Maskin, Social Choice and Individual Values: 3rd ed., London, U.K., Yale Univ. Press, 2012.
[2] X. Jiang, L.-H. Lim, Y. Yao, and Y. Ye, “Statistical ranking and combinatorial hodge theory,” Math. Program., vol. 127, no. 1, pp. 203–244, 2011.
[3] A. Korba, S. Clemenco, and E. Sibony, “A learning theory of ranking aggregation,” in Proc. Int. Conf. Artif. Intell. Statist., 2017, pp. 1001–1010.
[4] S. Negahban, S. Oh, and D. Shah, “Rank centrality: Ranking from pairwise comparisons,” Operation Res., vol. 65, no. 1, pp. 266–287, 2017.
[5] R. Herbrich, T. Minka, and T. Graepel, “Trueskill™: A Bayesian skill rating system,” in Proc. Adv. Neural Informat. Process. Syst., 2006, pp. 569–576.
[6] D. Cao, X. He, L. Miao, Y. An, C. Yang, and R. Hong, “Attentive group recommendation,” in Proc. ACM Int. Conf. Res. Develop. Informat. Retrieval, 2018, pp. 645–654.
[7] T. Liu, Learning to Rank for Information Retrieval, Berlin, Germany: Springer, 2011.
[8] D. E. Critchlow, M. A. Fligner, and J. S. Verducci, “Probability models on rankings,” J. Math. Psychol., vol. 35, no. 3, pp. 294–318, 1991.
[9] S. Kolde, S. Laur, P. Adler, and J. Vilo, “Robust rank aggregation for gene list integration and meta-analysis,” Bioinformatics, vol. 28, no. 4, pp. 573–580, 2012.
[10] N. B. Shah, S. Balakrishnan, J. K. Bradley, A. Parekh, K. Ramchandran, and M. J. Wainwright, “Estimation from pairwise comparisons: Sharp minimax bounds with topology dependence,” J. Mach. Learn. Res., vol. 17, no. 58, pp. 1–47, 2016.
[11] Y. Chen, J. Fan, C. Ma, and K. Wang, “Spectral method and regularized MLE are both optimal for top-k ranking,” Ann. Statist., vol. 47, no. 4, 2019, Art. no. 2204.
A. Agarwal, P. Patil, and S. Agarwal, “Accelerated spectral ranking,” in Proc. Int. Conf. Mach. Learn., 2018, pp. 70–79.

J. Wang, N. B. Shah, and R. Ravi, “Stretching the effectiveness of MLE from accuracy to bias for pairwise comparisons,” in Proc. Int. Conf. Artif. Intell. Statist., 2020, pp. 66–76.

B. Grofman and G. Owen, “Condorcet models, avenues for future research,” in Proc. Informat. Pooling Group Decis. Mak. Univ. Califor- nia, Irvine, Conf. Political Economy, 1986, pp. 93–102.

B. Biggio, B. Nelson, and P. Laskov, “Poisoning attacks against support vector machines,” in Proc. Int. Conf. Mach. Learn., 2012, pp. 1467–1474.

I. J. Goodfellow, J. Shlens, and C. Szegedy, “Explaining and harnessing adversarial examples,” in Proc. Int. Conf. Learn. Representa- tion., 2015, pp. 1–11.

R. A. Bradley and M. E. Terry, “Rank analysis of incomplete block designs: I. the method of paired comparisons,” Biometrika, vol. 39, no. 3, pp. 324–345, 1952.

R. Kannan and C. L. Monma, “On the computational complexity of integer programming problems,” in Optimization and Operations Research, Berlin, Germany: Springer, 1978, pp. 161–172.

F. Nash, “Equilibrium points in n-person games,” Proc. Nat. Acad. Sci. USA, vol. 36, no. 1, pp. 48–49, 1950.

J. Nash, “Non-cooperative games,” Ann. Math., vol. 54, no. 2, pp. 286–295, 1951.

S. Kakutani, “A generalization of Brouwer’s fixed point theorem,” Duke Math. J., vol. 8, no. 3, pp. 457–459, 1941.

I. L. Glicksberg, “A further generalization of the kakutani fixed point theorem with application to nash equilibrium points,” Proc. Amer. Math. Soc., vol. 3, no. 1, pp. 170–174, 1952.

R. B. Myerson, Game Theory: Analysis of Conflict. Cambridge, MA, USA: Harvard Univ. Press, 1997.

Z. Wen, D. Goldfarb, and W. Yin, “Alternating direction augmented lagrangian methods for semidefinite programming,” Math. Program., vol. 2, no. 3, pp. 203–230, 2010.

S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” Found. Trends Mach. Learn., vol. 3, no. 1, pp. 1–122, 2011.

M. Fazel, T. K. Pong, D. Sun, and P. Tseng, “Hankel matrix rank minimization with applications to system identification and realization,” SIAM J. Matrix Anal. Appl., vol. 34, no. 3, pp. 946–977, 2013.

L. Chen, D. Sun, and K.-C. Toh, “A note on the convergence of ADMM for linearly constrained convex optimization problems,” Comput. Optim. Appl., vol. 66, no. 2, pp. 327–343, 2017.

W. Deng and W. Yin, “On the global and linear convergence of the generalized alternating direction method of multipliers,” J. Sci. Comput., vol. 66, no. 3, pp. 889–916, 2016.

M. T. Chu, “Inverse eigenvalue problems,” SIAM Rev., vol. 40, no. 1, pp. 1–39, 1998.

W. Barnett et al., “The inverse eigenvalue problem of a graph: Multiplicities and minors,” J. Combinatorial Theory, Ser. B, vol. 142, pp. 276–306, 2020.

J. M. Flegal and G. L. Jones, “Batch means and spectral variance tests include object tracking, pattern recognition, computer vision. She has authored or coauthored 50+ academic papers in prestigious international journals and conferences (including IEEE Transactions on Pattern Analysis and Machine Intelligence, IJCV, IEEE Transactions on Image Processing, NeurIPS, ICML, CVPR, AAAI, etc). Moreover, she serves as an associate editor of IEEE Transactions on Circuits and Systems for Video Technology, ACM Transaction on Multimedia Computing, Communications, and Applications, and Multimedia Systems.

Jinshan Zeng received the PhD degree in applied mathematics from Xi’an Jiaotong University, Xi’an, China, in 2015. He is currently an associate professor with the School of Computer and Information Engineering, Jiangxi Normal University, Nanchang, China. His research interests include nonconvex optimization and machine learning.

Guorong Li received the BS degree in computer science from the Renmin University of China, in 2006, and the PhD degree in computer science from the Graduate University of the Chinese Academy of Sciences, in 2012. She is currently an associate professor with the University of Chinese Academy of Sciences. Her research interests include object tracking, pattern recognition, cross-media analysis, and multi-label learning.

Ke Ma (Member, IEEE) received the BS degree in mathematics from Tianjin University in 2009, the ME degree in software engineering from Beihang University (BUAA) in 2013, and the PhD degree in computer science from the Key Laboratory of Information Security (SKLOIS), Institute of Information Engineering (IIE), Chinese Academy of Sciences (CAS), in 2019. He is a postdoctoral research fellow with the School of Computer Science and Technology, University of Chinese Academy of Sciences (UCAS), Beijing, China. His research interests include rank aggregation and algorithmic game theory.

Qianxun Xu (Senior Member, IEEE) received the BS degree in computer science from the China University of Mining and Technology in 2007 and the PhD degree in computer science from the University of Chinese Academy of Sciences in 2013. She is currently an associate professor with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. Her research interests include statistical machine learning, with applications in multimedia and computer vision. She has authored or coauthored 50+ academic papers in prestigious international journals and conferences (including IEEE Transactions on Pattern Analysis and Machine Intelligence, IJCV, IEEE Transactions on Image Processing, NeurIPS, ICML, CVPR, AAAI, etc). Moreover, she serves as an associate editor of IEEE Transactions on Circuits and Systems for Video Technology, ACM Transaction on Multimedia Computing, Communications, and Applications, and Multimedia Systems.
Xiaochun Cao (Senior Member, IEEE) received the BE and ME degrees both in computer science from Beihang University (BUAA), China, and the PhD degree in computer science from the University of Central Florida, USA, with his dissertation nominated for the university level Outstanding Dissertation Award. He is a professor with the School of Cyber Science and Technology, Shenzhen Campus, Sun Yat-sen University. After graduation, he spent about three years with ObjectVideo Inc. as a Research Scientist. From 2008 to 2012, he was a professor with Tianjin University. Before joining SYSU, he was a professor with the Institute of Information Engineering, Chinese Academy of Sciences. He has authored and coauthored more than 200 journal and conference papers. In 2004 and 2010, he was the recipient of the Piero Zamperoni best student paper award at the International Conference on Pattern Recognition. He is on the editorial boards of IEEE Transactions on Image Processing and IEEE Transactions on Multimedia, and was on the editorial board of IEEE Transactions on Circuits and Systems for Video Technology.

Qingming Huang (Fellow, IEEE) received the bachelor’s and PhD degrees in computer engineering from the Harbin Institute of Technology, China, in 1988 and 1994, respectively. He is a chair professor with the University of Chinese Academy of Sciences and an adjunct research professor with the Institute of Computing Technology, Chinese Academy of Sciences. His research areas include multimedia computing, image processing, computer vision and pattern recognition. He has authored or coauthored more than 400 academic papers in prestigious international journals and top-level international conferences. He was the associate editor of IEEE Transaction on CSVT and Acta Automatica Sinica, and the reviewer of various international journals including IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE Transactions on Image Processing, IEEE Transaction on Multimedia, etc. He has served as general chair, program chair, track chair and TPC member for various conferences, including ACM Multimedia, CVPR, ICCV, ICME, ICMR, PCM, BigMM, PSIVT, etc.

For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/csdl.