Table 1: Comparison of this work to previous works in non-active pairwise ranking: extended-BTL model is due to (Cattelan 2012), chest-blade model is due to (Chen and Joachims 2016) and low-rank model is due to (Rajkumar and Agarwal 2016). Here, $\mathbf{w}$ and $\mathbf{W}$ are model parameters as given in Equation (1), and $\mathbf{F}$ denotes the feature matrix. Furthermore, *pairs* and *per* denote the state-of-the-art bounds known regarding the number of pairs compared and the number of comparisons per pair respectively. Note that we consider the practically important regime of $n \gg d$.

Model	$\mathbf{F}$	$\mathbf{w}$	$\mathbf{W}$	*pairs*	*per*
BTL	$\mathbf{F} = \mathbf{I}$	$\mathbf{w} \in \mathbb{R}^n$	$\mathbf{W} = \mathbf{0}$	$\Omega(n \log n)$	$\Omega(\log n)$
Item-feature	$\mathbf{F} \in \mathbb{R}^{d \times n}$	$\mathbf{w} \in \mathbb{R}^d$	$\mathbf{W} = \mathbf{0}$	many	many
Chest-blade	$\mathbf{F} = \mathbf{I}$	$\mathbf{w} \in \mathbb{R}^n$	rank$(\mathbf{W}) = O(d)$	many	many
Low-rank	$\mathbf{F} = \mathbf{I}$	$\mathbf{w} \in \mathbb{R}^n$	$\mathbf{W} \in \mathbb{R}^{n \times n}$	$\Omega(nr \log n)$	$\Omega(r \log n)$
This work	$\mathbf{F} \in \mathbb{R}^{d \times n}$	$\mathbf{w} \in \mathbb{R}^d$	$\mathbf{W} \in \mathbb{R}^{d \times d}$	$\Omega(d^2 \log n)$	$\Omega(d^2 \log^2 n / n^2)$

Proposition 2. *The unary RU models are special cases of the FLR model, ie,* $\mathcal{P}_n^{RU} \subseteq \mathcal{P}_n(\psi, r, A)$.

Corollary 2. *The BTL and Thurstone models are obtained as special cases of the FLR model under the logit and the probit transformations of* $\mathbf{P}$ *respectively. This follows from Proposition 1 (or Corollary 1-part (1)) above together with Propositions 6 and 7 of (Rajkumar and Agarwal 2016).*

Proposition 3. *Regression-based models with item-specific features in (Cattelan 2012) are special cases of the FLR model, ie,* $\mathcal{P}_n^{IF} \subseteq \mathcal{P}_n(\psi, r, \mathbf{A})$.

Corollary 3. *Let* $d \ll n$. *Then we recover the blade-chest model (Chen and Joachims 2016) as a special case of the FLR model by setting rank*$(\mathbf{W}) = O(d)$ *and* $\mathbf{w} = \mathbf{0}$. *Next, when* $d \geq n$, *it is clear from Theorem 1 of (Chen and Joachims 2016) that such preference matrices degenerate into matrices in* $\mathcal{P}_n(\psi, n, \mathbf{A})$ *where* ψ *is the logit function. Moreover, it is easy to see that the FLR model admits both stochastic-transitive and stochastic-intransitive preference matrices.*

Due to space constraints, proofs of Propositions 1, 2 and 3 are given in the appendix. To summarize, we have shown how to instantiate several previously proposed ranking models as special cases of our FLR model in Table 1.

Problem Setup and Solution Approach

Once we have the generative ranking model as developed in the previous section, the objective in our learning problem is then to find the permutation of n items that minimizes the number of violations with respect to the true underlying preference matrix $\mathbf{P}$, ie, to find the best ranking $\widehat{\sigma}$ in the sense that,

$$\widehat{\sigma} = \arg\min_{\sigma} \operatorname{dist}(\sigma, \mathbf{P})$$

The input is the pairwise comparison dataset $S = \{(i, j, y_{ij}^k)\}$ which consists of comparison results of pairs (i, j) from a survey involving K users where each user with index k assigns $y_{ij}^k = 1$ if he prefers i to j and $y_{ij}^k = 0$ if he prefers j to i. Note that it is not necessary that all pairs of items be compared; our algorithm is able to handle noisy and incomplete data. Since the true preference matrix $\mathbf{P}$ is unknown, our algorithm instead proceeds by using the empirical preference matrix $\widehat{\mathbf{P}}$ computed from the available y_{ij}^k; it is to be noted, even then, our analysis guarantees that $\operatorname{dist}(\widehat{\sigma}, \mathbf{P})$ is good as opposed to just $\operatorname{dist}(\widehat{\sigma}, \widehat{\mathbf{P}})$. Additionally, in our inductive setting, the feature information is encoded by $\mathbf{f}_i \in \mathbb{R}^d$ for every item i and concatenated to form the feature matrix $\mathbf{F} \in \mathbb{R}^{d \times n}$.

Algorithm

We present our main algorithm for inductive ranking in Algorithm 3. The input data consist of the set of pairwise comparison results $S = \{(i, j, \{y_{ij}^k\})\}$, $(i, j) \in \Xi \subseteq [n] \times [n]$, $k \in [K]$, $y_{ij}^k \in \{0, 1\}$ and the feature matrix $\mathbf{F} \in \mathbb{R}^{d \times n}$. The algorithm assumes the link function and the rank as input parameters. The subroutines used are:

1. *Noisy matrix completion with features (Subroutine 1):* Note that to solve our ranking problem and derive the associated recovery guarantee, it suffices, as we have done, to use the specified trace-norm program as a black-box method; hence, we assume that we have access to an oracle that gives us the solution to the convex program. The details of how the solution to this program may be found numerically is beyond the scope of this work – for further details regarding some possible sub-gradient algorithms, we refer the reader to (Chiang, Hsieh, and Dhillon 2015) and (Ji and Ye 2009).
2. *γ-approximate pairwise ranking procedure (Subroutine 2):* Let $\widehat{\sigma} \in \mathcal{S}_n$ be the output of any Pairwise Ranking (PR) procedure with respect to an underlying preference matrix $\mathbf{P}$. For a constant $\gamma > 1$, $\widehat{\sigma}$ is said to be γ-approximate if $\operatorname{dist}(\widehat{\sigma}, \mathbf{P}) \leq \gamma \min_{\sigma \in \mathcal{S}_n} \operatorname{dist}(\sigma, \mathbf{P})$. Any constant factor approximate ranking procedure maybe used. Specifically, we use the Copeland procedure (Copeland 1951) as a black-box method which has a 5-approximation guarantee (Coppersmith, Fleischer, and Rudra 2006). This method involves simply sorting the items according to a score which is computed for every item i as $\sum_{j=1}^{n} \mathbb{1}(\overline{\widehat{P}}_{ij} > 1/2)$.

Analysis

In this section, we state and prove our main result.

Theorem 1 **(Guaranteed rank aggregation with sub-linear sample complexity using item features).** *Let* $\mathbf{P} \in (\mathcal{P}_n(\psi, r, \mathbf{A}) \cap \mathcal{P}_n^{ST})$ *be the true underlying preference matrix according to which the pairwise comparison dataset* $S = \{(i, j, \{y_{ij}^k\})\}$ *is generated. Let* ψ *be L-Lipschitz*