SUPPLEMENTARY MATERIAL

While our analysis of the proposed approach and algorithm largely tracks the methodology in [Ding et al., 2014], here we develop a set of new analysis tools that can handle more general settings. Specifically, our new analysis tools can handle **any** isotropically distributed random projection directions. In contrast, the work in [e.g., Ding et al., 2014] can only handle special types of random projections, e.g., spherical Gaussian. Our new refined analysis can not only handle more general settings, it also gives an overall improved sample complexity bound.

We also analyse the post-processing step in Algorithm 4. This step accounts for the special constraints that a valid ranking representations must satisfy and guarantees a binary-valued estimate of $\boldsymbol{\sigma}$. It should also satisfy the property that either $\sigma^k(i) > \sigma^k(j)$ or $\sigma^k(i) < \sigma^k(j)$ for all distinct i, j and all k.

We note that the analysis framework that we present here for the solid angle can in fact be extended to handle other types distributions for the random projection directions. This is, however, beyond the scope this paper.

A On the generative model

Proposition 1. $\mathbf{B} = \mathbf{P}\boldsymbol{\sigma}$ *is column stochastic.*

Proof. Noting that $\sigma_{(i,j),k} + \sigma_{(j,i),k} = 1$ by definition, and $P_{(i,j),(i,j)} = P_{(j,i),(j,i)} = \mu_{i,j}$, therefore,

$$\begin{aligned}\sum_{(i,j)} B_{(i,j),k} &= \sum_{(i,j)\,:\,i<j} (\sigma_{(i,j),k} + \sigma_{(j,i),k})\mu_{i,j} \\ &= \sum_{(i,j)\,:\,i<j} \mu_{i,j} = 1\end{aligned}$$

□

B Connection to the model in FJS

Here we discuss in detail the connection to the probability model as well as the algorithm proposed in Jagabathula and Shah [2008]Farias et al. [2009] (denoted by FJS).

First, the generative model proposed in FJS can be viewed as a special case of our generative model. If we consider the prior distribution of $\boldsymbol{\theta}_m$ to be a pmf on the vertices of the K-dimensional probability simplex (so that $\boldsymbol{\theta}_m$ has only one nonzero component with probability one), i.e.,

$$\Pr(\boldsymbol{\theta}_m = \mathbf{e}_k) = b_k \tag{1}$$

where $\mathbf{e}_k$ is the k-th standard basis vector and $\sum_{k=1}^K b_k = 1$, then each user m is associated with only one of the K types with probability b_k for the k-th type. We note that under this prior, $\mathbf{a} \triangleq \mathbb{E}(\boldsymbol{\theta}_m) = \mathbf{b}$ and $\mathbf{R} \triangleq \mathbb{E}(\boldsymbol{\theta}_m\boldsymbol{\theta}_m^\top) = \mathrm{diag}(\mathbf{b})$ has full rank.

Second, the algorithm proposed in FJS can certainly be applied to our more general setting. Since the algorithm FJS only uses the first order statistic which corresponds to pooling the comparisons from all the users together, it suffices to consider only the probabilities of $p(w_1 = (i,j))$ by marginalizing over $\boldsymbol{\theta}$:

$$\begin{aligned}p(w_1 = (i,j)) &= \int_{\boldsymbol{\theta}_m} p(w_1 = (i,j)|\boldsymbol{\theta}_m)\Pr(\boldsymbol{\theta}_m)d\boldsymbol{\theta_m} \\ &= \sum_{k=1}^K \sigma_{(i,j),k}\int_{\boldsymbol{\theta}_m} \theta_{k,m}d\boldsymbol{\theta_m} \\ &= \sum_{k=1}^K \sigma_{(i,j),k}a_k \\ &= \sum_{k\ :\sigma^k(i)<\sigma^k(j)} a_k,\end{aligned}$$

where the last step is due to the definition of the ranking matrix $\boldsymbol{\sigma}$. The above derivation shows that if the expectation vector in our generative model equals that in the model of FJS, then the probability distribution of the first order statistic in both models will be identical and the two models will be indistinguishable in terms of the first order statistic. This shows that the comparison with FJS in the experiments conducted in Sections 6.1 and 6.2 of the main paper is both sensible and fair.

Indexing convention: For convenience, for the rest of this appendix we will index the $W = Q(Q-1)$ rows of $\mathbf{B}$ and $\mathbf{E}$ by just a single index i instead of an ordered pair (i,j) as in the main paper.

C Proof of Lemma 2 in the main paper

Lemma 2 in the main paper is a result about the almost sure convergence of the estimate of the normalized second order moments $\mathbf{E}$. Our proof of this result will also provide an attainable rate of convergence.

We first provide a generic method to establish the convergence rate for a function $\psi(\mathbf{X})$ of d random variables $X_1, \ldots, X_d$ given their individual convergence rates.

Proposition 2. *Let* $\mathbf{X} = [X_1, \ldots, X_d]$ *be* d *random variables and* $\mathbf{a} = [a_1, \ldots, a_d]$ *be positive constants. Let* $\mathcal{E} := \bigcup_{i\in\mathcal{I}} \{|X_i - a_i| \geq \delta_i\}$ *for some constants* $\delta_i > 0$, *and $\psi(\mathbf{X})$ be a continuously differentiable function in $\mathcal{C} := \mathcal{E}^c$. If for $i = 1, \ldots, d$, $\Pr(|X_i - a_i| \geq \epsilon) \leq f_i(\epsilon)$ and $\max_{X \in \mathcal{C}} |\partial_i \psi(\mathbf{X})| \leq C_i$, then,*

$$\Pr(|\psi(\mathbf{X}) - \psi(\mathbf{a})| \geq \epsilon) \leq \sum_i f_i(\gamma) + \sum_{i=1} f_i(\frac{\epsilon}{dC_i})$$

Proof. Since $\psi(\mathbf{X})$ is continuously differentiable in $\mathcal{C}$, $\forall \mathbf{X} \in \mathbf{C}, \exists \lambda \in (0,1)$ such that

$$\psi(\mathbf{X}) - \psi(\mathbf{a}) = \nabla^\top \psi((1-\lambda)\mathbf{a} + \lambda \mathbf{X}) \cdot (\mathbf{X} - \mathbf{a})$$

Therefore,

$$\begin{aligned}
&\Pr(|\psi(\mathbf{X}) - \psi(\mathbf{a})| \geq \epsilon) \\
\leq &\Pr(\mathbf{X} \in \mathcal{E}) \\
&+ \Pr(\sum_{i=1}^{d} |\partial_i \psi((1-\lambda)\mathbf{a} + \lambda \mathbf{X})||X_i - a_i| \geq \epsilon | \mathbf{X} \in \mathcal{C}) \\
\leq &\sum_{i \in \mathcal{I}} \Pr(|X_i - a_i| \geq \delta_i) \\
&+ \sum_{i=1}^{d} \Pr(\max_{\mathbf{x} \in \mathcal{C}} |\partial_i \psi(\mathbf{x})||X_i - a_i| \geq \epsilon/d) \\
= &\sum_{i \in \mathcal{I}} f_i(\delta_i) + \sum_{i=1} f_i(\frac{\epsilon}{dC_i})
\end{aligned}$$

□

Now we are ready to prove Lemma 2 of the main paper. Recall that $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{X}}'$ are obtained from $\mathbf{X}$ by first splitting each user's comparisons into two independent copies and then re-scaling the rows to make them row-stochastic. Therefore, $\widetilde{\mathbf{X}} = \operatorname{diag}^{-1}(\mathbf{X1})\mathbf{X}$. Since $\bar{\mathbf{B}} = \operatorname{diag}^{-1}(\mathbf{Ba})\mathbf{B}\operatorname{diag}(\mathbf{a})$, $\bar{\mathbf{R}} = \operatorname{diag}^{-1}(\mathbf{a})\mathbf{R}\operatorname{diag}^{-1}(\mathbf{a})$, and $\bar{\mathbf{B}}$ is row stochastic. From Lemma 2 of the main paper, we have

Lemma 1. *Let $\widehat{\mathbf{E}} = M\widetilde{\mathbf{X}'}\widetilde{\mathbf{X}}^\top$ and $\mathbf{E} = \bar{\mathbf{B}}\bar{\mathbf{R}}\bar{\mathbf{B}}^\top$. If $\eta = \min_{1 \leq i \leq W}(\mathbf{Ba})_i > 0$, then,*

$$\Pr(\|\widehat{\mathbf{E}} - \mathbf{E}\|_\infty \geq \epsilon) \leq 8W^2 \exp(-\epsilon^2 \eta^4 MN/20) \quad (2)$$

Proof. For any $1 \leq i, j \leq W$,

$$\begin{aligned}
\widehat{E}_{i,j} &= M \frac{1}{\sum_{m=1}^{M} X'_{i,m}} (\sum_{m=1}^{M} X'_{i,m} X_{j,m}) \frac{1}{\sum_{m=1}^{M} X_{i,m}} \\
&= \frac{1/M \sum_{m=1}^{M} (X'_{i,m} X_{j,m})}{(1/M \sum_{m=1}^{M} X'_{i,m})(1/M \sum_{m=1}^{M} X_{j,m})} \\
&= \frac{\frac{1}{MN^2} \sum_{m=1,n=1,n'=1}^{M,N,N} \mathbb{I}(w_{m,n} = i)\mathbb{I}(w'_{m,n'} = j)}{\frac{1}{MN} \sum_{m=1,n=1}^{M,N} \mathbb{I}(w_{m,n} = i) \frac{1}{MN} \sum_{m=1,n=1}^{M,N} \mathbb{I}(w'_{m,n} = i)} \\
&:= \frac{F_{i,j}(M,N)}{G_i(M,N) H_j(M,N)}
\end{aligned}$$

From the Strong Law of Large Numbers and equations (1), (2) in the main paper, we have

$$\begin{aligned}
F_{i,j}(M,N) &\xrightarrow{a.s.} \mathbb{E}(\mathbb{I}(w_{m,n} = i)\mathbb{I}(w'_{m,n'} = j)) \\
&= (\mathbf{BRB}^\top)_{i,j} := p_{i,j}
\end{aligned}$$

$$\begin{aligned}
G_i(M,N) &\xrightarrow{a.s.} \mathbb{E}(\mathbb{I}(w'_{m,n} = i)) = (\mathbf{Ba})_i := p_i \\
H_i(M,N) &\xrightarrow{a.s.} \mathbb{E}(\mathbb{I}(w_{m,n} = j)) = (\mathbf{Ba})_j := p_j
\end{aligned}$$

and $\frac{(\mathbf{BRB}^\top)_{i,j}}{(\mathbf{Ba})_i (\mathbf{Ba})_j} = \mathbf{E}_{i,j}$ by definition. Using McDiarmid's inequality, we obtain

$$\begin{aligned}
\Pr(|F_{i,j} - p_{i,j}| \geq \epsilon) &\leq 2\exp(-\epsilon^2 MN) \\
\Pr(|G_i - p_i| \geq \epsilon) &\leq 2\exp(-2\epsilon^2 MN) \\
\Pr(|H_j - p_j| \geq \epsilon) &\leq 2\exp(-2\epsilon^2 MN)
\end{aligned}$$

In order to calculate $\Pr\{|\frac{F_{i,j}}{G_i H_j} - \frac{p_{i,j}}{p_i p_j}| \geq \epsilon\}$, we apply the results from Proposition 2. Let $\psi(x_1, x_2, x_3) = \frac{x_1}{x_2 x_3}$ with $x_1, x_2, x_3 > 0$, and $a_1 = p_{i,j}$, $a_2 = p_i$, $a_3 = p_j$. Let $\mathcal{I} = \{2, 3\}$, $\delta_2 = \gamma p_i$, and $\delta_3 = \gamma p_j$. Then $|\partial_1 \psi| = \frac{1}{x_2 x_3}$, $|\partial_2 \psi| = \frac{x_1}{x_2^2 x_3}$, and $|\partial_3 \psi| = \frac{x_1}{x_2 x_3^2}$.

If $F_{i,j} = x_1$, $G_i = x_2$, and $H_j = x_3$, then $F_{i,j} \leq G_i$, $F_{i,j} \leq H_j$. Then note that

$$\begin{aligned}
C_1 &= \max_{\mathcal{C}} |\partial_1 \psi| = \max_{\mathcal{C}} \frac{1}{G_i H_j} \leq \frac{1}{(1-\gamma)^2 p_i p_j} \\
C_2 &= \max_{\mathcal{C}} |\partial_2 \psi| = \max_{\mathcal{C}} \frac{F_{i,j}}{G_i^2 H_j} \leq \max_{\mathcal{C}} \frac{1}{G_i H_j} \leq \frac{1}{(1-\gamma)^2 p_i p_j} \\
C_3 &= \max_{\mathcal{C}} |\partial_3 \psi| = \max_{\mathcal{C}} \frac{F_{i,j}}{G_i H_j^2} \leq \max_{\mathcal{C}} \frac{1}{G_i H_j} \leq \frac{1}{(1-\gamma)^2 p_i p_j}
\end{aligned}$$

By applying Proposition 2, we get

$$\begin{aligned}
&\Pr\{|\frac{F_{i,j}}{G_i H_j} - \frac{p_{i,j}}{p_i p_j}| \geq \epsilon\} \\
\leq &\exp(-2\gamma^2 p_i^2 MN) + \exp(-2\gamma^2 p_j^2 MN) \\
&+ 2\exp(-\epsilon^2(1-\gamma)^4(p_i p_j)^2 MN/9) \\
&+ 4\exp(-2\epsilon^2(1-\gamma)^4(p_i p_j)^2 MN/9) \\
\leq &2\exp(-2\gamma^2\eta^2 MN) + 6\exp(-\epsilon^2(1-\gamma)^4\eta^4 MN/9)
\end{aligned}$$

where $\eta = \min_{1\leq i\leq W} p_i$. There are many strategies for optimizing the free parameter γ. We set $2\gamma^2 = \frac{(1-\gamma)^4}{9}$ and solve for γ to obtain

$$\Pr\{|\frac{F_{i,j}}{G_i H_j} - \frac{p_{i,j}}{p_i p_j}| \geq \epsilon\} \leq 8\exp(-\epsilon^2\eta^4 MN/20)$$

Finally, by applying the union bound to the W^2 entries in $\widehat{\mathbf{E}}$, we obtain the claimed result. □

D Proof of Theorem 2 in the main paper

D.1 Outline

We focus on the case when the random projection directions are sampled from **any** *isotropic distribution*. Our proof is not tied to the special form of the distribution; just its isotropic nature. In contrast, the method in [e.g., Ding et al., 2014] can only handle special types of distributions such as the spherical Gaussian.

The proof of Theorem 2 in the main paper can be decoupled into two steps. First, we show that Algorithm 2 in the main paper can consistently identify all the novel words of the K distinct rankings. Then, given the success of the first step, we will show that Algorithm 3 proposed in the main paper can consistently estimate the ranking matrix $\boldsymbol{\sigma}$.

D.2 Useful propositions

We denote by $\mathcal{C}_k$ the set of all novel pairs of the ranking σ^k, for $k = 1, \ldots, K$, and denote by $\mathcal{C}_0$ the set of other non-novel pairs. We first prove the following result.

Proposition 3. *Let* $\mathbf{E}_i$ *be the i-th row of* $\mathbf{E}$. *Suppose* $\boldsymbol{\sigma}$ *is separable and* $\mathbf{R}$ *has full rank, then the following is true:*

	$\|\mathbf{E}_i - \mathbf{E}_j\|$	$E_{i,i} - 2E_{i,j} + E_{j,j}$
$i \in \mathcal{C}_1, j \in \mathcal{C}_1$	0	0
$i \in \mathcal{C}_1, j \notin \mathcal{C}_1$	$\geq (1-b)\lambda_{\min}$	$\geq (1-b)^2\lambda_{\min}^2/\lambda_{\max}$

where $b = \max_{j\in\mathcal{C}_0,k} \bar{B}_{j,k}$ *and* $\lambda_{\min}$, $\lambda_{\max}$ *are the minimum /maximum eigenvalues of* $\bar{\mathbf{R}}$

Proof. Let $\bar{\mathbf{B}}_i$ be the i-th row vector of matrix $\bar{\mathbf{B}}$. To show the above results, recall that $\mathbf{E} = \bar{\mathbf{B}}\mathbf{R}'\bar{\mathbf{B}}^\top$. Then

$$\begin{aligned}
&\|\mathbf{E}_i - \mathbf{E}_j\| = \|(\bar{\mathbf{B}}_i - \bar{\mathbf{B}}_j)\mathbf{R}'\bar{\mathbf{B}}^\top\| \\
&E_{i,i} - 2E_{i,j} + E_{j,j} = (\bar{\mathbf{B}}_i - \bar{\mathbf{B}}_j)\mathbf{R}'(\bar{\mathbf{B}}_i - \bar{\mathbf{B}}_j)^\top.
\end{aligned}$$

It is clear that when $i, j \in \mathcal{C}_1$, i.e., they are both novel pairs for the same ranking, $\bar{\mathbf{B}}_i = \bar{\mathbf{B}}_j$. Hence, $\|\mathbf{E}_i - \mathbf{E}_j\| = 0$ and $E_{i,i} - 2E_{i,j} + E_{j,j} = 0$.

When $i \in \mathcal{C}_1, j \notin \mathcal{C}_1$, we have $\bar{\mathbf{B}}_i = [1, 0, \ldots, 0]$, $\bar{\mathbf{B}}_j = [\bar{B}_{j,i}, \bar{B}_{j,2}, \ldots, \bar{B}_{j,K}]$ with $\bar{B}_{j,1} < 1$. Then,

$$\begin{aligned}
\bar{\mathbf{B}}_i - \bar{\mathbf{B}}_j &= [1 - \bar{B}_{j,i}, -\bar{B}_{j,2}, \ldots, -\bar{B}_{j,K}] \\
&= (1 - \bar{B}_{j,i})[1, -c_2, \ldots, -c_K] \\
&:= (1 - \bar{B}_{j,i})\mathbf{e}^\top
\end{aligned}$$

and $\sum_{l=2}^K c_l = 1$. Therefore, defining $\mathbf{Y} := \mathbf{R}'\bar{\mathbf{B}}^\top$, we get

$$\|\mathbf{E}_i - \mathbf{E}_j\|_2 = (1 - \bar{B}_{j,i})\|\mathbf{Y}_1 - \sum_{l=2}^K c_l \mathbf{Y}_l\|_2$$

Using the Proposition 1 in [Ding et al., 2013], if $\bar{\mathbf{R}}$ is full rank with minimum eigenvalue $\lambda_{\min} > 0$, then, $\bar{\mathbf{R}}$ is γ-(row)simplicial with $\gamma = \lambda_{\min}$, i.e., any row vector is at least γ distant from any convex combination of the remaining rows. Since $\bar{\mathbf{B}}$ is separable, $\mathbf{Y}$ is at least γ-simplicial (see Ding et al. [2014] Lemma 1). Therefore,

$$\|\mathbf{E}_i - \mathbf{E}_j\|_2 \geq (1 - \bar{B}_{j,1})\gamma \geq (1-b)\lambda_{\min}$$

where $b = \max_{j\in\mathcal{C}_0,k} \bar{B}_{j,k} < 1$.

Similarly, note that $\|\mathbf{e}^\top\bar{\mathbf{R}}\| \geq \gamma$ and let $\bar{\mathbf{R}} = \mathbf{U}\Sigma\mathbf{U}^\top$ be its singular value decomposition. If $\lambda_{\max}$ is the maximum eigenvalue of $\bar{\mathbf{R}}$, then we have

$$\begin{aligned}
E_{i,i} - 2E_{i,j} + E_{j,j} &= (1 - \bar{B}_{j,1})^2\mathbf{e}^\top\bar{\mathbf{R}}\mathbf{e} \\
&= (1 - \bar{B}_{j,1})^2(\mathbf{e}^\top\bar{\mathbf{R}})\mathbf{U}\Sigma^{-1}\mathbf{U}^\top(\mathbf{e}^\top\bar{\mathbf{R}})^\top \\
&\geq (1-b)^2\lambda_{\min}^2/\lambda_{\max}.
\end{aligned}$$

The inequality in the last step follows from the observation that $\mathbf{e}^\top\mathbf{R}'$ is within the column space spanned by $\mathbf{U}$. □

The results in Proposition 3 provide two statistics for identifying novel pairs of the same topic, $\|\mathbf{E}_i - \mathbf{E}_j\|$ and $E_{i,i} - 2E_{i,j} + E_{j,j}$. While the first is straightforward, the latter is efficient to calculate in practice with better computational complexity. Specifically, the set $\mathcal{J}_i$ in Algorithm 2 of the main paper

$$\mathcal{J}_i = \{j : \widehat{E}_{i,i} - \widehat{E}_{i,j} - \widehat{E}_{j,i} + \widehat{E}_{j,j} \geq d/2\}$$

can be used to discover the set of novel pairs of the same rankings asymptotically. Formally,

Proposition 4. *If $\|\widehat{\mathbf{E}}-\mathbf{E}\|_\infty \le d/8$, then,*

1. *For a novel pair $i \in \mathcal{C}_k$, $\mathcal{J}_i = \mathbf{C}_k^c$*
2. *For a non-novel pair $j \in \mathcal{C}_0$, $\mathcal{J}_i \supset \mathbf{C}_k^c$*

D.3 Consistency of Algorithm 2 in the main paper

Now we start to show that Algorithm 2 of the main paper can detect all the novel pairs of the K distinct rankings consistently. As a starting point, it is straightforward to show the following result.

Proposition 5. *Suppose $\boldsymbol{\sigma}$ is separable and $\mathbf{R}$ is full rank, then, $q_i > 0$ if and only if i is a novel pair.*

We denote the minimum solid angle of the K extreme points by $q_\wedge$. Proposition 5 shows that the novel pairs can be identified by simply sorting q_i.

The agenda is to show that the estimated solid angle in Alg. 2,

$$\hat{p}_i = \frac{1}{P}\sum_{r=1}^{P} \mathbb{I}\{\forall j \in \mathcal{J}_i,\ \widehat{\mathbf{E}}_j\mathbf{d}_r \le \widehat{\mathbf{E}}_i\mathbf{d}_r\} \qquad (3)$$

converges to the ideal solid angle

$$q_i = \Pr\{\forall j \in \mathcal{S}_i, (\mathbf{E}_i - \mathbf{E}_j)\mathbf{d} \ge 0\} \qquad (4)$$

hence the error event in Alg. 2 has vanishing probability as $M, P \to \infty$. $\mathbf{d}_1, \ldots, \mathbf{d}_P$ are iid directions drawn from a isotropic distribution. For a novel pair $i \in \mathcal{C}_k, k = 1, \ldots, K$, $\mathcal{S}_i = \mathcal{C}_k^c$, and for a non-novel pair $i \in \mathcal{C}_0$, let $\mathcal{S}_i = \mathcal{C}_0^c$.

To show the convergence of $\hat{p}_i$ to p_i, we consider an intermediate quantity,

$$p_i(\widehat{\mathbf{E}}) = \Pr\{\forall j \in \mathcal{J}_i, (\widehat{\mathbf{E}}_i - \widehat{\mathbf{E}}_j)\mathbf{d} \ge 0\}$$

First, by Hoeffding's lemma, we have the following result.

Proposition 6. $\forall t \ge 0, \forall i,$

$$\Pr\{|\hat{p}_i - p_i(\widehat{\mathbf{E}})|t\} \ge 2\exp(-2Pt^2) \qquad (5)$$

Next we show the convergence of $p_i(\widehat{\mathbf{E}})$ to solid angle q_i:

Proposition 7. *Consider the case when $\|\widehat{\mathbf{E}}-\mathbf{E}\|_\infty \le \frac{d}{8}$. If i is a novel pair, then,*

$$q_i - p_i(\widehat{\mathbf{E}}) \le \frac{W\sqrt{W}}{\pi d_2}\|\widehat{\mathbf{E}}-\mathbf{E}\|_\infty$$

Similarly, if j is a non-novel pair, we have,

$$p_j(\widehat{\mathbf{E}}) - q_i \le \frac{W\sqrt{W}}{\pi d_2}\|\widehat{\mathbf{E}}-\mathbf{E}\|_\infty$$

where $d_2 \triangleq (1-b)\lambda_{\min}$, $d = (1-b)^2\lambda_{\min}^2/\lambda_{\max}$.

Proof. First note that, by the definition of $\mathcal{J}_i$ and Proposition 3, if $\|\widehat{\mathbf{E}}-\mathbf{E}\|_\infty \le \frac{d}{8}$, then, for a novel pair $i \in \mathcal{C}_k$, $\mathcal{J}_i = \mathcal{S}(i)$. And for a non-novel pair $i \in \mathcal{C}_0$, $\mathcal{J}_i \supseteq \mathcal{S}(i)$. For convenience, let

$$A_j = \{\mathbf{d} : (\widehat{\mathbf{E}}_i - \widehat{\mathbf{E}}_j)\mathbf{d} \ge 0\} \quad A = \bigcap_{j\in\mathcal{J}_i} A_j$$

$$B_j = \{\mathbf{d} : (\mathbf{E}_i - \mathbf{E}_j)\mathbf{d} \ge 0\} \quad B = \bigcap_{j\in\mathcal{S}(i)} B_j$$

For i being a novel pair, we consider

$$q_i - p_i(\widehat{\mathbf{E}}) = \Pr\{B\} - \Pr\{A\} \le \Pr\{B \bigcap A^c\}$$

Note that $\mathcal{J}_i = \mathcal{S}(i)$ when $\|\widehat{\mathbf{E}}-\mathbf{E}\| \le d/8$,

$$\begin{aligned}
&\Pr\{B \bigcap A^c\} = \Pr\{B \bigcap (\bigcup_{j\in\mathcal{S}(i)} A_j^c)\} \\
&\le \sum_{j\in\mathcal{S}(i)} \Pr\{(\bigcap_{l\in\mathcal{S}(i)} B_l) \bigcap A_j^c\} \le \sum_{j\in\mathcal{S}(i)} \Pr\{B_j \bigcap A_j^c\} \\
&= \sum_{j\in\mathcal{S}(i)} \Pr\{(\widehat{\mathbf{E}}_i - \widehat{\mathbf{E}}_j)\mathbf{d} < 0, \text{and } (\mathbf{E}_i - \mathbf{E}_j)\mathbf{d} \ge 0\} \\
&= \sum_{j\in\mathcal{S}(i)} \frac{\phi_j}{2\pi}
\end{aligned}$$

where ϕ_j is the angle between $\mathbf{e}_j = \mathbf{E}_i - \mathbf{E}_j$ and $\widehat{\mathbf{e}}_j = \widehat{\mathbf{E}}_i - \widehat{\mathbf{E}}_j$ for any isotropic distribution on $\mathbf{d}$. Using the trigonometric inequality $\phi \le \tan(\phi)$,

$$\begin{aligned}
\Pr\{B \bigcap A^c\} &\le \sum_{j\in\mathcal{S}(i)} \frac{\tan(\phi_j)}{2\pi} \le \sum_{j\in\mathcal{S}(i)} \frac{1}{2\pi}\frac{\|\widehat{\mathbf{e}}_j - \mathbf{e}_j\|_2}{\|\mathbf{e}_j\|_2} \\
&\le \frac{W\sqrt{W}}{\pi d_2}\|\widehat{\mathbf{E}}-\mathbf{E}\|_\infty
\end{aligned}$$

where the last inequality is obtained by the relationship between the ℓ_∞ norm and theℓ_2 norm, and the fact that for $j \in \mathcal{S}(i)$, $\|\mathbf{e}_j\|_2 = \|\mathbf{E}_i - \mathbf{E}_j\|_2 \ge d_2 \triangleq (1-b)\lambda_{\min}$. Therefore for a novel word i, we have,

$$q_i - p_i(\widehat{\mathbf{E}}) \le \frac{W\sqrt{W}}{\pi d_2}\|\widehat{\mathbf{E}}-\mathbf{E}\|_\infty$$

Now for a non-novel word i, note the fact that $i \in \mathcal{C}_0$, $\mathcal{J}_i \supseteq \mathcal{S}(i)$,

$$\begin{aligned}
p_i(\widehat{\mathbf{E}}) - q_i &= \Pr\{A\} - \Pr\{B\} = \Pr\{A \bigcap B^c\} \\
&\le \sum_{j\in\mathcal{S}(i)} \Pr\{(\bigcap_{l\in\widehat{\mathcal{S}}(i)} A_l) \bigcap B_j^c\} \\
&\le \sum_{j\in\mathcal{S}(i)} \Pr\{A_j \bigcap B_j^c\} \\
&\le \frac{W\sqrt{W}}{\pi d_2}\|\widehat{\mathbf{E}}-\mathbf{E}\|_\infty
\end{aligned}$$

$\square$

A direct implication of Proposition 7 is,

Proposition 8. *$\forall\epsilon > 0$, let $\rho = \min\{\frac{d}{8}, \frac{\pi d_2 \epsilon}{W^{1.5}}\}$. If $\|\widehat{\mathbf{E}} - \mathbf{E}\|_\infty \leq \rho$, then, $q_i - p_i(\widehat{\mathbf{E}}) \leq \epsilon$ for a novel pair i and $p_j(\widehat{\mathbf{E}}) - q_j \leq \epsilon$ for a non-novel pair j.*

We now prove the consistency of Algorithm 2 of the main paper. Formally,

Lemma 2. *Algorithm 2 of the main paper can identify all the novel words from K distinct rankings with error probability,*

$$Pe \leq 2W^2 \exp(-Pq_\wedge^2/8) + 8W^2 \exp(-\rho^2\eta^4 MN/20)$$

where $\rho = \min\{\frac{d}{8}, \frac{\pi d_2 q_\wedge}{4W^{1.5}}\}$, $d_2 \triangleq (1-b)\lambda_{\min}$, $d = (1-b)^2\lambda_{\min}^2/\lambda_{\max}$, $b = \max_{j\in\mathcal{C}_0,k} \bar{B}_{j,k}$ and $\lambda_{\min}$, $\lambda_{\max}$ are the minimum /maximum eigenvalues of $\bar{\mathbf{R}}$. The result holds true for any isotropically distributed $\mathbf{d}$.

Proof. First of all, we decompose the error event to be the union of the following two types,

1. *Sorting error*, i.e., $\exists i \in \bigcup_{k=1}^K \mathcal{C}_k, \exists j \in \mathcal{C}_0$ such that $\hat{p}_i < \hat{p}_j$. This event is denoted as $A_{i,j}$ and let $A = \bigcup A_{i,j}$.

2. *Clustering error*, i.e., $\exists k, \exists i,j \in \mathcal{C}_k$ such that $i \notin \mathcal{J}_j$. This event is denoted as $B_{i,j}$ and let $B = \bigcup B_{i,j}$

According to Proposition 8, we also define $\rho = \min\{\frac{d}{8}, \frac{\pi d_2 q_\wedge}{4W^{1.5}}\}$ and $C = \{\|\mathbf{E} - \widehat{\mathbf{E}}\|_\infty \geq \rho\}$. Note that $B \subsetneq C$,

Therefore,

$$\begin{aligned} Pe &= \Pr\{A \bigcup B\} \\ &\leq \Pr\{A \bigcap C^c\} + \Pr\{C\} \\ &\leq \sum_{i\ novel, j\ non-novel} \Pr\{A_{i,j} \bigcap B^c\} + \Pr\{C\} \\ &\leq \sum_{i,j} \Pr(\hat{p}_i - \hat{p}_j < 0 \bigcap \|\widehat{\mathbf{E}} - \mathbf{E}\|_\infty \geq \rho) \\ &\quad + \Pr(\|\widehat{\mathbf{E}} - \mathbf{E}\|_\infty > \rho) \end{aligned}$$

The second term can be bound by Proposition 2. Now we focus on the first term. Note that

$$\begin{aligned} \hat{p}_i - \hat{p}_j &= \hat{p}_i - \hat{p}_j - p_i(\widehat{\mathbf{E}}) + p_i(\widehat{\mathbf{E}}) \\ &\quad -q_i + q_i - p_j(\widehat{\mathbf{E}}) + p_j(\widehat{\mathbf{E}}) - q_j + q_j \\ &= \{\hat{p}_i - p_i(\widehat{\mathbf{E}})\} + \{p_i(\widehat{\mathbf{E}}) - q_i\} \\ &\quad + \{p_j(\widehat{\mathbf{E}}) - \hat{p}_j\} + \{q_j - p_j(\widehat{\mathbf{E}})\} \\ &\quad + q_i - q_j \end{aligned}$$

and the fact that $q_i - q_j \geq q_\wedge$, then,,

$$\begin{aligned} &\Pr(\hat{p}_i < \hat{p}_j \bigcap \|\widehat{\mathbf{E}} - \mathbf{E}\|_\infty \leq \rho) \\ \leq\ & \Pr(p_i(\widehat{\mathbf{E}}) - \hat{p}_i \geq q_\wedge/4) + \Pr(\hat{p}_j - p_j(\widehat{\mathbf{E}}) \geq q_\wedge/4) \\ & + \Pr(q_i - p_i(\widehat{\mathbf{E}}) \geq q_\wedge/4) \bigcap \|\widehat{\mathbf{E}} - \mathbf{E}\|_\infty \leq \rho) \\ & + \Pr(p_j(\widehat{\mathbf{E}}) - q_j \geq q_\wedge/4) \bigcap \|\widehat{\mathbf{E}} - \mathbf{E}\|_\infty \leq \rho) \\ \leq\ & 2\exp(-Pq_\wedge^2/8) \\ & + \Pr(q_i - p_i(\widehat{\mathbf{E}}) \geq q_\wedge/4) \bigcap \|\widehat{\mathbf{E}} - \mathbf{E}\|_\infty \leq \rho) \\ & + \Pr(p_j(\widehat{\mathbf{E}}) - q_j \geq q_\wedge/4) \bigcap \|\widehat{\mathbf{E}} - \mathbf{E}\|_\infty \leq \rho) \end{aligned}$$

The last equality is by Proposition 6. For the last two terms, by Proposition 8 is 0. Therefore, applying Lemma 1 we obtain,

$$Pe \leq 2W^2 \exp(-Pq_\wedge^2/8) + 8W^2 \exp(-\rho^2\eta^4 MN/20)$$

□

D.4 Consistency of algorithm 3

Now we show that Algorithm 3 and 4 of the main paper can consistently estimate the ranking matrix $\boldsymbol{\sigma}$, given the success of the Algorithm 2. Without loss of generality, let $1, \ldots, K$ be the novel pairs of K distinct rankings. We first show that the solution of the constrained linear regression is consistent:

Proposition 9. *The solution to the following optimization problem*

$$\widehat{\mathbf{b}}^* = \arg \min_{b_j \geq 0, \sum b_j = 1} \|\widehat{\mathbf{E}}_i - \sum_{j=1}^K b_j \widehat{\mathbf{E}}_j\|$$

converges to the i-th row of $\bar{\mathbf{B}}$, $\bar{\mathbf{B}}_i$, as $M \to \infty$. Moreover,

$$\Pr(\|\widehat{\mathbf{b}}^* - \bar{\mathbf{B}}_i\|_\infty \geq \epsilon) \leq 8W^2 \exp(-\frac{\epsilon^2 MN\lambda_{\min}\eta^4}{80W^{0.5}})$$

Proof. We note that $\bar{\mathbf{B}}_i$ is the optimal solution to the following problem

$$\mathbf{b}^* = \arg \min_{b_j \geq 0, \sum b_j = 1} \|\mathbf{E}_i - \sum_{j=1}^K b_j \mathbf{E}_j\|$$

Define $f(\mathbf{E}, \mathbf{b}) = \|\mathbf{E}_i - \sum_{j=1}^K b_j \mathbf{E}_j\|$ and note the fact that $f(\mathbf{E}, \mathbf{b}^*) = 0$. Let $\mathbf{Y} = [\mathbf{E}_1^\top, \ldots, \mathbf{E}_K^\top]^\top$. Then,

$$\begin{aligned} & f(\mathbf{E}, \mathbf{b}) - f(\mathbf{E}, \mathbf{b}^*) = \|\mathbf{E}_i - \sum_{j=1}^K b_j \mathbf{E}_j\| - 0 \\ =& \|\sum_{j=1}^K (b_j - b_j^*)\mathbf{E}_j\| = \sqrt{(\mathbf{b} - \mathbf{b}^*)\mathbf{Y}\mathbf{Y}^\top(\mathbf{b} - \mathbf{b}^*)^\top} \\ \geq& \|\mathbf{b} - \mathbf{b}^*\|\lambda_{\min} \end{aligned}$$

where $\lambda_{\min} > 0$ is the minimum eigenvalue of $\bar{\mathbf{R}}$. Next, note that,

$$
\begin{aligned}
|f(\mathbf{E},\mathbf{b}) - f(\widehat{\mathbf{E}},\mathbf{b})| \leq & \|\mathbf{E}_i - \widehat{\mathbf{E}}_i + \sum b_j(\widehat{\mathbf{E}}_j - \mathbf{E}_j)\| \\
\leq & \|\mathbf{E}_i - \widehat{\mathbf{E}}_i\| + \sum b_j \|\widehat{\mathbf{E}}_j - \mathbf{E}_j\| \\
\leq & 2 \max_w \|\widehat{\mathbf{E}}_w - \mathbf{E}_w\|
\end{aligned}
$$

Combining the above inequalities, we obtain,

$$
\begin{aligned}
\|\widehat{\mathbf{b}}^* - \mathbf{b}^*\| \leq & \frac{1}{\lambda_{\min}}\{f(\mathbf{E},\widehat{\mathbf{b}}^*) - f(\mathbf{E},\mathbf{b}^*)\} \\
= & \frac{1}{\lambda_{\min}}\{f(\mathbf{E},\widehat{\mathbf{b}}^*) - f(\widehat{\mathbf{E}},\widehat{\mathbf{b}}^*) + f(\widehat{\mathbf{E}},\widehat{\mathbf{b}}^*) \\
& - f(\widehat{\mathbf{E}},\mathbf{b}^*) + f(\widehat{\mathbf{E}},\mathbf{b}^*) - f(\mathbf{E},\mathbf{b}^*)\} \\
\leq & \frac{1}{\lambda_{\min}}\{f(\mathbf{E},\widehat{\mathbf{b}}^*) - f(\widehat{\mathbf{E}},\widehat{\mathbf{b}}^*) \\
& + f(\widehat{\mathbf{E}},\mathbf{b}^*) - f(\mathbf{E},\mathbf{b}^*)\} \\
\leq & \frac{4W^{0.5}}{\lambda_{\min}}\|\widehat{\mathbf{E}} - \mathbf{E}\|_\infty
\end{aligned}
$$

where the last term converges to 0 almost surely. The convergence rate follows directly from Lemma 1. □

Now for the row-scaling step in algorithm 3,

$$
\begin{aligned}
\hat{\mathbf{B}}_i := & \hat{\mathbf{b}}^*(i)^\top (\frac{1}{M}\mathbf{X}\mathbf{1}_{M\times 1}) \\
& \to \bar{\mathbf{B}}_i(\mathbf{B}_i\mathbf{a}) = \mathbf{B}_i \operatorname{diag}(\mathbf{a}) \qquad (6)
\end{aligned}
$$

We point out that the "column-normalization" step in Ding et al. [2014] which was used to get rid of the diag($\mathbf{a}$) component in the above equation is not necessary in our approach. To show the convergence rate of the above equation, it is straightforward to apply the result in Lemma 1

Proposition 10. *For the row-scaled estimation $\hat{\mathbf{B}}_i$ as in Eq.* (6), *we have,*

$$\Pr(|\hat{\mathbf{B}}_{i,k} - \mathbf{B}_{i,k}a_k| \geq \epsilon) \leq 8W^2 \exp(-\frac{\epsilon^2 MN\lambda_{\min}\eta^4}{160W^{0.5}})$$

Proof. By Proposition 9, we have,

$$\Pr(|\widehat{\mathbf{b}}^*(i)_k - \bar{\mathbf{B}}_{i,k}| \geq \epsilon/2) \leq 8W^2 \exp(-\frac{\epsilon^2 MN\lambda_{\min}\eta^4}{160W^{0.5}})$$

Recall that,

$$\Pr(|\frac{1}{M}\mathbf{X}\mathbf{1}_{M\times 1} - \mathbf{B}_i\mathbf{a}| \geq \epsilon/2) \leq \exp(-\epsilon^2 MN/2)$$

Therefore,

$$
\begin{aligned}
& \Pr(|\hat{\mathbf{B}}_{i,k} - \mathbf{B}_{i,k}a_k| \geq \epsilon) \\
\leq & 8W^2 \exp(-\frac{\epsilon^2 MN\lambda_{\min}\eta^4}{80W^{0.5}}) + \exp(-\epsilon^2 MN/2)
\end{aligned}
$$

where the second term is dominated by the first term. □

For the rest of this section, we will use (i,j) to index the W rows of $\mathbf{E}, \mathbf{B}, \boldsymbol{\sigma}$. Recall in Eq. (6), $\widehat{\mathbf{B}}_{(i,j),k} \to \mathbf{B}_{(i,j),k}a_k = \mu_{i,j}\sigma_{(i,j),k}a_k$, and $\widehat{\mathbf{B}}_{(j,i),k} \to \mathbf{B}_{(j,i),k}a_k = \mu_{i,j}\sigma_{(j,i),k}a_k$, and in algorithm 1 of the main paper, we consider

$$
\begin{aligned}
\widehat{\sigma}_{(i,j),k} \leftarrow & \frac{\widehat{\mathbf{B}}_{(i,j),k}}{\widehat{\mathbf{B}}_{(i,j),k} + \widehat{\mathbf{B}}_{(j,i),k}} \\
\doteq & \frac{\sigma_{(i,j),k}\mu_{i,j}a_k}{\sigma_{(i,j),k}\mu_{i,j}a_k + \sigma_{(j,i),k}\mu_{i,j}a_k}
\end{aligned}
$$

Therefore, due to the rounding scheme of the last step, the estimation is consistent if $|\widehat{\mathbf{B}}_{(i,j),k} - \mathbf{B}_{(i,j),k}a_k| \leq 0.5\mu_{i,j}a_k$. η is a lower bound of $\mu_{i,j}a_k$. Putting the above results together, we have,

Lemma 3. *Given the success in Lemma 2, Algorithm 3 and the remaining post-processing steps in Algorithm 1 of the main paper can consistently estimate the ranking matrix $\boldsymbol{\sigma}$ as $M \to \infty$. Moreover, the error probability is less than $8W^2\exp(-\frac{MN\lambda_{\min}\eta^6}{160W^{0.5}})$.*

D.5 Proof of Theorem 2

We now formally prove the sample complexity Theorem 2 in the main paper.

Theorem 2 Let $\boldsymbol{\sigma}$ be separable and $\mathbf{R}$ be full rank. Then the overall Algorithm 1 consistently recovers $\boldsymbol{\sigma}$ up to a column permutation as the number of users $M \to \infty$ and number of projections $P \to \infty$. Furthermore, $\forall \delta > 0$, if

$$M \geq \max\left\{40\frac{\log(3W/\delta)}{N\rho^2\eta^4},\ 320\frac{W^{0.5}\log(3W/\delta)}{N\eta^6\lambda_{\min}}\right\}$$

and for

$$P \geq 16\frac{\log(3W/\delta)}{q_\wedge^2}$$

then Algorithm 1 fails with probability at most δ. The other model parameters are defined as $\eta = \min_{1\leq w\leq W}[\mathbf{B}\mathbf{a}]_w$, $\rho = \min\{\frac{d}{8}, \frac{\pi d_2 q_\wedge}{4W^{1.5}}\}$, $d_2 \triangleq (1-b)\lambda_{\min}$, $d = (1-b)^2\lambda_{\min}^2/\lambda_{\max}$, $b = \max_{j\in\mathcal{C}_0,k}\bar{B}_{j,k}$ and $\lambda_{\min}, \lambda_{\max}$ are the minimum /maximum eigenvalues of $\bar{\mathbf{R}}$. $q_\wedge$ is the minimum normalized solid angle of the extreme points of the convex hull of the rows of $\mathbf{E}$.

Proof. We combine the results in Lemmas 2 and 3, i.e., the error probability of alg. 1 can be upper bounded by

$$
\begin{aligned}
Pe \leq & 2W^2\exp(-Pq_\wedge^2/8) + 8W^2\exp(-\rho^2\eta^4 MN/20) \\
& + 8W^2\exp(-\frac{MN\lambda_{\min}\eta^6}{160W^{0.5}})
\end{aligned}
$$

This leads to the sample complexity results in the theorem. □

E Algorithm 2 and Theorem 2 for Gaussian Random Directions

The proof in Section D holds for any isotropic distribution on $\mathbf{d}$. If we assume $\mathbf{d}$ to be the standard spherical Gaussian distribution, we can have better sample complexity bounds following the steps in [Ding et al., 2014, Theorem 2]. First note that,

Proposition 11. *Let $\mathbf{X}^n, \mathbf{X} \in \mathbb{R}^m$ be two random vectors, $\mathbf{a}, \boldsymbol{\epsilon} \in \mathbb{R}^m$ be two vectors and $\boldsymbol{\epsilon} > \mathbf{0}$.*

$$\begin{aligned}&|\Pr\{\mathbf{X}^n \leq \mathbf{a}\} - \Pr\{\mathbf{X} \leq \mathbf{a}\}| \\ &\leq \Pr(\exists i : |X_i^n - X_i| \geq \epsilon_i) + \Pr(\mathbf{a} - \boldsymbol{\epsilon} \leq \mathbf{X} \leq \mathbf{a} + \boldsymbol{\epsilon})\end{aligned}$$

The inequality is element-wise.

Proof. Note that

$$\begin{aligned}\Pr\{\mathbf{X}^n \leq \mathbf{a}\} \leq &\Pr\{\mathbf{X}^n \leq \mathbf{a}, \forall i : |X_i^n - X_i| \leq \epsilon_i\} \\ &+ \Pr\{\mathbf{X}^n \leq \mathbf{a}, \exists i : |X_i^n - X_i| \geq \epsilon_i\} \\ \leq &\Pr\{\mathbf{X} \leq \mathbf{a} + \boldsymbol{\epsilon}\} + \Pr\{\exists i : |X_i^n - X_i| \geq \epsilon_i\}\end{aligned}$$

Similarly, by swapping $\mathbf{X}^n$ and $\mathbf{X}$, we have,

$$\Pr\{\mathbf{X} \leq \mathbf{a} - \boldsymbol{\epsilon}\} \leq \Pr\{\mathbf{X}^n \leq \mathbf{a}\} + \Pr\{\exists i : |X_i^n - X_i| \geq \epsilon_i\}$$

Combining them concludes the proof. □

Proposition 12. *Let the random projection directions be $\mathbf{d} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_W)$ in Algorithm 2 of the main paper. Then, $\forall\ \epsilon > 0$, let $\rho = \min\{\frac{d}{8}, \frac{\sqrt{\pi}\epsilon d_2}{4K\sqrt{W \log(2W/\epsilon)}}\}$. If $\|\widehat{\mathbf{E}} - \mathbf{E}\|_\infty \leq \rho$, then, $q_i - p_i(\widehat{\mathbf{E}}) \leq \epsilon$ for a novel pair i and $p_j(\widehat{\mathbf{E}}) - q_j \leq \epsilon$ for a non-novel pair j.*

Proof. Recall the definition of q_i and $p_i(\widehat{\mathbf{E}})$,

$$\begin{aligned}q_i &= \Pr\{\forall j \in \mathcal{S}(i),\ \mathbf{E}_i\mathbf{d} \geq \mathbf{E}_j\mathbf{d}\} \\ p_i(\widehat{\mathbf{E}}) &= \Pr\{\forall j \in \mathcal{J}_i,\ \widehat{\mathbf{E}}_i\mathbf{d} \geq \widehat{\mathbf{E}}_i\mathbf{d}\}\end{aligned}$$

When i is a novel word, $\mathcal{S}(i) = \mathcal{J}_i$ for $\|\widehat{\mathbf{E}} - \mathbf{E}\|_\infty \leq \rho \leq d/8$, therefore, by Proposition 11, we have,

$$\begin{aligned}|q_i - p_i(\widehat{\mathbf{E}})| \leq &\Pr(\exists j \in \mathcal{J}_i : |\mathbf{e}_{i,j}\mathbf{d}| \geq \delta) \\ &+ \Pr(\forall j \in \mathcal{J}_i : |\mathbf{z}_{ij}\mathbf{d}| \leq \delta)\end{aligned} \tag{7}$$

where $\mathbf{e}_{i,j} = \mathbf{E}_i - \widehat{\mathbf{E}}_i + \widehat{\mathbf{E}}_j - \mathbf{E}_j$ and $\mathbf{z}_{ij} = \mathbf{E}_i - \mathbf{E}_j$. To apply the union bound to the second term in Eq. (7), it suffice to consider only $j \in \bigcup_{k=1}^K \mathcal{C}_k$. Therefore, by union bounding both the first and second terms, we obtain,

$$\begin{aligned}&|q_i - p_i(\widehat{\mathbf{E}})| \\ &\leq \sum_j \Pr(|\mathbf{e}_{i,j}\mathbf{d}| \geq \delta) + \sum_j \Pr(|\mathbf{z}_{ij}\mathbf{d}| \leq \delta)\end{aligned}$$

Note that $\mathbf{e}_{ij}\mathbf{d} \sim \mathcal{N}(0, \|\mathbf{z}_{ij}\|_2^2)$ and $\mathbf{z}_{ij}\mathbf{d} \sim \mathcal{N}(0, \|\mathbf{a}_{ij}\|_2^2)$ conditioned on $\widehat{\mathbf{E}}$. Using the properties of the Gaussian distribution we have,

$$\begin{aligned}\Pr(|\mathbf{z}_{ij}\mathbf{d}| \leq \delta) &= \int_{-\delta}^{\delta} \frac{1}{\sqrt{2\pi}\|\mathbf{z}_{ij}\|} e^{-t^2/2\|\mathbf{z}_{ij}\|^2} dt \\ &\leq \frac{\sqrt{2/\pi}}{\|\mathbf{z}_{ij}\|}\delta\end{aligned}.$$

By Proposition 3, $\|\mathbf{z}_{ij}\| \geq d_2$ for $j \in \mathcal{J}_i$, therefore, $\Pr(|\mathbf{z}_{ij}\mathbf{d}| \leq \delta) \leq \frac{\sqrt{2/\pi}}{d_2}\delta$. Similarly, note that

$$\Pr(|\mathbf{e}_{i,j}\mathbf{d}| \geq \delta | \widehat{\mathbf{E}}) = 2Q(\delta/\|\mathbf{e}_{i,j}\|) \leq \exp(-\delta^2/2\|\mathbf{e}_{i,j}\|^2)$$

by the property of the Q-function. Note that

$$\begin{aligned}\|\mathbf{e}_{i,j}\| \leq &\|\mathbf{E}_i - \widehat{\mathbf{E}}_i\| + \|\widehat{\mathbf{E}}_j - \mathbf{E}_j\| \\ \leq &2W^{0.5}\|\mathbf{E} - \widehat{\mathbf{E}}\|_\infty\end{aligned}$$

Then, by marginalizing over $\widehat{\mathbf{E}}$ we obtain, $\Pr(|\mathbf{e}_{i,j}\mathbf{d}| \geq \delta) \leq \exp(-\delta^2/8W\|\mathbf{E} - \widehat{\mathbf{E}}\|_\infty^2)$. Combining these results, we obtain,

$$|q_i - p_i(\widehat{\mathbf{E}})| \leq K\frac{\sqrt{2/\pi}}{d_2}\delta + W\exp(-\delta^2/8W\|\mathbf{E} - \widehat{\mathbf{E}}\|_\infty^2)$$

hold true for any $\delta > 0$. Therefore, if we set $\delta = \frac{\epsilon_0\rho}{2K\sqrt{2/\pi}}$, and require

$$\|\mathbf{E} - \widehat{\mathbf{E}}\|_\infty \leq \frac{\sqrt{\pi}\epsilon d_2}{4K\sqrt{W\log(2W/\epsilon)}}$$

then $|q_i - p_i(\widehat{\mathbf{E}})| \leq \epsilon$. In summary, we require $\|\mathbf{E} - \widehat{\mathbf{E}}\|_\infty \leq \min\{\frac{\sqrt{\pi}\epsilon d_2}{4K\sqrt{W\log(2W/\epsilon)}}, d/8\}$. We note that the argument above holds true for a non-novel pair as well. □

In Proposition 12, the bound on $\|\mathbf{E} - \widehat{\mathbf{E}}\|_\infty$ is,

$$\min\{\frac{d}{8}, \frac{\sqrt{\pi}\epsilon d_2}{4K\sqrt{W\log(2W/\epsilon)}}\}$$

which is an improvement over the result in Proposition 8,

$$\min\{\frac{d}{8}, \frac{\pi d_2 \epsilon}{W^{1.5}}\}$$

where we could reduce the dependence on W from $W\sqrt{W}$ to $K\sqrt{W}$. Since $K \ll W$, we obtain a gain over the general isotropic distribution. This leads to lightly improved results for the overall sample complexity bounds:

Theorem 2(Gaussian Random Projections) Let $\boldsymbol{\sigma}$ be separable and $\mathbf{R}$ be full rank. Then the overall Algorithm 1 consistently recovers $\boldsymbol{\sigma}$ up to a column permutation as the number of users $M \to \infty$ and number of projections $P \to \infty$. Furthermore, if the random directions for projections are drawn from a spherical Gaussian distribution, then $\forall \delta > 0$, if

$$M \geq \max\left\{40\frac{\log(3W/\delta)}{N\rho^2\eta^4},\ 320\frac{W^{0.5}\log(3W/\delta)}{N\eta^6\lambda_{\min}}\right\}$$

and for

$$P \geq 16\frac{\log(3W/\delta)}{q_\wedge^2}$$

then Algorithm 1 fails with probability at most δ. The other model parameters are defined as $\eta = \min_{1\leq w\leq W}[\mathbf{Ba}]_w$, $\rho = \min\{\frac{d}{8}, \frac{\sqrt{\pi}d_2 q_\wedge}{4K\sqrt{W\log(2W/q_\wedge)}}\}$, $d_2 \triangleq (1-b)\lambda_{\min}$, $d = (1-b)^2\lambda_{\min}^2/\lambda_{\max}$, $b = \max_{j\in\mathcal{C}_0,k}\bar{B}_{j,k}$ and $\lambda_{\min}$, $\lambda_{\max}$ are the minimum /maximum eigenvalues of $\bar{\mathbf{R}}$. $q_\wedge$ is the minimum normalized solid angle of the extreme points of the convex hull of the rows of $\mathbf{E}$.

F Proof of Theorem 1

The stated computational efficiency can be achieved in the same way as discussed in Proposition 1 and 2 in Ding et al. [2014]. We need to point out that the post-processing steps in Algorithm 4 requires a computation time of $\mathcal{O}(WK)$ which is dominated by that of the Algorithm 2 and 3.

References

W. Ding, P. Ishwar, M. H. Rohban, and V. Saligrama. Necessary and Sufficient Conditions for Novel Word Detection in Separable Topic Models. In *Advances in on Neural Information Processing Systems (NIPS), Workshop on Topic Models: Computation, Application*, Lake Tahoe, NV, USA, Dec. 2013.

W. Ding, M. H. Rohban, P. Ishwar, and V. Saligrama. Efficient Distributed Topic Modeling with Provable Guarantees. In *Proc. ot the 17th International Conference on Artificial Intelligence and Statistics*, Reykjavik, Iceland, Apr. 2014.

V. Farias, S. Jagabathula, and D. Shah. A data-driven approach to modeling choice. In *Advances in Neural Information Processing Systems*. Vancouver, Canada, Dec. 2009.

S. Jagabathula and D. Shah. Inferring rankings under constrained sensing. In *Advances in Neural Information Processing Systems*, pages 753–760. Vancouver, Canada, Dec. 2008.