Controlling the distance to a Kemeny consensus without computing it

Yunlong Jiao*, Anna Korba† and Eric Sibony†
* Mines ParisTech – CBIO † Télécom ParisTech, CNRS LTCI

Introduction

We study the problem of **ranking aggregation**: agents provide a collection of ranked preferences over a set of alternatives, we wish to aggregate them into one consensus ranking.

One popular approach in ranking aggregation follows Kemeny's rule. Given a collection of rankings/permutations $\mathcal{D}_N := (\sigma_1, \dots, \sigma_N) \in \mathfrak{S}_n^N$ over n alternatives, **Kemeny consensus(es)** are the solution to

$$\mathcal{K}_N := \operatorname*{arg\,min}_{\sigma \in \mathfrak{S}_n} \sum_{i=1}^{N} d(\sigma, \sigma_t)\,,$$

where d is the Kendall's tau distance between permutations, i.e.,

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

Kemeny consensuses satisfy many desirable properties but are NP-hard to compute even for $n = 4$. It thus calls for study on apprehending the complexity of Kemeny aggregation and theoretical guarantees of approximation procedures commonly used in practice.

Two-minute summary

The problem: Let $\mathcal{D}_N \in \mathfrak{S}_n^N$ be a dataset and $\sigma \in \mathfrak{S}_n$ be a permutation, typically output by a computationally efficient aggregation procedure on $\mathcal{D}_N$. Can we give a (tractable) upper bound for the distance $d(\sigma, \sigma^*)$ between σ and a Kemeny consensus $\sigma^* \in \mathcal{K}_N$?

The bound: Denote by $0 \le \theta_N(\sigma) \le \pi$ the angle between the Kemeny embeddings $\phi(\sigma)$ and $\phi(\mathcal{D}_N)$ in an Euclidean space. If $\theta_N(\sigma) < \frac{\pi}{2}$, then for all $\sigma^* \in \mathcal{K}_N$,

$$d(\sigma, \sigma^*) \le \left\lfloor \binom{n}{2} \sin^2(\theta_N(\sigma)) \right\rfloor =: k_{min}(\sigma; \mathcal{D}_N)\,,$$

where $\lfloor x \rfloor$ denotes the integer part of the real x.

The distinguishing merits:

- k_{min} is simple to code and efficient to compute in time $O(Nn^2)$.
- Generality with no assumption on the dataset $\mathcal{D}_N$ or the aggregation procedure giving σ.
- Depends on geometric understanding of the combinatorial problem of Kemeny aggregation.

Results on the *sushi* dataset

Voting rule	$\cos(\theta_N(\sigma))$	$k_{min}(\sigma)$
Borda	0.820	14
Copeland	0.822	14
QuickSort	0.822	14
Plackett-Luce	0.80	15
2-approval	0.745	20
1-approval	0.710	22
Pick-a-Perm	0.383†	34.85†
Pick-a-Random	0.377†	35.09†

Table 1: k_{min} with different aggregation procedures on the original *sushi* dataset.

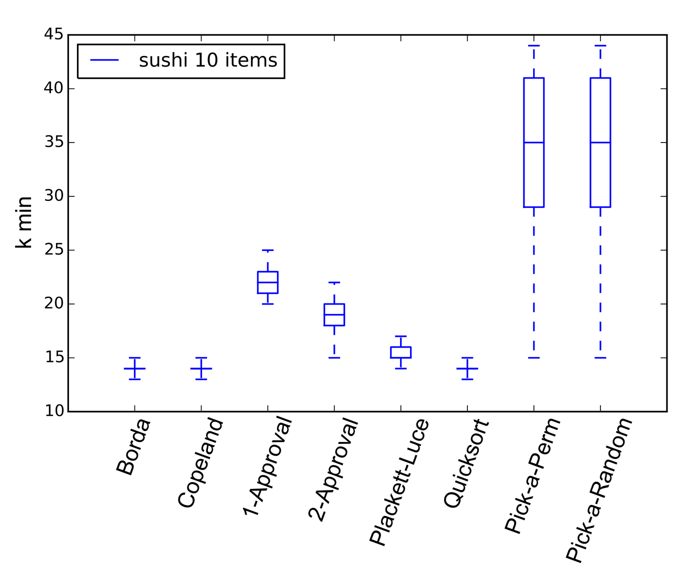

Figure 1: k_{min} with different aggregation procedures on 500 bootstrapped pseudo-samples of the *sushi* dataset ($N = 5000, n = 10$).

Geometric analysis of Kemeny aggregation

Kemeny embedding: The Kemeny embedding of a single permutation $\sigma \in \mathfrak{S}_n$ is defined by:

$$\phi : \mathfrak{S}_n \to \mathbb{R}^{\binom{n}{2}}; \sigma \mapsto (\operatorname{sign}(\sigma(j) - \sigma(i)))_{1 \le i < j \le n}\,,$$

In particular, $\|\phi(\sigma)\| = \sqrt{\frac{n(n-1)}{2}}$ for all $\sigma \in \mathfrak{S}_n$, i.e., $\phi(\sigma)$ lies on the sphere centered at origin with radius $R := \sqrt{\frac{n(n-1)}{2}}$.

The Kemeny embedding induces a space where the squared Euclidean distance recovers the Kendall's tau distance, i.e., for all $\sigma, \sigma' \in \mathfrak{S}_n$,

$$d(\sigma, \sigma') = \frac{1}{4}\|\phi(\sigma) - \phi(\sigma')\|^2\,.$$

Kemeny aggregation in the embedded space: For any dataset $\mathcal{D}_N := (\sigma_1, \dots, \sigma_N) \in \mathfrak{S}_n^N$, Kemeny aggregation is equivalent to solving:

$$\mathcal{K}_N = \operatorname*{arg\,min}_{\sigma \in \mathfrak{S}_n} \|\phi(\sigma) - \phi(\mathcal{D}_N)\|^2 = \operatorname*{arg\,min}_{\sigma \in \mathfrak{S}_n} \theta_N(\sigma)\,, \tag{1}$$

where

$$\phi(\mathcal{D}_N) := \frac{1}{N}\sum_{t=1}^{N} \phi(\sigma_t)$$

denotes the barycenter of the point cloud $\mathcal{D}_N$, and $\theta_N(\sigma)$ denotes the Euclidean angle between the Kemeny embeddings $\phi(\sigma)$ and $\phi(\mathcal{D}_N)$.

Kemeny aggregation decomposes in two steps:

1. Compute the mean embedding of the dataset $\phi(\mathcal{D}_N)$ in time $O(Nn^2)$.
2. Find the consensus permutation σ^* that minimizes (1).

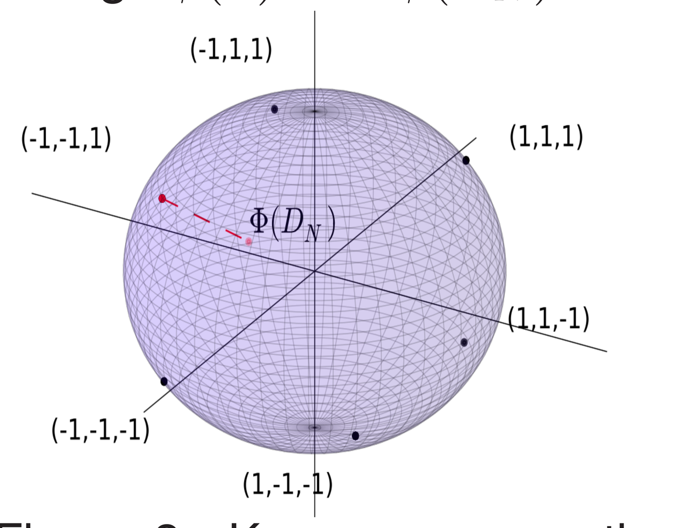

Figure 2: Kemeny aggregation for $n = 3$.

Geometric interpretation of the bound

Fix $\mathcal{D}_N \in \mathfrak{S}_n^N$ and let $\sigma \in \mathfrak{S}_n$ be a permutation and $\sigma^* \in \mathcal{K}_N$ be a Kemeny consensus. Since $\theta_N(\sigma^*) \le \theta_N(\sigma) < \frac{\pi}{2}$, applying laws of cosines in the 2-dimensional subspace spanned by $\phi(\mathcal{D}_N)$ and $\phi(\sigma)$ gives

$$r^2 > 2R^2(1 - 2\cos(2\theta_N(\sigma))) = 4R^2 \sin^2(\theta_N(\sigma))\,.$$

The minimum integer of $\frac{r^2}{4}$ satisfying the inequality recovers k_{min} exactly.

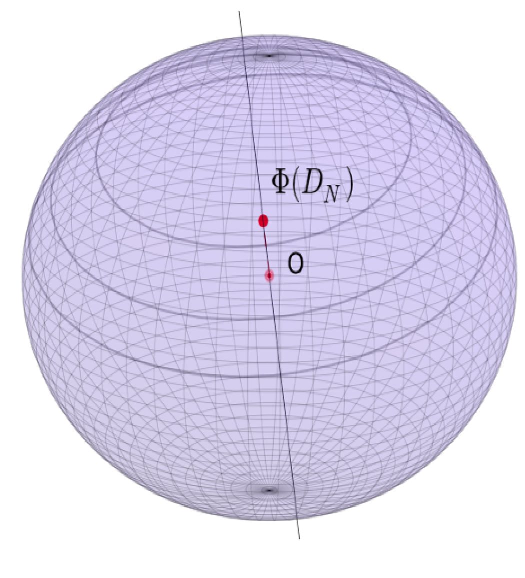

Figure 3: Level sets of the cost function in (1) over the sphere for $n = 3$.

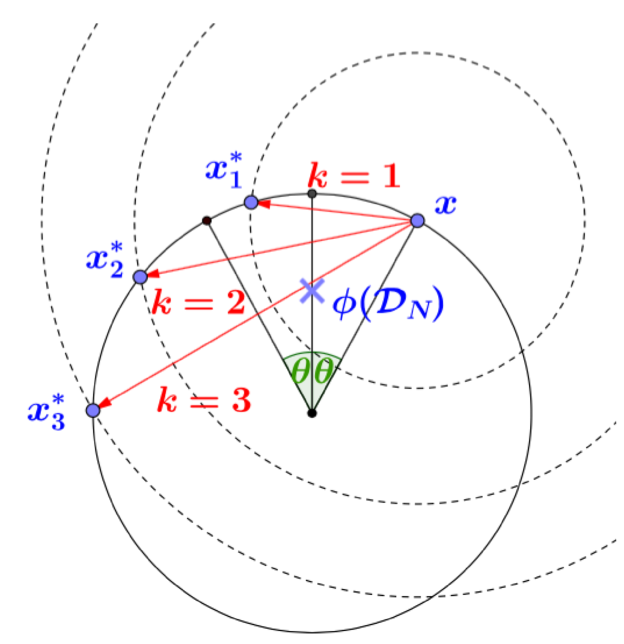

Figure 4: Geometric illustration of the bound taking $x = \phi(\sigma), k = \frac{r^2}{4}$.

Conclusion

- **Methodological:** geometric properties of Kemeny aggregation.
- **Theoretical:** simple and tractable quantity controlling the distance to a Kemeny consensus without computing it.
- **Extensible:** ranking aggregation from partial orders, etc.

References

[1] J. P. Barthelemy and B. Monjardet. *The median procedure in cluster analysis and social choice theory.* Mathematical Social Sciences, 1981.

[2] W. S. Zwicker. *Consistency without neutrality in voting rules: When is a vote an average?* Mathematical and Computer Modelling, 2008.