# How the Sausage is Made

## LLM-embedding-based Heuristic Hierarchical Clustering For Organizing TeaP2025 Talk & Poster Sessions

MAX PLANCK INSTITUTE FOR EMPIRICAL AESTHETICS

**Seung-Goo Kim**[1], **Daniela Sammler**[1,2]

[1] Research Group Neurocognition of Music and Language, Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany
[2] Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

## Introduction

Organizing talks and poster presentations into scientifically coherent sessions is non-trivial at a large scale. Motivated by recent applications of large language models (LLMs) in topic extraction and clustering of scientific literature [1], we were interested in whether an LLM could reduce human efforts (i.e., ~40 HIWI-hours) at a reasonable cost (i.e., <€1).

## Conclusion

While the automated clustering algorithm yielded better-than-nothing groundwork, it still required a considerable amount of human post-processing (**Fig 5,6**). This suggests that LLM-based tools can be helpful to some extent when combined with human expertise.

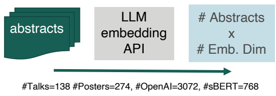



**Fig 2. Embedding extraction.** OpenAI model running cost was 0.69 USD for 447K tokens. sBERT model was freely hosted from 🤗.

## Methods & Materials

**Text.** Human reviewers accepted 138 talks and 274 posters. Input texts were the title, abstract, author keyword, and reviewer keyword.

**LLM embedding.** We used OpenAI embedding model "`text-embedding-3-large`" (https://platform.openai.com/docs/models#embeddings). For comparison, open-source Sentence BERT model "`all-mpnet-base-v2`" (https://huggingface.co/sentence-transformers/all-mpnet-base-v2) was used. We used squared Euclidean distance for Ward linkage (**Fig 2**).

**Iterative hierarchical clustering.** Since the clustering is restricted by the number of talks per session, an iterative algorithm was used [2,3] that prunes leaves in the best branch at each iteration (**Fig 3**).

**Post-processing.** A human expert (D.S.) adjusted clustering to increase within-cluster coherence (~10 PhD/PD-hours).

**Optimal spatiotemporal curation.** Since no attendee is omnipresent😇, parallel sessions were made maximally distant to avoid conflicts. Moreover, the poster clusters in the same room (and posters within each cluster) were placed to align physical and semantic proximities (**Fig 4**).

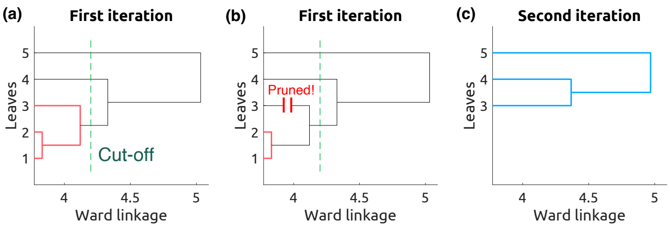



**Fig 3. Iterative hierarchical clustering.** As a toy example, we are to find 2 clusters with 2 leaves and 3 leaves, respectively. **(a)** The best cluster (red) is found in the first iteration. **(b)** The cluster is too big (3 leaves > 2 leaves), thus the algorithm prunes one distant leaf. **(c)** The best cluster (blue) is found in the second iteration. The algorithm ends with {1,2}, {3,4,5}.

**Fig 1. Methods overview.** `Model="GPT-4o"; Prompt="can you make it a bit cuter not gross?"`. Of course, this has nothing to do with the actual methods but is very cute. 🥰

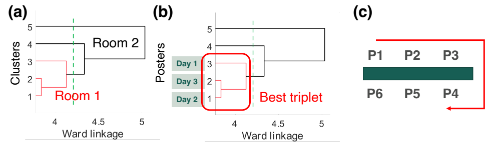
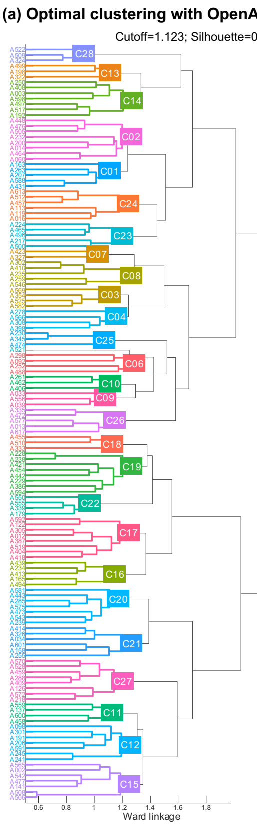



**Fig 4. Optimal spatiotemporal curation. (a)** A linkage structure was found on cluster-wise (averaged) embedding vectors. Then, the clusters were hyper-clustered into two rooms. This was for attendees to be able to see all the interesting posters in one room (i.e., path minimization). **(b)** In each room-topic cluster, the closest triplet of posters was iteratively found, and then they were randomly assigned to either of three days to ensure that poster presenters could also visit other posters of their interest. **(c)** Finally, posters (P1, P2, …) within each day-room-topic cluster were again ordered based on their semantic similarity (i.e., physically close posters are also semantically close along the path of viewing [red arrow]; also minimizing the motion path).



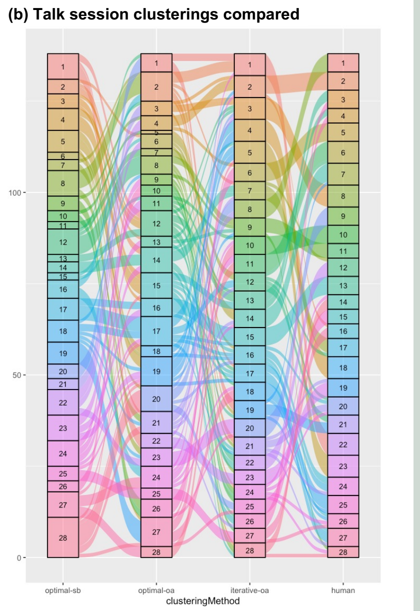
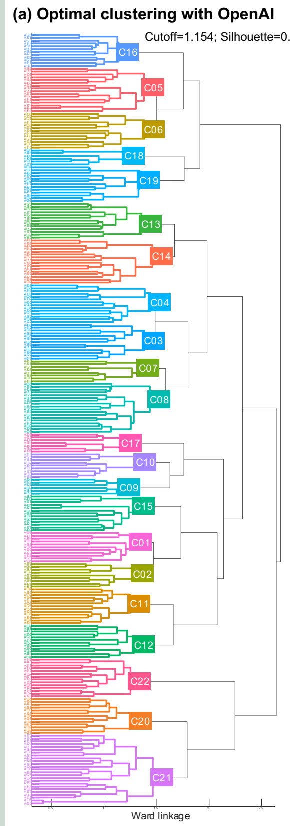




**Fig 5. Talk session clustering. (a)** Optimal clustering based on OpenAI for 28 clusters in a dendrogram. Abstract submission numbers (A*) are shown vertically, Ward linkage distance between clusters is shown horizontally. **(b)** Alluvial plot [4] displays cluster changes across clustering methods (**optimal-sb**, optimal clustering with sBERT-embedding; **optimal-oa**, optimal clustering with OpenAI-embedding, **iterative-oa**, iterative hierarchical clustering with cluster-size constraints; **human**, clustering after human post-processing). Cluster indices were arbitrarily assigned. Thicker connections across methods indicate more preservation of clustering structures.



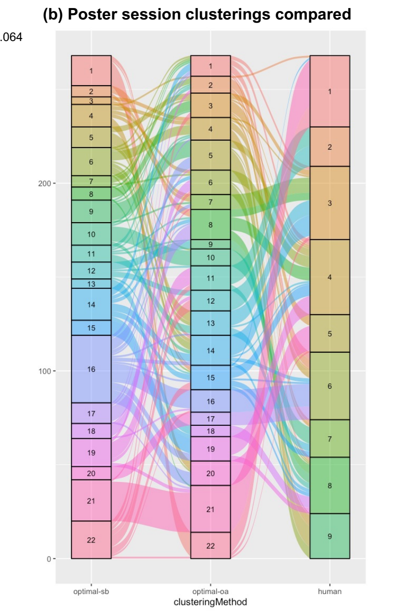




**Fig 6. Poster session clustering. (a)** Optimal clustering based on OpenAI for 22 clusters in a dendrogram. The optimal number of clusters (K=22) was determined based on the silhouette score. **(b)** Alluvial plot displays cluster changes across clustering methods (**optimal-sb**, optimal clustering with sBERT-embedding; **optimal-oa**, optimal clustering with OpenAI-embedding, **human**, clustering after human post-processing).

**We hope you enjoy the tour!**

**Contact:** Seung-Goo Kim
seung-goo.kim@ae.mpg.de
https://www.ae.mpg.de/
+49 69 8300 479 851

**Acknowledgment:** LLMs were strictly used only in defining abstract-wise distances. All selection procedures were based on human decisions. All text inputs are publicly available on the TeaP2025 official website (https://coms.app/TeaP2025/). This study was funded privately (OpenAI API token fee by S.G.K) and publicly (printing of this poster by Max Planck Society). Sponsors of this study did not play any role in study design, data collection, analysis, and interpretation (well, at least not as a sponsor).

**References:** **[1]** Tshitoyan et al., 2019, Nature, https://doi.org/10.1038/s41586-019-1335-8 **[2]** https://jmonlong.github.io/Hippocamplus/2018/06/09/cluster-same-size/ **[3]** https://de.mathworks.com/help/stats/cluster.html
**[4]** https://cran.r-project.org/web/packages/ggalluvial/vignettes/ggalluvial.html