

Learning meta-features from **MAGIC** cards.

THE GATHERING

Cory Bennett, Murtaza Husain

Overview

- Our dataset contains **card data** (text, stats) as well as role tags from **Magic: The Gathering cards**
- In this project, we want to explore how to cluster cards **conceptually**. We expect that assigning conceptual tags to cards will allow for effective clustering into different **game features** and **strategies**.
- Our clusters will be evaluated using metrics such as **silhouette scores**, **accuracy**, **precision**, and **recall**. This is possible because a subset of cards comes with tags, meaning we can evaluate the results by comparing **learned groupings vs known groupings**. We generally expect the model will work well considering the **many different groupings** that exist for strategies and the **high quality** of the dataset.

Datasets

Project Datasets

Card dataset (30,837 rows x 3 columns)

	id	name	text
0	d61fb9e8-d05a-481a-a90f-5def300c9abb	Cordial Vampire	Whenever this creature or another creature die...
1	bd637245-52bc-4cc9-810d-253e69dc3394	Butcher's Glee	Target creature gets +3/+0 and gains lifelink ...
2	a1fb6dfc-1dab-4337-8311-b4ab2ef336c0	Delusions of Mediocrity	When this enchantment enters, you gain 10 life...
3	852d2a7f-5082-4668-978d-51776cad2cfd	Jeska and Kamahl	Haste, first strike\n{T}: Jeska and Kamahl dea...
4	a9b9e798-debc-4f2c-9770-e16f670082f7	Koskun Falls	At the beginning of your upkeep, sacrifice thi...



Card text represents mechanics and abilities.

Tag dataset (2,257 rows x 3 columns)

	id	name	description
0	827b8666-9058-457d-a732-f2373fa0b2c7	hate-color	Color hate for one or more colors. See also [h...
1	abdb5417-eee7-4ba2-a314-a75b08d8db71	freeze	A permanent doesn't untap during its next N un...
2	5a0a0531-5462-477d-add9-886148c4413f	card-names	A collection of tags identifying properties an...
3	22728b1f-f880-44d6-a6ef-66f2784c22ed	helper-card	Cards printed to assist you with using mechanics.
4	0641a74c-4dd5-426d-be58-2ab86d71995d	burn	Effects that deal damage, whether to creatures...

Tags describe game concepts and strategies.

Some describe non-text features. Other tags are *categorical* and contain descendants, like “burn” contains sub-tags “burn-player”, “burn-creature”, “burn-any”, etc.

Each tag contains a set of tagged cards as examples.

Card Embeddings - Generation Process

Card dataset (30,837 rows x 3 columns)

	id	name	text
0	d61fb9e8-d05a-481a-a90f-5def300c9abb	Cordial Vampire	Whenever this creature or another creature die...
1	bd637245-52bc-4cc9-810d-253e69dc3394	Butcher's Glee	Target creature gets +3/+0 and gains lifelink ...
2	a1fb6dfc-1dab-4337-8311-b4ab2ef336c0	Delusions of Mediocrity	When this enchantment enters, you gain 10 life...
3	852d2a7f-5082-4668-978d-51776cad2cfd	Jeska and Kamahl	Haste, first strike\n{T}: Jeska and Kamahl dea...
4	a9b9e798-debc-4f2c-9770-e16f670082f7	Koskun Falls	At the beginning of your upkeep, sacrifice thi...

Card Metadata

Learned embeddings capture only semantic information of the card text. The model has no understanding of the game rules, keywords, or concepts.

We want to learn this from the tag dataset.

emb_0 ... emb_1023
[embedding vector]
[embedding vector]
[embedding vector]
[embedding vector]
[embedding vector]

New Features

BAAI/bge-large-en-v1.5

Embedding Model
(batch size 64)




Card text represents mechanics and abilities.

Tag Embeddings - Generation Process

Card embeddings (30,837 rows x 1027 columns)

	id	name	text	emb_0 ... emb_1023
0	d61fb9e8-d05a-481a-a90f-5def300c9abb	Cordial Vampire	Whenever this creature or another creature die...	[embedding vector]
1	bd637245-52bc-4cc9-810d-253e69dc3394	Butcher's Glee	Target creature gets +3/+0 and gains lifelink ...	[embedding vector]
2	a1fb6dfc-1dab-4337-8311-b4ab2ef336c0	Delusions of Mediocrity	When this enchantment enters, you gain 10 life...	[embedding vector]
3	852d2a7f-5082-4668-978d-51776cad2cfd	Jeska and Kamahl	Haste, first strike\n{T}: Jeska and Kamahl dea...	[embedding vector]
4	a9b9e798-debc-4f2c-9770-e16f670082f7	Koskun Falls	At the beginning of your upkeep, sacrifice thi...	[embedding vector]

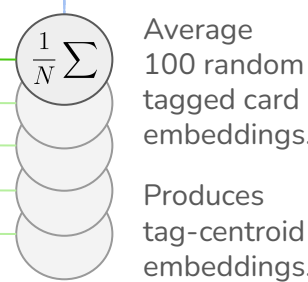


Tag embeddings (2,257 rows x 1026 columns)

	id	name	description	emb_0 ... emb_1023
0	827b8666-9058-457d-a732-f2373fa0b2c7	hate-color	Color hate for one or more colors. See also [h...	[embedding vector]
1	abdb5417-eee7-4ba2-a314-a75b08d8db71	freeze	A permanent doesn't untap during its next N un...	[embedding vector]
2	5a0a0531-5462-477d-add9-886148c4413f	card-names	A collection of tags identifying properties an...	[embedding vector]
3	22728b1f-f880-44d6-a6ef-66f2784c22ed	helper-card	Cards printed to assist you with using mechanics.	[embedding vector]
4	0641a74c-4dd5-426d-be58-2ab86d71995d	burn	Effects that deal damage, whether to creatures...	[embedding vector]

Tag Metadata

New Features



Data Cleaning

Tag Embeddings - Feature Selection

Tag ngrams (30,837 rows x 1027 columns)

	tag	ngram	bm25	gini
200	burn	creature deals damage equal	180.356371	0.450908
201	burn	control deals damage equal	146.810756	0.718176
202	burn	opponent controls deals damage equal	135.097132	0.682989
203	burn	creature deals damage divided	108.498427	0.640603
204	burn	controls deals damage equal	100.830561	0.629047
2517	removal	creature deals damage equal	172.399504	0.692925
2518	removal	control deals damage equal	161.396654	0.745258
2519	removal	destroy target artifact destroy target	149.770545	0.820185
2520	removal	destroy target attacking creature draw	140.049085	0.903502
2521	removal	target attacking creature draw card	136.032233	0.960930
1417	hate	create white spirit creature token	127.141643	0.959722
1418	hate	creature deals damage equal	111.067971	0.885240
1419	hate	creature fights target creature token	103.862515	0.843336
1420	hate	destroy target artifact destroy target	103.819382	0.932747
1421	hate	enters return target creature card	97.610903	0.860413



Lightning Bolt

name	description	category	similarity
burn	Effects that deal damage, whether to creatures, players, or planeswalkers.	True	0.759430
removal	Spot or limited removal. See also [sweeper] for removal effects that remove everything.	True	0.718648
hate	Cards that hate on things — colors, card types, zones, etc.	True	0.701495



Instant

Lightning Bolt deals 3 damage to target creature or player.

The sparkmage shrieked, calling on the rage of the storms of his youth. To his surprise, the sky responded with a fierce energy he'd never thought to see again.

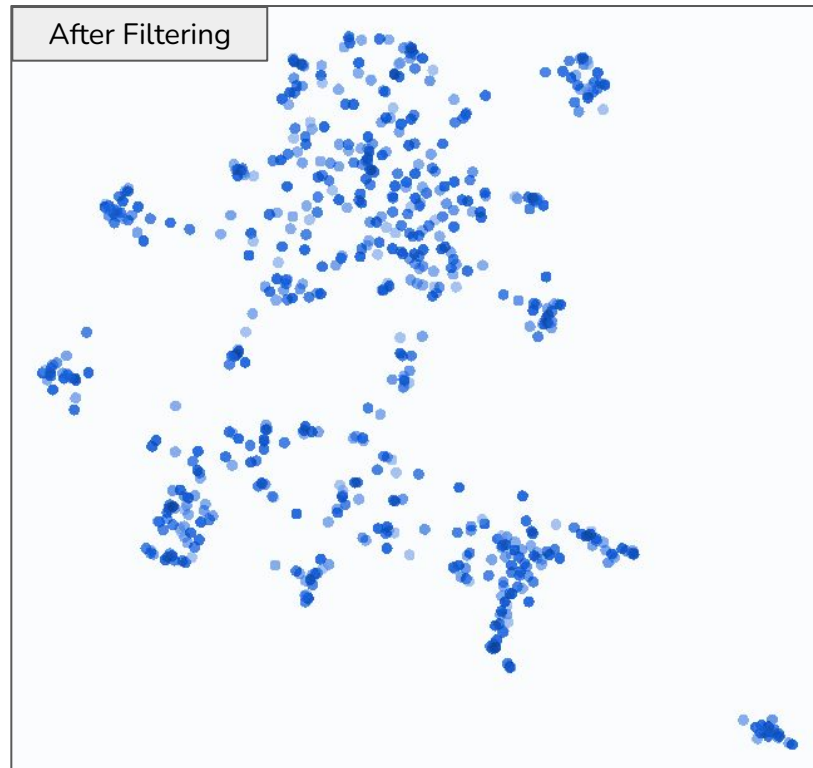
Christopher Moeller

™ & © 1993-2009 Wizards of the Coast LLC 146/240

Tag Embeddings - Data Filtering



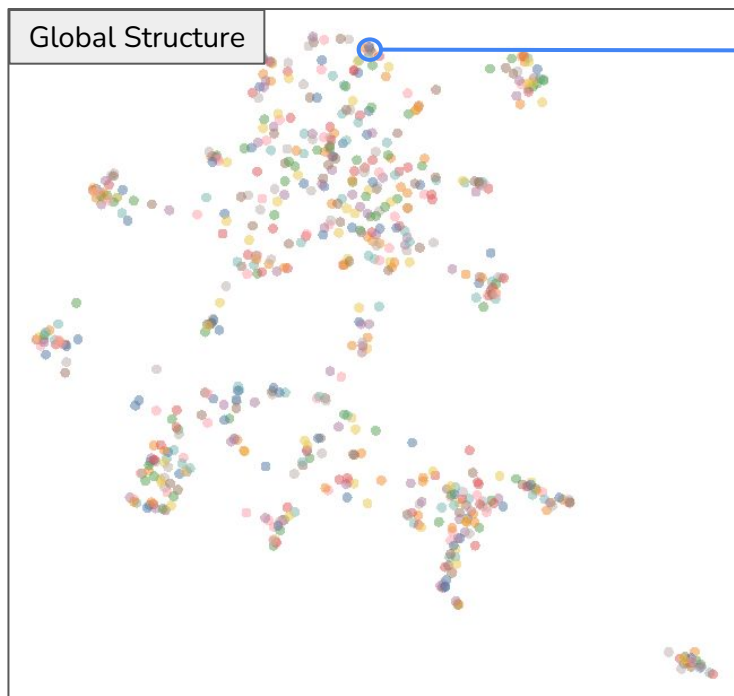
2,257 tags



644 tags

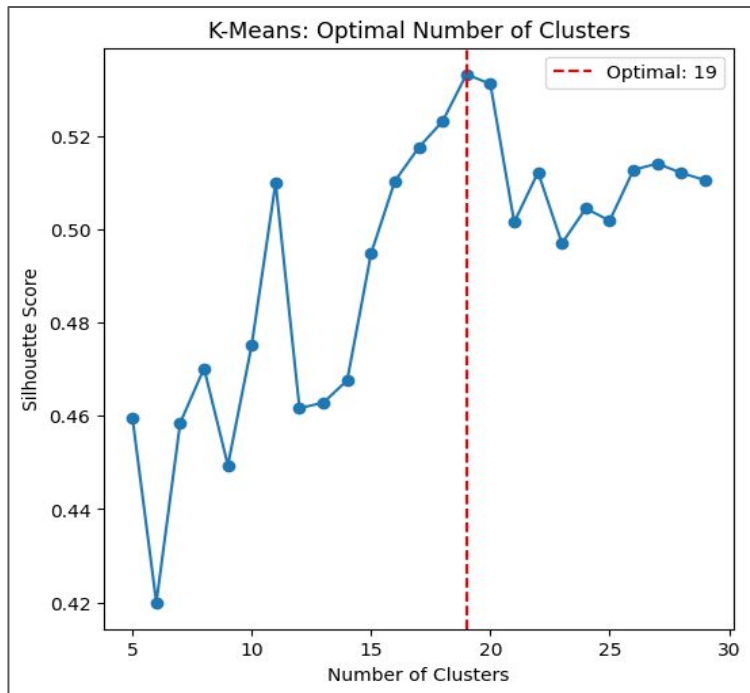
Dimensionality Reduction

Tag Embeddings - UMAP Projection

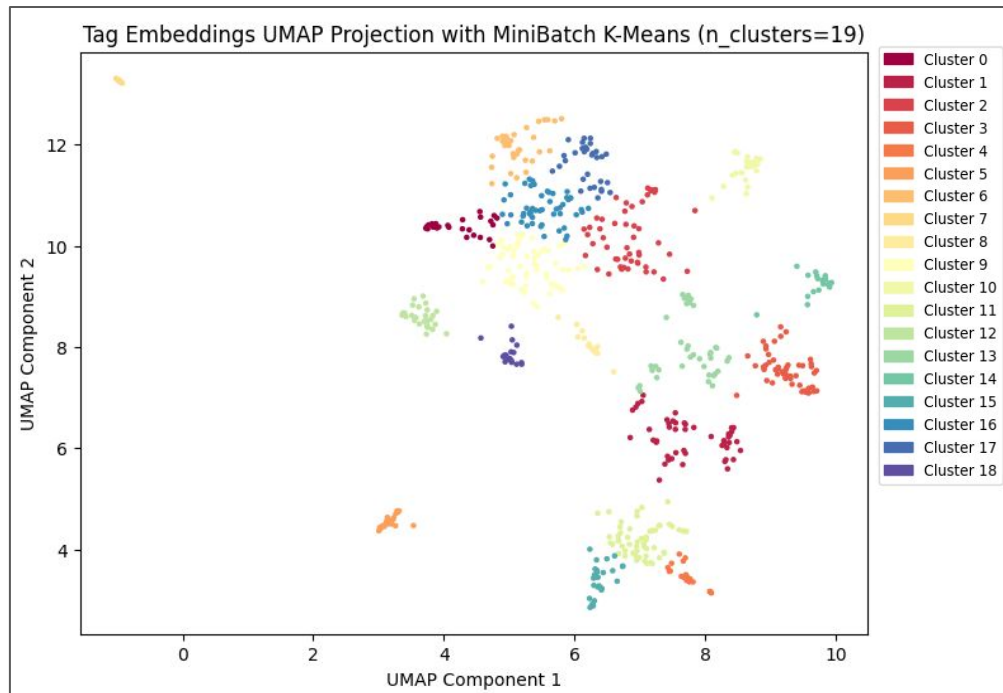


Clustering

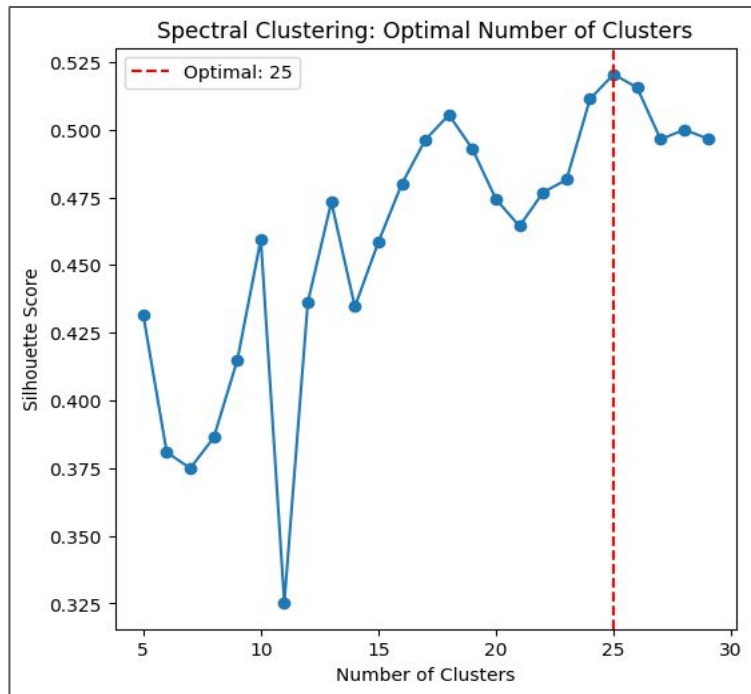
Mini-Batch KMeans Clusters



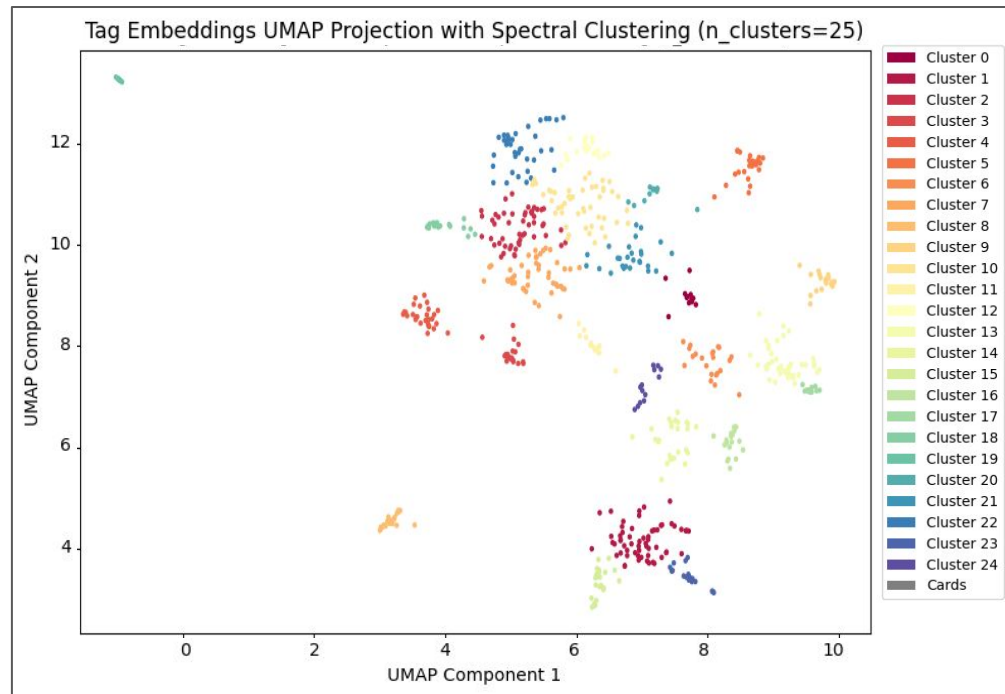
n_clusters=19



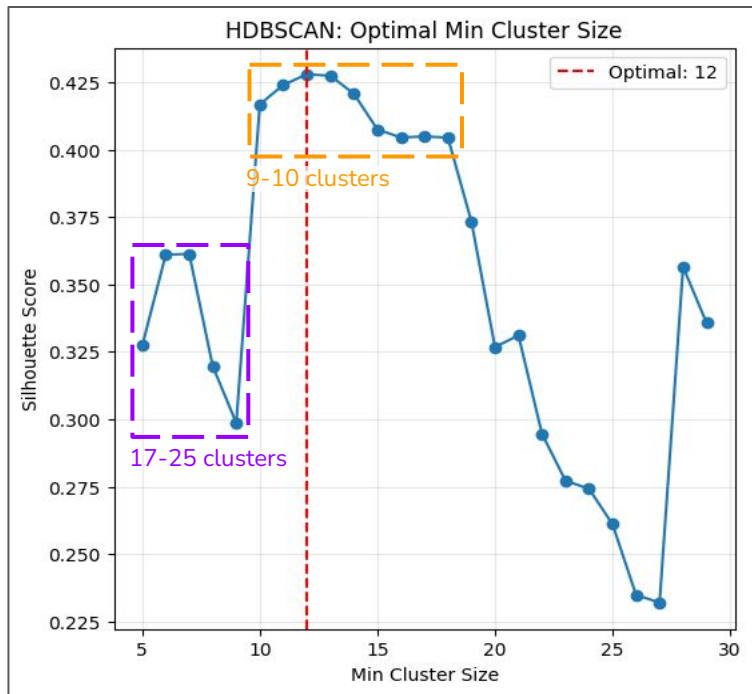
Spectral Clusters



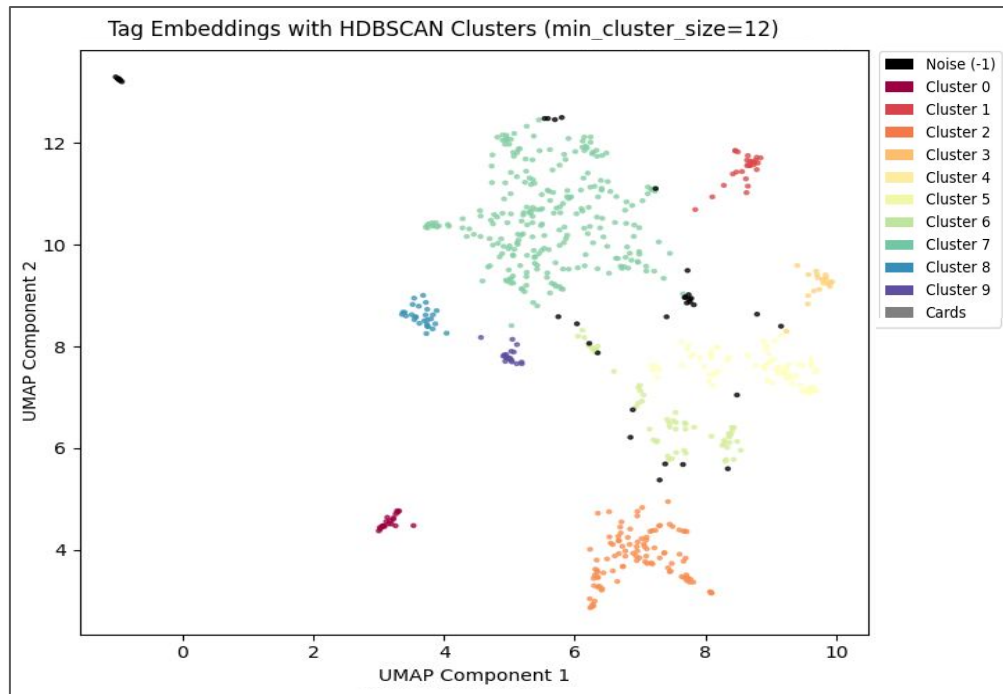
n_clusters=25



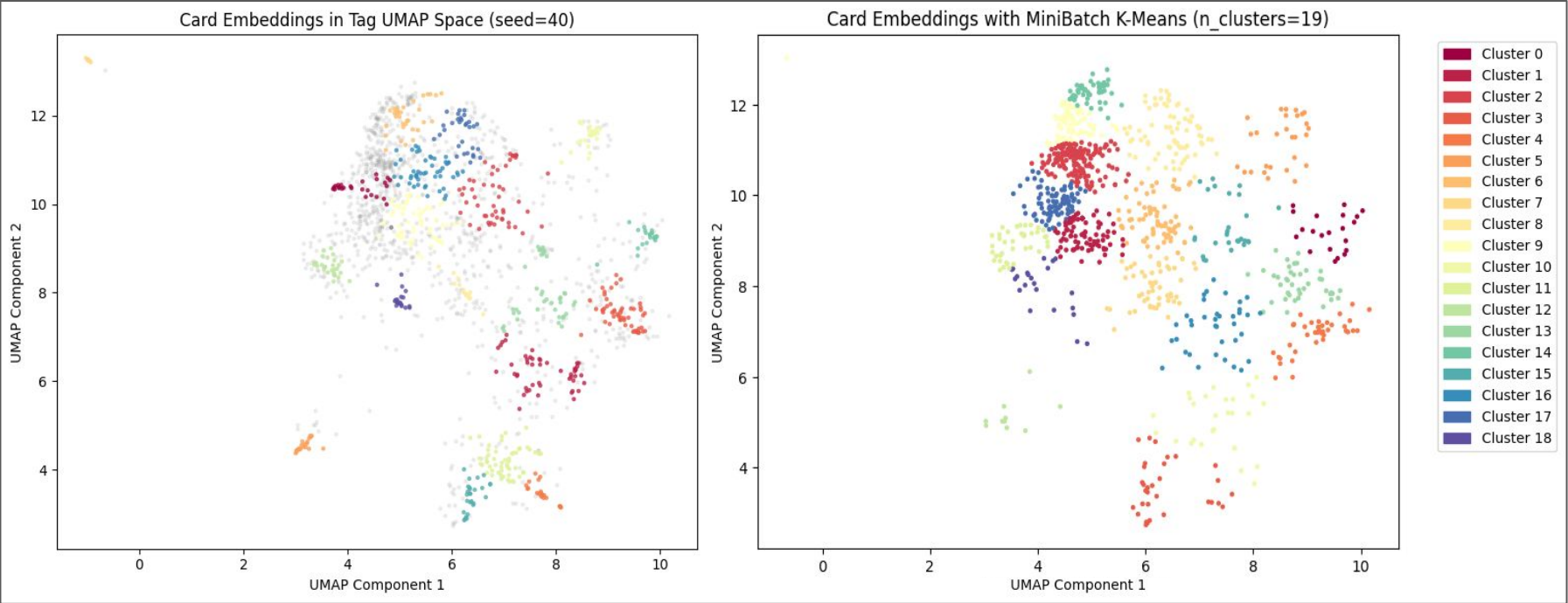
HDBSCAN Clusters



n_clusters=10

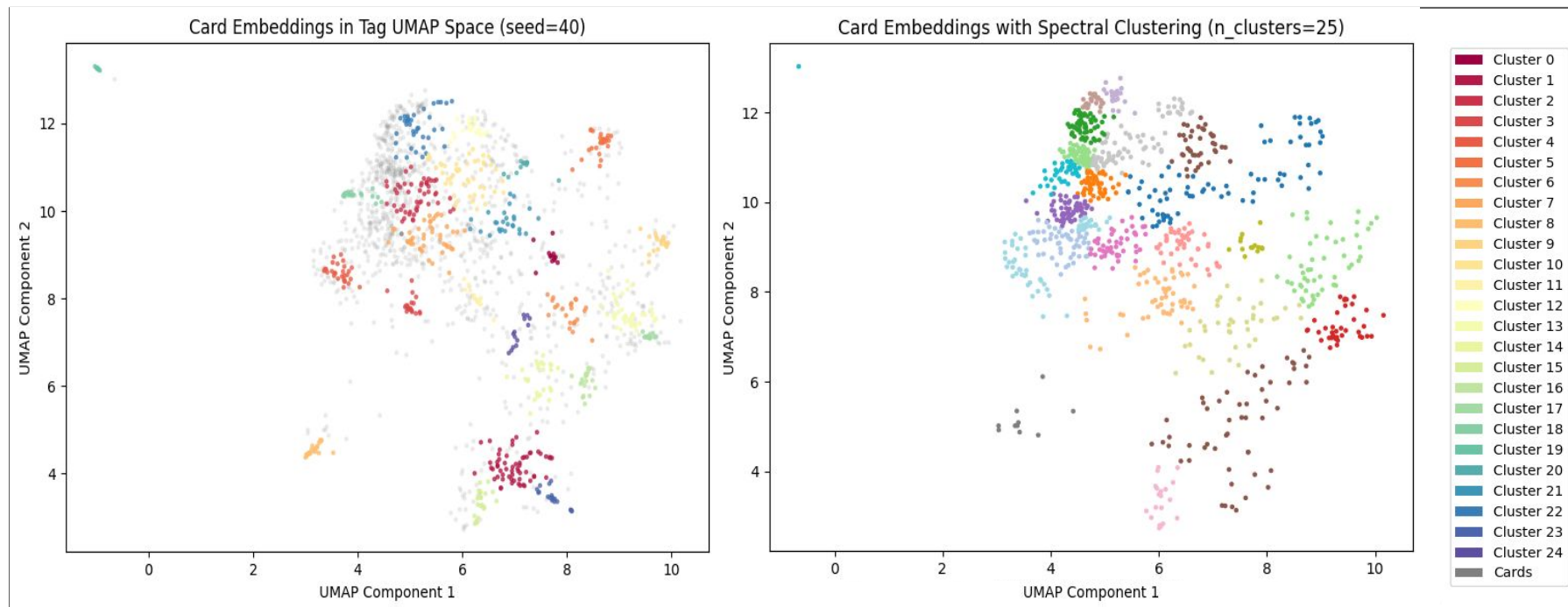


Mini-Batch KMeans Clusters - Prediction



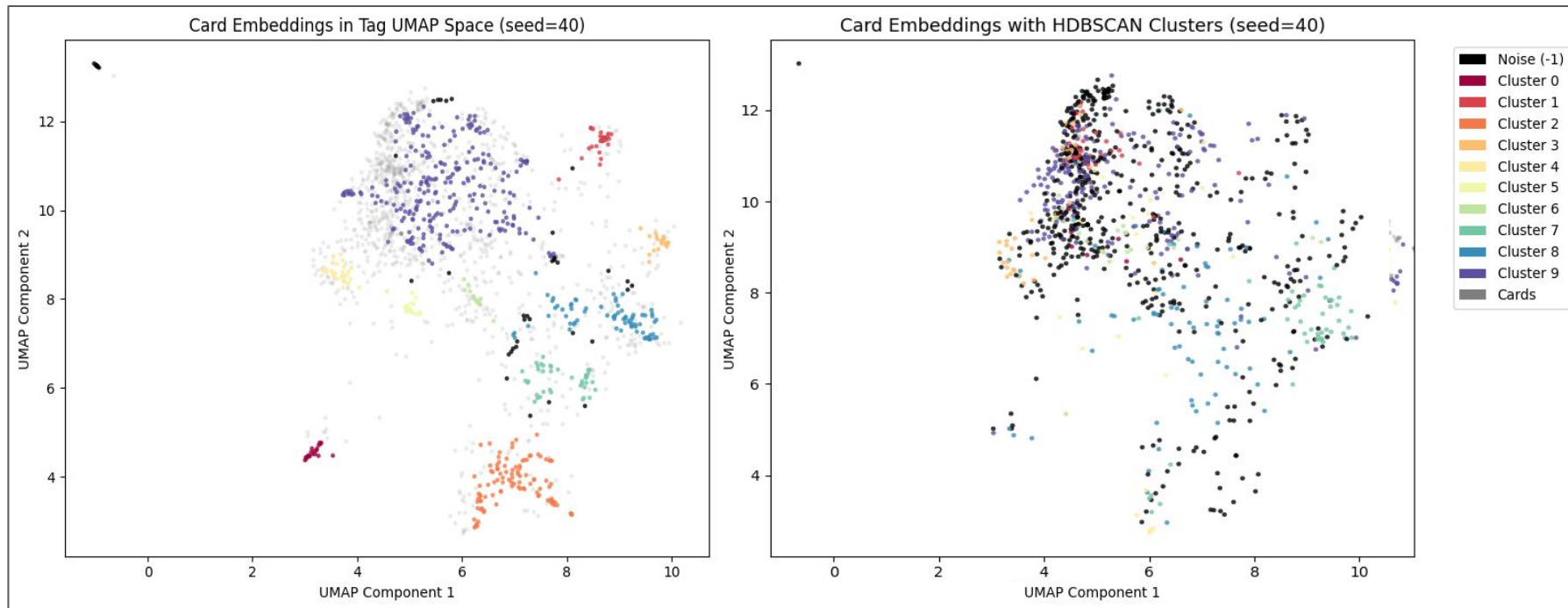
n_clusters=19

Spectral Clusters - Prediction



n_clusters=25

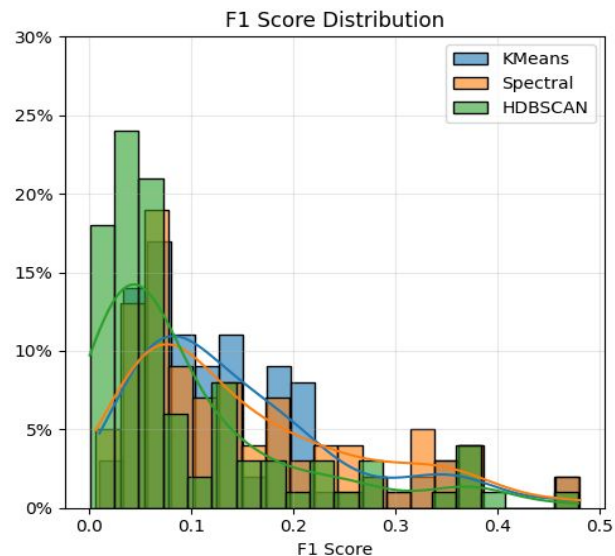
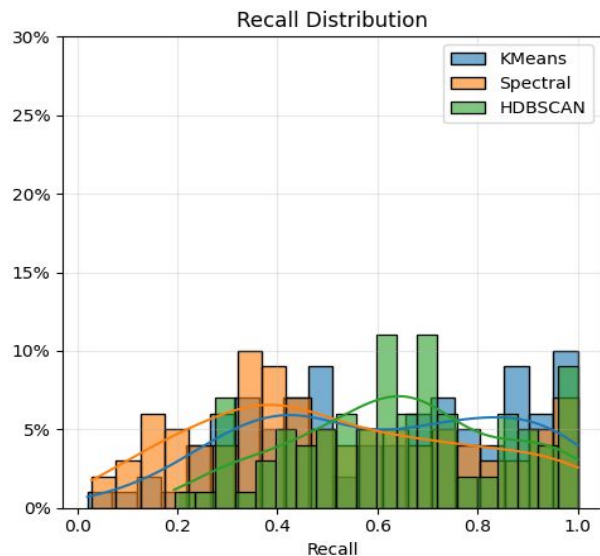
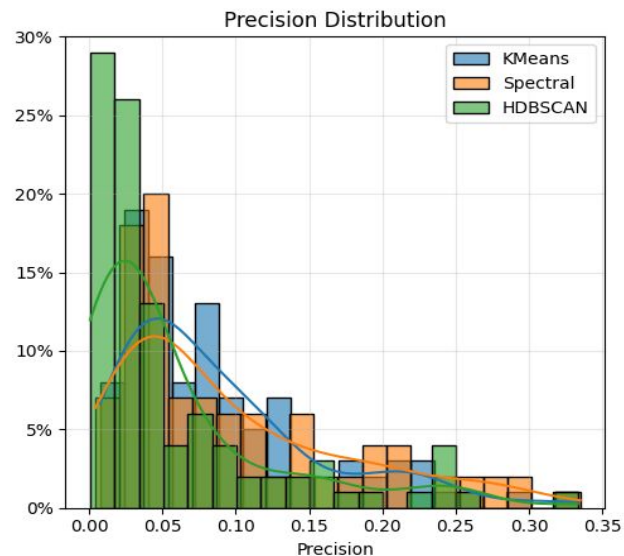
HDBSCAN Clusters - Prediction



n_clusters=10

Results

Clustering Results Comparison



tag	description	avg_precision	avg_recall	avg_f1	f1_kmeans	f1_spectral	f1_hdbscan	avg_auc
0	conjure-to-hand Cards that conjure cards to your hand.	0.333889	0.848101	0.479143	0.478571	0.480287	0.478571	0.912888
1	conjure-creature Cards that conjure creatures.	0.259138	0.732394	0.382823	0.382353	0.383764	0.382353	0.853798
2	banish-creature NaN	0.237134	0.785714	0.362812	0.347578	0.375839	0.365019	0.877931
3	counterspell-soft A "soft counterspell" gives the other player a...	0.245834	0.746528	0.356606	0.366255	0.323651	0.379913	0.854224
4	tutor-land-to-battlefield NaN	0.218170	0.781362	0.329177	0.354701	0.381295	0.251534	0.865540

Summary

Overall we found good diversity of features in the tag clusters with high coherency with the card embeddings.

Comparing the ground truth labels to our predictions, the clusters learned from our models are also very consistent semantically. We saw a large improvement in the dataset quality after filtering, which after inspecting these results visually, gave us good confidence into the quality of these tags.

We found that these tags effectively explore the concept space of the card dataset, and the clustering results showed that these tags are well enough separated to make accurate predictions.

References

Card and tag data provided courtesy of Scryfall and the Tagger project:

- <https://scryfall.com/docs/api/cards>
- <https://scryfall.com/docs/tagger-tags>

Embedding model provided by the Beijing Academy of Artificial Intelligence (BAAI) on Hugging Face:

- <https://huggingface.co/BAAI/bge-large-en-v1.5>

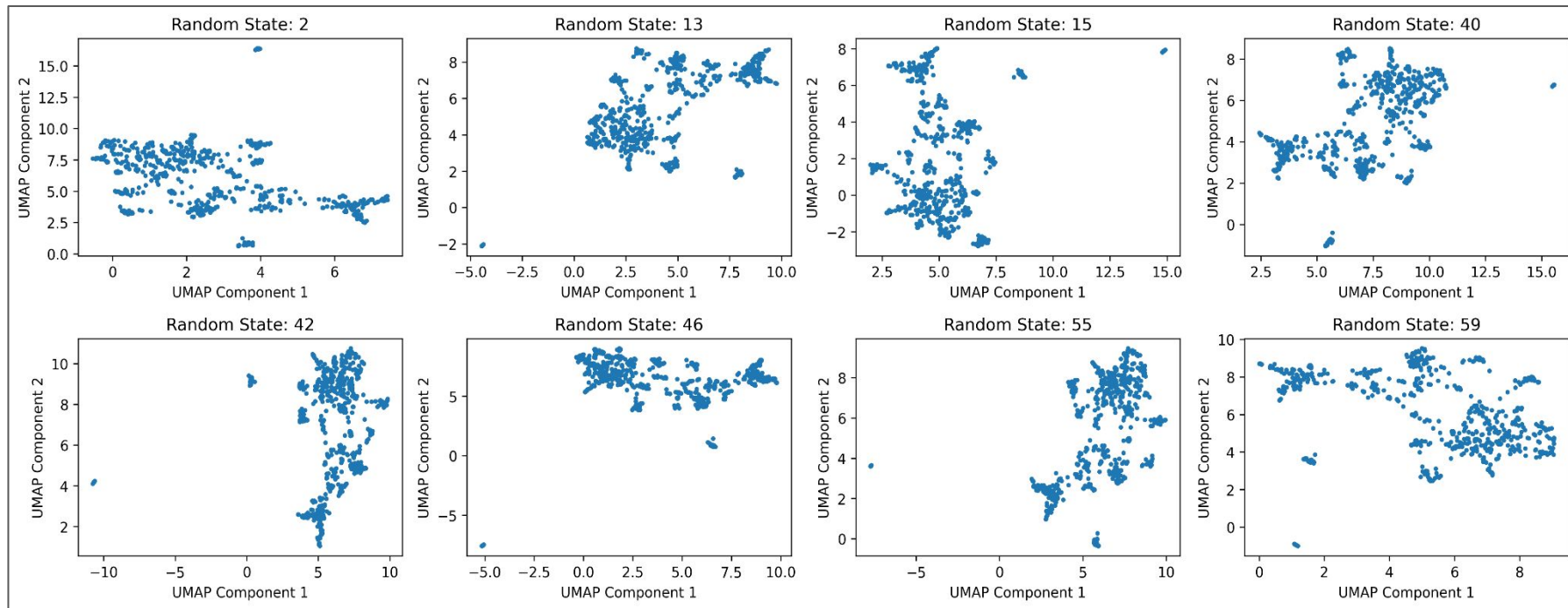
Libraries used:

- **pandas** for data manipulation and analysis.
- **numpy** for high-dimensional numerical computations.
- **matplotlib** and **seaborn** for data visualization.
- **sklearn** for clustering methods, metrics, and similarity functions.
- **hdbscan** for the HDBSCAN hierarchical clustering algorithm.
- **umap-learn** for UMAP dimensionality reduction.
- **emblaze** for interactive embedding visualization.

Model training and inference code available on GitHub:

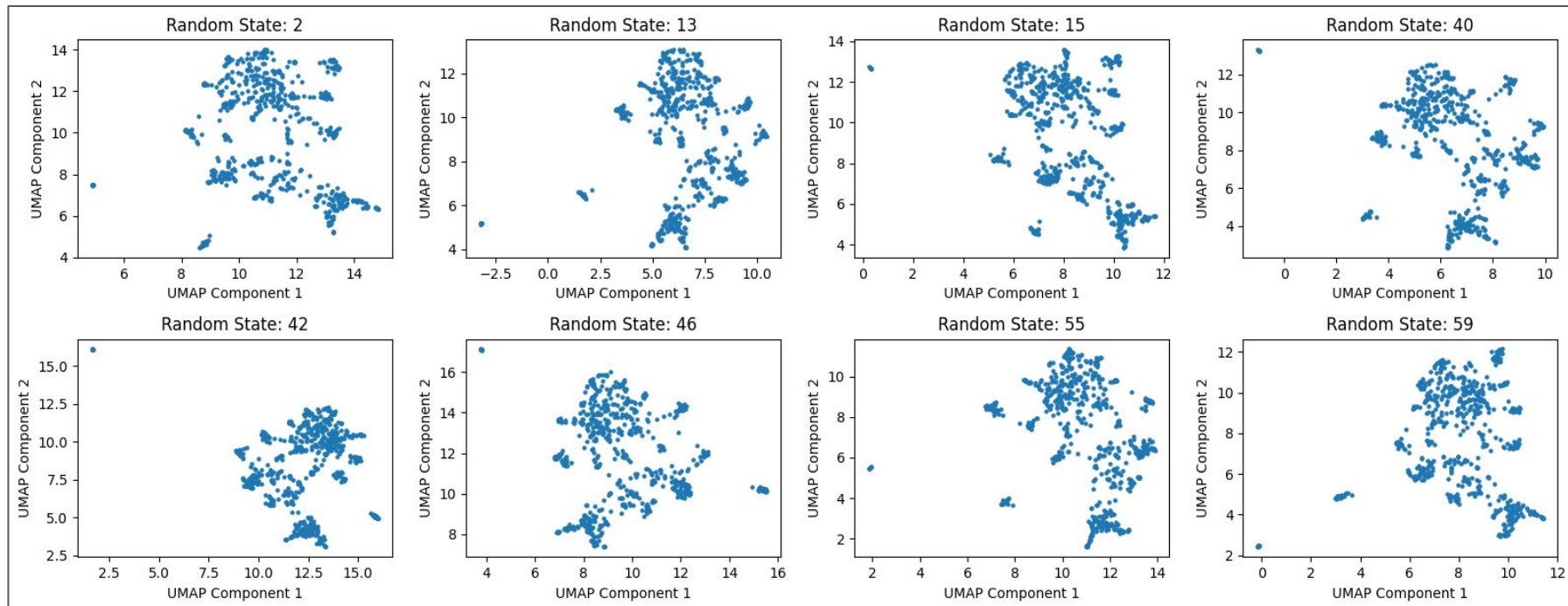
- <https://github.com/videre-project/mlm>

Tag Embeddings - UMAP Projection (Random Initialization)



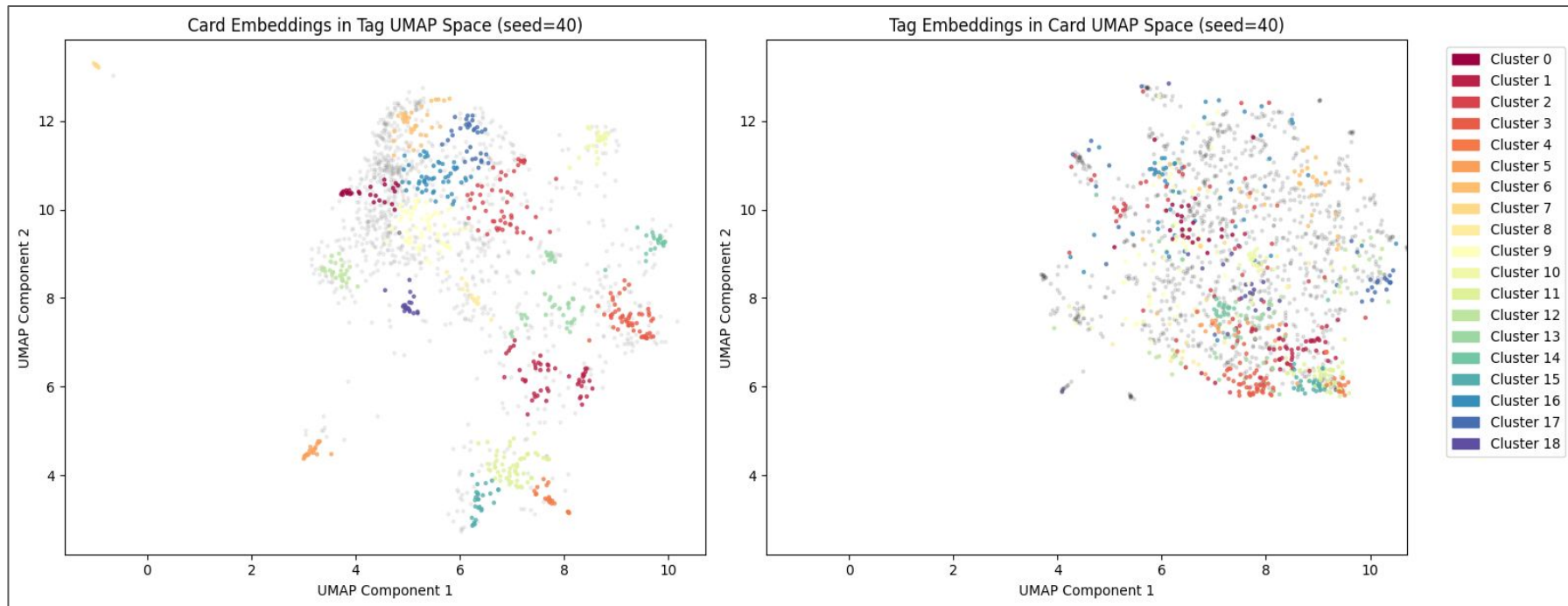
Using Cosine Distance Metric

Tag Embeddings - UMAP Projection (Spectral Initialization)



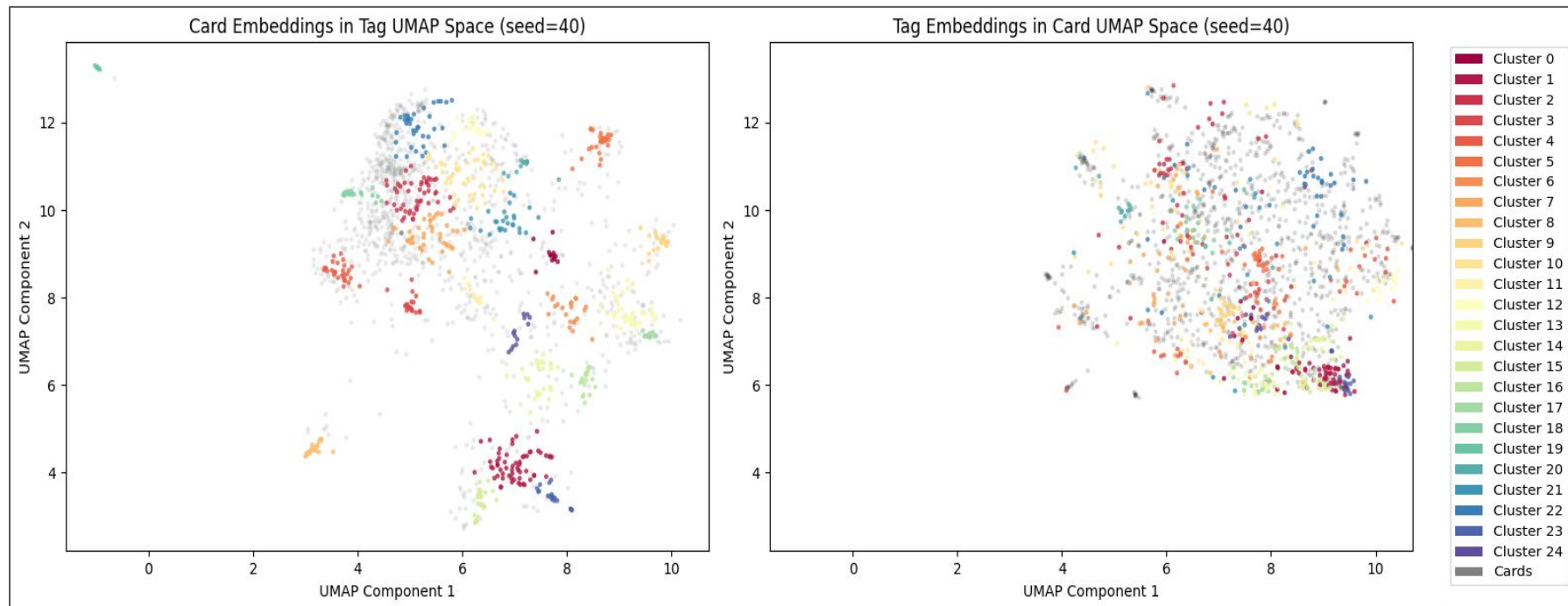
Using Cosine Distance Metric

Mini-Batch KMeans Clusters - Embedding Spaces



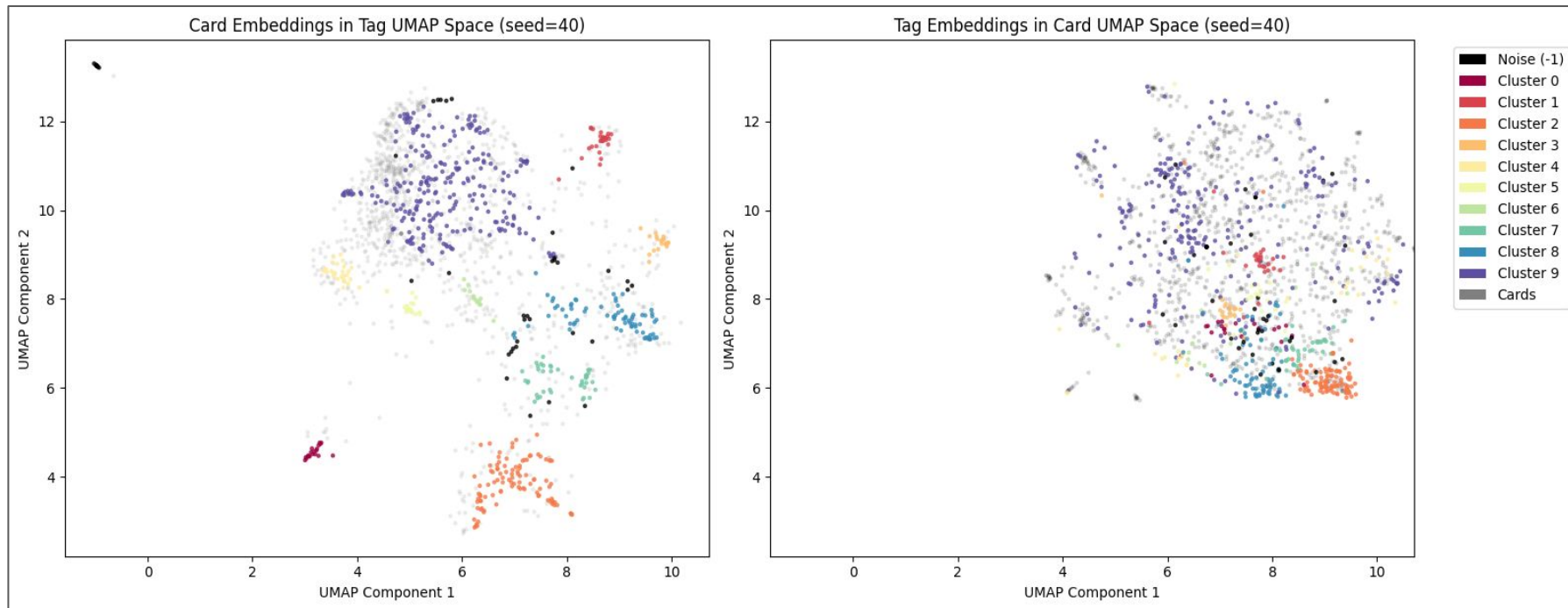
n_clusters=19

Spectral Clusters - Embedding Spaces



n_clusters=25

HDBSCAN Clusters - Embedding Spaces



min_cluster_size=12