A Poisson Multiplex Graph Model for Clustering Scientific Abstracts

Eve Fraczkiewicz, Alfonso Landeros

Department of Statistics, University of California, Riverside

UCR Research in Science & Engineering (RISE)

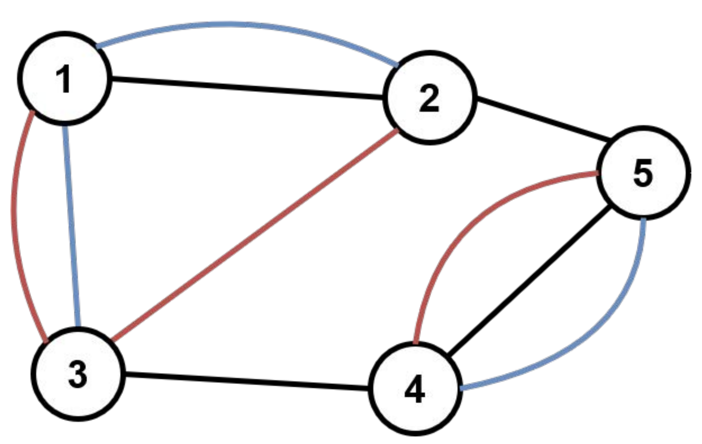
Abstract

- Generative AI models like ChatGPT are notorious for regurgitating false information, and as they become more sophisticated, it becomes increasingly difficult to differentiate between AI-generated and "human-generated" text.
- As more evidence of these models being used in academic writing comes to light, it is important to develop methods for detecting AI-generated text to protect academic integrity.
- We propose a Poisson multiplex graph model that clusters scientific abstracts into three groups: AI-generated, human-written, or ambiguous, by utilizing n-grams.
- AI-generated text, like human-written text, tends to have distinct features. By observing shared n-grams across and between generated and written text, we can find these unique features and use them to cluster abstracts into their respective groups.
- We tested our method on two datasets, the first being a collection of 28,000 abstracts from COVID-19 research papers, and the second being a collection of 12,000 abstracts of various topics scrapped from *Nature*.
- We utilized the Jaccard index and F-1 scores to assess the quality and accuracy of our model.

Background

Multiplex Graphs

- A mathematical model which contains nodes that are connected by edges.
- Multigraphs can have multiple edges between two nodes.
- Multiplex graphs are multigraphs that can have different types of edges.



n-grams

- A list of n-sized consecutive word groups from a piece of text.

Sentence Embeddings

- Word embeddings are special vectors that capture the semantic and contextual information of a word.
- To calculate sentence embeddings we compute the weighted average of all the word embeddings in a sentence:

$$v_s = \frac{1}{|s|} \sum_{w \in s} \frac{\alpha}{\alpha + p(w)} v_w$$

- And remove the projections of the weighted average vector on their first singular vector:

$$v_s = v_s - uu^T v_s$$

Methods

Collecting n-gram Data

- After generating n-grams for each abstract we calculate the sentence embeddings of each n-gram.
- In every abstract pair we calculate the cosine distance between each embedding and add to a count matrix if the distance is smaller than 0.1.

Poisson Clustering Model

- The resulting count matrix is a representation of our multiplex graph, each abstract is a node and their shared n-gram counts are the edges.
- To determine if the connections between nodes are statistically significant we assume the number of edges between nodes follows a Poisson distribution.
- We assign each node i a propensity p_i to form edges with other nodes and a cluster assignment c_i to account for the two different types of abstracts.
- To estimate the expected parameters we use the Poisson log-likelihood:

$$Poisson(c, p, R) = \sum_i \sum_{j \neq i} [x_{ij} \ln(r_{c_i c_j} p_i p_j) - r_{c_i c_j} p_i p_j - \ln(x_{ij}!)]$$

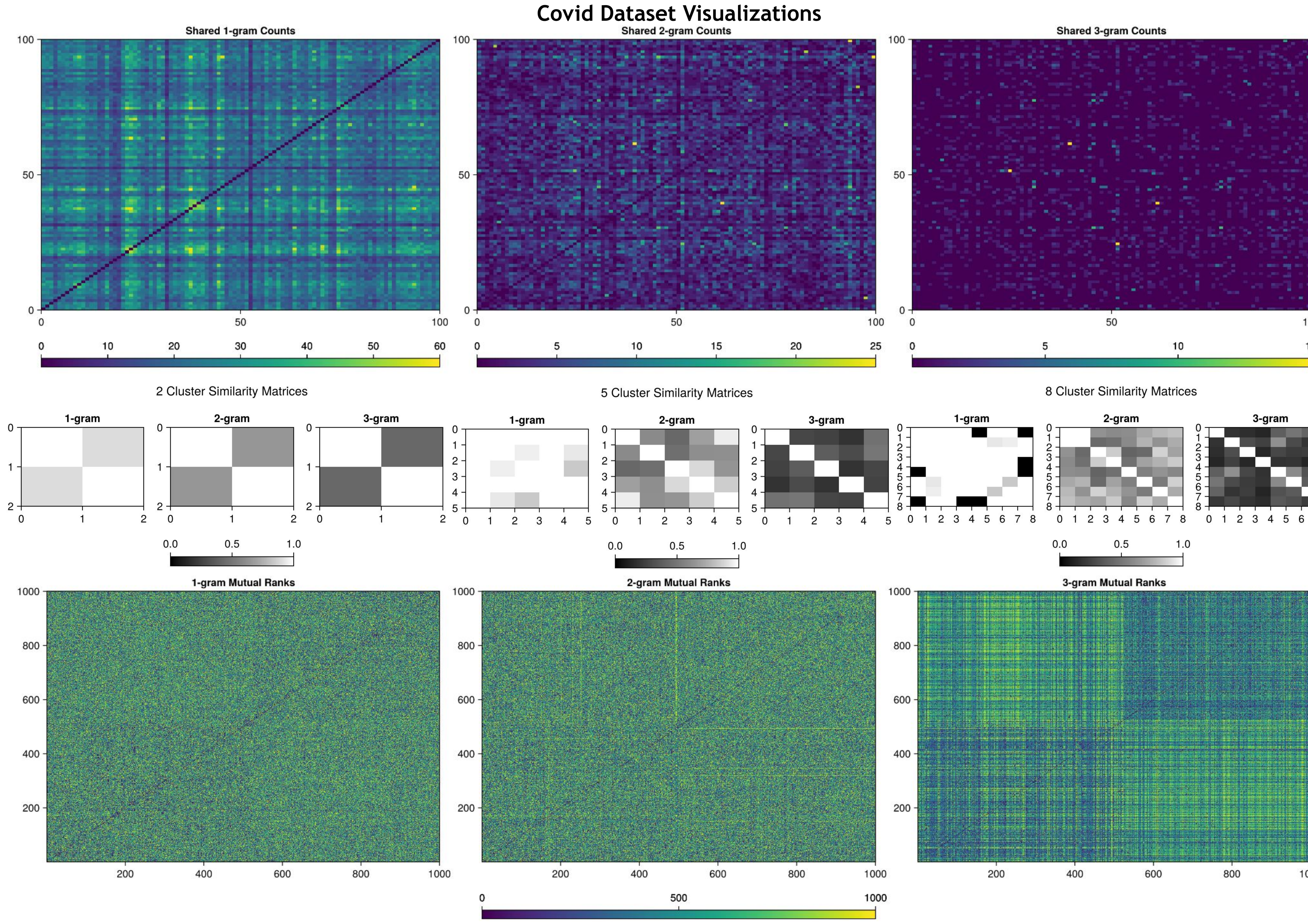
Mutual Ranking

- To calculate statistical significance we calculate the p-values of the residuals for each node pair and rank them for each node and all its node pairs.

$$R_{ij} = h(Z_{ij}) - [h(\hat{\mu}_{ij}) + \tfrac{1}{2} h''(\hat{\mu}_{ij}) \hat{\mu}_{ij}]$$

Results

Covid Dataset Visualizations



Covid Dataset Performance

2 Cluster Jaccard Index and F-1 Score

Row	n_gram	cluster_label	ground_truth	j_score
	String	String	String	Float64
1	1-gram	human	human	0.567839
2	1-gram	human	machine	0.0873309
3	1-gram	machine	human	0.201292
4	1-gram	machine	machine	0.609682
5	2-gram	human	human	0.882246
6	2-gram	human	machine	0.0270552
7	2-gram	machine	human	0.0400411
8	2-gram	machine	machine	0.873294
9	3-gram	human	human	0.710824
10	3-gram	human	machine	0.101751
11	3-gram	machine	human	0.0948181
12	3-gram	machine	machine	0.680357

Row	n_gram	f1_score
	String	Float64
1	1-gram	0.757519
2	2-gram	0.932362
3	3-gram	0.809777

5 Cluster Jaccard Index and F-1 Score

Row	n_gram	cluster_label	ground_truth	j_score
	String	String	String	Float64
1	1-gram	ambiguous	human	0.287066
2	1-gram	ambiguous	machine	0.164634
3	1-gram	human	human	0.449912
4	1-gram	human	machine	0.0589041
5	1-gram	machine	human	0.103651
6	1-gram	machine	machine	0.574733
7	2-gram	human	human	0.842756
8	2-gram	human	machine	0.042061
9	2-gram	machine	human	0.0510417
10	2-gram	machine	machine	0.829828
11	3-gram	human	human	0.667758
12	3-gram	human	machine	0.0963719
13	3-gram	machine	human	0.128962
14	3-gram	machine	machine	0.657095

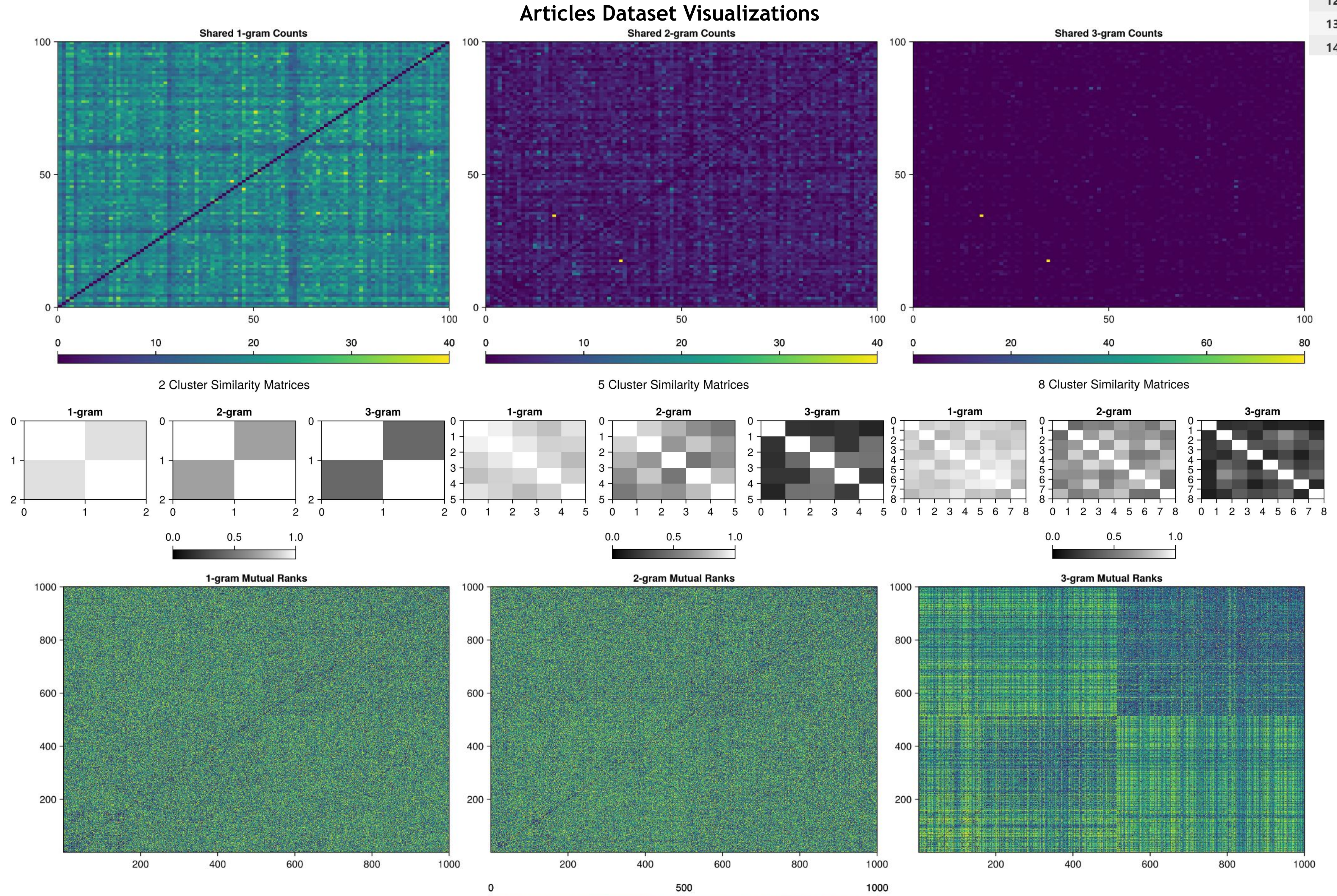
Row	n_gram	f1_score
	String	Float64
1	1-gram	0.831403
2	2-gram	0.907001
3	3-gram	0.793068

8 Cluster Jaccard Index and F-1 Score

Row	n_gram	cluster_label	ground_truth	j_score
	String	String	String	Float64
1	1-gram	ambiguous	human	0.299051
2	1-gram	ambiguous	machine	0.159879
3	1-gram	human	human	0.422339
4	1-gram	human	machine	0.0656425
5	1-gram	machine	human	0.112161
6	1-gram	machine	machine	0.564148
7	2-gram	human	human	0.87406
8	2-gram	human	machine	0.00638978
9	2-gram	machine	human	0.0613682
10	2-gram	machine	machine	0.874766
11	3-gram	ambiguous	human	0.150794
12	3-gram	ambiguous	machine	0.182777
13	3-gram	human	human	0.623746
14	3-gram	human	machine	0.0850059
15	3-gram	machine	human	0.0703883
16	3-gram	machine	machine	0.56015

Row	n_gram	f1_score
	String	Float64
1	1-gram	0.818878
2	2-gram	0.9332
3	3-gram	0.820937

Articles Dataset Visualizations



Articles Dataset Performance

2 Cluster Jaccard Index and F-1 Score

Row	n_gram	cluster_label	ground_truth	j_score
	String	String	String	Float64
1	1-gram	ambiguous	human	0.512
2	1-gram	ambiguous	machine	0.488
3	1-gram	human	human	0.0
4	1-gram	human	machine	0.0
5	1-gram	machine	human	0.0
6	1-gram	machine	machine	0.0
7	2-gram	human	human	0.876173
8	2-gram	human	machine	0.0219895
9	2-gram	machine	human	0.0459653
10	2-gram	machine	machine	0.876173
11	3-gram	human	human	0.729875
12	3-gram	human	machine	0.0524554
13	3-gram	machine	human	0.109129
14	3-gram	machine	machine	0.744932

Row	n_gram	f1_score
	String	Float64
1	1-gram	NaN
2	2-gram	0.934
3	3-gram	0.853824

5 Cluster Jaccard Index and F-1 Score

Row	n_gram	cluster_label	ground_truth	j_score
	String	String	String	Float64
1	1-gram	human	human	0.895753
2	1-gram	human	machine	0.00630252
3	1-gram	machine	human	0.0482897
4	1-gram	machine	machine	0.899254
5	2-gram	ambiguous	human	0.167213
6	2-gram	ambiguous	machine	0.166102
7	2-gram	human	human	0.758285
8	2-gram	human	machine	0.00114025
9	2-gram	machine	human	0.0233074
10	2-gram	machine	machine	0.764244
11	3-gram	human	human	0.722772
12	3-gram	human	machine	0.101512
13	3-gram	machine	human	0.0816777
14	3-gram	machine	machine	0.701068

Row	n_gram	f1_score
	String	Float64
1	1-gram	0.99
2	2-gram	0.972781
3	3-gram	0.860656

8 Cluster Jaccard Index and F-1 Score

Row	n_gram	cluster_label	ground_truth	j_score
	String	String	String	Float64
1	1-gram	human	human	0.906764
2	1-gram	human	machine	0.0355691
3	1-gram	machine	human	0.0165803
4	1-gram	machine	machine	0.89881
5	2-gram	ambiguous	human	0.047619
6	2-gram	ambiguous	machine	0.0661479
7	2-gram	human	human	0.847145
8	2-gram	human	machine	0.0327004
9	2-gram	machine	human	0.0278075
10	2-gram	machine	machine	0.822957
11	3-gram	ambiguous	human	0.139535
12	3-gram	ambiguous	machine	0.157343
13	3-gram	human	human	0.667254
14	3-gram	human	machine	0.0645905
15	3-gram	machine	human	0.057377
16	3-gram	machine	machine	0.636872

Row	n_gram	f1_score
	String	Float64
1	1-gram	0.946708
2	2-gram	0.936877
3	3-gram	0.86692

- We tested 2-8 different clusters to see if there was any clustering happening on a level beyond just written and generated text.
- While the Jaccard indices were the best for 2 clusters, the F-1 scores improved with higher order clusters.
- This suggests that globally we are not accurately clustering written and generated abstracts, but our within-cluster groupings are good.

Conclusion

Project Goal

- We successfully built an accessible model that directly accounts for n-gram statistics and is able to achieve reasonable clustering into generated and written abstracts.

Reflection

- We initially intended on only collecting raw shared n-gram counts and using a purely propensity-based Poisson model, but we shifted to sentence embeddings and clustering to better reflect our data.

Future Directions

- In the future, we anticipate improving our initial clustering and investigating any possible hierarchical clustering.
- We will also experiment with using adjective information from the abstracts to improve our model.
- Our end goal is to use our insights to pool information across layers in the multiplex graph to guide the design of a classification model.

References

Ranola, John M., et al. "A Poisson model for random multigraphs." *Bioinformatics*, vol. 26, no. 16, 16 June 2010, pp. 2004-2011, https://doi.org/10.1093/bioinformatics/btq309.

Ranola, John Michael et al. "Cluster and propensity based approximation of a network." *BMC systems biology* vol. 7 21. 14 Mar. 2013, doi:10.1186/1752-0509-7-21

Arora, Sanjeev, et al. "A Simple but Tough-to-Beat Baseline for Sentence Embeddings", 5 May 2023, https://openreview.net/group?id=ICLR.cc/2017/conference.