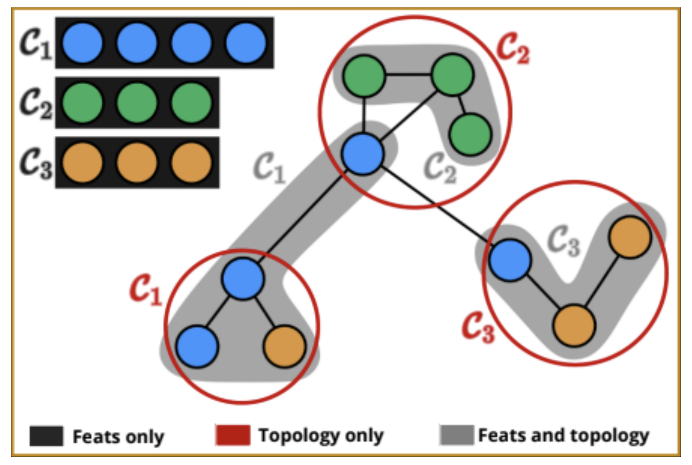
Cluster-Aware Graph Summaries through Graph Matching Neural Networks

Achilleas Tsimichodimos, Angeliki Dimitriou, Nikolaos Chaidos, Giorgos Stamou

National Technical University of Athens

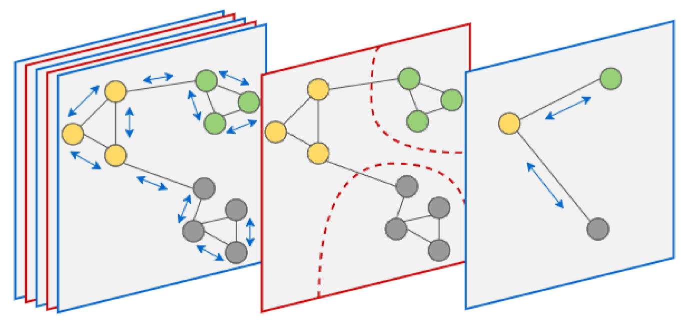
Embeddings-Based Clustering

- Models generate embeddings:
  - Based only on adjacency matrix: Spectral Clustering, Node2Vec
  - Based on adjacency matrix and node features: VGAE
- Embeddings clustered with k-means



Clustering with GNNs

- Cluster nodes of an attributed graph
- GNNs accounts both for topology and node features



Losses:

MinCutPool:

$$-\frac{\mathrm{Tr}\,(S^T\tilde{A}S)}{\mathrm{Tr}\,(S^T\tilde{D}S)} + \left\| \frac{S^TS}{\|S^TS\|_F} - \frac{I_K}{\sqrt{K}} \right\|_F$$

JustBalance:

$$-\mathrm{Tr}(\sqrt{S^TS})$$

DMoN:

$$\frac{\mathrm{Tr}\,(S^T\tilde{A}S)}{2E} - \frac{\sqrt{K}}{N}\left\| \sum_i S_i^T \right\|_F - 1$$

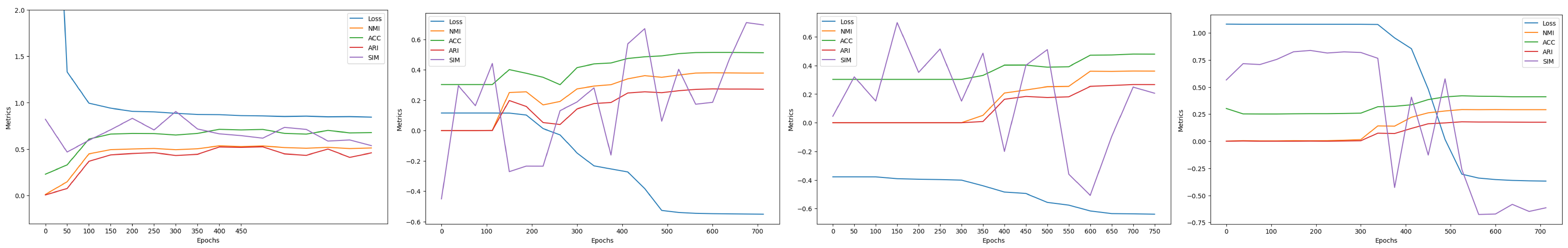
Results

Name	Nodes	Edges	Features	Classes
Cora	2708	10 556	1433	7
Citeseer	3327	9104	3703	6
PubMed	19 717	88 648	500	3

Datasets

	Cora				CiteSeer				PubMed			
	NMI	ACC	ARI	SIM	NMI	ACC	ARI	SIM	NMI	ACC	ARI	SIM
SC	0.041	0.291	−0.002	-	0.024	0.242	0.016	-	0.182	0.587	0.129	-
Node2vec	0.081	0.243	0.225	−0.231	0.06	0.197	0.02	−0.382	0.002	0.368	0.001	0.087
VGAE	0.535	0.712	0.523	0.665	0.262	0.437	0.136	0.558	0.284	0.665	0.271	0.591
MinCut	0.382	0.528	0.293	0.711	0.247	0.470	0.230	0.513	0.248	0.578	0.244	0.330
DMoN	0.292	0.419	0.179	−0.405	0.138	0.330	0.106	0.218	0.118	0.437	0.093	−0.408
JustBalance	0.361	0.479	0.266	0.253	0.162	0.394	0.112	0.259	0.202	0.570	0.160	0.518

Clustering Metrics



(a) VGAE (b) MinCutPool (c) JustBalance (d) DMoN

Evaluation Metrics on Cora

References

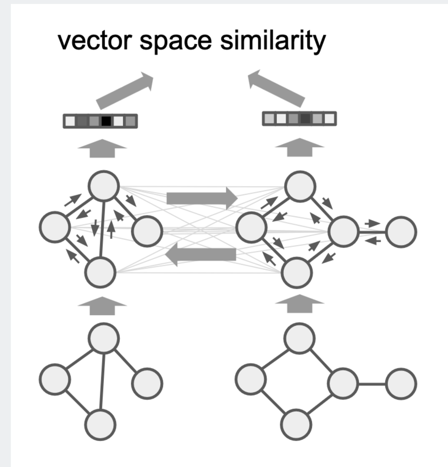
[1] Yujia Li et al. *Graph Matching Networks for Learning the Similarity of Graph Structured Objects*. 2019. arXiv: `1904.12787 [cs.LG]`.

[2] Filippo Maria Bianchi. "Simplifying Clustering with Graph Neural Networks". In: *Proceedings of the Northern Lights Deep Learning Workshop* 4 (Jan. 2023). ISSN: 2703-6928. DOI: `10.7557/18.6790`. URL: `http://dx.doi.org/10.7557/18.6790`.

[3] Filippo Maria Bianchi, Daniele Grattarola, and Cesare Alippi. *Spectral Clustering with Graph Neural Networks for Graph Pooling*. 2020. arXiv: `1907.00481 [cs.LG]`.

Graph Matching Networks

Graph Matching Networks (GMNs) are a specialized type of Graph Neural Networks that take a pair of graphs as an input and compute a similarity score between them.

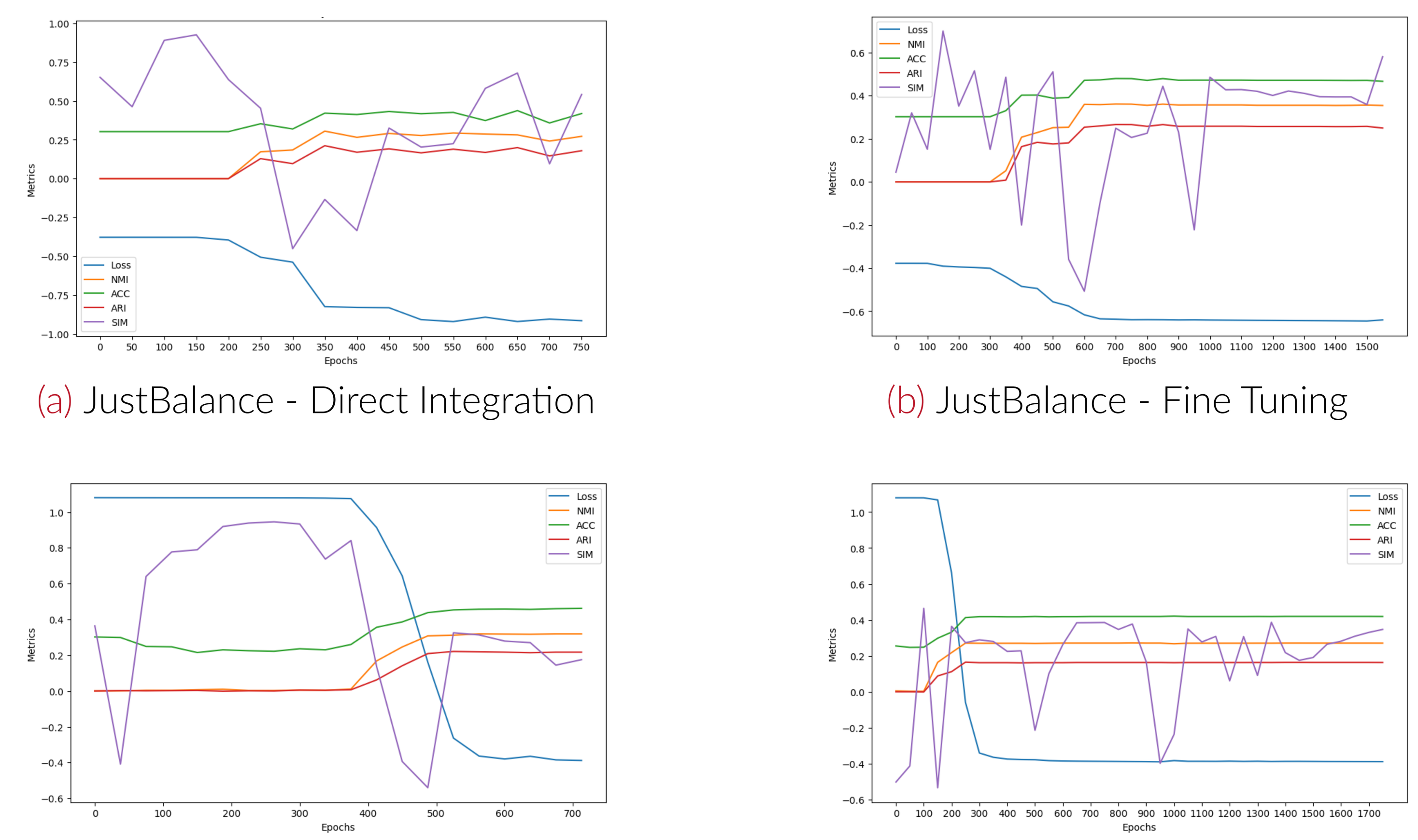


GMNs utilize a cross-graph matching vector in each propagation layer, which evaluates how well a node in one graph matches with nodes in another graph, taking into account both the aggregated messages on the edges and the cross-graph relationships.

Improving Graph Clustering with GMN

- **Direct Integration**: From the beginning of training, the similarity score provided by the GMN is combined with the pooling loss.
- **Fine-Tuning**: In this approach, the models are initially trained using only the pooling loss. After this initial training phase, the models are fine-tuned by adding the GMN-based loss to the existing pooling loss.

Impact of GMN in Graph Clustering



(a) JustBalance - Direct Integration (b) JustBalance - Fine Tuning

(c) DMoN - Direct Integration (d) DMoN - Fine Tuning

Conclusions

Integrating GMNs in the training process of graph clustering models leads to:

- similarity stabilization and improvements for all models
- traditional metric improvement for DMoN models