Subgraph Embeddings and Graph-level ML

I. Makarov & V. Pozdnyakov & D. Kiselev

BigData Academy MADE from Mail.ru Group

Graph Neural Networks and Applications

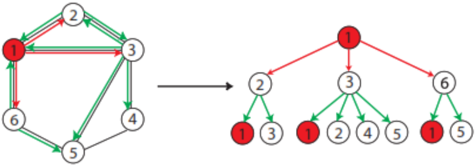
Topics

1. Hashing graphs
2. Deep Sets
3. Graph classification and clustering
4. Tricks for efficient graph pooling and beyond

Embedding Graphs

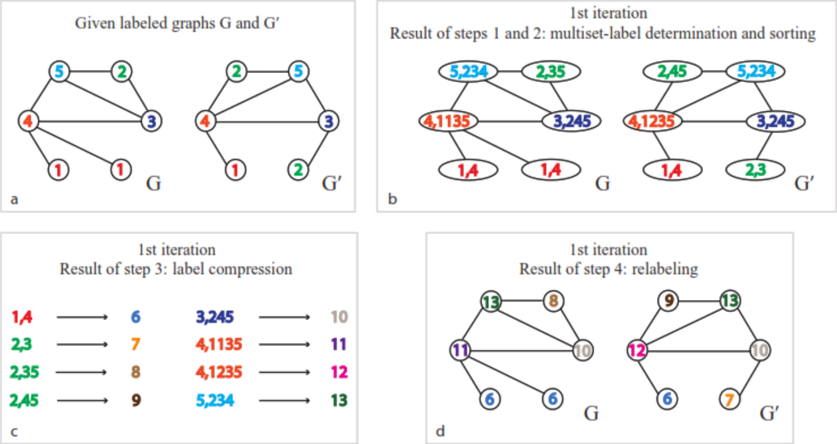
Weisfeiler-Lehman Graph Kernels

1. How to extract features for embedding graphs?
2. we need hashing and aggregation framework



from Borgwardt et al., 2011

Weisfeiler-Lehman Graph Kernels



from Borgwardt et al., 2011

Weisfeiler-Lehman Graph Kernels

1. Weisfeiler-Lehman subtree kernel with h = 1 for two graphs.
2. Compressed labels denote rooted subtree patterns.

End of the 1st iteration

Feature vector representations of G and G′

$$\varphi^{(1)}_{WLsubtree}(G) = (2, 1, 1, 1, 1, 2, 0, 1, 0, 1, 1, 0, 1)$$

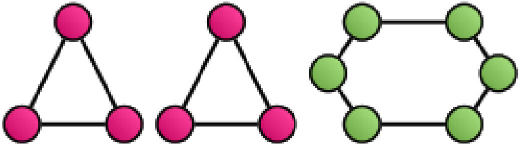
$$\varphi^{(1)}_{WLsubtree}(G') = (\underbrace{1, 2, 1, 1, 1}_{\text{Counts of original node labels}}, \underbrace{1, 1, 0, 1, 1, 0, 1, 1}_{\text{Counts of compressed node labels}})$$

$$k^{(1)}_{WLsubtree}(G,G') = \langle \varphi^{(1)}_{WLsubtree}(G), \varphi^{(1)}_{WLsubtree}(G') \rangle = 11.$$

e

Weisfeiler-Lehman Graph Kernels

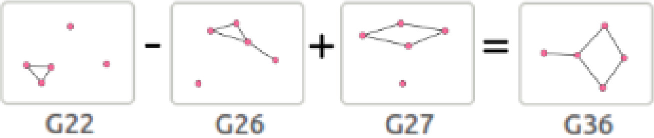
1. Why always there will be drawbacks?
2. Two graphs (pink and green) cannot be distinguished by the 1-WL.



from Kriege et al., 2021

Deep Graph Kernels

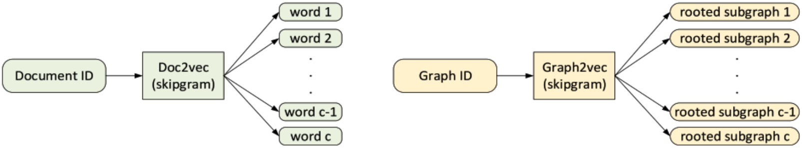
1. Let us use CBOW for small graphlets
2. Edit-based, WL, Shortest path kernels supported



from Vishwanathan et al., 2015

Graph2vec

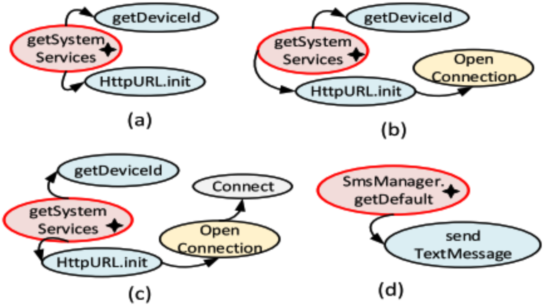
1. doc2vec's skipgram model takes a document d, samples c words from d and considers them as co-occurring in the same context d
2. graph2vec takes a graph G, samples c rooted subgraphs around different nodes and uses them analogous to doc2vec's context words
3. possible edge label support via Line graph (GL2vec)



from Jaiswal et al., 2017

Subgraph2vec

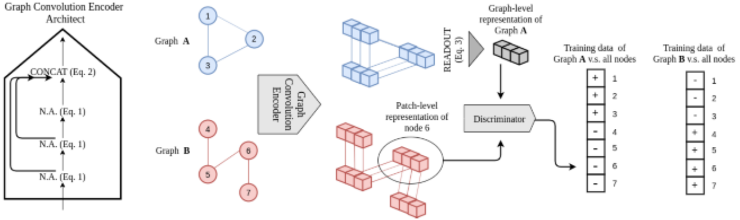
1. The rooted subgraphs where root nodes are marked with a star.
2. Graph (b) can be derived from (a) by adding a node and an edge. Graph (c) can be derived from (b) in a similar fashion.
3. Graph (d) is highly dissimilar from all the other graphs and is not readily derivable from any of them.



from Saminathan et al., 2016

InfoGraph

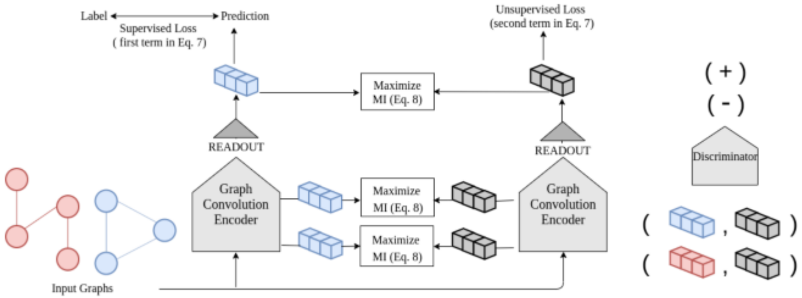
1. Concatenations with skip connections in GNN encoder
2. Discriminator takes a (global representation, patch representation) pair as input and decides whether they are from the same graph.



from Tang et al., 2020

InfoGraph

1. Two separate encoders with the same architecture, one for the supervised task and the other trained using both labeled and unlabeled data with an unsupervised objective.
2. Mutual Information of the two representations learned by the two encoders should be maximized



from Tang et al., 2020

Deep Sets

1. objective functions defined on sets that are invariant to permutations
2. main theorem characterizes the permutation invariant functions and provides a family of functions to which any permutation invariant objective function must belong
3. designing a deep network architecture that can operate on sets

Property 1 *A function $f : 2^{\mathfrak{X}} \rightarrow \mathcal{Y}$ acting on sets must be permutation* ***invariant*** *to the order of objects in the set,* i.e. *for any permutation* $\pi : f(\{x_1, \ldots, x_M\}) = f(\{x_{\pi(1)}, \ldots, x_{\pi(M)}\})$.

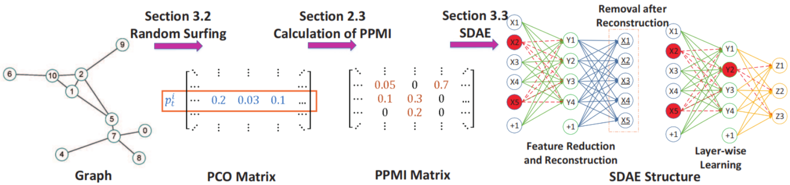
$$\mathbf{f}(\mathbf{x}) \doteq \boldsymbol{\sigma} \left(\lambda \mathbf{I} \mathbf{x} + \gamma \ \mathrm{maxpool}(\mathbf{x}) \mathbf{1}\right)$$

from Smola et al., 2017

Graph Classification

Deep Neural Networks for Learning Graph Representations (DNGR)

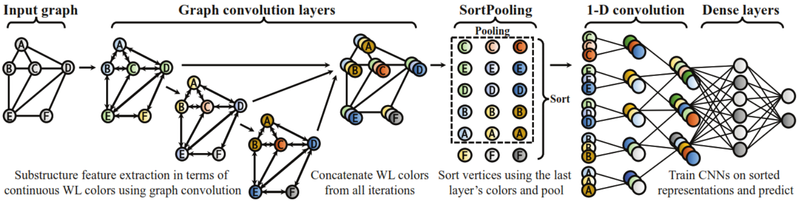
1. Random surfing
2. Calculation of PPMI matrix
3. Feature reduction by stacked denoising autoencoders (SDAE)



from Xu et al., 2016

Deep Graph Convolutional Neural Network (DGCNN)

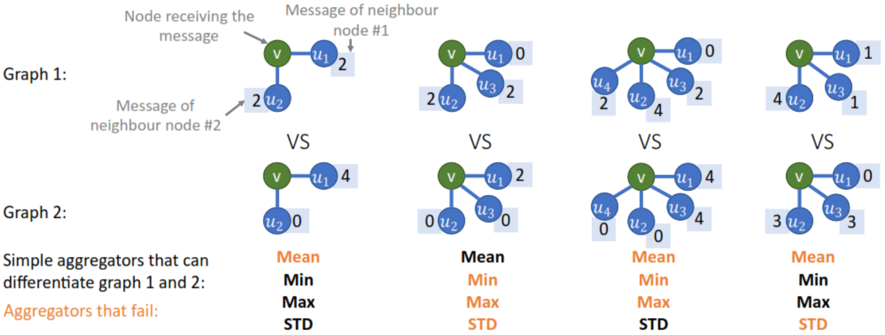
1. GCN layers are used to construct embeddings
2. Node features (colored) are sorted and pooled with a SortPooling layer, and passed to traditional CNN structures to learn a predictive model.



from Chen et al., 2016

Principal Neighbourhood Aggregation for Graph Nets

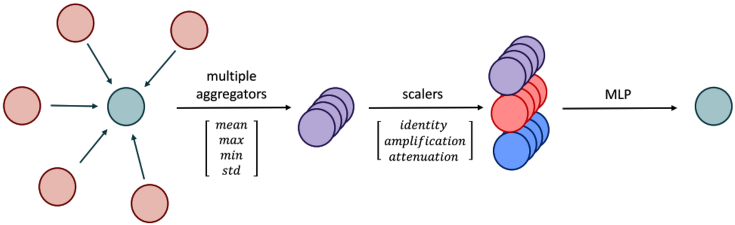
1. GCN layers may wrongly aggregate
2. Different schema alone do not solve the problem



from Veličković et al., 2020

Principal Neighbourhood Aggregation for Graph Nets

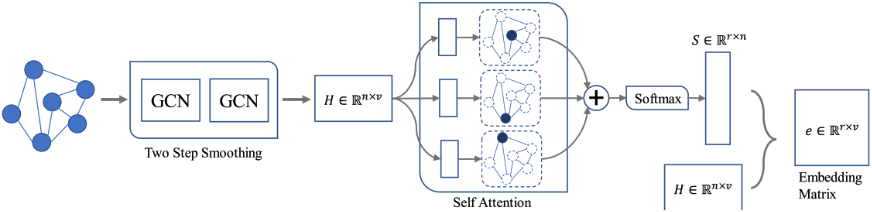
1. Use ensembles or attention via different aggregations with further MLP/pooling.

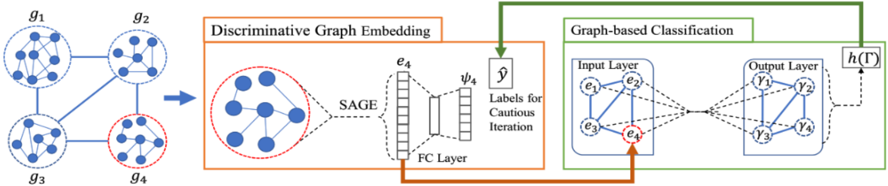


from Veličković et al., 2020

Semi-Supervised Graph Classification: A Hierarchical Graph Perspective

1. Learning subgraph node embeddings and aggregating for classes
2. Cautious/Active Iterations for coarsened graph model for subgraph classification

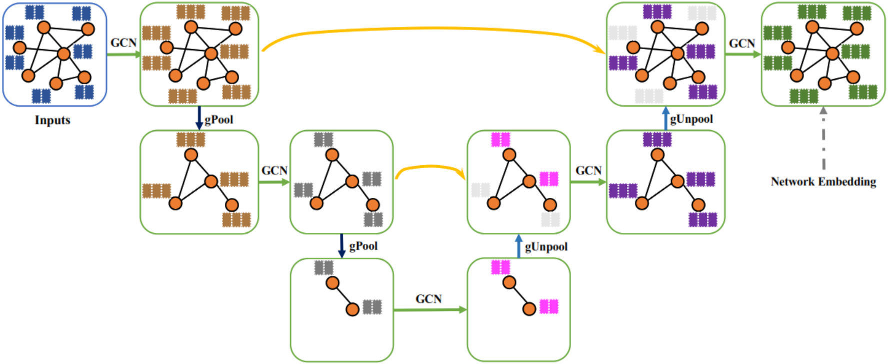




Various techniques for regularization

Graph U-Nets

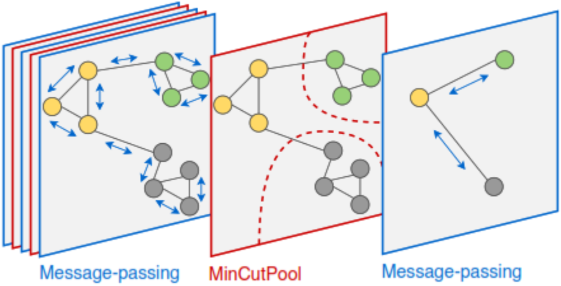
1. Residual autoencoders for graphs
2. Various feature augmentations may be added



from Ji et al., 2019

Spectral Clustering with Graph Neural Networks for Graph Pooling

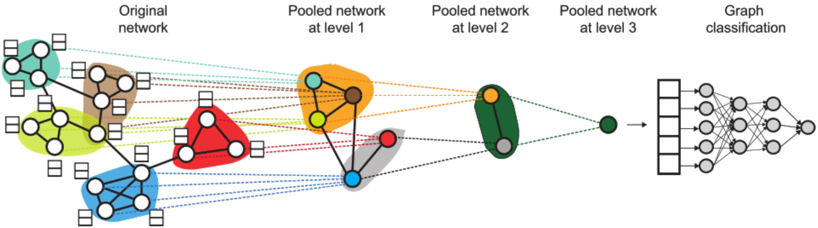
1. Aggregate, coarsen, propagate
2. Spectral clustering via normalized minCUT using GNN to compute cluster assignments minimizing cut value



from Alippi et al., 2019

Hierarchical Graph Representation Learning with Differentiable Pooling

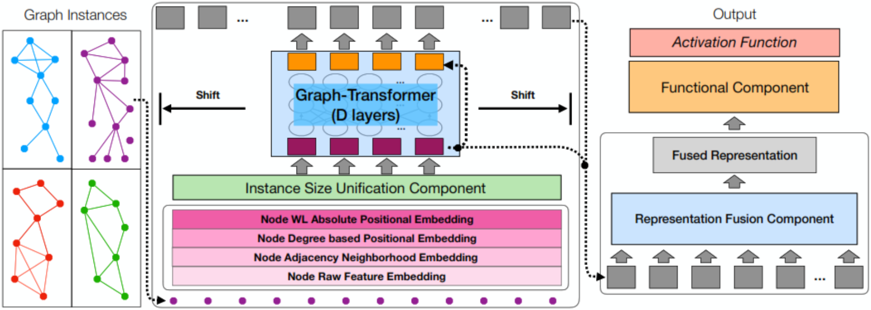
1. DiffPool for graph coarsening
2. Link prediction and hierarchical views for regularization



from Leskovec et al., 2018

Segmented GRAPH-BERT for Graph Instance Modeling (SSL)

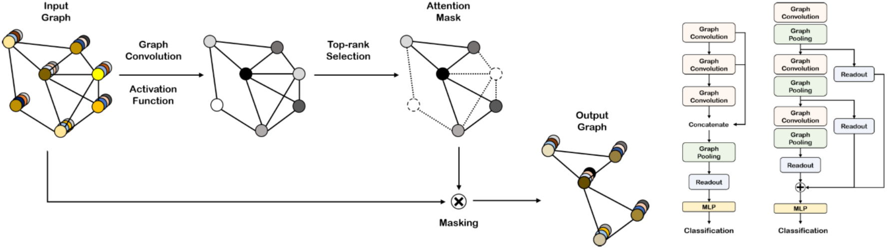
1. SSL for graph instance
2. Node equivariance



from Zhang et al., 2020

Self-Attention Graph Pooling

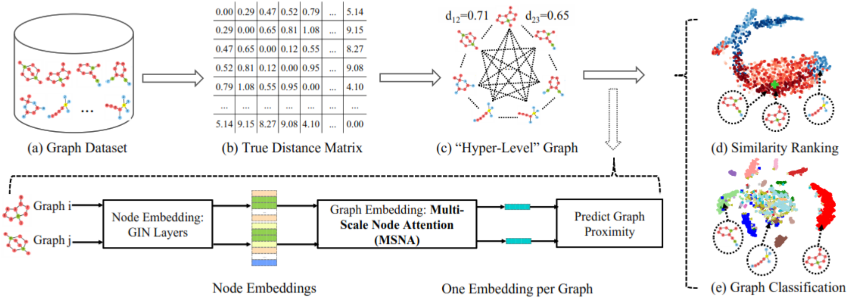
1. Self-attention masking for graph simplification
2. Various feature augmentations may be added



from Kang et al., 2019

Unsupervised Inductive Graph-Level Representation Learning via Graph-Graph Proximity

1. UGRAPHEMB computes the graph-graph proximity scores
2. “Hyper-level graph” trained for proximity preserving.
3. Similarity ranking based on a given query
4. Finetuning for graph classification.

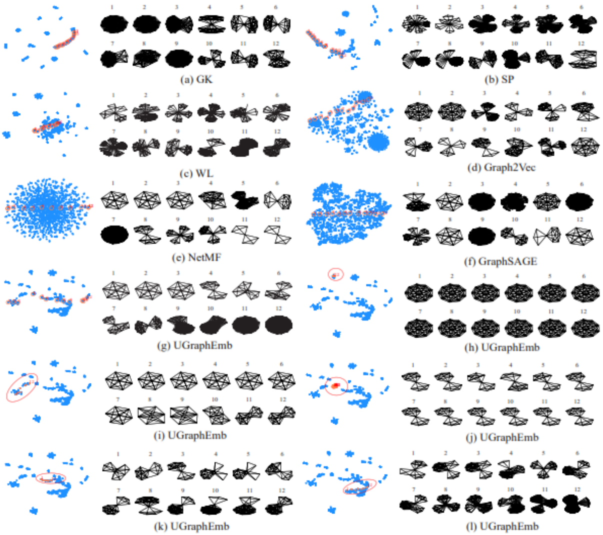


(a) Graph Dataset (b) True Distance Matrix (c) “Hyper-Level” Graph (d) Similarity Ranking (e) Graph Classification

from Wang et al., 2019

UGRAPHEMB

1. 12 graphs from IMDBM sampled from each cluster.



References

- Narayanan, A., Chandramohan, M., Venkatesan, R., Chen, L., Liu, Y. and Jaiswal, S., 2017. graph2vec: Learning distributed representations of graphs. arXiv preprint arXiv:1707.05005.
- Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R. and Smola, A., 2017. Deep sets. arXiv preprint arXiv:1703.06114.
- Narayanan, A., Chandramohan, M., Chen, L., Liu, Y. and Saminathan, S., 2016. subgraph2vec: Learning distributed representations of rooted sub-graphs from large graphs. arXiv preprint arXiv:1606.08928.
- Zhang, M., Cui, Z., Neumann, M. and Chen, Y., 2018, April. An end-to-end deep learning architecture for graph classification. In Thirty-Second AAAI Conference on Artificial Intelligence.

References

- Shervashidze, N., Schweitzer, P., Van Leeuwen, E.J., Mehlhorn, K. and Borgwardt, K.M., 2011. Weisfeiler-Lehman graph kernels. Journal of Machine Learning Research, 12(9).
- Morris, C., Fey, M. and Kriege, N.M., 2021. The Power of the Weisfeiler-Leman Algorithm for Machine Learning with Graphs. arXiv preprint arXiv:2105.05911.
- Cao, S., Lu, W. and Xu, Q., 2016, February. Deep neural networks for learning graph representations. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 30, No. 1).
- Yanardag, P. and Vishwanathan, S.V.N., 2015, August. Deep graph kernels. In Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1365-1374).

References

- Sun, F.Y., Hoffmann, J., Verma, V. and Tang, J., 2019. Infograph: Unsupervised and semi-supervised graph-level representation learning via mutual information maximization. arXiv preprint arXiv:1908.01000.
- Chen, H. and Koga, H., 2019, December. Gl2vec: Graph embedding enriched by line graphs with edge features. In International Conference on Neural Information Processing (pp. 3-14). Springer, Cham.
- Corso, G., Cavalleri, L., Beaini, D., Liò, P. and Veličković, P., 2020. Principal neighbourhood aggregation for graph nets. arXiv preprint arXiv:2004.05718.
- Zhang, J., 2020. Segmented GRAPH-BERT for Graph Instance Modeling. arXiv preprint arXiv:2002.03283.

References

- Ying, R., You, J., Morris, C., Ren, X., Hamilton, W.L. and Leskovec, J., 2018. Hierarchical graph representation learning with differentiable pooling. arXiv preprint arXiv:1806.08804.
- Lee, J., Lee, I. and Kang, J., 2019, May. Self-attention graph pooling. In International Conference on Machine Learning (pp. 3734-3743). PMLR.
- Gao, H. and Ji, S., 2019, May. Graph u-nets. In international conference on machine learning (pp. 2083-2092). PMLR.
- Bianchi, F.M., Grattarola, D. and Alippi, C., 2020, November. Spectral clustering with graph neural networks for graph pooling. In International Conference on Machine Learning (pp. 874-883). PMLR.

References

- Bai, Y., Ding, H., Qiao, Y., Marinovic, A., Gu, K., Chen, T., Sun, Y. and Wang, W., 2019. Unsupervised inductive graph-level representation learning via graph-graph proximity. arXiv preprint arXiv:1904.01098.
- Li, J., Rong, Y., Cheng, H., Meng, H., Huang, W. and Huang, J., 2019, May. Semi-supervised graph classification: A hierarchical graph perspective. In The World Wide Web Conference (pp. 972-982).