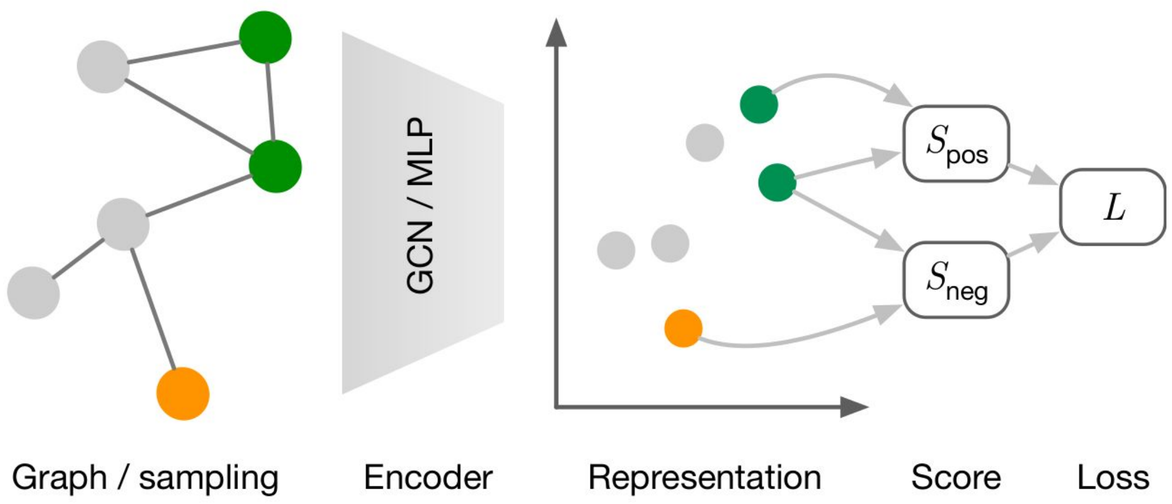
A Modular Framework for Graph Representation Learning

Daniel Daza & Thomas Kipf
University of Amsterdam

Graph Unsupervised Learning

- We are interested in unsupervised methods for learning representations of nodes in undirected graphs in the absence of labeled data.
- Representations can be used in downstream tasks like link prediction, node and graph classification, and for visualization purposes.
- Recent approaches have been proposed aiming to learn useful representations for graphs of different sizes and structure.



Framework Description

- We identify a set of fundamental components used throughout existing unsupervised learning methods:

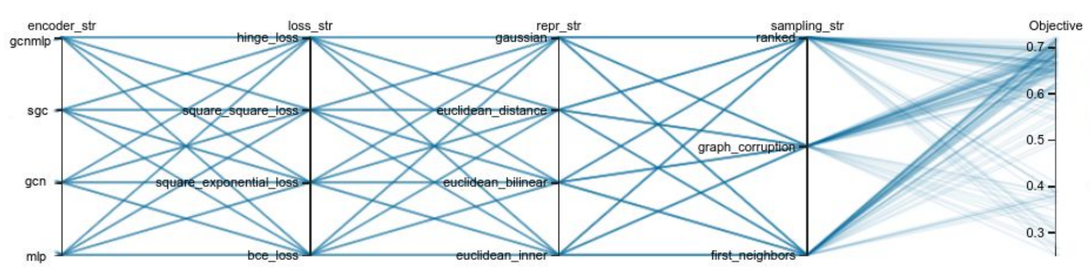
Method	Encoder	Representation	Score	Loss	Sampling
DeepWalk	LUT	$\mathbf{z}_i \in \mathbb{R}^D$	$\sigma(\mathbf{z}_i^\top \mathbf{z}_j)$	$-\log S - \log(1-\tilde{S})$	(+) random walk neighbors (-) non-neighbors
GAE	GCN	$\mathbf{z}_i \in \mathbb{R}^D$	$\sigma(\mathbf{z}_i^\top \mathbf{z}_j)$	$-\log S - \log(1-\tilde{S})$	(+) 1st order neighbors (-) non-neighbors
$\mathcal{S}$-VGAE	GCN	$\mathbf{z}_i \sim \text{vMF}(\mathbf{z})$	$\sigma(\mathbf{z}_i^\top \mathbf{z}_j)$	$-\log S - \log(1-\tilde{S})$	(+) 1st order neighbors (-) non-neighbors
DGI	GCN	$\mathbf{z}_i \in \mathbb{R}^D$	$\sigma\left(\mathbf{z}_i^T \mathbf{W} \mathbf{s}\right)$	$-\log S - \log(1-\tilde{S})$	(+) original graph (-) corrupted graph
G2G	MLP	$\mathbf{z}_\mu \in \mathbb{R}^D$ $\mathbf{z}_\Sigma \in \mathbb{R}^D$	$\exp(-\text{KL}(\mathcal{N}_i \| \mathcal{N}_j))$	$(\log S)^2 + \tilde{S}$	(+) 1st order neighbors (-) higher order neighbors

- **Graph encoder:** assigns a node a low dimensional representation. It can leverage graph structure.
- **Representation:** determines properties of the embedding: e.g. vector in a Euclidean space, Gaussian distribution, point clouds.
- **Score:** a symmetric function that gives a measure of similarity between nodes.
- **Loss:** most methods use a *contrastive learning* approach so that score is high for positive samples and low for negative samples.
- **Sampling:** a strategy to choose positive and negative samples for a given node in the graph.
- This description motivates a study of existing methods and potential variations of them.

Results

Method	Cora	Citeseer	Pubmed	Cora Full	Coauthor CS	Coauthor Physics
Raw features	76.8 ± 1.2	88.2 ± 0.9	89.6 ± 0.2	89.9 ± 0.2	92.3 ± 0.2	93.3 ± 0.1
DeepWalk	92.0 ± 0.6	91.3 ± 0.5	91.3 ± 0.3	96.4 ± 0.1	95.7 ± 0.2	96.1 ± 0.1
MLP-DGI	62.3 ± 2.2	69.1 ± 2.7	75.4 ± 0.9	66.8 ± 0.8	79.9 ± 0.8	80.9 ± 0.8
GCN-DGI	88.5 ± 1.2	91.7 ± 1.2	93.2 ± 0.4	89.5 ± 0.4	87.4 ± 1.3	86.7 ± 0.9
SGC-DGI	$\mathbf{92.4 \pm 0.7}$	$\mathbf{92.4 \pm 1.1}$	$\underline{\mathbf{96.2 \pm 0.3}}$	$\mathbf{96.5 \pm 0.2}$	$\mathbf{87.7 \pm 2.3}$	$\mathbf{92.9 \pm 0.5}$
GCN-DGI-B	93.2 ± 0.8	93.9 ± 0.7	92.8 ± 1.0	95.8 ± 0.4	94.1 ± 1.1	96.1 ± 0.5
MLP-GAE	79.9 ± 1.5	83.1 ± 1.5	87.5 ± 0.6	83.7 ± 0.7	94.1 ± 0.2	94.3 ± 0.1
GCN-GAE	91.9 ± 0.8	92.9 ± 0.9	94.9 ± 0.2	95.4 ± 0.2	96.2 ± 0.2	97.0 ± 0.1
SGC-GAE	$\underline{\mathbf{93.1 \pm 0.8}}$	$\underline{\mathbf{94.6 \pm 0.8}}$	$\mathbf{96.0 \pm 0.4}$	$\underline{\mathbf{96.5 \pm 0.2}}$	$\underline{\mathbf{96.7 \pm 0.1}}$	$\underline{\mathbf{97.2 \pm 0.1}}$
MLP-G2G	$\mathbf{92.7 \pm 0.8}$	$\mathbf{94.3 \pm 0.7}$	92.8 ± 0.3	97.7 ± 0.2	63.2 ± 0.6	71.7 ± 0.7
MLP-G2V	92.3 ± 0.8	93.9 ± 0.8	92.9 ± 0.3	$\underline{\mathbf{97.8 \pm 0.1}}$	65.7 ± 0.4	71.5 ± 0.4
GCN-G2G	92.5 ± 0.9	92.6 ± 0.8	93.7 ± 0.4	97.5 ± 0.2	$\mathbf{96.6 \pm 0.1}$	$\mathbf{97.2 \pm 0.1}$
GCN-G2V	$\mathbf{92.7 \pm 0.8}$	92.5 ± 0.9	$\mathbf{94.0 \pm 0.4}$	97.7 ± 0.2	$\mathbf{96.6 \pm 0.1}$	$\mathbf{97.2 \pm 0.1}$

- Variants of existing methods can exhibit significantly improved performance.
- Under some configurations, simplified convolutional encoders can exhibit lower generalization.
- Graph properties can have an important effect depending on the method of choice.
- Certain modeling choices made in previous works are less important than previously believed, and other optimization-related choices are sometimes responsible for improved performance.
- The right method is data-dependent: **methods can be tailored to a specific graph and task** through hyperparameter search across components:



Experiments

- We perform a comparative study of methods in the tasks of **link prediction** and **node classification**, and we introduce novel variants of these approaches based on different graph encoders, representations and scoring functions.
- We run experiments with **DeepWalk** [1], **Graph Autoencoders** [2], **Deep Graph Infomax** [3], and **Graph2Gauss** [4].
- Datasets include graphs containing on the order of thousands to hundreds of thousands of nodes and edges, with varying degrees of connectivity

Conclusions

- The described modular framework is of practical use for devising methods for unsupervised learning on graphs, and highlightinging strengths and weaknesses of existing approaches.
- Our framework motivates hyperparameter search strategies so that learned node embeddings are better suited to datasets and tasks of interest.
- Directions for future work include extending our analysis to large scale graphs, and additional downstream tasks for evaluation, such as graph classification.

References:
[1] Perozzi et al. "Deepwalk: Online learning of social representations." SIGKDD (2014).
[2] Kipf and Welling. "Variational graph autoencoders." arXiv preprint arXiv:1611.07308 (2016).
[3] Veličković et al. "Deep graph infomax." arXiv preprint arXiv:1809.10341 (2018).
[4] Bojchevski and Günnemann. "Deep Gaussian embedding of attributed graphs: unsupervised inductive learning via ranking." arXiv preprint arXiv:1707.03815 (2017).

Project supported by SAP SE Berlin