Wasserstein Embedding for Graph Learning (WEGL)

This repository contains the code accompanying our paper titled *Wasserstein Embedding for Graph Learning (WEGL)*. *WEGL* is a novel and fast framework for embedding entire graphs in a vector space, in which various machine learning models are applicable for graph-level prediction tasks. Our proposed approach embeds a graph into a Hilbert space, where the ℓ_2 distance between two embedded graphs provides a true metric between the graphs that approximates their 2-Wasserstein distance. For a set of M graphs, the proposed method provides:

- Reduced computational complexity of estimating the graph Wasserstein distance [togninalli2019wasserstein] for a dataset of M graphs from a quadratic complexity in the number of graphs, i.e., $\frac{M(M-1)}{2}$ calculations, to linear complexity, i.e., M calculations of the Wasserstein distance; and
- An explicit Hilbertian embedding for graphs, which is not restricted to kernel methods, and therefore can be used in conjunction with any downstream classification framework.

The Figure shown below, visualizes *WEGL*'s process for embedding graphs into a linear Wasserstein embedding (aka, the linear optimal transport embedding).

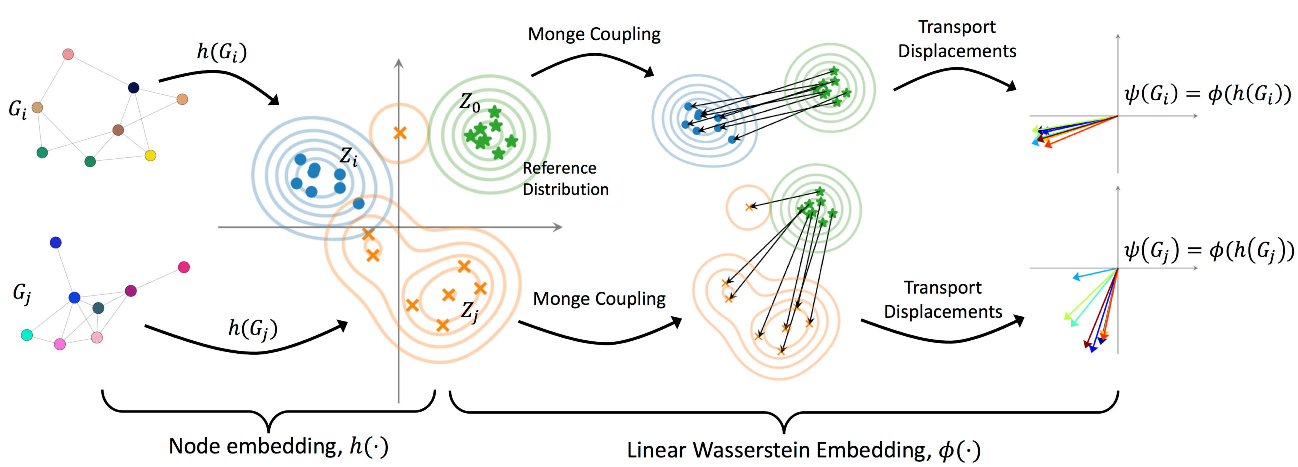

Fig 1. Our proposed graph embedding framework, WEGL, combines node embedding methods with the linear Wasserstein embedding framework described.

Requirements

- PyTorch
- PyTorch Geometric
- Scikit-Learn (pip install scikit-learn)
- OGB (pip install ogb)
- POT (pip install pot)
- PrettyTable (pip install prettytable)
- TQDM (pip install tqdm)

Code

- LOT.ipynb: An i-PythonNotebook tutorial for the Linear Optimal Transport framework.
- WEGL__RandomForest__ogbg__molhiv__VirtualNode.ipynb: Demo of *WEGL* on the OGB-MOLHIV dataset.

References

[togninalli2019wasserstein] Togninalli, M., Ghisu, E., Llinares-López, F., Rieck, B. and Borgwardt, K., 2019. Wasserstein weisfeiler-lehman graph kernels. In *Advances in Neural Information Processing Systems* (pp. 6436-6446).

License