The Journal of Open Source Software

GEM: A Python package for graph embedding methods

Palash Goyal[1] **and Emilio Ferrara**[1]

1 USC Information Sciences Institute

DOI: 10.21105/joss.00876

Software

- Review
- Repository
- Archive

Submitted: 09 July 2018
Published: 01 September 2018

Summary

Many physical systems in the world involve interactions between different entities and can be represented as graphs. Understanding the structure and analyzing properties of graphs are hence paramount to developing insights into the physical systems. Graph embedding, which aims to represent a graph in a low dimensional vector space, takes a step in this direction. The embeddings can be used for various tasks on graphs such as visualization, clustering, classification and prediction.

`GEM` is a Python package which offers a general framework for graph embedding methods. It implements many state-of-the-art embedding techniques including Locally Linear Embedding (Roweis & Saul, 2000), Laplacian Eigenmaps (Belkin & Niyogi, 2003), Graph Factorization (Ahmed, Shervashidze, Narayanamurthy, Josifovski, & Smola, 2013), HOPE (Ou, Cui, Pei, Zhang, & Zhu, 2016), SDNE (Wang, Cui, & Zhu, 2016) and node2vec (Grover & Leskovec, 2016). It is formatted such that new methods can be easily added for comparison. Furthermore, the framework implements several functions to evaluate the quality of obtained embedding including graph reconstruction, link prediction, visualization and node classification. It supports many edge reconstruction metrics including cosine similarity, euclidean distance and decoder based. For node classification, it defaults to one-vs-rest logistic regression classifier and supports other classifiers. For faster execution, C++ backend is integrated using Boost for supported methods.

`GEM` was designed to be used by researchers studying graphs. It has already been used in a number of scientific publications to compare novel methods against the state-of-the-art and general evaluation (Salehi Rizi, Granitzer, & Ziegler, 2017, Lyu, Zhang, & Zhang (2017)). A paper showcasing the results using `GEM` on various real world datasets can be accessed (Goyal & Ferrara, 2018). The source code of `GEM` is made available at https://github.com/palash1992/GEM. Bug reports and feedback can be directed to the Github issues page (https://github.com/palash1992/GEM/issues).

References

Ahmed, A., Shervashidze, N., Narayanamurthy, S., Josifovski, V., & Smola, A. J. (2013). Distributed large-scale natural graph factorization. In *Proceedings of the 22nd international conference on world wide web* (pp. 37–48). ACM. doi:10.1145/2488388.2488393

Belkin, M., & Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. In *Neural computation* (pp. 1373–1396). doi:10.1162/089976603321780317

Goyal, P., & Ferrara, E. (2018). Graph embedding techniques, applications, and performance: A survey. *Knowledge-Based Systems*. doi:10.1016/j.knosys.2018.03.022

Grover, A., & Leskovec, J. (2016). Node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 855–864). ACM. doi:10.1145/2939672.2939754

Lyu, T., Zhang, Y., & Zhang, Y. (2017). Enhancing the network embedding quality with structural similarity. In *Proceedings of the 2017 ACM on conference on information and knowledge management* (pp. 147–156). ACM. doi:10.1145/3132847.3132900

Ou, M., Cui, P., Pei, J., Zhang, Z., & Zhu, W. (2016). Asymmetric transitivity preserving graph embedding. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1105–1114). ACM. doi:10.1145/2939672.2939751

Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, *290*(5500), 2323–2326. doi:10.1126/science.290.5500.2323

Salehi Rizi, F., Granitzer, M., & Ziegler, K. (2017). Properties of vector embeddings in social networks. *Algorithms*, *10*(4), 109. doi:10.3390/a10040109

Wang, D., Cui, P., & Zhu, W. (2016). Structural deep network embedding. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1225–1234). ACM. doi:10.1145/2939672.2939753