Representation Learning in Large Attributed Graphs

Nesreen K. Ahmed
Intel Labs
nesreen.k.ahmed@intel.com

Ryan A. Rossi
PARC
rrossi@parc.com

Rong Zhou
Google
rongzhou@google.com

John Boaz Lee
WPI
jtlee@wpi.edu

Xiangnan Kong
WPI
xkong@wpi.edu

Theodore L. Willke
Intel Labs
ted.willke@intel.com

Hoda Eldardiry
PARC
heldardiry@parc.com

Graphs (networks) are ubiquitous and allow us to model entities (nodes) and the dependencies (edges) between them. Graph data is often observed directly in the natural world (*e.g.*, biological or social networks) [21] or constructed from non-relational data by deriving a metric space between entities and retaining only the most significant edges [5, 23]. Learning a useful feature representation from graph data lies at the heart and success of many machine learning tasks such as classification, anomaly detection, link prediction, among many others. Many existing techniques use *random walks* as a basis for learning features or estimating the parameters of a graph model for a downstream prediction task. Examples include recent node embedding methods such as DeepWalk [12], node2vec [4], as well as graph-based deep learning algorithms. However, the simple random walk used by these methods is fundamentally tied to the *identity* of the node. This has three main disadvantages. First, these approaches are inherently transductive and do not generalize to unseen nodes and other graphs [16]. Second, they are not space-efficient as a feature vector is learned for each node which is impractical for large graphs. Third, most of these approaches lack support for *attributed graphs*.

To make these methods more generally applicable, we propose a framework based on the notion of attributed random walk that is not tied to node identity and is instead based on learning a function $\phi : \mathbf{x} \rightarrow w$ that maps a node attribute vector to a type. This framework serves as a basis for generalizing existing methods such as DeepWalk, node2vec, LINE, and many other existing (and future) methods that leverage traditional random walks. Given an (un)directed graph $G = (V, E)$, the framework consists of two general steps:

1. **Function Mapping Nodes to Types**: The first step is to learn a function ϕ that maps nodes to types based on a $N \times K$ matrix $\mathbf{X}$ of attributes. Note $\mathbf{X}$ may be given as input and/or computed based on the structure of the graph.
2. **Attributed Random Walks**: The second step uses the types derived by the function ϕ for generating *attributed random walks*. An *attributed walk* S of length L is defined as a sequence of adjacent node types $\phi(\mathbf{x}_{i_1}), \phi(\mathbf{x}_{i_2}), \ldots, \phi(\mathbf{x}_{i_{L+1}})$ associated with a sequence of indices $i_1, i_2, \ldots, i_{L+1}$ such that $(v_{i_t}, v_{i_{t+1}}) \in E$ for all $1 \leq t \leq L$.

The set of attributed random walks can then be given as input into the Skip-Gram model (or other representation learning methods) to learn embeddings for the node types (as opposed to the nodes themselves). Our proposed framework has the following key properties:

- **Space-efficient**: It requires on average 853x less space than existing methods.
- **Accurate**: It is accurate with an average improvement of 16.1% across a variety of graphs from several domains.[1]
- **Inductive**: It is an inductive learning approach that is able to learn embeddings for new nodes and graphs.
- **Attributed**: It naturally supports graphs with attributes (if available) and serves as a foundation for generalizing existing methods for use on attributed graphs.

[1] Graph data is from [14]

Table 1: AUC scores for Link Prediction. See text for discussion.

GRAPH	**Our Method**	**Node2Vec** [4]	**DeepWalk** [12]	**LINE** [20]
biogrid–human	**0.877**	0.869	0.864	0.744
bn–cat	**0.710**	0.627	0.627	0.672
bn–rat–brain	**0.748**	0.716	0.716	0.691
bn–rat–cerebral	**0.867**	0.813	0.811	0.709
ca–CSphd	**0.838**	0.768	0.735	0.620
eco–everglades	**0.762**	0.739	0.739	0.704
eco–fweb–baydry	**0.681**	0.655	0.627	0.660
ia–radoslaw–email	**0.867**	0.756	0.745	0.769
inf–USAir97	**0.884**	0.881	0.834	0.843
soc–anybeat	**0.961**	0.854	0.848	0.850
soc–dolphins	**0.656**	0.580	0.498	0.551
fb–Yale4	**0.793**	0.742	0.728	0.763
fb–nips–ego	**0.998**	0.997	0.996	0.743
web–EPA	**0.926**	0.804	0.738	0.768

To evaluate the effectiveness of the proposed framework we use it to generalize node2vec. Additional details are discussed in [2]. Experimental results are provided in Table 1. The results in Table 1 use the MEAN OP $(\mathbf{z}_i + \mathbf{z}_j)/2$ to construct edge features from the learned node embedding vectors $\mathbf{z}_i$ and $\mathbf{z}_j$ of node i and j. Similar results were observed using other operators. In all cases, the generalized approach outperforms all other methods across a wide variety of networks from biology to information and social networks. The experimental results in Table 1 demonstrate the effectiveness of the proposed framework for generalizing existing methods making them more powerful and practical for attributed and heterogeneous/typed networks [7, 18, 19] as well as for inductive learning tasks [6, 9, 10, 13, 16, 17, 22]. Notably, the generalization enables these methods to leverage any available intrinsic or self-attributes as well as any arbitrary set of structural features derived from the graph (*e.g.*, higher-order subgraph features such as graphlets [1]) or even relational features that leverage the graph structure as well as the initial intrinsic/self-attributes [15]. Furthermore, the learned embeddings from our approach naturally generalize as they are no longer tied to identity and instead represent general functions that can easily be extracted on another arbitrary graph. Therefore, the approach naturally supports a wide range of graph-based transfer learning tasks such as cross-temporal link prediction [10], classification in dynamic networks [11], across-network prediction and modeling [9], role discovery [13], graph matching [17] & similarity [22], or more generally inductive feature learning for networks [16], among others [3, 8]. Finally, the generalized approach using the proposed framework is guaranteed to perform at least as well as the original method since it is recovered as a special case of the framework.

Conclusion: This work proposed a flexible framework for generalizing an important class of embedding and representation learning methods for graphs that leverage random walks. The framework serves as a basis for generalizing existing methods for use with attributed graphs, unseen nodes, inductive learning/graph-based transfer learning tasks, while also scaling up these approaches for handling significantly larger graphs due to the inherent space-efficiency of the approach. Finally, the framework was shown to have the following desired properties: space-efficient, accurate, inductive, and able to support graphs with attributes (and more generally heterogeneous/typed networks).

References

[1] N. K. Ahmed, J. Neville, R. A. Rossi, and N. Duffield. Efficient graphlet counting for large networks. In *ICDM*, page 10, 2015.

[2] N. K. Ahmed, R. A. Rossi, R. Zhou, J. B. Lee, X. Kong, T. L. Willke, and H. Eldardiry. A framework for generalizing graph-based representation learning methods. In *arXiv:1709.04596*, pages 1–8, 2017.

[3] L. Akoglu, H. Tong, and D. Koutra. Graph based anomaly detection and description: a survey. *DMKD*, 29(3):626–688, 2015.

[4] A. Grover and J. Leskovec. node2vec: Scalable feature learning for networks. In *SIGKDD*, pages 855–864, 2016.

[5] M. Henaff, J. Bruna, and Y. LeCun. Deep convolutional networks on graph-structured data. *arXiv:1506.05163*, 2015.

[6] P. Holme and J. Saramäki. Temporal networks. *Physics Reports*, 2012.

[7] X. Kong, P. S. Yu, Y. Ding, and D. J. Wild. Meta path-based collective classification in heterogeneous information networks. In *CIKM*, pages 1567–1571. ACM, 2012.

[8] J. B. Lee, R. Rossi, and X. Kong. Deep graph attention model. In *arXiv:1709.06075*, pages 1–8, 2017.

[9] J. Neville, D. Jensen, L. Friedland, and M. Hay. Learning relational probability trees. In *SIGKDD*, pages 625–630, 2003.

[10] S. Oyama, K. Hayashi, and H. Kashima. Cross-temporal link prediction. In *ICDM*, pages 1188–1193. IEEE, 2011.

[11] H. Park, J. Moore, and J. Neville. Deep dynamic relational classifiers: Exploiting dynamic neighborhoods in complex networks. pages 1–7, 2017.

[12] B. Perozzi, R. Al-Rfou, and S. Skiena. Deepwalk: Online learning of social representations. In *SIGKDD*, pages 701–710, 2014.

[13] R. A. Rossi and N. K. Ahmed. Role discovery in networks. *Transactions on Knowledge and Data Engineering*, 27(4):1112–1131, April 2015. ISSN 1041-4347. doi: 10.1109/TKDE.2014.2349913.

[14] R. A. Rossi and N. K. Ahmed. An interactive data repository with visual analytics. *SIGKDD Explor.*, 17(2):37–41, 2016. URL `http://networkrepository.com`.

[15] R. A. Rossi, L. K. McDowell, D. W. Aha, and J. Neville. Transforming graph data for statistical relational learning. *Journal of Artificial Intelligence Research*, 45(1):363–441, 2012.

[16] R. A. Rossi, R. Zhou, and N. K. Ahmed. Deep feature learning for graphs. In *arXiv:1704.08829*, 2017.

[17] R. Singh, J. Xu, and B. Berger. Pairwise global alignment of protein interaction networks by matching neighborhood topology. In *Proceedings of the 11th Annual International Conference on Research in Computational Molecular Biology*, volume 4453 of *Lecture Notes in Computer Science*, pages 16–31. Springer, 2007.

[18] Y. Sun and J. Han. Mining heterogeneous information networks: a structural analysis approach. *ACM SIGKDD Explorations Newsletter*, 14(2):20–28, 2013.

[19] Y. Sun, J. Han, X. Yan, and P. S. Yu. Mining knowledge from interconnected data: a heterogeneous information network analysis approach. *Proceedings of the VLDB Endowment*, 5(12): 2022–2023, 2012.

[20] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei. LINE: Large-scale Information Network Embedding. In *WWW*, pages 1067–1077, 2015.

[21] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi. On the evolution of user interaction in Facebook. In *OSN*, pages 37–42, 2009.

[22] L. A. Zager and G. C. Verghese. Graph similarity scoring and matching. *Applied mathematics letters*, 21(1):86–94, 2008.

[23] X. Zhu, Z. Ghahramani, and J. D. Lafferty. Semi-supervised learning using gaussian fields and harmonic functions. In *ICML*, pages 912–919, 2003.