Node Embedding Samples for Expressive Graph Substructure Representations

Ishaan Saxena

Abstract

Positional node embeddings are used frequently for prediction tasks on Graphs, but usually lack structural information when taken in the familiar sense. This work discusses the ineffectiveness of node embeddings in the traditional sense - as lower dimensional embeddings of data by utilizing a model-free definition of node embeddings provided by Srinivasan & Ribeiro (2019). An alternative approach to use Monte Carlo samples of node embeddings is discussed. Finally, we present the tool of ensemble node embeddings to make the structural representations obtained from node embeddings more expressive.

1. Introduction

Positional node embeddings have found widespread use in various learning tasks that involve multi-ary relationships within graphs, such as link prediction, clustering, etc. However, they seem to be ineffective in situations involving structural information about the graph, such as role discovery or node classification. This is partly a consequence of the way (positional) node embeddings have been thought of in the body of literature surrounding them.

A more familiar description of node embeddings would suggest that they are lower-dimensional representations of nodes (or sets of nodes) that preserve relative position within the graph structure (Graham & Winkler, 1985; Cai et al., 2018). Such an interpretation of node embeddings implies that they rely on some model that defines closeness.

To understand why this presents a problem for learning tasks concerning the structure of nodes (or sets of nodes), consider the graph in Figure 1.

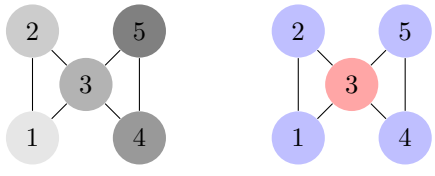
An embedding obtained for this graph via a positional node embedding, say the SVD, would result in a d-dimensional vector representation of the nodes which would capture the closeness of the nodes to each other but would produce distinct embeddings for nodes 1 and 2 (Wang et al., 2017; Yan et al., 2006). However, the two nodes have the same structural roles in the graph. Naturally, with this notion, multi-ary relationships may be modeled better than the individual role of (a set of) nodes in the graph, as properties of two (sets of) nodes are preserved relative to each other in the embedding space, while the individual structural information may be lost. As a result, the positional embeddings when taken in a more familiar sense may not be of much help when the structural properties of the graphs must be exploited.

Srinivasan & Ribeiro (2019) present a more nuanced, model-free description of node embeddings. Subsequently, they show that node embeddings are related to structural representations of a graph in the same way as a sample is to a distribution. This has some interesting consequences for the problem at hand. Firstly, this suggests that all tasks which may be done with structural representations can also be done with node embeddings (and vice-versa). Secondly, this presents a method to obtain structural representations of graphs from multiple (samples of) node embeddings. As such, this interpretation of node embeddings leads to a new framework that enables the use of node embeddings to more effectively obtain information about the structural roles of sets of nodes within a graph.

Following a brief discussion of this notion of node embeddings, we will use it to show exactly why the more familiar notion fails to capture structural information. Finally, we explore what a more generalized framework to utilize node embeddings in an inductive setting would look like, and how we can make node embeddings and structural representations more expressive.

2. Preliminaries

In this section, some notation and definitions are introduced. For a more detailed description of some of these definitions, we recommend Srinivasan & Ribeiro (2019); Bloem-Reddy & Teh (2019).



(a) Positional (b) Structural

Figure 1. Five-node graph illustrating (left) its positional node embeddings and (right) its structural node representations.

A graph, $G = (V, E, \mathbf{X}, \mathbf{E})$ is a collection of the set of nodes V, the set of edges $E \subseteq V \times V$, with $\mathbf{X} \in \mathbb{R}^{n \times k}$ and $\mathbf{E} \in \mathbb{R}^{n \times n \times k'}$ representing node and edge features respectively. Suppose that the edge set has an associated adjacency matrix $A : V \times V \to \{0, 1\}$. We may compress $\mathbf{E}$ and A into a single 3-mode adjacency tensor $\mathbf{A} \in \mathbb{R}^{n \times n \times (k'+1)}$, and just write $G = (\mathbf{A}, \mathbf{X})$ for any graph G.

A permutation action π is a function that acts on any tensor defined over the nodes V and outputs an equivalent tensor with the order of the nodes permuted. We define Π_n as the set of all $n!$ such permutations.

Let Σ be the set of all possible attributed graphs G. We say that a function $g : \Sigma \to \mathbb{R}^{n \times \cdot}$ is $\mathcal{G}$-equivariant w.r.t. all valid permutations of the set of nodes whenever any permutation action $\pi \in \Pi_n$ in the symmetric space Σ is associated with the same permutation action of the nodes in the $\mathbb{R}^{n \times \cdot}$ symmetric space. We say that $g : \Sigma \to \mathbb{R}^{\cdot}$ is $\mathcal{G}$-invariant if it is invariant to any permutation action $\pi \in \Pi_n$ in the Σ symmetric space.

2.1. Vertex Subset Orbits

An orbit is the result of a group action Π_n acting on elements of a group corresponding to bijective transformations of the space that preserves some structure of the space. The orbit of an element $\omega \in \Omega$ is the set of equivalent elements under action Π_n, i.e., $\Pi_n(x) = \{\pi(x) \mid \pi \in \Pi_n\}$.

The equivalence classes of k-sized subsets of vertices $S \subset V$ of a graph G under the action of automorphism between the subsets are called vertex subset orbits (or simply vertex orbits for $k = 1$). For any $S_1, S_2 \subset V$ in the same equivalence class, we say that S_1, S_2 are jointly isomorphic.

The orbit of a set of nodes can be thought of as a collection of all other sets of nodes in a graph that they may be 'exchanged with' without any changes in the graph structure. Thus, sets of nodes in the same orbit essentially have the same structural roles within the graph.

2.2. Structural Representations and Node Embeddings

Consider some graph $G = (\mathbf{A}, \mathbf{X})$.

The structural representation of $S \subset V$ in G, as in Srinivasan & Ribeiro (2019), is the G-invariant representation $\Gamma(S, \mathbf{A}, \mathbf{X})$. That is, $\forall S \subset V$ and $\forall (\mathbf{A}, \mathbf{X}) \in \Sigma$,

$$\Gamma(S, \mathbf{A}, \mathbf{X}) = \Gamma(\pi(S), \pi(\mathbf{A}), \pi(\mathbf{X}))$$

holds true for any $\pi \in \Pi_n$. Clearly, for any two jointly isomorphic subsets $S_1, S_2 \subset V$, $\Gamma(S_1, \mathbf{A}, \mathbf{X}) = \Gamma(S_2, \mathbf{A}, \mathbf{X})$.

A structural representation Γ is called most-expressive iff there exists a bijective measurable map between $\Gamma(S, \mathbf{A}, \mathbf{X})$ and the orbit of S in G. We write this as $\Gamma^\star$.

Following Definition 12 of Srinivasan & Ribeiro (2019), node embeddings of a graph $G = (\mathbf{A}, \mathbf{X})$ are defined as joint samples of random variables $(Z_i)_{i \in V} \mid \mathbf{A}, \mathbf{X} \sim p(\cdot \mid \mathbf{A}, \mathbf{X})$ where $Z_i \in \mathbb{R}^d$ for some $d \geq 1$, and $p(\cdot \mid \mathbf{A}, \mathbf{X})$ is a G-equivariant probability distribution. That is,

$$\pi(p(\cdot \mid \mathbf{A}, \mathbf{X})) = p(\cdot \mid \pi(\mathbf{A}), \pi(\mathbf{X}))$$

for any permutation action $\pi \in \Pi_n$.

2.3. Prediction Task

Let $\mathcal{S}$ be a set of size k subsets of V. Let $\mathbf{Y}(\mathcal{S}, \mathbf{A}, \mathbf{X})$ or equivalently $(Y(S, \mathbf{A}, \mathbf{X}))_{S \in \mathcal{S}}$ be a set of random variables defined over the sets $S \in \mathcal{S}$, that are invariant to the ordering of S, $S \in \mathcal{S}$, such that $Y(S_1, \mathbf{A}, \mathbf{X}) \overset{d}{=} Y(S_2, \mathbf{A}, \mathbf{X})$ for any two jointly isomorphic subsets S_1, S_2 (where $\overset{d}{=}$ means equality in their marginal distributions).

We wish to use some notion of node embeddings to predict $\mathbf{Y}(\mathcal{S}, \mathbf{A}, \mathbf{X})$ for all $S \in \mathcal{S}$ in the graph.

3. Model-free Node Embeddings

3.1. Ineffectiveness of Single-Sample Node Embeddings

To understand why the traditional definition and usage of node embeddings falls short, let's consider the example of using the Singular Value Decomposition (SVD) of a graph's adjacency matrix to obtain node embeddings. Consider two graphs $G = (\mathbf{A}, \mathbf{X})$ and $H = (\mathbf{B}, \mathbf{X}')$. We know that G and H are isomorphic graphs if and only if there is some permutation matrix P which corresponds to a permutation in Π_n for which $PAP^T = B$. That is, there is some valid relabelling of the nodes of G to obtain an equivalent graph H. This indicates that the eigenvalues A and B must be the same. Note that for all eigenvalues of multiplicity 1, eigenvectors may be determined up to sign which depends on the permutation P (Spielman, 2018). As for all the degenerate eigenvalues, if the matrix is permuted randomly upon input, the distribution of eigenvector solutions corresponding to the degenerate eigenvalues would be $\mathcal{G}$-equivariant, since the distribution of the eigenvector solutions does not change with unitary rotations in the corresponding degenerate eigenspaces (like a permutation from A to B would result in). This clearly shows that the embeddings produced by the SVD of the adjacency matrix of a graph are a single sample of some $\mathcal{G}$-equivariant probability distribution, satisfying the model-free definition of node embeddings.

Now, consider some non-empty, isomorphic sets of nodes $S_1, S_2 \subset V$ in a graph $G = (\mathbf{A}, \mathbf{X})$. From this, we may deduce that for any valid structural representation Γ,

$$\Gamma(S_1, \mathbf{A}, \mathbf{X}) = \Gamma(S_2, \mathbf{A}, \mathbf{X})$$

Furthermore, we know for any $\mathcal{G}$-equivariant distribution

$p(\cdot \mid \mathbf{A}, \mathbf{X})$ that

$$p((Z_i)_{i \in S_1} \mid \mathbf{A}, \mathbf{X}) = \pi(p((Z_j)_{j \in S_2} \mid \mathbf{A}, \mathbf{X}))$$

This has an important consequence. Suppose $\mathbf{Z}$ is the set of embeddings obtained by a single sample from a $\mathcal{G}$-equivariant probability distribution (e.g. a set of embeddings obtained by the left singular vectors). In such a case, even though the distribution from which the embedding has been sampled must be $\mathcal{G}$-equivariant, the embedding $\mathbf{Z}$ itself may not be $\mathcal{G}$-equivariant. This may result in different embeddings for isomorphic sets of nodes S_1, S_2 even though the distribution of possible embeddings for the two nodes, as well as their structural representation in the graph, must be identical.

Thus, using a single sample of node embeddings $\mathbf{Z}$ may result in different (and inherently poor) representations for sets of nodes that share the same role in a graph. Such a method may thus be deemed transductive (Cai et al., 2018). However, with some repeated Monte Carlo sampling, we should be able to obtain a $\mathcal{G}$-invariant representation for jointly isomorphic sets of nodes, as long as the distribution from which $\mathbf{Z}$ is sampled is $\mathcal{G}$-equivariant. Clearly, node embeddings can be inductive since they may reveal structural information about the data.

3.2. Structural Representations of Graphs from Node Embedding Samples

Srinivasan & Ribeiro (2019, Corollary 3) show that we may learn a k-node structural representation of a subset of k nodes S by sampling Z and simply learning a function $f^{(k)}$ whose average

$$\Gamma(S, \mathbf{A}, \mathbf{X}) = \mathbb{E}[f^{(k)}((Z_v)_{v \in S})]$$

can be used to predict $Y(S, \mathbf{A}, \mathbf{X})$. As such, this provides for a neat framework to utilize node embeddings in such a prediction task effectively, as this sampling and aggregation help obtain a G-invariant structural representation for some given set of nodes.

Recall that the most expressive structural representation $\Gamma^\star$ for some graph G is one for which there is a bijective measurable map between $\Gamma^\star(S, \mathbf{A}, \mathbf{X})$ and the orbit of S in G. Thus, $\Gamma^\star(S_1, \mathbf{A}, \mathbf{X}) = \Gamma^\star(S_2, \mathbf{A}, \mathbf{X})$ if and only if S_1, S_2 are jointly isomorphic subsets. Srinivasan & Ribeiro (2019, Theorem 2) suggests that $\forall (\mathbf{A}, \mathbf{X}) \in \Sigma$, there is some most expressive node embedding $\mathbf{Z}^\star \mid \mathbf{A}, \mathbf{X}$ such that $\Gamma^\star(S, \mathbf{A}, \mathbf{X}) = \mathbb{E}[f^{(k)}((Z^\star_v)_{v \in S})]$ for some appropriate collection of functions $\{f^{(k)}(\cdot)\}_{k=1,\ldots,n}$. Furthermore, it is argued that given the most expressive structural representation $\Gamma^\star(S, \mathbf{A}, \mathbf{X})$, the results of the prediction task $Y(S, \mathbf{A}, \mathbf{X})$ for any $S \in \mathcal{S}$ are independent of the node embeddings $\mathbf{Z} \mid \mathbf{A}, \mathbf{X}$ as defined previously. Clearly, $\Gamma^\star(S, \mathbf{A}, \mathbf{X})$ is a representation of the orbit of S in G. As a result, $\Gamma^\star(S, \mathbf{A}, \mathbf{X})$ gives us the desired representation of S for any prediction task on S where we need to predict $Y(S, \mathbf{A}, \mathbf{X})$, as it encodes information about the structural role of any $S \subset V$ in the graph. It should be obvious that any structural representation Γ learned from samples of an embedding $\mathbf{Z} \mid \mathbf{A}, \mathbf{X}$ is most expressive if and only if the embedding is most expressive.

Notably, finding the most expressive representation might is, in general, a hard problem, since it is equivalent to finding the node orbits, or all automorphisms of a Graph. As such, one may have to utilize less expressive structural representations for prediction tasks. Now, one obvious question to ask is whether we can compare the expressiveness of distinct structural representations or node embeddings. Moreover, how can we obtain more expressive structural representations? We will explore these further in the following section.

4. Finding Expressive Structural Representations

Suppose some structural representation Γ is not most expressive. Then, there is not a bijective measurable map between $\Gamma(S, \mathbf{A}, \mathbf{X})$ and the orbit of S in G. Since Γ is a structural representation, it cannot be true for any jointly isomorphic S_1, S_2 that $\Gamma(S_1, \mathbf{A}, \mathbf{X}) \neq \Gamma(S_2, \mathbf{A}, \mathbf{X})$. This indicates that $\Gamma(S_1, \mathbf{A}, \mathbf{X}) = \Gamma(S_2, \mathbf{A}, \mathbf{X})$ for non-isomorphic S_1, S_2. This would result in a situation where $\Gamma(U, \mathbf{A}, \mathbf{X})$ would be the same for all U in the orbits of S_1 or S_2. This gives a natural approach to define what a more expressive structural representation would be. Any structural representation Γ' which would result in fewer orbits mapping to the same representation than in Γ would be more expressive than Γ. This is consistent with the definition of the most expressive structural representation, as there can be no structural representation that is more expressive.

Keeping in mind this notion of a more expressive structural representation relative to some other structural representation, let's explore how we can attempt to find more expressive structural representations, and make better use of node embeddings in an inductive learning setting. We discuss two methods – one combines multiple structural representations to find a new, more expressive structural representation (or equivalently, combines multiple node embeddings and uses them to learn a structural representation), and the other utilizes systematic perturbations to increase the expressiveness of node embeddings.

4.1. Ensemble Structural Representations

Suppose $\Gamma_1, \ldots, \Gamma_m$ are structural representations obtained from distinct node embeddings methods such that $\Gamma_i : \mathcal{S} \times \Sigma \to \mathbb{R}^{d_i}$, and Γ_i is not most expressive for any $i = 1, \ldots, m$. It is easy to show that a structural repre-

sentation Γ formed by stacking each of the m structural representation, that is,

$$\Gamma(S, \mathbf{A}, \mathbf{X}) = \begin{bmatrix} \Gamma_1(S, \mathbf{A}, \mathbf{X}) \\ \vdots \\ \Gamma_m(S, \mathbf{A}, \mathbf{X}) \end{bmatrix}$$

would result in a structural representation which is at least as expressive as any one of the m structural representations which it was formed from.

Equivalently, one may use multiple samples of an ensemble of node embeddings $\mathbf{Z}_1, \dots \mathbf{Z}_m$ to learn a structural representation Γ such that $\Gamma(S, \mathbf{A}, \mathbf{X}) = \mathbb{E}[f^{(|S|)}((Z_v)_{v \in S})]$ where

$$Z_v = \begin{bmatrix} (Z_1)_v \\ \vdots \\ (Z_m)_v \end{bmatrix}$$

Again, it is easy to show that the structural representation Γ learned from this ensemble of node embeddings would be at least as expressive as the structural representations Γ_i learned from any one of the $\mathbf{Z}_i$.

For instance, suppose that Ω_1, Ω_2 are two distinct vertex subset orbits for subsets of size k in the graph $G = (\mathbf{A}, \mathbf{X})$. Assume that there is some structural representation Γ_1 for which $\Gamma_1(S_1, \mathbf{A}, \mathbf{X}) = \Gamma_1(S_2, \mathbf{A}, \mathbf{X})$ for any $S_1 \in \Omega_1$ and $S_2 \in \Omega_2$. If there is another structural representation Γ_2 for which $\Gamma_2(S_1, \mathbf{A}, \mathbf{X}) \neq \Gamma_2(S_2, \mathbf{A}, \mathbf{X})$ for all $S_1 \in \Omega_1$ and all $S_2 \in \Omega_2$, then the structural representation Γ formed by the same stacking procedure will have distinct representations for any members of Ω_1, Ω_2. As such, Γ is more expressive than Γ_1, and at least as expressive as Γ_2.

It should be noted that for some combinations of node embeddings, we could assert that the ensemble would result in a more expressive structural representation with a very high probability (instead of just one which is at least as expressive). This assertion is stronger. Characterizing these node embeddings is an interesting question for future work.

4.2. Merging Orbits with Systematic Perturbations

Better representations for vertex subsets may also be obtained by aggregating representations for the same vertex subsets with appropriately chosen perturbations applied to them. This has the effect of 'blurring' the structural orbits which a vertex subset is likely to be a part of. Let's formalize this notion before we proceed with a more in-depth discussion.

Consider some graph $G = (\mathbf{A}, \mathbf{X})$. Suppose there is some graph $G_p = (\mathbf{A}_p, \mathbf{X})$ defined on the same set of vertices. We say that graph obtained by taking the symmetric difference of the two edge sets is called the perturbed graph H, on the application of G_p to G. That is, $H = (\mathbf{A} \triangle \mathbf{A}_p, \mathbf{X})$. Note that taking the symmetric difference is simply flipping the relationship (u, v) in G for all (u, v) in G_p. Thus, to ensure that we do not stray too far away from the current topology, we restrict our perturbation graph G_p to a budget β of the number of edges which may be flipped by enforcing the constraint $\sum_{e \in E(G_p)} 1 \leq \beta$.

We can characterize these perturbation graphs (G_p) with respect to their effects on the given graph (G).

1. Trivial action: if G_p is an empty graph, $H = G$.
2. Splitting perturbation: there exist some vertex subsets which were isomorphic in G but are not in H.
3. Merging perturbations: there exist some vertex subsets which were not isomorphic in G but are in H.
4. Mering and splitting perturbations: both merging and splitting occurs.

Suppose that we utilize a single method to obtain node embeddings repeatedly (say the SVD of the graph Laplacian). Let Γ_0 be the structural representation learned from repeated samples of $\mathbf{Z} \mid \mathbf{A}, \mathbf{X}$, such that it is not most expressive. Now, suppose that Γ is a structural representation learned from samples of a series of node embeddings obtained from some perturbed graphs formed by applying some perturbation actions to the original graph. Let's consider the expressiveness of Γ relative to Γ_0 based on the kind of perturbation which is applied.

Perturbations which are splitting or merging and splitting are not desirable. This is easy to see if we consider some S_1, S_2 which are in the same orbit in the original graph, but in two distinct orbits in the perturbed graph. If a structural representation is learned from the embeddings of S_1, S_2 obtained from the perturbed graph, they may have distinct structural representations. This is undesirable since they are in the same orbits in the original graph.

If the perturbation applied is trivial, there is no difference between the expressiveness of Γ and Γ_0, since there is no change in the orbits. That is, if two orbits map to the same representation before the perturbation action, they will still map to the same representation after the perturbation action.

However, if the perturbation action applied is one which merges orbits, we may obtain a more expressive representation (or obtain one which is at least as expressive). To see this, consider some graph $G = (\mathbf{A}, \mathbf{X})$ with two distinct orbits Ω_1, Ω_2 such that $\Gamma_0(S_1, \mathbf{A}, \mathbf{X}) = \Gamma_0(S_2, \mathbf{A}, \mathbf{X})$ for any $S_1 \in \Omega_1$ and $S_2 \in \Omega_2$. Now suppose that some merging perturbation action is applied which causes Ω_1 to merge with some orbit Ω_3 to result in the corresponding orbit $\Omega_{1,3}$ in the perturbed graph $H = (\mathbf{A}', \mathbf{X}')$. Now, if

the structural representation Γ is learned from the embeddings of sets S_1, S_2 in the original and perturbed graphs, it may be the case that $\Gamma(S_1, \mathbf{A}, \mathbf{X}) \neq \Gamma(S_2, \mathbf{A}, \mathbf{X})$ or $\Gamma(S_1, \mathbf{A}, \mathbf{X}) = \Gamma(S_2, \mathbf{A}, \mathbf{X})$. Since no orbits were split, the expressiveness cannot decrease. As such, the structural representation Γ learned from a series of merging perturbations applied to $\mathbf{A}, \mathbf{X}$ will be at least as expressive as Γ_0.

It should be noted that finding such merging perturbations might not be an easy task, and is an interesting line of research for future work on this topic.

5. Related Work

Previous works have categorized themselves into either ones that use node embedding methods or ones that use structural representations. This separation, although not necessary as we have shown, restricts the methods to only a few downstream tasks. (Cai et al., 2018) surveys various node embedding methods, and shows that most of them rely on embedding node subsets into metric spaces, and thus depend on a notion of closeness. However, Nematzadeh et al. (2017) shows that even non-metric methods such as node2vec (Grover & Leskovec, 2016) have properties simliar to those of metric spaces.

There are several attempts to obtain structure preserving node embeddings in the literature (Shaw & Jebara, 2009; Abu-El-Haija et al., 2018; Tsitsulin et al., 2018). These are merely heuristic based approaches which try to detect similarity in role to find metric space embeddings. This method still has a little merit though – one may think of similarities in structural neighborhood as a model-free idea of closeness (Donnat et al., 2018).

Structural representations have an increasing body of literature surrounding them focused on node and graph classification tasks. These utilize a group theoretic approach instead of embeddings into metric spaces Srinivasan & Ribeiro (2019). Several rely on utilizing GNNs to learn structural representations for graphs (Cao et al., 2016; Wang et al., 2016).

Srinivasan & Ribeiro (2019) provide the first theoretical framework uniting structural representations and node embeddings. This framework justifies the perfromance improvements seen by (Epasto & Perozzi, 2019; Goyal et al., 2019; Chen et al., 2019). These works combine multiple node embeddings in a simliar fashion to the one described in Section 4 to obtain a representation for k-sized subsets of vertices.

6. Conclusions

This work provided justification for the ineffectiveness of single samples of (positional) node embeddings for certain prediction tasks on graphs utilizing a group theoretic, model-free definition of node set embeddings provided by Srinivasan & Ribeiro (2019). An alternative approach, using Monte Carlo samples of node-embeddings to learn a structural representation for size k subsets of nodes is briefly discussed. Finally, we discuss how more expressive structural representations may be obtained from an ensemble of node embeddings, or alternatively (in principle) through systematic graph perturbations.

References

Abu-El-Haija, S., Perozzi, B., Al-Rfou, R., and Alemi, A. A. Watch your step: Learning node embeddings via graph attention. In *Advances in Neural Information Processing Systems*, pp. 9180–9190, 2018.

Bloem-Reddy, B. and Teh, Y. W. Probabilistic symmetry and invariant neural networks. *arXiv preprint arXiv:1901.06082*, 2019.

Cai, H., Zheng, V. W., and Chang, K. C.-C. A comprehensive survey of graph embedding: Problems, techniques, and applications. *IEEE Transactions on Knowledge and Data Engineering*, 30(9):1616–1637, 2018.

Cao, S., Lu, W., and Xu, Q. Deep neural networks for learning graph representations. In *Thirtieth AAAI conference on artificial intelligence*, 2016.

Chen, Y., Pu, J., Liu, X., and Zhang, X. Gaussian mixture embedding of multiple node roles in networks. *World Wide Web*, pp. 1–24, 2019.

Donnat, C., Zitnik, M., Hallac, D., and Leskovec, J. Learning structural node embeddings via diffusion wavelets. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1320–1329, 2018.

Epasto, A. and Perozzi, B. Is a single embedding enough? learning node representations that capture multiple social contexts. In *The World Wide Web Conference*, pp. 394–404, 2019.

Goyal, P., Huang, D., Chhetri, S. R., Canedo, A., Shree, J., and Patterson, E. Graph representation ensemble learning. *arXiv preprint arXiv:1909.02811*, 2019.

Graham, R. L. and Winkler, P. M. On isometric embeddings of graphs. *Transactions of the American mathematical Society*, 288(2):527–536, 1985.

Grover, A. and Leskovec, J. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 855–864, 2016.

Nematzadeh, A., Meylan, S. C., and Griffiths, T. L. Evaluating vector-space models of word representation, or, the unreasonable effectiveness of counting words near other words. In *CogSci*, 2017.

Shaw, B. and Jebara, T. Structure preserving embedding. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 937–944, 2009.

Spielman, D. Testing isomorphism of graphs with distinct eigenvalues. "https://www.cs.yale.edu/homes/spielman/561/lect08-18.pdf", September 2018.

Srinivasan, B. and Ribeiro, B. On the equivalence between node embeddings and structural graph representations. *arXiv preprint arXiv:1910.00452*, 2019.

Tsitsulin, A., Mottin, D., Karras, P., and Müller, E. Verse: Versatile graph embeddings from similarity measures. In *Proceedings of the 2018 World Wide Web Conference*, pp. 539–548, 2018.

Wang, D., Cui, P., and Zhu, W. Structural deep network embedding. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1225–1234, 2016.

Wang, X., Cui, P., Wang, J., Pei, J., Zhu, W., and Yang, S. Community preserving network embedding. In *Thirty-first AAAI conference on artificial intelligence*, 2017.

Yan, S., Xu, D., Zhang, B., Zhang, H.-J., Yang, Q., and Lin, S. Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE transactions on pattern analysis and machine intelligence*, 29(1):40–51, 2006.