vGraph: A Generative Model for Joint Community Detection and Node Representation Learning

Fan-Yun Sun$^{1,2}$, Meng Qu$^2$, Jordan Hoffmann$^{2,3}$, Chin-Wei Huang$^{2,4}$, Jian Tang$^{2,5,6}$

$^1$National Taiwan University,
$^2$Mila-Quebec Institute for Learning Algorithms, Canada
$^3$Harvard University, USA
$^4$Element AI, Canada
$^5$HEC Montreal, Canada
$^6$CIFAR AI Research Chair
b04902045@ntu.edu.tw

Abstract

This paper focuses on two fundamental tasks of graph analysis: community detection and node representation learning, which capture the global and local structures of graphs, respectively. In the current literature, these two tasks are usually independently studied while they are actually highly correlated. We propose a probabilistic generative model called vGraph to learn community membership and node representation collaboratively. Specifically, we assume that each node can be represented as a mixture of communities, and each community is defined as a multinomial distribution over nodes. Both the mixing coefficients and the community distribution are parameterized by the low-dimensional representations of the nodes and communities. We designed an effective variational inference algorithm which regularizes the community membership of neighboring nodes to be similar in the latent space. Experimental results on multiple real-world graphs show that vGraph is very effective in both community detection and node representation learning, outperforming many competitive baselines in both tasks. We show that the framework of vGraph is quite flexible and can be easily extended to detect hierarchical communities.

1 Introduction

Graphs, or networks, are a general and flexible data structure to encode complex relationships among objects. Examples of real-world graphs include social networks, airline networks, protein-protein interaction networks, and traffic networks. Recently, there has been increasing interest from both academic and industrial communities in analyzing graphical data. Examples span a variety of domains and applications such as node classification [3, 26] and link prediction [8, 32] in social networks, role prediction in protein-protein interaction networks [16], and prediction of information diffusion in social and citation networks [22].

One fundamental task of graph analysis is community detection, which aims to cluster nodes into multiple groups called communities. Each community is a set of nodes that are more closely connected to each other than to nodes in different communities. A community level description is able to capture important information about a graph’s global structure. Such a description is useful in many real-world applications, such as identifying users with similar interests in social networks [22] or proteins with similar functionality in biochemical networks [16]. Community detection has been extensively studied in the literature, and a number of methods have been proposed, including algorithmic approaches [1, 5] and probabilistic models [10, 20, 36, 37]. A classical approach to
detect communities is spectral clustering [34], which assumes that neighboring nodes tend to belong to the same communities and detects communities by finding the eigenvectors of the graph Laplacian.

Another important task of graph analysis is node representation learning, where nodes are described using low-dimensional features. Node representations effectively capture local graph structure and are often used as features for many prediction tasks. Modern methods for learning node embeddings [11, 24, 26] have proved effective on a variety of tasks such as node classification [3, 26], link prediction [8, 32] and graph visualization [27, 31].

Clustering, which captures the global structure of graphs, and learning node embeddings, which captures local structure, are typically studied separately. Clustering is often used for exploratory analysis, while generating node embeddings is often done for predictive analysis. However, these two tasks are very correlated and it may be beneficial to perform both tasks simultaneously. The intuition is that (1) node representations can be used as good features for community detection (e.g., through K-means) [4, 25, 29], and (2) the node community membership can provide good contexts for learning node representations [33]. However, how to leverage the relatedness of node clustering and node embedding in a unified framework for joint community detection and node representation learning is under-explored.

In this paper, we propose a novel probabilistic generative model called vGraph for joint community detection and node representation learning. vGraph assumes that each node \( v \) can be represented as a mixture of multiple communities and is described by a multinomial distribution over communities \( z \), i.e., \( p(z|v) \). Meanwhile, each community \( z \) is modeled as a distribution over the nodes \( v \), i.e., \( p(v|z) \). vGraph models the process of generating the neighbors for each node. Given a node \( u \), we first draw a community assignment \( z \) from \( p(z|u) \). This indicates which community the node is going to interact with. Given the community assignment \( z \), we generate an edge \( (u, v) \) by drawing another node \( v \) according to the community distribution \( p(v|z) \). Both the distributions \( p(z|v) \) and \( p(v|z) \) are parameterized by the low-dimensional representations of the nodes and communities. As a result, this approach allows the node representations and the communities to interact in a mutually beneficial way.

We also design a very effective algorithm for inference with backpropagation. We use variational inference for maximizing the lower-bound of the data likelihood. The Gumbel-Softmax [13] trick is leveraged since the community membership variables are discrete. Inspired by existing spectral clustering methods [6], we added a smoothness regularization term to the objective function of the variational inference routine to ensure that community membership of neighboring nodes is similar.

The whole framework of vGraph is very flexible and general. We also show that it can be easily extended to detect hierarchical communities.

In the experiment section, we show results on three tasks: overlapping community detection, non-overlapping community detection, and node classification– all using various real-world datasets. Our results show that vGraph is very competitive with existing state-of-the-art approaches for these tasks. We also present results on hierarchical community detection. Relevant source codes have been made public 1.

2 Related Work

Community Detection. Many community detection methods are based on matrix factorization techniques. Typically, these methods try to recover the node-community affiliation matrix by performing a low-rank decomposition of the graph adjacency matrix or other related matrices [17, 18, 32, 36]. These methods are not scalable due to the complexity of matrix factorization, and their performance is restricted by the capacity of the bi-linear models. Many other studies develop generative models for community detection. Their basic idea is to characterize the generation process of graphs and cast community detection as an inference problem [37, 38, 39]. However, the computational complexity of these methods is also high due to complicated inference. Compared with these approaches, vGraph is more scalable and can be efficiently optimized with backpropagation and Gumbel-Softmax [13, 19]. Additionally, vGraph is able to learn and leverage the node representations for community detection.

Node Representation Learning. The goal of node representation learning is to learn distributed representations of nodes in graphs so that nodes with similar local connectivity tend to have sim-

1https://github.com/fanyun-sun/vGraph
Figure 1: The diagram on the left represents the graphical model of vGraph and the diagram on the right represents the graphical model of the hierarchical extension. \( \phi_n \) is the embedding of node \( w_n \), \( \psi \) denotes the embedding of communities, and \( \varphi \) denotes the embeddings of nodes used in \( p(c|z) \). Refer to Eq. 2 and Eq. 3.

ilar representations. Some representative methods include DeepWalk \[24\], LINE \[26\], node2vec \[11\] and GraphRep \[3\]. Typically, these methods explore the local connectivity of each node by conducting random walks with either breadth-first search \[24\] or depth-first search \[26\]. Despite their effectiveness in a variety of applications, these methods mainly focus on preserving the local structure of graphs, therefore ignoring global community information. In vGraph, we address this limitation by treating the community label as a latent variable. This way, the community label can provide additional contextual information which enables the learned node representations to capture the global community information.

Framework for node representation learning and community detection. There exists previous work \[4, 14, 29, 30, 33\] that attempts to solve community detection and node representation learning jointly. However, their optimization process alternates between community assignment and node representation learning instead of simultaneously solving both tasks \[4, 30\]. Compared with these methods, vGraph is scalable and the optimization is done end-to-end.

Mixture Models. Methodologically, our method is related to mixture models, particularly topic models (e.g. PSLA \[12\] and LDA \[2\]). These methods simulate the generation of words in documents, in which topics are treated as latent variables, whereas we consider generating neighbors for each node in a graph, and the community acts as a latent variable. Compared with these methods, vGraph parameterizes the distributions with node and community embeddings, and all the parameters are trained with backpropagation.

3 Problem Definition

Graphs are ubiquitous in the real-world. Two fundamental tasks on graphs are community detection and learning node embeddings, which focus on global and local graph structures respectively and hence are naturally complementary. In this paper, we study jointly solving these two tasks. Let \( \mathcal{G} = (V, E) \) represent a graph, where \( V = \{v_1, \ldots, v_V\} \) is a set of vertices and \( E = \{e_{ij}\} \) is the set of edges. Traditional graph embedding aims to learn a node embedding \( \phi_i \in \mathbb{R}^d \) for each \( v_i \in V \) where \( d \) is predetermined. Community detection aims to extract the community membership \( \mathcal{F} \) for each node. Suppose there are \( K \) communities on the graph \( \mathcal{G} \), we can denote the community assignment of node \( v_i \) as \( \mathcal{F}(v_i) \subseteq \{1, \ldots, K\} \). We aim to jointly learn node embeddings \( \phi \) and community affiliation of vertices \( \mathcal{F} \).

4 Methodology

In this section, we introduce our generative approach vGraph, which aims at collaboratively learning node representations and detecting node communities. Our approach assumes that each node can belong to multiple communities representing different social contexts \[7\]. Each node should generate different neighbors under different social contexts. vGraph parameterizes the node-community distributions by introducing node and community embeddings. In this way, the node representations
can benefit from the detection of node communities. Similarly, the detected community assignment can in turn improve the node representations. Inspired by existing spectral clustering methods [6], we add a smoothness regularization term that encourages linked nodes to be in the same communities.

4.1 vGraph

vGraph models the generation of node neighbors. It assumes that each node can belong to multiple communities. For each node, different neighbors will be generated depending on the community context. Based on the above intuition, we introduce a prior distribution $p(z|w)$ for each node $w$ and a node distribution $p(c|z)$ for each community $z$. The generative process of each edge $(w, c)$ can be naturally characterized as follows: for node $w$, we first draw a community assignment $z \sim p(z|w)$, representing the social context of $w$ during the generation process. Then, the linked neighbor $c$ is generated based on the assignment $z$ through $c \sim p(c|z)$. Formally, this generation process can be formulated in a probabilistic way:

$$p(c|w) = \sum_z p(c|z)p(z|w).$$

(1)

vGraph parameterizes the distributions $p(z|w)$ and $p(c|z)$ by introducing a set of node embeddings and community embeddings. Note that different sets of node embeddings are used to parametrize the two distributions. Specifically, let $\phi_i$ denote the embedding of node $i$ used in the distribution $p(z|w)$, $\varphi_i$ denote the embedding of node $i$ used in $p(c|z)$, and $\psi_j$ denote the embedding of the $j$-th community. The prior distribution $p_{\varphi, \varphi}(z|w)$ and the node distribution conditioned on a community $p_{\varphi, \varphi}(c|z)$ are parameterized by two softmax models:

$$p_{\varphi, \varphi}(z = j|w) = \frac{\exp(\phi_w^T \varphi_j)}{\sum_{i=1}^K \exp(\phi_w^T \varphi_i)},$$

(2)

$$p_{\varphi, \varphi}(c = j) = \frac{\exp(\psi_j^T \varphi_c)}{\sum_{c' \in V} \exp(\psi_j^T \varphi_{c'})}.$$  

(3)

Calculating Eq. 3 can be expensive as it requires summation over all vertices. Thus, for large datasets we can employ negative sampling as done in LINE [26] using the following objective function:

$$\log \sigma(\varphi_c^T \cdot \psi_j) + \sum_{i=1}^K \mathbb{E}_{v \sim P_n(v)}[\log \sigma(-\varphi_v^T \cdot \psi_j)],$$

(4)

where $\sigma(x) = 1/(1 + \exp(-x))$, $P_n(v)$ is a noise distribution, and $K$ is the number of negative samples. This, combined with stochastic optimization, enables our model to be scalable.

To learn the parameters of vGraph, we try to maximize the log-likelihood of the observed edges, i.e., $\log p_{\varphi, \varphi}(c|w)$. Since directly optimizing this objective is intractable for large graphs, we instead optimize the following evidence lower bound (ELBO) [15]:

$$\mathcal{L} = E_{z \sim q(z|c, w)}[\log p_{\varphi, \varphi}(c|z)] - \text{KL}(q(z|c, w)||p_{\varphi, \varphi}(z|w))$$

(5)

where $q(z|c, w)$ is a variational distribution that approximates the true posterior distribution $p(z|c, w)$, and $\text{KL}(\cdot||\cdot)$ represents the Kullback-Leibler divergence between two distributions.

Specifically, we parametrize the variational distribution $q_{\varphi, \varphi}(z = j|w, c)$ with a neural network as follows:

$$q_{\varphi, \varphi}(z = j|w, c) = \frac{\exp((\phi_w \odot \varphi_c)^T \psi_j)}{\sum_{i=1}^K \exp((\phi_w \odot \varphi_c)^T \psi_i)}.$$  

(6)

where $\odot$ denotes element-wise multiplication. We chose element-wise multiplication because it is symmetric and it forces the representation of the edge to be dependent on both nodes.

The variational distribution $q(z|c, w)$ represents the community membership of the edge $(w, c)$. Based on this, we can easily approximate the community membership distribution of each node $w$, i.e., $p(z|w)$ by aggregating all its neighbors:

$$p(z|w) = \sum_c p(z, c|w) = \sum_c p(z|w, c)p(c|w) \approx \frac{1}{|N(w)|} \sum_{c \in N(w)} q(z|w, c),$$

(7)
where \( N(w) \) is the set of neighbors of node \( w \). To infer non-overlapping communities, we can simply take the \( \arg \max \) of \( p(z|w) \). However, when detecting overlapping communities instead of thresholding \( p(z|w) \) as in [14], we use

\[
\mathcal{F}(w) = \{ \arg \max_k q(z = k|w, c) \}_{c \in N(w)}.
\]

That is, we assign each edge to one community and then map the edge communities to node communities by gathering nodes incident to all edges within each edge community as in [1].

**Complexity.** Here we show the complexity of vGraph. Sampling an edge takes constant time, thus calculating Eq. (4) takes \( O(d(M + 1)) \) time, where \( M \) is the number of negative samples and \( d \) is the dimension of embeddings (the node embeddings and community embeddings have the same dimension). To calculate Eq. (6), it takes \( O(dK) \) time where \( K \) is the number of communities. Thus, an iteration with one sample takes \( O(\max(dM, dK)) \) time. In practice the number of updates required is proportional to the number of edges \( O(|\mathcal{E}|) \), thus the overall time complexity of vGraph is \( O(|\mathcal{E}|d\max(M, K)) \).

### 4.2 Community-smoothness Regularized Optimization

For optimization, we need to optimize the lower bound (5) w.r.t. the parameters in the variational distribution and the generative parameters. If \( z \) is continuous, the reparameterization trick [15] can be used. However, \( z \) is discrete in our case. In principle, we can still estimate the gradient using a score function estimator [9, 35]. However, the score function estimator suffers from a high variance, even when used with a control variate. Thus, we use the Gumbel-Softmax reparametrization [13, 19] to obtain gradients for the evidence lower bound. More specifically, we use the straight-through Gumbel-Softmax estimator [13].

A community can be defined as a group of nodes that are more similar to each other than to those outside the group [23]. For a non-attributed graph, two nodes are similar if they are connected and share similar neighbors. However, vGraph does not explicitly weight local connectivity in this way. To resolve this, inspired by existing spectral clustering studies [6], we augment our training objective with a smoothness regularization term that encourages the learned community distributions of linked nodes to be similar. Formally, the regularization term is given below:

\[
\mathcal{L}_{reg} = \lambda \sum_{(w,c) \in \mathcal{E}} \alpha_{w,c} \cdot d(p(z|c), p(z|w))
\]

where \( \lambda \) is a tunable hyperparameter, \( \alpha_{w,c} \) is a regularization weight, and \( d(\cdot, \cdot) \) is the distance between two distributions (squared difference in our experiments). Motivated by [25], we set \( \alpha_{w,c} \) to be the Jaccard’s coefficient of node \( w \) and \( c \), which is given by:

\[
\alpha_{w,c} = \frac{|N(w) \cap N(c)|}{|N(w) \cup N(c)|},
\]

where \( N(w) \) denotes the set of neighbors of \( w \). The intuition behind this is that \( \alpha_{w,c} \) serves as a similarity measure of how similar the neighbors are between two nodes. Jaccard’s coefficient is used for this metric and thus the higher the value of Jaccard’s coefficient, the more the two nodes are encouraged to have similar distribution over communities.

By combining the evidence lower bound and the smoothness regularization term, the entire loss function we aim to minimize is given below:

\[
\mathcal{L} = -E_{z \sim q_{\phi, \psi}(z|c, w)}[\log p_{\phi, \psi}(c|z)] + \text{KL}(q_{\phi, \psi}(z|c, w) \| p_{\phi, \psi}(z|w)) + \mathcal{L}_{reg}
\]

For large datasets, negative sampling can be used for the first term.

### 4.3 Hierarchical vGraph

One advantage of vGraph’s framework is that it is very general and can be naturally extended to detect hierarchical communities. In this case, suppose we are given a \( d \)-level tree and each node is associate with a community, the community assignment can be represented as a \( d \)-dimensional path vector \( \vec{z} = (z^{(1)}, z^{(2)}, ..., z^{(d)}) \), as shown in Fig. 1. Then, the generation process is formulated as
below: (1) a tree path \( \vec{z} \) is sampled from a prior distribution \( p_{\phi, \psi}(\vec{z} | w) \). (2) The context \( c \) is decoded from \( \vec{z} \) with \( p_{\psi, \phi}(c | \vec{z}) \). Under this model, the likelihood of the network is

\[
p_{\phi, \psi, \psi}(c | w) = \sum_{\vec{z}} p_{\phi, \psi}(c | \vec{z}) p_{\phi, \psi}(\vec{z} | w).
\]

(12)

At every node of the tree, there is an embedding vector associated with the community. Such a method is similar to the hierarchical softmax parameterization used in language models [21].

5 Experiments

As vGraph can detect both overlapping and non-overlapping communities, we evaluate it on three tasks: overlapping community detection, non-overlapping community detection, and vertex classification.

5.1 Datasets

We evaluate vGraph on 20 standard graph datasets. For non-overlapping community detection and node classification, we use 6 datasets: Citeseer, Cora, Cornell, Texas, Washington, and Wisconsin. For overlapping community detection, we use 14 datasets, including Facebook, Youtube, Amazon, Dblp, Coauthor-CS. For Youtube, Amazon, and Dblp, we consider subgraphs with the 5 largest ground-truth communities due to the runtime of baseline methods. To demonstrate the scalability of our method, we additionally include visualization results on a large dataset – Dblp-full. Dataset statistics are provided in Table 1. More details about the datasets is provided in Appendix A.

5.2 Evaluation Metric

For overlapping community detection, we use F1-Score and Jaccard Similarity to measure the performance of the detected communities as in [37, 18]. For non-overlapping community detection, we use Normalized Mutual Information (NMI) [28] and Modularity. Note that Modularity does not utilize ground truth data. For node classification, Micro-F1 and Macro-F1 are used.

5.3 Comparative Methods

For overlapping community detection, we choose four competitive baselines: BigCLAM [36], a nonnegative matrix factorization approach based on the Bernoulli-Poisson link that only considers the graph structure; CESNA [37], an extension of BigCLAM, that additionally models the generative process for node attributes; Circles [20], a generative model of edges w.r.t. attribute similarity to detect communities; and SVI [10], a Bayesian model for graphs with overlapping communities that uses a mixed-membership stochastic blockmodel.

To evaluate node embedding and non-overlapping community detection, we compare our method with the five baselines: MF [32], which represents each vertex with a low-dimensional vector obtained through factoring the adjacency matrix; DeepWalk [24], a method that adopts truncated random walk and Skip-Gram to learn vertex embeddings; LINE [26], which aims to preserve the first-order and second-order proximity among vertices in the graph; Node2vec [11], which adopts biased random walk and Skip-Gram to learn vertex embeddings; and ComE [4], which uses a Gaussian mixture model to learn an embedding and clustering jointly using random walk features.

5.4 Experiment Configuration

For all baseline methods, we use the implementations provided by their authors and use the default parameters. For methods that only output representations of vertices, we apply K-means to the
learned embeddings to get non-overlapping communities. Results report are averaged over 5 runs. **No node attributes** are used in all our experiments. We generate node attributes using node degree features for those methods that require node attributes such as CESNA [37] and Circles [20]. It is hard to compare the quality of community results when the numbers of communities are different for different methods. Therefore, we set the number of communities to be detected, K, as the number of ground-truth communities for all methods, as in [18]. For vGraph, we use full-batch training when the dataset is small enough. Otherwise, we use stochastic training with a batch size of 5000 or 10000 edges. The initial learning rate is set to 0.05 and is decayed by 0.99 after every 100 iterations. We use the Adam optimizer and we trained for 5000 iterations. When smoothness regularization is used, λ is set to 100. For community detection, the model with the lowest loss is chosen. For node classification, we evaluate node embeddings after 1000 iterations of training. The dimension of node embeddings is set to 128 in all experiments for all methods. For the node classification task, we randomly select 70% of the labels for training and use the rest for testing.

5.5 Results

Table 2 shows the results on overlapping community detection. Some of the methods are not very scalable and cannot obtain results in 24 hours on some larger datasets. Compared with these studies, vGraph outperforms all baseline methods in 11 out of 14 datasets in terms of F1-score or Jaccard Similarity, as it is able to leverage useful representations at node level. Moreover, vGraph is also very efficient on these datasets, since we use employ variational inference and parameterize the model with node and community embeddings. By adding the smoothness regularization term (vGraph+), we see a farther increase performance, which shows that our method can be combined with concepts from traditional community detection methods.

The results for non-overlapping community detection are presented in Table 3. vGraph outperforms all conventional node embeddings + K-Means in 4 out of 6 datasets in terms of NMI and outperforms all 6 in terms of modularity. ComE, another framework that jointly solves node embedding and community detection, also generally performs better than other node embedding methods + K-Means. This supports our claim that learning these two tasks collaboratively instead of sequentially can further enhance performance. Compare to ComE, vGraph performs better in 4 out of 6 datasets in terms of NMI and 5 out of 6 datasets in terms of modularity. This shows that vGraph can also outperform frameworks that learn node representations and communities together.

Table 4 shows the result for the node classification task. vGraph significantly outperforms all the baseline methods in 9 out of 12 datasets. The reason is that most baseline methods only consider the local graph information without modeling the global semantics. vGraph solves this problem by representing node embeddings as a mixture of communities to incorporate global context.
Table 3: Evaluation (in terms of NMI and Modularity) on networks with non-overlapping ground-truth communities.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>NMI MF</th>
<th>deepwalk</th>
<th>LINE</th>
<th>node2vec</th>
<th>ComE</th>
<th>vGraph</th>
<th>NMI MF</th>
<th>deepwalk</th>
<th>LINE</th>
<th>node2vec</th>
<th>ComE</th>
<th>vGraph</th>
</tr>
</thead>
<tbody>
<tr>
<td>cornell</td>
<td>0.0632</td>
<td>0.0789</td>
<td>0.0697</td>
<td>0.0712</td>
<td>0.0732</td>
<td>0.0803</td>
<td>0.4220</td>
<td>0.4055</td>
<td>0.2372</td>
<td>0.3793</td>
<td>0.5748</td>
<td>0.5792</td>
</tr>
<tr>
<td>washington</td>
<td>0.0599</td>
<td>0.0752</td>
<td>0.0910</td>
<td>0.0538</td>
<td>0.0504</td>
<td>0.0649</td>
<td>0.3679</td>
<td>0.1841</td>
<td>0.1655</td>
<td>0.4311</td>
<td>0.4862</td>
<td>0.5169</td>
</tr>
<tr>
<td>wisconsin</td>
<td>0.0530</td>
<td>0.0759</td>
<td>0.0680</td>
<td>0.0749</td>
<td>0.0689</td>
<td>0.0852</td>
<td>0.3892</td>
<td>0.3384</td>
<td>0.1651</td>
<td>0.5338</td>
<td>0.5500</td>
<td>0.5706</td>
</tr>
<tr>
<td>citeseer</td>
<td>0.0552</td>
<td>0.1190</td>
<td>0.0340</td>
<td>0.1592</td>
<td>0.2499</td>
<td>0.1030</td>
<td>0.6963</td>
<td>0.6819</td>
<td>0.4014</td>
<td>0.4657</td>
<td>0.7324</td>
<td>0.7711</td>
</tr>
</tbody>
</table>

Table 4: Results of node classification on 6 datasets.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Macro-F1 MF</th>
<th>DeepWalk</th>
<th>LINE</th>
<th>Node2Vec</th>
<th>ComE</th>
<th>vGraph</th>
<th>Micro-F1 MF</th>
<th>DeepWalk</th>
<th>LINE</th>
<th>Node2Vec</th>
<th>ComE</th>
<th>vGraph</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cornell</td>
<td>13.05</td>
<td>22.79</td>
<td>23.78</td>
<td>20.90</td>
<td>19.86</td>
<td>20.76</td>
<td>15.45</td>
<td>33.05</td>
<td>37.29</td>
<td>34.78</td>
<td>35.00</td>
<td>34.78</td>
</tr>
<tr>
<td>Texas</td>
<td>8.74</td>
<td>21.32</td>
<td>16.33</td>
<td>14.95</td>
<td>15.46</td>
<td>26.00</td>
<td>14.03</td>
<td>40.35</td>
<td>27.19</td>
<td>25.44</td>
<td>33.33</td>
<td>47.37</td>
</tr>
<tr>
<td>Washington</td>
<td>15.88</td>
<td>18.45</td>
<td>13.99</td>
<td>21.23</td>
<td>15.80</td>
<td>30.36</td>
<td>15.94</td>
<td>34.06</td>
<td>25.36</td>
<td>28.99</td>
<td>33.33</td>
<td>34.78</td>
</tr>
<tr>
<td>cora</td>
<td>14.77</td>
<td>23.44</td>
<td>19.06</td>
<td>18.47</td>
<td>14.63</td>
<td>29.91</td>
<td>18.75</td>
<td>38.75</td>
<td>28.12</td>
<td>25.00</td>
<td>32.50</td>
<td>35.00</td>
</tr>
</tbody>
</table>

5.6 Visualization

In order to gain more insight, we present visualizations of the facebook107 dataset in Fig. 2(a). To demonstrate that our model can be applied to large networks, we present results of vGraph on a co-authorship network with around 100,000 nodes and 330,000 edges in Fig. 2(b). More visualizations are available in appendix B. We can observe that the community structure, or “social context”, is reflected in the corresponding node embedding (node positions in both visualizations are determined by t-SNE of the node embeddings). To demonstrate the hierarchical extension of our model, we visualize a subset of the co-authorship dataset in Fig. 3. We visualize the first-tier communities and second-tier communities in panel (a) and (b) respectively. We can observe that the second-tier communities grouped under the same first-tier communities interact more with themselves than they do with other second-tier communities.

6 Conclusion

In this paper, we proposed vGraph, a method that performs overlapping (and non-overlapping) community detection and learns node and community embeddings at the same time. vGraph casts the generation of edges in a graph as an inference problem. To encourage collaborations between community detection and node representation learning, we assume that each node can be represented by a mixture of communities, and each community is defined as a multinomial distribution over nodes. We also design a smoothness regularizer in the latent space to encourage neighboring nodes to
Figure 3: We visualize the result on a subset of Dblp dataset using two-level hierarchical vGraph. The coordinates of the nodes are determined by t-SNE of the node embeddings. In panel (a) we visualize the first-tier communities. In panel (b), we visualize the second-tier communities. In panel (c) we show the corresponding hierarchical tree structure.

Empirical evaluation on 20 different benchmark datasets demonstrates the effectiveness of the proposed method on both tasks compared to competitive baselines. Furthermore, our model is also readily extendable to detect hierarchical communities.

Acknowledgments

This project is supported by the Natural Sciences and Engineering Research Council of Canada, as well as the Canada CIFAR AI Chair Program.

References


