

# RaftGP: Random Fast Graph Partitioning

Yu Gao<sup>\*†</sup> Meng Qin<sup>\*†§</sup> Yibin Ding<sup>†</sup> Li Zeng<sup>‡</sup> Chaorui Zhang<sup>†</sup> Weixi Zhang<sup>†</sup> Wei Han<sup>†</sup> Rongqian Zhao<sup>‡</sup> Bo Bai<sup>†</sup>

<sup>†</sup>Theory Lab, Central Research Institute, 2012 Labs, Huawei Technologies Co., Ltd.

<sup>‡</sup>Algorithm & Technology Development Department, Global Technical Service Department, Huawei Technologies Co., Ltd.

<sup>§</sup>Department of Computer Science & Engineering, Hong Kong University of Science & Technology

**Abstract**—Graph partitioning (GP), a.k.a. community detection, is a classic problem that divides the node set of a graph into densely-connected blocks. Following prior work on the IEEE HPEC Graph Challenge benchmark and recent advances in graph machine learning, we propose a novel RANdom FasT Graph Partitioning (RaftGP) method based on an efficient graph embedding scheme. It uses the Gaussian random projection to extract community-preserving features from classic GP objectives. These features are fed into a graph neural network (GNN) to derive low-dimensional node embeddings. Surprisingly, our experiments demonstrate that a randomly initialized GNN even without training is enough for RaftGP to derive informative community-preserving embeddings and support high-quality GP. To enable the derived embeddings to tackle GP, we introduce a hierarchical model selection algorithm that simultaneously determines the number of blocks and the corresponding GP result. We evaluate RaftGP on the Graph Challenge benchmark and compare the performance with five baselines, where our method can achieve a better trade-off between quality and efficiency. In particular, compared to the baseline algorithm [1] of the IEEE HPEC Graph Challenge, our method is 6.68x – 23.9x faster on graphs with 1E3 – 5E4 nodes and at least 64.5x faster on larger (1E5 node) graphs on which the baseline takes more than 1E4 seconds. Our method achieves better accuracy on all test cases. We also develop a new graph generator to address some limitations of the original generator in the benchmark.

**Index Terms**—Graph Partitioning, Graph Clustering, Community Detection, Graph Embedding

## I. INTRODUCTION

For many real-world complex systems (e.g., social, communication, and biological networks), graphs serve as a generic abstraction of system entities and their relations in terms of nodes and edges. Graph partitioning (GP), a.k.a. disjoint community detection, is a classic problem that divides the node set of a graph into disjoint blocks (communities) with dense linkages distinct from other blocks. The partitioning of blocks (communities) can reveal the structures and functions of a graph [2]–[4] while supporting advanced applications (e.g., parallel task assignment [5], wireless network decomposition [6], and traffic classification [7]).

GP on large graphs is usually challenging because it can be formulated as several NP-hard combinatorial optimization objectives. Previous studies on the IEEE HPEC Graph Challenge benchmark [8] have provided a series of solutions to alleviating the NP-hard challenge of GP. Knyazev et al. [9] solved the eigen decomposition of graph Laplacians via

locally optimal block preconditioned conjugate gradient and accelerated the spectral clustering algorithm. Durbeck et al. [10] explored the ability of graph summarization techniques to preserve the underlying community structures of graphs. Uppal et al. [11] and Wanye et al. [12] proposed a top-down divide-and-conquer algorithm and a sampling strategy to speed up the Bayes inference of the stochastic block model (SBM).

Different from the aforementioned studies that focus on the engineering improvements of existing GP techniques (e.g., spectral clustering and Bayesian SBM), we propose a novel RANdom FasT Graph Partitioning (RaftGP) method following recent advances in graph machine learning [13]. It adopts an efficient community-preserving graph embedding scheme including (i) random feature extraction, (ii) random embedding derivation, and (iii) hierarchical model selection.

We first use the efficient Gaussian random projection [14] to extract two sets of low-dimensional features from classic GP objectives of normalized cut (NCut) minimization [15] and modularity maximization [16], which induces two variants of RaftGP. The extracted features are believed to encode the community structures of graphs because the random projection can preserve the geometric structures (e.g., in terms of distances) between inputs (i.e., components about graph structures in the two GP objectives) with rigorous theoretical guarantees [14].

To derive low-dimensional embeddings, we then feed the extracted features into a graph neural network (GNN), which further enhances their abilities to preserve community structures via multi-hop feature aggregation while reducing the dimensionality. In contrast to existing GNN-based methods [17]–[19] that require time-consuming training, we get inspiration from the efficient random projection and consider an extreme design of RaftGP, where we do not apply any training procedures to GNN. Surprisingly, our experiments demonstrate that one feedforward propagation of a randomly initialized GNN, even without training, is enough for RaftGP to derive informative community-preserving embeddings.

To enable the embedding scheme to tackle GP, we introduce a hierarchical model selection algorithm that simultaneously determines the proper number of blocks and corresponding block membership (i.e., the GP result). It recursively partitions the node set into two subsets by applying  $K$ Means to the derived node embeddings, where a novel *local modularity* metric is used to stop the recursion.

We follow the experiment settings of the IEEE HPEC Graph Challenge to evaluate RaftGP and compare the performance

\*The first two authors contributed equally.

with five baselines. However, we notice some limitations of the original data generators in the benchmark. To address these limitations, we develop a new algorithm that can generate graphs from the exact SBM distribution in nearly linear time.

## II. PROBLEM STATEMENTS & PRELIMINARIES

In this study, we consider undirected unweighted graphs. In general, a graph can be represented as  $G = (V, E)$ , where  $V = \{v_1, v_2, \dots, v_N\}$  and  $E = \{(v_i, v_j) | v_i, v_j \in V\}$  are the sets of nodes and edges. The topology structure of a graph can be described by an adjacency matrix  $\mathbf{A} \in \{0, 1\}^{N \times N}$ , where  $\mathbf{A}_{ij} = \mathbf{A}_{ji} = 1$  if  $(v_i, v_j) \in E$  and  $\mathbf{A}_{ij} = \mathbf{A}_{ji} = 0$  otherwise. We study the following *graph partitioning* (GP) problem.

**Definition 1 (Graph Partitioning, GP).** Given a graph  $G$ , GP (a.k.a. disjoint community detection) aims to partition the node set  $V$  into  $K$  disjoint subsets (i.e., blocks or communities)  $C = \{C_1, \dots, C_K\}$  such that (i) within each block the linkage is dense but (ii) between different blocks the linkage is relatively loose. In this study, we assume that  $K$  is not given and should be determined by a model selection procedure.

Our RaftGP method follows a novel community-preserving embedding scheme based on some classic combinatorial optimization objectives of GP (e.g., NCut minimization [15] and modularity maximization [16]).

**Definition 2 (Graph Embedding).** Given a graph  $G$ , graph embedding (a.k.a. graph representation learning) learns a function  $f: V \mapsto \mathbb{R}^d$  that maps each node  $v_i$  to a low-dimensional vector representation (a.k.a. embedding)  $\mathbf{z}_i \in \mathbb{R}^d$  ( $d \ll N$ ). The derived embeddings  $\{\mathbf{z}_i\}$  are expected to preserve some major properties (e.g., community structures) of graph topology. For instance, nodes  $(v_i, v_j)$  with similar properties (e.g., in the same block) should have close embeddings  $(\mathbf{z}_i, \mathbf{z}_j)$  (e.g., in terms of distance). It is a unified framework that can support various graph inference tasks by combining  $\{\mathbf{z}_i\}$  with specific downstream modules. For GP, one can apply the  $K$ Means clustering algorithm to  $\{\mathbf{z}_i\}$ , which derives the GP result  $C$  w.r.t. the determined number of blocks  $K$ .

**Definition 3 (NCut Minimization).** Given a graph  $G$  and the number of blocks  $K$  to be partitioned, NCut minimization aims to derive a GP result  $C$  that minimizes the NCut metric:

$$\min_C \text{NCut}(G, K) \equiv \frac{1}{2} \sum_{r=1}^K [\text{cut}(C_r, \bar{C}_r) / \text{vol}(C_r)], \quad (1)$$

where  $\bar{C}_r = V - C_r$  denotes the complementary set of  $C_r$ ;  $\text{cut}(C_r, \bar{C}_r) = \sum_{v_i \in C_r, v_j \in \bar{C}_r} \mathbf{A}_{ij}$  is the cut between  $C_r$  and  $\bar{C}_r$ ;  $\text{vol}(C_r) = \sum_{v_i \in C_r, v_j \in V} \mathbf{A}_{ij}$  is the volume of  $C_r$ . (1) can be further rewritten into the following matrix form:

$$\min_{\mathbf{H}} \text{tr}(\mathbf{H}^T \mathbf{L} \mathbf{H}) \text{ s.t. } \mathbf{H}_{ir} = \begin{cases} [\text{deg}(v_i) / \text{vol}(C_r)]^{0.5}, & v_i \in C_r \\ 0, & \text{otherwise} \end{cases}, \quad (2)$$

where  $\mathbf{L} \equiv \mathbf{I}_N - \mathbf{D}^{-0.5} \mathbf{A} \mathbf{D}^{-0.5}$  is the normalized Laplacian matrix of  $\mathbf{A}$ ;  $\mathbf{D} \equiv \text{diag}(\text{deg}(v_1), \dots, \text{deg}(v_N))$  is defined as a degree diagonal matrix with  $\text{deg}(v_i)$  as the degree of node  $v_i$ ;  $\mathbf{I}_N$  is an  $N$ -dimensional identity matrix;  $\mathbf{H}$  consists of the scaled block membership indicators.

**Definition 4 (Modularity maximization).** For the given  $G$  and  $K$ , modularity maximization aims to derive a GP result  $C$  that maximizes the following modularity metric:

$$\max_C \text{Mod}(C, K) \equiv \frac{1}{2e} \sum_{r=1}^K \sum_{v_i, v_j \in C_r} [\mathbf{A}_{ij} - \frac{\text{deg}(v_i)\text{deg}(v_j)}{2e}], \quad (3)$$

where  $\text{deg}(v_i)$  is the degree of node  $v_i$ ;  $e$  is the number of edges. Similarly, (3) can also be rewritten into a matrix form:

$$\min_{\mathbf{H}} -\text{tr}(\mathbf{H}^T \mathbf{Q} \mathbf{H}) \text{ s.t. } \mathbf{H}_{ir} = \begin{cases} 1, & v_i \in C_r \\ 0, & \text{otherwise} \end{cases}, \quad (4)$$

where  $\mathbf{Q} \in \mathbb{R}^{N \times N}$  is defined as the modularity matrix with  $\mathbf{Q}_{ij} = \mathbf{A}_{ij} - \text{deg}(v_i)\text{deg}(v_j)/(2e)$ ;  $\mathbf{H}$  consists of the block membership indicators.

Although NCut minimization and modularity maximization are NP-hard and need a pre-set number of blocks  $K$ , our method does not directly solve these combinatorial optimization problems. Instead, we only focus on how to efficiently extract informative community-preserving features from (2) and (4), which is detailed later in Section III-A.

The benchmarks of the IEEE HPEC Graph Challenge contain graphs drawn from the SBM with different settings. One baseline algorithm [1] is also based on maximizing the posterior probability of the SBM.

**Definition 5 (Stochastic Block Model, SBM).** Let  $\theta \in \mathbb{R}_+^N$  and  $\mathbf{c} \in [K]^N$  be the degree correction vector and the block assignment vector. Let  $\Omega \in \mathbb{R}_+^{K \times K}$  be the block interaction matrix. The SBM independently generates an edge for each pair of nodes  $(v_i, v_j)$  (i.e., each entry of adjacency matrix  $\mathbf{A}$ ) via a Poisson distribution  $\mathbf{A}_{ij} \sim \text{Pois}(\lambda_{ij} \equiv \theta_i \theta_j \Omega_{c_i c_j})$ .

## III. METHODOLOGY

To alleviate the NP-hard challenge of GP, we propose a novel RaftGP method following an efficient community-preserving embedding scheme. In this section, we elaborate on the three major procedures of RaftGP, which are (i) *random feature extraction*, (ii) *random embedding derivation*, and (iii) *hierarchical model selection*.

### A. Random Feature Extraction

We first extract informative features from the key components regarding graph structures in classic GP objectives. For NCut minimization defined in (2),  $\mathbf{M} \equiv \mathbf{D}^{-0.5} \mathbf{A} \mathbf{D}^{-0.5}$  in the Laplacian matrix  $\mathbf{L}$  is used to describe the graph topology. For modularity maximization, the modularity matrix  $\mathbf{Q}$  in (4) is the primary component regarding graph structures.

In particular,  $\{\mathbf{M}, \mathbf{Q}\}$  can be considered as a reweighting of the original neighbor structures described by the adjacency matrix  $\mathbf{A}$ , where the similarity between  $(\mathbf{M}_{i,:}, \mathbf{M}_{j,:})$  (or  $(\mathbf{Q}_{i,:}, \mathbf{Q}_{j,:})$ ) indicates the neighbor similarity between nodes  $(v_i, v_j)$ . In general, nodes  $(v_i, v_j)$  with a higher neighbor similarity (i.e., denser local linkage) are more likely to be partitioned into the same block. Hence, we believe that  $\{\mathbf{M}, \mathbf{Q}\}$  encode key characteristics about the community structures of a graph. Tian et al. [20] and Yang et al. [21] have also validated

the potential of  $\mathbf{M}$  and  $\mathbf{Q}$  to derive community-preserving embeddings via deep auto-encoders. Instead of using auto-encoders, we propose a more efficient strategy to extract community-preserving features (denoted by  $\mathbf{Y} \in \mathbb{R}^{N \times L}$ ) from  $\{\mathbf{M}, \mathbf{Q}\}$ , with the feature dimensionality  $L \ll N$ .

Since most real-world graphs have sparse topology,  $\mathbf{M}$  can be treated as a sparse matrix. However,  $\mathbf{Q}$  is still a dense matrix. To fully utilize the sparsity of topology, we introduce a reduced modularity matrix  $\tilde{\mathbf{Q}} \in \mathbb{R}^{N \times N}$ , where  $\tilde{\mathbf{Q}}_{ij} = \mathbf{Q}_{ij}$  if  $(v_i, v_j) \in E$  and  $\tilde{\mathbf{Q}}_{ij} = 0$  otherwise. Although the reduction of  $\mathbf{Q}$  may lose some information, our experiments demonstrated that  $\tilde{\mathbf{Q}}$  is enough to derive informative community-preserving embeddings to support high-quality GP.

Let  $\mathbf{X} \in \{\mathbf{M}, \tilde{\mathbf{Q}}\}$  be the structural component from NCut minimization or modularity maximization, which induces two variants of RaftGP. We extract the features  $\mathbf{Y}$  via

$$\mathbf{Y} = \mathbf{X}\Theta \text{ with } \Theta \in \mathbb{R}^{N \times L}, \Theta_{ir} \sim \mathcal{N}(0, L^{-0.5}), \quad (5)$$

where  $\mathbf{Y}_{i,:}$  is the extracted feature of node  $v_i$ . Concretely, (5) is the Gaussian random projection [14] of  $\mathbf{X} \in \mathbb{R}^{N \times N}$ , an efficient dimension reduction technique that can preserve the geometry structures (e.g., in terms of distances) of input features in finite-dimensional  $l_1$  [22] and  $l_2$  [23] spaces with rigorous theoretical guarantees. For instance, nodes  $(v_i, v_j)$  with close  $(\mathbf{X}_{i,:}, \mathbf{X}_{j,:})$  are guaranteed to have close  $(\mathbf{Y}_{i,:}, \mathbf{Y}_{j,:})$ .

Since  $\mathbf{X}$  is sparse, the time complexity of random projection is  $O(|E|L)$  by using the sparse-dense matrix multiplication. Given a graph  $G$ , the complexities to compute  $\mathbf{M}$  and  $\tilde{\mathbf{Q}}$  are both  $O(|E|)$ . In summary, the overall complexity of the random feature extraction step is  $O(|E|L)$ .

### B. Random Embedding Derivation

Although the extracted features (described by  $\mathbf{Y} \in \mathbb{R}^{N \times L}$ ) have the initial ability to encode characteristics about community structures, we derive the final community-preserving embeddings (denoted by  $\mathbf{Z} \in \mathbb{R}^{N \times d}$ ) by feeding  $\mathbf{Y}$  into a multi-layer GNN, which further enhances the extracted features via the feature aggregation of GNN while reducing the feature dimensionality to  $d < L$ . We adopt GCN [24] to build the multi-layer GNN because it is easy to implement and has a low complexity of feature aggregation.

Let  $\mathbf{Z}^{(k-1)} \in \mathbb{R}^{N \times L_{k-1}}$  and  $\mathbf{Z}^{(k)} \in \mathbb{R}^{N \times L_k}$  be the input and output of the  $k$ -th GNN layer, where  $\mathbf{Z}^{(0)} = \mathbf{Y}$ ;  $L_{k-1}$  and  $L_k$  are the input and output feature dimensionality. The  $k$ -th GNN layer can be described as

$$\mathbf{Z}^{(k)} = f_{\text{act}}(\hat{\mathbf{D}}^{-0.5} \hat{\mathbf{A}} \hat{\mathbf{D}}^{-0.5} \mathbf{Z}^{(k-1)} \mathbf{W}^{(k)}), \quad (6)$$

where  $\mathbf{Z}_{i,:}^{(k)}$  is the (intermediate) representation of node  $v_i$ ;  $f_{\text{act}}(\cdot)$  is an activation function to be specified (e.g., tanh in our implementation);  $\hat{\mathbf{A}} \equiv \mathbf{A} + \mathbf{I}_N$  is the adjacency matrix with self-edges;  $\hat{\mathbf{D}}$  is the degree diagonal matrix of  $\hat{\mathbf{A}}$ ;  $\mathbf{W}^{(k)} \in \mathbb{R}^{L_{k-1} \times L_k}$  is a trainable parameter. In particular,  $\mathbf{Z}_{i,:}^{(k)}$  is the non-linear aggregation (i.e., weighted mean) of features w.r.t.  $\{v_i\} \cup \mathcal{N}(v_i)$  from the previous layer, with  $\mathcal{N}(v_i)$  as the neighbor set of  $v_i$ . The feature aggregation forces nodes  $(v_i, v_j)$  with similar neighbor sets  $(\mathcal{N}(v_i), \mathcal{N}(v_j))$

(i.e., dense local linkage) to have similar features  $(\mathbf{Z}_{i,:}^{(k)}, \mathbf{Z}_{j,:}^{(k)})$ , thus enhancing the ability to encode characteristics regarding community structures. The multi-layer structure can further extend the feature aggregation to the local topology induced by multi-hop neighbors.

Before feeding  $\mathbf{Z}^{(k)}$  into the next GNN layer, we recommend conducting the row-wise  $l_2$ -normalization via  $\mathbf{Z}_{i,:}^{(k)} \leftarrow \mathbf{Z}_{i,:}^{(k)} / |\mathbf{Z}_{i,:}^{(k)}|_2$ , which controls the scale of the (intermediate) representation of each node  $v_i$ . We treat the normalized representations from the last GNN layer (denoted by  $\mathbf{Z} \in \mathbb{R}^{N \times d}$ ) as the final community-preserving embeddings. In the rest of this paper, we use  $\mathbf{z}_i \in \mathbb{R}^d$  to represent the embedding of  $v_i$ .

The powerful inference abilities of some state-of-the-art GNN-based GP methods (e.g., *ClusterNet* [17], *LGNN* [18], and *ICD* [19]) rely on the offline training of model parameters (e.g.,  $\{\mathbf{W}^{(k)}\}$ ) via the time-consuming gradient descent.

Inspired by the efficient random projection described in (5), we consider an extreme design of RaftGP, where we do not apply any training procedures to the multi-layer GNN. Our experiments demonstrate that one feedforward propagation through a randomly initialized GNN without any training is enough for RaftGP to derive informative community-preserving embeddings and support high-quality GP. A reasonable interpretation is that each GNN layer described in (6) can be considered as a special random projection similar to (5). In the  $k$ -th layer, the geometric properties (e.g., relative distances) between the aggregated representations (i.e.,  $\hat{\mathbf{D}}^{-0.5} \hat{\mathbf{A}} \hat{\mathbf{D}}^{-0.5} \mathbf{Z}^{(k-1)}$ ) can be preserved with theoretical guarantees, even though it is multiplied by a random matrix  $\mathbf{W}^{(k)}$  and mapped to another space with lower dimensionality. The nonlinear activation and  $l_2$ -normalization will also not largely change the geometric properties.

For the GNN feature aggregation in (6), one can treat  $\hat{\mathbf{D}}^{-0.5} \hat{\mathbf{A}} \hat{\mathbf{D}}^{-0.5}$  as a sparse matrix and use the efficient sparse-dense matrix multiplication for implementation. In this setting, the complexity of GNN feedforward propagation is  $O(|E|L + NLL_1 + |E|L_1 + NL_1L_2 + \dots) = O(|E|L + NLL_1)$ , where we usually set  $L_k > L_{k+1}$ . Assume that  $\{L, L_1\}$  have the same magnitude with the embedding dimensionality  $d$ . The complexity of the embedding derivation step is  $O(|E|d + Nd^2)$ .

### C. Hierarchical Model Selection

Graph embedding is a unified framework that can support various inference tasks by combining the derived embeddings  $\{\mathbf{z}_i\}$  with specific downstream modules. To enable RaftGP to support the GP task stated in Definition 1, the model selection problem (i.e., determining a proper number of blocks  $K$ ) is critical. We introduce an efficient hierarchical model selection algorithm based on a novel *local modularity* metric.

**Definition 6 (Local Modularity).** Let  $U \subseteq V$  be a subset of nodes in graph  $G = (V, E)$ . Let  $\deg_G(v)$  be the degree of node  $v$  in  $G$ . Suppose  $U$  is partitioned into  $K$  disjoint blocks  $(U_1, \dots, U_K)$ . We define the local modularity metric as

$$\text{LMod}(U_1, \dots, U_K; G) \equiv \sum_{r=1}^K \left[ \frac{m_r}{e} - \left( \frac{\bar{m}_r}{2e} \right)^2 \right], \quad (7)$$

---

**Algorithm 1:** Hierarchical Model Selection

---

```
1 global variables
2   Graph  $G = (V, E)$ ; Derived node embeddings  $\{z_i\}$ ;
   Current GP result  $C$  (with  $C \leftarrow \emptyset$  at the beginning)
3 procedure PARTITION( $U, G$ )
4    $m_1 \leftarrow \text{LMod}(U; G)$  via (7)
5   Apply  $K$ Means to  $\{z_i | v_i \in U\}$  with  $K = 2$  and get two
   subsets  $\{U_1, U_2\}$ 
6    $m_2 \leftarrow \text{LMod}(U_1, U_2; G)$  via (7)
7   if  $|U_1| \leq \epsilon$  or  $|U_2| \leq \epsilon$  or  $m_1 > m_2$  then
8     Add  $U$  as a new block to  $C$ 
9   else
10    PARTITION( $U_1, G$ )
11    PARTITION( $U_2, G$ )
```

---

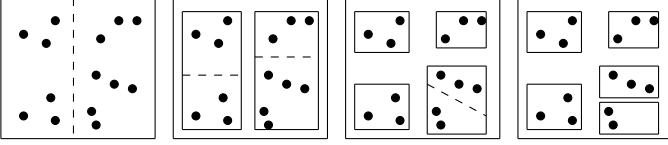


Fig. 1: The hierarchical model selection procedure on a sample embedding. Dots represent vertex embeddings. Rectangles represent final or intermediate blocks. Dashed lines represent partitions returned by the  $K$ means algorithm.

where  $m_r$  is the number of intra-block edges of  $U_r$ ;  $\bar{m}_r = \sum_{v \in U_r} \deg_G(v)$  is total degree of nodes in  $U_r$ ;  $2e = \sum_{r=1}^K \bar{m}_r$  is the total degree of all nodes in  $U$ .

Different from the standard modularity on  $G[U]$  which considers only edges induced by  $U$ , local modularity also takes account of the edges from  $U$  to  $V - U$ . When partitioning a subgraph, local modularity keeps its interaction with the remaining graph and thus ensures that we do not further partition a true block into sub-blocks.

Our hierarchical model selection procedure based on the local modularity metric is summarized in Algorithm 1. We demonstrate it by an example (Fig. 1). For a set of nodes  $U \subseteq V$  to be partitioned, we first compute the local modularity  $m_1$  that treats  $U$  as a single block (i.e., Line 4). The  $K$ Means clustering algorithm is then applied to embeddings  $\{z_i | v_i \in U\}$  (i.e., Line 5), which tries to divide  $U$  into two blocks  $\{U_1, U_2\}$  based on the distances between embeddings. One can also compute the local modularity  $m_2$  w.r.t. the partition of  $\{U_1, U_2\}$  (i.e., Line 6). We then introduce a heuristic stopping rule based on the sizes of  $\{U_1, U_2\}$  and values of  $\{m_1, m_2\}$ . Concretely, we adopt the set  $U$  as a new block, if  $|U_1|$  or  $|U_2|$  are less than or equal to a threshold  $\epsilon$  (i.e.,  $\epsilon = 5$  in our implementation) or  $m_1 > m_2$  (i.e., Lines 7,8). Otherwise, we repeat the aforementioned procedures to further partition  $U_1$  and  $U_2$  (i.e., Lines 10, 11).

In particular, the stopping rule in Line 7 can effectively avoid partitioning  $U$  into too small or imbalanced blocks (e.g., with  $|U_1| < \epsilon$  or  $|U_2| < \epsilon$ ). Consistent with the modularity maximization objective, it also tries to avoid the decrease of

modularity after the partitioning.

The complexity of this algorithm depends on the quality of the embedding. If the embedding is of good quality so that the  $K$ Means algorithm always partitions the node set  $U$  into two balanced sets, we need to run the  $K$ Means algorithm on a total of  $O(N \log N)$  nodes. The complexity in this case is  $O(NT \log N)$  where  $T$  is the number of iterations in  $K$ Means. In the worst case, the set  $U$  is always partitioned into a single block and the remaining blocks. In this case, the  $K$ Means algorithm is run on  $O(KN)$  nodes where  $K$  is the number of blocks. The complexity in this case is  $O(KTN)$ .

#### D. Extension to Streaming Graphs

The aforementioned designs of RaftGP can be easily extended to streaming graphs. Let  $E'$  be the set of newly added edges in the streaming setting. Since the addition of  $E'$  only changes the degrees of at most  $2|E'|$  nodes, one can use an incremental strategy to update the input statistics  $\mathbf{X} \in \{\mathbf{M}, \mathbf{Q}\}$  with a complexity of  $O(|E'|)$ . Let  $V'$  be the set of nodes incident to  $E'$ . Note that only the multi-hop neighbors of  $V'$  (denoted by  $V''$ ) will participate in the GNN feature aggregation of  $V'$ . Let  $E''$  be the set of edges induced by  $V' \cup V''$ . One can also incrementally update embeddings  $\{z_i\}$  via the random projection and GNN feedforward propagation induced by  $(V' \cup V'', E'')$ . Therefore, the overall complexity of embedding derivation is  $O(|E''|d + |V' \cup V''|d)$ , where we usually have  $|E''| < |E|$  and  $|V' \cup V''| < N$ .

For the hierarchical model selection, one can formulate the partitioning procedure as a tree, where each tree node represents a subset of nodes (i.e.,  $U \subseteq V$ ) to be partitioned; the children of this tree node represent a partition of  $U$  (e.g.,  $(U_1, U_2)$ ); each leaf corresponds to a block  $C_r$ . The tree structure allows one to update only a tree path (path from a leaf to the root) once the embedding  $z_i$  of node  $v_i$  is updated.

## IV. FAST GRAPH GENERATING FROM THE SBM

The stochastic block partition benchmark of the IEEE HPEC Graph Challenge generates test graphs using the SBM generators (i.e., ‘blockmodel’ and ‘blockmodel-degree’) implemented by graph-tool [25]. We argue that there are some limitations in the implementations of these generators.

For instance, ‘blockmodel’ ignores the degree correction terms (i.e.,  $\theta$  in Definition 5). ‘blockmodel-degree’ uses the Markov Chain Monte Carlo (MCMC) sampling algorithm to rewire the edges generated from another random model. As a result, the distribution  $\mathcal{X}$  of its output approaches the SBM distribution  $\mathcal{S}$  only after sufficient iterations of MCMC. In practice, the number of iterations is set to a constant. This makes the distribution  $\mathcal{X}$  an approximation of  $\mathcal{S}$  without a provable distance guarantee.

To address the limitations, we introduce a new fast graph generating algorithm. It generates graphs from the exact SBM distribution  $\mathcal{S}(\theta, \mathbf{c}, \mathbf{\Omega})$  and has a nearly linear time complexity of  $O((|E| + K^2) \log N)$ , where  $\{\theta, \mathbf{c}, \mathbf{\Omega}\}$  are defined in Definition 5;  $K$ ,  $N$ , and  $|E|$  are the numbers of blocks, nodes, and edges in the generated graph.

**Algorithm 2:** Fast Graph Generating from the SBM

```

1 procedure RANGESUM( $r, s, x, y$ )
2   return Sum of  $\{\lambda_{ij} \equiv \theta_i \theta_j \Omega_{c_i c_j}\}$  over the  $x$ -th to the
    $y$ -th edge between blocks  $C_r$  and  $C_s$ 
3 procedure SAMPLE( $N, K, c, \theta, \Omega$ )
4   Let  $C_r \leftarrow \{v_i | c_i = r\}$  be the  $r$ -th block
5   for all pairs  $(r, s)$  from  $[K] \times [K]$  do
6      $m \leftarrow 1$  // Counter of edge
7     while  $m \leq |C_r| |C_s|$  do
8       Sample  $p \sim U[0, 1]$ 
9       Let  $t$  be the smallest integer in  $[m, |C_r| |C_s|]$  s.t.
        $p > \exp(-\text{RANGESUM}(r, s, m, t))$ 
10      if  $t$  does not exist then
11        break
12      Let  $(v_i, v_j)$  be the  $t$ -th edge from  $C_r$  to  $C_s$ 
13      Generate  $\mathbf{A}_{ij} \sim \text{Pois}(\lambda_{ij})$  s.t.  $\mathbf{A}_{ij} \neq 0$ 
14       $m \leftarrow t + 1$ 

```

TABLE I: Details of the Generated Benchmark Datasets

$N$	$ E $	$K$	deg	density	Layer Configuration
1E3	1.69E4~1.74E4	11	9~112	3E-2	[256,128,64]
5E3	9.51E4~9.69E4	19	5~164	8E-3	[1024,512,256,128]
1E4	1.95E5~1.96E5	25	6~180	4E-3	[4096,2048,1024,512,256]
5E4	9.97E5~1.01E6	44	5~205	8E-4	[4096,2048,1024,512,256]
1E5	2.01E6~2.02E6	56	4~209	4E-4	[4096,2048,1024,512,256]
2E5	4.04E6~4.06E6	71	5~230	2E-4	[4096,2048,1024,512,256]

Algorithm 2 summarizes the graph generating procedure, where we enumerate all pairs of blocks  $(C_r, C_s)$  ( $r \leq s$ ) and sample edges between these two blocks. The algorithm repeatedly finds the next edge with nonzero weight (Lines 9 and 10) and determines its weight from a Poisson distribution (Line 13). To implement Line 9, we perform a binary search on  $t$ . To compute  $\text{RANGESUM}(r, s, x, y)$ , we arrange all the possible edges between  $(C_r, C_s)$  in a virtual  $|C_r| \times |C_s|$  table. Then we can see that a precomputed 2D prefix sum gives  $\text{RANGESUM}(r, s, x, y)$  in  $O(1)$  time. Note that for undirected graphs, we should ignore the lower diagonal entries in  $\mathbf{A}$ .

## V. EXPERIMENTS

## A. Experiment Setups

1) **Datasets:** We followed the stochastic block partitioning benchmark of the IEEE HPEC Graph Challenge to generate test datasets from the SBM. As mentioned in Section IV, we develop a new generator to overcome the limitations of the original implementations in the benchmark.

We set (i) the ratio between the numbers of within-block edges and between-block edges to 2.5 and (ii) block size heterogeneity to 3. These configurations correspond to the hardest setting of the benchmark. We generated graphs with different scales by respectively setting the number of nodes  $N$  to 1E3, 5E3, 1E4, 5E4, 1E5, and 2E5. For each setting of  $N$ , we independently generated five graphs. Statistics of the generated datasets are shown in Table I, where  $|E|$  and  $K$  are the numbers of edges and blocks; deg is the node degree.

2) **Baselines:** We compared RaftGP with five baselines, including *MC-SBM* [1], *greedy modularity maximization (GMod)* [26], *spectral modularity maximization (SMod)* [16], *Par-SBM* [27], and *BP-Mod* [28]. As stated in Definition 1,

we consider the GP task where the number of blocks  $K$  is unknown and determined by the method to be evaluated. Hence, some other baselines that need a pre-set  $K$  (e.g., *metis* [29], *GraClus* [30], and *spectral clustering* [15]) could not be included in our experiments.

Note that RaftGP has two variants with their features extracted from NCut minimization (i.e.,  $\mathbf{X} = \mathbf{M}$ ) and modularity maximization (i.e.,  $\mathbf{X} = \tilde{\mathbf{Q}}$ ). They are denoted as *RaftGP-C* and *RaftGP-M*, respectively.

3) **Evaluation Criteria:** We followed the IEEE HPEC Graph Challenge to adopt *accuracy*, *adjusted random index (ARI)*, *precision*, and *recall* as quality metrics. Given precision and recall, we also computed the corresponding *F1-score* as a comprehensive metric considering both aspects. The *runtime* (in terms of seconds) of a method to output its GP result was used as the efficiency metric. We also define that a method encounters the out-of-time (OOT) exception if it cannot obtain a GP result within 1E4 seconds.

4) **Parameter & Environment Settings:** Layer configurations of RaftGP on all the datasets are shown in Table I, where the first and last numbers denote the input feature dimensionality  $L$  and embedding dimensionality  $d$ . We used Python 3.7 and PyTorch 1.13 to implement RaftGP and adopted the official or widely-used implementations of baselines (implemented by C++ and Python). All the experiments were conducted on a server with two Intel Xeon 6130 CPUs (16 cores), one GeForce RTX3090 GPU (24GB GPU memory), 32GB main memory, and Ubuntu 18.04.5 LTS OS.

## B. Quantitative Evaluation &amp; Discussions

The average quality metrics and runtime over five generated graphs w.r.t. each setting of  $N$  are reported in Table II, where the quality metrics of RaftGP are in **bold** if they perform the best; OOT and OOM denote the out-of-time and out-of-memory exceptions. In addition to the total runtime, Table III reports the time of (i) random feature extraction, (ii) embedding derivation (i.e., GNN feedforward propagation), and (iii) model selection for RaftGP.

Note that RaftGP does not include any training procedures. Surprisingly, both variants of RaftGP can achieve the best quality metrics in most cases, significantly outperforming *MC-SBM*, *GMod*, *SMod*, and *BP-SBM*. It verifies our discussions in Section III-B that *one feedforward propagation through a random initialized GNN is a special random projection with the geometric properties of the aggregated features preserved, and thus can still derive informative embeddings*. Although *Par-SBM* has close quality and slightly faster runtime compared with RaftGP when  $N$  is small, our methods can achieve much better quality metrics in cases with larger  $N$ s. In this sense, we believe that *RaftGP can achieve a better trade-off between quality and efficiency especially when  $N$  is large*.

On all the datasets, the two variants of RaftGP have close quality metrics with best performance. It validates our motivation of *using  $\mathbf{M}$  and  $\tilde{\mathbf{Q}}$  (extracted from the NCut minimization and modularity maximization objectives) as informative statistics for the derivation of community-preserving embeddings*.

TABLE II: Quantitative Evaluation Results

$N$	Methods	Acc $\uparrow$	ARI $\uparrow$	F1 $\uparrow$ (Recall, Precision)	Time $\downarrow$ (s)	$N$	Methods	Acc $\uparrow$	ARI $\uparrow$	F1 $\uparrow$ (Recall, Precision)	Time $\downarrow$ (s)
1E3	MC-SBM	0.9764	0.9689	0.9736 (0.9486, 1.0000)	10.4589	5E4	MC-SBM	0.9769	0.9701	0.9711 (0.9439, 1.0000)	1658.27
	GMod	0.7098	0.6813	0.7309 (0.9823, 0.5820)	4.7274		GMod	OOM	OOM	OOM	>10,000
	SMod	0.7098	0.5879	0.6531 (0.8675, 0.5237)	42.4847		SMod	OOM	OOM	OOM	OOM
	Par-SBM	0.9984	0.9976	0.9978 (0.9959, 0.9998)	0.4022		Par-SBM	0.9748	0.9747	0.9758 (0.9998, 0.9529)	14.2388
	BP-SBM	0.6904	0.6924	0.7508 (0.9999, 0.6010)	261.7399		BP-SBM	OOM	OOM	OOM	>10,000
	<b>RaftGP-C</b>	<b>0.9996</b>	<b>0.9994</b>	<b>0.9994</b> (0.9997, 0.9992)	1.6172		<b>RaftGP-C</b>	<b>0.9998</b>	<b>0.9998</b>	<b>0.9998</b> (1.0000, 0.9996)	71.1237
<b>RaftGP-M</b>	<b>0.9996</b>	<b>0.9994</b>	<b>0.9994</b> (0.9997, 0.9992)	1.5634	<b>RaftGP-M</b>	<b>0.9998</b>	<b>0.9998</b>	<b>0.9998</b> (1.0000, 0.9996)	69.4147		
5E3	MC-SBM	0.9841	0.9858	0.9868 (0.9739, 1.0000)	83.0489	1E5	MC-SBM	OOM	OOM	OOM	>10,000
	GMod	0.6172	0.5930	0.6289 (0.9627, 0.4670)	120.4837		GMod	OOM	OOM	OOM	>10,000
	SMod	0.3639	0.1939	0.2848 (0.7149, 0.1778)	1234.89		SMod	OOM	OOM	OOM	OOM
	Par-SBM	0.9916	0.9894	0.9904 (0.9984, 0.9825)	1.1748		Par-SBM	0.9061	0.9096	0.9130 (0.9998, 0.8401)	38.9002
	BP-SBM	0.4746	0.4755	0.5282 (0.9999, 0.3589)	3994.94		BP-SBM	OOM	OOM	OOM	>10,000
	<b>RaftGP-C</b>	<b>0.9998</b>	<b>0.9998</b>	<b>0.9997</b> (0.9999, 0.9996)	5.9843		<b>RaftGP-C</b>	<b>1.0000</b>	<b>1.0000</b>	<b>1.0000</b> (1.0000, 1.0000)	150.6443
<b>RaftGP-M</b>	<b>0.9998</b>	<b>0.9998</b>	<b>0.9997</b> (0.9999, 0.9996)	6.1395	<b>RaftGP-M</b>	<b>1.0000</b>	<b>1.0000</b>	<b>1.0000</b> (1.0000, 1.0000)	154.8510		
1E4	MC-SBM	0.9608	0.9495	0.9539 (0.9119, 0.9999)	313.1485	2E5	MC-SBM	OOM	OOM	OOM	>10,000
	GMod	OOM	OOM	OOM	>10,000		GMod	OOM	OOM	OOM	>10,000
	SMod	0.4693	0.3066	0.3721 (0.7592, 0.2464)	6609.61		SMod	OOM	OOM	OOM	OOM
	Par-SBM	0.9908	0.9959	0.9962 (0.9992, 0.9933)	4.4148		Par-SBM	0.9341	0.9445	0.9461 (0.9999, 0.8978)	109.4829
	BP-SBM	OOM	OOM	OOM	>10,000		BP-SBM	OOM	OOM	OOM	>10,000
	<b>RaftGP-C</b>	<b>0.9981</b>	<b>0.9974</b>	<b>0.9975</b> (0.9999, 0.9952)	15.6803		<b>RaftGP-C</b>	<b>1.0000</b>	<b>1.0000</b>	<b>1.0000</b> (1.0000, 1.0000)	377.0465
<b>RaftGP-M</b>	<b>0.9981</b>	0.9966	0.9968 (0.9999, 0.9938)	15.6424	<b>RaftGP-M</b>	<b>1.0000</b>	<b>1.0000</b>	<b>1.0000</b> (1.0000, 1.0000)	367.5187		

TABLE III: Detailed Runtime of RaftGP in terms of Seconds

$N$	Methods	Total (s)	Feat	Emb	Model
1E3	<b>RaftGP-C</b>	1.6172	0.1172	0.2454	1.2547
	<b>RaftGP-M</b>	1.5634	0.1136	0.2320	1.2177
5E3	<b>RaftGP-C</b>	5.9843	0.5228	0.2525	5.2089
	<b>RaftGP-M</b>	6.1395	0.4374	0.2537	5.4484
1E4	<b>RaftGP-C</b>	15.6803	2.7233	0.3040	12.6531
	<b>RaftGP-M</b>	15.6424	2.4969	0.2788	12.8667
5E4	<b>RaftGP-C</b>	71.1237	11.2939	0.6683	59.1615
	<b>RaftGP-M</b>	69.4147	10.4094	0.4556	58.5497
1E5	<b>RaftGP-C</b>	150.6443	23.1460	0.7190	126.7793
	<b>RaftGP-M</b>	154.8510	24.0318	0.7432	130.0760
2E5	<b>RaftGP-C</b>	377.0465	62.1050	1.3236	313.6178
	<b>RaftGP-M</b>	367.5187	57.6826	1.5124	308.3236

TABLE IV: Ablation Study of RaftGP

$N$	Methods	Acc $\uparrow$	ARI $\uparrow$	F1 $\uparrow$ (Recall, Precision)
1E3	<b>RaftGP-C</b>	<b>0.9996</b>	<b>0.9994</b>	<b>0.9994</b> (0.9997, 0.9992)
	(i) w/o Feat	0.9996	0.9994	0.9994 (0.9997, 0.9992)
	(ii)w/o Emb	0.9806	0.9903	0.9916 (0.9997, 0.9836)
	<b>RaftGP-M</b>	<b>0.9996</b>	<b>0.9994</b>	<b>0.9994</b> (0.9997, 0.9992)
	(i) w/o Feat	0.9996	0.9994	0.9994 (0.9997, 0.9992)
	(ii)w/o Emb	0.9000	0.8936	0.9190 (0.9998, 0.8502)
5E3	<b>RaftGP-C</b>	<b>0.9998</b>	<b>0.9998</b>	<b>0.9997</b> (0.9999, 0.9996)
	(i) w/o Feat	0.9998	0.9998	0.9997 (0.9999, 0.9996)
	(ii)w/o Emb	0.9846	0.9784	0.9798 (0.9999, 0.9605)
	<b>RaftGP-M</b>	<b>0.9998</b>	<b>0.9998</b>	<b>0.9997</b> (0.9999, 0.9996)
	(i) w/o Feat	0.9998	0.9998	0.9997 (0.9999, 0.9996)
	(ii)w/o Emb	0.1257	0.0067	0.1266 (0.9996, 0.0676)
1E4	<b>RaftGP-C</b>	<b>0.9981</b>	<b>0.9974</b>	<b>0.9975</b> (0.9999, 0.9952)
	(i) w/o Feat	0.9981	0.9967	0.9968 (0.9999, 0.9938)
	(ii)w/o Emb	0.8452	0.7023	0.7534 (0.9998, 0.6045)
	<b>RaftGP-M</b>	<b>0.9981</b>	<b>0.9966</b>	<b>0.9968</b> (0.9999, 0.9938)
	(i) w/o Feat	0.9981	0.9966	0.9968 (0.9999, 0.9938)
	(ii)w/o Emb	0.2725	0.0914	0.1966 (1.0000, 0.1090)
5E4	<b>RaftGP-C</b>	<b>0.9998</b>	<b>0.9998</b>	<b>0.9998</b> (1.0000, 0.9996)
	(i) w/o Feat	OOM	OOM	OOM
	(ii)w/o Emb	0.0537	0.0000	0.0543 (1.0000, 0.0279)
	<b>RaftGP-M</b>	<b>0.9998</b>	<b>0.9998</b>	<b>0.9998</b> (1.0000, 0.9996)
	(i) w/o Feat	OOM	OOM	OOM
	(ii)w/o Emb	0.0829	0.0261	0.0791 (0.9998, 0.0412)

As in Table III, model selection is the major bottleneck for both variants of RaftGP, which accounts for more than 80% of the total runtime. It implies that *the random feature extraction and GNN-based embedding derivation are efficient designs for RaftGP with runtime better than Par-SBM.*

### C. Ablation Study

We also validated the effects of (i) random feature extraction and (ii) embedding derivation (with GNN) of RaftGP by removing the corresponding components. In case (i),  $\mathbf{X} \in \{\mathbf{M}, \tilde{\mathbf{Q}}\}$  was directly used as the input of GNN (without random projection). In case (ii), we directly treated the features  $\mathbf{Y}$  derived via random projection as the embedding input of model selection (without using GNN). Results of our ablation study on all the datasets are shown in Table IV.

Although case (i) can achieve the quality competitive to the original RaftGP model when  $N$  is small, we encounter the OOM exception when  $N \geq 5E4$ . It indicates that *the Gaussian random projection can reduce the feature dimensionality while preserving the key geometric properties of input features* as discussed in Section III-A. Moreover, the GNN-based embedding derivation is essential to ensure the high quality of RaftGP because there are significant quality declines in case (ii) with the increase of  $N$ . It verifies our discussions in Section III-B that *the feature aggregation of GNN can enhance the ability of RaftGP to preserve community structures.*

## VI. CONCLUSION

In this paper, we focused on the GP problem and proposed a RaftGP method based on an efficient community-preserving embedding scheme, including the random feature extraction, random embedding derivation, and hierarchical model selection. We found that a randomly initialized GNN, even without training, is enough to derive informative community-preserving embeddings and support high-quality GP. We evaluated our method on the stochastic block partitioning benchmark of the IEEE HPEC Graph Challenge and compared the performance with five baselines, where a new graph generating algorithm was developed to address some limitations of the original data generators in the benchmark. Compared to the baseline provided by the IEEE HPEC Graph Challenge, our method achieves better accuracy on all test cases and takes 64.5x shorter time on large graphs. In our future work, we plan to extend RaftGP to dynamic graphs [31]–[34].

## REFERENCES

- [1] T. Peixoto, "Efficient monte carlo and greedy heuristic for the inference of stochastic block models," *Physical review E, Statistical, nonlinear, and soft matter physics*, vol. 89, p. 012804, 01 2014.
- [2] M. Qin, D. Jin, K. Lei, B. Gabrys, and K. Musial-Gabrys, "Adaptive community detection incorporating topology and content in social networks," *Knowledge-Based Systems*, vol. 161, pp. 342–356, 2018.
- [3] W. Li, M. Qin, and K. Lei, "Identifying interpretable link communities with user interactions and messages in social networks," in *Proceedings of the 2019 IEEE International Conference on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom)*. IEEE, 2019, pp. 271–278.
- [4] M. Qin and K. Lei, "Dual-channel hybrid community detection in attributed networks," *Information Sciences*, vol. 551, pp. 146–167, 2021.
- [5] B. Hendrickson and T. G. Kolda, "Graph partitioning models for parallel computing," *Parallel computing*, vol. 26, no. 12, pp. 1519–1534, 2000.
- [6] L. Dai and B. Bai, "Optimal decomposition for large-scale infrastructure-based wireless networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 8, pp. 4956–4969, 2017.
- [7] M. Qin, K. Lei, B. Bai, and G. Zhang, "Towards a profiling view for unsupervised traffic classification by exploring the statistic features and link patterns," in *Proceedings of the 2019 ACM SIGCOMM Workshop on Network Meets AI & ML*, 2019, pp. 50–56.
- [8] E. K. Kao, V. Gadelally, M. B. Hurley, M. Jones, J. Kepner, S. Mohindra, P. Monticciolo, A. Reuther, S. Samsi, W. Song, D. Staheli, and S. T. Smith, "Streaming graph challenge: Stochastic block partition," in *Proceedings of 2017 IEEE High Performance Extreme Computing Conference (HPEC)*, 2017, pp. 1–12.
- [9] D. Zhuzhunashvili and A. Knyazev, "Preconditioned spectral clustering for stochastic block partition streaming graph challenge (preliminary version at arxiv)," in *Proceedings of the 2017 IEEE High Performance Extreme Computing Conference (HPEC)*. IEEE, 2017, pp. 1–6.
- [10] L. Durbeck and P. Athanas, "Dpqs graph summarization preserves community structure," in *Proceedings of the 2021 IEEE High Performance Extreme Computing Conference (HPEC)*. IEEE, 2021, pp. 1–9.
- [11] A. J. Uppal, J. Choi, T. B. Rolinger, and H. H. Huang, "Faster stochastic block partition using aggressive initial merging, compressed representation, and parallelism control," in *Proceedings of the 2021 IEEE High Performance Extreme Computing Conference (HPEC)*. IEEE, 2021, pp. 1–7.
- [12] F. Wanye, V. Gleyzer, and W.-c. Feng, "Fast stochastic block partitioning via sampling," in *Proceedings of the 2019 IEEE High Performance Extreme Computing Conference (HPEC)*. IEEE, 2019, pp. 1–7.
- [13] Z. Zhang, P. Cui, and W. Zhu, "Deep learning on graphs: A survey," *IEEE Transactions on Knowledge & Data Engineering (TKDE)*, vol. 34, no. 1, pp. 249–270, 2020.
- [14] R. I. Arriaga and S. Vempala, "An algorithmic theory of learning: Robust concepts and random projection," *Machine Learning*, vol. 63, pp. 161–182, 2006.
- [15] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics & Computing*, vol. 17, no. 4, pp. 395–416, 2007.
- [16] M. E. Newman, "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences (PNAS)*, vol. 103, no. 23, pp. 8577–8582, 2006.
- [17] B. Wilder, E. Ewing, B. Dilkina, and M. Tambe, "End to end learning and optimization on graphs," *Proceedings of the 2019 Advances in Neural Information Processing Systems (NIPS)*, vol. 32, 2019.
- [18] Z. Chen, J. Bruna, and L. Li, "Supervised community detection with line graph neural networks," in *Proceedings of the 7th International Conference on Learning Representations (ICLR)*, 2019.
- [19] M. Qin, C. Zhang, B. Bai, G. Zhang, and D.-Y. Yeung, "Towards a better trade-off between quality and efficiency of community detection: An inductive embedding method across graphs," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 2023.
- [20] F. Tian, B. Gao, Q. Cui, E. Chen, and T.-Y. Liu, "Learning deep representations for graph clustering," in *Proceedings of the 28th AAAI Conference on Artificial Intelligence*, vol. 28, no. 1, 2014.
- [21] L. Yang, X. Cao, D. He, C. Wang, X. Wang, and W. Zhang, "Modularity based community detection with deep learning," in *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*, vol. 16, 2016, pp. 2252–2258.
- [22] J. Bourgain, "On lipschitz embedding of finite metric spaces in hilbert space," *Israel Journal of Mathematics*, vol. 52, no. 1-2, pp. 46–52, 1985.
- [23] W. J. J. Lindenstrauss, "Extensions of lipschitz maps into a hilbert space," *Contemporary Mathematics*, vol. 26, no. 189, pp. 189–206, 1984.
- [24] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proceedings of the 5th International Conference on Learning Representations (ICLR)*, 2017, p. 1609.02907.
- [25] T. P. Peixoto, "The graph-tool python library," *figshare*, 2014. [Online]. Available: [http://figshare.com/articles/graph\\_tool/1164194](http://figshare.com/articles/graph_tool/1164194)
- [26] A. Clauset, M. E. Newman, and C. Moore, "Finding community structure in very large networks," *Physical review E*, vol. 70, no. 6, p. 066111, 2004.
- [27] C. Peng, Z. Zhang, K.-C. Wong, X. Zhang, and D. E. Keyes, "A scalable community detection algorithm for large graphs using stochastic block models," *Intelligent Data Analysis*, vol. 21, no. 6, pp. 1463–1485, 2017.
- [28] P. Zhang and C. Moore, "Scalable detection of statistically significant communities and hierarchies, using message passing for modularity," *Proceedings of the National Academy of Sciences (PNAS)*, vol. 111, no. 51, pp. 18 144–18 149, 2014.
- [29] B. Hendrickson, R. W. Leland *et al.*, "A multi-level algorithm for partitioning graphs," *Proceedings Supercomputing*, vol. 95, no. 28, pp. 1–14, 1995.
- [30] I. S. Dhillon, Y. Guan, and B. Kulis, "Weighted graph cuts without eigenvectors a multilevel approach," *IEEE Transactions on Pattern Analysis & Machine Intelligence (TPAMI)*, vol. 29, no. 11, pp. 1944–1957, 2007.
- [31] K. Lei, M. Qin, B. Bai, and G. Zhang, "Adaptive multiple non-negative matrix factorization for temporal link prediction in dynamic networks," in *Proceedings of the 2018 SIGCOMM NetAI Workshop*, 2018, pp. 28–34.
- [32] K. Lei, M. Qin, B. Bai, G. Zhang, and M. Yang, "Gcn-gan: A non-linear temporal link prediction model for weighted dynamic networks," in *Proceedings of the 2019 IEEE Conference on Computer Communications (INFOCOM)*. IEEE, 2019, pp. 388–396.
- [33] M. Qin and D.-Y. Yeung, "Temporal link prediction: A unified framework, taxonomy, and review," *arXiv preprint arXiv:2210.08765*, 2022.
- [34] M. Qin, C. Zhang, B. Bai, G. Zhang, and D.-Y. Yeung, "High-quality temporal link prediction for weighted dynamic graphs via inductive embedding aggregation," *IEEE Transactions on Knowledge & Data Engineering (TKDE)*, vol. 35, no. 9, pp. 9378–9393, 2023.