Finding Dense Subgraphs via Low-Rank Bilinear Optimization

Dimitris S. Papailiopoulos DIMITRIS@UTEXAS.EDU
Ioannis Mitliagkas IOANNIS@UTEXAS.EDU
Alexandros G. Dimakis DIMAKIS@AUSTIN.UTEXAS.EDU
Constantine Caramanis CONSTANTINE@UTEXAS.EDU
The University of Texas at Austin

Abstract

Given a graph, the Densest k-Subgraph (DkS) problem asks for the subgraph on k vertices that contains the largest number of edges. In this work, we develop a new algorithm for DkS that searches a low-dimensional space for provably good solutions. Our algorithm comes with novel performance bounds that depend on the graph spectrum. Our graph-dependent bounds are in practice significantly tighter than worst case *a priori* bounds: for most tested real-world graphs we find subgraphs with density provably within 70% of the optimum.

Our algorithm runs in nearly linear time, under spectral assumptions satisfied by most graphs found in applications. Moreover, it is highly scalable and parallelizable. We demonstrate this by implementing it in MapReduce and executing numerous experiments on massive real-world graphs that have up to billions of edges. We empirically show that our algorithm can find subgraphs of significantly higher density compared to the previous state of the art.

1. Introduction

Given a graph $\mathcal{G}$ on n vertices with m edges and a parameter k, we are interested in finding an induced subgraph on k vertices with the largest average degree, also known as the *maximum density*. This is the *Densest k-Subgraph* (DkS) – a fundamental problem in combinatorial optimization with applications in numerous fields including social sciences, communication networks, and biology (see *e.g.* (Hu et al., 2005; Gibson et al., 2005; Dourisboure et al., 2007; Saha et al., 2010; Miller et al., 2010; Bahmani et al., 2012)).

Proceedings of the 31st International Conference on Machine Learning, Beijing, China, 2014. JMLR: W&CP volume 32.

DkS is a notoriously hard problem. It is NP-hard by reduction to MaxClique. Moreover, Khot showed in (Khot, 2004) that, under widely believed complexity theoretic assumptions, DkS cannot be approximated within an arbitrary constant factor.[1] The best known approximation ratio was $n^{1/3+\epsilon}$ (for some small ϵ) due to (Feige et al., 2001). Recently, (Bhaskara et al., 2010) introduced an algorithm with approximation ratio $n^{1/4+\epsilon}$, that runs in time $n^{O(1/\epsilon)}$. Such results, where the approximation factor scales as a polynomial in the number of vertices, are too pessimistic for real-world applications. The resistance to better approximation despite the long history of the problem suggests that DkS is probably very hard in the worst case.

Our Contributions. In this work we move beyond the worst case framework. We present a novel DkS algorithm that has two key features: *i)* it comes with approximation guarantees that are surprisingly tight on real-world graphs and *ii*) it is fully parallelizable and can scale up to graphs with billions of edges.

Our algorithm combines spectral and combinatorial techniques; it relies on examining candidate subgraphs obtained from vectors lying in a low-dimensional subspace of the adjacency matrix of the graph. This is accomplished through a framework called the *Spannogram*, which we define below. At its core our algorithm has an exact solver for constant rank *bilinear* programs with combinatorial constraints.

Our approximation guarantees are *data-dependent*: they are related to the spectrum of the adjacency matrix. Let opt denote the average degree (i.e., the density) of the densest k-subgraph, where $0 \leq \mathsf{opt} \leq k-1$. Our algorithm takes as input the graph, the subgraph size k and an accuracy parameter $d \in \{1, \ldots, n\}$ and runs in time $O(kn^{d+1})$. Note that under a spectral condition satisfied by many real graphs, the computational time is reduced to nearly-linear as we subse-

[1] approximation ratio ρ means that there exists an algorithm that produces in polynomial time a number A, such that $1 \leq \frac{\mathsf{opt}}{\mathsf{A}} \leq \rho$, where opt is the optimal density.

quently discuss. The output is a subgraph on k vertices with density opt_d, for which we obtain the following approximation result:

Theorem 1. *For any unweighted graph*

$$\mathsf{opt}_d \geq \mathsf{opt}/(2 + o_n(1)) - 2 \cdot |\lambda_{d+1}|.$$

where λ_i is the ith largest, in magnitude, eigenvalue of the adjacency matrix of the graph. Moreover, if the graph is bipartite, or if the largest d eigenvalues of the graph are positive, then $\mathsf{opt}_d \geq \mathsf{opt} - 2 \cdot |\lambda_{d+1}|$.

Our bounds come close to $2 + \epsilon$ and $1 + \epsilon$ factor approximations, when λ_{d+1} is significantly smaller than the density of the densest k-subgraph. In the following theorem, we give such an example. However, we would like to note that in the worst case our bounds might not yield something meaningful.

Theorem 2. *If the densest-k-subgraph contains a constant fraction of all the edges, and $k = \Theta(\sqrt{E})$, then we can approximate DkS within a factor of $2 + \epsilon$, in time $n^{O(1/\epsilon^2)}$. If additionally the graph is bipartite, we can approximate DkS within a factor of $1 + \epsilon$.*

The above result is similar to the $1 + \epsilon$ approximation ratio of (Arora et al., 1995) for dense graphs, where the densest-k-subgraph contains a constant fraction of the $\Omega(n^2)$ edges, where $k = \Omega(n)$. The innovation here is that our ratio also applies to *sparse graphs* with sublinear number of edges.

Computable upper bounds. In addition to these theoretical guarantees, our analysis allows us to obtain a graph-dependent upper bound for the optimal subgraph density. This is shown in Fig. 3 in our experimental section, where for many graphs our algorithm is provably within 70% from the upper bound of opt. These are far stronger guarantees than the best available *a priori* bounds. This illustrates the potential power of graph-dependent guarantees that, however, require the execution of an algorithm.

Nearly-linear time approximation. Our algorithm has a worst-case running time of $O(kn^{d+1})$, where d controls the quality of the approximation. However, under some mild spectral assumptions, a randomized version of our algorithm runs in nearly-linear time.

Theorem 3. *Let the d largest eigenvalues of the graph be positive and let the d and $d+1$ largest have constant ratio: $\lambda_d/\lambda_{d+1} = \Theta(1)$. Then, we can modify our algorithm to output, with high probability, a subgraph with density $(1-\frac{1}{\log n})\cdot\mathsf{opt}_d$ in nearly-linear time $O(m \cdot \log n + n \cdot \log^{d+1} n)$, where m is the number of edges.*

We found that the above spectral condition holds for all $d \leq 5$, in many real-world graphs that we tested.

Scalability. We develop two key scalability features that allow us to scale up efficiently on massive graphs.

Vertex sparsification: We introduce a pre-processing step that eliminates vertices that are unlikely to be part of the densest k-subgraph. The elimination is based on the vertices' *weighted leverage scores* (Mahoney & Drineas, 2009; Boutsidis et al., 2009) and admits a provable bound on the introduced error. We empirically found that even with a negligible additional error, the elimination dramatically reduced problem sizes in all tested datasets.

MapReduce implementation: We show that our algorithm is *fully-parallelizable* and tailor it for the MapReduce framework. We use our MapReduce implementation to run experiments on Elastic MapReduce (EMR) on Amazon. In our large-scale experiments, we were able to scale out to thousands of mappers and reducers in parallel over 800 cores, and find large dense subgraphs in graphs with billions of edges.

1.1. Related work

DkS algorithms: One of the few positive results for DkS is a $1 + \epsilon$ approximation for dense graphs where $m = \Omega(n^2)$, and in the linear subgraph setting $k = \Omega(n)$ (Arora et al., 1995). For some values of $m = o(n^2)$ a $2 + \epsilon$ approximation was established by (Suzuki & Tokuyama, 2005). Moreover, for any $k = \Omega(n)$ a constant factor approximation is possible via a greedy approach by (Asahiro et al., 2000), or via semidefinite relaxations by (Srivastav & Wolf, 1998) and (Feige & Langberg, 2001). Recently, (Alon et al., 2013) established new approximation results for graphs with small "ϵ-rank," using an approximate DkS solver for low-rank perturbed versions of the adjacency matrix.

There is a vast literature on algorithms for detecting communities and well-connected subgraphs: greedy schemes (Ravi et al., 1994), optimization approaches (Jethava et al., 2012; d'Aspremont et al., 2010; Ames, 2011), and the truncated power method (Yuan & Zhang, 2011). We compare with various of these algorithms in our evaluation section.

The Spannogram framework: We present an exact solver for *bilinear* optimization problems on matrices of constant rank, under $\{0, 1\}$ *and* sparsity constraints on the variables. Our theory is a generalization of the Spannogram framework, originally introduced in the foundational work of (Karystinos & Liavas, 2010) and further developed in (Asteris et al., 2014; Papailiopoulos et al., 2013), that obtains exact solvers for low-rank *quadratic optimization* problems with combinatorial constraints, such as sparse PCA.

MapReduce algorithms for graphs: The design of

MapReduce algorithms for massive graphs is an active research area as Hadoop becomes one of the standards for storing large data sets. The related work by Bahmani *et al.* (Bahmani et al., 2012) designs a novel MapReduce algorithm for the *densest subgraph* problem. This densest subgraph problem requires finding a subgraph of highest *normalized density* without enforcing a specific subgraph size k. Surprisingly, without a subgraph size restriction, the densest subgraph becomes polynomially solvable and therefore fundamentally different from what we consider in this paper.

2. Proposed Algorithm

The *density* of a subgraph indexed by a vertex set $\mathcal{S} \subseteq \{1, \ldots, n\}$ is equal to the average degree of the vertices within $\mathcal{S}$:

$$\mathsf{den}(\mathcal{S}) = \frac{\mathbf{1}_{\mathcal{S}}^T \mathbf{A} \mathbf{1}_{\mathcal{S}}}{|\mathcal{S}|}$$

where $\mathbf{A}$ is the adjacency matrix ($A_{i,j} = 1$ if (i,j) is an edge, else $A_{i,j} = 0$) and the indicator vector $\mathbf{1}_{\mathcal{S}}$ has 1s in the entries indexed by $\mathcal{S}$ and 0 otherwise. Observe that $\mathbf{1}_{\mathcal{S}}^T \mathbf{A} \mathbf{1}_{\mathcal{S}} = \sum_{i,j \in \mathcal{S}} A_{i,j}$ is twice the number of edges in the subgraph with vertices in $\mathcal{S}$.

For a fixed subgraph size $|\mathcal{S}| = k$, we can express DkS as a quadratic optimization:

$$\mathsf{DkS}: \quad \mathsf{opt} = (1/k) \cdot \max_{|\mathcal{S}|=k} \mathbf{1}_{\mathcal{S}}^T \mathbf{A} \mathbf{1}_{\mathcal{S}}$$

where $|\mathcal{S}| = k$ denotes that the optimization variable is a k-vertex subset of $\{1, \ldots, n\}$.

The bipartite version of DkS. We approximate DkS via approximating its bipartite version. This problem can be expressed as a bilinear maximization:

$$\mathsf{DBkS}: \quad \mathsf{opt}^{\mathsf{B}} = (1/k) \cdot \max_{|\mathcal{X}|=k} \max_{|\mathcal{Y}|=k} \mathbf{1}_{\mathcal{X}}^T \mathbf{A} \mathbf{1}_{\mathcal{Y}}.$$

As we see in the following lemma, the two problems are fundamentally related: a good solution for the bipartite version of the problem maps to a "half as good" solution for DkS. The proof can be found in the supplemental material, where we describe how to convert an algorithm for DBkS to one for DkS.

Lemma 1. *A ρ-approximation algorithm for DBkS implies a 2ρ-approximation algorithm for DkS.*

2.1. DBkS through low rank approximations.

At the core of our approximation lies a constant rank solver: we show that DBkS can be solved in polynomial time on constant rank matrices. We solve constant rank instances of DBkS instead of DkS due to an important implication: DkS is NP-hard even for rank-1 matrices, as we show in the supplemental material.

The exact steps of our low-rank DBkS algorithm are given in the pseudo-code tables referred to as Algorithms 1 and 2.[2] We continue with a high-level description.

Step 1: *Obtain $\mathbf{A}_d = \sum_{i=1}^{d} \lambda_i \mathbf{v}_i \mathbf{v}_i^T$, a rank-$d$ approximation of $\mathbf{A}$.* Here, λ_i is the i-th largest in magnitude eigenvalue and $\mathbf{v}_i$ the corresponding eigenvector.

Step 2: *Use $\mathbf{A}_d$ to obtain $O(n^d)$ candidate subgraphs.* For any matrix $\mathbf{A}$ we can solve DBkS by exhaustively checking all $\binom{n}{k} \cdot \binom{n}{k}$ pairs $(\mathcal{X}, \mathcal{Y})$ of k-subsets of vertices. Surprisingly, if we want to find the $\mathcal{X}, \mathcal{Y}$ pairs that maximize $\mathbf{1}_{\mathcal{X}}^T \mathbf{A}_d \mathbf{1}_{\mathcal{Y}}$, i.e., the bilinear problem on the rank-d matrix $\mathbf{A}_d$, then we show that only $O(n^d)$ candidate pairs need to be examined.

In the next section, we derive the constant rank-solver using two key facts. First, for each fixed vertex set $\mathcal{Y}$, we show that it is easy to find the optimal set $\mathcal{X}$ that maximizes $\mathbf{1}_{\mathcal{X}}^T \mathbf{A}_d \mathbf{1}_{\mathcal{Y}}$. Since this turns out to be easy, then the challenge is to find the number of different vertex sets $\mathcal{Y}$ that we need to check. Do we need to exhaustively check all $\binom{n}{k}$ supports $\mathcal{Y}$? We show that this question is equivalent to searching the span of the first d eigenvectors of $\mathbf{A}$, and collecting in a set $\mathcal{S}_d$ the top-k coordinates of all vectors in that d-dimensional space. By modifying the Spannogram theory of (Karystinos & Liavas, 2010; Asteris et al., 2014; Papailiopoulos et al., 2013), we show how this set has size $O(n^d)$ and can be constructed in time $O(n^{d+1})$. This will imply that DBkS can be solved in time $O(n^{d+1})$ on $\mathbf{A}_d$.

Computational Complexity. The worst-case time complexity of our algorithm is $O(\mathsf{T}_d + n^{d+1})$, where T_d is the time to compute the first d eigenvectors of $\mathbf{A}$. Under conditions satisfied by many real world graphs, we show that the complexity reduces to *nearly linear* in the size of $\mathcal{G}$: $O(m \cdot \log n + n \cdot \log^{d+1} n)$.

Algorithm 1 `lowrankDBkS(`k`, `d`, `$\mathbf{A}$`)`

1: $[\mathbf{V}_d, \mathbf{\Lambda}_d] = \texttt{EVD}(\mathbf{A}, d)$
2: $\mathcal{S}_d = \texttt{Spannogram}(k, \mathbf{V}_d)$
3: $\{\mathcal{X}_d, \mathcal{Y}_d\} = \arg\max_{|\mathcal{X}|=k} \max_{\mathcal{Y} \in \mathcal{S}_d} \mathbf{1}_{\mathcal{X}}^T \mathbf{V}_d \mathbf{\Lambda}_d \mathbf{V}_d^T \mathbf{1}_{\mathcal{Y}}$
4: **Output:** $\{\mathcal{X}_d, \mathcal{Y}_d\}$

`Spannogram(`k`, `$\mathbf{V}_d$`)`

$\mathcal{S}_d = \{\mathrm{top}_k(\mathbf{v}) : \mathbf{v} \in \mathrm{span}(\mathbf{v}_1, \ldots, \mathbf{v}_d)\}$
Output: $\mathcal{S}_d$.

Remark 1. *In the following we present an exact solver for DBkS on constant rank approximations of $\mathbf{A}$. Our general algorithm for DkS makes a number of $O(k)$ calls to the DBkS low-rank solver, on slightly different matrices that are obtained by sub-sampling the*

[2] In the pseudocode of Algorithm 1, $\mathrm{top}_k(\mathbf{v})$, denotes the indices of the k largest signed elements of $\mathbf{v}$.

entries of a low-rank approximation of $\mathbf{A}$. *The details of our general algorithm can be found in the supplemental material.*

2.2. Approximation Guarantees

We approximate DBkS by finding a solution to the constant rank problem

$$\max_{|\mathcal{X}|=k} \max_{|\mathcal{Y}|=k} \mathbf{1}_{\mathcal{X}}^T \mathbf{A}_d \mathbf{1}_{\mathcal{Y}}.$$

We output a pair of vertex sets, $\mathcal{X}_d, \mathcal{Y}_d$, which we refer to as the *rank-d optimal solution*, that has density

$$\mathsf{opt}_d^{\mathsf{B}} = (1/k) \cdot \mathbf{1}_{\mathcal{X}_d}^T \mathbf{A} \mathbf{1}_{\mathcal{Y}_d}.$$

Our approximation guarantees measure how far $\mathsf{opt}_d^{\mathsf{B}}$ is from $\mathsf{opt}^{\mathsf{B}}$, the optimal density for DBkS. Our bounds capture a simple core idea: the loss in our approximation comes due to solving the problem on $\mathbf{A}_d$ instead of solving it on the full rank matrix $\mathbf{A}$. This loss is quantified in the next lemma. The proofs of the following results are in the supplemental material.

Lemma 2. *For any matrix* $\mathbf{A}$

$$\mathsf{opt}_d^{\mathsf{B}} \geq \mathsf{opt}^{\mathsf{B}} - 2 \cdot |\lambda_{d+1}|.$$

where λ_i *is the ith largest eigenvalue of* $\mathbf{A}$.

Using an appropriate pre-processing step and then running Algorithm 1 as a subroutine on a sub-sampled and low-rank version of $\mathbf{A}$, we output a k-subgraph $\mathcal{Z}_d$ that has density opt_d. By essentially combining Lemmata 1 and 2 we obtain the following bounds.

Theorem 1. *The* k*-subgraph indexed by* $\mathcal{Z}_d$ *has density*

$$\mathsf{opt}_d \geq \mathsf{opt}/(2 + o_n(1)) - 2 \cdot |\lambda_{d+1}|.$$

where λ_i *is the ith largest, in magnitude, eigenvalue of the adjacency matrix of the graph. Moreover, if the graph is bipartite, or if the largest* d *eigenvalues of the graph are positive, then* $\mathsf{opt}_d \geq \mathsf{opt} - 2 \cdot |\lambda_{d+1}|$.

Using bounds on eigenvalues of graphs, Theorem 1 translates to the following approximation guarantees.

Theorem 2. *If the densest-*k*-subgraph contains a constant fraction of all the edges, and* $k = \Theta(\sqrt{E})$, *then we can approximate DkS within a factor of* $2 + \epsilon$, *in time* $n^{O(1/\epsilon^2)}$. *If additionally the graph is bipartite, then we can approximate DkS within a factor of* $1 + \epsilon$.

Remark 2. *The above results are similar to the* $1 + \epsilon$ *ratio of (Arora et al., 1995), which holds for graphs where the densest-*k*-subgraph contains* $\Omega(n^2)$ *edges.*

Graph dependent bounds. For any given graph, after running our algorithm, we can compute an upper bound to the optimal density opt via bounds on $\mathsf{opt}^{\mathsf{B}}$, since it is easy to see that $\mathsf{opt}^{\mathsf{B}} \geq \mathsf{opt}$. Our graph-dependent bound is the minimum of three upper bounds on the unknown optimal density:

Lemma 3. *The optimal density of DkS can be bounded as*

$$\mathsf{opt} \leq \min\left\{(1/k) \cdot \mathbf{1}_{\mathcal{X}_d}^T \mathbf{A}_d \mathbf{1}_{\mathcal{Y}_d} + |\lambda_{d+1}|,\ k-1,\ \lambda_1\right\}.$$

In our experimental section, we plot the above upper bounds, and show that for most tested graphs our algorithm performs *provably* within 70% from the upper bound on the optimal density. These are far stronger guarantees than the best available *a priori* bounds.

3. The Spannogram Framework

In this section, we describe how our constant rank solver operates by examining candidate vectors in a low-dimensional span of $\mathbf{A}$.

Here, we work on a rank-d matrix $\mathbf{A}_d = \mathbf{v}_1\mathbf{u}_1^T + \ldots + \mathbf{v}_d\mathbf{u}_d^T$ where $\mathbf{u}_i = \lambda_i \mathbf{v}_i$, and we wish to solve:

$$\max_{|\mathcal{X}|=|\mathcal{Y}|=k} \mathbf{1}_{\mathcal{X}}^T \left(\mathbf{v}_1\mathbf{u}_1^T + \ldots + \mathbf{v}_d\mathbf{u}_d^T\right) \mathbf{1}_{\mathcal{Y}}. \tag{1}$$

Observe that we can rewrite (1) in the following way

$$\max_{|\mathcal{X}|=|\mathcal{Y}|=k} \mathbf{1}_{\mathcal{X}}^T \left[\mathbf{v}_1 \cdot \underbrace{(\mathbf{u}_1^T \mathbf{1}_{\mathcal{Y}})}_{c_1} + \ldots + \mathbf{v}_d \cdot \underbrace{(\mathbf{u}_d^T \mathbf{1}_{\mathcal{Y}})}_{c_d}\right]$$
$$= \max_{|\mathcal{Y}|=k} \left(\max_{|\mathcal{X}|=k} \mathbf{1}_{\mathcal{X}}^T \mathbf{v}_{\mathcal{Y}}\right), \tag{2}$$

where $\mathbf{v}_{\mathcal{Y}} = \mathbf{v}_1 \cdot c_1 + \ldots + \mathbf{v}_d \cdot c_d$ is an n-dimensional vector generated by the d-dimensional subspace spanned by $\mathbf{v}_1, \ldots, \mathbf{v}_d$.

We will now make a key observation: for every fixed vector $\mathbf{v}_{\mathcal{Y}}$ in (2), the index set $\mathcal{X}$ that maximizes $\mathbf{1}_{\mathcal{X}}^T \mathbf{v}_{\mathcal{Y}}$ can be easily computed. It is not hard to see that for any fixed vector $\mathbf{v}_{\mathcal{Y}}$, the k-subset $\mathcal{X}$ that maximizes

$$\mathbf{1}_{\mathcal{X}}^T \mathbf{v}_{\mathcal{Y}} = \sum_{i \in \mathcal{X}} [\mathbf{v}_{\mathcal{Y}}]_i$$

corresponds the set of k largest signed coordinates of $\mathbf{v}_{\mathcal{Y}}$. That is, the locally optimal k-set is $\mathrm{top}_k(\mathbf{v}_{\mathcal{Y}})$.

We now wish to find all possible locally optimal sets $\mathcal{X}$. If we could possibly check all vectors $\mathbf{v}_{\mathcal{Y}}$, then we could find all locally optimal index sets $\mathrm{top}_k(\mathbf{v}_{\mathcal{Y}})$.

Let us denote as $\mathcal{S}_d$ the set of all k-subsets $\mathcal{X}$ that are the optimal solutions of the inner maximization of (2) for *any* vector $\mathbf{v}$ in the span of $\mathbf{v}_1, \ldots, \mathbf{v}_d$

$$\mathcal{S}_d = \{\mathrm{top}_k([\mathbf{v}_1 \cdot c_1 + \ldots + \mathbf{v}_d \cdot c_d]) : c_1, \ldots, c_d \in \mathbb{R}\}.$$

Clearly, this set contains all possible locally optimal $\mathcal{X}$ sets of the form $\text{top}_k(\mathbf{v}_\mathcal{Y})$. Therefore, we can rewrite DBkS on $\mathbf{A}_d$ as

$$\max_{|\mathcal{Y}|=k} \max_{\mathcal{X}\in\mathcal{S}_d} \mathbf{1}_\mathcal{X}^T \mathbf{A}_d \mathbf{1}_\mathcal{Y}. \tag{3}$$

The above problem can now be solved in the following way: for every set $\mathcal{X} \in \mathcal{S}_d$ find the locally optimal set $\mathcal{Y}$ that maximizes $\mathbf{1}_\mathcal{X}^T \mathbf{A}_d \mathbf{1}_\mathcal{Y}$, that is, this will be $\text{top}_k(\mathbf{A}_d \mathbf{1}_\mathcal{X})$. Then, we simply need to test all such $\mathcal{X}, \mathcal{Y}$ pairs on $\mathbf{A}_d$ and keep the optimizer.

Due to the above, the problem of solving DBkS on $\mathbf{A}_d$ is equivalent to constructing the set of k-supports $\mathcal{S}_d$, and then finding the optimal solution in that set. How large can $\mathcal{S}_d$ be and can we construct it in polynomial time? Initially one could expect that the set $\mathcal{S}_d$ could have size as big as $\binom{n}{k}$. Instead, we show that the set $\mathcal{S}_d$ will be tremendously smaller.

Lemma 4. *The set $\mathcal{S}_d$ has size at most $O(n^d)$ and can be built in time $O(n^{d+1})$ using Algorithm 2.*

3.1. Constructing the set $\mathcal{S}_d$

We build up to the general rank-d algorithm by explaining special cases that are easier to understand.

Rank-1 case. We start with the $d = 1$ case, where we have $\mathcal{S}_1 = \{\text{top}_k(c_1 \cdot \mathbf{v}_1) : c_1 \in \mathbb{R}\}$. It is not hard to see that there are only two supports to include in $\mathcal{S}_1$: $\text{top}_k(\mathbf{v}_1)$ and $\text{top}_k(-\mathbf{v}_1)$. These two sets can be constructed in time in time $O(n)$, via a partial sorting and selection algorithm (Cormen et al., 2001). Hence, $\mathcal{S}_1$ has size 2 and can be constructed in time $O(n)$.

Rank-2 case. This is the first non-trivial d which exhibits the details of the Spannogram algorithm.

Let an auxiliary angle $\phi \in \Phi = [0,\ \pi)$ and let

$$\mathbf{c} = \left[\begin{smallmatrix} c_1 \\ c_2 \end{smallmatrix}\right] = \left[\begin{smallmatrix} \sin\phi \\ \cos\phi \end{smallmatrix}\right].^{3}$$

Then, we re-express $c_1 \cdot \mathbf{v}_1 + c_2 \cdot \mathbf{v}_2$ in terms of ϕ as

$$\mathbf{v}(\phi) = \sin\phi \cdot \mathbf{v}_1 + \cos\phi \cdot \mathbf{v}_2. \tag{4}$$

This means that we can rewrite the set $\mathcal{S}_2$ as:

$$\mathcal{S}_2 = \{\text{top}_k(\pm(\mathbf{v}(\phi)), \phi \in [0, \pi)\}.$$

Observe that each element of $\mathbf{v}(\phi)$ is a continuous *spectral curve* in ϕ: $[\mathbf{v}(\phi)]_i = [\mathbf{v}_1]_i \sin(\phi) + [\mathbf{v}_2]_i \cos(\phi)$. Consequently, the top/bottom-k supports of $\mathbf{v}(\phi)$ (*i.e.*, $\text{top}_k(\pm\mathbf{v}(\phi))$) are themselves a function of ϕ. How can we find all possible supports?

[3] Observe that when we scan ϕ, the vectors $\mathbf{c}, -\mathbf{c}$ express all possible unit norm vectors on the circle.

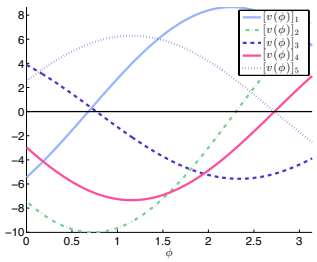


Figure 1. A rank $d = 2$ spannogram for $n = 5$ and two random vectors $\mathbf{v}_1, \mathbf{v}_2$. Observe that every two curves intersect in exactly one point. These intersection points define intervals in which a top-k set is invariant.

The Spannogram. In Fig. 1, we draw an example plot of five curves $[\mathbf{v}(\phi)]_i$, $i = 1, \ldots, 5$, which we call a spannogram. From the spannogram in Fig. 1, we can see that the continuity of these sinusoidal curves implies a "local invariance" property of the top/bottom k supports $\text{top}_k(\pm\mathbf{v}(\phi))$, in a small neighborhood around a fixed ϕ. So, when does a top/bottom-k support change? The index sets $\text{top}_k(\pm\mathbf{v}(\phi))$ change *if and only if* two curves cross, *i.e.*, when the ordering of two elements $[\mathbf{v}(\phi)]_i$,$[\mathbf{v}(\phi)]_j$ changes.

Finding all supports: There are n curves and each pair intersects at exactly one point in the Φ domain[4]. Therefore, there are exactly $\binom{n}{2}$ intersection points. These $\binom{n}{2}$ intersection points define $\binom{n}{2} + 1$ intervals. Within an interval the top/bottom k supports $\text{top}_k(\pm\mathbf{v}(\phi))$ *remain the same*. Hence, it is now clear that $|\mathcal{S}_2| \leq 2\binom{n}{2} = O(n^2)$.

A way to find all supports in $\mathcal{S}_2$ is to compute the $\mathbf{v}(\phi_{i,j})$ vectors on the intersection points of two curves i, j, and then the supports in the two adjacent intervals of such intersection point. The $\mathbf{v}(\phi_{i,j})$ vector on an intersection point of two curves i and j can be easily computed by first solving a set of linear equations $[\mathbf{v}(\phi_{i,j})]_i = [\mathbf{v}(\phi_{i,j})]_j \Rightarrow (\mathbf{e}_i - \mathbf{e}_j)^T[\mathbf{v}_1\ \ \mathbf{v}_2]\mathbf{c}_{i,j} = \mathbf{0}_{2\times 1}$ for the unknown vector $\mathbf{c}_{i,j}$, where $\mathbf{e}_i$ is the i-th column of the $n \times n$ identity matrix, i.e.,

$$\mathbf{c}_{i,j} = \text{nullspace}((\mathbf{e}_i - \mathbf{e}_j)^T[\mathbf{v}_1\ \ \mathbf{v}_2]).$$

Then, we compute $\mathbf{v}(\phi_{i,j}) = [\mathbf{v}_1\ \mathbf{v}_2]\mathbf{c}_{i,j}$. Further details on breaking ties in $\text{top}_k(\mathbf{v}(\phi_{i,j}))$ can be found in the supplemental material.

Computational cost: We have $\binom{n}{2}$ intersection points, where we calculate the top/bottom k supports for each $\mathbf{v}(\phi_{i,j})$. The top/bottom k elements of every $\mathbf{v}(\phi_{i,j})$ can be computed in time $O(n)$ using a partial sorting

[4] Here we assume that the curves are in *general position*. This can be always accomplished by infinitesimally perturbing the curves as in (Papailiopoulos et al., 2013).

and selection algorithm (Cormen et al., 2001). Since we perform this routine a total of $O(\binom{n}{2})$ times, the total complexity of our rank-2 algorithm is $O(n^3)$.

General Rank-d case. The algorithm generalizes to arbitrary dimension d, as we show in the supplemental material; its pseudo-code is given as Algorithm 2.

Remark 3. *Observe that the computation of each loop in line 2 of Algorithm 2 can be computed in parallel. This will allow us to parallelize the Spannogram.*

Algorithm 2 `Spannogram`$(k, \mathbf{V}_d)$

1: $\mathcal{S}_d = \emptyset$
2: **for** all $(i_1, \ldots, i_d) \in \{1, \ldots, n\}^d$ and $s \in \{-1, 1\}$ **do**
3: $\quad \mathbf{c} = s \cdot \text{nullspace}\left(\begin{bmatrix} [\mathbf{V}_d]_{i_1,:} - [\mathbf{V}_d]_{i_2,:} \\ \vdots \\ [\mathbf{V}_d]_{i_1,:} - [\mathbf{V}_d]_{i_d,:} \end{bmatrix}\right)$
4: $\quad \mathbf{v} = \mathbf{V}_d^T \mathbf{c}$
5: $\quad \mathcal{S} = \text{top}_k(\mathbf{v})$
6: $\quad \mathcal{T} = \mathcal{S} - \{i_1, \ldots, i_d\}$
7: $\quad$ **for** all $\binom{d}{d-|\mathcal{T}|}$ subsets $\mathcal{J}$ of $(i_1, \ldots, i_d)$ **do**
8: $\quad\quad \mathcal{S}_d = \mathcal{S}_d \bigcup (\mathcal{T} \cup \mathcal{J})$
9: $\quad$ **end for**
10: **end for**
11: **Output:** $\mathcal{S}_d$.

3.2. An approximate $\mathcal{S}_d$ in nearly-linear time

In our exact solver, we solve DBkS on $\mathbf{A}_d$ in time $O(n^{d+1})$. Surprisingly, when $\mathbf{A}_d$ has only positive eigenvalues, then we can tightly approximate DBkS on $\mathbf{A}_d$ in nearly linear time.

Theorem 3. *Let the d largest eigenvalues of the graph be positive and let $\lambda_d/\lambda_{d+1} = \Theta(1)$. Then, using Algorithm 3 as a subroutine in Algorithm 1 yields, with high probability, a subgraph with density $(1 - \frac{1}{\log n}) \cdot \mathsf{opt}_d$ in nearly-linear time $O(m \cdot \log n + n \cdot \log^{d+1} n)$, where m is the number of edges.*

The main idea is that instead of checking *all* $O(n^d)$ possible k sets in $\mathcal{S}_d$, we can approximately solve the problem by randomly sampling $(\log^d n)$ vectors in the span of $\mathbf{v}_1, \ldots, \mathbf{v}_d$. Our proof is based on the fact that we can "approximate" the surface of the d-dimensional sphere with $M = O(\log^{d+1} n)$ randomly sampled vectors from the span of $\mathbf{v}_1, \ldots, \mathbf{v}_d$, which then allows us to identify, with high probability, near-optimal candidates in $\mathcal{S}_d$. The modified algorithm is very simple and is given below; its analysis can be found in the supplemental material.

Algorithm 3 `Spannogram_approx`$(k,\ \mathbf{V}_d, \mathbf{\Lambda}_d)$

1: **for** $i = 1 : d \cdot 2^d \cdot \log^{d+1} n$ **do**
2: $\quad \mathbf{v} = (\mathbf{\Lambda}_d^{1/2} \cdot \mathbf{V}_d)^T \cdot \texttt{randn}(d, 1)$
3: $\quad \mathcal{S}_d = \mathcal{S}_d \cup \text{top}_k(\mathbf{v}) \cup \text{top}_k(-\mathbf{v})$
4: **end for**
5: **Output:** $\mathcal{S}_d$.

4. Scaling up

In this section, we present the two key scalability features that allow us to scale up to graphs with billions of edges.

4.1. Vertex Sparsification

We introduce a very simple and efficient pre-processing step for discarding vertices that are unlikely to appear in a top k set in $\mathcal{S}_d$. This step runs after we compute $\mathbf{A}_d$ and uses the leverage score, $\ell_i = \sum_{j=1}^{d} [\mathbf{V}_d]_{i,j}^2 |\lambda_j|$, of the i-th vertex to decide whether we will discard it or not. We show in the supplemental material, that by appropriately setting a threshold, we can guarantee a provable bound on the error introduced. In our experimental results, the above elimination is able to reduce n to approximately $\hat{n} \approx 10 \cdot k$ for a provably small additive error, even for data sets where $n = 10^8$.

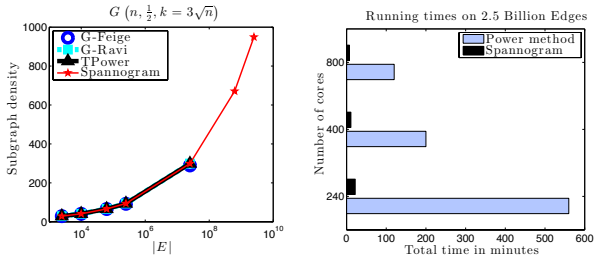
4.2. MapReduce Implementation

A MapReduce implementation allows scaling out to a large number of compute nodes that can work in parallel. The reader can refer to (Meng & Mahoney; Bahmani et al., 2012)) for a comprehensive treatment of the MapReduce paradigm. In short, the Hadoop/MapReduce infrastructure stores the input graph as a distributed file spread across multiple machines; it provides a tuple streaming abstraction, where each `map` and `reduce` function receives and emits tuples as $(key, value)$ pairs. The role of the keys is to ensure information aggregation: all the tuples with the same key are processed by the same reducer.

For the spectral decomposition step of our scheme we design a simple implementation of the power method in MapReduce. The details are beyond the scope of this work; high-performance implementations are already available in the literature, e.g. (Lin & Schatz, 2010). We instead focus on the novel implementation of the Spannogram.

Our MapReduce implementation of the rank-2 Spannogram is outlined in Algorithm 4. The Mapper is responsible for the duplication and dissemination of the eigenvectors, $\mathbf{V}_2, \mathbf{U}_2 = \mathbf{V}_2 \mathbf{\Lambda}_2$, to all reducers. Line 3 emits the j-th row of $\mathbf{V}_2$ and $\mathbf{U}_2$ once for every node i. Since i is used as the key, this ensures that every reducer receives $\mathbf{V}_2, \mathbf{U}_2$ in their entirety.

From the breakdown of the Spannogram in Section 3, it is understood that, for the rank-2 case, it suffices to solve a simple system of equations for every pair of nodes. The Reducer for node i receives the full eigenvectors $\mathbf{V}_2, \mathbf{U}_2$ and is responsible for solving the problem for every pair (i, j), where $j > i$. Then, Line 6 emits the best candidate computed at Reducer i. A



(a) Densities of the recovered subgraph v.s. the expected number of edges.

(b) Running times of the Spannogram and power iteration for two top eigenvectors.

Figure 2. Planted clique experiments for random graphs.

trivial final step, not outlined here, collects all n^2 candidate sets and keeps the best one as the final solution.

The basic outline in Algorithm 4 comes with heavy communication needs and was chosen here for ease of exposition. The more efficient version that we implement, does not replicate $\mathbf{V}_2, \mathbf{U}_2$ n times. Instead, the number of reducers – say $R = n^\alpha$ – is fine-tuned to the capabilities of the cluster. The mappers emit $\mathbf{V}_2, \mathbf{U}_2$ R times, once for every reducer. Then, reducer r is responsible for solving for node pairs (i, j), where $i \equiv r \pmod R$ and $j > i$. Depending on the performance bottleneck, different choices for α are more appropriate. We divide the construction of the $O(n^2)$ candidate sets in $\mathcal{S}_2$ to $O(n^\alpha)$ reducers and each of them computes $O(n^{2-\alpha})$ candidate subgraphs. The total communication cost for this parallelization scheme is $O(n^{1+\alpha})$: n^α reducers need to have access to the entire $\mathbf{V}_2, \mathbf{U}_2$ that has $2 \cdot 2 \cdot n$ entries. Moreover, the total computation cost for each reducer is $O(n^{3-\alpha})$.

Algorithm 4 `SpannogramMR`$(\mathbf{V}_2, \mathbf{U}_2)$

Map$(\{[\mathbf{V}_2]_{j,:}, [\mathbf{U}_2]_{j,:}, j\})$:
for $i = 1 : n$ **do**
  emit: $\langle i, \{[\mathbf{V}_2]_{j,1}, [\mathbf{V}_2]_{j,2}, [\mathbf{U}_2]_{j,1}, [\mathbf{U}_2]_{j,2}, j\}\rangle$
end for

Reduce$_i(\langle i, \{[\mathbf{V}_2]_{j,1}, [\mathbf{V}_2]_{j,2}, [\mathbf{U}_2]_{j,1}, [\mathbf{U}_2]_{j,2}, j\}\rangle, \forall j)$:
for each $j \geq i + 1$ **do**
  $\mathbf{c} = \text{nullspace}([\mathbf{V}]_{i,:} - [\mathbf{V}]_{j,:})$
  $[\mathsf{den}_j,\ \{\mathcal{X}_j, \mathcal{Y}_j\}] = \max_{|\mathcal{Y}|=k, \mathcal{X} \in \text{top}_k(\pm \mathbf{V}_2 \mathbf{c})} \mathbf{1}_\mathcal{X} \mathbf{V}_2 \mathbf{U}_2^T \mathbf{1}_\mathcal{Y}$
end for
emit: $\langle i, \{\mathcal{X}_i, \mathcal{Y}_i\} = \max_j \mathbf{1}_{\mathcal{X}_j} \mathbf{V}_2 \mathbf{U}_2^T \mathbf{1}_{\mathcal{Y}_j} \rangle$

5. Experimental Evaluation

We experimentally evaluate the performance of our algorithm and compare it to the truncated power method (TPower) of (Yuan & Zhang, 2011), a greedy algorithm by (Feige et al., 2001) (GFeige) and another greedy algorithm by (Ravi et al., 1994) (GRavi). We performed experiments on synthetic dense subgraphs and also massive real graphs from multiple sources. In all experiments we compare the density of the subgraph obtained by the Spannogram to the density of the output subgraphs given by the other algorithms.

Our experiments illustrate three key points: (1) for all tested graphs, our method outperforms – some times significantly – all other algorithms compared; (2) our method is highly scalable, allowing us to solve far larger problem instances; (3) our data-dependent upper bound in many cases provide a certificate of near-optimality, far more accurate and useful, than what *a priori* bounds are able to do.

Planted clique. We first consider the so-called (and now much studied) Planted Clique problem: we seek to find a clique of size k that has been planted in a graph where all other edges are drawn independently with probability 1/2. We scale our randomized experiments from $n = 100$ up to 10^5. In all cases we set the size of the clique to $k = 3 \cdot \sqrt{n}$ – close to what is believed to be the critical computability threshold. In all our experiments, GRavi, TPower, and the Spannogram successfully recovered the hidden clique. However, as can be seen in Fig. 2, the Spannogram algorithm is the only one able to scale up to $n = 10^5$– a massive dense graph with about 2.5 billion edges. The reason is that this graph does not fit in the main memory of one machine and caused all centralized algorithms to crash after several hours. Our MapReduce implementation scales out smoothly, since it splits the problem over multiple smaller problems solved in parallel.

Specifically, we used Amazon Wireless Services' Elastic MapReduce framework (aws). We implemented our `map` and `reduce` functions in Python and used the MRJob class (mrj). For our biggest experiments we used a 100-machine strong cluster, consisting of m1.xlarge AWS instances (a total of 800 cores).

The running times of our experiments over MapReduce are shown in Fig. 2(b). The main bottleneck is the computation of the first two eigenvectors which is performed by repeating the power iteration for few (typically 4) iterations. This step is not the emphasis of this work and has not been optimized. The Spannogram algorithm is significantly faster and the benefits of parallelization are clear since it is CPU intensive.

In principle, the other algorithms could be also implemented over MapReduce, but that requires non-trivial distributed algorithm design. As is well-known, *e.g.*, (Meng & Mahoney), implementing iterative machine learning algorithms over MapReduce can be a significant task and schemes which perform worse in standard metrics can be highly preferable for this parallel

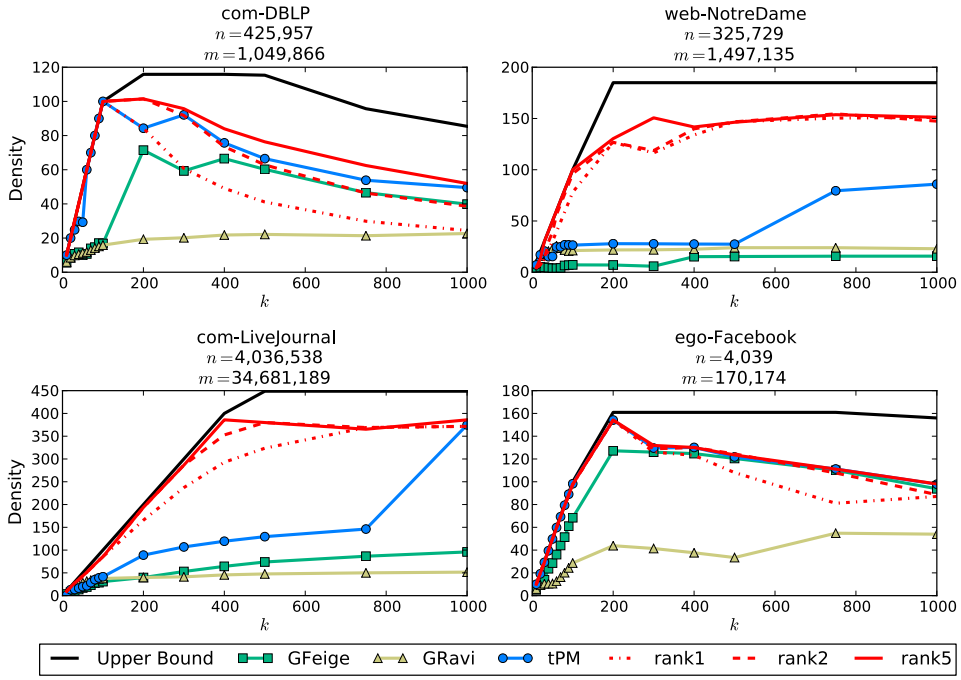


Figure 3. Subgraph density vs. subgraph size (k). We compare our DkS Spannogram algorithm with the algorithms from (Feige et al., 2001) (GFeige), (Ravi et al., 1994) (GRavi), and (Yuan & Zhang, 2011) (tPM). Across all subgraph sizes k, we obtain higher subgraph densities using Spannograms of rank $d = 2$ or 5. We also obtain a *provable data-dependent upper bound* (solid black line) on the objective. This proves that for these data sets, our algorithm is typically within 80% from optimality, for all sizes up to $k = 250$, and indeed for small subgraph sizes we find a clique which is clearly optimal. Further experiments on multiple other data sets are shown in the supplemental material.

framework. Careful MapReduce algorithmic design is needed especially for dense graphs like the one in the hidden clique problem.

Real Datasets. Next, we demonstrate our method's performance in real datasets and also illustrate the power of our data-dependent bounds. We run experiments on large graphs from different applications and our findings are presented in Fig. 3. The figure compares the density achieved by the Spannogram algorithm for rank 1, 2 and 5 to the performance of GFeige, GRavi and TPower. The figure shows that the rank-2 and rank-5 versions of our algorithm, improve – sometimes significantly – over the other techniques. Our novel data-dependent upper-bound shows that our results on these data sets are provably near-optimal.

The experiments are performed for two community graphs (com-LiveJournal and com-DBLP), a web graph (web-NotreDame), and a subset of the Facebook graph. A larger set of experiments is included in the supplemental material. Note that the largest graph in Figure 3 contains no more than 35 million edges; these cases fit in the main memory of a single machine and the running times are presented in the supplemental material, all performed on a standard Macbook Pro laptop using Matlab. In summary, rank-2 took less than one second for all these graphs while prior work methods took approximately the same time, up to a few seconds. Rank-1 was significantly faster than all other methods in all tested graphs and took fractions of a second. Rank-5 took up to 1000 seconds for the largest graph (LiveJournal).

We conclude that our algorithm is an efficient option for finding dense subgraphs. Different rank choices give a tradeoff between accuracy and performance while the parallel nature allows scalability when needed. Further, our theoretical upper-bound can be useful for practitioners investigating dense structures in large graphs.

6. Acknowledgments

The authors would like to acknowledge support from NSF grants CCF 1344364, CCF 1344179, DARPA XDATA, and research gifts by Google, Docomo and Microsoft.

References

Amazon Web Services, Elastic Map Reduce. URL http://aws.amazon.com/elasticmapreduce/.

MRJob. URL http://pythonhosted.org/mrjob/.

Alon, Noga, Lee, Troy, Shraibman, Adi, and Vempala, Santosh. The approximate rank of a matrix and its algorithmic applications: approximate rank. In *Proceedings of the 45th annual ACM symposium on Symposium on theory of computing*, pp. 675–684. ACM, 2013.

Ames, Brendan PW. *Convex relaxation for the planted clique, biclique, and clustering problems*. PhD thesis, University of Waterloo, 2011.

Arora, Sanjeev, Karger, David, and Karpinski, Marek. Polynomial time approximation schemes for dense instances of np-hard problems. In *STOC*, 1995.

Asahiro, Yuichi, Iwama, Kazuo, Tamaki, Hisao, and Tokuyama, Takeshi. Greedily finding a dense subgraph. *Journal of Algorithms*, 34(2):203–221, 2000.

Asteris, Megasthenis, Papailiopoulos, Dimitris S, and Karystinos, George N. The sparse principal component of a constant-rank matrix. *IEEE Trans. IT*, 60(4):228–2290, 2014.

Bahmani, Bahman, Kumar, Ravi, and Vassilvitskii, Sergei. Densest subgraph in streaming and mapreduce. *Proceedings of the VLDB Endowment*, 5(5):454–465, 2012.

Bhaskara, Aditya, Charikar, Moses, Chlamtac, Eden, Feige, Uriel, and Vijayaraghavan, Aravindan. Detecting high log-densities: an $\mathsf{O}(n^{1/4})$ approximation for densest k-subgraph. In *STOC*, 2010.

Boutsidis, Christos, Mahoney, Michael W, and Drineas, Petros. An improved approximation algorithm for the column subset selection problem. In *Proceedings of the twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 968–977. Society for Industrial and Applied Mathematics, 2009.

Cormen, Thomas H, Leiserson, Charles E, Rivest, Ronald L, and Stein, Clifford. *Introduction to algorithms*. MIT press, 2001.

d'Aspremont, Alexandre et al. Weak recovery conditions using graph partitioning bounds. 2010.

Dourisboure, Yon, Geraci, Filippo, and Pellegrini, Marco. Extraction and classification of dense communities in the web. In *WWW*, 2007.

Feige, Uriel and Langberg, Michael. Approximation algorithms for maximization problems arising in graph partitioning. *Journal of Algorithms*, 41(2):174–211, 2001.

Feige, Uriel, Peleg, David, and Kortsarz, Guy. The dense k-subgraph problem. *Algorithmica*, 29(3):410–421, 2001.

Gibson, David, Kumar, Ravi, and Tomkins, Andrew. Discovering large dense subgraphs in massive graphs. In *PVLDB*, 2005.

Hu, Haiyan, Yan, Xifeng, Huang, Yu, Han, Jiawei, and Zhou, Xianghong Jasmine. Mining coherent dense subgraphs across massive biological networks for functional discovery. *Bioinformatics*, 21(suppl 1):i213–i221, 2005.

Jethava, Vinay, Martinsson, Anders, Bhattacharyya, Chiranjib, and Dubhashi, Devdatt. The lovasz theta function, svms and finding large dense subgraphs. In *NIPS*, 2012.

Karystinos, George N and Liavas, Athanasios P. Efficient computation of the binary vector that maximizes a rank-deficient quadratic form. *IEEE Trans. IT*, 56(7):3581–3593, 2010.

Khot, Subhash. Ruling out ptas for graph min-bisection, densest subgraph and bipartite clique. In *FOCS*, 2004.

Lin, Jimmy and Schatz, Michael. Design patterns for efficient graph algorithms in mapreduce. In *Proceedings of the Eighth Workshop on Mining and Learning with Graphs*, pp. 78–85. ACM, 2010.

Mahoney, Michael W and Drineas, Petros. Cur matrix decompositions for improved data analysis. *Proceedings of the National Academy of Sciences*, 106(3):697–702, 2009.

Meng, Xiangrui and Mahoney, Michael W. Robust regression on mapreduce. *ICML 2013, (to appear)*.

Miller, B, Bliss, N, and Wolfe, P. Subgraph detection using eigenvector l1 norms. In *NIPS*, 2010.

Papailiopoulos, Dimitris S, Dimakis, Alexandros G, and Korokythakis, Stavros. Sparse pca through low-rank approximations. *arXiv preprint arXiv:1303.0551*, 2013.

Ravi, Sekharipuram S, Rosenkrantz, Daniel J, and Tayi, Giri K. Heuristic and special case algorithms for dispersion problems. *Operations Research*, 42(2):299–310, 1994.

Saha, Barna, Hoch, Allison, Khuller, Samir, Raschid, Louiqa, and Zhang, Xiao-Ning. Dense subgraphs with restrictions and applications to gene annotation graphs. In *Research in Computational Molecular Biology*, pp. 456–472. Springer, 2010.

Srivastav, Anand and Wolf, Katja. *Finding dense subgraphs with semidefinite programming*. Springer, 1998.

Suzuki, Akiko and Tokuyama, Takeshi. Dense subgraph problems with output-density conditions. In *Algorithms and Computation*, pp. 266–276. Springer, 2005.

Yuan, Xiao-Tong and Zhang, Tong. Truncated power method for sparse eigenvalue problems. *arXiv preprint arXiv:1112.2679*, 2011.

Supplemental Material for: Finding Dense Subgraphs via Low-Rank Bilinear Optimization

1. Proof of Lemma 4: Building the set $\mathcal{S}_d$ for arbitrary d-dimensional subspaces

In our general case, we solve DBkS on

$$\mathbf{A}_d = \mathbf{V}_d \mathbf{U}_d^T = \sum_{i=1}^{d} \mathbf{v}_i \mathbf{u}_i^T$$

where

$$\mathbf{V}_d = [v_1 \ \ \ldots \ \ v_d] \text{ and } \mathbf{U}_d = [\lambda_1 \cdot v_1 \ \ \ldots \ \ \lambda_d \cdot v_d] .$$

Solving the problem on $\mathbf{A}_d$ is equivalent to answering the following combinatorial question:

"how many different top-k supports are there in a d-dimensional subspace: $top_k(c_1 \cdot \mathbf{v}_1 + \ldots + c_d \cdot \mathbf{v}_d)$?"

Here we define $d-1$ auxiliary angles $\phi_1, \ldots, \phi_{d-1} \in \Phi = [0, \ \pi)$ and we rewrite the coefficients $c_1, \ldots, c_d$ as

$$\mathbf{c} = \begin{bmatrix} c_1 \\ \vdots \\ c_d \end{bmatrix} = \begin{bmatrix} \sin\phi_1 \\ \cos\phi_1 \sin\phi_2 \\ \vdots \\ \cos\phi_1 \cos\phi_2 \ldots \sin\phi_{d-1} \\ \cos\phi_1 \cos\phi_2 \ldots \cos\phi_{d-1} \end{bmatrix}.$$

Clearly we can express every vector in the span of $\mathbf{V}_d$ as a linear combination $c_1 \cdot \mathbf{v}_1 + \ldots + c_d \cdot \mathbf{v}_d$ in terms of ϕ:

$$\mathbf{v}(\phi_1, \ldots, \phi_{d-1}) = (\sin\phi_1) \cdot \mathbf{v}_1 + (\cos\phi_1 \sin\phi_2) \cdot \mathbf{v}_2 + \ldots + (\cos\phi_1 \cos\phi_2 \ldots \cos\phi_{d-1}) \cdot \mathbf{v}_d. \tag{1}$$

For notation simplicity let us define a vector that contains all $d-1$ auxiliary phase variables

$$\varphi = [\phi_1, \ldots, \phi_{d-1}].$$

We can use the above derivations to rewrite the set $\mathcal{S}_d$ that contains all top k coordinates in the span of $\mathbf{V}_d$ as:

$$\begin{aligned} \mathcal{S}_d &= \{\text{top}_k(c_1 \cdot \mathbf{v}_1 + \ldots + c_d \cdot \mathbf{v}_d) : c_1, \ldots, c_d \in \mathbb{R}\} \\ &= \{\text{top}_k \pm (\mathbf{v}(\varphi)) : \varphi \in \Phi^{d-1}\} \\ &= \{\text{top}_k \pm ((\sin\phi_1) \cdot \mathbf{v}_1 + (\cos\phi_1 \sin\phi_2) \cdot \mathbf{v}_2 + \ldots + (\cos\phi_1 \cos\phi_2 \ldots \cos\phi_{d-1}) \cdot \mathbf{v}_d), \varphi \in \Phi^{d-1}\} \end{aligned}$$

Observe again that each element of $\mathbf{v}(\varphi)$ is a continuous *spectral curve* in the $d-1$ auxiliary variables:

$$[\mathbf{v}(\varphi)]_i = (\sin\phi_1) \cdot [\mathbf{v}_1]_i + (\cos\phi_1 \sin\phi_2) \cdot [\mathbf{v}_2]_i + \ldots + (\cos\phi_1 \cos\phi_2 \ldots \cos\phi_{d-1}) \cdot [\mathbf{v}_d]_i.$$

Consequently, the top/bottom-k supports of $\mathbf{v}(\varphi)$ (*i.e.*, $\text{top}_k(\pm\mathbf{v}(\varphi))$) are themselves a function of the $d-1$ variables in φ. How can we find all possible supports?

Remark 1. *In our general problem we wish to find all top and bottom k coordinates that appear in a d-dimensional subspace. In the following discussion, for simplicity we handle the top k coordinates problem. Finding the bottom k trivially follows, by just checking the smallest k coordinates of each vector $c_1 \cdot \mathbf{v}_1 + \ldots + c_d \cdot \mathbf{v}_d$ that we construct using our algorithm.*

1.1. Ranking regions for a single coordinate $[\mathbf{v}(\varphi)]_i$

We now show that for each single coordinate $[\mathbf{v}(\varphi)]_i$, we can partition Φ^{d-1} in regions, wherein the ith coordinate $[\mathbf{v}(\varphi)]_i$ retains the same ranking relative to the other $n-1$ coordinates in the vector $\mathbf{v}(\varphi)$.

Let us first consider for simplicity $[\mathbf{v}(\varphi)]_1$. We aim to find all values of φ where $[\mathbf{v}(\varphi)]_1$ is in one of the the k largest coordinates of $\mathbf{v}(\varphi)$. We observe that this region can be characterized by using n boundary tests:

$$\begin{gathered}[\mathbf{v}(\varphi)]_1 \gtrless [\mathbf{v}(\varphi)]_2 \\ [\mathbf{v}(\varphi)]_1 \gtrless [\mathbf{v}(\varphi)]_3 \\ \vdots \\ [\mathbf{v}(\varphi)]_1 \gtrless [\mathbf{v}(\varphi)]_n\end{gathered}$$

Each of the above boundary tests defines a bounding curve that partitions the Φ^{d-1} domain. We refer to this bounding curve as $\mathcal{B}_{1,j}(\varphi) : \Phi^{d-1} \mapsto \Phi^{d-2}$. A $\mathcal{B}_{1,j}(\varphi)$ curve partitions Φ and defines two regions of φ angles:

$$\mathcal{R}_{1>j} = \{\varphi \in \Phi^{d-1} : [\mathbf{v}(\varphi)]_1 > [\mathbf{v}(\varphi)]_j\} \text{ and } \mathcal{R}_{1\leq j} = \{\varphi \in \Phi^{d-1} : [\mathbf{v}(\varphi)]_1 \leq [\mathbf{v}(\varphi)]_j\} \tag{2}$$

such that $\mathcal{R}_{1>j} \cup \mathcal{R}_{1\leq j} = \Phi^{d-1}$.

Observe that these $n-1$ curves $\mathcal{B}_{1,1}(\varphi), \ldots, \mathcal{B}_{1,n}(\varphi)$ partition Φ in disjoint *cells*, $\mathcal{C}^1_1, \ldots, \mathcal{C}^1_T$, such that

$$\bigcup_{i=1}^{T} \mathcal{C}^1_i = \Phi^{d-1}.$$

Within each cell $\mathcal{C}^1_i$, the first coordinate $[\mathbf{v}(\varphi)]_1$ retains a fixed ranking relative to the rest of the elements in $\mathbf{v}(\varphi)$, e.g., for a specific cell it might be the largest element, and in another cell it might be the 10th smallest, etc. This happens because for all values of φ in a single cell, the respective ordering $[\mathbf{v}(\varphi)]_1 \gtrless [\mathbf{v}(\varphi)]_2, \ \ldots, \ [\mathbf{v}(\varphi)]_1 \gtrless [\mathbf{v}(\varphi)]_n$ remains the same.

If we have access to a single point, say φ_0, that belongs to a specific cell, say $\mathcal{C}^1_j$, then we can calculate $[\mathbf{v}(\varphi_0)]$ and find the ranking of the first coordinate $[\mathbf{v}(\varphi)]_1$, that remains invariant for all $\varphi \in \mathcal{C}^1_j$. Hence, if we visit all these cells, then we can find all possible rankings that the first coordinate $[\mathbf{v}(\varphi)]_1$ takes in the d-dimensional span of $\mathbf{v}_1, \ldots, \mathbf{v}_d$. In the following subsections, we show that the number of these cells is bounded by $T \leq 2^d \binom{n-1}{d-1}$.

Observe that each bounding curve $\mathcal{B}_{1,i}(\varphi)$ has a one-to-one correspondence to an equation $[\mathbf{v}(\varphi)]_1 = [\mathbf{v}(\varphi)]_j$, which is linear in $\mathbf{c}$:

$$[\mathbf{v}(\varphi)]_1 = [\mathbf{v}(\varphi)]_j \Rightarrow \mathbf{e}_1^T \mathbf{V}_d \mathbf{c} - \mathbf{e}_i^T \mathbf{V}_d \mathbf{c} = 0 \Rightarrow (\mathbf{e}_1 - \mathbf{e}_j)^T \mathbf{V}_d \mathbf{c} = 0. \tag{3}$$

Due to their linear characterization with respect to $\mathbf{c}$, it is easy to see that each $(d-1)$-tuple of bounding curves intersects on a single point in Φ^{d-1}:[1]

$$\begin{array}{rcl} [\mathbf{v}(\varphi)]_1 = [\mathbf{v}(\varphi)]_{i_1} & (\mathbf{e}_1 - \mathbf{e}_{i_1})^T \mathbf{V}_d \mathbf{c} = 0 & \\ [\mathbf{v}(\varphi)]_1 = [\mathbf{v}(\varphi)]_{i_2} & (\mathbf{e}_1 - \mathbf{e}_{i_2})^T \mathbf{V}_d \mathbf{c} = 0 & \\ \vdots & \vdots & \\ [\mathbf{v}(\varphi)]_1 = [\mathbf{v}(\varphi)]_{i_{d-1}} & (\mathbf{e}_1 - \mathbf{e}_{i_{d-1}})^T \mathbf{V}_d \mathbf{c} = 0 & \end{array} \Rightarrow \begin{bmatrix} (\mathbf{e}_1 - \mathbf{e}_{i_1})^T \\ (\mathbf{e}_1 - \mathbf{e}_{i_2})^T \\ \vdots \\ (\mathbf{e}_1 - \mathbf{e}_{i_{d-1}})^T \end{bmatrix} \mathbf{V}_d \mathbf{c} = \mathbf{0}_{(d-1)\times 1}.$$

Let us denote the solution of the above linear inverse problem as $\mathbf{c}_{1,i_1,\ldots,i_{d-1}}$. We refer to $\mathbf{c}_{1,i_1,\ldots,i_{d-1}}$ as an intersection vector. For each intersection vector $\mathbf{c}_{1,i_1,\ldots,i_{d-1}}$, we can compute its polar expression and solve for the angles φ that generate it. These $d-1$ input angles correspond exactly to the intersection point of $d-1$ curves specified by the above $d-1$ equations. We denote these $d-1$ angles that generate $\mathbf{c}_{1,i_1,\ldots,i_{d-1}}$, as $\varphi_{1,i_1,\ldots,i_{d-1}}$ which we refer to as the intersection point of the $d-1$ curves $\mathcal{B}_{1,i_1}(\varphi), \ldots, \mathcal{B}_{1,i_{d-1}}(\varphi)$.

[1] as a matter of fact, due to the sign ambiguity of the solution, this corresponds to two intersection points. However, the following discussion omits this technical detail for simplicity.

Since, the $\varphi_{1,i_1,\ldots,i_{d-1}}$ intersection points are defined for every $d-1$ curves, the total number of intersection points is $\binom{n-1}{d-1}$. In the following subsections, we show how we can visit all cells by just examining these intersection points.

We proceed to show that if we visit the adjacent cells of the intersection points defined for all coordinates, then we can find all top-k supports in the span of $\mathbf{V}_d$.

1.2. Visiting all cells = finding all top k supports

Our goal is to find all top-k supports that can appear in the span of $\mathbf{V}_d$. To do so, it is sufficient to visit the cells where $[\mathbf{v}(\varphi)]_1$ is the k-th largest coordinate, then the cells where $[\mathbf{v}(\varphi)]_2$ is the k-largest, and so on. Within such cells, one coordinate (say $[\mathbf{v}(\varphi)]_i$) remains always the k-th largest, while the identities of the bottom $n-k$ coordinates remain the same. This means that in such a cell, we have that

$$[\mathbf{v}(\varphi)]_i \geq [\mathbf{v}(\varphi)]_{j_1}, \ldots, [\mathbf{v}(\varphi)]_i \geq [\mathbf{v}(\varphi)]_{j_{n-k}}$$

for all φ in that cell and some scecific $n-k$ other coordinates indexed by $j_1, \ldots, j_{n-k}$. Hence, although the sorting of the top $k-1$ elements might change in that cell (i.e., the first might become the second largest, and vice versa), the coordinates that participate in the top $k-1$ support will be the same, while at the same time the k-th largest will be $[\mathbf{v}(\varphi)]_i$.

Hence, for each coordinate $[\mathbf{v}(\varphi)]_i$, we need to visit the cells wherein it is the k-th largest. We do this by examining *all* cells wherein $[\mathbf{v}(\varphi)]_i$ retains a fixed ranking. Visiting all these cells (T for each coordinate), is possible by visiting all $n \cdot \binom{n-1}{d-1}$ intersection points of $\mathcal{B}_{i,j}(\varphi)$ curves as defined earlier. Since we know that each cell is adjacent to at least 1 intersection point, then at each of these points we visit all adjacent cells. For each cell that we visit, we compute the support of the largest k coordinates of a vector $\mathbf{v}(\varphi_0)$ with a φ_0 that lies in that cell. We include this top k index set in $\mathcal{S}_d$ and carry the same procedure for all cells. Since we visit all coordinates and all their adjacent cells, this means that we visit all cells $\mathcal{C}_j^i$. This means that this procedure will construct all possible supports in

$$\mathcal{S}_d = \{\text{top}_k(c_1 \cdot \mathbf{v}_1 + \ldots + c_d \cdot \mathbf{v}_d) : c_1, \ldots, c_d \in \mathbb{R}\}$$

1.3. Constructing the set $\mathcal{S}_d$

To visit all possible cells $\mathcal{C}_j^i$, we now have to check the intersection points, which are obtained by solving the system of $d-1$ equations

$$\begin{aligned} &[\mathbf{v}(\varphi)]_{i_1} = [\mathbf{v}(\varphi)]_{i_2} = \ldots = [\mathbf{v}(\varphi)]_{i_d} \\ \Leftrightarrow &[\mathbf{v}(\varphi)]_{i_1} = [\mathbf{v}(\varphi)]_{i_2}, \ldots, [\mathbf{v}(\varphi)]_{i_1} = [\mathbf{v}(\varphi)]_{i_d}. \end{aligned} \tag{4}$$

We can rewrite the above as

$$\begin{bmatrix} \mathbf{e}_{i_1}^T - \mathbf{e}_{i_2}^T \\ \vdots \\ \mathbf{e}_{i_1}^T - \mathbf{e}_{i_d}^T \end{bmatrix} \mathbf{V}_d \mathbf{c} = \mathbf{0}_{(d-1)\times 1} \tag{5}$$

where the solution is the nullspace of the matrix, which has dimension 1.

To explore all possible candidate vectors, we need to visit all cells. To do so, we compute all possible $\binom{n}{d}$ solution intersection vectors $\mathbf{c}_{i_1,\ldots,i_d}$. On each intersection vector we need to compute the locally optimal support set

$$\text{top}_k\left(\mathbf{V}_d \mathbf{c}_{i_1,\ldots,i_d}\right).$$

Then observe that the coordinates $i_1, \ldots, i_d$ of $\mathbf{V}_d \mathbf{c}_{i_1,\ldots,i_d}$ have the same value, since they all satisfy equation (5). Let us assume that t of them appear in the set $\text{top}_k\left(\mathbf{V}_d \mathbf{c}_{i_1,\ldots,i_d}\right)$. The, finding the top k supports of all neighboring cell is equivalent to checking all different supports that can be generated by taking all $\binom{d}{t}$ possible t-subsets of the $i_1, \ldots, i_d$ coordinates with respect to $\mathbf{V}_d \mathbf{c}_{i_1,\ldots,i_d}$, while keeping the rest of the elements in $\mathbf{V}_d \mathbf{c}_{i_1,\ldots,i_d}$ in their original ranking, as computed in $\text{top}_k\left(\mathbf{V}_d \mathbf{c}_{i_1,\ldots,i_d}\right)$. This, induces at most $O(\binom{d}{d/2})$ local sortings, i.e.,

top k supports. All these sortings will eventually be the elements of the $\mathcal{S}_d$ set. The number of all candidate support sets will now be $O(\binom{d}{d/2}\binom{n-1}{d}) = O(n^d)$ and the total computation complexity is $O\left(n^{d+1}\right)$, since for each point we compute the top-k support in linear time $O(n)$.

For completeness the algorithm of the spannogram framework that generates $\mathcal{S}_d$ is given below.

Algorithm 1 Spannogram Algorithm for $\mathcal{S}_d$.

$\mathcal{S}_d = \emptyset$
for all $(i_1, \ldots, i_d) \in \{1, \ldots, n\}^d$ and $s \in \{-1, 1\}$ **do**
  $\mathbf{c} = s \cdot \text{nullspace}\left(\begin{bmatrix} [(\mathbf{V}_d]_{i_1,:} - [\mathbf{V}_d]_{i_2,:} \\ \vdots \\ [\mathbf{V}_d]_{i_1,:} - [\mathbf{V}_d]_{i_d,:} \end{bmatrix}\right)$
  $\mathbf{v} = \mathbf{V}_d^T \mathbf{c}$
  $\mathcal{S} = \text{top}_k(\mathbf{v})$
  $\mathcal{T} = \mathcal{S} - \{i_1, \ldots, i_d\}$
  for all $\binom{d}{d-|\mathcal{T}|}$ subsets $\mathcal{J}$ of $(i_1, \ldots, i_d)$ **do**
    $\mathcal{S}_d = \mathcal{S}_d \bigcup (\mathcal{T} \cup \mathcal{J})$
  end for
end for
Output: $\mathcal{S}_d$.

1.4. Resolution of singularities

In our proofs, we assumed that the curves in $\mathbf{v}(\phi)$ are in general position. This is needed so that no more than $d-1$ curves intersect at a single point. This assumption is equivalent to requiring that every $d \times d$ submatrix of $\mathbf{V}_d$ is full rank. This "general position" requirement can be handled by introducing infinitesimal perturbations in V_d. The details of the analysis of this method can be found in (Papailiopoulos et al., 2013).

2. Going from DkS to DBkS and back

In this subsection we show how a ρ-approximation algorithm for DBkS for arbitrary matrices, implies a 2ρ-approximation for DkS. Our proof goes through a randomized sampling argument.

At the end of this section we also show a deterministic scheme that coverts a DBkS algorithm to an algorithm for DkS. This deterministic translation only guarantees that a ρ-approximation algorithm for DBkS for arbitrary matrices, can be converted to 4ρ-approximation for DkS. Our deterministic method uses much simpler vertex pruning techniques.

Algorithm 2 `randombipartite`$(\mathcal{G})$

$\mathcal{L} = \emptyset$, $\mathcal{R} = \emptyset$
for Each vertex v in $\mathcal{G}$ **do**
  Z = Bernoulli(1/2).
  if Z==1 **then**
    Put v in $\mathcal{L}$
  else
    Put v in $\mathcal{R}$
  end if
end for
$\mathcal{G}_{\mathsf{B}} = \mathcal{G}$
delete all edges in $\mathcal{G}_{\mathsf{B}}(\mathcal{L})$ and $\mathcal{G}_{\mathsf{B}}(\mathcal{R})$
Output: $\mathcal{G}_{\mathsf{B}}$

2.1. Proof of Lemma 1: Randomized Reduction

Let us denote by $\mathcal{G}(\mathcal{S})$ the subgraph in $\mathcal{G}$ induced by a vertex set $\mathcal{S}$. Let the adjacency matrix of the bipartite graph created by $\texttt{randombipartite}(\mathcal{G})$ be $\mathcal{G}_{\mathsf{B}}$

$$\mathbf{A}_{\mathsf{B}} = \begin{bmatrix} \mathbf{0}_{n_1 \times n_2} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{0}_{n_2 \times n_1} \end{bmatrix},$$

where $n_1 + n_2 = n$. In the following, we refer to $\mathbf{B}$ as the bi-adjacency matrix of the bipartite graph $\mathcal{G}_{\mathsf{B}}$. Moreover, we denote as $\mathcal{L}$ and $\mathcal{R}$ the two disjoint vertex sets of a bipartite graph.

Before we proceed let us state a simple property on the quadratic form of bipartite graphs.

Proposition 1. *Let $\mathbf{A}_{\mathsf{B}} = \left[\begin{smallmatrix} \mathbf{0}_{n_1 \times n_2} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{0}_{n_2 \times n_1} \end{smallmatrix}\right]$ be the adjacency matrix of a bipartite graph. Then, for any subset of vertices $\mathcal{S}$, we have that $\mathcal{S} = \mathcal{S}_l \cup \mathcal{S}_r$, with $\mathcal{S}_l = \mathcal{S} \cap \mathcal{L}$ and $\mathcal{S}_r = \mathcal{S} \cap \mathcal{R}$ Moreover,*

$$\mathbf{1}_{\mathcal{S}}^T \mathbf{A}_{\mathsf{B}} \mathbf{1}_{\mathcal{S}} = 2 \cdot \mathbf{1}_{\mathcal{S}_l}^T \mathbf{B} \mathbf{1}_{\mathcal{S}_r}$$

Proof. It is easy to see that $\mathcal{S}_l$ and $\mathcal{S}_r$ are the vertex subsets of $\mathcal{S}$ that correspond to either the left or right nodes of the bipartite graph. Since the two sets are disjoint, we have

$$\mathbf{1}_{\mathcal{S}} = \mathbf{1}_{\mathcal{S}_l} + \mathbf{1}_{\mathcal{S}_r}.$$

Then, the quadratic forms on $\mathbf{A}_{\mathsf{B}}$ can be equivalently rewritten as bilinear forms on $\mathbf{B}$:

$$\mathbf{1}_{\mathcal{S}}^T \mathbf{A}_{\mathsf{B}} \mathbf{1}_{\mathcal{S}} = (\mathbf{1}_{\mathcal{S}_l} + \mathbf{1}_{\mathcal{S}_r})^T \begin{bmatrix} \mathbf{0}_{n_1 \times n_2} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{0}_{n_2 \times n_1} \end{bmatrix} (\mathbf{1}_{\mathcal{S}_l} + \mathbf{1}_{\mathcal{S}_r}) = \mathbf{1}_{\mathcal{S}_r} \mathbf{B}^T \mathbf{1}_{\mathcal{S}_l} + \mathbf{1}_{\mathcal{S}_l} \mathbf{B} \mathbf{1}_{\mathcal{S}_r} = 2 \cdot \mathbf{1}_{\mathcal{S}_l}^T \mathbf{B} \mathbf{1}_{\mathcal{S}_r}.$$

□

Due to the above, we will consider the following DBkS problem

$$\{\mathcal{X}_{\mathsf{B}}, \mathcal{Y}_{\mathsf{B}}\} = \arg \max_{|\mathcal{X}|=k_1} \max_{|\mathcal{Y}|=k_2} \mathbf{1}_{\mathcal{X}}^T \mathbf{B} \mathbf{1}_{\mathcal{Y}},$$

where $k_1 + k_2 = k$. Due to Proposition 1, $\mathcal{X}_{\mathsf{B}} \cap \mathcal{Y}_{\mathsf{B}} = \emptyset$, since the columns and rows of $\mathbf{B}$ index two disjoint vertex sets $\mathcal{L}$ and $\mathcal{R}$, respectively. Then, our approximate solution with respect to the original DkS problem on $\mathbf{A}$ will then be

$$\mathcal{S}_{\mathsf{B}} = \mathcal{X}_{\mathsf{B}} \cup \mathcal{Y}_{\mathsf{B}}.$$

Clearly $|\mathcal{S}_{\mathsf{B}}| = k$.

Proposition 2. *The density of the above approximate solution is*

$$\mathsf{den}(\mathcal{S}_{\mathsf{B}}) = \frac{\mathbf{1}_{\mathcal{S}_{\mathsf{B}}}^T \mathbf{A} \mathbf{1}_{\mathcal{S}_{\mathsf{B}}}}{k} \geq \frac{\mathbf{1}_{\mathcal{S}_{\mathsf{B}}}^T \mathbf{A}_{\mathsf{B}} \mathbf{1}_{\mathcal{S}_{\mathsf{B}}}}{k} = 2 \cdot \frac{\mathbf{1}_{\mathcal{X}_{\mathsf{B}}}^T \mathbf{B} \mathbf{1}_{\mathcal{Y}_{\mathsf{B}}}}{k}.$$

Proof. The result follows immediately by the nonnegativity of the entries in $\mathbf{A}$, and the fact that $\mathbf{A}_{\mathsf{B}}$ contains a subset of the entries of $\mathbf{A}$. The last equality follows from Proposition 1. □

We will now show that $\mathsf{den}(\mathcal{S}_{\mathsf{B}})$ is at least $\mathsf{opt}/(2+\delta)$ with high probability, if we solve DBkS on $\log n$ graphs independently created using $\texttt{randombipartite}(\mathcal{G})$, and by keeping the best solution among all $\log n$ extracted sets $\mathcal{S}_{\mathsf{B}}$.

Proposition 3. *Let us partition the vertices of $\mathcal{G}$ into two sets $\mathcal{L}, \mathcal{R}$ according to $\texttt{randombipartite}(\mathcal{G})$: We flip a fair coin for each vertex of $\mathcal{G}$ and put it with probability 1/2 in either of the two sets. Then, we create a bipartite graph $\mathcal{G}_{\mathsf{B}}$ that has as left and right vertex sets the sets $\mathcal{L}$ and $\mathcal{R}$, respectively. The edges that we maintain from the original graph, are only those that cross from $\mathcal{L}$ to $\mathcal{R}$, while we delete all edges in the subgraphs induced by both sets $\mathcal{L}$ and $\mathcal{R}$. Then, there exists in a k-subgraph in $\mathcal{G}_{\mathsf{B}}$, that contains $0.5 \cdot k \cdot \mathsf{opt}$ edges, in expectation.*

Proof. First observe that we can represent the edges of $\mathcal{G}_{\mathsf{B}}$ as random variables $X_{i,j}$. If (i,j) is not an edge in $\mathcal{G}$, then $X_{i,j}$ will be 0 with probability 1. If however (i,j) is an edge in $\mathcal{G}$, then $X_{i,j}$ is 1, i.e., appears in $\mathcal{G}_{\mathsf{B}}$, with the same probability that one of its vertices lands in $\mathcal{L}$, while the second is in $\mathcal{R}$. It is easy to find that this probability is $\Pr\{X_{i,j}=1\}=1/2$. Hence,

$$X_{i,j}=\begin{cases} 0, & \text{if } (i,j) \text{ not an edge in } \mathcal{G},\\ Z, & \text{if } (i,j) \text{is an edge in } \mathcal{G},\end{cases} \tag{6}$$

where Z is a Bernoulli(1/2) random variable.

Now let $\mathcal{S}_*$ denote the vertex set of the densest k-subgraph on the original graph $\mathcal{G}$, that has density $\mathsf{den}(\mathcal{S}_*)=\mathsf{opt}$. Observe that for that subgraph we have

$$\mathbf{1}_{\mathcal{S}_*}^T\mathbf{A}\mathbf{1}_{\mathcal{S}_*}=\sum_{i,j\in\mathcal{S}_*}A_{i,j}=k\cdot\mathsf{opt}.$$

Let $\mathbf{A}_{\mathsf{B}}$ denote the adjacency matrix of the bipartite graph $\mathcal{G}_{\mathsf{B}}$. Then, we have that the expected quadratic form on the new adjacency $\mathbf{1}_{\mathcal{S}_*}^T\mathbf{A}_{\mathsf{B}}\mathbf{1}_{\mathcal{S}_*}$ is:

$$\begin{aligned}E\left\{\mathbf{1}_{\mathcal{S}_*}^T\mathbf{A}_{\mathsf{B}}\mathbf{1}_{\mathcal{S}_*}\right\}&=E\left\{\sum_{i,j\in\mathcal{S}_*}X_{i,j}\right\}=E\left\{\sum_{i,j\in\mathcal{S}_*\text{and }(i,j)\in\mathcal{G}}X_{i,j}\right\}=E\left\{\sum_{i,j\in\mathcal{S}_*\text{and }(i,j)\in\mathcal{G}}Z\right\}\\&=\frac{1}{2}\cdot\sum_{i,j\in\mathcal{S}_*}A_{i,j}=0.5\cdot k\cdot\mathsf{opt}.\end{aligned} \tag{7}$$

□

We will now show that if we run $\texttt{randombipartite}(\mathcal{G})$ a total number of $3\log n\cdot\log\log n$ times, then with high probability, at least one $\mathcal{G}_{\mathsf{B}}$ will contain a k-subgraph with density at least $0.5\cdot\mathsf{opt}$. This will imply that the densest k subgraph of $\mathcal{G}_{\mathsf{B}}$ will have density *at least* $0.5\cdot\mathsf{opt}$.

Proposition 4. *Let us run* $\texttt{randombipartite}(\mathcal{G})$ $m=3\log\cdot\log\log n$ *times, and let us obtain each time a graph* $\mathcal{G}_{\mathsf{B}}^i$. *For each bipartite graph let* $\mathbf{A}_{\mathsf{B}}^i$ *denote its adjacency and* $\mathbf{B}_i$ *its bi-adjacency, let*

$$\{\mathcal{X}_i,\mathcal{Y}_i\}=\arg\max_{k1+k2=k}\max_{|\mathcal{X}|=k_1}\max_{|\mathcal{Y}|=k_2}\mathbf{1}_{\mathcal{X}}^T\mathbf{B}_i\mathbf{1}_{\mathcal{Y}},$$

and let $\mathcal{S}_i=\mathcal{X}_i\cup\mathcal{Y}_i$ *denote the optimal k-subgraph of* $\mathcal{G}_{\mathsf{B}}^i$. *Then, with probability at least* $1-\frac{1}{\log n}$, *we have*

$$\max_{i=1,\ldots,m}\mathsf{den}(\mathcal{S}_i)=\max_{i=1,\ldots,m}\mathbf{1}_{\mathcal{S}_i}^T\mathbf{A}\mathbf{1}_{\mathcal{S}_i}\geq\frac{\mathsf{opt}}{2+1/\log n}.$$

Proof. In this proof we will use the reverse Markov Inequality which states that for any random variable X, such that $X\leq m$, then, for any $a\leq E\{X\}$, we have

$$\Pr\{X\leq a\}\leq\frac{m-E\{X\}}{m-a}.$$

Let $\mathcal{S}_*$ denote the densest k-subgraph for $\mathcal{G}$. Here, our random variable will be the the quadratic form

$$X=\mathbf{1}_{\mathcal{S}_*}^T\mathbf{A}_{\mathsf{B}}^i\mathbf{1}_{\mathcal{S}_*}.$$

Due to Proposition 3, we have that $E\{X\}=0.5\cdot k\cdot\mathsf{opt}$. Moreover, we know that

$$\mathbf{1}_{\mathcal{S}_*}^T\mathbf{A}_{\mathsf{B}}^i\mathbf{1}_{\mathcal{S}_*}\leq\mathbf{1}_{\mathcal{S}_*}^T\mathbf{A}\mathbf{1}_{\mathcal{S}_*}=k\cdot\mathsf{opt}.$$

Hence, we will set $\delta = 1/\log n$ and use the reverse Markov inequality:

$$\begin{aligned}\Pr\left\{\mathbf{1}_{\mathcal{S}_*}^T \mathbf{A}_{\mathsf{B}}^i \mathbf{1}_{\mathcal{S}_*} \leq k \cdot \mathsf{opt}/(2+1/\log n)\right\} &\leq \frac{k\cdot\mathsf{opt} - k\cdot\mathsf{opt}/2}{k\cdot\mathsf{opt} - k\cdot\mathsf{opt}/(2+1/\log n)} \\ &= \frac{1-1/2}{1-1/(2+1/\log n)} = \frac{1/2}{\frac{1+1/\log n}{2+1/\log n}} \\ &= \frac{2+1/\log n}{2+2/\log n} = \frac{2\log n+1}{2\log n+2} = 1 - \frac{1}{2(\log n+1)} \leq 1 - \frac{1}{3\log n}.\end{aligned}$$

Now, let B_i denote the event that $\{\mathbf{1}_{\mathcal{S}_*}^T \mathbf{A}_{\mathsf{B}}^i \mathbf{1}_{\mathcal{S}_*} \leq k\cdot\mathsf{opt}/(2+1/\log n)\}$. We would like to find a number of draws m such that, with high probability $1 - \frac{1}{\log n}$, we will have one graph $\mathcal{G}_{\mathsf{B}}^i$ satisfying

$$\mathbf{1}_{\mathcal{S}_*}^T \mathbf{A}_{\mathsf{B}}^i \mathbf{1}_{\mathcal{S}_*} > k\cdot\mathsf{opt}/(2+1/\log n).$$

Since these m events are independent for different graphs, it is easy to see find the number of draws m by solving the equation:

$$\left(1 - \frac{1}{3\log n}\right)^m \leq \frac{1}{\log n} \Rightarrow m\cdot\log\left(1-\frac{1}{3\log n}\right) \leq -\log(\log n) \Rightarrow m \geq \frac{-\log(\log n)}{\log\left(1-\frac{1}{3\log n}\right)} \Rightarrow m \geq 3\log n\cdot\log\log n.$$

where in the last step we used the fact that $-\log(1-\epsilon) > \epsilon$, for $\epsilon < 1$.

The above derivations tell us that if we draw $m = 3\log n\cdot\log\log n$ graphs $\mathcal{G}_{\mathsf{B}}^i$ using $\texttt{randombipartite}(\mathcal{G})$, then at least one of them contains a subgraph of density

$$\frac{\mathsf{opt}}{2+1/\log n}.$$

Hence, the subgraph $\mathcal{S}_i = \mathcal{X}_i \cup \mathcal{Y}_i$ with the highest density among the m densest k-subgraph of the graphs $\mathcal{G}_{\mathsf{B}}^i$ has density at least $\frac{\mathsf{opt}}{2+1/\log n}$. □

Using Proposition 4, we can now establish Lemma 1 in the following way. Suppose that we have an approximation algorithm such that for any k_1, k_2, we can obtain a subgraph with density $\mathcal{A}$

$$\frac{1}{\rho}\cdot\arg\max_{|\mathcal{X}|=k_1}\max_{|\mathcal{Y}|=k_2} \mathbf{1}_{\mathcal{X}}^T\mathbf{B}\mathbf{1}_{\mathcal{Y}} \leq \mathcal{A} \leq \arg\max_{|\mathcal{X}|=k_1}\max_{|\mathcal{Y}|=k_2} \mathbf{1}_{\mathcal{X}}^T\mathbf{B}\mathbf{1}_{\mathcal{Y}}$$

Then, we could use it with Proposition 4, to obtain a $(2+\delta)\rho$ approximation algorithm for DkS, for any $\delta > 1/n^\gamma$, and $\gamma = \Theta(1)$, in polynomial time.

What we show in our approximation results is that we can compute a solution with density at least

$$\arg\max_{|\mathcal{X}|=k_1}\max_{|\mathcal{Y}|=k_2} \mathbf{1}_{\mathcal{X}}^T\mathbf{B}\mathbf{1}_{\mathcal{Y}} - 2|\lambda_{d+1}|,$$

where λ_{d+1} is the $d+1$ absolutely largest eigenvalue of the adjacency matrix $\mathbf{A}$.

2.2. Deterministic Reduction

Here, we show a simper translation of a DBkS algorithm to a DkS algorithm.

Proposition 5. *Let $\mathcal{X}, \mathcal{Y}$ be two vertex sets of size k, and let $\mathcal{Z} = \mathcal{X} \cup \mathcal{Y}$. Then, we can use* $\texttt{prune}(\mathcal{G}, \mathcal{X}_*, \mathcal{Y}_*, k)$ *to construct in time $O(k^2)$ a vertex set $\mathcal{Z}'$ of size k, that has density at least*

$$\left(\frac{1}{4} - O\left(\frac{1}{k}\right)\right)\mathsf{den}(\mathcal{Z}).$$

Algorithm 3 prune($\mathcal{G}$, $\mathcal{X}$, $\mathcal{Y}$, k)

Input: $\mathcal{G}$, $\mathcal{X}$, $\mathcal{Y}$
$\mathcal{Z}_1 = \mathcal{X} \cup \mathcal{Y}$, $k' = |\mathcal{Z}_1|$
for $i = 1 : k' - k$ **do**
  find v smallest degree vertex in $\mathcal{G}(\mathcal{Z}_1)$
  $\mathcal{Z}_{i+1} = \mathcal{Z}_i \backslash \{v\}$.
end for
Output: $\mathcal{Z} = \mathcal{Z}_{k'-k}$

Proof. Let k' denote the size of the set $\mathcal{Z}_1 = \mathcal{X} \cup \mathcal{Y}$. Clearly $k \leq k' \leq 2k$. Let us denote as $d_{\min}(\mathcal{S})$, the smallest degree in the subgraph induced by the vertex set $\mathcal{S}$. Then, observe that its minimum degree can be bounded as

$$d_{\min}(\mathcal{S}) \leq \frac{\mathsf{den}(\mathcal{S})}{|\mathcal{S}|}.$$

During the first iteration of prune, we delete the smallest degree vertex from $\mathcal{Z}_1$ and obtain $\mathcal{Z}_2$. This leads to a loss in density that is equal to $2 \cdot d_{\min}(\mathcal{Z}_1)$. Hence, the density of $\mathcal{Z}_2$ is lower bounded as

$$\mathsf{den}(\mathcal{Z}_2) = \mathsf{den}(\mathcal{Z}_1) - 2 \cdot d_{\min}(\mathcal{Z}_1) \geq \mathsf{den}(\mathcal{Z}_1) - 2 \cdot \frac{\mathsf{den}(\mathcal{Z}_1)}{k} = \frac{k-2}{k} \cdot \mathsf{den}(\mathcal{Z}_1).$$

This means that for any iteration i of the algorithm the density of the current subgraph induced by $\mathcal{Z}_i$ is at least

$$\mathsf{den}(\mathcal{Z}_{i+1}) \geq \frac{(k'-i+1)-2}{k-i+1} \cdot \mathcal{Z}_i = \frac{k-i-1}{k-i+1} \cdot \mathcal{Z}_i.$$

Hence, the output vertex set $\mathcal{Z}$ satisfies

$$\mathsf{den}(\mathcal{Z}) \geq \left(\prod_{i=1}^{k'-k} \frac{k'-i-1}{k'-i+1} \right) \mathsf{den}(\mathcal{Z}_1).$$

We can now compute the product using factorials. For the numerator we have

$$(k'-2) \cdot (k'-3) \cdot \ldots \cdot (k'-l-1) = \frac{k'!}{k' \cdot (k'-1) \cdot [(k'-l-2)!]}.$$

And for the denominator

$$k' \cdot (k'-1) \cdot \ldots \cdot (k'-l+1) = \frac{k'!}{(k'-l)!}.$$

Hence,

$$\prod_{i=1}^{l} \frac{k'-i-1}{k'-i+1} = \frac{(k'-l)!}{k' \cdot (k'-1) \cdot [(k'-l-2)]!} = \frac{(k'-l-1)(k'-l)}{k'(k'-1)}.$$

Hence, for $l = k' - k$ we get

$$\mathsf{den}(\mathcal{Z}) = \frac{k \cdot (k-1)}{k' \cdot (k'-1)} \geq \frac{k \cdot (k-1)}{2k \cdot (2k-1)} = \frac{1}{4} - \frac{1}{8k-2}$$

which establishes the result. □

Hence, if we had an algorithm that obtains an approximate solution $\{\hat{\mathcal{X}}, \hat{\mathcal{Y}}\}$ such that

$$\mathbf{1}_{\hat{\mathcal{X}}}^T \mathbf{A} \mathbf{1}_{\hat{\mathcal{Y}}} \geq \frac{\mathbf{1}_{\mathcal{X}_*}^T \mathbf{A} \mathbf{1}_{\hat{\mathcal{Y}}_*}}{\rho}.$$

Then, we could use prune to obtain a k-subgraph $\hat{\mathcal{Z}}$, such that

$$\mathsf{den}(\hat{\mathcal{Z}}) \geq \frac{1}{4} \frac{\mathbf{1}_{\hat{\mathcal{X}}}^T \mathbf{A} \mathbf{1}_{\hat{\mathcal{Y}}}}{k} \geq \frac{1}{4 \cdot \rho} \frac{\mathbf{1}_{\mathcal{X}_*}^T \mathbf{A} \mathbf{1}_{\hat{\mathcal{Y}}_*}}{k} \geq \frac{1}{4 \cdot \rho} \frac{\mathbf{1}_{\mathcal{S}_*}^T \mathbf{A} \mathbf{1}_{\hat{\mathcal{S}}_*}}{k} \geq \frac{1}{4 \cdot \rho} \mathsf{opt}.$$

3. Approximation Guarantees

3.1. Proof of Theorem 1

The first important technical proposition that we show, is that we can solve DBkS for any constant rank *rectangular* matrix $\mathbf{B}$ of dimensions $n_1 \times n_2$.

Proposition 6. *Let* $\mathbf{B}$ *be any matrix of size* $n_1 \times n_2$ *and let*

$$\mathbf{B}_d = \sum_{i=1}^{d} \sigma_i \mathbf{v}_i \mathbf{u}_i^T$$

be its singular value decomposition, where $\mathbf{v}_i$ *and* $\mathbf{u}_i$ *is the left and right singular vectors corresponding to the* i*th largest singular value* $\sigma_i(\mathbf{B})$*. Then, we can solve the following problem*

$$\{\mathcal{X}_d, \mathcal{Y}_d\} = \arg \max_{|\mathcal{X}|=k_1, |\mathcal{Y}|=k_2} \mathbf{1}_{\mathcal{X}}^T \mathbf{B}_d \mathbf{1}_{\mathcal{Y}}.$$

in time $O(\min\{n_1, n_2\}^{d+1})$.

Proof. For simplicity we assume that the left singular vectors are scaled by their singular values, hence

$$\mathbf{B}_d = \mathbf{v}_1 \mathbf{u}_1^T + \ldots + \mathbf{v}_d \mathbf{u}_d^T.$$

Let us without loss of generality assume that $n_1 \leq n_2$.

We wish to solve:

$$\max_{|\mathcal{X}|=k_1, |\mathcal{Y}|=k_2} \mathbf{1}_{\mathcal{X}}^T \left(\mathbf{v}_1 \mathbf{u}_1^T + \ldots + \mathbf{v}_d \mathbf{u}_d^T \right) \mathbf{1}_{\mathcal{Y}}. \tag{8}$$

Observe that we can rewrite (8) in the following way

$$\max_{|\mathcal{X}|=k_1, |\mathcal{Y}|=k_2} \mathbf{1}_{\mathcal{X}}^T \left[\mathbf{v}_1 \cdot \underbrace{(\mathbf{u}_1^T \mathbf{1}_{\mathcal{Y}})}_{c_1} + \ldots + \mathbf{v}_d \cdot \underbrace{(\mathbf{u}_d^T \mathbf{1}_{\mathcal{Y}})}_{c_d} \right] = \max_{|\mathcal{Y}|=k_2} \left(\max_{|\mathcal{X}|=k_1} \mathbf{1}_{\mathcal{X}}^T \mathbf{v}_{\mathcal{Y}} \right),$$

where $\mathbf{v}_{\mathcal{Y}} = \mathbf{v}_1 \cdot c_1 + \ldots + \mathbf{v}_d \cdot c_d$ is an n_1-dimensional vector generated by the d-dimensional subspace spanned by $\mathbf{v}_1, \ldots, \mathbf{v}_d$.

We will now make a key observation: for every fixed vector $\mathbf{v}_{\mathcal{Y}}$, the index set $\mathcal{X}$ that maximizes $\mathbf{1}_{\mathcal{X}}^T \mathbf{v}_{\mathcal{Y}}$ can be easily computed. It is not hard to see that for any fixed vector $\mathbf{v}_{\mathcal{Y}}$, the k_1-subset $\mathcal{X}$ that maximizes

$$\mathbf{1}_{\mathcal{X}}^T \mathbf{v}_{\mathcal{Y}} = \sum_{i \in \mathcal{X}} [\mathbf{v}_{\mathcal{Y}}]_i$$

corresponds to either the set of k_1 largest or k_1 smallest signed coordinates of $\mathbf{v}_{\mathcal{Y}}$. That is, the locally optimal sets are either $\text{top}_{k_1}(\mathbf{v}_{\mathcal{Y}})$ or $\text{top}_{k_1}(-\mathbf{v}_{\mathcal{Y}})$.

We now wish to find all possible locally optimal sets $\mathcal{X}$. If we could possibly check all vectors $\mathbf{v}_{\mathcal{Y}}$, then we could find all locally optimal index sets $\text{top}_{k_1}(\pm \mathbf{v}_{\mathcal{Y}})$.

Let us denote as $\mathcal{S}_d$ the set of all k_1-sized sets $\mathcal{X}$ that are the optimal solutions of the inner maximization of in the above, for *any* vector $\mathbf{v}$ in the span of $\mathbf{v}_1, \ldots, \mathbf{v}_d$

$$\mathcal{S}_d = \{\text{top}_{k_1}(\pm [\mathbf{v}_1 \cdot c_1 + \ldots + \mathbf{v}_d \cdot c_d]) : c_1, \ldots, c_d \in \mathbb{R}\}.$$

Clearly, this set contains all possible locally optimal $\mathcal{X}$ sets of the form $\text{top}_{k_1}(\mathbf{v}_{\mathcal{Y}})$. Therefore, we can rewrite DBkS on $\mathbf{B}_d$ as

$$\max_{|\mathcal{Y}|=k_2} \max_{\mathcal{X} \in \mathcal{S}_d} \mathbf{1}_{\mathcal{X}}^T \mathbf{B}_d \mathbf{1}_{\mathcal{Y}}. \tag{9}$$

The above problem can now be solved in the following way: for every set $\mathcal{X} \in \mathcal{S}_d$ find the locally optimal set $\mathcal{Y}$ that maximizes $\mathbf{1}_{\mathcal{X}}^T \mathbf{B}_d \mathbf{1}_{\mathcal{Y}}$. Again, this will either be $\text{top}_{k_2}(-\mathbf{B}_d \mathbf{1}_{\mathcal{X}})$ or $\text{top}_{k_2}(\mathbf{B}_d \mathbf{1}_{\mathcal{X}})$. Then, we simply need to test all such $\mathcal{X}, \mathcal{Y}$ pairs on $\mathbf{B}_d$ and keep the optimizer.

Due to the above, the problem of solving DBkS on the rectangular matrix $\mathbf{B}_d$ is equivalent to constructing the set of k_1-supports $\mathcal{S}_d$, and then finding the optimal solution in that set. How large can $\mathcal{S}_d$ be and can we construct it in polynomial time? As we showed in the first section of the supplemental material this set has size $O(\binom{n_1}{d})$ and can be constructed in time $O(n_1^{d+1})$.

Observe that in the above we could have equivalently solved the problem by finding all the top k_2 sets in the span of $\mathbf{u}_1, \ldots, \mathbf{u}_d$, say that they belong in set $\mathcal{S}'_d$. Then, we could solve the problem by finding for each k_2 sized set $\mathcal{Y} \in \mathcal{S}'_d$ the optimal k_1 sized set $\mathcal{X}$. Both approaches are the same, and the one with the smallest dimension is selected to reduce the computational complexity.

The algorithm that solves the problem for rectangular matrices is given below.

Algorithm 4 low-rank approximations for DBkS

`lowrankDBkS(`k_1`,` k_2`,` d`,` $\mathbf{B}$`)`
$[\mathbf{V}_d, \mathbf{\Sigma}_d, \mathbf{U}_d] = \texttt{SVD}(\mathbf{B}, d)$
$\mathcal{S}_d = \texttt{Spannogram}(k_1, \mathbf{V}_d)$
$\{\mathcal{X}_d, \mathcal{Y}_d\} = \arg\max_{|\mathcal{Y}|=k_2} \max_{\mathcal{X} \in \mathcal{S}_d} \mathbf{1}_{\mathcal{X}}^T \mathbf{V}_d \mathbf{\Sigma}_d \mathbf{U}_d^T \mathbf{1}_{\mathcal{Y}}$
Output: $\{\mathcal{X}_d, \mathcal{Y}_d\}$

`Spannogram(`k_1`,` $\mathbf{V}_d$`)`
$\mathcal{S}_d = \{\text{top}_k(\mathbf{v}) : \mathbf{v} \in \text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_d)\}$
Output: $\mathcal{S}_d$.

□

In our following derivations, for both cases of a rectangular and square symmetric matrices, we consider the same notation of the output solution and output density for simplicity:

$$\{\mathcal{X}_d, \mathcal{Y}_d\} = \arg \max_{|\mathcal{X}|=k, |\mathcal{Y}|=k} \mathbf{1}_{\mathcal{X}}^T \mathbf{A}_d \mathbf{1}_{\mathcal{Y}} \text{ and } \mathsf{opt}_d^{\mathsf{B}} = \frac{\mathbf{1}_{\mathcal{X}_d}^T \mathbf{A} \mathbf{1}_{\mathcal{Y}_d}}{k},$$

$$\{\mathcal{X}_d, \mathcal{Y}_d\} = \arg \max_{k_1,k_2:k_1+k_2=k} \max_{|\mathcal{X}|=k_1, |\mathcal{Y}|=k_2} \mathbf{1}_{\mathcal{X}}^T \mathbf{B}_d \mathbf{1}_{\mathcal{Y}} \text{ and } \mathsf{opt}_d^{\mathsf{B}} = 2\frac{\mathbf{1}_{\mathcal{X}_d}^T \mathbf{B} \mathbf{1}_{\mathcal{Y}_d}}{k}.$$

Moreover, as a reminder the optimal solutions and densities for the problems of interest (DkS, DBkS on $\mathbf{A}$, and DBkS on $\mathbf{B}$) are

$$\mathcal{S}* = \arg \max_{|\mathcal{S}|=k} \mathbf{1}_{\mathcal{S}}^T \mathbf{A} \mathbf{1}_{\mathcal{S}} \text{ and } \mathsf{opt} = \frac{\mathbf{1}_{\mathcal{S}_*}^T \mathbf{A} \mathbf{1}_{\mathcal{S}_*}}{k},$$

$$\{\mathcal{X}_*, \mathcal{Y}_*\} = \arg \max_{|\mathcal{X}|=k, |\mathcal{Y}|=k} \mathbf{1}_{\mathcal{X}}^T \mathbf{A} \mathbf{1}_{\mathcal{Y}} \text{ and } \mathsf{opt}^{\mathsf{B}} = \frac{\mathbf{1}_{\mathcal{X}_*}^T \mathbf{A} \mathbf{1}_{\mathcal{Y}_*}}{k},$$

$$\{\mathcal{X}_*, \mathcal{Y}_*\} = \arg \max_{k_1,k_2:k_1+k_2=k} \max_{|\mathcal{X}|=k_1, |\mathcal{Y}|=k_2} \mathbf{1}_{\mathcal{X}}^T \mathbf{B} \mathbf{1}_{\mathcal{Y}} \text{ and } \mathsf{opt}^{\mathsf{B}} = 2\frac{\mathbf{1}_{\mathcal{X}_*}^T \mathbf{B} \mathbf{1}_{\mathcal{Y}_*}}{k}.$$

We continue with bounding the distance between the optimal solution for DBkS and rank-d optimal solution pair $\{\mathcal{X}_d, \mathcal{Y}_d\}$. We have the following result.

Proposition 7. *For any matrix* $\mathbf{A}$*, we have*

$$\mathsf{opt}_d^{\mathsf{B}} \geq \mathsf{opt}^{\mathsf{B}} - 2 \cdot |\lambda_{d+1}|. \tag{10}$$

Moreover, for any rectangular matrix $\mathbf{B}$*, we have*

$$\mathsf{opt}_d^{\mathsf{B}} \geq \mathsf{opt}^{\mathsf{B}} - 2 \cdot \sigma_{d+1}. \tag{11}$$

Proof. Let $\mathcal{X}_*, \mathcal{Y}_*$ be the optimal solution of DBkS on $\mathbf{A}$ and let $\mathcal{X}_d, \mathcal{Y}_d$ be the optimal solution of DBkS on the rank-d matrix $\mathbf{A}_d$. Then, we have

$$\begin{aligned}\mathsf{opt}_d^{\mathsf{B}} &= \frac{\mathbf{1}_{\mathcal{X}_d}^T \mathbf{A} \mathbf{1}_{\mathcal{Y}_d}}{k} = \frac{\mathbf{1}_{\mathcal{X}_d}^T (\mathbf{A}_d + \mathbf{A} - \mathbf{A}_d) \mathbf{1}_{\mathcal{Y}_d}}{k} = \frac{\mathbf{1}_{\mathcal{X}_d}^T \mathbf{A}_d \mathbf{1}_{\mathcal{Y}_d}}{k} + \frac{\mathbf{1}_{\mathcal{X}_d}^T (\mathbf{A} - \mathbf{A}_d) \mathbf{1}_{\mathcal{Y}_d}}{k} \\ &\geq \frac{\mathbf{1}_{\mathcal{X}_d}^T \mathbf{A}_d \mathbf{1}_{\mathcal{Y}_d}}{k} - \frac{\|\mathbf{1}_{\mathcal{X}_d}\|_2 \cdot \|(\mathbf{A} - \mathbf{A}_d)\mathbf{1}_{\mathcal{Y}_d}\|_2}{k} \geq \frac{\mathbf{1}_{\mathcal{X}_d}^T \mathbf{A}_d \mathbf{1}_{\mathcal{Y}_d}}{k} - |\lambda_{d+1}|, \end{aligned} \tag{12}$$

where the first inequality comes due to Cauchy-Schwarz and the second due to the fact that the norm of the indicator vector is k and the operator norm of $\mathbf{A} - \mathbf{A}_d$ is equal to the $d+1$ largest eigenvalue of $\mathbf{A}$.

Moreover, we have that

$$\begin{aligned}\mathsf{opt}^{\mathsf{B}} &= \frac{\mathbf{1}_{\mathcal{X}_*}^T \mathbf{A} \mathbf{1}_{\mathcal{Y}_*}}{k} = \frac{\mathbf{1}_{\mathcal{X}_*}^T (\mathbf{A}_d + \mathbf{A} - \mathbf{A}_d) \mathbf{1}_{\mathcal{Y}_*}}{k} = \frac{\mathbf{1}_{\mathcal{X}_*}^T \mathbf{A}_d \mathbf{1}_{\mathcal{Y}_*}}{k} + \frac{\mathbf{1}_{\mathcal{X}_*}^T (\mathbf{A} - \mathbf{A}_d) \mathbf{1}_{\mathcal{Y}_*}}{k} \\ &\leq \frac{\mathbf{1}_{\mathcal{X}_d}^T \mathbf{A}_d \mathbf{1}_{\mathcal{Y}_d}}{k} + \frac{\mathbf{1}_{\mathcal{X}_*}^T (\mathbf{A} - \mathbf{A}_d) \mathbf{1}_{\mathcal{Y}_*}}{k} \leq \frac{\mathbf{1}_{\mathcal{X}_d}^T \mathbf{A}_d \mathbf{1}_{\mathcal{Y}_d}}{k} + \frac{\|\mathbf{1}_{\mathcal{X}_*}\|_2 \|(\mathbf{A} - \mathbf{A}_d)\mathbf{1}_{\mathcal{Y}_*}\|_2}{k} \leq \frac{\mathbf{1}_{\mathcal{X}_d}^T \mathbf{A}_d \mathbf{1}_{\mathcal{Y}_d}}{k} + |\lambda_{d+1}| \end{aligned} \tag{13}$$

where the first inequality comes due to the fact that $\mathbf{1}_{\mathcal{X}_d}^T \mathbf{A}_d \mathbf{1}_{\mathcal{Y}_d} \geq \mathbf{1}_{\mathcal{X}_*}^T \mathbf{A}_d \mathbf{1}_{\mathcal{Y}_*}$ and the second and third are similar to the previous bound. We can now combine the above two bound to obtain:

$$\mathsf{opt}_d^{\mathsf{B}} \geq \mathsf{opt}^{\mathsf{B}} - 2 \cdot |\lambda_{d+1}|. \tag{14}$$

In the exact same way, we can obtain the result for rectangular matrices. □

The above proposition, combined with Proposition 1 give us the bipartite part of Theorem 1, where $\mathsf{opt}^{\mathsf{B}} = \mathsf{opt}$, that is

$$\mathsf{opt}_d^{\mathsf{B}} \geq \mathsf{opt}^{\mathsf{B}} - 2 \cdot |\lambda_{d+1}| = \mathsf{opt} - 2 \cdot |\lambda_{d+1}|.$$

To establish the part about graphs with the d largest eigenvalues being positive, we use the following result.

Proposition 8. *If* $\mathbf{A}_d$ *is positive semidefinite, then*

$$\max_{|\mathcal{X}|=k} \frac{\mathbf{1}_{\mathcal{X}}^T \mathbf{A}_d \mathbf{1}_{\mathcal{X}}}{k} = \max_{|\mathcal{X}|=k} \max_{|\mathcal{Y}|=k} \frac{\mathbf{1}_{\mathcal{X}}^T \mathbf{A}_d \mathbf{1}_{\mathcal{Y}}}{k}$$

Proof. This is easy to see by the fact that for any two sets $\mathcal{X}, \mathcal{Y}$ we have

$$\begin{aligned}\max_{|\mathcal{S}|=k} \mathbf{1}_{\mathcal{X}}^T \mathbf{A}_d \mathbf{1}_{\mathcal{Y}} &\leq \max_{|\mathcal{X}|=k, |\mathcal{Y}|=k} \mathbf{1}_{\mathcal{X}}^T \mathbf{A}_d \mathbf{1}_{\mathcal{Y}} = \max_{|\mathcal{X}|=k, |\mathcal{Y}|=k} \mathbf{1}_{\mathcal{X}}^T \mathbf{V}_d \mathbf{\Lambda}_d^{1/2} \mathbf{\Lambda}_d^{1/2} \mathbf{V}_d^T \mathbf{1}_{\mathcal{Y}} \\ &\leq \max_{|\mathcal{X}|=k, |\mathcal{Y}|=k} \max \left\{ \|\mathbf{\Lambda}_d^{1/2} \mathbf{V}_d \mathbf{1}_{\mathcal{X}}\|^2, \ \|\mathbf{\Lambda}_d^{1/2} \mathbf{V}_d \mathbf{1}_{\mathcal{Y}}\| \right\} \leq \max_{|\mathcal{X}|=k, |\mathcal{Y}|=k} \max \left\{ \mathbf{1}_{\mathcal{X}}^T \mathbf{A}_d \mathbf{1}_{\mathcal{X}}, \ \mathbf{1}_{\mathcal{Y}}^T \mathbf{A}_d \mathbf{1}_{\mathcal{Y}} \right\} \\ &\leq \max_{|\mathcal{S}|=k} \mathbf{1}_{\mathcal{S}}^T \mathbf{A}_d \mathbf{1}_{\mathcal{S}} \end{aligned}$$

where the second inequality comes due to the Cauchy-Schwarz inequality. □

We can combine the above proposition with the first part of Proposition 7 to obtain that

$$\mathsf{opt}_d^{\mathsf{B}} \geq \mathsf{opt} - 2|\lambda_{d+1}(\mathbf{A})|$$

when $\mathbf{A}_d$ is positive semidefinite.

In the next proposition, we show how to translate a low-rank approximation of $\mathbf{A}$ after we used the random sampling of `randombipartite`($\mathcal{G}$). We need this result to to establish the general result of Theorem 1, by connecting the previous spectral bound, with the 2 loss in approximation between DBkS and DkS.

Proposition 9. *Let* $\mathbf{A}$ *be the adjacency matrix of a graph. Moreover, let the matrices* $\mathbf{P}_1$ *and* $\mathbf{P}_2$ *be such that* $\mathbf{B} = \mathbf{P}_1\mathbf{A}\mathbf{P}_2$ *is the bi-adjecency created by each loop of* `randombipartite`$(\mathcal{G})$*, where* $\mathbf{P}_1$ *is an* $n_1 \times n$ *matrix indexing the left vertices of the graph, and* $\mathbf{P}_2$ *is an* $n \times n_2$ *sampling matrix that indexes the right vertices of the sub-sampled graph. Then,*

$$\mathsf{opt}_d^{\mathsf{B}} \geq \mathsf{opt}^{\mathsf{B}} - 2|\lambda_{d+1}(\mathbf{A})|,$$

where $\mathsf{opt}^{\mathsf{B}}$ is the maximum density on $\mathbf{B} = \mathbf{P}_1\mathbf{A}\mathbf{P}_2$.

Proof. Let without loss of generality assume that $\mathbf{B}$ will be the bipartite subgraph between the first n_1 and the remaining $n_2 = n - n_1$ vertices, such that

$$\mathbf{A} = \begin{bmatrix} \mathbf{C} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{D} \end{bmatrix} \tag{15}$$

Then there are two sampling matrices that pick the corresponding columns and rows

$$\mathbf{P}_1 = [\mathbf{I}_{n_1 \times n_1} \quad \mathbf{0}_{n_1 \times n_2}] \text{ and } \mathbf{P}_2 = \begin{bmatrix} \mathbf{0}_{n_1 \times n_2} \\ \mathbf{I}_{n_2 \times n_2} \end{bmatrix}$$

Then, instead of working on the matrix that is the rank-d best fit for $\mathbf{B}$, we work on

$$\mathbf{B}_d = \mathbf{P}_1\mathbf{A}_d\mathbf{P}_2.$$

Now, we use the bounding techniques of our previous derivations:

$$\begin{aligned} \mathsf{opt}_d = \frac{\mathbf{1}_{\mathcal{X}_d}^T \mathbf{B} \mathbf{1}_{\mathcal{Y}_d}}{k} = \frac{\mathbf{1}_{\mathcal{X}_d}^T \mathbf{P}_1(\mathbf{A}_d + \mathbf{A} - \mathbf{A}_d)\mathbf{P}_2\mathbf{1}_{\mathcal{Y}_d}}{k} = \frac{\mathbf{1}_{\mathcal{X}_d}^T \mathbf{P}_1\mathbf{A}_d\mathbf{P}_2\mathbf{1}_{\mathcal{Y}_d}}{k} + \frac{\mathbf{1}_{\mathcal{X}_d}^T \mathbf{P}_1(\mathbf{A} - \mathbf{A}_d)\mathbf{P}_2\mathbf{1}_{\mathcal{Y}_d}}{k} \\ \geq \frac{\mathbf{1}_{\mathcal{X}_d}^T \mathbf{P}_1\mathbf{A}_d\mathbf{P}_2\mathbf{1}_{\mathcal{Y}_d}}{k} - \frac{\|\mathbf{1}_{\mathcal{X}_d}\|_2 \cdot \|\mathbf{P}_1(\mathbf{A} - \mathbf{A}_d)\mathbf{P}_2\mathbf{1}_{\mathcal{Y}_d}\|_2}{k} \geq \frac{\mathbf{1}_{\mathcal{X}_d}^T \mathbf{P}_1\mathbf{A}_d\mathbf{P}_2\mathbf{1}_{\mathcal{Y}_d}}{k} - |\lambda_{d+1}|, \end{aligned} \tag{16}$$

where the last step comes due to the fact that $\mathbf{P}_1, \mathbf{P}_2$ their singular values are 1. We can use a similar bound to obtain

$$\mathsf{opt}^{\mathsf{B}} \leq \frac{\mathbf{1}_{\mathcal{X}_d}^T \mathbf{P}_1\mathbf{A}_d\mathbf{P}_2\mathbf{1}_{\mathcal{Y}_d}}{k} + |\lambda_{d+1}|,$$

where $\mathsf{opt}^{\mathsf{B}}$ is the density of the densest k-subgraph on the graph with bi-adjacency matrix $\mathbf{P}_1\mathbf{A}\mathbf{P}_2$. and combine the above to establish the result. □

We can now use our random sampling Proposition 4 and combine that with Propositions 9, and 7, to establish Theorem 1 for arbitrary graphs.

The general algorithm is given below

Algorithm 5 low-rank approximations for DkS

$[\mathbf{V}_d, \mathbf{\Lambda}_d] = \texttt{EVD}(\mathbf{A}, d)$
if $\mathcal{G}$ is bipartite **then**
  $\mathbf{B}$=bi-adjacency of $\mathcal{G}$
  $[\mathbf{V}_d, \mathbf{\Sigma}_d, \mathbf{U}_d] = \texttt{SVD}(\mathbf{B}, d)$
  $\mathbf{B}_d = \mathbf{V}_d \mathbf{\Sigma}_d \mathbf{U}_d^T$
  $\{\mathcal{X}_d, \mathcal{Y}_d\} = \arg\max_{|\mathcal{X}|+|\mathcal{Y}|=k} \mathbf{1}_{\mathcal{X}}^T \mathbf{B}_d \mathbf{1}_{\mathcal{Y}}$.
  $\mathcal{Z}_d = \mathcal{X}_d \cup \mathcal{Y}_d$
else if $\mathcal{G}$ is not bipartite and first d eigenvalues are positive **then**
  $\{\mathcal{X}_d, \mathcal{X}_d\} = \arg\max_{|\mathcal{X}|=k} \max_{|\mathcal{Y}|=k} \mathbf{1}_{\mathcal{X}}^T \mathbf{A}_d \mathbf{1}_{\mathcal{Y}}$.
  $\mathcal{Z}_d = \mathcal{X}_d$
else if $\mathcal{G}$ is not bipartite and first d eigenvalues not all positive **then**
  for $i = 1 : O(\text{poly}(\log n))$ **do**
    draw n fair coins and assign them to vertices
    $\mathcal{L}$ = the set of the vertices with heads.
    $\mathcal{R} = \{1, \ldots, n\} - \mathcal{L}$
    $\mathbf{B}_d^i = [\mathbf{A}_d]_{\mathcal{L},\mathcal{R}}$
    $\{\mathcal{X}^i, \mathcal{Y}^i\} = \arg\max_{|\mathcal{X}|+|\mathcal{Y}|=k} \mathbf{1}_{\mathcal{X}}^T \mathbf{B}_d^i \mathbf{1}_{\mathcal{Y}}$.
  end for
  $\{\mathcal{X}^i, \mathcal{Y}^i\} = \arg\max_{1 \le i \le n} \mathbf{1}_{\mathcal{X}^i}^T \mathbf{B}_d^i \mathbf{1}_{\mathcal{Y}^i}$
  $\mathcal{Z}_d = \mathcal{X}_d \cup \mathcal{Y}_d$
end if
Output: $\mathcal{Z}_d$

3.2. Proof of Theorem 2

We now proceed with with a simple bound that we get for the d-th largest eigenvalue of any adjacency matrix. This bound in combination with Theorem 1 will establish Theorem 2.

Proposition 10. *For any un-weighted adjacency matrix* $\mathbf{A}$*, we have that*

$$\lambda_d \le \sqrt{\frac{2E}{d}}.$$

Moreover, for any un-weighted bi-adjacency matrix $\mathbf{B}$*, we have that*

$$\sigma_d \le \sqrt{\frac{E}{d}}.$$

Proof. First, observe that First, observe that

$$d\lambda_d^2 \le \sum_{i=1}^{d} \lambda_i^2 \le \sum_{i=1}^{n} \lambda_i^2 = \|\mathbf{A}\|_F^2.$$

Observe moreover that

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i,j} A_{i,j}^2} = \sqrt{\sum_{i,j} A_{i,j}} = \sqrt{2 \cdot E},$$

where the second equality comes due to the fact that $A_{i,j}^2$ can only be 1 or 0. We now combine the above inequalities to obtain $\lambda_d \le \sqrt{\frac{2E}{d}}$. For the rectangular case, we similarly have

$$d\sigma_d^2 \le \sum_{i=1}^{d} \sigma_i^2 \le \sum_{i=1}^{\min n_1, n_2} \lambda_i^2 = \|\mathbf{B}\|_F^2. \tag{17}$$

Observe moreover that

$$\|\mathbf{B}\|_F = \sqrt{\sum_{i,j} B_{i,j}^2} = \sqrt{\sum_{i,j} B_{i,j}} = \sqrt{E}.$$

Hence, $\sigma_d \leq \sqrt{\frac{E}{d}}$. □

We can now establish our $1+\epsilon$ result for bipartite matrices.

Claim 1. *If the densest-k-subgraph of $\mathcal{G}$ contains a constant fraction of all the edges and $\mathcal{G}$ is bipartite, and $k = \Theta(\sqrt{E})$, then we can approximate DkS within a factor of $1+\epsilon$, in time $n^{O(1/\epsilon^2)}$.*

Proof. Let

$$\mathbf{A} = \begin{bmatrix} \mathbf{0}_{n_1 \times n_2} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{0}_{n_2 \times n_1} \end{bmatrix},$$

where $n_1 + n_2 = n$, and $\mathbf{B}$ is the bi-adjacency matrix of the bipartite graph $\mathcal{G}$. Then, as we showed in Section 7

$$\mathsf{opt} = \max_{|\mathcal{S}|=k} \mathbf{1}_{\mathcal{S}}^T \mathbf{A} \mathbf{1}_S = 2 \max_{k_1,k_2:k=k_1+k_2} \max_{|\mathcal{X}|=k_1} \max_{|\mathcal{Y}|=k_2} \mathbf{1}_{\mathcal{X}}^T \mathbf{B} \mathbf{1}_{\mathcal{Y}}.$$

We approximate the above optimization, by solving the one on the rank-d matrix $\mathbf{B}_d$:

$$\max_{k_1,k_2:k=k_1+k_2} \max_{|\mathcal{X}|=k_1} \max_{|\mathcal{Y}|=k_2} \mathbf{1}_{\mathcal{X}}^T \mathbf{B}_d \mathbf{1}_{\mathcal{Y}}.$$

and among all $O(k)$ pairs for all cases $k_1 + k_2 = k$, we output the one that is best. For each pair $k_1 \times k_2$, we solve

$$\max_{|\mathcal{X}|=k_1} \max_{|\mathcal{Y}|=k_2} \mathbf{1}_{\mathcal{X}}^T \mathbf{B}_d \mathbf{1}_{\mathcal{Y}}.$$

Let $\{\mathcal{X}_d, \mathcal{Y}_d\}$ denote the solution, among all pairs of sparsity levels k_1, k_2.

Then, due to Proposition 10, we have that

$$\mathbf{1}_{\mathcal{X}_d}^T \mathbf{B} \mathbf{1}_{\mathcal{Y}_d} \geq \mathbf{1}_{\mathcal{X}_*}^T \mathbf{B} \mathbf{1}_{\mathcal{Y}_*} - 2 \cdot \sigma_{d+1}.$$

We can now combine the spectral bound of Proposition 10 $\sigma_i \leq \sqrt{\frac{E}{i}}$ with our assumption that the optimal k-subgraph contains a number of edges equal to $k \cdot \mathsf{opt} = c \cdot E$, for some constant $c > 0$. Since $\sigma_d \leq \sqrt{\frac{E}{d}}$ and $k = c\sqrt{E}$ we have

$$2\mathsf{opt}_d = 2\mathbf{1}_{\mathcal{X}_d}^T \mathbf{B} \mathbf{1}_{\mathcal{Y}_d} \geq 2(\mathbf{1}_{\mathcal{X}_*}^T \mathbf{B} \mathbf{1}_{\mathcal{Y}_*} - 2 \cdot \sigma_{d+1}) \tag{18}$$

$$\geq \mathsf{opt} - 4 \cdot \sigma_{d+1} \geq c' \cdot \sqrt{E} - 4\sqrt{\frac{E}{d}} = \left(1 - 4\sqrt{\frac{2}{c^2 d}}\right) \mathsf{opt} \tag{19}$$

Hence, if we want $4\sqrt{\frac{2}{c^2 d}} = \epsilon$, we need

$$d = \frac{16}{c^2 \epsilon^2} = O(1/\epsilon^2),$$

which establishes the result. □

Now, it is simple to establish our $2+\epsilon$ result on general graphs that contain a dense subgraph.

Claim 2. *If the densest-k-subgraph of $\mathcal{G}$ contains a constant fraction of all the edges and $\mathcal{G}$, and $k = \Theta(\sqrt{E})$, then we can approximate DkS within a factor of $2+\epsilon$, in time $n^{O(1/\epsilon^2)}$.*

Proof. A simple combination of Proposition 10 and Proposition 4, give us our result. We first convert $\mathcal{G}$ into a bipartite with our random sampling Algorithm of Proposition 4. Then, we solve the bipartite problem on each of the sampled graphs. The solution to the bipartite will be within $1+\epsilon$, since the bipartite graph will contain a constant fraction of the edges. Then converting it to a solution for the original graph, we lose an extra factor of at most $2+\delta$, as we showed in Proposition 4. □

3.3. Proof of Lemma 3: the data dependent bound

Finally, we prove our data dependent bounds, that bound the performance of opt, relative to polynomially computable quantities.

Claim 3. *The optimal density of can be bounded as*

$$\mathsf{opt} \leq \min\left\{\frac{2+o_n(1)}{k}\cdot \mathbf{1}_{\mathcal{Z}_d}^T\mathbf{A}_d\mathbf{1}_{\mathcal{Z}_d} - |\lambda_{d+1}|,\ k-1,\ \lambda_1\right\}.$$

Additionally, if the graph is bipartite, or if the largest d eigenvalues of the graph are positive, then

$$\mathsf{opt} \leq \min\left\{\frac{1}{k}\cdot \mathbf{1}_{\mathcal{Z}_d}^T\mathbf{A}_d\mathbf{1}_{\mathcal{Z}_d} - |\lambda_{d+1}|,\ \ k-1,\ \ \lambda_1\right\}.$$

Proof. Let $\mathcal{X}_*, \mathcal{Y}_*$ be the optimal solution of DBkS on $\mathbf{A}$ and let Let $\mathcal{X}_d, \mathcal{Y}_d$ be the optimal solution of DBkS on the rank-d matrix $\mathbf{A}_d$. Then, for the first bound we have

$$\begin{aligned}\mathsf{opt}^{\mathsf{B}} &= \frac{\mathbf{1}_{\mathcal{X}_*}^T\mathbf{A}\mathbf{1}_{\mathcal{Y}_*}}{k} = \frac{\mathbf{1}_{\mathcal{X}_*}^T(\mathbf{A}_d+\mathbf{A}-\mathbf{A}_d)\mathbf{1}_{\mathcal{Y}_*}}{k} = \frac{\mathbf{1}_{\mathcal{X}_*}^T\mathbf{A}_d\mathbf{1}_{\mathcal{Y}_*}}{k} + \frac{\mathbf{1}_{\mathcal{X}_*}^T(\mathbf{A}-\mathbf{A}_d)\mathbf{1}_{\mathcal{Y}_*}}{k}\\ &\leq \frac{\mathbf{1}_{\mathcal{X}_d}^T\mathbf{A}_d\mathbf{1}_{\mathcal{Y}_d}}{k} + \frac{\mathbf{1}_{\mathcal{X}_*}^T(\mathbf{A}-\mathbf{A}_d)\mathbf{1}_{\mathcal{Y}_*}}{k} \leq \frac{\mathbf{1}_{\mathcal{X}_d}^T\mathbf{A}_d\mathbf{1}_{\mathcal{Y}_d}}{k} + \frac{\|\mathbf{1}_{\mathcal{X}_*}\|_2\,\|(\mathbf{A}-\mathbf{A}_d)\mathbf{1}_{\mathcal{Y}_*}\|_2}{k} \leq \frac{\mathbf{1}_{\mathcal{X}_d}^T\mathbf{A}_d\mathbf{1}_{\mathcal{Y}_d}}{k} + |\lambda_{d+1}|. \quad (20)\end{aligned}$$

Applying this bound for all cases of graphs, as described by Theorem 1, and using the same bounds in the proofs of Proposition 7-9, we obtain the result. The upper bound $k-1$ is trivial by the fact that for any $\mathcal{X}$ and $\mathcal{Y}$ we have

$$\frac{\mathbf{1}_{\mathcal{X}}^T\mathbf{A}\mathbf{1}_{\mathcal{Y}}}{k} \leq \frac{\mathbf{1}_{\mathcal{X}}^T(\mathbf{1}\mathbf{1}^T-\mathbf{I}_n)\mathbf{1}_{\mathcal{Y}}}{k} \leq \frac{k(k-1)}{k} = k-1.$$

The last bound is simply due to the spectral bound on the bilinear form $\frac{\mathbf{1}_{\mathcal{X}}^T\mathbf{A}\mathbf{1}_{\mathcal{Y}}}{k} \leq \frac{\|\mathbf{1}_{\mathcal{X}}\|_2\cdot\|\mathbf{A}\mathbf{1}_{\mathcal{Y}}\|_2}{k} \leq \lambda_1$. If we combine now the above with the PSD part of Theorem 1, and the results of Proposition 4, we obtain the claim. □

4. Proof of Theorem 3: Nearly-linear Time Solver for Positive Semidefinite Matrices

When the matrix $\mathbf{A}_d$ has mixed signs of eigenvalues, then we have to go through the route of DBkS. However, when $\mathbf{A}_d$ has only positive eigenvalues, then it is easy to show that solving the bilinear problem on $\mathbf{A}_d$ is equivalent to solving

$$\max_{|\mathcal{X}|=k}\frac{\mathbf{1}_{\mathcal{X}}^T\mathbf{A}_d\mathbf{1}_{\mathcal{X}}}{k}.$$

This is the DkS low-rank problem, that now can be solved by our algorithm. We show that when this spectral scenario holds, we can speed up computations tremendously, by the use of a simple randomization.

Let us first remind the fact that DBkS and DkS are equivalent for positive semidefinite matrices. We will show here how $\max_{|\mathcal{S}|=k}\mathbf{1}_{\mathcal{S}}^T\mathbf{A}_d\mathbf{1}_{\mathcal{S}}$ can be approximately in time nearly-linear in n, by only introducing a small relative approximation error. Our approximation will use ϵ-nets.

Definition 1. *(ϵ-net) Let $\mathbb{S}^d = \{\mathbf{c}\in\mathbb{R}^d : \|\mathbf{c}\|^2 = 1\}$ be the surface of the d-dimensional sphere. An ϵ-net of $\mathbb{S}^d$ is a finite set $\mathcal{N}_\epsilon^d \subset \mathbb{S}^d$ such that*

$$\forall \mathbf{c}\in\mathbb{S}^d\ \ \exists\ \widehat{\mathbf{c}}\in\mathcal{N}_\epsilon^d : \|\mathbf{c}-\widehat{\mathbf{c}}\|_2 \leq \epsilon.$$

We now show that we can solve our optimization, via the use of ϵ-nets.

Proposition 11. *Let $\mathcal{N}_\epsilon^d$ be an ϵ-net of $\mathbb{S}^d$. Then,*

$$(1-\epsilon)^2\cdot\max_{|\mathcal{S}|=k}\mathbf{1}_{\mathcal{S}}^T\mathbf{A}_d\mathbf{1}_{\mathcal{S}} \leq \max_{\mathbf{c}\in\mathcal{N}_\epsilon^d}\max_{|\mathcal{S}|=k}\left(\mathbf{c}^T\mathbf{\Lambda}_d^{1/2}\mathbf{V}_d^T\mathbf{1}_{\mathcal{S}}\right)^2 \leq \max_{|\mathcal{S}|=k}\mathbf{1}_{\mathcal{S}}^T\mathbf{A}_d\mathbf{1}_{\mathcal{S}}. \quad (21)$$

Proof. Let $\mathbf{c}$ be a $d \times 1$ unit length vector, *i.e.*, $\|\mathbf{c}\|_2 = 1$. Then, by the Cauchy-Schwartz inequality we have

$$(\mathbf{c}^T \mathbf{\Lambda}_d^{1/2} \mathbf{V}_d^T \mathbf{1}_{\mathcal{S}})^2 \leq \|\mathbf{\Lambda}_d^{1/2} \mathbf{V}_d^T \mathbf{1}_{\mathcal{S}}\|_2^2.$$

We can get equality in the previous bound for a unit norm $\mathbf{c}$ co-linear to $\mathbf{\Lambda}_d^{1/2} \mathbf{V}_d^T \mathbf{1}_{\mathcal{S}}$. Therefore, we have

$$\|\mathbf{\Lambda}_d^{1/2} \mathbf{V}_d^T \mathbf{1}_{\mathcal{S}}\|_2^2 = \max_{\|\mathbf{c}\|_2=1} (\mathbf{c}^T \mathbf{\Lambda}_d^{1/2} \mathbf{V}_d^T \mathbf{1}_{\mathcal{S}})^2. \tag{22}$$

Hence,

$$\max_{|\mathcal{S}|=k} \mathbf{1}_{\mathcal{S}}^T \mathbf{A}_d \mathbf{1}_{\mathcal{S}} = \max_{|\mathcal{S}|=k} \|\mathbf{\Lambda}_d^{1/2} \mathbf{V}_d^T \mathbf{1}_{\mathcal{S}}\|_2^2 = \max_{|\mathcal{S}|=k} \max_{\|\mathbf{c}\|_2=1} (\mathbf{c}^T \mathbf{\Lambda}_d^{1/2} \mathbf{V}_d^T \mathbf{1}_{\mathcal{S}})^2. \tag{23}$$

We can now obtain the upper bound of the proposition, since $\mathcal{N}_\epsilon^d \subseteq \mathbb{S}^d$. Now for the lower bound, let $(\mathbf{1}_{\mathcal{S}_d}, \mathbf{c}_d)$ denote the optimal solution of the above maximization, such that

$$\max_{|\mathcal{S}|=k} \mathbf{1}_{\mathcal{S}}^T \mathbf{A}_d \mathbf{1}_{\mathcal{S}} = (\mathbf{c}_d \mathbf{V}_d{}^T \mathbf{1}_{\mathcal{S}_d})^2.$$

Then, there exists a vector $\widehat{\mathbf{c}}$ in the ϵ-net $\mathcal{N}_\epsilon^d$, such that $\mathbf{c}_d = \widehat{\mathbf{c}} + \mathbf{r}$, with $\|\mathbf{r}\| \leq \epsilon$. Then,

$$\sqrt{\max_{|\mathcal{S}|=k} \mathbf{1}_{\mathcal{S}}^T \mathbf{A}_d \mathbf{1}_{\mathcal{S}}} = \mathbf{c}_d{}^T \mathbf{V}_d^T \mathbf{1}_{\mathcal{S}_d} = (\widehat{\mathbf{c}} + \mathbf{r})^T \mathbf{V}_d^T \mathbf{1}_{\mathcal{S}_d} \overset{(\alpha)}{\leq} \widehat{\mathbf{c}}^T \mathbf{V}_d^T \mathbf{1}_{\mathcal{S}_d} + \epsilon \cdot \|\mathbf{V}_d^T \mathbf{1}_{\mathcal{S}_d}\| = \widehat{\mathbf{c}}^T \mathbf{V}_d^T \mathbf{1}_{\mathcal{S}_d} + \epsilon \cdot \sqrt{\max_{|\mathcal{S}|=k} \mathbf{1}_{\mathcal{S}}^T \mathbf{A}_d \mathbf{1}_{\mathcal{S}}},$$

where (α) is due to the triangle inequality, then the Cauchy-Schwartz inequality, and then the fact that $\|\mathbf{r}\| \leq \epsilon$. From the above inequality we get

$$(1-\epsilon)^2 \max_{|\mathcal{S}|=k} \mathbf{1}_{\mathcal{S}}^T \mathbf{A}_d \mathbf{1}_{\mathcal{S}} \leq \left(\widehat{\mathbf{c}}^T \mathbf{V}_d^T \mathbf{1}_{\mathcal{S}_d}\right)^2 \leq \max_{\mathbf{c} \in \mathcal{N}_\epsilon^d} \max_{|\mathcal{S}|=k} \left(\mathbf{c}^T \mathbf{V}_d^T \mathbf{1}_{\mathcal{S}}\right)^2$$

which concludes the proof. □

The importance of the above proposition is because for a fixed $\mathbf{c}$ we can easily solve the problem

$$\max_{|\mathcal{S}|=k} \left(\mathbf{c}^T \mathbf{V}_d^T \mathbf{1}_{\mathcal{S}}\right)^2.$$

Observe that the above optimization is the same problem that we had to solve for a fixed $\mathcal{Y}$ in Section 3. This inner product is maximized when $\mathcal{S}$ picks the largest, or smallest k elements of the n-dimensional vector $\mathbf{c}^T \mathbf{V}_d^T$. The complexity to do that is linear $O(n)$ (Cormen et al., 2001).

It is now obvious that the number of elements, and the complexity to construct $\mathcal{N}_\epsilon^d$ is important. In (Wyner, 1967), it is shown how to randomly create an ϵ-net on the sphere, by randomly drawing m vectors on the sphere, where

$$m_\epsilon = \epsilon^{-d} \cdot e^{o(d)}.$$

This approach gives with probability $1 - o_d(1)$ an ϵ-net. Let $\mathcal{M}_\epsilon$ be a set of m_ϵ vectors drawn uniformly on the sphere. Our algorithm operates as follows: First we draw if we draw a set of $\mathcal{M}_\epsilon$ random vectors, and then we find the corresponding optimal $\mathcal{S}$, by solving

$$\max_{\mathbf{c} \in \mathcal{M}_\epsilon} \max_{|\mathcal{S}|=k} \left(\mathbf{c}^T \mathbf{\Lambda}_d^{1/2} \mathbf{V}_d^T \mathbf{1}_{\mathcal{S}}\right)^2.$$

This can be done in time $O(m_\epsilon n)$. Then, we draw the set $\mathcal{M}_\epsilon$ a total of $\log n$ times, and iterate this process of finding the best $\mathcal{S}$ for each of these sets $\mathcal{M}_\epsilon$. Then, among all these solutions, with probability $1 - c/n$, for come $c > 0$, the best solution satisfies the bound of the proposition. The randomized algorithm is given below for completeness.

Algorithm 6 Randomized Spannogram

`Spannogram_approx`$(k,\ \mathbf{V}_d, \mathbf{\Lambda}_d)$
$\mathcal{S}_d = \emptyset$
for $i = 1 : m_\epsilon \cdot \log n$ **do**
  $\mathbf{v} = (\mathbf{\Lambda}_d^{1/2} \cdot \mathbf{V}_d)^T \cdot \texttt{randn}(d, 1)$
  $\mathcal{S}_d = \mathcal{S}_d \cup \text{top}_k(\mathbf{v}) \cup \text{top}_k(-\mathbf{v})$
end for
Output: $\arg\max_{\mathcal{S}\in\mathcal{S}_d} \left\|\mathbf{\Lambda}_d^{1/2}\mathbf{V}_d^T\mathbf{1}_\mathcal{S}\right\|$.

Computing $\mathbf{A}_d$ in the first step of the algorithm, can also be done in nearly linear-time in the size of the input $\mathbf{A}$. There is extensive literature on approximating $\mathbf{A}_d$ by a rank-d matrix $\widehat{\mathbf{A}}_d$ such that $\|\mathbf{A}_d - \widehat{\mathbf{A}}_d\|_2 \leq \delta$. in time proportional to the nonzero entries of $\mathbf{A}$ times a poly-logarithmic term, as long as λ_d/λ_{d+1} is at least a constant (Rokhlin et al., 2009; Halko et al., 2011; Gittens et al., 2013).

By further setting $\epsilon = \frac{1}{\log n}$, we obtain Theorem 3.

5. Vertex Sparsification via Simple Leverage Score Sampling

Our algorithm comes together with a vertex elimination step: after we compute the low-rank approximation matrix $\mathbf{A}_d$, we discard rows and columns of $\mathbf{A}_d$, i.e., vertices, depending on their *weighted leverage scores*. Leverage score sampling has been extensively studied in the literature, for many different applications, where it can provably provide small error bounds, while keeping a small number of features from the original matrix (Mahoney & Drineas, 2009; Boutsidis et al., 2009).

As we see in the following, this pre-processing step comes with an error guarantee. We show that by throwing away vertices with small leverage scores, can only introduce a provably small error.

Let us define as

$$\ell_i = \left\| \left[\mathbf{V}_d|\mathbf{\Lambda}|^{1/2}\right]_{i,:} \right\| = \sqrt{\sum_{j=1}^{d} [\mathbf{V}_d]_{i,j}^2 |\lambda_j|},$$

the weighted leverage score of a vertex i. Then, our elimination step is simple. Let $\hat{\mathbf{A}}_d$ be a subset of $\mathbf{A}_d$, where we have eliminated all vertices with $\ell_i \leq \frac{\eta}{3k\ell_1}$. Let $\mathbf{P}_\mathcal{H}$ be a diagonal matrix of 1s and 0s, with a 1 only in the (i,i) indices such that $\ell_i > \frac{\eta}{3k\ell_1}$. Then,

$$\hat{\mathbf{A}}_d = \mathbf{P}_\mathcal{H}\mathbf{A}_d\mathbf{P}_\mathcal{H}.$$

We can now guarantee the following upper bound on the error introduced by the elimination.

Proposition 12. *Let $\hat{\mathbf{A}}_d$ be created as above,*

$$\hat{\mathbf{A}}_d = \mathbf{P}_\mathcal{H}\mathbf{A}_d\mathbf{P}_\mathcal{H}.$$

where $\mathbf{P}_\mathcal{H}$ is a diagonal matrix of 1s and 0s, with a 1 on (i,i) indices such that $\ell_i > \frac{\eta}{3k\ell_1}$. Then,

$$\frac{\mathbf{1}_\mathcal{X}^T\mathbf{A}_d\mathbf{1}_\mathcal{Y}}{k} - \eta \leq \frac{\mathbf{1}_\mathcal{X}^T\hat{\mathbf{A}}_d\mathbf{1}_\mathcal{Y}}{k} \leq \frac{\mathbf{1}_\mathcal{X}^T\mathbf{A}_d\mathbf{1}_\mathcal{Y}}{k} + \eta, \tag{24}$$

for all subsets of k vertices $\mathcal{X}, \mathcal{Y}$.

Proof. Let for brevity $\theta = \frac{\eta}{3k\ell_1}$. Moreover, Let $\mathbf{P}_\mathcal{H}$ be a diagonal matrix of 1s and 0s, with a 1 on (i,i) indices such that $\ell_i > \frac{\eta}{3k\ell_1}$, and let $\mathbf{P}_\mathcal{L}$ be a diagonal matrix of 1s and 0s, with a 1 on (i,i) indices such that $\ell_i \leq \frac{\eta}{3k\ell_1}$. Clearly,

$$\mathbf{P}_\mathcal{H} + \mathbf{P}_\mathcal{L} = \mathbf{I}_{n\times n}.$$

Then, we can rewrite $\mathbf{A}_d$ as:

$$\mathbf{A}_d = (\mathbf{P}_\mathcal{H} + \mathbf{P}_\mathcal{L})\mathbf{A}_d(\mathbf{P}_\mathcal{H} + \mathbf{P}_\mathcal{L}) = \mathbf{P}_\mathcal{H}\mathbf{A}_d\mathbf{P}_\mathcal{H} + \mathbf{P}_\mathcal{L}\mathbf{A}_d\mathbf{P}_\mathcal{L} + \mathbf{P}_\mathcal{H}\mathbf{A}_d\mathbf{P}_\mathcal{L} + \mathbf{P}_\mathcal{L}\mathbf{A}_d\mathbf{P}_\mathcal{H}.$$

Then, we have the following

$$|\mathbf{1}_{\mathcal{X}}^T(\mathbf{A}_d - \hat{\mathbf{A}}_d)\mathbf{1}_{\mathcal{Y}}| = |\mathbf{1}_{\mathcal{X}}^T(\mathbf{P}_{\mathcal{L}}\mathbf{A}_d\mathbf{P}_{\mathcal{L}} + \mathbf{P}_{\mathcal{H}}\mathbf{A}_d\mathbf{P}_{\mathcal{L}} + \mathbf{P}_{\mathcal{L}}\mathbf{A}_d\mathbf{P}_{\mathcal{H}})\mathbf{1}_{\mathcal{Y}}| \tag{25}$$

$$\leq |\mathbf{1}_{\mathcal{X}}^T\mathbf{P}_{\mathcal{L}}\mathbf{A}_d\mathbf{P}_{\mathcal{L}}\mathbf{1}_{\mathcal{Y}}| + |\mathbf{1}_{\mathcal{X}}^T\mathbf{P}_{\mathcal{H}}\mathbf{A}_d\mathbf{P}_{\mathcal{L}}\mathbf{1}_{\mathcal{Y}}| + 2|\mathbf{1}_{\mathcal{X}}^T\mathbf{P}_{\mathcal{L}}\mathbf{A}_d\mathbf{P}_{\mathcal{H}}y| \tag{26}$$

$$\tag{27}$$

Observe that for the first error term we have

$$|\mathbf{1}_{\mathcal{X}}^T\mathbf{P}_{\mathcal{L}}\mathbf{A}_d\mathbf{P}_{\mathcal{L}}\mathbf{1}_{\mathcal{Y}}| = |\mathbf{1}_{\mathcal{X}}^T\mathbf{P}_{\mathcal{L}}\mathbf{V}_d\mathbf{\Lambda}_d^{1/2}\mathbf{S}\mathbf{\Lambda}_d^{1/2}\mathbf{V}_d^T\mathbf{P}_{\mathcal{L}}\mathbf{1}_{\mathcal{Y}}| \tag{28}$$

$$\leq \|\mathbf{1}_{\mathcal{X}}^T\mathbf{P}_{\mathcal{L}}\mathbf{V}_d\mathbf{\Lambda}_d^{1/2}\| \cdot \|\mathbf{S}\mathbf{\Lambda}_d^{1/2}\mathbf{V}_d^T\mathbf{P}_{\mathcal{L}}\mathbf{1}_{\mathcal{Y}}\| \leq \|\mathbf{1}_{\mathcal{X}}^T\mathbf{P}_{\mathcal{L}}\mathbf{V}_d\mathbf{\Lambda}_d^{1/2}\| \cdot \|\mathbf{S}\| \cdot \|\mathbf{\Lambda}_d^{1/2}\mathbf{V}_d^T\mathbf{P}_{\mathcal{L}}\mathbf{1}_{\mathcal{Y}}\| \tag{29}$$

$$= \|\mathbf{1}_{\mathcal{X}}^T\mathbf{P}_{\mathcal{L}}\mathbf{V}_d\mathbf{\Lambda}_d^{1/2}\| \cdot 1 \cdot \|\mathbf{\Lambda}_d^{1/2}\mathbf{V}_d^T\mathbf{P}_{\mathcal{L}}\mathbf{1}_{\mathcal{Y}}\| \leq k \cdot \theta \cdot k \cdot \theta = k^2 \cdot \theta^2. \tag{30}$$

where the second and third inequalities come due to the Cauchy-Schwarz inequality. In the above $\mathbf{S}$ denotes the diagonal matrix that contains the signs of the eigenvalues. Clearly, its operator norm is 1. Hence, the last inequality in the above is due to the fact that $\mathbf{1}_{\mathcal{X}}, \mathbf{1}_{\mathcal{Y}}$ have k entries with 1, and each picks the rows of $\mathbf{V}_d\mathbf{\Lambda}_d^{1/2}$ with the highest leverage score. Then, due to the triangle inequality on the k-largest row norms (i.e., leverage scores) of $\mathbf{V}_d\mathbf{\Lambda}_d^{1/2}$ we get the final result.

Similarly, we can bound the remaining two error terms

$$|\mathbf{1}_{\mathcal{X}}^T\mathbf{P}_{\mathcal{L}}\mathbf{A}_d\mathbf{P}_{\mathcal{H}}\mathbf{1}_{\mathcal{Y}}| = |\mathbf{1}_{\mathcal{X}}^T\mathbf{P}_{\mathcal{L}}\mathbf{V}_d\mathbf{\Lambda}_d^{1/2}\mathbf{S}\mathbf{\Lambda}_d^{1/2}\mathbf{V}_d^T\mathbf{P}_{\mathcal{H}}\mathbf{1}_{\mathcal{Y}}| \tag{31}$$

$$\leq \|\mathbf{1}_{\mathcal{X}}^T\mathbf{P}_{\mathcal{L}}\mathbf{V}_d\mathbf{\Lambda}_d^{1/2}\| \cdot \|\mathbf{S}\mathbf{\Lambda}_d^{1/2}\mathbf{V}_d^T\mathbf{P}_{\mathcal{H}}\mathbf{1}_{\mathcal{Y}}\| \leq \|\mathbf{1}_{\mathcal{X}}^T\mathbf{P}_{\mathcal{L}}\mathbf{V}_d\mathbf{\Lambda}_d^{1/2}\| \cdot \|\mathbf{S}\| \cdot \|\mathbf{\Lambda}_d^{1/2}\mathbf{V}_d^T\mathbf{P}_{\mathcal{H}}\mathbf{1}_{\mathcal{Y}}\| \tag{32}$$

$$= \|\mathbf{1}_{\mathcal{X}}^T\mathbf{P}_{\mathcal{L}}\mathbf{V}_d\mathbf{\Lambda}_d^{1/2}\| \cdot 1 \cdot \|\mathbf{\Lambda}_d^{1/2}\mathbf{V}_d^T\mathbf{P}_{\mathcal{H}}\mathbf{1}_{\mathcal{Y}}\| \leq k \cdot \theta \cdot k \cdot \ell_1 = k^2 \cdot \theta \cdot \ell_1. \tag{33}$$

Since, $\eta \leq \ell_1$, we conclude that the above error can be bounded as

$$\frac{|\mathbf{1}_{\mathcal{X}}^T(\mathbf{A}_d - \hat{\mathbf{A}}_d)\mathbf{1}_{\mathcal{Y}}|}{k} \leq 3 \cdot k \cdot \theta \cdot \ell_1 = \eta.$$

Hence, we obtain the proposition. □

6. NP-hardness of DkS on rank-1 matrices

In this section, we establish the hardness of the quadratic formulation of DkS, even for rank-1 matrices. Interestingly the problem is not hard when we relax it to its bilinear form as we showed in our main result. The claim follows.

Claim 4. *DkS is NP-hard for rank-1 matrices* $\mathbf{A}$ *with one negative eigenvalue.*

Proof. Observe that a rank-1 matrix with 1 negative eigenvalue can be written as

$$\mathbf{A} = -\mathbf{v}\mathbf{v}$$

where $\lambda_1 = \|\mathbf{v}\|^2$. Then, see that

$$\max_{|\mathcal{S}|=k} \mathbf{1}_{\mathcal{S}}^T\mathbf{A}\mathbf{1}_{\mathcal{S}} = \max_{|\mathcal{S}|=k} -\mathbf{1}_{\mathcal{S}}^T\mathbf{v}\mathbf{v}^T\mathbf{1}_{\mathcal{S}} = \min_{|\mathcal{S}|=k} \mathbf{1}_{\mathcal{S}}^T\mathbf{v}\mathbf{v}^T\mathbf{1}_{\mathcal{S}} = \left(\min_{|\mathcal{S}|=k} |\mathbf{1}_{\mathcal{S}}^T\mathbf{v}|\right)^2 = \left(\min_{|\mathcal{S}|=k} \left|\sum_{i\in S} v_i\right|\right)^2.$$

An algorithm that can solve the above problem, can be used to solve SubsetSum. In SubsetSum we are given a set of integers and we wish to decide whether there exists a non-empty subset of these integers that sums to zero. In the following algorithm we show how this can be trivially done, by solving $\min_{|\mathcal{S}|=k} \left|\sum_{i\in S} v_i\right|$ for all values of k. If for some value of k the sum in the optimizaton is zero, then we decide YES as the output for the SubsetSum. Hence, solving $\max_{|\mathcal{S}|=k} \mathbf{1}_{\mathcal{S}}^T\mathbf{A}\mathbf{1}_{\mathcal{S}}$ is NP-hard even for rank-1 matrices, in the general case. □

7. Additional Experiments

In Fig. 1, we show additional experiment on 9 more large-graphs. The description of the graphs can be found in Table 1. Moreover, in Fig. 2. The description of the experiments can be found in the figure captions.

Algorithm 7 SubsetSum via rank-1 DkS

Input: $\mathbf{v} = [v_1, \ldots, v_n]$
for $i = 1 : k$ **do**
  $s_i = \min_{|\mathcal{S}|=k} \left|\sum_{i \in S} v_i\right|$
  if $s_i == 0$ **then**
    Output: YES
  else
    Output: NO
  end if
end for

Table 1. Datasets used in our experiments

Data set	Nodes	Edges	Description
com-DBLP	317,080	1,049,866	DBLP collaboration network
com-LiveJournal	3,997,962	34,681,189	LiveJournal online social network
web-NotreDame	325,729	1,497,134	Web graph of Notre Dame
ego-Facebook	4,039	88,234	Social circles from Facebook (anonymized)
ca-AstroPh	18,772	396,160	Collaboration network of Arxiv Astro Physics
ca-HepPh	12,008	237,010	Collaboration network of Arxiv High Energy Physics
ca-CondMat	23,133	186,936	Collaboration network of Arxiv Condensed Matter
ca-GrQc	5,242	28,980	Collaboration network of Arxiv General Relativity
ca-HepTh	9,877	51,971	Collaboration network of Arxiv High Energy Physics Theory
loc-Brightkite	58,228	214,078	Brightkite location based online social network
roadNet-CA	1,965,206	5,533,214	Road network of California
email-Enron	36,692	367,662	Email communication network from Enron
com-Orkut	3,072,441	117,185,083	Orkut online social network
flickr	105,938	2,316,948	Network of Flickr images sharing Common Metadata
FBMedium	63,731	817,090	Friendship Data of Facebook Users

References

Boutsidis, Christos, Mahoney, Michael W, and Drineas, Petros. An improved approximation algorithm for the column subset selection problem. In *Proceedings of the twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 968–977. Society for Industrial and Applied Mathematics, 2009.

Cormen, Thomas H, Leiserson, Charles E, Rivest, Ronald L, and Stein, Clifford. *Introduction to algorithms*. MIT press, 2001.

Gittens, Alex, Kambadur, Prabhanjan, and Boutsidis, Christos. Approximate spectral clustering via randomized sketching. *arXiv preprint arXiv:1311.2854*, 2013.

Halko, Nathan, Martinsson, Per-Gunnar, and Tropp, Joel A. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM review*, 53(2):217–288, 2011.

Mahoney, Michael W and Drineas, Petros. Cur matrix decompositions for improved data analysis. *Proceedings of the National Academy of Sciences*, 106(3):697–702, 2009.

Papailiopoulos, Dimitris S, Dimakis, Alexandros G, and Korokythakis, Stavros. Sparse pca through low-rank approximations. *arXiv preprint arXiv:1303.0551*, 2013.

Rokhlin, Vladimir, Szlam, Arthur, and Tygert, Mark. A randomized algorithm for principal component analysis. *SIAM Journal on Matrix Analysis and Applications*, 31(3):1100–1124, 2009.

Wyner, Aaron D. Random packings and coverings of the unit n-sphere. *Bell System Technical Journal*, 46(9): 2111–2118, 1967.

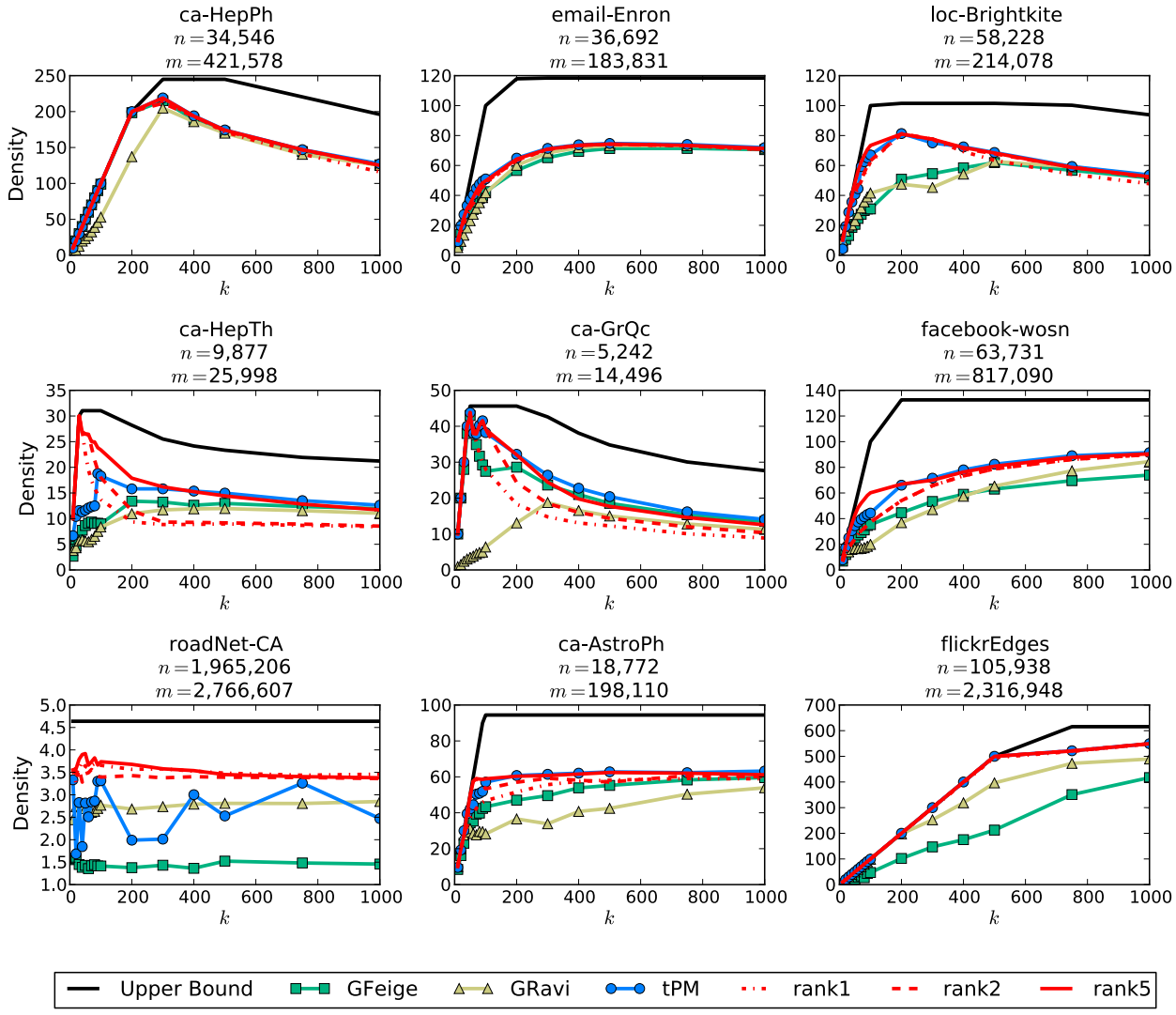


Figure 1. Subgraph density vs. subgraph size (k). We show the comparison of densest subgraph algorithms on several additional datasets: Academic collaboration graphs from Arxiv (ca-HepTh, ca-HepPh, ca-GrQc, Ca-Astro), Geographic location-based networks (roadNet, loc-Brightkite), The Enron email communication graph (email-Enron) and a facebook subgraph (facebook-wosn). The number of vertices and edges are shown in each plot. As can be seen, in almost all cases rank-2 and rank-5 spannograms match or outperform previous algorithms. One notable exception is the ca-GrQc where, for subgraphs of size above $k = 400$ or above, T-power performs better. Another observation is that the spannogram benefits are often more significant for smaller subgraph sizes. It can also be seen that the tightness of our data-dependent bound (solid black line) varies for different data sets and subgraph sizes.

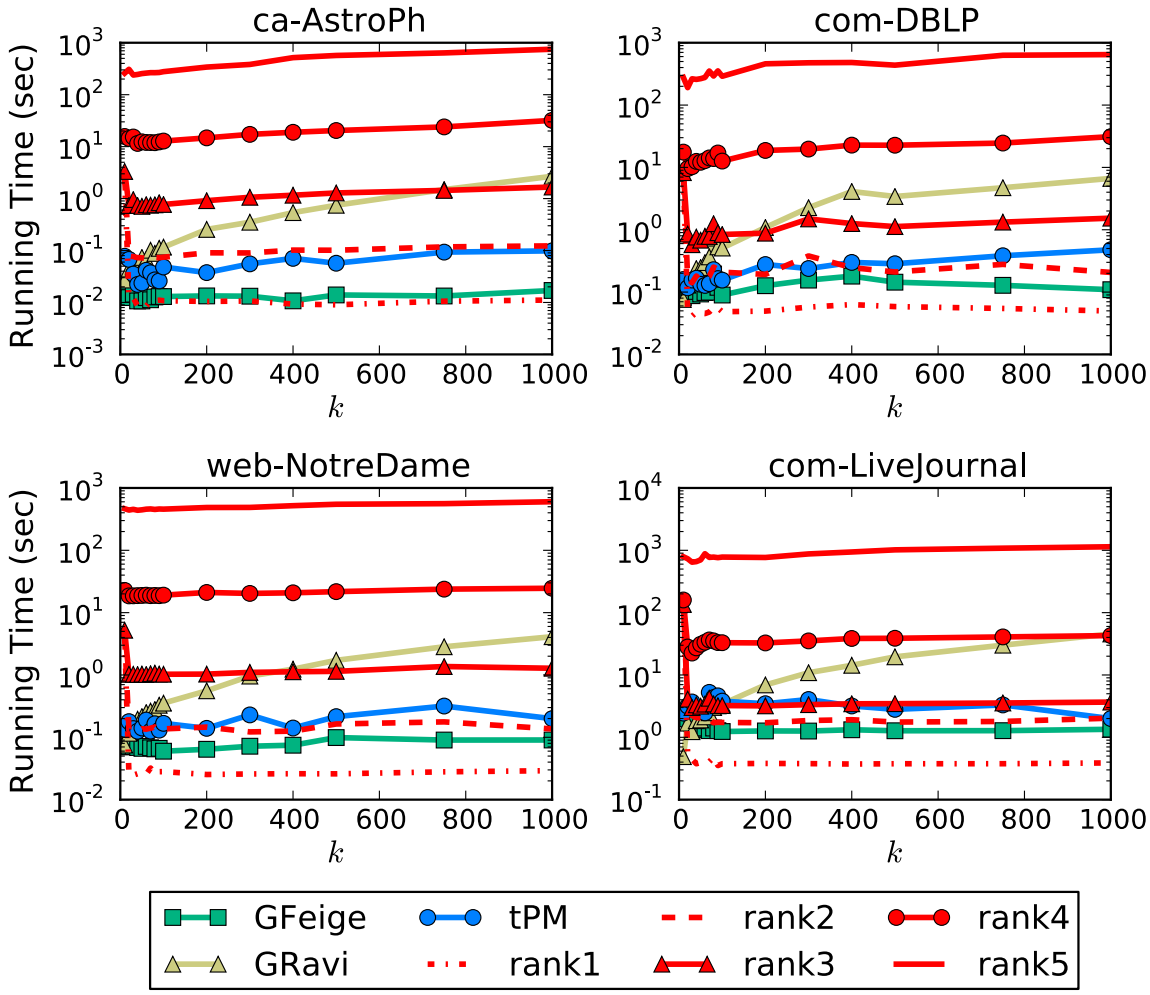


Figure 2. Running times on a MacBook Pro 10.2, with Intel Core i5 @ 2.5 GHz (2 cores), 256 KB L2 Cache per core, 3 MB L3 Cache, and 8 GB RAM. Experiments were run on MATLAB R2011b (7.13.0.564). As can be seen, Rank-1 is significantly faster than all other algorithms for all tested cases. Rank-2 is comparable to prior work, having running times of a few seconds. Rank-5 was the highest accuracy setting we tested. It can take several minutes on large graphs and seems useful only when high accuracy is desired or other methods are far from the upper bound.