Semirandom Planted Bipartite Subgraphs

Anand Louis
Indian Institute of Science, Bengaluru, India

Kirtan Vora
Indian Institute of Science, Bengaluru, India

Abstract

There have been many recent works studying planted subgraphs problems. The *semirandom planted bipartite subgraph problem* is defined as follows. Starting with a vertex set V, an arbitrary subset $S \subset V$ of size k is chosen, then an arbitrary bipartite graph is added on S. After this between each pair of vertices in $S \times (V \setminus S)$ an edge is added independently with probability p, then an arbitrary graph is added on $V \setminus S$. The analogous semirandom planted clique problem, where S forms a clique, has been studied starting with the work of Feige and Kilian [6]; recent work by [3, 11] gave an algorithm for this problem when $k = \Omega(\sqrt{n \log n})$. We give an algorithm for semirandom planted bipartite subgraph problem when $k = \Omega(\sqrt{n \log n})$ and the two color classes are roughly balanced.

Our algorithms are essentially the same as the elegant greedy algorithm of [3]. We generalize their idea to our setting. Handling the arbitrary nature of the bipartite graph requires some new technical ideas and is our main technical contribution.

2012 ACM Subject Classification Theory of computation → Graph algorithms analysis; Theory of computation → Approximation algorithms analysis

Keywords and phrases Semirandom Models, Spectral Algorithms, Planted Subgraphs, Random Graphs, Approximate Recovery Algorithms

Digital Object Identifier 10.4230/LIPIcs.SWAT.2026.32

Funding *Anand Louis*: Supported in part by SERB Award CRG/2023/002896 and the Walmart Center for Tech Excellence at IISc (CSR Grant WMGT-23-0001).
Kirtan Vora: Supported by Prime Minister's Research Fellowship, India.

1 Introduction

One of the well known NP-Hard problem in computer science is the problem of finding the maximum independent set (MIS) in a graph. Consider a graph $G = (V, E)$ then unless $P = NP$, [12] showed that for every $\epsilon > 0$ it is hard to approximate MIS within $n^{1-\epsilon}$. Another such problem which is known to be NP-Hard [18] is to find the largest induced bipartite subgraph of the graph $G = (V, E)$.

A series of works on semirandom models have partially explained the tractability of finding the maximum clique problem. An instance of the planted clique problem is sampled as follows: we start with the set of n vertices from which we choose an arbitrary subset S of k vertices and add a clique on them, for any pair of vertices, both of them simultaneously, not in S we add an edge independently with probability p (e.g. $p = 1/2$). [2] designed a spectral algorithm to find the planted clique when $k = \Omega(\sqrt{n})$ but the spectral algorithm is not robust to the significant change in the edges outside of the clique. To study the robustness of clique finding algorithms, Feige and Kilian [6] introduced an insightful and influential semirandom model. In one version of this model, the edges between S and $V \setminus S$ (where V is the set of vertices), also known as cut edges or $cut(S)$, are added independently with probability p, and the edges within $V \setminus S$ can be configured adversarially with access to the random edges between S and $V \setminus S$. In this model S cannot be uniquely recovered since $V \setminus S$ can have $(n - k)/k$ cliques of size k whose cut edges are drawn independently with probability p.

© Anand Louis and Kirtan Vora;
licensed under Creative Commons License CC-BY 4.0

The problem of finding a small list (of size $\frac{n}{k}(1+o(1))$) of cliques of size k in which the planted clique is present has been studied in [6, 5, 16, 4, 3, 11]. [3] gave an algorithm for this problem that works with probability at least $1-1/n$ over the input when $k=\Omega(\sqrt{n}\log^2 n)$. [11] gave an algorithm for this problem that works with probability 0.99 over the input when $k=\Omega(\sqrt{n\log n})$. [17] proved that it is information-theoretically impossible to recover the planted clique when $k=o(\sqrt{n})$

[15] considered the problem of finding an arbitrary planted r-colorable k size subgraph S of G on n vertices such that the edges with both endpoints contained in $V\setminus S$ are sampled independently with probability p, and the edges between S and $V\setminus S$ can be configured adversarially with access to random edges contained in $V\setminus S$. This work designed a polynomial time algorithm to find a subgraph T of G of size at most $(1+\delta)k$ whose intersection with S is at least $(1-\delta)k$ whenever $k=\Omega_{p,\delta}(r\sqrt{n/p})$. Their proof follows by showing that for a novel SDP relaxation of the problem, for which most of the mass is supported on S. Their results also hold in the presence of monotone adversary, i.e., adversary who can arbitrary add edges in $V\setminus S$.

[9] considers the model in which an arbitrary bipartite subgraph on S is planted and the graph on $V\setminus S$ is arbitrary, and then the edges between S and $V\setminus S$ are sampled independently at random with probability p. Note that the model considered by [9] is different from the model proposed by [6] since in this model, adversary doesn't have access to the random edges between S and $V\setminus S$. [9] gave an algorithm for the problem when S is a bipartite graph of size $(1-\epsilon)n$ and recovered a bipartite graph of size $(1-O(\epsilon/p^2))n$ whenever $\epsilon=\Omega(\log n/n)$.

In this paper inspired from [6], we consider the model when the cut edges between the 2-colorable graph and the rest of graph are sampled uniformly and independently at random with probability 1/2. The recent progress on the random planted r-colorable graph problem naturally motivates us to study the Semirandom Planted Bipartite subgraph (SRPB) problem and Semirandom Planted Tripartite subraph (SRPT) problem mentioned below.

▶ **Definition 1** (Semirandom Planted Bipartite Graph, SRPB(n,k,p,t)). *(See figure 1) To an empty graph G with vertex set $V=[n]$:*

1. *Plant a bipartite graph $G(S_1,S_2,E)$: Plant a bipartite graph on a random subset of vertices $S\subseteq V$ with $|S|=k, S=S_1\cup S_2$, where S_1 and S_2 are two independent disjoint components of S in which $1\le |S_2|/|S_1|\le t$.*
2. *Include cut edges at random: Add each edge to $cut(S)$ independently with probability p.*
3. *Choose the rest of the edges adversarially: Adaptively choose any induced graph on $V\setminus S$.*

Our main results are following:

▶ **Theorem 2.** *There exists a polynomial time algorithm such that for every $\epsilon>0$ and $t\in[1,3)$, given an instance of SRPB$(n,k,1/2,t)$, the algorithm outputs a n^4 size list of graphs which contains a subgraph W of G such that $|S\setminus W|=O(\epsilon k)$ and $|W\setminus S|=o(k)$ whenever $k\ge\Omega\left(\frac{1}{\sqrt{\epsilon^{1/2}(3-t)^4}}n^{3/4}\log n\right)$ with probability $1-1/n$ over the instance of SRPB$(n,k,1/2,t)$.*

▶ **Theorem 3.** *There exists a polynomial time algorithm such that for every $\epsilon>0$ and $t\in[1,3)$, given an instance of SRPB$(n,k,1/2,t)$, the algorithm outputs a n^{12} size list of graphs which contains a subgraph W of G such that $|S\setminus W|=O(\epsilon k)$ and $|W\setminus S|=o(k)$ whenever $k\ge\Omega\left(\frac{1}{\epsilon^{1/2}(3-t)^2}\sqrt{n\log n}\right)$ with probability 0.99 over then instance of SRPB$(n,k,1/2,t)$.*

We also study the 3-colorable version of the problem which is added here.

▶ **Definition 4** (Semirandom Planted Tripartite Subgraph, SRPT(n,k,p)). *To an empty graph G with vertex set $V = [n]$:*

1. *Plant a tripartite graph $G(S_1, S_2, S_3, E)$: Plant a tripartite graph on a random subset of vertices $S \subseteq V$ with $|S| = k, S = S_1 \cup S_2 \cup S_3$, where S_1, S_2 and S_3 are three independent disjoint components of S in which $|S_2| = |S_1| = |S_3| = k/3$.*
2. *Include cut edges at random: Add each edge to $cut(S)$ independently with probability p.*
3. *Choose the rest of the edges adversarially: Adaptively choose any induced graph on $V \setminus S$.*

▶ **Theorem 5.** *There exists a polynomial time algorithm such that for every $\epsilon > 0$, given an instance of $SRPT(n, k, 1/2)$, the algorithm outputs a n^{18} size list of graphs that contains a subgraph W of G such that $|S \setminus W| = O(\epsilon k)$ and $|W \setminus S| = o(k)$ whenever $k \geq \Omega\left(\frac{1}{\epsilon^{1/2}}\sqrt{n \log n}\right)$ with probability 0.99 over the instance of $SRPT(n, k, 1/2)$.*

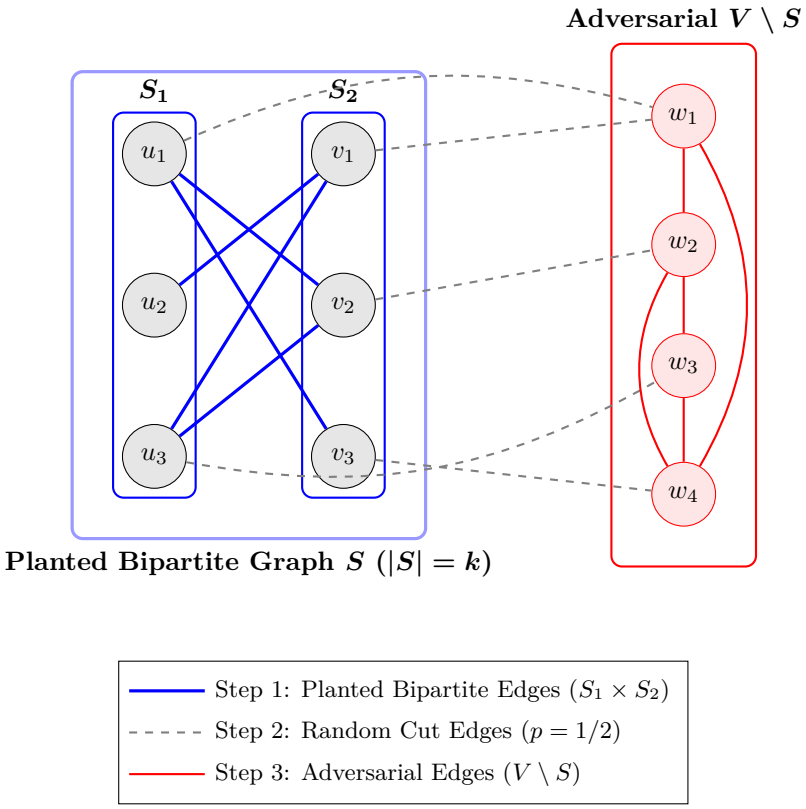

Figure 1 Structure of the Semirandom Planted Bipartite Graph, SRPB($n, k, 1/2, 1$).

1.1 Other Related Work

Random and Semirandom Models. [13] proposed a model in which a d-regular bipartite graph is planted on a set S of size k in G and rest of the edges are added independently with probability p, they designed an algorithm which recovers the planted bipartite subgraph with high probability when $k = \Omega_p(\sqrt{n \log n})$. Their algorithm also works for monotone adversary i.e. their algorithm can also recover the planted bipartite subgraph even if arbitrary edges are added in $G[V \setminus S]$ with addition to the edges that are already randomly generated. They show that the natural SDP formulation of this problem is integral. [14] studied this problem without monotone adversary and gave an simple spectral algorithm to recover when $k = \Omega(\sqrt{n})$.

[9] also study a natural extension of their semirandom model for 2-coloring (defined earlier in this section) to 3-coloring. Assuming that there is an algorithm to color 3-colorable graphs using n^θ colors, they gave an algorithm to compute a set of size $(1-\epsilon f(p,\theta))n$ (for some function f) and an n^θ coloring of it, where $(1-\epsilon)n$ is the size of the planted 3-colorable subgraph.

Vertex Deletion Models. In [1] a polynomial time algorithm was designed for the following problem. Given a graph $G=(V,E)$ where there exists a $(1-\epsilon)|V|$ size subset U of V such that the induced subgraph on U is bipartite, produce a subset U' of V of size at least $(1-O(\epsilon\sqrt{\log n}))|V|$ such that the subgraph induced on U' is bipartite. In [7], for the case where the maximum degree of G is d, they gave an algorithm to produce a subset U' of size at least $(1-O(\sqrt{\epsilon \log d}))|V|$ such that the subgraph induced on U' is bipartite; assuming the Unique Games conjecture, they prove a matching (up to constant factors) hardness. [8] studied the problem when the bipartite graph induced on U is a regular and low threshold-rank graph; they gave an algorithm, running in time exponential in $poly(1/\epsilon)$, to recover a $(1-\epsilon^{1/12})n$ sized subset of $U' \subset V$ such that the graph induced on U' is bipartite. Given a graph $G=(V,E)$ in which there exists a subset U of size $(1-\epsilon)|V|$ such that the graph induced on U is 3-colorable. [9] designed a polynomial time algorithm to produce a subset $S \subset V$ such that $|S| \leq 3\sqrt{\epsilon}|V|$ and color $V \setminus S$ using $\tilde{O}(n^{0.25+O(\epsilon^{0.25})})$. They gave an algorithm to give a more general result.

Edge Deletion Models. In [1], the problem of Min UnCut was studied which is equivalent to the problem of removing a minimum set of edges M from a graph $G=(V,E)$ such that the graph $G'=(V,E\setminus M)$ is bipartite. [1] designed a $O(\sqrt{\log n})$ approximation algorithm based on semi definite programming. [10] gave an algorithm to delete $O(\sqrt{OPT})$ fraction of edges to obtain a bipartite graph where OPT is the optimal fraction of edges in Min UnCut.

2 Proof Overview

We are going to follow the strategy introduced in [3] and will try to extend it to the case of planted bipartite subgraphs.

Let G denote the ± 1 adjacency matrix of the graph $G=(V,E)$ with $V=[n]$, i.e., $G(i,j)=1$ iff $\{i,j\}$ is an edge in G. We also let G_i be the *i-th* row of G, $G_i^{\{j\}} \in \{\pm 1\}^{|S_j|}$ be the projection of G_i to coordinates in the independent components S_1, S_2 of the planted bipartite subgraph S, $G_i^{\text{out}} \in \{\pm 1\}^{n-k}$ be the projection of G_i to the coordinates in $V \setminus S$ and $G_i^S \in \{\pm 1\}^k$ be the projection of G_i to the coordinates in S. N_i denotes the neighbors of vertex i in S. For two sets A and B, $A \Delta B$ denote their symmetric difference. For two vectors $u, v \in \mathbb{R}^m$, $\langle u, v\rangle$ denotes their Euclidean inner product.

Strong correlation between vectors present in same color class of S

We observe that i,j are non-trivially correlated if i,j are both in S_1 (or S_2).

$$\langle G_i, G_j\rangle = \langle G_i^S, G_j^S\rangle + \langle G_i^{\text{out}}, G_j^{\text{out}}\rangle = \langle G_i^{\{1\}}, G_j^{\{1\}}\rangle + \langle G_i^{\{2\}}, G_j^{\{2\}}\rangle + \langle G_i^{\text{out}}, G_j^{\text{out}}\rangle.$$

Since if $i,j \in S_1$, the first term is $|S_1|$ and the second term depends on the number of common neighbors of i and j in S_2 i.e.

$$\left\langle G_i^{\{2\}}, G_j^{\{2\}}\right\rangle = |S_2| - 2|N_i \Delta N_j|.$$

The third term is the sum of $n-k$ $\{\pm 1\}$ terms sampled uniformly and independently at random because every term in $cut(S)$ is chosen independently to be in G. Therefore the third term satisfies $|\langle G_i^{\text{out}}, G_j^{\text{out}}\rangle| \leq O(\sqrt{n\log n})$ with high probability. Thus

$$\langle G_i, G_j\rangle = |S_1| + |S_2| - 2|N_i \Delta N_j| \pm O(\sqrt{n\log n}) = k - 2|N_i\Delta N_j| \pm O(\sqrt{n\log n}).$$

Thus $i,j \in S_1$ will be highly correlated if $k \gg O(\sqrt{n\log n})$ and $|N_i\Delta N_j|$ is smaller than some constant fraction of k. In case of the semirandom planted clique problem [3], we directly get that $\langle G_i, G_j\rangle = k \pm \sqrt{n\log n}$ for $i,j \in S$ because S is a clique. Using the following Lemma 6, which uses the principle of inclusion-exclusion, we will show that many of the pairs $i,j \in S_1$ will be highly correlated. We need to carefully track the arbitrary edges between S_1 and S_2; we do this in Lemma 6 by showing that for three vertices $i,j,w, \in S_1$, all three pairs cannot simultaneously have large symmetric difference of their neighborhood in S.

▶ **Lemma 6.** *For all $i,j,w \in S_1$ (or S_2) if $\langle G_i^S, G_j^S\rangle < \frac{k}{12/(3-t)}$ and $\langle G_i^S, G_w^S\rangle < \frac{k}{12/(3-t)}$, then $\langle G_j^S, G_w^S\rangle \geq \frac{k}{12/(3-t)}$.*

Weak correlation between vectors when $i \in S$ and $j \in V \setminus S$

If $i \in S$ and $j \in V\setminus S$, then $\langle G_i^{\{1\}}, G_j^{\{1\}}\rangle + \langle G_i^{\{2\}}, G_j^{\{2\}}\}\rangle$ will be the sum of k $\{\pm 1\}$ terms sampled uniformly and independently at random whose sum will be $O(\sqrt{k\log k})$ with high probability. Thus i,j will be highly correlated when $\langle G_i^{\text{out}}, G_j^{\text{out}}\rangle$ is sufficiently large. But using Lemma 10 we will show that most of such pairs cannot be highly correlated with high probability; one such easy case to check this is to sample a $\mathcal{G}(n-k,1/2)$ graph on $V\setminus S$ vertices.

In [3] it was shown that for most of the vertices in $S(99\ \%)$ only a small amount $o(k)$ of the vertices of $V\setminus S$ can form a large inner product with a vertex in S, when $k \gg n^{3/4}$. For a vertex $i \in V$, define $V_i = \{j \in V | \langle G_i, G_j\rangle \geq \frac{k}{18/(3-t)}\}$. We show in Lemma 7 using a subtle counting argument that there exist four vertices $i,j,w,l \in S$, using which we recover most of S, i.e. $|(V_i \cup V_j \cup V_w \cup V_l)\Delta S| = O(\epsilon|S|)$.

▶ **Lemma 7.** *For $k = \Omega\left(\sqrt{\frac{24}{\epsilon^{1/2}(3-t)^4}} n^{3/4}\log n\right)$ with probability at least $1-1/n$ over the input of SRPB$(n,k,1/2,t)$ there exist 4 vertices i,j,w,l of S such that $|(V_i \cup V_j \cup V_w \cup V_l)\Delta S| = O(\epsilon k)$.*

To make this work for smaller k, [3] used the Hadamard product of the G_i vectors instead of working with G_i vectors: For $\alpha \subseteq V, |\alpha| = 3$ define G_α as point wise product (Hadamard product) of the vectors G_i, G_j, G_w where $\alpha = \{i,j,w\}$. We note that using the above argument that if $\alpha \subseteq S_1$ and $l \in S$ then

$$\langle G_\alpha, G_l\rangle = k - 2|N_\alpha \Delta N_l| \pm O(\sqrt{n\log n})$$

where N_α corresponds to a neighborhood of a natural imaginary vertex formed by α in S_1 which will be explained later. Similarly if $l \in V\setminus S$ then

$$\langle G_\alpha^{S_1}, G_l^{S_1}\rangle + \langle G_\alpha^{S_2}, G_l^{S_2}\rangle = O(\sqrt{k\log k}).$$

The analogue of the previous two equations for the semirandom planted clique problem was shown in [3]. In [3, 11] it was also shown that for most of the 3 size subset of $S(99\ \%)$, at most $o(k)$ vertices of $V\setminus S$ satisfies $\langle G_\alpha^{\text{out}}, G_l^{\text{out}}\rangle \geq k/3$ they showed this using restricted isometric property (RIP) of random matrices. To recover the planted bipartite subgraph when $k \gg \sqrt{n\log n}$, we use this strategy to prove results analogous to Lemma 6 and Lemma 7 (see Lemma 12).

3 Algorithm and Analysis

3.1 Semirandom Planted Bipartite Subgraph for size $k = \Omega(n^{3/4})$

In this section we prove Theorem 2 using Algorithm 1.

▶ **Theorem 2.** *There exists a polynomial time algorithm such that for every $\epsilon > 0$ and $t \in [1,3)$, given an instance of SRPB$(n,k,1/2,t)$, the algorithm outputs a n^4 size list of graphs which contains a subgraph W of G such that $|S \setminus W| = O(\epsilon k)$ and $|W \setminus S| = o(k)$ whenever $k \geq \Omega\left(\frac{1}{\sqrt{\epsilon^{1/2}(3-t)^4}} n^{3/4} \log n\right)$ with probability $1-1/n$ over the instance of SRPB$(n,k,1/2,t)$.*

Algorithm 1 Greedy algorithm for a n^4 sized list containing the planted bipartite graph.

1: **for** $i = 1$ to n **do**
2: Collect all the vertices j such that $\langle G_i, G_j \rangle \geq \frac{k}{18/(3-t)}$, call this set V_i.
3: **end for**
4: **for** $i,j,w,l = 1$ to n **do**
5: Output the subgraph induced by the vertex set $V_i \cup V_j \cup V_w \cup V_l$.
6: **end for**

First we show that no three vertices $i,j,w \in S_1$ (or S_2) can be simultaneously uncorrelated with each other.

▶ **Lemma 6.** *For all $i,j,w \in S_1$ (or S_2) if $\langle G_i^S, G_j^S \rangle < \frac{k}{12/(3-t)}$ and $\langle G_i^S, G_w^S \rangle < \frac{k}{12/(3-t)}$, then $\langle G_j^S, G_w^S \rangle \geq \frac{k}{12/(3-t)}$.*

Proof. Assume $i,j,w \in S_1$ and $\langle G_i^S, G_j^S \rangle, \langle G_j^S, G_w^S \rangle, \langle G_i^S, G_w^S \rangle < \frac{k}{12/(3-t)}$. Then

$$\langle G_i^S, G_j^S \rangle = k - 2|N_i \Delta N_j| < k/\frac{12}{3-t} \implies |N_i \Delta N_j| > \frac{k}{2}\left(1 - \frac{3-t}{12}\right).$$

Similarly the same will be also true for $|N_i \Delta N_w|$ and $|N_j \Delta N_w|$. By the formula of inclusion-exclusion, we have

$$\begin{aligned} 2|N_i \cup N_j \cup N_w| &= 2|N_i| + 2|N_j| + 2|N_w| - 2|N_i \cap N_j| - 2|N_i \cap N_w| - 2|N_w \cap N_j| \\ &\quad + 2|N_i \cap N_j \cap N_w| \\ &= |N_i \Delta N_j| + |N_i \Delta N_w| + |N_j \Delta N_w| + 2|N_i \cap N_j \cap N_w| \end{aligned}$$

Since $N_i \cup N_j \cup N_w \subseteq S_2$ and $|S_2| \leq tk/(1+t)$, we have

$$\begin{aligned} 2tk/(1+t) \geq 2|N_i \cup N_j \cup N_w| &> \frac{3k}{2}\left(1 - \frac{3-t}{12}\right) + 2|N_i \cap N_j \cap N_w| \\ \implies |N_i \cap N_j \cap N_w| &< -\frac{(3-t)^2}{16(1+t)} < 0 \text{ if } t < 3, \end{aligned}$$

which gives us a contradiction. Similarly if $i,j,k \in S_2$ then $N_i \cup N_j \cup N_w \subset S_1$ and $|S_1| \leq k/2$. Therefore

$$2|N_i \cap N_j \cap N_w| < k - \frac{3k}{2}\left(1 - \frac{3-t}{12}\right) = \frac{-k}{8}(1+t) < 0. \qquad \blacktriangleleft$$

▶ Remark 8. Note that we can prove a much tighter bound if $i,j,w \in S_2$, since $|S_2| \geq |S_1|$. By following the same argument as above we get that for all $i,j,w \in S_2$ if $\langle G_i^S, G_j^S \rangle < \frac{k}{3}$ and $\langle G_i^S, G_w^S \rangle < \frac{k}{3}$, then $\langle G_j^S, G_w^S \rangle \geq \frac{k}{3}$.

Call a pair of vertices $i, j \in S$ good if $\langle G_i^S, G_j^S \rangle \geq \frac{k}{12/(3-t)}$, otherwise call them bad. Lemma 9 shows that all but one vertex of S_1 forms a good pair with some other vertex of S_1.

▶ **Lemma 9.** *At least $|S_1| - 1$ vertices of S_1 forms a good pair with some other vertex in S_1; the same is true for the vertices of S_2.*

Proof. To see this define

$$T = \{i \in S_1 \mid i \text{ forms a good pair with some other vertex in } S_1\}$$

if $S_1 \setminus T$ is nonempty then choose a vertex i from it. Since $i \notin T$ it means i forms bad pairs with all other vertices in S_1. Then by Lemma 6 we can conclude that all the vertices in $S_1 \setminus \{i\}$ forms good pairs in each other, so $|T| \geq |S_1| - 1$. Similar argument can be applied to S_2. ◀

The following lemma is taken directly from [3] which will imply that a vertex from S cannot be highly correlated with many vertices of $V \setminus S$. We reproduce their proof here to make dependence on t explicit since this case doesn't arise in their setting.

▶ **Lemma 10** ([3])**.** *Let $v_1, v_2, v_3, ..., v_k \in \{\pm 1\}^n$ be vectors sampled uniformly and independently at random for $k = \Omega(\sqrt{n \log n})$ Then with probability at least $1 - 1/n$ over the draw of vectors v_i, for every $u \in \{\pm 1\}^n$ there are at most $O\left(\frac{24^2}{(3-t)^2} n^2/k^2\right)$ vectors v_i such that $\langle u, v_i \rangle \geq \frac{k}{24/(3-t)}$.*

Proof. Consider the $n \times k$ matrix H with columns $v_1, v_2, ..., v_k$. Then, by standard results in random matrix theory, $||H||_2 \leq O(\sqrt{n} + \sqrt{k})$ with probability $1 - 1/n$. Let $\mathcal{B} \subset [k]$ be such that, for every $i \in \mathcal{B}, \langle u, v_i \rangle \geq \frac{k}{24/(3-t)}$. Then, by the Cauchy-Schwartz inequality, we have

$$\langle u, H1_{\mathcal{B}} \rangle \leq ||u||_2 ||H1_{\mathcal{B}}||_2 \leq \sqrt{n}\sqrt{|\mathcal{B}|}||H||_2 \leq O(n\sqrt{|\mathcal{B}|}).$$

On the other hand, by the choice of set $\mathcal{B}$, we have

$$\langle u, H1_{\mathcal{B}} \rangle = \sum_{i \in \mathcal{B}} \langle u, v_i \rangle \geq |\mathcal{B}| \frac{k}{24/(3-t)}.$$

Combining those two inequalities and rearranging, we get $|\mathcal{B}| \leq O\left(\frac{24^2}{(3-t)^2} n^2/k^2\right)$. ◀

Greedy Procedure. For $i \in V$, define $V_i = \{j \in V | \langle G_i, G_j \rangle \geq \frac{k}{18/(3-t)}\}$. We show that we can reconstruct a bipartite graph which is close to the planted bipartite subgraph using $i, j, w, l \in S$.

We apply the Lemma 10 with G_i^{out} for each $i \in S$ as a uniform random element of $\{\pm 1\}^{n-k}$ and by taking $u = G_j^{\text{out}}$ for any $j \in V \setminus S$. By Lemma 10, for each $j \in V \setminus S$, at most $O\left(\frac{24^2}{(3-t)^2} n^2/k^2\right)$ elements of i of S will satisfy $\langle G_i^{\text{out}}, G_j^{\text{out}} \rangle \geq \frac{k}{24/(3-t)}$. Using this and the fact that for all $i \in S$ and $j \in V \setminus S$, $\langle G_i^{\text{S}}, G_j^{\text{S}} \rangle = O(\sqrt{k \log k})$ with probability $1 - 1/n$, we get that there are in total $O\left(\frac{24^2}{(3-t)^2} n^3/k^2\right)$ pairs (i, j) such that $i \in S, j \notin S$ satisfying

$$\langle G_i, G_j \rangle \geq \frac{k}{18/(3-t)} \geq \frac{k}{24/(3-t)} + O(\sqrt{k \log k}).$$

By Markov's inequality, a uniformly random $i \in S$ satisfies $\langle G_i, G_j \rangle \geq \frac{k}{18/(3-t)}$ for at most $O\left(\frac{24^2}{\epsilon(3-t)^2} n^3/k^3\right)$ vertices $j \in V \setminus S$ with probability $1 - \epsilon$. If $k \gg O\left(\frac{24^2}{\epsilon(3-t)^2} n^3/k^3\right)$ e.g.

$k = \Omega\left(\sqrt{\frac{24}{\epsilon^{1/2}(3-t)}} n^{3/4}\gamma(n)\right)$ such that $\lim_{n\to\infty}\gamma(n) = \infty$, then at least $k(1-\epsilon)$ vertices i of S will satisfy $|V_i \setminus S| = O\left(\frac{k}{\gamma^4(n)}\right) = o(k)$, which means V_i will contain at most $o(k)$ vertices not from S, call such a set of vertices i of S as S^*. Therefore

$$|S \setminus S^*| \leq \epsilon k. \tag{1}$$

Note that to make the contribution from random terms small we need $\frac{k}{18/3-t} - \frac{k}{24/3-t} > \Omega(\sqrt{k\log k})$. Therefore we also need $k = \Omega(\log k/(3-t)^2)$.

In [3] because S was a clique, the above argument implies that for a vertex $i \in S^*$, V_i contains the whole of S and $o(k)$ vertices from $V \setminus S$. To see this, observe that each vertex $j \in S$ will satisfy $\langle G_i, G_j\rangle = k \pm O(\sqrt{n\log n}) \geq k/6$ and all but $o(k)$ vertices of $V \setminus S$ will not be in V_i because $i \in S^*$. Our goal is to extend this idea to recover bipartite subgraphs. We show that in SRPB$(n, k, 1/2, t)$ we can recover $1 - O(\epsilon)$ fraction of vertices of S using at most 4 vertices of S.

▶ **Lemma 7.** *For $k = \Omega\left(\sqrt{\frac{24}{\epsilon^{1/2}(3-t)^4}} n^{3/4}\log n\right)$ with probability at least $1-1/n$ over the input of SRPB$(n, k, 1/2, t)$ there exist 4 vertices i, j, w, l of S such that $|(V_i \cup V_j \cup V_w \cup V_l)\Delta S| = O(\epsilon k)$.*

Proof. Let $i \in S_1 \cap S^*$. Intuitively Lemma 6 gives us that if V_i doesn't contain all vertices of S_1 i.e. if $S_1 \setminus V_i \neq \phi$, then any two vertices $j, w \in S_1 \setminus V_i$ will satisfy $\langle G_i^S, G_w^S\rangle \geq k/\frac{12}{3-t}$ so $V_j \supset S_1 \setminus V_i$ i.e. $V_i \cup V_j \supset S_1$. But we also need j to be in S^* for $|(V_i \cup V_j) \setminus S| = o(k)$ which brings us to two cases.

Case 1. Assume that $(S_1 \setminus V_i) \setminus (S \setminus S^*) \neq \phi$. Then by Lemma 6, any vertex $j \in (S_1 \setminus V_i) \setminus (S \setminus S^*)$ (i.e. $j \notin V_i$ and $j \in S_1 \cap S^*$) will satisfy $V_j \supset S_1 \setminus V_i$ because for any other vertex $w \in S_1 \setminus V_i$, it must be that $\langle G_i, G_w\rangle = \langle G_i^S, G_w^S\rangle + \langle G_i^{\text{out}}, G_w^{\text{out}}\rangle < \frac{k}{18/(3-t)}$, and since $\langle G_i^{\text{out}}, G_w^{\text{out}}\rangle = O(\sqrt{n\log n}) \ll \frac{k}{36/(3-t)}$ with very high probability, it implies that $\langle G_i^S, G_w^S\rangle < \frac{k}{12/(3-t)}$. This implies that by Lemma 6 $\langle G_j^S, G_w^S\rangle \geq \frac{k}{12/(3-t)}$ which means $\langle G_j, G_w\rangle \geq \frac{k}{18/(3-t)}$ for any $w \in S_1 \setminus V_i$, so $V_i \cup V_j \supset S_1$. Also since $j \in S_1 \cap S^*$ we have $|(V_1 \cup V_j) \setminus S| \leq |V_i \setminus S| + |V_j \setminus S| = o(k)$.

Case 2. $(S_1 \setminus V_i) \setminus (S \setminus S^*) = \phi$, then V_i has almost all the vertices of S_1 because $S_1 \setminus V_i \subset S \setminus S^*$, so $|S_1 \setminus V_i| \leq |S \setminus S^*| = O(\epsilon k)$ and since $i \in S_1 \cap S^*$ we have $|V_1 \setminus S| = o(k)$.

Therefore either by case 1 or case 2 there exists either one or two vertices of S_1 such that $|S_1 \setminus (V_i \cup V_j)| = O(\epsilon k)$ and $|(V_i \cup V_j) \setminus S| = o(k)$. Similarly we can find two vertices $w, l \in S_2$ which satisfies $|S_2 \setminus (V_w \cup V_l)| = O(\epsilon k)$ and $(|V_w \cup V_l) \setminus S| = o(k)$, which shows that there exists at most four vertices such that $|(V_i \cup V_j \cup V_w \cup V_l)\Delta S| = O(\epsilon k)$. ◀

From the above four lemmas we can finish the proof of Theorem 2.

Proof of Theorem 2. Algorithm 1 goes through all the four set of vertices and take their union. Then by Lemma 7 one of them will be close to the planted bipartite subgraph. Since Algorithm 1 enumerates over all choice of i, the size of list is at most n^4. ◀

3.2 Semirandom Planted Bipartite Subgraph for size $k = \Omega(\sqrt{n\log n})$

In this subsection we prove Theorem 3.

▶ **Theorem 3.** *There exists a polynomial time algorithm such that for every $\epsilon > 0$ and $t \in [1,3)$, given an instance of SRPB$(n,k,1/2,t)$, the algorithm outputs a n^{12} size list of graphs which contains a subgraph W of G such that $|S \setminus W| = O(\epsilon k)$ and $|W \setminus S| = o(k)$ whenever $k \geq \Omega\left(\frac{1}{\epsilon^{1/2}(3-t)^2}\sqrt{n \log n}\right)$ with probability 0.99 over then instance of SRPB$(n,k,1/2,t)$.*

Similar to [3, 11], we will show that we can recover S using a constant no. of vertices in S. For $\alpha \subset V$, define $G_\alpha \in \{\pm 1\}^n$ as $G_\alpha(i) = \Pi_{j\in\alpha} G_j(i)$ i.e. G_α is a Hadamard product of all the vectors G_i such that $i \in \alpha$. For $\alpha = \{i,j,w\}$, we will interchange between the notation G_α and G_{ijw}. For $\alpha \subset V$ define $V_\alpha = \left\{ i \in V \mid \langle G_\alpha, G_i \rangle \geq \frac{k}{18/(3-t)} \right\}$.

Algorithm 2 Greedy algorithm for a n^{12} sized list containing the planted bipartite graph.

for $i,j,w = 1$ to n **do**
  Collect all the vertices t such that $\langle G_{ijw}, G_t \rangle \geq \frac{k}{18/(3-t)}$, call that set V_{ijw}
end for
for $\alpha_i, \alpha_j, \alpha_w, \alpha_l = 1$ to $\binom{n}{3}$ **do** ▷ Go through all 3-size subsets of V
  Output the subgraph induced by the vertex set $V_{\alpha_i} \cup V_{\alpha_j} \cup V_{\alpha_w} \cup V_{\alpha_l}$.
end for

A small observation tells us that if $\alpha_1 = \{i,j,l\}$ and $\alpha_2 = \{l,w,t\}$ then

$$\langle G_{\alpha_1}, G_{\alpha_2} \rangle = \langle G_{ijw}, G_t \rangle. \tag{2}$$

The next lemma uses the 1-norm analogue of RIP of matrices. For reference see proposition 2 and section 4 of [11].

▶ **Lemma 11** ([11]). *For every $i \in V \setminus S$, at most $O\left(\frac{24^2}{(3-t)^2} n^2/k^2\right)$ subsets $\alpha \subset S$ of size 3 satisfy $\langle G_\alpha^{\text{out}}, G_i^{\text{out}} \rangle \geq \frac{k}{24/(3-t)}$ with probability 0.99 whenever $k \geq \Omega(\frac{1}{3-t}\sqrt{n \log n})$.*

Thus in total $O\left(\frac{24^2}{(3-t)^2} n^3/k^2\right)$ pairs (α, j) such that $\alpha \in \binom{S}{3}, j \notin S$ satisfying $\langle G_\alpha, G_j \rangle \geq \frac{k}{18/(3-t)} \geq \frac{k}{24/(3-t)} + O(\sqrt{k \log k})$, by Markov's inequality, a uniformly random $\alpha \in \binom{S}{3}$ satisfies $\langle G_\alpha, G_j \rangle \geq \frac{k}{18/(3-t)}$ for at most $O\left(\frac{24^2}{\epsilon^3(3-t)^2} n^3/k^5\right)$ with probability $1-\epsilon^3$ and if $k \gg O\left(\frac{24^2}{\epsilon^3(3-t)^2} n^3/k^5\right)$ and $k = \Omega\left(\frac{1}{3-t}\sqrt{n \log n}\right)$ e.g. $\left(k = \Omega\left(\frac{24^{1/3}}{\epsilon^{1/2}(3-t)}\sqrt{n \log n}\right)\right.$ then for a uniformly chosen $\alpha \in \binom{S}{3}$ will satisfy $|V_\alpha \setminus S| = o(k)$, i.e. $(1-\epsilon^3)$ fractions elements of $\binom{S}{3}$ will form the vertex set V_α which contains at most $o(k)$ vertices not from S, call such a set of $(1-\epsilon^3)\binom{k}{3}$ elements of $\binom{S}{3}$ as T^*. Therefore

$$|T/T^*| \leq O(\epsilon^3 k^3). \tag{3}$$

In [3, 11] because S was a clique, the above argument implies that for a triplet $\alpha \in T^*$, V_α contains whole of S and $o(k)$ vertices from $V \setminus S$. To see this, observe that each vertex $j \in S$ will satisfy $\langle G_\alpha, G_j \rangle = k \pm O(\sqrt{n \log n}) \geq k/6$ and because $\alpha \in T^*$, all but $o(k)$ vertices of $V \setminus S$ will not be in V_α. Our goal is to extend this idea to recover bipartite subgraphs. We show that in SRPB$(n,k,t,1/2)$ we can recover $1 - O(\epsilon)$ fraction of vertices of S using at most 12 vertices of S.

▶ **Lemma 12.** *Whenever $k = \Omega\left(\frac{24^{1/3}}{\epsilon^{1/2}(3-t)^2}\sqrt{n \log n}\right)$, with probability 0.99 there exists $\alpha_1, \alpha_2 \in \binom{S_1}{3} \cap T^*$ such that $|S_1 \setminus (V_{\alpha_1} \cup V_{\alpha_2})| \leq O(\epsilon k)$ and $|(V_{\alpha_1} \cup V_{\alpha_2}) \setminus S| \leq o(k)$, $\alpha_3, \alpha_4 \in \binom{S_2}{3} \cap T^*$ such that $|S_2 \setminus (V_{\alpha_3} \cup V_{\alpha_4})| \leq O(\epsilon k)$ and $|(V_{\alpha_3} \cup V_{\alpha_4}) \setminus S| \leq o(k)$, which implies $|S \Delta (V_{\alpha_1} \cup V_{\alpha_2} \cup V_{\alpha_3} \cup V_{\alpha_4})| \leq O(\epsilon k)$.*

Proof.

Case 1. There exists $\alpha_1, \alpha_2 \in T^* \cap \binom{S_1}{3}$ such that $S_1 \subset V_{\alpha_1} \cup V_{\alpha_2}$, which means that $S_1 \setminus (V_{\alpha_1} \cup V_{\alpha_2}) = \phi$. Since $V_{\alpha_1}, V_{\alpha_2} \in T^* \cap \binom{S_1}{3}$, it implies that $|V_{\alpha_1} \setminus S|, |V_{\alpha_2} \setminus S| = o(k)$. Therefore $|(V_{\alpha_1} \cup V_{\alpha_2}) \setminus S| \leq o(k)$.

Case 2. For all pairs of $\alpha_1, \alpha_2 \in T^* \cap \binom{S_1}{3}$, there exists a vertex $i_{\alpha_1,\alpha_2} \in S_1$ such that $\langle G_{\alpha_1}, G_{i_{\alpha_1,\alpha_2}} \rangle < \frac{k}{18/(3-t)}$ and $\langle G_{\alpha_2}, G_{i_{\alpha_1,\alpha_2}} \rangle < \frac{k}{18/(3-t)}$. Note that this implies that for α_1, $\langle G^S_{\alpha_1}, G^S_{i_{\alpha_1,\alpha_2}} \rangle < \frac{k}{12/(3-t)}$ because $\langle G^{\text{out}}_{\alpha_1}, G^{\text{out}}_{i_{\alpha_1,\alpha_2}} \rangle = O(\sqrt{n \log n}) \ll \frac{k}{18/(3-t)}$, and the same is true for α_2 .

We will now show that $\langle G^S_{\alpha_1}, G^S_{\alpha_2} \rangle \geq \frac{k}{12/(3-t)}$. Observe that $G_{\alpha_1}(i) = -1$ whenever $i \in S_1$, and the same is true for α_2. Consider a new instance of bipartite graph T consisting of color classes W_1, W_2 of size $|S_1|, |S_2|$ respectively such that W_1 consists of three vertices a, b, c whose neighbors in W_2 are determined by the vectors $G^{S_2}_{\alpha_1}, G^{S_2}_{\alpha_2}, G^{S_2}_{i_{\alpha_1,\alpha_2}}$ respectively, i.e. a is connected to a vertex $p \in W_2$ if and only if $G_{\alpha_1}(p) = 1$ and rest of the vertices of W_1 have no neighbors in W_2. Recall that in our notation T_a is the a^{th} row of the signed adjacency matrix of T(similarly for b and c). By the construction of T we get that

$$\langle T_a, T_c \rangle = \langle G^S_{\alpha_1}, G^S_{i_{\alpha_1,\alpha_2}} \rangle < \frac{k}{12/(3-t)}, \ \langle T_b, T_c \rangle = \langle G^S_{\alpha_2}, G^S_{i_{\alpha_1,\alpha_2}} \rangle < \frac{k}{12/(3-t)}, \text{ and}$$
$$\langle T_a, T_b \rangle = \langle G^S_{\alpha_1}, G^S_{\alpha_2} \rangle$$

Then by Lemma 6 $\langle T_a, T_b \rangle \geq \frac{k}{12/(3-t)}$, therefore for all $\alpha_1, \alpha_2 \in T^*$

$$\langle G^S_{\alpha_1}, G^S_{\alpha_2} \rangle \geq \frac{k}{12/(3-t)}. \tag{4}$$

For $\beta \in \binom{S_1}{2}$, let b_β be the number of vertices $i \in S_1$ such that $\beta \cup \{i\} \in \binom{S_1}{3} \setminus T^*$. Then

$$\frac{1}{3} \sum_{\beta \in \binom{S_1}{2}} b_\beta \leq |\binom{S_1}{3} \setminus T^*| \leq O(\epsilon^3 k^3).$$

Then by Markov's inequality, $1 - \epsilon$ fraction elements β of $\binom{S_1}{2}$ will have

$$b_\beta \leq O(\epsilon^2 k). \tag{5}$$

For $i \in S_1$, let b_i be the number of elements of $\binom{S_1}{2}$ such that $\beta \cup i \in \binom{S_1}{3} \setminus T^*$, then by similar argument as above, $1 - \epsilon$ fraction of b_i will satisfy

$$b_i \leq O(\epsilon^2 k^2). \tag{6}$$

Choose an $m \in S_1$ for which $b_m \leq O(\epsilon^2 k^2)$, now for $t \in S_1$ call c_t to be the number of elements t' such that $\{m, t, t'\} \in \binom{S_1}{3} \setminus T^*$. Next we observe that $b_m = \frac{1}{2} \sum_{t \in S_1} c_t$, therefore $\sum_{t \in S_1} c_t \leq O(\epsilon^2 k^2)$. This implies that $1 - \epsilon$ fraction elements of S_1 satisfy

$$c_t \leq O(\epsilon k). \tag{7}$$

Now choose $\beta = \{i, j\} \subset S_1$ such that $\{m\} \cup \beta \in T^*$ and $b_\beta \leq O(\epsilon^2 k)$ (such a β exists because number of pairs $\beta \in \binom{S_1}{2}$ that satisfies $\{m\} \cup \beta \in T^*$ is $\binom{|S_1|-1}{2} - b_m \approx \frac{|S_1|^2}{2}(1 - O(\epsilon^2))$ (Equation 6), and most 2 size subsets β of S_1, i.e. $\binom{|S_1|}{2}(1 - \epsilon)$ elements of S_1 satisfy

$b_\beta \leq O(\epsilon^2 k)$ (Equation 5). Therefore the number of sets $\beta \in \binom{S_1}{2}$ that satisfies $\beta \cup \{m\} \in T^*$ are $\approx \frac{|S_1|^2}{2}(1-\epsilon-O(\epsilon^2)))$. Now, choose a $l \in S_1$ for which $c_l \leq O(\epsilon k)$ and $\{l\} \cup \beta \in T^*$ (such a l exist because at least $(1-\epsilon)|S_1|$ elements $l \in S_1$ satisfy $c_l \leq O(\epsilon k)$ (Equation 7), and $(1-O(\epsilon^2))|S_1|$ elements $l \in S_1$ satisfy $\beta \cup \{l\} \in T^*$ (Equation 5), so number of elements $l \in S_1$ satisfying both of the conditions simultaneously are $(1-\epsilon-O(\epsilon^2))|S_1|)$.

Since $|c_l| \leq O(\epsilon k)$, for at least $(1-O(\epsilon))|S_1|$ elements $x \in S_1$ we have $\{m,l,x\} \in T^* \cap \binom{S_1}{3}$. Moreover $\beta \cup \{l\} = \{i,j,l\} \in T^* \cap \binom{S_1}{3}$. Therefore by Equation 2 and Equation 4

$$\langle G^S_{i,j,m}, G^S_x \rangle = \langle G^S_{i,j,l}, G^S_{m,l,x} \rangle \geq \frac{k}{12/(3-t)}.$$

This implies that for at least $(1-O(\epsilon))|S_1|$ vertices x of S_1,

$$\langle G_{i,j,m}, G_x \rangle = \langle G^S_{i,j,m}, G^S_x \rangle + \langle G^{\text{out}}_{i,j,m}, G^{\text{out}}_x \rangle \geq \frac{k}{12/(3-t)} - O(\sqrt{n \log n}) \geq \frac{k}{18/(3-t)}$$

and therefore $x \in V_{ijm}$. This shows that $|S_1 \setminus V_{ijm}| \leq O(\epsilon k)$, and since $\{i,j,m\} \in T^*$ we have $V_{ijm} \setminus S = o(k)$. Therefore choosing $\alpha_1 = \{i,j,m\}$ we get the desired result.

Similarly there exist $\alpha_3, \alpha_4 \in \binom{S_2}{3} \cap T^*$ satisfying the desired property. Therefore there exist $\alpha_1, \alpha_2, \alpha_3, \alpha_4 \in T^*$ such that $|S \Delta (V_{\alpha_1} \cup V_{\alpha_2} \cup V_{\alpha_3} \cup V_{\alpha_4})| \leq O(\epsilon k)$. ◀

From the above two lemmas we can finish proof of Theorem 3 observation the Algorithm 2 gives a n^{12} size list of almost bipartite graphs which will surely contain a graph close to the planted bipartite subgraph on S.

Proof of Theorem 3. Using Algorithm 2 and Lemma 12 there exists $\alpha_i, \alpha_j, \alpha_w, \alpha_l \in \binom{S}{3}$ such that the subgraph induced on the set $V_{\alpha_i} \cup V_{\alpha_j} \cup V_{\alpha_w} \cup V_{\alpha_l}$ satisfies the conditions of Theorem 3. Since Algorithm 1 enumerates over all choice of α_i, the size of list is at most n^{12}. ◀

▶ Remark 13. For any $t \geq 1$, in particular for $t \geq 3$, we can still recover most of S_2 as follows. Using the arguments of Lemma 12, we can show that there exist $\alpha, \beta \in \binom{S_2}{3}$ such that $(S_2 \setminus (V_\alpha \cup V_\beta) \leq O(\epsilon k)$ where $V_\alpha = \left\{ i \in V \mid \langle G_\alpha, G_i \rangle \geq \frac{k}{6} \right\}$. Therefore Algorithm 2 can be modified to output a list of subgraphs from which at least one subgraph W will satisfy $|W \cap S| \geq (1-O(\epsilon))tk/(1+t)$ and $|W \setminus S| \leq o(k)$.

3.3 Semirandom Planted Tripartite Subgraph

Most of the work required to prove Theorem 5 is already done in the previous subsections. We use Algorithm 3 to prove Theorem 5.

Algorithm 3 Greedy algorithm for a n^{18} sized list containing the planted tripartite graph.

for $i,j,w = 1$ to n **do**
    Collect all the vertices t such that $\langle G_{ijw}, G_t \rangle \geq \frac{k}{18/(3-t)}$, call that set V_{ijw}
end for
for $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6 = 1$ to $\binom{n}{3}$ **do**
    Output the subgraph induced by the vertex set $V_{\alpha_1} \cup V_{\alpha_2} \cup V_{\alpha_3} \cup V_{\alpha_4} \cup V_{\alpha_5} \cup V_{\alpha_6}$.
end for

Proof of Theorem 5. Observe that if $i, j \in S_1$ then

$$\langle G_i^S, G_j^S \rangle = k - 2|N_i \Delta N_j|$$

where N_i denotes the set of neighbors of i in S. $\langle G_i^S, G_j^S \rangle$ doesn't depend on the edges between S_2 and S_3, this case is similar to the case of SRPB$(n, k, 1/2, 2)$. Therefore using Lemma 12, there exists six subsets of S $\alpha_1, \alpha_2 \in \binom{S_1}{3}, \alpha_3, \alpha_4 \in \binom{S_2}{3}, \alpha_5, \alpha_6 \in \binom{S_3}{3}$ which satisfies $|S_1 \setminus (V_{\alpha_1} \cup V_{\alpha_2})| = O(\epsilon k)$ and $|(V_{\alpha_1} \cup V_{\alpha_2}) \setminus S_1| = o(k)$, and similarly for $V_{\alpha_3} \cup V_{\alpha_4}$ with S_2 and $V_{\alpha_5} \cup V_{\alpha_6}$ with S_3. So the Algorithm 3 gives the desired result. ◀

4 Conclusion

In this work, we gave algorithms for recovering planted SRPB. In Theorem 2 we output a list of n^4 subgraphs in which one subgraph is guaranteed to have to have small symmetric difference with S, whereas in Theorem 3 the list size is n^{12}. A natural open question is to obtain list size of $(1 + o(1))\binom{\lfloor n/k \rfloor}{2}$. [3, 11] obtain an optimal list size of $(1 + o(1))n/k$ for the semirandom planted clique problem by using some combinatorial methods to reduce the size of their list obtained from their greedy procedure. Such techniques don't seem to directly extend to SRPB.

Our algorithms also require the planted bipartite subgraph to be close to being balanced, i.e., $t \in [1, 3)$. Proving an analogue of Theorem 2 and Theorem 3 is also an open question whenever $t \geq 3$.

References

1. Amit Agarwal, Moses Charikar, Konstantin Makarychev, and Yury Makarychev. O($\sqrt{\log n}$) approximation algorithms for min uncut, min 2cnf deletion, and directed cut problems. In *Proceedings of the Thirty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '05, pages 573–581, New York, NY, USA, 2005. Association for Computing Machinery. doi:10.1145/1060590.1060675.
2. Noga Alon, Michael Krivelevich, and Benny Sudakov. Finding a large hidden clique in a random graph. In *Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '98, pages 594–598, USA, 1998. Society for Industrial and Applied Mathematics. URL: http://dl.acm.org/citation.cfm?id=314613.315014.
3. Jaroslaw Blasiok, Rares-Darius Buhai, Pravesh K. Kothari, and David Steurer. Semirandom Planted Clique and the Restricted Isometry Property . In *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 959–969, Los Alamitos, CA, USA, October 2024. IEEE Computer Society. doi:10.1109/FOCS61266.2024.00064.
4. Rares-Darius Buhai, Pravesh K. Kothari, and David Steurer. Algorithms approaching the threshold for semi-random planted clique. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, STOC 2023, pages 1918–1926, New York, NY, USA, 2023. Association for Computing Machinery. doi:10.1145/3564246.3585184.
5. Moses Charikar, Jacob Steinhardt, and Gregory Valiant. Learning from untrusted data. In *Proceedings of the 49th annual ACM SIGACT symposium on theory of computing*, pages 47–60, 2017. doi:10.1145/3055399.3055491.
6. Uriel Feige and Joe Kilian. Heuristics for semirandom graph problems. *Journal of Computer and System Sciences*, 63(4):639–671, 2001. doi:10.1006/jcss.2001.1773.
7. Suprovat Ghoshal and Anand Louis. Approximation algorithms and hardness for strong unique games. In *Proceedings of the Thirty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '21, pages 414–433, USA, 2021. Society for Industrial and Applied Mathematics. doi:10.1137/1.9781611976465.26.

8 Suprovat Ghoshal and Anand Louis. Approximating csps with outliers. In Amit Chakrabarti and Chaitanya Swamy, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2022, University of Illinois, Urbana-Champaign, USA (Virtual Conference), September 19-21, 2022*, volume 245 of *LIPIcs*, pages 43:1–43:16. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022. doi:10.4230/LIPIcs.APPROX/RANDOM.2022.43.

9 Suprovat Ghoshal, Anand Louis, and Rahul Raychaudhury. Approximation Algorithms for Partially Colorable Graphs. In Dimitris Achlioptas and László A. Végh, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019)*, volume 145 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 28:1–28:20, Dagstuhl, Germany, 2019. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.APPROX-RANDOM.2019.28.

10 Michel X. Goemans and David P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42(6):1115–1145, November 1995. doi:10.1145/227683.227684.

11 Venkatesan Guruswami and Hsin-Po Wang. Semirandom planted clique via 1 norm isometry property. *International Conference on Integer Programming and Combinatorial Optimization*, pages 270–282, 2025.

12 Johan Håstad. Clique is hard to approximate within $n^{1-\epsilon}$. *Acta Mathematica*, 182(1):105–142, 1999. doi:10.1007/BF02392825.

13 Akash Kumar, Anand Louis, and Rameesh Paul. Exact Recovery Algorithm for Planted Bipartite Graph in Semi-Random Graphs. In Mikołaj Bojańczyk, Emanuela Merelli, and David P. Woodruff, editors, *49th International Colloquium on Automata, Languages, and Programming (ICALP 2022)*, volume 229 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 84:1–84:20, Dagstuhl, Germany, 2022. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.ICALP.2022.84.

14 Anand Louis and Rameesh Paul. No sdp needed for efficiently recovering planted regular bipartite graphs. In *2026 SIAM Symposium on Simplicity in Algorithms (SOSA)*, pages 538–550. SIAM, 2026.

15 Anand Louis, Rameesh Paul, and Prasad Raghavendra. Robust algorithms for recovering planted r-colorable graphs. In Nika Haghtalab and Ankur Moitra, editors, *Proceedings of Thirty Eighth Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pages 3766–3794. PMLR, 30 June–04 July 2025. URL: https://proceedings.mlr.press/v291/louis25a.html.

16 Theo McKenzie, Hermish Mehta, and Luca Trevisan. A new algorithm for the robust semi-random independent set problem. In *Proceedings of the Thirty-First Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '20, pages 738–746, USA, 2020. Society for Industrial and Applied Mathematics. doi:10.1137/1.9781611975994.45.

17 Jacob Steinhardt. Does robustness imply tractability? A lower bound for planted clique in the semi-random model. *Electron. Colloquium Comput. Complex.*, TR17-069, 2017. URL: https://eccc.weizmann.ac.il/report/2017/069.

18 Mihalis Yannakakis. Node-and edge-deletion np-complete problems. In *Proceedings of the Tenth Annual ACM Symposium on Theory of Computing*, STOC '78, pages 253–264, New York, NY, USA, 1978. Association for Computing Machinery. doi:10.1145/800133.804355.