Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 6 (2014)

Inapproximability of Feedback Vertex Set for Bounded Length Cycles

Venkatesan Guruswami* Euiwoong Lee†

Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213.

Abstract

The Feedback Vertex Set problem (FVS), where the goal is to find a small subset of vertices that intersects every cycle in an input directed graph, is among the fundamental problems whose approximability is not well-understood. One can efficiently find an $\widetilde{O}(\log n)$ factor approximation, but the best NP-hardness result is only a factor of ≈ 1.36 (via a simple reduction from Vertex Cover). A constant-factor approximation is ruled out under the Unique Games Conjecture (UGC), and we give a simpler proof of this in the paper.

Our main result concerns a natural variant of FVS, where the goal is to find a small subset of vertices that intersects every cycle of *bounded length*. For this variant, we prove strong *NP-hardness* of approximation results: For any integer constant $k \geqslant 3$ and $\epsilon > 0$, it is hard to find a $(k-1-\epsilon)$-approximate solution to the problem of intersecting every cycle of length at most k. The hardness result almost matches the trivial factor k approximation algorithm for the problem. In fact, the hardness holds also for the problem of hitting every cycle of length at most a parameter $k' \geqslant k$ where k' can be taken to be $\Omega(\frac{\log n}{\log \log n})$. Taking $k' = \omega(\log n \log \log n)$ would be enough to prove a hardness for FVS (for arbitrary length cycles). Our work thus identifies the problem of hitting cycles of length $\approx \log n$ as the key towards resolving the approximability of FVS.

Our result is based on reductions from k-uniform Hypergraph Vertex Cover with random matching and labeling techniques. As byproducts of our techniques, we also prove a factor $(k-1-\epsilon)$ hardness of approximation result for k-Clique Transversal, where one must hit every k-clique in the graph with fewest possible vertices, and a factor $\Omega(k)$ hardness result for finding a minimum-sized set of *edges* to hit all k-cycles. We also obtain almost tight $\widetilde{\Omega}(k)$ factor hardness results for the dual problem of packing vertex-disjoint k-cycles and k-cliques in a graph, albeit relying on the UGC for k-Cycle Packing (but we do get a weaker factor $\widetilde{\Omega}(\sqrt{k})$ NP-hardness result).

*Supported in part by NSF grant CCF-1115525. guruswami@cmu.edu

†Supported by a Samsung Fellowship, US-Israel BSF grant 2008293, and NSF CCF-1115525. euiwoonl@cs.cmu.edu

ISSN 1433-8092

1 Introduction

Feedback Vertex Set (FVS) is a fundamental combinatorial optimization problem. Given a (directed) graph G, the problem asks to find a subset F of vertices[1] with the minimum cardinality that intersects every cycle in the graph (equivalently, the induced subgraph $G \setminus F$ is acyclic). One of Karp's 21 NP-complete problems, FVS has been a subject of active research for many years. Recent results on the problem study *approximability* and *fixed-parameter tractibility*. In fixed-parameter tractibility, both undirected and directed FVS are shown to be in FPT [9, 11]. See recent results on a generalization of FVS [18, 12] and references therein. In this work, we focus on approximability.

FVS in undirected graphs has a 2-approximation algorithm [5, 8, 13], but the same problem is not well-understood in directed graphs. The best approximation algorithm [35, 22, 21] achieves an approximation factor of $O(\min(\log \tau^* \log\log \tau^*, \log n \log\log n))$, where τ^* is the optimal fractional solution in the natural LP relaxation[2]. The best hardness result follows from a simple approximation preserving reduction from Vertex Cover, which implies that it is NP-hard to approximate FVS within a factor of 1.36 [20]. Assuming the Unique Games Conjecture (UGC) [27], it is NP-hard (called *UG-hard*) to approximate FVS in directed graphs within any constant factor [24, 36]. The main challenge is to *bypass* the UGC and to show a super-constant inapproximability result for FVS assuming only $\mathsf{P} \neq \mathsf{NP}$ or $\mathsf{NP} \not\subseteq \mathsf{BPP}$.

While there is a huge gap between inapproximability results with and without the UGC for many problems (see [28] for a survey), there are also many notable problems where the NP-hardness results almost match their counterparts that rely on the UGC. The k-uniform Hypergraph Vertex Cover problem (Ek-HVC) is one of these problems. Given a k-uniform hypergraph, Ek-HVC asks to find a subset of vertices with the minimum cardinality that intersects every hyperedge in the hypergraph. There is a natural k-approximation algorithm and matching hardness results based on the UGC [29]. Even without relying on the UGC, it is NP-hard to approximate the same problem within a factor less than $k-1$ [19]. This strong hardness, as well as its similarity to FVS (both problems ask to find a subset of vertices that intersects every given set in a certain family) makes Ek-HVC a natural starting point for a reduction to show inapproximability of FVS.

Our result shows that this method is indeed effective, at least for a natural variant of FVS — intersecting all cycles of *bounded* length. Our main result is stated in the following two theorems, very similar to each other but incomparable (and based on different reduction techniques). Their differences are highlighted in boldface. As the results are based on reductions from Ek-HVC as a black box, *the completeness guarantee can be improved to $\frac{1}{k} + \epsilon$ for both results under the UGC.* But as the focus of this work is to bypass the UGC in hardness results for FVS, we only state the NP-hardness versions.

Theorem 1.1. *Fix an integer $k \geqslant 3$ and $\epsilon \in (0,1)$. Given a graph $G = (V_G, E_G)$* **(directed or undirected)***, unless $\mathsf{NP} \subseteq \mathsf{BPP}$, there is no polynomial time algorithm to tell apart the following two cases.*

- *Completeness: There exists $F \subseteq V_G$ with $\frac{1}{k-1} + \epsilon$ fraction of vertices that intersects every cycle of* **length** *$O(\frac{\log n}{\log\log n})$ (hidden constant in O depends on k and ϵ).*

[1] The related Feedback *Arc* Set problem asks for a subset of *edges* to intersect every cycle. This problem is easy on undirected graphs, and equivalent to FVS for directed graphs. In this paper, we deal with the vertex variant.

[2] In unweighted cases, τ^* is always at most n. In weighted cases, we assume all weights are at least 1.

- *Soundness: Every subset F with less than $1-\epsilon$ fraction of vertices does not intersect at least one cycle of length k. Equivalently, any subset with more than ϵ fraction of vertices has a cycle of length exactly k in the induced subgraph.*

Theorem 1.2. *Fix an integer $k \geqslant 3$ and $\epsilon \in (0,1)$. Given a* **directed** *graph $G = (V_G, E_G)$, unless $\mathsf{NP} \subseteq \mathsf{P}$, there is no polynomial time algorithm that distinguishes between the following two cases.*

- *Completeness: There exists $F \subseteq V_G$ with $\frac{1}{k-1} + \epsilon$ fraction of vertices that intersects every cycle of* **length** *$O(\log \log n)$ (hidden constant in O depends on k and ϵ).*

- *Soundness: Any subset with more than ϵ fraction of vertices has a cycle of length exactly k in the induced subgraph.*

Finding a small subset of vertices to intersect all short cycles is a natural and well-motivated optimization problem as the complement of such a subset induces a subgraph of large girth which is a useful property in many settings. The above theorems also give another (very structured) class of covering problems for which a NP-hardness result almost matches the integrality gap of the simple linear programming relaxation.

The *edge-version* of the problem, where the goal is to pick the smallest subset of *edges* to hit every r-cycle was studied in [30], where a factor $(r-1)$ algorithm is given for odd r, generalizing a factor 2 algorithm for $r = 3$ (hitting triangles) from [31]. The best known hardness factor for any r was 1.36 (or $2-\epsilon$ under UGC) via a simple reduction from Vertex Cover, and improving this was left as an open problem in [30]. Our methods, with simple modifications, can be used to show a factor $\Omega(r)$ NP-hardness for this problem (as well as for picking edges to hit all cycles of length *up to* r) on both directed and undirected graphs. We discuss the edge-version of the problem in Appendix A.

To state a final reason for our interest in the bounded length version of FVS, if there exists a set F that has c fraction of vertices and intersects every cycle of length slightly bigger than $\Omega(\log n \log\log n)$ (say $\log^{1.1} n$), then $c + o_n(1)$ fraction of vertices are enough to intersect *every* cycle. Our result shows that intersecting cycles of length around $\log n$ is the main obstacle to overcome in extending our result to FVS for arbitrary cycles.

The above two theorems use quite different techniques: *random matching* for Theorem 1.1 and *labeling gadget* for Theorem 1.2. Ignoring the difference between randomized and deterministic reductions, Theorem 1.1 is stronger than Theorem 1.2. However, we include both proofs since the proof using labeling gadget seems interesting in its own right and might have future uses. The random matching techniques also applies to undirected graphs, and as FVS on undirected graphs admits a factor 2 approximation [5, 8, 13], the labeling gadget based approach, which exploits edge directions, might be more suited to proving inapproximability of the original FVS problem. See Section 2 for a more detailed discussion.

Simpler Unique Games-Hardness. The first UG-hardness of approximating FVS within a constant factor [24] is the corollary the fairly complicated result on the Maximum Acyclic Subgraph (MAS). Svensson [36] gave a simpler proof tailored for FVS with a stronger statement in completeness — deleting $\frac{1+\epsilon}{k}$ fraction of vertices ensures that there is no walk of length k. We obtained a simpler proof of the same statement (presented in Appendix E), that differs from [36] in two aspects:

- The ingenious application of *It Ain't Over Till It's Over* Theorem is replaced by the more general *Invariance principle* of Mossel [33].

- The reduction from the Unique Games is simpler, introducing only one long code for each vertex of the Unique Games instance, while [36] used multiple long codes for each tuple of vertices of a certain length. Instead we rely on the stronger (but equivalent) UGC proposed by Khot and Regev [29] for Ek-HVC.

k-Clique Transversal. Our technique for Theorem 1.1 can also be applied to k-Clique Transversal. Given an undirected graph $G = (V_G, E_G)$, the problem asks to find the smallest subset of vertices that intersects every clique of size k (equivalent to intersecting every clique of size at least k).

Assuming the UGC, we have a strong hardness result of Bansal and Khot [6] for Ek-HVC showing factor $k - \epsilon$ inapproximability even when the hypergraph has a small set whose removal makes it k-partite (a formal statement appears in Appendix C). A simple reduction (replacing each hyperedge by a k-clique) then implies a factor $k-\epsilon$ inapproximability for k-Clique Transversal assuming the UGC. Applying the random matching technique again, we obtain the following theorem that matches the best hardness result for Ek-HVC *without relying on the UGC*. The proof is almost identical to the proof of Theorem 1.1, except for a few details due to combinatorial differences between cycles and cliques. We defer it to Appendix C.

Theorem 1.3. *Fix an integer $k \geqslant 3$ and $\epsilon \in (0, 1)$. Given an undirected graph $G = (V_G, E_G)$, unless* $\mathsf{NP} \subseteq \mathsf{BPP}$, *there is no polynomial time algorithm that distinguishes between the following cases.*

- *Completeness: There is a subset $F \subseteq V_G$ with $\frac{1}{k-1} + \epsilon$ fraction of vertices that intersects every clique of size k.*
- *Soundness: Every subset F with less than $1 - \epsilon$ fraction of vertices does not intersect at least one clique of size k.*

k-Cycle Packing and k-Clique Packing. As another application of our technique, we show inapproximability of *k-Cycle Packing.* Given an undirected graph, the problem asks to find the maximum number of vertex-disjoint cycles of length at most k. For the problem of packing cycles of any length, called Vertex-Disjoint Cycle Packing (VDCP), the results of [32, 23] imply that the best approximation factor by any polynomial time algorithm lies between $\Omega(\sqrt{\log n})$ and $O(\log n)$. In a closely related problem Edge-Disjoint Cycle Packing (EDCP), the same papers showed that $\Theta(\log n)$ is the best possible. *k-Clique Packing* is a similar problem where we want to find the maximum number of vertex-disjoint cliques of size exactly k. Both k-Cycle Packing and k-Clique Packing are special cases of k-Set Packing, where the instance consists of a family of k-sets (without any structure relating the sets). k-Set Packing admits a $(k + 1 + \epsilon)/3$-approximation algorithm [17], and is NP-hard to approximate within $\Omega(k/\log k)$ [26].

The problem we reduce from is another special case of k-Set Packing, which is Maximum Independent Set in degree-k graphs (MIS-k). Given a graph with the maximum degree at most k, the problem asks to find the largest subset of vertices such that there is no edge in the induced subgraph. A restricted version of k-Set Packing, MIS-k has also almost matching hardness results assuming $\mathsf{P} \neq \mathsf{NP}$ ($\Omega(k/\log^4 k)$ [10]) or the UGC ($\Omega(k/\log^2 k)$ [4]). Using elementary but elegant properties of graphs (compared to hypergraphs) and the random matching technique, we show the following results. The details are deferred to Appendix D.

Theorem 1.4. *For every sufficiently large k, it is NP-hard to approximate k-Clique Packing within a factor of $\Omega(k/\log^4 k)$, and it is Unique Games-hard to approximate the same problem within a factor of $\Omega(k/\log^2 k)$.*

Theorem 1.5. *For every sufficiently large k, unless $\mathsf{NP} \not\subseteq \mathsf{BPP}$, no polynomial time algorithm can approximate k-Cycle Packing within a factor of $\Omega(\sqrt{k}/\log^3 k)$. It is Unique Games-hard (under randomized reductions) to approximate the same problem within a factor of $\Omega(k/\log^4 k)$.*

2 Techniques

We use NP-hardness of approximating Ek-HVC as a black box, and reduce it to FVS for bounded length cycles. The simplest try will be, given a hypergraph $H = (V_H, E_H)$ (let $n = |V_H|, m = |E_H|$), to produce a graph $G = (V_G, E_G)$ where $V_G = V_H$, and for each hyperedge $e = (v^1, \dots, v^k)$ add k edges that form a *canonical cycle* $(v^1, v^2, \dots, v^k)$ to E_G. While the soundness follows directly (if $F \subseteq V_H$ contains a hyperedge, the same F also contains a k-cycle), the completeness property does not hold since edges that belong to different canonical cycles may form an unintended non-canonical cycle. To prevent this, a natural strategy is to replace one vertex by a set of many vertices (call it a *cloud*), and for each hyperedge $(v^1, \dots, v^k)$, add many canonical cycles on the k clouds (each cycle consists of one vertex from each cloud). If we have too many canonical cycles, soundness works easily but completeness is hard to show due to the risk posed by non-canonical cycles, and in the other extreme, having too few canonical cycles could result in the violation of the soundness property. Therefore, it is important to control the structure (number) of canonical cycles that ensure both completeness and soundness at the same time.

Random Matching. Our first technique, which we call *random matching*, proceeds by creating many *random* cycles for each hyperedge.

- Each cloud consists of B vertices (just a set of vertices without any additional structure)
- For each hyperedge $(v^1, \dots, v^k)$, for aB times,
  - Sample one vertex from each cloud, and add a canonical cycle between k picked vertices.

It can be shown that by cleverly choosing a and B, the above scheme works to ensure both soundness and completeness, thus proving Theorem 1.1. The soundness follows from that of Ek-HVC combined with a standard technique. Unlike typical reductions, the completeness is more complicated. The number of canonical cycles is aBm, but they all intersect $C \times [B]$, where C is a vertex cover of H. We show that the number of non-canonical cycles of length up to k' is $O(n(k')^{k'})$ (hidden constant depends on k). By taking large B and $k' = O(\frac{\log n}{\log\log n})$, this number becomes $o(N)$, where $N = nB$ is the number of vertices in the final graph. Therefore, hitting such non-canonical cycles using one additional vertex per cycle does not increase the size of the feedback vertex set too much. We remark that properties of random matchings are also used to bound the number of short non-canonical paths in inapproximability results for edge-disjoint paths on undirected graphs [2, 1]. The details in our case are different as we create random *hypergraphs* based on *many* random matchings.

Thus, in the completeness case, we can ensure the existence of a set of small measure that intersects all cycles of length up to $O(\frac{\log n}{\log\log n})$. This is close to showing inapproximability of FVS in the following sense. Consider the following standard linear programming relaxation for FVS.

$$\min \sum_{v \in V_G} x_v \quad \text{subject to} \quad \sum_{v \in C} x_v \geqslant 1 \quad \forall \text{ cycle } C\,, \quad \text{and} \quad 0 \leqslant x_v \leqslant 1 \;\; \forall v \in V_G$$

The integrality gap of the above LP is upper bounded by $O(\log n)$ for undirected graphs [7] and $O(\log n \log\log n)$ for directed graphs [22]. Suppose in the completeness case, there exists a set of measure c that intersects every cycle of length bigger than the known integrality gaps, say $\log^{1.1} n$. If we remove these vertices and consider the above LP on the remaining graphs, since every cycle is of length at least $\log^{1.1} n$, setting $x_v = 1/\log^{1.1} n$ is a feasible solution, implying that the optimal solution to the LP is at most $n/\log^{1.1} n$. Since the integrality gap is at most $O(\log n \log\log n)$, we can conclude that the remaining cycles can be hit by at most $O(n \log\log n/\log^{0.1} n) = o_n(n)$ vertices, extending completeness result to every cycle. Thus, improving our result to hit cycles of length $\omega(\log n \log\log n)$ in the completeness case will prove inapproximability of FVS.

Another interesting aspect about Theorem 1.1 is that it also holds for undirected graphs. This should be contrasted with the fact that undirected graphs admit a 2-approximation algorithm for FVS, suggesting that to overcome $\log n$-cycle barrier mentioned above, some properties of directed graphs must be exploited. Towards developing a directed graph specific approach, we now present a different reduction technique.

Labeling Gadget. Our second technique, *labeling gadget*, explicitly controls the structure of every cycle. The idea of labeling gadgets to prove hardness of approximation has been used previously to show inapproximability of edge-disjoint paths problem with congestion and directed cut problems [3, 15, 16].

In this work, the labeling gadget is a directed graph $L = (V_L, E_L)$ with roughly the following properties: (i) Its girth is k, and (ii) Every subset of vertices of measure at least δ has at least one cycle of length k.

Given H and L, $G = (V_G, E_G)$ is constructed as the following way. $V_G = V_H \times V_L^m$, where m is the number of hyperedges in H (let's say the hyperedges are $e^1, e^2, \ldots, e^m$), so that the cloud for each vertex $v \in V$ becomes V_L^m. Each copy of L corresponds to one of the m hyperedge. Consider the naive approach introduced earlier where we added k edges for each hyperedge (multiple edges possible), without duplicating vertices. Call this graph $H' = (V_H, E'_H)$. In G, we add an edge from $(v, x^1, \ldots, x^m)$ to $(u, y^1, \ldots, y^m)$ if and only if

- There is an edge $(v, u) \in E'_H$ created by a hyperedge e^i for some i.
- $x^j = y^j$ for all $j \neq i$, and
- $(x^i, y^i) \in E_L$.

Intuitively, if we want to move from $(u, \ldots)$ to $(v, \ldots)$ where the edge $(u, v) \in E_{H'}$ is created by a hyperedge e^i, then we need to move the ith coordinate by an edge of L (other coordinates stay put). Once we changed the ith coordinate, since L has girth k, we have to use an edge formed by e^i at least k times to move ith coordinate back to the original solution.

Suppose $\mathcal{C} = ((v^1, \ldots), \cdots, (v^{k'}, \ldots))$ is a cycle in G. By the above argument, $(v^1, \ldots, v^{k'})$ is a cycle of H', and must use at least k edges formed by a single hyperedge, say e^l. This is not quite enough to argue that this cycle intersects a vertex cover of H as the same edge of H' that is created by hyperedge e^i may be used multiple times. To fix this problem, we color each edge of L by one of k colors and associate a different color to the k edges formed by a hyperedge. If we ensure the stronger property in the labeling gadget that every cycle of L must be colorful (which implies that the girth is at least k), then the cycle $\mathcal{C} = ((v^1, \ldots), \cdots, (v^{k'}, \ldots))$ uses all k edges formed by a single hyperedge, so it must intersect any vertex cover of H. See Figure 1 for an example.

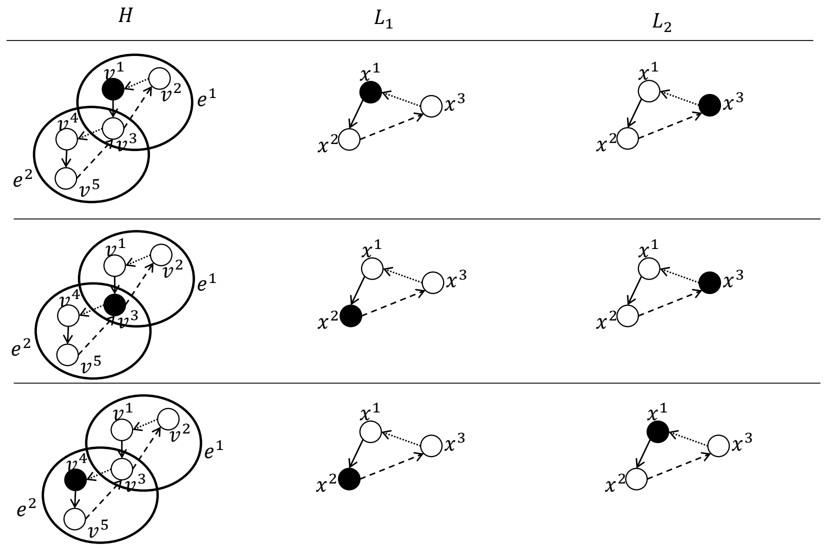

Figure 1: Example with $k = 3$. Each row corresponds to a vertex of G ((v^1, x^1, x^3) in the first row), and each edge of H and L has one of 3 types. From (v^1, x^1, x^3), we used e_1 and the solid edge to get to (v^3, x^2, x^3). The position in L_2 stays the same. From (v^3, x^2, x^3), we used e_2 and the dotted edge to get to (v^4, x^2, x^1).

For soundness, given a subset $F \subseteq V_G$ of measure δ, we find a hyperedge $e = (v^1, \ldots, v^k)$ such that $(\bigcap \mathsf{cloud}(v^i)) \cap F$ is large. This follows from averaging arguments and needs a *density* guarantee in the soundness case of Ek-HVC. Then we focus on the copy of L associated with e, find a colorful k-cycle in L, and produce the final cycle by combining two cycles $(v^1, v^2), \cdots, (v^k, v^1)$ (from V_H) and the colorful cycle in L.

This is a complete and sound reduction from Ek-HVC to the *original* FVS problem, except that it blows up the size of the instance exponentially. To get a polynomial time reduction, we compress the construction by coalescing different copies of L, retaining only a constant number (dependent on the degree of the original hypergraph) out of the m coordinates. However, as a result we are not able to control the behavior of long cycles, and we may not intersect all cycles in the completeness case of Theorem 1.2. Since we have good control over the structure of cycles using labeling gadgets, and the only issue is to reduce the size of labels, we hope that more sophisticated variants of this technique might be able to prove inapproximability of FVS itself.

3 Preliminaries

Ek-HVC. A k-uniform hypergraph is denoted by $H = (V_H, E_H)$ such that each $e \in E_H$ is a k-subset of V_H. An instance of Ek-HVC consists of k-uniform hypergraph H, where the goal is to find a set $C \subseteq V_H$ with the minimum cardinality such that it intersects every edge; $C \cap e \neq \emptyset$ for every $e \in E_H$. Both in Ek-HVC and FVS, we assume that vertices are unweighted. Since we show hardness of approximation, dealing with unweighted cases yields a stronger statement.

The NP-hardness results of Ek-HVC we use have the following form. Given a k-uniform hypergraph H, it is NP-hard to distinguish

- Completeness: There exists a vertex cover of measure at most $c(k)$.[3]
- Soundness: Every vertex cover has to have measure at least $s(k)$.

In some cases, we need a fact that H has small degrees (only function of k) as well as the following additional *density* property in the soundness case. In a hypergraph, we define the degree of a vertex to be the number of hyperedges containing it.

- The maximum degree of H is bounded by $d(k)$.
- In the soundness case above, every set of measure at least $\delta(k)$ contains $\rho(k)$ fraction of hyperedges in the induced subgraph for some $\delta(k) \geqslant 1 - s(k)$.

For example, the NP-hardness of Ek-HVC with $c(k) = \frac{3}{k}$, $s(k) = 1 - \frac{1}{k}$, $d(k) = 2^{k^\beta}$, $\delta(k) = \frac{2}{k}$, and $\rho(k) = \frac{1}{k2^{k^\beta}}$ for some β is made explicit in [14]. However, careful examination of other results, especially that of [19], yields a better result. See Appendix B for the proof.

Theorem 3.1 ([19]). *For any rational function $\epsilon(k) > 0$, hardness of Ek-HVC holds with $c(k) = \frac{1}{k-1} + \epsilon(k), s(k) = 1 - \epsilon(k), \delta(k) = 2\epsilon(k)$ and $d(k), \rho(k)$ are some positive functions of k (and $\epsilon(k)$).*

FVS and k-(Directed)-CT. A graph is denoted by $G = (V_G, E_G)$ where each edge is denoted as $e = (v^i, v^j)$. If G is undirected, (v^i, v^j) is an unordered pair, and if G is directed, (v^i, v^j) is an ordered pair and represents an edge from v^i to v^j. Let $[k] := \{1, 2, \ldots, k\}$. For notational convenience, for $i \in \mathbb{Z}$, let $(i) \in [k]$ be the unique integer such that $i \equiv (i) \mod k$. Given a graph $G = (V_G, E_G)$, let $(v^1, \ldots, v^k)$ denote a cycle that consists of edges $(v^i, v^{(i+1)})_{1 \leqslant i \leqslant k}$.

An instance of k-(Directed)-Cycle Transversal (k-(Directed)-CT) consists of a (directed) graph $G = (V_G, E_G)$, where the goal is to find a set $F \subseteq V_G$ with the minimum cardinality such that it intersects every cycle of length at most k; in other words, the induced subgraph $G \setminus F$ does not have a cycle of length at most k.[4] If k is unbounded ($k = |V_G|$ suffices), it becomes the well-known (directed) Feedback Vertex Set (FVS) problem.

Notation. We now discuss the notational convention used throughout the paper. An (hyper)edge e of a k-uniform hypergraph H is a subset of V_H of size k. We denote e as an *ordered* k-tuple $e = (v^1, \ldots, v^k)$. The ordering can be chosen arbitrarily given H, but should be fixed throughout. For an integer m, let $[m] = \{1, 2, \ldots, m\}$. For $i \in [m]$, let $(i+1)$ be the unique integer such that $i + 1 \equiv (i+1) \mod m$ (m will be clear in the context). In many cases, vertices are represented as a vector (i.e. in n-dimensional hypercube, $V = \{0,1\}^n$ and $v = (v_1, \ldots, v_n)$ is a n-dimensional vector). To avoid confusion, if v indicates a vertex of some graph, we use a superscript v^i to denote another vertex of the same graph, and subscripts v_i to denote the ith coordinate of v. We also use a superscript e^i to denote the ith (hyper)edge. For any set V and its subset F, let $\mu_V(F)$ be the measure of F under the uniform measure in V (subscripts might be omitted if clear from the context).

[3]Usually, the result is stated as $c(k) - \epsilon$ for arbitrary $\epsilon > 0$. For simplicity, we consider ϵ as a function of k, e.g. $1/k^2$.

[4]It is also natural to find a set that intersects of every cycle of length *exactly* k; both our results hold for this case as well.

4 Random Matching

In this section, we give a reduction from Ek-HVC to k-CT and prove Theorem 1.1. Fix k and let $c := c(k), s := s(k), d := d(k)$ be the parameters from Theorem 3.1 (we do not need density). Let a and B be integer constants greater than 1, which will be determined later. Lemma 4.1 and 4.3 with these parameters imply Theorem 1.1. We focus on undirected graphs in this section, but the result with the same proof holds for directed graphs.

4.1 Reduction

Given a hypergraph $H = (V_H, E_H)$, construct an undirected graph $G = (V_G, E_G)$ such that

- $V_G = V_H \times [B]$. Let $n = |V_H|$ and $N = |V_G| = nB$. For $v \in V_H$, let $\mathsf{cloud}(v) := \{v\} \times [B]$ be the copy of $[B]$ associated with v.
- For each hyperedge $e = (v^1, \ldots, v^k)$, for aB times, take $(l^1, \ldots, l^k)$ independently and uniformly from $[B]$. Add $((v^i, l^i), (v^{(i+1)}, l^{(i+1)}))$ for $i \in [k]$. Each time $((v^1, l^1), \ldots, (v^k, l^k))$ is a cycle of length k, and we have aB of such cycles per each hyperedge. Call such cycles *canonical*.

4.2 Completeness

Lemma 4.1. *Suppose H has a vertex cover C of measure c. For any $\epsilon > 0$, with probability at least $3/4$, there exists a subset $F \subseteq G$ of measure at most $c + \epsilon$ such that the induced subgraph $G \setminus F$ has no cycle of length $O(\frac{\log n}{\log \log n})$. The constant hidden in O depends on ϵ and the parameters from Ek-HVC (which depend only on k).*

Proof. Let $F = C \times [B]$. We consider the expected number of cycles that avoid F and argue that a small fraction of additional vertices intersect all of these cycles. Let k' be the length of a purported cycle. Choose k' vertices $(v^1, l^1), \ldots, (v^{k'}, l^{k'})$ which satisfy

- $v^1 \in V_H$ can be any vertex.
- $l^1, \ldots, l^{k'} \in B$ can be arbitrary labels.
- For each $1 \leqslant i < k'$, there must be a hyperedge $e = (u^1, \ldots, u^k)$ and $j \in [k]$ such that ($v^i = u^j$ and $v^{i+1} = u^{(j+1)}$) or ($v^i = u^{(j+1)}$ and $v^{i+1} = u^j$). Equivalently, there are edges between $\mathsf{cloud}(v^i)$ and $\mathsf{cloud}(v^{i+1})$.

There are n possible choices for v^1, B choices for each l^i, and $2d$ choices for each v^i $(i > 1)$ (there are at most d hyperedges containing one vertex, and for each canonical cycle, there are two possibilities to choose a neighbor). The number of possibilities to choose such $(v^1, l^1), \ldots, (v^{k'}, l^{k'})$ is bounded by $n(2d)^{k'-1}B^{k'}$. Note that no other k'-tuple of vertices can form a cycle. Further discard the tuple when two vertices are the same (the resulting cycle is not simple and its simple pieces will be considered for smaller k').

We calculate the probability that $((v^1, l^1), \ldots, (v^{k'}, l^{k'}))$ form a cycle (i.e. all k' edges exist) that does not intersect F. For a set of purported edges, we say that this set can be *covered by a single canonical cycle* if one copy of canonical cycle can contain all k' edges with nonzero probability. Suppose that all k' edges in the purported cycle can be covered by a single canonical cycle. It is only possible when $k' = k$ and there is a hyperedge e such that after an appropriate shifting, $e = (v^1, \ldots, v^{k'})$. In this case, $((v^1, l^1), \ldots, (v^{k'}, l^{k'}))$ intersects F (right case of Figure 2). When k' edges of the purported cycle have to be covered by more than one canonical cycle, some vertices must be covered by more than one canonical cycle, and each canonical cycle covering the same vertex should give the same label to that vertex. This redundancy makes it unlikely to have all k' edges exist at the same time (left case of Figure 2). The below lemma formalizes this intuition.

Claim 4.2. *Suppose that $((v^1, l^1), \ldots, (v^{k'}, l^{k'}))$ cannot be covered by a single canonical cycle. Then the probability that it forms a cycle is at most $k'(\frac{adk'}{B})^{k'}$.*

Proof. Fix $2 \leqslant p \leqslant k'$. Partition k' purported edges into p nonempty *groups* $I_1, \ldots, I_p$ such that each group can be covered by a single canonical cycle. There are at most $p^{k'}$ possibilities to partition. For each $v \in V_H$, there are at most d hyperedges containing v and at most aBd canonical cycles intersecting $\mathsf{cloud}(v)$. Therefore, all edges in one group can be covered simultaneously by at most aBd copies of canonical cycles. There are at most $(aBd)^p$ possibilities to assign a canonical cycle to each group. Assume that one canonical cycle is responsible for exactly one group. This is without loss of generality since if one canonical cycle is responsible for many groups, we can merge them and this case can be dealt with smaller p.

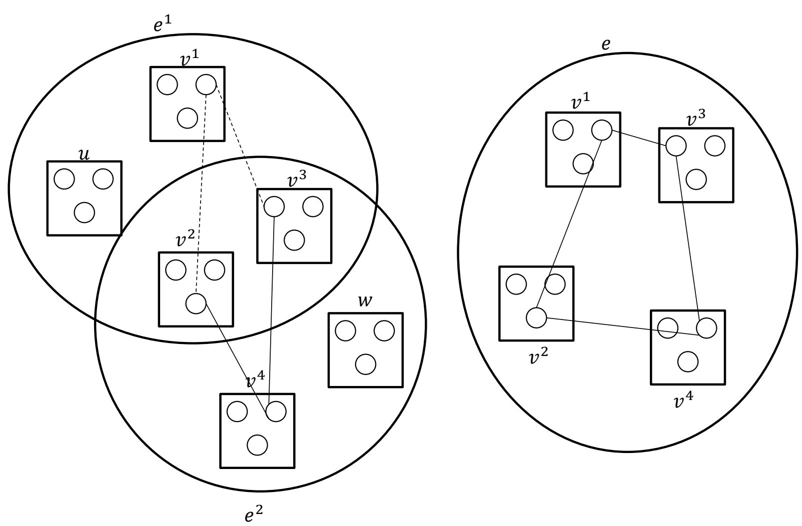

Figure 2: Two examples where $k' = k = 4$. On the left, purported edges are divided into two groups (dashed and solid edges). Each copy of canonical cycle should match the labels of three vertices to ensure it covers 2 designated edges (6 labels total). On the right, one canonical cycle can cover all the edges, and it only needs to match the labels of four vertices (4 labels total).

Focus on one group I of purported edges, and one canonical cycle L which is supposed to cover them. Let $I' \subseteq V_G$ be the set of vertices which are incident on the edges in I. Suppose $L = ((u^1, l'^1), \ldots, (u^k, l'^k))$, which is produced by a hyperedge $f = (u^1, \ldots, u^k) \in E_H$. We calculate the probability that L contains all edges in I over the choice of labels $l'^1, \ldots, l'^k$ for L. One necessary condition is that $\{v | (v, l) \in I' \text{ for some } l \in [B]\}$ (I' projected to V_H) is contained in f. Otherwise, some vertices of I' cannot be covered by L. Another necessary condition is $v^i \neq v^j$ for any $(v^i, l^i) \neq (v^j, l^j) \in I'$. Otherwise ($(v, l^i), (v, l^j) \in I'$ for $l^i \neq l^j$), since L gives only one label to each vertex in $f \subseteq V_H$, (v, l^i) and (v, l^j) cannot be contained in L simultaneously. Therefore, we have a nice characterization of I': It consists of at most one vertex from the cloud of each vertex in f.

Now we make a crucial observation that $|I'| \geqslant |I| + 1$. This is because I is a proper subset of the edges that form a simple cycle. Formally, in the graph with vertices I' and edges I, the maximum degree is at most 2, and there are at least two vertices of degree 1. The probability that L contains I is at most the probability that for each $(v^i, l^i) \in I'$, l^i is equal to the label L assigns to v^i, which is $B^{-|I'|} \leqslant B^{-|I|-1}$.

We conclude that for each partition, the probability of having all the edges is at most

$$(aBd)^p \prod_{q=1}^{p} B^{-|I_q|-1} = \frac{(aBd)^p}{B^{k'+p}} = \frac{(ad)^p}{B^{k'}}.$$

The probability that $((v^1, l^1), \ldots, (v^{k'}, l^{k'}))$ forms a cycle is therefore bounded by

$$\sum_{p=2}^{k'} p^{k'} \frac{(ad)^p}{B^{k'}} \leqslant k' \big(\frac{adk'}{B}\big)^{k'}. \qquad \square$$

Therefore, the expected number of cycles of length k' that avoid F is bounded by $n(2d)^{k'-1} B^{k'} \cdot k' (\frac{adk'}{B})^{k'} \leqslant n(Rk')^{k'}$ where R is a constant depending only on k (not k'). With probability at least $3/4$, the number of such cycles of length up to k' is at most $4n(Rk')^{k'+1}$. Let $B \geqslant \frac{4(Rk')^{k'+1}}{\epsilon}$. Then these cycles can be covered by at most $\epsilon n B = \epsilon N$ vertices. If $k' = \frac{\log n}{\log \log n}$, then $k'^{k'} = \exp(k' \log k')$ is also $o(n)$, we can take B linear in n and $k' \geqslant \Omega(\frac{\log N}{\log \log N})$. $\square$

4.3 Soundness

The soundness claim above is easier to establish. By an averaging argument, a measure 2ϵ subset I of V_G must contain ϵB vertices from the clouds corresponding to a subset S of measure ϵ in V_H. There must be a hyperedge e contained within S, and the chosen parameters ensure that one of the canonical cycles corresponding to e is likely to lie within I.

Lemma 4.3. *If every subset of V_H of measure at least ϵ contains a hyperedge in the induced subgraph, with probability at least $3/4$, every subset of V_G with measure 2ϵ contains a canonical cycle.*

Proof. Fix one hyperedge $e = (v^1, \ldots, v^k)$. We want to ensure that if a subset of vertices I has at least ϵ fraction from each $\mathsf{cloud}(v^i)$, then I will contain a canonical cycle. Fix $A^1 \subseteq \mathsf{cloud}(v^1), \ldots, A^k \subseteq \mathsf{cloud}(v^k)$ be such that for each i, $|A^i| \geqslant \epsilon B$. There are at most 2^{kB} ways to choose such A's. The probability that one canonical cycle associated with e is not contained in $(v^1, A^1) \times \cdots \times (v^k, A^k)$ is at most $1 - \epsilon^k$. The probability that none of canonical cycle associated with e is contained in $(v^1, A^1) \times \cdots \times (v^k, A^k)$ is $(1 - \epsilon^k)^{aB} \leqslant \exp(-aB\epsilon^k)$.

By union bound over all $A^1, \ldots, A^k$, the probability that there exists $A^1, \ldots, A^k$ containing no canonical cycle is at most $\exp(kB - aB\epsilon^k) = \exp(-B) \leqslant \frac{1}{4|E_H|}$ by taking a large enough constant depending on k and ϵ, and $B = \Omega(\log |E_H|)$ (note that B was already taken to be linear in $|V_H|$). Therefore, with probability at least $3/4$, the desired property holds for all hyperedges.

Let I be a subset of V_G of measure at least 2ϵ. By an averaging argument, at least ϵ fraction of *good* vertices $v \in V_H$ satisfy that $|\mathsf{cloud}(v^i) \cap I| \geqslant \epsilon B$. By the soundness property of H, there is a hyperedge contained in the subgraph induced by the good vertices. $\square$

5 Labeling Graph Based Reduction

In this section, we show NP-hardness of k-Directed-CT by a reduction from Ek-HVC, proving Theorem 1.2. Fix k and let $c := c(k), s := s(k), d := d(k), \delta := \delta(k), \rho := \rho(k)$ be the parameters we have from Theorem 3.1. Lemma 5.5 and 5.6 with these parameters imply Theorem 1.2.

5.1 Labeling Gadget

A (k, δ)-labeling gadget is a directed graph $L = (V_L, E_L)$ with each edge colored with a color from $[k]$ that satisfies the following three properties.

1. Its girth is exactly k.
2. Every cycle has at least one edge for each color.
3. Every subset of vertices of measure at least δ has at least one cycle $(x^1, x^2, \ldots, x^k)$ such that
   - Its length is exactly k.
   - After an appropriate shifting, the color of $(x^i, x^{(i+1)})$ is i.

Let $V_L = [B]^k$, where B will be determined later depending on δ and k. For each $1 \leqslant i \leqslant k$, and for each $x_1, \ldots, x_k$ and $y_i > x_i, y_{(i+1)} > x_{(i+1)}$, we add an edge of color i from

$$(x_1, \ldots, x_i, y_{(i+1)}, \ldots, x_k) \quad \text{to} \quad (x_1, \ldots, y_i, x_{(i+1)}, \ldots, x_k)\ .$$

Intuitively, edges of color i *strictly increase ith coordinate, strictly decrease* $(i+1)$*th coordinate, and do not change the others*.

With this construction, properties 1. and 2. can be shown easily. If a cycle uses an edge of color i, the ith coordinate was decreased by using this edge, and the cycle should use at least one edge of color $(i+1)$ to return. The same argument can be applied to color $(i+1)$, $(i+2)$, $\ldots$, until the cycle uses all the colors. The following lemma shows property 3.

Lemma 5.1. *For $k \in \mathbb{N}$ and $\delta > 0$, there exists an integer $B := B(k, \delta)$ such that a subset $S \subseteq [B]^k$ with measure at least δ contains a k-cycle that has one edge of each color.*

Proof. Fix a subset $S \subseteq [B]^k$ of measure at least δ. For each $x \in [B]^k$ and $i \in [k]$, define $\mathsf{line}(x, i) := \left\{y \in [B]^k : (y)_j = (x)_j \text{ for all } j \neq i\right\}$ to be the axis-parallel line containing x and parallel to the ith unit vector e_i. Let

$$\mathsf{surface}_S(x, i) := \begin{cases} \underset{y \in S \cap \mathsf{line}(x,i)}{\operatorname{argmax}} (y_i) & S \cap \mathsf{line}(x, i) \neq \emptyset \\ \emptyset & S \cap \mathsf{line}(x, i) = \emptyset, \end{cases}$$

and $S' := \cup_{x,i} \mathsf{surface}_S(x, i)$. There are $k \cdot B^{k-1}$ lines total (B points for each line), so $|S'| \leqslant k \cdot B^{k-1}$. If $B > \frac{k}{\delta}$, there is an element in $S \setminus S'$. Call this point $(x_1, \ldots, x_k)$. For any $i \in [k]$, $(x_1, \ldots, x_{i-1}, y_i, x_{i+1}, x_k)$ is also in S for some $y_i > x_i$. $((x_1, x_2, \ldots, x_{k-1}, y_k), (x_1, x_2, \ldots, y_{k-1}, x_k)$, $\ldots$, $(x_1, y_2, \ldots, x_{k-1}, x_k)$, $(y_1, x_2, \ldots, x_{k-1}, x_k)$, $(x_1, x_2, \ldots, x_{k-1}, y_k))$ is a cycle we wanted. $\square$

5.2 Reduction

Let k' be the maximum length of cycles that we want to intersect in the completeness case, which will be determined later. Let $L = (V_L, E_L)$ be a $(k, \rho\delta)$-labeling gadget. We are given a hypergraph $H = (V_H, E_H)$ with the maximum degree d. Since each vertex has a degree at most d, each hyperedge shares a vertex with at most dk other hyperedges. Consider a graph $H' = (V_{H'}, E_{H'})$ where $V_{H'} = E_H$ and there exists an edge between e and f if and only if intersect. Define the distance between two hyperedges e and f to be the minimum distance between e and f in H'. The maximum degree of H' is bounded by dk, and for each $e \in V_{H'}$, there are at most $(dk)^{k'}$ neighbors within distance k'. Therefore, each hyperedge can be colored with $d' = (dk)^{k'} + 1$ colors so that two hyperedges within distance k' are assigned different colors. To distinguish it from the coloring of L, we call the former *outer coloring* and the latter *inner coloring*. We use letters u, v to denote the vertices of V_H, x, y for V_L, and a, b for $(V_L)^{d'}$. Furthermore, since some vertices are indexed by a vector, we use superscripts to denote different vertices (e.g. $x^1, x^2 \in V_L$) and subscripts to denote different coordinates of a single vertex (e.g. $x = (x_1, \ldots, x_k)$).

Our reduction will produce a directed graph $G = (V_G, E_G)$ where $V_G = V_H \times (V_L)^{d'} = V_H \times ([B]^k)^{d'}$. The number of vertices (from H to G) is increased by a factor of $|V_L|^{d'} = |V_L|^{(dk+1)^{k'}}$. Since $|V_L|$ and $dk+1$ only depend on k, this quantity is polynomial in $|V_H|$ if $k' = \Omega(\log\log|V_H|)$. The edges of G are constructed as the following:

- For any $e = (v^1, \ldots, v^k) \in E_H$, let $q \in [d']$ be its (outer) color.
- For any $i \in [k]$,
- For any $x, y \in V_L$ such that $(x, y) \in E_L$ with inner color i,
- For any $a \in (V_L)^{d'}$,
- We put an edge $(v^i, a_{q \mapsto x})$ to $(v^{(i+1)}, a_{q \mapsto y})$ with outer color q and inner color i, where $a_{q \mapsto x}$ means that the qth *outer* coordinate of a (which is an element of V_L) is replaced by x.

For a vertex $(v, a) \in V_G$, consider $a = (x^1, \ldots, x^{d'})$ as a label which is a d'-dimensional vector and each coordinate x^i corresponds to a vertex of L. Following one edge with outer color q changes only x^q (according to L), while leaving the other coordinates unchanged. Based on this fact, it is easy to prove the following lemmas.

Lemma 5.2. *G has girth at least k.*

Proof. From the above discussion, each edge of G acts like an edge for an exactly one copy of L and acts like a self-loop for the other copies of L. If $((v^1, a^1), \ldots, (v^l, a^l))$ is a cycle in G, then each coordinate of a^i is a cycle in L as well. Since L has girth k, G also has girth at least k. □

Definition 5.3 (Canonical cycles)**.** *For any hyperedge $e = (v^1, \ldots, v^k)$ of H with outer color q, for any cycle $x^1, \ldots, x^k$ of L such that $(x^i, x^{(i+1)})$ is colored i for $i = 1, 2, \ldots, k$, and for any $a \in (V_L)^{d'}$, $((v^1, a_{q \to x^1}), \ldots, (v^k, a_{q \to x^k}))$ is also a cycle of G of length exactly k. Call such cycles* canonical.

Lemma 5.4. *Suppose $k \leqslant l \leqslant k'$, and $((u^1, a^1), \ldots, (u^l, a^l))$ be a cycle. Then, there exists a hyperedge e such that $e \subseteq \{u^1, \ldots, u^l\}$.*

Proof. Let one of the edges of the cycle have outer color q. By the properties of L (corresponding to outer color q), for each $i \in [k]$, there must be an edge with outer color q and inner color i. Since the distance between two hyperedges with the same outer color is at least k', every edge with outer color q must be from the same hyperedge, say $e = (v^1, \ldots, v^k)$.

By the property 2. of the labeling gadget corresponding to outer color q (equivalently hyperedge e), for every inner color j, $((u^1, a^1), \ldots, (u^l, a^l))$ must use an edge with inner color j and outer color q. Notice that if $((u^i, a^i), (u^{(i+1)}, a^{(i+1)}))$ is with outer color q and inner color j, $u^i = v^j$ and $u^{(i+1)} = v^{(j+1)}$. Therefore, $e \subseteq \{u^1, \ldots, u^l\}$. $\square$

5.3 Completeness

Lemma 5.5. *If H has a vertex cover of measure c, G has a k'-cycle transversal of measure c.*

Proof. Let $C \subseteq V_H$ be such that it has measure c and intersects every hyperedge $e \in E_H$. Let $F = C \times (V_L)^{d'} \subseteq V_G$. It is clear that F has measure c. We argue that F indeed intersects every cycle of length at most k'. For every cycle $((u^1, u^1), \ldots, (u^l, u^l))$ of length $k \leqslant l \leqslant k'$, by Lemma 5.4, there exists a hyperedge $e = (v^1, \ldots, v^k)$ such that $e \subseteq \{u^1, \ldots, u^l\}$. Since C is a vertex cover for H, there exists $v^i \in C$, so $F \supseteq v^i \times (V_L)^{d'}$ intersects this cycle. $\square$

5.4 Soundness

Lemma 5.6. *If every subset of V_H with measure at least δ contains a ρ fraction of hyperedges in the induced subgraph, every subset of V_G with measure 2δ contains a canonical cycle.*

Proof. Let $I \subseteq V_G$ has measure at least 2δ. For $a \in (V_L)^{d'}$, we let $\mathsf{slice}(a) := V_H \times a$ to be the copy of V_H associated with a. Let $A = \left\{a \in (V_L)^{d'} : \mu_H(\mathsf{slice}(a) \cap I) \geqslant \delta\right\}$. An averaging argument shows that $\mu_{(V_L)^{d'}}(A) \geqslant \delta$. By the soundness property (with density) of Ek-HVC, for each $a \in A$, $\mathsf{slice}(a) \cap I \subseteq V(H)$ contains at least ρ fraction of hyperedges. Therefore, if we consider the product space $E(H) \times V(G)^{d'}$, at least $\rho\delta$ fraction of tuples (e, a) satisfy $e \subseteq \mathsf{slice}(a) \cap I$.

By an averaging argument with respect to E_H, we can conclude that there exists a hyperedge $e = (v^1, \ldots, v^k)$ such that $\rho\delta$ fraction of $a = (x^1, \ldots, x^{d'}) \in (V_L)^{d'}$ satisfies $e \subseteq \mathsf{slice}(a) \cap I$. Without loss of generality, assume that its outer color is 1. Another averaging argument with respect to $x^2, \ldots, x^{d'}$ shows that there exists $(y^2, \ldots, y^{d'})$ such that $X := \left\{x \in L \mid e \subseteq \mathsf{slice}((x, y^2, \ldots, y^m)) \cap I\right\}$ satisfies $\mu_L(X) \geqslant \rho\delta$.

Since L is a $(k, \rho\delta)$-labeling gadget, there exists a cycle $(x^1, \ldots, x^k) \subseteq X$ such that (x^i, x^{i+1}) is colored with i. Our final cycle of G consists of $((v^i, x^i, y^2, \ldots, y^{d'}), (v^{(i+1)}, x^{(i+1)}, y^2, \ldots, y^{d'}))$ for each $i \in [k]$. Note that $(v^i, x^i, y^2, \ldots, y^{d'}) \in I$ for each i since by the definition of X, for each $x^i \in X$, $e \subseteq \mathsf{slice}(x^i, y^2, \ldots, y^{d'}) \cap I$. The edge $((v^i, x^i, y^2, \ldots, y^{d'}), (v^{(i+1)}, x^{(i+1)}, y^2, \ldots, y^{d'}))$ exists by the construction. $\square$

6 Concluding Remarks

We have a proved a near-optimal inapproximability result for FVS problem when restricted to cycles of length k (which is the vertex cover problem on a highly structured type of k-uniform hypergraph). In fact, our results yield a super-constant inapproximability result for the problem of intersecting cycles of length at most $O(\log n / \log\log n)$. Improving this bound on cycle length to $\omega(\log n \log\log n)$ would imply inapproximability for the normal FVS problem. We hope that our results renew the hope to prove a super-constant factor inapproximability for FVS without relying on the Unique Games Conjecture. In particular, the general labeling gadget approach, instantiated with other kinds of label graphs, might hold some promise in this regard.

We close by noting that the soundness guarantee in our results is very strong, implying that any FVS must contain nearly all the vertices. While this is a desirable location for the gap, it is overkill if one just seeks a large inapproximability factor. In this vein, if it helps executing a labeling gadget based approach, we can also reduce from vertex cover on k-partite k-uniform hypergraphs, which is also hard to approximate within $\Omega(k)$ factors [25].

References

[1] M. Andrews, J. Chuzhoy, V. Guruswami, S. Khanna, K. Talwar, and L. Zhang. Inapproximability of edge-disjoint paths and low congestion routing on undirected graphs. *Combinatorica*, 30(5):485–520, 2010. 5

[2] M. Andrews and L. Zhang. Logarithmic hardness of the undirected edge-disjoint paths problem. *J. ACM*, 53(5):745–761, 2006. 5

[3] M. Andrews and L. Zhang. Almost-tight hardness of directed congestion minimization. *J. ACM*, 55(6), 2008. 6

[4] P. Austrin, S. Khot, and M. Safra. Inapproximability of vertex cover and independent set in bounded degree graphs. *Theory of Computing*, 7(1):27–43, 2011. 4, 24, 31

[5] V. Bafna, P. Berman, and T. Fujito. Constant ratio approximations of the weighted feedback vertex set problem for undirected graphs. In *Proceedings of the 6th International Symposium on Algorithms and Computation*, ISAAC '95, pages 142–151, 1995. 2, 3

[6] N. Bansal and S. Khot. Inapproximability of hypergraph vertex cover and applications to scheduling problems. In *Proceedings of the 37th International Colloquium on Automata, Languages and Programming*, ICALP'10, pages 250–261, 2010. 4, 21, 32, 35

[7] R. Bar-Yehuda, D. Geiger, J. Naor, and R. Roth. Approximation algorithms for the feedback vertex set problem with applications to constraint satisfaction and bayesian inference. *SIAM Journal on Computing*, 27(4):942–959, 1998. 6

[8] A. Becker and D. Geiger. Optimization of pearl's method of conditioning and greedy-like approximation algorithms for the vertex feedback set problem. *Artificial Intelligence*, 83(1):167–188, 1996. 2, 3

[9] H. L. Bodlaender. On disjoint cycles. *International Journal of Foundations of Computer Science*, 5(01):59–68, 1994. 2

[10] S. Chan. Approximation resistance from pairwise independent subgroups. In *Proceedings of the 45th annual ACM Symposium on Symposium on Theory of Computing*, STOC '13, pages 447–456, 2013. 4, 23, 29

[11] J. Chen, Y. Liu, S. Lu, B. O'sullivan, and I. Razgon. A fixed-parameter algorithm for the directed feedback vertex set problem. *J. ACM*, 55(5):21:1–21:19, 2008. 2

[12] R. Chitnis, M. Cygan, M. Hajiaghayi, and D. Marx. Directed subset feedback vertex set is fixed-parameter tractable. In *Proceedings of the 39th International Colloquium on Automata, Languages, and Programming*, ICALP '12, pages 230–241, 2012. 2

[13] F. A. Chudak, M. X. Goemans, D. S. Hochbaum, and D. P. Williamson. A primal-dual interpretation of two 2-approximation algorithms for the feedback vertex set problem in undirected graphs. *Operations Research Letters*, 22(45):111–118, 1998. 2, 3

[14] J. Chuzhoy, S. Guha, E. Halperin, S. Khanna, G. Kortsarz, R. Krauthgamer, and J. Naor. Asymmetric k-center is log* n-hard to approximate. *J. ACM*, 52(4):538–551, 2005. 8

[15] J. Chuzhoy, V. Guruswami, S. Khanna, and K. Talwar. Hardness of routing with congestion in directed graphs. In *Proceedings of the 39th annual ACM Symposium on Theory of Computing*, STOC '07, pages 165–178, 2007. 6

[16] J. Chuzhoy and S. Khanna. Polynomial flow-cut gaps and hardness of directed cut problems. *J. ACM*, 56(2), 2009. 6

[17] M. Cygan. Improved approximation for 3-dimensional matching via bounded pathwidth local search. In *Proceedings of the 54th annual IEEE symposium on Foundations of Computer Science*, FOCS '13, pages 509–518, 2013. 4

[18] M. Cygan, M. Pilipczuk, M. Pilipczuk, and J. O. Wojtaszczyk. Subset feedback vertex set is fixed-parameter tractable. In *Proceedings of the 38th International Colloquim on Automata, Languages and Programming*, ICALP '11, pages 449–461, 2011. 2

[19] I. Dinur, V. Guruswami, S. Khot, and O. Regev. A new multilayered PCP and the hardness of hypergraph vertex cover. *SIAM J. Comput.*, 34(5):1129–1146, 2005. 2, 8, 19, 24

[20] I. Dinur and S. Safra. On the hardness of approximating minimum vertex cover. *Annals of Mathematics*, pages 439–485, 2005. 2

[21] G. Even, J. Naor, S. Rao, and B. Schieber. Divide-and-conquer approximation algorithms via spreading metrics. *J. ACM*, 47(4):585–616, 2000. 2

[22] G. Even, J. Naor, B. Schieber, and M. Sudan. Approximating minimum feedback sets and multicuts in directed graphs. *Algorithmica*, 20:151–174, 1998. 2, 6

[23] Z. Friggstad and M. R. Salavatipour. Approximability of packing disjoint cycles. In *Proceedings of the 18th International conference on Algorithms and Computation*, ISAAC'07, pages 304–315, 2007. 4

[24] V. Guruswami, R. Manokaran, and P. Raghavendra. Beating the random ordering is hard: Inapproximability of maximum acyclic subgraph. In *Proceedings of the 49th annual IEEE symposium on Foundations of Computer Science*, FOCS '08, pages 573–582, 2008. 2, 3

[25] V. Guruswami, S. Sachdeva, and R. Saket. Inapproximability of minimum vertex cover on k-uniform k-partite hypergraphs. *Electronic Colloquium on Computational Complexity (ECCC)*, 20:71, 2013. 15, 18

[26] E. Hazan, S. Safra, and O. Schwartz. On the complexity of approximating k-set packing. *Computational Complexity*, 15(1):20–39, 2006. 4

[27] S. Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the 34th annual ACM Symposium on Theory of Computing*, STOC '02, pages 767–775, 2002. 2, 34

[28] S. Khot. On the unique games conjecture (invited survey). In *Proceedings of the 25th annual IEEE Conference on Computational Complexity*, CCC '10, pages 99–121. IEEE, 2010. 2

[29] S. Khot and O. Regev. Vertex cover might be hard to approximate to within $2-\epsilon$. *J. Comput. Syst. Sci.*, 74(3):335–349, 2008. 2, 4, 34

[30] G. Kortsarz, M. Langberg, and Z. Nutov. Approximating maximum subgraphs without short cycles. *SIAM J. Discrete Math.*, 24(1):255–269, 2010. 3, 18

[31] M. Krivelevich. On a conjecture of Tuza about packing and covering of triangles. *Discrete Mathematics*, 142(1-3):281–286, 1995. 3

[32] M. Krivelevich, Z. Nutov, M. R. Salavatipour, J. V. Yuster, and R. Yuster. Approximation algorithms and hardness results for cycle packing problems. *ACM Trans. Algorithms*, 3(4), Nov. 2007. 4

[33] E. Mossel. Gaussian bounds for noise correlation of functions. *Geometric and Functional Analysis*, 19(6):1713–1756, 2010. 3, 31, 33, 34

[34] S. Sachdeva and R. Saket. Optimal inapproximability for scheduling problems via structural hardness for hypergraph vertex cover. In *Proceedings of the 28th annual IEEE Conference on Computational Complexity*, CCC '13, pages 219–229, 2013. 21

[35] P. D. Seymour. Packing directed circuits fractionally. *Combinatorica*, 15(2):281–288, 1995. 2

[36] O. Svensson. Hardness of vertex deletion and project scheduling. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, volume 7408 of *Lecture Notes in Computer Science*, pages 301–312. 2012. 2, 3, 4, 31

A Reduction from Vertex k-Cycle Transversal to Edge Version

We present simple reductions from Vertex k-Cycle Transversal to *Edge* r-Cycle Transversal where the goal is to pick a set of edges to intersect all cycles of length (up to) r. There are several natural versions of Edge r-Cycle Transversal, depending on

- Whether the graph is *directed* or *undirected*
- Whether r is *odd* or *even*
- Whether we want to hit every cycle of length *exactly* r or *up to* r.

To distinguish variants, let Edge ($\leqslant r$)-Cycle Transversal be the version where we want to intersect all cycles of length up to r, and Edge ($= r$)-Cycle Transversal to be the version we want to intersect all cycles of length exactly r.

Theorem A.1. *Unless* $\mathsf{NP} \not\subseteq \mathsf{BPP}$*, there is no polynomial time algorithm that does any the following tasks, whether the graph is directed or not.*

- *For* $r \geqslant 6$*, approximate Edge* $(\leqslant r)$*-Cycle Transveral within a factor of* $\lfloor \frac{r}{2} \rfloor - 1$.
- *For even* $r \geqslant 6$*, approximate Edge* $(= r)$*-Cycle Transveral within a factor of* $\lfloor \frac{r}{2} \rfloor - 1$.
- *For odd* $r \geqslant 9$*, approximate Edge* $(= r)$*-Cycle Transveral within a factor of* $\frac{\lfloor \frac{r}{2} \rfloor}{2} - 1 + \frac{1}{r-1}$.

The only case we obtain the worse $(\frac{\lfloor \frac{r}{2} \rfloor}{2} - 1 + \frac{1}{r-1})$-hardness, Edge $(= r)$-Cycle Transversal with odd r has a $(r-1)$-approximation algorithm [30] in undirected graphs. This is better than the trivial r-approximation algorithm and is not known in other variants. Although the results in directed graphs and undirected graphs are the same, we present them in different subsections due to slightly different techniques involved.

A.1 Directed Graphs

Given a directed graph $G = (V_G, E_G)$ as an instance of Vertex k-Cycle Transversal, our basic reduction to the edge version is simple. *Split* every vertex $v \in V_G$ into two vertices v_{in}, v_{out}, and an edge (v_{in}, v_{out}) (call it a *vertex-edge*). Replace every edge $(u, v) \in E_G$ by (u_{out}, v_{in}). Let G' be the resulting graph as an instance of Edge $2k$-Cycle Transversal. Note that there is one-to-one correspondence between cycles of G of length l and cycles of length $2l$ in G'. In partiulcar, the length of every cycle of G' is even. Also note that our hardness result for Vertex k-Cycle Transversal is strong in the sense that in the completeness case, a small fraction of vertices hit every cycle of length *up to* k, while in the soundenss case, every small fraction of vertices must contain a cycle of length *exactly* k.

For completeness, if removing vertices in $C \subseteq V_G$ gets rid of every cycle of length up to k in G, then removing the same number of edges (vertex-edges corresponding to C) gets rid of every cycle of length up to $2k$ in G'.

For soundness, given $C' \subseteq E_{G'}$ such that the subgraph $(V_{G'}, E_{G'} \setminus C')$ does not have a cycle of length exactly $2k$, we can easily convert it into $C \subseteq V_G$ with the same cardinality such that the induced subgraph of G on $V_G \setminus C$ does not have a cycle of length exactly k. For $(v_{in}, v_{out}) \in C'$ put v in C, and for (u_{out}, v_{in}) pick either u or v. This works since for any cycle of G $(v_1, \dots, v_l)$ of length l, $(v_{1,in}, v_{1,out}, \dots, v_{l,in}, v_{l,out})$ is a cycle of length $2l$ in G', and C' must contain $(v_{i,in}, v_{i,out})$ or $(v_{i,out}, v_{(i+1),in})$ for some i.

The above reduction shows that if we want to hit every cycle of length up to r, we have a factor of $\lfloor \frac{r}{2} \rfloor - 1$ hardness from Vertex k-CT, whether r is even or odd. If we want to hit every cycle of length exactly r, the above reduction works only for even r to give a $\frac{r}{2} - 1$-factor hardness. This proves the first two items of Theorem A.1 in directed graphs.

When r is odd, the above reduction requires a hardness result for the vertex version from a more structured Hypergraph Vertex Cover. In Vertex Cover in k-uniform k-partite hypergraphs, the set of verticies V is partitioned into k parts $V_1, \dots, V_k$ (which are given) and each hyperedge contains exactly one vertex from V_i. This problem is also NP-hard to approximate within a factor of $\frac{k}{2} - 1 + \frac{1}{2k}$ [25]. For each hyperedge $e = (v^1, \dots, v^k)$ such that $v^i \in V_i$, we add a canonical cycle $(v^i, v^{(i+1)})$ (previously we allowed any ordering of vertices, but here we require each edge go from V_i to $V_{(i+1)}$). An analogue to Theorem 1.1 will work to give the same factor of hardness for the vertex version.

Given a hardness result for the vertex version from k-uniform k-partite graphs, our final reduction to the edge version is to split vertices in $V_1, \ldots, V_{k-1}$ as usual (two vertices and one edge), and split each vertex $v \in V_k$ as v_{in}, v_{mid}, v_{out} and an edge $(v_{in}, v_{mid}), (v_{mid}, v_{out})$ (two vertex-edges per one vertex). By similar arguments to the above, there is one-to-one correspondence of cycles of G of length l and cycles of G' of length $2l + 1$. Therefore, we also get a $(\frac{\lfloor \frac{r}{2} \rfloor}{2} - 1 + \frac{1}{r-1})$-factor hardness for Edge $(= r)$-CT for odd r, completing the proof of Theorem A.1 in directed graphs.

A.2 Undirected Graphs

Given an undirected graph $G = (V_G, E_G)$ as an instance of Vertex k-Cycle Transversal, the basic reduction to G' as an instance of the edge version is almost identical to that of directed graphs. Split each vertex $v \in V_G$ into v_{in} and v_{out}, and a vertex-edge (v_{in}, v_{out}). For each canonical cycle $(v_1, \ldots, v_k)$ of G, fix the orientation (which is already fixed in directed graphs), and add $(v_{i,out}, v_{(i+1),in})$. Call $(v_{1,in}, v_{1,out}, \ldots, v_{k,in}, v_{k,out})$ canonical in G'.

For the completeness analysis, if G has a small set of vertices C intersecting every canonical cycle of G, then vertex-edges corresponding to C intersect every canonical cycle of G'. The proof of completeness in Theorem 1.1 essentially shows that the number of non-canonical cycles in G' of length up to $2k$ is $o(|V_{G'}|)$. The only difference in the analysis is that we might additionally have a cycle like $(\ldots, u_{in}, v_{out}, w_{in}, \ldots)$ where u, v, w are different, but the fact that this cycle cannot covered by a single canonical edge means that the expected number of these cycles is $o(|V_{G'}|)$. Hitting each non-canonical cycle by one edge, we can conclude that if G has a vertex k-cycle transversal of size m, G' has an edge $2k$-cycle transversal of size $m + o(|V_{G'}|)$.

The soundness analysis works as in directed graphs. From this reduction we prove the first two items of Theorem A.1 in undirected graphs.

To deal with the odd r, we use the same hardness of the vertex version from k-HVC on k-partite graphs as in directed graphs to show that Edge $(= r)$-CT is also hard to approximate within a factor of $\frac{\lfloor \frac{r}{2} \rfloor}{2} - 1 + \frac{1}{r-1}$ in undirected graphs, completing the proof of Theorem A.1.

B NP-hardness of k-HVC with density

In this section, we show that the best known NP-hardness result for Ek-HVC [19] yields an unweighted instance with completeness and soundness parameters we want (depending on k and $\epsilon(k)$), proving Theorem 3.1.

Their Theorem 4.1 requires a multi-layered PCP with parameters l (number of layers) and R (number of labels), which both depend on k and ϵ. Note that in the original Raz verifier, the degree d_R is a function of R. Given a Raz verifier which consists of a bipartite graph $G = (V_G, E_G)$ such that $V_G = Y \cup Z$, Theorem 3.3 yields a multilayered PCP where variables of layer i are of the form $(z_1, \ldots, z_i, y_{i+1}, \ldots, y_l)$ where $z_j \in Z$ and $y_j \in Y$. The number of labels for any vertex is bounded by R^l. For $i < j$, there exists a constraint between $(z_1, \ldots, z_i, y_{i+1}, \ldots, y_l)$ and $(z'_1, \ldots, z'_j, y'_{j+1}, \ldots, y'_l)$ if and only if

- $z_q = z'_q$ where $q \leqslant i$.
- $y_q = y'_q$ where $q > j$.
- $(y_q, z'_q) \in E_G$ for $i < q \leqslant j$.

Therefore, the degree is at most $l(d_R)^l$, which is still a function of k and $\epsilon(k)$. After the reduction from a multilayered PCP to a weighted hypergraph, the degree of each vertex is still bounded by a function of k and $\epsilon(k)$, since each variable of the PCP is replaced by at most 2^{R^l} vertices and each PCP constraint is replaced by at most 2^{kR^l} hyperedges.

Given such a weighted instance, we convert it to an unweighted instance by duplicating vertices according to their weights. The weight of each vertex in the ith layer is of the form

$$\frac{1}{l|X_i|} p^r (1-p)^{R_i - r}$$

where $X_i = |Z_i|^i |Y_i|^{l-i}$ is the set of vertices in the ith layer, $R_i = R^{O(l)}$ is the number of labels in ith layer, $p = 1 - \frac{1}{k-1-\epsilon}$, and $0 \leqslant r \leqslant R_i$. The original paper set the weight as above so that the sum of weights becomes 1. Multiply weight of each vertex by $|Y_i|^l$ so that the weight of each vertex in the ith layer is of the form

$$\frac{1}{l} \left(\frac{|Y_i|}{|Z_i|} \right)^i p^r (1-p)^{R_i - r}$$

Let α be a rational that divides both p and $1-p$ with both quotients bounded. Then α^{R^l} divides any $p^r (1-p)^{R_i - r}$ as well with quotient bounded by a function of ϵ and k. Therefore, if we set the minimum weight to be

$$\frac{1}{l} \cdot \frac{|Y_i|}{|Z_i|} \cdot \alpha$$

the weight of each vertex must be divisible by the minimum weight, and the quotient will be bounded by a function of k and ϵ. We replace each weighted vertex by (weight / minimum weight) number of unweighted vertices, and for each hyperedge $(v^1, \ldots, v^k)$, add all hyperedges $(u^1, \ldots, u^k)$ where u^i is a copy v^i. Since each quotient and the original degree of the weighted instance are bounded by a function of k and ϵ, so is the degree of the unweighted instance.

Now we have an unweighted problem with completeness c, soundness s, and degree bounded by d. Let $\delta = 2(1-s)$. Suppose in soundness case, we have $1-\delta$ fraction of vertices cover more than $1 - \frac{k(1-s)}{d}$ fraction of hyperedges. Cover the remaining hyperedges with one vertex each. Since $|E_H| \leqslant \frac{d}{k}|V_H|$, this process requires less than $\frac{k(1-s)}{d} \cdot \frac{d}{k} = 1 - s$ fraction of vertices, and we have a vertex cover of measure less than $1 - \delta + (1-s) = s$. This contradicts the original soundness, so any $\delta := 2(1-s)$ fraction of vertices should contain at least $\rho := \frac{k\delta}{2d}$ fraction of edges, both depending only on k and ϵ.

C k-Clique Transversal

In this section, our goal is to prove Theorem 1.3 showing a near-optimal hardness result for k-Clique Transversal, the problem of hitting k-cliques in an undirected graph.

If we assume the UGC, the following structured hardness result of Bansal and Khot [6] for Ek-HVC implies a tight hardness result for k-Clique Transversal.

Theorem C.1 ([6]). *Fix an integer $k \geqslant 3$ and $\epsilon \in (0,1)$. Given a k-uniform graph $H = (V_H, E_H)$, assuming the UGC, there is no polynomial time algorithm that distinguishes the following cases.*

- *Completeness: There exist disjoint subsets $V_1, \ldots, V_k \subseteq V_H$, each with $\frac{1-\epsilon}{k}$ fraction of vertices, such that each hyperedge has at most one vertex in each V_i.*
- *Soundness: Every subset C with less than $1-\epsilon$ fraction of vertices does not intersect at least one hyperedge. Equivalently, every subset of ϵ fraction of vertices wholly contains a hyperedge.*

Replacing each hyperedge by a k-clique (without duplicating vertices) gives an easy reduction to k-HVC to k-Clique Transversal. In completeness case, taking one V_i and $V \setminus (V_1 \cup \cdots \cup V_k)$ gives a transversal of measure at most $1/k + \epsilon$. Soundness immediately follows from that of k-HVC.

The above reduction crucially uses the fact that in completeness case, each V_i (as a subset of the graph after reduction) is still an independent set. A similar but weaker statement can be obtained without relying on the UGC [34]. However, this result does not have the above property, so at least as a black-box, cannot be applied to prove hardness of k-Clique Transversal.

Our approach will be to be reduce a general instance of Ek-HVC to k-Clique Transversal via the random matching technique. Fix k and let $c := c(k), s := s(k), d := d(k)$ be the parameters we have from Theorem 3.1 (we do not need density). Let a and B be integer constants greater than 1, which will be determined later. Lemmas C.2 and C.4 with these parameters imply Theorem 1.3.

C.1 Reduction

The idea is similar to k-Cycle Transversal — instead of adding canonical cycles, we create canonical cliques. Given a hypergraph $H = (V_H, E_H)$, construct an undirected graph $G = (V_G, E_G)$ such that

- $V_G = V_H \times [B]$. Let $n = |V_H|$ and $N = |V_G| = nB$.
- For each edge $e = (v^1, \ldots, v^k)$, for aB times, take $(l^1, \ldots, l^k)$ independently and uniformly from $[B]$. Add a k-clique between (v^i, l^i)'s. Call such cliques *canonical*.

C.2 Completeness

Lemma C.2. *Suppose H has a vertex cover C of measure c. For any $\epsilon > 0$, with probability at least $3/4$, there exists a subset $F \subseteq G$ of measure at most $c + \epsilon$ such that the induced subgraph $G \setminus F$ has no k-clique.*

Proof. Let $F = C \times [B]$. It is easy to see that every canonical clique intersects F. We consider the number of non-canonical cliques and argue that a small fraction of additional vertices intersect all of these cliques. Choose k vertices $(v^1, l^1), \ldots, (v^k, l^k)$ which satisfy

- $v^1 \in V_H$ can be any vertex.
- $l^1, \ldots, l^k \in B$ can be arbitrary labels.
- For each $1 \leqslant i < k$, there must be a hyperedge e that contains both v^i and v^{i+1}.

The number of possibilities to choose such $(v^1, l^1), \ldots, (v^k, l^k)$ is bounded by $n(kd)^{k-1}B^k$. Note that no other k-tuple of vertices can form a clique. Further discard the tuple when two vertices are the same.

We calculate the probability that $(v^1, l^1), \ldots, (v^k, l^k)$ form a clique that does not intersect F. To form a clique, $\binom{k}{2}$ edges must exist. For $2 \leqslant p \leqslant k$, partition $\binom{k}{2}$ purported edges into p *groups* $I_1, \ldots, I_p$. There are at most $p^{\binom{k}{2}}$ possibilities. For each group, there are at most aBd copies of canonical cliques that can possibly contain all the edges in that group. There are at most $(aBd)^p$ possibilities. Assume that one canonical clique is responsible for exactly one group. This is without loss of generality since if one canonical clique is responsible for many groups, we can merge them and this case can be dealt with smaller p.

Focus on one group (say qth) of purported edges, and one canonical clique which is supposed to contain them. Let I'_q be the set of vertices incident on edges in I_q. Without loss of generality, $v^i \neq v^j$ for all $(v^i, l^i), (v^j, l^j) \in I'_q$, since one canonical clique assigns at most one label to each vertex of H.

Lemma C.3. *For each* $2 \leqslant p \leqslant k$,

$$\sum_{q=1}^{p} |I'_q| \geqslant k + p \ .$$

Proof. Since $|I'_q| \geqslant 2$ for each q, the lemma holds when $p \geqslant k$. Suppose $p < k$. For each vertex (v^i, l^i) $(1 \leqslant i \leqslant k)$, let t_i be the number of q's such that $i \in I'_q$. If $t_i \geqslant 2$ for every i,

$$\sum_q |I'_q| = \sum_i t_i \geqslant 2k \geqslant k + p \ .$$

Further suppose there exists i such that $t_i = 1$. It means that every edge incident on (v^i, l^i) must belong to the same group q'. Therefore $|I'_{q'}| = k$ and

$$\sum_q |I'_q| \geqslant k + 2(p-1) \geqslant k + p \ . \qquad \square$$

We can conclude that for each partition, the probability of having all the edges is at most $\frac{(aBd)^p}{B^{k+p}} = \frac{(ad)^p}{B^k}$. The probability that $(v^1, l^1), \ldots, (v^k, l^k)$ form a clique is bounded by

$$\sum_{p=2}^{k} p^{\binom{k}{2}} \frac{(ad)^p}{B^k} \leqslant k^{k^2} \left(\frac{ad}{B}\right)^k$$

Therefore, the expected number of non-canonical k-cliques is bounded by $n(kd)^{k-1}B^k \cdot k^{k^2}(\frac{ad}{B})^k \leqslant nR$, where R is a constant depending only on k. With probability at least $3/4$, the number of non-canonical k-cliques is at most $4nR$. Let $B \geqslant \frac{4R}{\epsilon}$. Then these cliques can be hit by at most $\epsilon nB = \epsilon N$ vertices. $\square$

C.3 Soundness

The soundness is very similar to the argument used for cycle transversal (Lemma 4.3).

Lemma C.4. *If every subset of V_H of measure at least ϵ contains a hyperedge in the induced subgraph, with probability at least $3/4$, every subset of V_G with measure 2ϵ contains a canonical k-clique.*

Proof. For each vertex $v \in V_H$, let $\mathsf{cloud}(v) := v \times [B]$ be the set of vertices in V_G that corresponds to v. Fix one hyperedge $e = (v^1, \ldots, v^k)$. We want to ensure that if a subset of vertices I has at least ϵ fraction from each $\mathsf{cloud}(v^i)$, then I will contain a canonical clique. Fix $A^1 \subseteq \mathsf{cloud}(v^1), \ldots, A^k \subseteq \mathsf{cloud}(v^k)$ be such that for each i, $|A^i| \geqslant \epsilon B$. There are at most 2^{kB} ways to choose such A's. The probability that one canonical clique associated with e is not contained in $(v^1, A^1) \times \ldots \times (v^k, A^k)$ is at most $1 - \epsilon^k$. The probability that none of canonical clique associated with e is contained in $(v^1, A^1) \times \ldots \times (v^k, A^k)$ is $(1-\epsilon^k)^{aB} \leqslant exp(-aB\epsilon^k)$.

By union bound over all $A^1, \ldots, A^k$, the probability of the bad event is at most $\exp(kB - aB\delta^k) = exp(-B) \leqslant \frac{1}{4|E_H|}$ by taking a large enough constant depending on k and ϵ, and $L = \Omega(\log |E_H|)$. Therefore, with probability at least $3/4$, the desired property holds for all hyperedges.

Let I be a subset of V_G of measure at least 2ϵ. By an averaging argument, at least ϵ fraction of *good* vertices $v \in V_H$ satisfy that $\mathsf{cloud}(v^i) \cap I \geqslant \epsilon$. By the soundness property of H, there is a hyperedge contained in the subgraph induced by the good vertices. $\square$

D k-Cycle Packing and k-Clique Packing

We now turn to the problem of packing vertex-disjoint cliques of size *exactly* k, and packing vertex-disjoint cycles of length at most k in an undirected graph. The starting point for our reductions will be the independent set problem on bounded-degree graphs, so we begin with a discussion of the best known hardness results for this problem.

D.1 Hardness of MIS-k

A degree-k graph is a graph with maximum degree at most k. The MIS-k denotes the maximum independent set problem restricted to degree-k graphs. We begin by stating the known hardness results for MIS-k. We state both NP- and UG-hardness results below since the UG-hardness result is slightly stronger (by a factor of $\log^2 k$) and has other useful properties that are exploited in some our reductions.

Theorem D.1 ([10]). *For every sufficiently large k, given an instance of MIS-k it is NP-hard to distinguish the following cases.*

- *Completeness: There is an independent set of measure $\Omega(1/\log k)$.*
- *Soundness: Every independent set has measure at most $O(\log^3 k/k)$.*

As a result, it is NP-hard to approximate MIS-k within a factor of $\Omega(\frac{k}{\log^4 k})$.

An important property that we exploit in the UG-hardness result is density in the soundness case. In Theorem 1.2, we also used the density argument implied by the NP-hardness of [19]. However, here we need a stronger density such that the graph almost looks like a random graph in the soundness case, and such a strong density is known only assuming the UGC. Following the notation of [4], call a graph G (β, α)-dense if for every $S \subseteq V_G$ with measure at least β, the total fraction of edges inside S is at least α (α might be a function of β).

Theorem D.2 ([4]). *For every sufficiently large k, given a degree-k graph, it is Unique Games-hard (under randomized reductions) to distinguish the following cases.*

- *Completeness: There is an independent set of measure $\Omega(1/\log k)$.*
- *Soundness: The graph is $(\beta, \alpha(\beta))$-dense for $\beta \geqslant \beta_0 := \Theta(\log k/k)$ and $\alpha(\beta) = \Omega(\beta^2)$.*

As a result, it is UG-hard (under randomized reductions) to approximate MIS-k within a factor of $\Omega(\frac{k}{\log^2 k})$.

Because this is slightly stronger than the statement stated in [4] (they do not prove the density property in the final low-degree graph), we prove how their result implies this stronger statement.

Proof. Let $\rho := \Theta(-\frac{1}{\log k})$, $\epsilon := \Gamma_\rho(\beta_0)/2$, where Γ_ρ is defined by

$$\Gamma_\rho(\mu) = \Pr[X \leqslant \Phi^{-1}(\mu) \wedge Y \leqslant \Phi^{-1}(\mu)]$$

where X are jointly normal variables with mean 0 and covariance matrix $\left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$. After the reduction from Unique Games and removing weights (Section 3 and Step 1 of Section 4 of [4]), we have a graph G_1 (regular unweighted graph with large degree) such that

- In the completeness case, it has an independent set of measure $\Omega(1/\log k)$.
- In the soundness case, G_1 is $(\beta, \alpha(\beta))$-dense for $\beta \geqslant \beta_0$ and $\alpha(\beta) = \Omega(\beta^2)$.

To make the soundness property holds for every $\beta \geqslant \beta_0$, we used the fact that the reduction from Unique Games in [4] shows that G_1 is $(\beta, \Gamma_\rho(\beta) - \epsilon)$-dense in the soundness case (so that $\Gamma_\rho(\beta) - \epsilon \geqslant \Gamma_\rho(\beta)/2$), and that $\Gamma_\rho(\beta) = \Theta(\beta^2)$ for $\beta \geqslant \beta_0$ and the choice $\rho = \Theta(-\frac{1}{\log k})$ we made.

Fix $\beta \geqslant \beta_0$ and a set S of measure β. Their sparsification step picks $kn/4$ edges of G_1 at random. Since there are at least $\alpha := \alpha(\beta)$ fraction of edges in the subgraph induced by S, the expected number of picked edges in this subgraph is at least $\frac{\alpha kn}{4}$. By Chernoff bound, the probability that it is less than $\frac{\alpha kn}{8}$ is at most $\exp(-\frac{\alpha kn}{32})$. By union bound over all sets of measure exactly β (there are at most $\binom{n}{n\beta} \leqslant \exp(2\beta\log(1/\beta)n)$ of them), and over all possible values of β (there are at most n possible sizes), the desired property fails with probability at most

$$n \cdot \max_{\beta\in[\beta_0,1]} \left\{\exp(-\alpha kn/32) \cdot \exp(2\beta\log(1/\beta)n)\right\} \leqslant n \cdot e^{-n}$$

since $k \geqslant \Omega(\frac{\beta\log(1/\beta)}{\alpha})$ holds for our choice of $k = \Omega(\frac{\log(1/\beta_0)}{\beta_0})$. Note that the degree here is increased from the degree in the original paper by a (absolute) constant factor. In the last step where we throw all the edges incident on vertices of degree more than k (average degree is $k/2$), the density property still holds since with constant probability, the fraction of edges thrown out is very small compared to $\Omega(\beta_0^2)$. $\square$

D.2 k-Clique Packing

We prove hardness of k-Clique Packing by a reduction from MIS-k. Given a degree-k graph $G = (V_G, E_G)$, our basic reduction is to produce a *line graph* $L(G) = (V_{L(G)}, E_{L(G)})$ such that $V_{L(G)} = E_G$ and there is an edge between $e, f \in V_{L(G)}$ if and only if they share an endpoint (as edges of G). The following lemma gives a simple reduction from MIS-k to k-Clique Packing. Together with Theorem D.1 and D.2, it proves Theorem 1.4.

Lemma D.3. *For $k \geqslant 4$, there is a reduction from MIS-k to k-Clique Packing such that an instance of MIS-k has an independent set of cardinality p if and only if the reduced instance of k-Clique Packing has p disjoint k-cliques.*

Proof. Given an instance G of MIS-k, produce its line graph $L := L(G)$. For each vertex $v \in V_G$ such that $\mathsf{deg}(v) < k$, we add $k - \mathsf{deg}(v)$ vertices to V_L, make them connected one another, and also to the edges (of G) incident on v. Each added vertex has degree exactly $k - 1$. Now each $v \in V_G$ has exactly k vertices of L which are either edges (of G) incident on v or newly added by v, and they form a k-clique in L. Let $\mathsf{star}(v)$ denote these k vertices. If $e \in V_L$ is an original edge of G, it belongs to two stars. Otherwise, it belongs to only one star.

Therefore, $\mathsf{star}(v)$ and $\mathsf{star}(u)$ intersect if and only if $(u, v) \in E_G$. Suppose there exists k-clique in L. If it contains a newly added vertex, then the clique has to be $\mathsf{star}(v)$ for some $v \in V$, since the newly added vertex has neighbors in only one star. Otherwise, all of k vertices correspond to original edges. We use a basic fact that when if G is simple and $k \geqslant 4$, the only way that k edges of G form a clique in $L(G)$ is that they all share the same endpoint. In any case, any k-clique of L is $\mathsf{star}(v)$ for some $v \in V$.

From the above facts, it is clear that G has an independent set I of cardinality p if and only if $P = \{\mathsf{star}(v) \mid v \in I\}$ is a set of p disjoint k-cliques in L. $\square$

D.3 k-Cycle Packing

In this subsection, we show hardness of k-Cycle Packing by a reduction from MIS-k, towards the goal of proving Theorem 1.5. However, the reduction is not as simple as the one we used for k-Clique Packing. A line graph has a nice property that it does not have a big clique except those formed by the edges incident on the same vertex, even though the original graph has a big clique. This nice property is not preserved when we consider cycles; when $(v^1, \dots, v^k)$

forms a cycle in G, $((v^1, v^2), \ldots, (v^k, v^1))$ forms a cycle in $L(G)$ as well. To circumvent these issues, large girth is a natural requirement to impose on the hard instances of MIS-k.

Definition D.4. *MIS-k-l is the problem of finding the maximum independent set on degree-k graphs that have girth* strictly greater than l.

If MIS-k-k is also hard to approximate, it is easy to show hardness of k-Cycle Packing, as shown in Lemma D.5 below. In Section D.4 (resp. Section D.5), we will prove an NP-hardness (resp. UG-hardness) result for MIS-k-k, which together with the below lemma will yield Theorem 1.5.

Lemma D.5. *There is a polynomial time approximation-ratio preserving reduction from MIS-k-k to k-Cycle Packing.*

Proof. Given a degree-k graph G with girth strictly greater than k, we construct a variant of the line graph $L' := L'(G) = (V_{L'}, E_{L'})$ where

- $V_{L'} = E_G$.
- For each vertex $v \in V(G)$, order the edges $e^1, \ldots, e^{k'}$ incident on it arbitrarily ($k' \leqslant k$), and add k' edges $((e^1, e^2)), \ldots, (e^{k'}, e^1))$ to L'. Call this cycle *canonical cycle*, and let v be *pivot* for all these edges.
- The above construction does not work properly for vertices with degree $0, 1$, or 2, if we want L' to be a simple graph. For these vertices,
  - $\mathsf{deg}(v) = 0$: Add a isolated triangle to L'.
  - $\mathsf{deg}(v) = 1$ with incident edge e: Add a new vertex f, g to $V_{L'}$ and add a triangle between e, f, g.
  - $\mathsf{deg}(v) = 2$ with incident edge e, f: Add a new vertex g to $V_{L'}$ and add a triangle between e, f, g.
  - We also call each triangle *canonical cycle* and let v be the pivot for these edges.

We argue that there is no non-canonical cycle in L' of length at most k. Suppose that $(e^1, \ldots, e^{k'})$ forms a simple cycle of L'. For each $i \in [k']$, let v^i be the unique pivot (shared endpoint) of $(e^i, e^{(i+1)})$. Also, if e^i and e^j share an endpoint v, the edge (e^i, e^j) has to have pivot v, or does not exist. There are three cases to consider.

- $v^1 = \cdots = v^{k'}$: Since the set of edges with pivot v form a simple cycle of length $\max(3, \mathsf{deg}(v))$, it means that $e^1, \ldots, e^{k'}$ are all the edges incident on v (plus newly created edges if $\mathsf{deg}(v) < 3$), and $(e^1, \ldots, e^{k'})$ is a canonical cycle.
- $\left|\{v^1, \ldots, v^{k'}\}\right| = 2$: Let u, w be the two pivots. There are at least two i's such that $v^i \neq v^{(i+1)}$. For those i's, $e^{(i+1)}$ used both v^i and $v^{(i+1)}$ as pivots, so it is the unique edge (u, w). This contradicts the assumption that $(e^1, \ldots, e^{k'})$ is simple.
- $\left|\{v^1, \ldots, v^{k'}\}\right| \geqslant 3$: *Compress* the sequence $(v^1, \ldots, v^{k'})$ so that

– While $\exists i$ such that $v^i = v^{(i+1)}$ (here, $(i+1) \equiv i+1$ mod (current length of sequence)),
  * Delete $v^{(i+1)}$ from the sequence (so that the length is decreased by 1).
– After compression,
  * The length of sequence is at least 3, and at most k'.
  * There is an edge between v^i and $v^{(i+1)}$.
  * Since the original cycle $(e^1, \ldots, e^{k'})$ is simple, none of edge $(v^i, v^{(i+1)})$ is used more than once.
– Therefore, we have a closed walk in G of length k'. As the girth of G is greater than k, we must have $k' > k$.

Therefore, all cycles of length at most k are canonical cycles. If G has an independent set I, L' has a k-cycle packing P of the same size by simply taking all the canonical cycles whose pivot is in I. Conversely, if L' has a k-cycle packing P, all cycles has to be canonical, and we can simply take their pivots as an independent set. □

D.4 NP-hardness of independent set on bounded-degree high-girth graphs

Finally, we prove that MIS-k-k is also hard to approximate. Given a k-regular graph with possibly small girth, we replace each vertex by a cloud of B vertices, and replace each edge by a a copies of random matching between the two clouds. While maintaining the soundness guarantee, we show that there are only a few small cycles, and by deleting a vertex from each of them we obtain a hard instance for MIS-k-k. The below lemma, together with Lemma D.5, proves the first half (NP-hardness) of Theorem 1.5.

Lemma D.6. *It is NP-hard to approximate MIS-k-k within a factor of $\Omega(\frac{\sqrt{k}}{\log^3 k})$.*

Proof. Given an instance $G_0 = (V_{G_0}, E_{G_0})$ of MIS-d, we construct $G = (V_G, E_G)$ and $G' = (V_{G'}, E_{G'})$ by the following procedure:

- $V_G = V_{G_0} \times [B]$. As usual, let $\mathsf{cloud}(v) = \{v\} \times [B]$.
- For each edge $(u, v) \in E_{G_0}$, for a times, add a *random matching* as follows.
  – Take a random permutation $\pi : [B] \to [B]$.
  – Add an edge $((u, i), (v, \pi(i))$ for all $i \in [B]$.
- Call the resulting graph G. To get the final graph G',
  – For any cycle of length at most ad, delete an arbitrary vertex from the cycle. Repeat until there is no cycle of length at most ad.

Let $n = |V_{G_0}|, m = |E_{G_0}|, N = nB = |V_G| \geqslant |V_{G'}|, M = m \cdot aB = |E_G| \geqslant |E_{G'}|$. The maximum degree of G and G' is at most ad. By construction, girth of G' is at least $ad + 1$.

Girth Control. We calculate the expected number of small cycles in G, and argue that the number of these cycles is much smaller than the total number of vertices, so that $|V_G|$ and $|V_{G'}|$ are almost the same. Let k' be the length of a purported cycle. Choose k' vertices $(v^1, l^1), \ldots, (v^{k'}, l^{k'})$ which satisfy

- $v^1 \in V_{G_0}$ can be any vertex.
- For each $1 \leqslant i < k'$, $(v^i, v^{i+1}) \in E_{G_0}$.
- $l^1, \ldots, l^k \in B$ can be arbitrary labels.

There are n possible choices for v^1, B choices for each l^i, and d choices for each v^i ($i > 1$). The number of possibilities to choose such $(v^1, l^1), \ldots, (v^{k'}, l^{k'})$ is bounded by $nd^{k'-1}B^{k'}$. Without loss of generality, assume that no vertices appear more than once.

For each edge $e = (u, w) \in G_0$, consider the purported cycle induced by $\mathsf{cloud}(u) \cup \mathsf{cloud}(w)$. It is a bipartite graph with the maximum degree 2. Suppose there are q purported edges $e^1, \ldots, e^q$ (ordered arbitrarily). By slightly abusing notation, let e^i also denote the event that e^i exists in G. For each e^i, we upper bound $\Pr[e^i | e^1, \ldots, e^{i-1}]$.

Claim D.7. $\Pr[e^i | e^1, \ldots, e^{i-1}] \leqslant \frac{a}{B-i}$.

Proof. There are a random matchings between $\mathsf{cloud}(u)$ and $\mathsf{cloud}(w)$, and for each $j < i$, there is at least one random matching including e^j. We fix one random matching and calculate the probability that the random matching contains e^i, conditioned on the fact that it already contains some of $e^1, \ldots, e^{i-1}$.

If there is e^j ($j < i$) that shares a vertex with e^i, e^i cannot be covered by the same random matching with e^j. If a random matching covers p of $e^1, \ldots, e^{i-1}$ which are disjoint from e^i, the probability that e^i is covered by that random matching is $\frac{1}{B-p}$, and this is maximized when $p = i - 1$.

By a union bound over the a random matchings, $\Pr[e^i | e^1, \ldots, e^{i-1}] \leqslant \frac{a}{B-i}$. $\square$

The probability that all of $e^1, \ldots, e^q$ exist is at most

$$\prod_{i=1}^{q} \frac{a}{B-i} \leqslant \left(\frac{a}{B-q}\right)^q \leqslant \left(\frac{a}{B-k'}\right)^q .$$

Since edges of G_0 are processed independently, the probability of success for one fixed purported cycle is $(\frac{a}{B-k'})^{k'}$. The expected number of cycles of length k' is

$$nd^{k'-1}B^{k'} \cdot \left(\frac{a}{B-k'}\right)^{k'} = nd^{k'-1}a^{k'}\left(1 + \frac{k'}{B-k'}\right)^{k'} \leqslant nd^{k'-1}a^{k'} \exp\left(\frac{k'^2}{B-k'}\right) \leqslant en(ad)^{k'}$$

by taking $B - k' \geqslant k'^2$. Summing over $k' = 1, \ldots, ad$, the expected number of cycles of length up to ad, is bounded by $en(ad)^{ad+1}$ if $B \geqslant (ad)^2 + ad$. Take $B \geqslant 4d^2 \cdot e(ad)^{ad+1}$. Then with probability at least $3/4$, the number of cycles of length at most ad is at most $\frac{Bn}{d^2}$. By taking $1/d^2$ fraction of vertices away (one for each short cycle), we have a girth at least $ad+1$, which implies

$$\left(1 - \frac{1}{d^2}\right)|V_G| \leqslant |V_{G'}| \leqslant |V_G| .$$

Completeness. Let I_0 be an independent set of G_0 of measure c. Then $I = I_0 \times [B]$ is also an independent set of G of measure c. Let $I' = I \cap V_{G'}$. I' is independent in both G and G', and the measure of I' in G is at least $c - 1/d^2$, implying that the measure in G' is even bigger.

Soundness. Suppose that every subset of V_{G_0} of measure at least s contains an edge. Fix a subset S of $V_{G'}$ with measure $\frac{s+\epsilon}{1-1/d^2}$. It has measure at least $s+\epsilon$ in G. By an averaging argument, at least a fraction s of the vertices of G_0 satisfy $|\mathsf{cloud}(v) \cap S| \geqslant \epsilon B$; let's call these vertices *good*. We know that there exists an $(u, w) \in E_{G_0}$ both of whose endpoints are good.

Claim D.8. *For an appropriate choice of parameters a and B, with probability at least $3/4$ the following holds: For every $(u, w) \in E_{G_0}, X \subseteq \mathsf{cloud}(u), Y \subseteq \mathsf{cloud}(w)$ such that $|X| = |Y| = \epsilon B$, there is an edge (of E_G) between X and Y.*

Proof. Fix (u, w) first. The possibilities of choosing X and Y is

$$\binom{B}{\epsilon B}^2 \leqslant \exp(O(\epsilon \log(1/\epsilon) B)) \,.$$

The probability that one random matching does not have any edge between X and Y is at most $(1-\epsilon)^{\epsilon L} \leqslant \exp(-\epsilon^2 L)$, as all ϵL vertices in X should avoid the Y, and the conditional probability is at most $(1-\epsilon)$, since it decreases as we condition more. By considering a independent random matchings and taking a union bound over all possibilities of choosing X and Y, the probability that the claim fails for (u, w) is at most

$$\exp(\epsilon \log(1/\epsilon) B) \cdot \exp(-a\epsilon^2 B) \leqslant \frac{1}{4m}$$

by taking $a = \Omega\left(\frac{\log(1/\epsilon)}{\epsilon}\right)$ and $B = \Omega\left(\frac{\log m}{\epsilon}\right)$. With a final union bound over all m edges of E_0, the claim follows. $\square$

Combining Results and Parameters. From the result of [10], given a d-regular graph, it is NP-hard to distinguish

- Completeness: There is an independent set of measure $c = \Omega(\frac{1}{\log d})$.
- Soundness: Every independent set has measure at most $s = O(\frac{\log^3 d}{d})$.

Take $\epsilon = s$, which implies that

$$a = O\left(\frac{\log(1/\epsilon)}{\epsilon}\right) = O\left(\frac{d}{\log^2 d}\right) \,.$$

Therefore, the degree of G' is bounded by $ad = O\left(\frac{d^2}{\log^2 d}\right)$. By the above analysis, with probability at least half,

- Completeness: There is an independent set of measure $c - \frac{1}{d^2} = \Omega\left(\frac{1}{\log d}\right)$.
- Soundness: Every independent set of measure $\frac{s+\epsilon}{1-1/d^2} = O(s) = O\left(\frac{\log^3 d}{d}\right)$.

Let $k := ad = O\left(\frac{d^2}{\log^2 d}\right)$ be the maximum degree. Given a graph with maximum degree at most k and girth greater than k, we conclude that it is NP-hard to approximate MIS within a factor of $\Omega\left(\frac{d}{\log^4 d}\right) = \Omega\left(\frac{\sqrt{k}}{\log^3 k}\right)$. $\square$

D.5 Unique-Games hardness of MIS-k on high-girth graphs

We now turn to the result yielding a near-optimal inaproximability factor for MIS-k-k, relying on the UGC. Together with Lemma D.5, this proves the second half (UG-hardness) of Theorem 1.5.

Lemma D.9. *It is Unique Games-hard (under randomized reductions) to approximate MIS-k-k within a factor of $\Omega(\frac{k}{\log^4 k})$.*

Proof. By Theorem D.2, we know that given a degree-d graph G_0, it is UG-hard to distinguish whether there is an independent set of measure $\Omega(1/\log d)$ or G_0 is $(\beta, \Omega(\beta^2))$-dense for any $\beta \geqslant \beta_0 := \Theta(\log d/d)$.

Our reduction is identical to Lemma D.6, which showed NP-hardness of the same problem. We replace each vertex by B copies, and each edge by a random matchings to obtain a graph G, where a and B are parameters to be determined later. Finally, we remove cycles of length at most ad by simply getting rid of one vertex for each such cycle, to obtain G'.

In Lemma D.6, a was almost d and the degree was blown up almost quadratically, which resulted in a worse inapproximability factor. Our strategy is to show that the final hard instance graph with large girth is again dense, so we can sparsify again to produce a low-degree graph with similar completeness and soundness. That is why we only obtained UG-hardness because the NP-hardness result does not give such a strong density of the graph G_0 that we start the reduction with.

Soundness. The only part where the analysis differs from that of Lemma D.6 is the soundness. We prove the following stronger version of Claim D.8. Say a bipartite graph is a (β, α)-dense if we take β fraction of vertices from each side, at least α fraction of edges lie within the induced subgraph.

Claim D.10. *For certain choices of a and B, the following holds with probability at least $3/4$: For every $(u, w) \in E_{G_0}$, the bipartite graph between $\mathsf{cloud}(u)$ and $\mathsf{cloud}(w)$ is $(\epsilon, \epsilon^2/8)$-dense for all $\epsilon \geqslant \beta_0$.*

Proof. Fix (u, w), and $\epsilon \in [\beta_0, 1]$, and $X \subseteq \mathsf{cloud}(u)$ and $Y \subseteq \mathsf{cloud}(w)$ be such that $|X| = |Y| = \epsilon B$. The possibilities of choosing X and Y is

$$\binom{B}{\epsilon B}^2 \leqslant \exp(O(\epsilon \log(1/\epsilon) B))$$

Without loss of generality, let $X = Y = [\epsilon B]$. In one random matching, let X_i $(i \in [\epsilon B])$ be the random variable indicating whether vertex $(u, i) \in X$ is matched with a vertex in Y or not. $\Pr[X_1 = 1] = \epsilon$, and $\Pr[X_i = 1 | X_1, \ldots, X_{i-1}] \geqslant \epsilon/2$ for $i \in [\epsilon B/2]$ and any $X_1, \ldots, X_{i-1}$. Therefore, the expected number of edges between X and Y is at least $\epsilon^2 B/4$. With a random matchings, the expected number is at least $a\epsilon^2 B/4$. By Chernoff bound, the probability that it is less than $a\epsilon^2 B/8$ is at most $\exp(\frac{a\epsilon^2 B}{32})$. By union bound over all possibilities of choosing X and Y, the probability that the bipartite graph is not $(\epsilon, \epsilon^2/8)$-dense is

$$\exp(\epsilon \log(1/\epsilon) B) \cdot \exp\Big(-\frac{a\epsilon^2 B}{32}\Big) \leqslant \frac{1}{4mB}$$

by taking $a = \Omega(\frac{\log(1/\beta_0)}{\beta_0})$ and $B = \Omega\big(\frac{\log(mB)}{\beta_0}\big)$. Note that a is increased by a (universal) constant factor compared to Lemma D.6. A union bound over all possible choices of ϵ (B possibilities) and m edges of E_0 implies the claim. $\square$

Claim D.11. *With the parameters a and B decided above, G is $(4\beta_0 \log(1/\beta_0), \Omega(\beta_0^2))$-dense.*

Proof. Fix a subset S of measure $4\beta_0 \log(1/\beta_0)$. For a vertex v of G_0, let $\mu(v) := \frac{|\mathsf{cloud}(v) \cap S|}{B}$. Note that $\mathbb{E}_v[\mu(v)] = 4\beta_0 \log(1/\beta_0)$. Partition V_{G_0} into $t+1$ buckets $B_0, \ldots, B_t$ ($t := \lceil \log_2(1/\beta_0) \rceil$), such that B_0 contains v such that $\mu(v) \leqslant \beta_0$, and B_i contains v such that $\mu(v) \in (2^{i-1}\beta_0, 2^i\beta_0]$. Denote

$$\mu(B_i) := \frac{\sum_{v \in B_i} \mu(v)}{|V_{G_0}|} \ .$$

Clearly $\mu(B_0) \leqslant \beta_0$. Pick $i \in \{1, \ldots, t\}$ with the largest $\mu(B_i)$. We have $\mu(B_i) \geqslant 2\beta_0$. Let $\gamma = 2^{i-1}\beta_0$. All vertices of B_i has $\mu(v) \in [\gamma, 2\gamma]$, so $|B_i| \geqslant (\beta_0/\gamma)|V_{G_0}|$. Since G_0 is $(\beta, \Omega(\beta^2))$-dense for any $\beta \geqslant \beta_0$, at least $\Omega((\beta_0/\gamma)^2)$ fraction of edges lie in the subgraph induced by B_i (since $\beta_0/\gamma \geqslant \beta_0$). For each of these edges, by Claim D.10, at least $\gamma^2/8$ fraction of the edges from the bipartite graph connecting the clouds of its two endpoints, lie in the subgraph induced by S (since $\gamma \geqslant \beta_0$). Overall, we conclude that there are at least a fraction $\Omega((\beta_0/\gamma)^2) \cdot \gamma^2/8 = \Omega(\beta_0^2)$ of edges inside the subgraph induced by S. $\square$

Final Result. Once the parameter a is fixed, girth control and completeness work exactly in the same way as Lemma D.6. As a result, G' is obtained from G by deleting at most $1/d^2$ fraction of vertices, and it has no cycle of length at most ad (which is actually much better than we need as the final degree will be much less than that). In the completeness case, G' has an indepenent set of size $O(1/\log d)$ and in the soundness case, G' is $(\frac{4\beta_0 \log(1/\beta_0)}{1-1/d^2}, \Omega(\beta_0^2))$-dense, or $(5\beta_0 \log(1/\beta_0), \Omega(\beta_0^2))$-dense.

Using density of G', we sparsify G' again by taking kn edges. Following the analysis of [4] again, maximum degree $k = \Theta(\frac{\log^2(1/\beta_0)}{\beta_0})$ is enough to preserve the completeness and soundness parameters by a constant factor. Sparsifying does not decrease girth, so we can conclude that it is UG-hard to approximate MIS in degree-k graphs with girth more than k within a factor of $\Omega\Big(\frac{1/\log d}{\beta_0 \log(1/\beta_0)}\Big) = \Omega\big(\frac{k}{\log^4 k}\big)$. $\square$

E Simple Unique Games Hardness

In this section, we give a simpler proof of Svensson's structured UG-hardness result for FVS [36], using the powerful tool of Mossel's invariance principle [33].

Theorem E.1. *Fix an integer $k \geqslant 3$ and $\epsilon \in (0, 1)$. Given a directed graph $G = (V_G, E_G)$, it is Unique Games-hard to distinguish the following cases.*

- *Completeness: The vertex is partitioned into $V_0, \ldots, V_k$ such that $\mu(V_i) \geqslant \frac{1-\epsilon}{k}$ and each edge not incident on V_0 goes from V_i to $V_{(i+1)}$ for some $i \in [k]$.*
- *Soundness: Any subset of measure ϵ contains a k-cycle.*

As a result, it is UG-hard (under randomized reductions) to approximate FVS within a factor of k, for any constant k.

The idea of using the invariance principle to prove a structured hardness of Ek-HVC was first used in the elegant paper of Bansal and Khot [6]. Our main idea for this result is to use a more restricted distribution for the dictatorship test than the one used in [6] to ensure more structures in the completeness case. At the same time we also ensure that the distribution satisfies certain properties so that the same soundness analysis can be applied.

E.1 Dictatorship Test

We propose a simple dictatorship test for FVS, which is used to prove that it is UG-hard to approximate FVS within any constant factor. Given positive integers k and R and $\epsilon > 0$, our dictatorship test is a vertex-weighted graph $G = (V_G, E_G)$ where $V_G = ([k] \cup \{0\})^R$ and edges in E_G are carefully chosen to prove the following properties.

- Completeness: For each $1 \leqslant j \leqslant R$, *depending only on the jth coordinate*, V_G can be partitioned to $k+1$ parts $V_0, \ldots, V_k$ such that $\mu(V_0) = \epsilon$, $\mu(V_1) = \cdots = \mu(V_k) = \frac{1-\epsilon}{k}$, where the induced subgraph on $V_1 \cup \cdots \cup V_k$ has each edge going from V_i to $V_{(i+1)}$ for some $1 \leqslant i \leqslant k$. It is easy to see that $V_0 \cup V_i$ for any $1 \leqslant i \leqslant k$ gives a feedback vertex set with measure $\epsilon + \frac{1-\epsilon}{k}$.
- Soundness: Any subset of measure at least ϵ that does not reveal any *influential coordinate* must contain a k-cycle.

Before defining G, we first define a k-uniform hypergraph $H = (V_H, E_H)$ with $V_H = V_G = [k]^R$. The graph G is then simply obtained by replacing a hyperedge $(x^1, \ldots, x^k)$ by k edges $(x^1, x^2), \ldots, (x^k, x^1)$. Unlike the rest of the paper, we have to make these k edges in a *specified order*. The hypergraph H is vertex-weighted and edge-weighted. Both weights sum to 1 and induce probability distributions, where the weight of vertex x is the sum of the weight of the edges containing x divided by k. The hyperedges of H are described by the following procedure to sample k vertices $(x^1, \ldots, x^k)$ from $[k]^R$.

- For each coordinate $1 \leqslant j \leqslant R$, sample $(x^1)_j, \ldots, (x^k)_j$ as follows, independently of the other coordinates.
  - Sample $a \in [k]$ uniformly at random.
  - Set $(x^1)_j = a, (x^2)_j = (a+1), \ldots, (x^k)_j = (a+k-1)$.
  - For each $(x^i)_j$, set $(x^i)_j = 0$ with probability ϵ independently.

This defines the hypergraph E_H. In the above distribution to sample $(x^1, \ldots, x^k)$, the marginal on each x^i is the same: $\Pr[x^i = (a_1, \ldots, a_R)] = \prod_{j=1}^R \mu(a_j)$, where $\mu : [k] \cup \{0\} \to \mathbb{R}$ is defined by $\mu(0) = \epsilon$ and $\mu(i) = \frac{1-\epsilon}{k}$ for $i \in [k]$. Let the weight of $(x_1, \ldots, x_R)$ be this quantity. The sum of the vertex weights is also 1.

With nonzero probability a randomly sampled hyperedge $(x^1, \ldots, x^k)$ might have $x^i = x^j$ for some $i \neq j$. We call such hyperedges *defective* since they do not make H k-uniform.

However, $x^i = x^j$ means $x^i = x^j = (0, 0, \ldots, 0)$, so the probability that it happens is at most ϵ^{2R} and the sum of the weights of the defective hyperedges is at most $k^2\epsilon^{2R}$.

Finally, we define G. The vertex set $V_G = V_H$ with the same vertex weights, and for each non-defective hyperedge $(x^1, \ldots, x^k) \in E_H$, we add k edges $(x^1, x^2), \ldots, (x^k, x^1)$ to E_G. The analysis dealing with edge weights will be done in H, so we do not add edge weights to the edges of G.

E.2 Analysis of Dictatorship Test

Completeness Fix a coordinate $1 \leqslant j \leqslant R$. For all $0 \leqslant i \leqslant k$, let $V_i = \{(x_1, \ldots, x_R) \in V_G : x_j = i\}$. $\mu(V_0) = \epsilon, \mu(V_i) = \frac{1-\epsilon}{k}$ by definition. The distribution on $(x^1, \ldots, x^k)$ satisfies that for any $1 \leqslant i \leqslant k$, $(x^{(i+1)})_j = ((x^i)_j + 1)$ or at least one of $(x^i)_j, (x^{(i+1)})_j$ is 0. This proves that if we delete V_0 and the edges incident on it, all the remaining edges will go from V_i to $V_{(i+1)}$.

Soundness We introduce some definitions and properties of correlated spaces and Fourier analysis of functions defined on (product of) these spaces. See Mossel [33] for details.

Let $\Omega := [k] \cup \{0\}$ and $\mu : \Omega \to \mathbb{R}$ such that $\mu(0) = \epsilon$ and $\mu(i) = \frac{1-\epsilon}{k}$ as defined previously. Let (Ω^k, μ') be the probability space defined by the distribution of $(x^1)_j, \ldots, (x^k)_j$ for some j from our hyperedge sampling. Note that the marginal distribution of each copy of Ω is μ.

Given two probability spaces $(\Omega_1 \times \Omega_2, \nu)$, we define the correlation

$$\rho(\Omega_1, \Omega_2; \nu) = \sup \left\{ \mathsf{Cov}[f, g] : f \in \mathbb{R}^{\Omega_1}, g \in \mathbb{R}^{\Omega_2}, \mathsf{Var}[f] = \mathsf{Var}[g] = 1 \right\}.$$

With more than two spaces, the correlation $(\Omega_1 \times \cdots \times \Omega_k, \nu)$ is defined by

$$\rho(\Omega_1, \ldots, \Omega_k; \nu) = \max_{1 \leqslant i \leqslant k} \rho(\prod_{j=1}^{i-1} \Omega_j \times \prod_{j=i+1}^{k} \Omega_j, \Omega_i; \nu).$$

Going back to our distribution (Ω^k, μ), note that $(0, 0, \ldots, 0)$ has probability $\alpha := \epsilon^k$, and this is indeed the smallest nonzero probability assuming $\epsilon < \frac{1}{k+1}$. Furthermore, every $(x_1, \ldots, x_R) \in \Omega^k$ with nonzero probability is *connected* to $(0, 0, \ldots, 0)$ in a sense that $(x_1, x_2, \ldots, x_R)$, $(0, x_2, x_3, \ldots, x_R)$, $(0, 0, x_3, \ldots, x_R)$, $\ldots$, $(0, 0, 0, \ldots, x_R)$, $(0, 0, 0, \ldots, 0)$ is a sequence of elements with nonzero probability where each consecutive elements differ by at most 1 coordinate. In this case, Lemma 2.9 of [33] ensures that $\rho(\Omega^k; \mu) \leqslant 1 - \alpha^2/2 < 1$.

Let $\chi_0, \ldots, \chi_k \in \mathbb{R}^{\Omega}$ be orthonormal random variables satisfying that $\chi_0 \equiv 1$. Given $f : \Omega^R \to [0, 1]$, its multilinear decomposition is

$$f(x_1, \ldots, x_R) = \sum_{\alpha \in \Omega^R} \hat{f}(\alpha) \prod_{j=1}^{R} \chi_{\alpha(j)}(x_j).$$

Let $\mathsf{Supp}(\alpha)$ be the number of nonzero coordinates of α. The d-degree influence of the jth coordinate of f is defined by

$$\mathsf{Inf}_j^{\leqslant d}(f) = \sum_{\alpha \in \Omega^R : \alpha_j \neq 0, \mathsf{Supp}(\alpha) \leqslant d} \hat{f}(\alpha)^2.$$

It is well-know that $\sum_{j=1}^{R} \mathsf{Inf}_j^{\leqslant d}(f) \leqslant d$ for $[0,1]$-valued f.

We establish the soundness property using the invariance principle. Let A be the subset of V_G of measure at least ϵ, and f its indicator function. The following is the statement of the invariance principle tailored to our setting.

Theorem E.2. *[33] For any $k \geqslant 3, 0 < \epsilon < \frac{1}{k+1}$ (so that $\rho < 1$ is fixed), $\mu > 0$, there exist $\delta > 0, \tau > 0$ and integer d independent of R such that for any indicator function f satisfying $\mathbb{E}[f] \geqslant \mu$ and $\mathsf{Inf}_j^{\leqslant d}(f) \leqslant \tau$ for all j, we have*

$$\Pr_{x^1,\ldots,x^k}[x^1,\ldots,x^k \in A] = \mathop{\mathbb{E}}_{x^1,\ldots,x^k}[\prod_{i=1}^{k} f(x^i)] > \delta$$

We can conclude that as long as δ is greater than the sum of the weights of the defective hyperedges, which is at most $k^2\epsilon^{2R}$, such A contains a non-defective hyperedge $(x^1,\ldots,x^k)$ of H and the corresponding k-cycle of G.

E.3 Reduction from the Unique Games

We introduce the Unique Games Conjecture and its equivalent variant.

Definition E.3. *An instance $\mathcal{L}(B(V_B \cup W_B, E_B), [R], \{\pi(v,w)\}_{(v,w)\in E_B})$ of Unique Games consists of a regular bipartite graph $B(V_B \cup W_B, E_B)$ and a set $[R]$ of labels. For each edge $(v,w) \in E_B$ there is a constraint specified by a permutation $\pi(v,w) : [R] \to [R]$. The goal is to find a labeling $l : V_B \cup W_B \to [R]$ of the vertices such that as many edges as possible are satisfied, where an edge $e = (v,w)$ is said to be satisfied if $l(v) = \pi(v,w)(l(w))$.*

Definition E.4. *Given a Unique Games instance $\mathcal{L}(B(V_B \cup W_B, E_B), [R], \{\pi(v,w)\}_{(v,w)\in E_B})$, let $\mathsf{Opt}(\mathcal{L})$ denote the maximum fraction of simultaneously satisfied edges of $\mathcal{L}$ by any labeling, i.e.*

$$\mathsf{Opt}(\mathcal{L}) := \frac{1}{|E|} \max_{l:V_B\cup W_B\to[R]} |\{e \in E : l \text{ satisfies } e\}|$$

Conjecture E.5 ([27]). *For any constants $\eta > 0$, there is $R = R(\eta)$ such that, for a Unique Games instance $\mathcal{L}$ with label set $[R]$, it is NP-hard to distinguish between*

- $\mathsf{opt}(\mathcal{L}) \geqslant 1-\eta$.
- $\mathsf{opt}(\mathcal{L}) \leqslant \eta$.

To show the optimal hardness result for Min-Vertex-Cover, Khot and Regev [29] introduced the following stronger conjecture, and proved that it is in fact equivalent to the original Unique Games Conjecture.

Conjecture E.6 ([29]). *For any constants $\eta > 0$, there is $R = R(\eta)$ such that, for a Unique Games instance $\mathcal{L}$ with label set $[R]$, it is NP-hard to distinguish between*

- *There is a set $W' \subseteq W_B$ such that $|W'| \geqslant (1-\eta)|W_B|$ and a labeling $l : V_B \cup W_B \to [R]$ that satisfies every edge (v,w) for $v \in V'$ and $w \in W_B$.*

- $\mathsf{opt}(\mathcal{L}) \leqslant \eta$.

We use the following reduction from the Unique Games. Given an instance $\mathcal{L}$ to the Unique Games, we assign to each vertex $w \in W$ the hypercube Ω_w^R. $V_G = V_H := \cup_{w \in W_B} \Omega^R$. The weight of each vertex (w, x) is the weight of x in Ω^R divided by W, so that the sum of the weights is again 1.

For a permutation $\sigma : [R] \to [R]$, let $x \circ \sigma := (x_{\sigma(1)}, \dots, x_{\sigma(R)})$. The weighted hyperedges of H is again defined by the procedure to sample k vertices $(w^1, x^1), \dots, (w^k, x^k)$.

- Sample $v \in V_B$ uniformly at random.
- Sample k vertices $w^1, \dots, w^k \in W_B$ i.i.d. from neighbors of v.
- Sample $x^1, \dots, x^k \in \Omega^R$ from the dictatorship distribution.
- Return the hyperedge $((w^1, x^1 \circ \sigma(v, w^1)), \dots, (w^k, x^k \circ \sigma(v, w^k)))$.

For each hyperedge $((w^1, x^1), \dots, (w^k, x^k))$, we add k edges $((w^1, x^1), (w^2, x^2)), \dots, ((w^k, x^k), (w^1, x^1))$ to G.

Completeness Suppose there exist a labeling l and a subset $W' \subseteq W_B$ with $|W'| \geqslant (1-\eta)|W_B|$ such that l satisfy every edge incident on W'. For $1 \leqslant i \leqslant k$, let

$$V_i := \cup_{w \in W'} \left\{(w, x) : x_{l(w)} = i\right\}$$

and $V_0 := V_G \backslash (\cup_{i=1}^k V_i)$. Let G' be the induced subgraph on $V_G \backslash V_0$. For any edge $((w^1, x^1), (w^2, x^2)) \in E_{G'}$, we know $w^1, w^2 \in W'$ and they share a neighbor $v \in V_B$. By the property of our dictatorship test, for each $1 \leqslant j \leqslant R$, $a := (x^1)_{\sigma(v,w^1)^{-1}(j)}$ and $b := (x^2)_{\sigma(v,w^2)^{-1}(j)}$ satisfy that either one of them is zero or $b = (a+1)$. Therefore, if $(w^1, x^1), (w^2, x^2) \notin V_0$, which implies $(x^1)_{\sigma(v,w^1)^{-1}(l(v))} = (x^1)_{l(w^1)}, (x^2)_{\sigma(v,w^2)^{-1}(l(v))} = (x^2)_{l(w^2)}$ are nonzero, we can conlude that $(w^1, x^1) \in V_i$ and $(w^2, x^2) \in V_{(i+1)}$ for some $1 \leqslant i \leqslant k$.

Soundness The soundness anlaysis is standard and closely follows [6]. Suppose $A \subseteq V_H$ of measure at least β such that it does not contain any non-defective hyperedge. Let $A_w = \Omega_w^R \cap A$ be the vertices of A that lie in Ω_w^R for $w \in W_B$. Let $f_w : \Omega^R \to \{0, 1\}$ be the indicator function of A_w. Define

$$f_v(x) = \mathop{\mathbb{E}}_{w \in N(v)}[f_w(x \circ \sigma(v, w))]$$

where $N(v)$ is the set of neighbors of $v \in V_B$. Since B is regular, $\mathbb{E}_{v,x}[f_v(x)] \geqslant \beta$. By averaging argument, at least $\beta/2$ fraction of vertices in V_B satisfy $\mathbb{E}_x[f_v(x)] \geqslant \beta/2$. Call such vertices *good*.

Since A is independent set, for any $v \in V$ and its k neighbors $w^1, \dots, w^k$, we have

$$\mathop{\mathbb{E}}_{x^1, \dots, x^k}[\prod_{i=1}^k f_{w^i}(x^i \circ \sigma(v, w^i))] \leqslant k^2 \epsilon^{2R}.$$

Averaging over all k-tuples $w^1, \ldots, w^k$ of neighbors of v, we have

$$\mathop{\mathbb{E}}_{x^1,\ldots,x^k}[\prod_{i=1}^{k} f_v(x^i)] = \mathop{\mathbb{E}}_{x^1,\ldots,x^k} \mathop{\mathbb{E}}_{w^1,\ldots,w^k \in N(v)}[\prod_{i=1}^{k} f_{w^i}(x^i \circ \sigma(v, w^i))] \leqslant k^2 \epsilon^{2R}.$$

Applying Theorem E.2 (take R large enough to make sure that $k^2\epsilon^{2R} \ll \delta$), there exist τ and d such that f_v has a coordinate j with $\mathsf{Inf}_j^{\leqslant d} \geqslant \tau$. Set $l(v) = j$. Since

$$\begin{aligned}\mathsf{Inf}_j^{\leqslant d}(f_v) &= \sum_{\alpha_j \neq 0, |\alpha| \leqslant d} \hat{f}_v(\alpha)^2 = \sum_{\alpha_j \neq 0, |\alpha| \leqslant d} (\mathop{\mathbb{E}}_{w}[\hat{f}_w(\sigma(v,w)^{-1}(\alpha))]^2) \\ &\leqslant \sum_{\alpha_j \neq 0, |\alpha| \leqslant d} \mathop{\mathbb{E}}_{w}[\hat{f}_w(\sigma(v,w)^{-1}(\alpha))^2] = \mathop{\mathbb{E}}_{w}[\mathsf{Inf}^{\leqslant d}_{\sigma(v,w)^{-1}(j)}(f_w)],\end{aligned}$$

at least $\tau/2$ fraction of v's neighbors satisfy $\mathsf{Inf}^{\leqslant d}_{\sigma(v,w)^{-1}(j)} \geqslant \tau/2$. There are at most $2d/\tau$ coordinates with degree-d influence at most $\tau/2$, and $l(w)$ is chosen uniformly among those coordinates (if there is none, set it arbitrarily). The above probabilistic strategy satisfies at least $(\beta/2)(\tau/2)(\tau/2d)$ fraction of all edges, completing the proof of soundness.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de