Approximating Maximum Subgraphs Without Short Cycles

Guy Kortsarz[1*], Michael Langberg[2], and Zeev Nutov[3]

[1] Rutgers University, Camden. guyk@camden.rutgers.edu
Currently visiting IBM Research Yorktown Heights.
[2] The Open University of Israel mikel@openu.ac.il
[3] The Open University of Israel nutov@openu.ac.il

Abstract. We study approximation algorithms, integrality gaps, and hardness of approximation, of two problems related to cycles of "small" length k in a given graph. The instance for these problems is a graph $G = (V, E)$ and an integer k. The k-Cycle Transversal problem is to find a minimum edge subset of E that intersects every k-cycle. The k-Cycle-Free Subgraph problem is to find a maximum edge subset of E without k-cycles.

The 3-Cycle Transversal problem (covering all triangles) was studied by Krivelevich [Discrete Mathematics, 1995], where an LP-based 2-approximation algorithm was presented. The integrality gap of the underlying LP was posed as an open problem in the work of Krivelevich. We resolve this problem by showing a sequence of graphs with integrality gap approaching 2. In addition, we show that if 3-Cycle Transversal admits a $(2-\varepsilon)$-approximation algorithm, then so does the Vertex-Cover problem, and thus improving the ratio 2 is unlikely. We also show that k-Cycle Transversal admits a $(k-1)$-approximation algorithm, which extends the result of Krivelevich from $k = 3$ to any k. Based on this, for odd k we give an algorithm for k-Cycle-Free Subgraph with ratio $\frac{k-1}{2k-3} = \frac{1}{2} + \frac{1}{4k-6}$; this improves over the trivial ratio of 1/2.

Our main result however is for the k-Cycle-Free Subgraph problem with even values of k. For any $k = 2r$, we give an $\Omega\left(n^{-\frac{1}{r}+\frac{1}{r(2r-1)}-\varepsilon}\right)$-approximation scheme with running time $\varepsilon^{-\Omega(1/\varepsilon)}\mathsf{poly}(n)$. This improves over the ratio $\Omega(n^{-1/r})$ that can be deduced from extremal graph theory. In particular, for $k = 4$ the improvement is from $\Omega(n^{-1/2})$ to $\Omega(1/n^{-1/3-\varepsilon})$. Similar results are shown for the problem of covering cycles of length $\leq k$ or finding a maximum subgraph without cycles of length $\leq k$.

1 Introduction

In this work, we study approximation algorithms, integrality gaps, and hardness of approximation, of two problems related to cycles of a given "small" length k (henceforth k-cycles) in a graph. The instance for each one of these problems is an undirected graph $G = (V, E)$ and an integer k. The goal is:

* Partially supported by NSF Award Grant number 072887.

k-Cycle Transversal:
Find a minimum edge subset of E that intersects every k-cycle.

k-Cycle Free Subgraph:
Find a maximum edge subset of E without k-cycles.

Note that k-Cycle Transversal and k-Cycle-Free Subgraph are complementary problems, as the sum of their optimal values equals $|E| = m$; hence they are equivalent with respect to their optimal solutions. However, they differ substantially when considering approximate solutions. Also note that for $k = O(\log n)$ the number of k cycles in a graph can be computed in polynomial time, c.f., [3], and that it is polynomial for any fixed k. The k-Cycle Transversal problem is sometimes referred to as the "k-cycle cover" problem (as one seeks to cover k-cycles by edges). We adapt an alternative name, to avoid any mixup with an additional problem that has the same name – the problem of covering the edges of a given graph with a minimum family of k-cycles.

We will also consider problems of covering cycles of length $\leq k$ or finding a maximum subgraph without cycles of length $\leq k$. We will elaborate on the relation of these problems to our problems later. Most of our results extend to the case when edges have weights, but for simplicity of exposition, we consider unweighted and simple graphs only. We will also assume w.l.o.g. that G is connected.

1.1 Previous and related work

Problems related to k-cycles are among the most fundamental in the fields of Extremal Combinatorics, Combinatorial Optimization, and Approximation Algorithms, and they were studied extensively for various values of k. See for example [5, 1, 2, 17, 4, 8, 10, 12, 11, 13, 14, 16, 15, 6] for only a small sample of papers on the topic. 3-Cycle Transversal was studied by Krivelevich [12]. Erdös et al. [6] considered 3-Cycle Transversal and 3-Cycle-Free Subgraph and their connections to related problems. Pevzner et al. [18] studied the problem of finding a maximum subgraph without cyles of lengt $\leq k$ in the context of computational biology, and suggested some heuristics for the problem, without analyzing their approximation ratio. However, most of the related papers studied k-Cycle-Free Subgraph in the context of extremal graph theory, and address the maximum number of edges in a graph without k-cycles (or without cycles of length $\leq k$). This is essentially the k-Cycle-Free Subgraph problem on complete graphs. In this work we initiate the study of k-Cycle-Free Subgraph in the context of approximation algorithms on general graphs.

As the state of the art differs substantially for odd and even values of k, we consider these cases separately. But for both odd and even k, note that k-Cycle Transversal is a particular case of the problem of finding a minimum transversal in a k-uniform hypergraph (which is exactly the Hitting-Set problem). Thus a simple greedy algorithm which repeatedly removes a k-cycle until no k-cycles remain, has approximation ratio k.

Odd k: For k-Cycle Transversal, an improvement over the trivial ratio of k was obtained for $k = 3$ by Krivelevich [12]. Let $\mathcal{C}_k(G)$ denote the set of k cycles in G, and let $\tau^*(G)$ denote the optimal value of the following LP-relaxation for k-Cycle Transversal:

$$\begin{aligned} \min\ & \textstyle\sum_{e \in E} x_e \\ \text{s.t.}\ & \textstyle\sum_{e \in C} x_e \geq 1 \ \forall C \in \mathcal{C}_k(G) \\ & x_e \geq 0 \qquad\quad \forall e \in E \end{aligned} \tag{1}$$

Theorem 1 (Krivelevich [12]). 3-Cycle Transversal *admits a* 2*-approximation algorithm, that computes a solution of size at most* $2\tau^*(G)$*.*

For odd values of k, k-Cycle-Free Subgraph admits an easy $1/2$-approximation algorithm, as it is well known that any graph G has a subgraph without odd cycles (namely, a bipartite subgraph) containing at least half of the edges (such a subgraph can be computed in polynomial time). In fact, the problem of computing a maximum bipartite subgraph is exactly the Max-Cut problem, for which Goemans and Williamson [9] gave an 0.878-approximation algorithm. Note however that the solution found by the Goemans-Williamson algorithm has size at least 0.878 times the size of an optimal subgraph without odd cycles at all, and the latter can be much smaller than the optimal subgraph without k-cycles only.

Even k: For k-Cycle Transversal with even values of k we are not aware of any improvements over the trivial ratio of k. For k-Cycle-Free Subgraph with even k, it is no longer the case that G has a k-cycle free subgraph containing at least half of the edges. The maximum number $\mathrm{ex}(n, C_{2r})$ of edges in a graph with n nodes and without cycles of length $k = 2r$ has been extensively studied. This is essentially the $2r$-Cycle-Free Subgraph problem on complete graphs. This line of research in extremal graph theory was initiated by Erdös [5]. The first major result is known as the "Even Circuit Theorem", due to Bondy and Simonovits [4], states that any undirected graph without even cycles of length $\leq 2r$ has at most $O(rn^{1+1/r})$ edges. This bound was subsequently improved. To the best of our knowledge, the currently best known upper bound on $\mathrm{ex}(n, C_{2r})$ due to Lam and Verstraëte [15] is $\frac{1}{2}n^{1+1/r} + 2^{r^2}n$. We note that the best lower bounds on $\mathrm{ex}(n, C_{2r})$ are as follows. For $r = 2, 3, 5$ it holds that $\mathrm{ex}(n, C_{2r}) = \Theta(n^{1+1/r})$. For other values of r, the existence of a $2r$-cycle-free graph with $\Theta(n^{1+1/r})$ has not been established, and the best lower bound known is $\mathrm{ex}(n, C_{2r}) = \Omega\left(n^{1+\frac{2}{3k-3+\varepsilon}}\right)$ where $\varepsilon = 0$ if r is odd and $\varepsilon = 1$ if r is even; we refer the reader to [16] for a summary of results of this type. All this implies that on complete graphs (a case which was studied extensively), the best known ratios for $2r$-Cycle-Free Subgraph are: constant for $r = 2, 3, 5$, and $\Omega\left(n^{-\frac{1}{r}+\frac{2}{6r-3+\varepsilon}}\right)$ otherwise. For general graphs, the bound $\mathrm{ex}(n, C_{2r}) = O\left(n^{1+1/r}\right)$ implies an $\Omega(n^{-1/r})$-approximation by taking a spanning tree of G as a solution. In particular, for $k = 4$, the approximation ratio is $\Omega(1/\sqrt{n})$, and no better approximation ratio was known for this case.

1.2 Our results

Our main result is for the k-Cycle-Free Subgraph problem on even values of k. It can be summarized by the following theorem:

Theorem 2. *For $k = 2r$, k-Cycle-Free Subgraph admits an $\Omega\left(n^{-\frac{1}{r}+\frac{1}{r(2r-1)}-\varepsilon}\right)$-approximation scheme with running time $\varepsilon^{-\Omega(1/\varepsilon)}\mathsf{poly}(n)$. In particular, 4-Cycle-Free Subgraph admits an $\Omega(1/n^{-1/3-\varepsilon})$-approximation scheme.*

For dense graphs, we obtain better ratios that are close to the ones known for complete graphs. Proof of the following statement will appear in the full version of this paper.

Theorem 3. *Let $G = (V, E)$ be a graph with n nodes and at least εn^2 edges. Then G contains a $2r$-cycle-free subgraph with at least $\varepsilon \cdot \mathrm{ex}(n, C_{2r})$ edges.*

On the negative side, the only hardness of approximation result we obtain (again proof will appear in the full version of this paper) is APX-hardness. Thus for even values of k there is a large gap between the upper and lower bounds we present. Resolving this large gap is an intriguing question left open in our work.

Our next results are for odd k. Krivelevevich [12] posed as an open question if his (upper) bound of 2 on the integrality gap of LP (1) is tight for $k = 3$. We resolve this question, and in addition show that the ratio 2 achieved by Krivelevich for $k = 3$ is essentially the best possible.

Theorem 4.

(i) *If 3-Cycle Transversal admits a $2 - \varepsilon$ approximation ratio for some positive universal constant $\varepsilon < 1/2$, then so does the Vertex-Cover problem.*
(ii) *For any $\varepsilon > 0$ there exist infinitely many undirected graphs G for which the integrality gap of LP (1) with $k = 3$ is at least $2 - \varepsilon$.*

We note that Theorem 4 holds also for any $k \geq 4$. We also extend the 2-approximation algorithm of Krivelevich [12] for 3-Cycle Transversal to arbitrary k which is odd, and use it to improve the trivial ratio of $1/2$ for k-Cycle-Free Subgraph.

Theorem 5. *For any odd k the following holds:*

(i) *k-Cycle Transversal admits a $(k-1)$-approximation algorithm.*
(ii) *k-Cycle-Free Subgraph admits a $\left(\frac{1}{2} + \frac{1}{4k-6}\right)$-approximation algorithm.*

Some remarks are in place: Theorem 5 is valid also for digraphs, for any value of k. Our results can be used to give approximation algorithms for the problem of covering cycles of length $\leq k$, or finding a maximum subgraph without cycles of length $\leq k$. For $k = 3$ we have for both problems the same ratios as in Theorem 5. For $k \geq 4$, the problem of covering cycles of length $\leq k$ admits a k-approximation algorithm (via the trivial reduction to the Hitting Set problem). For the problem of finding a maximum subgraph without cycles of length $\leq k$, we can show the ratio $\Omega(n^{-1/3-\varepsilon})$ for any k. For $k \geq 6$ this follows from extremal graph theory results mentioned, while for $k = 4, 5$ this is achieved by first computing a bipartite subgraph G' of G with at least $|E|/2$ edges, and then applying on G' the algorithm from Theorem 2 for 4-cycles.

1.3 Techniques

The proof of Theorem 2 is the main technical contribution of this paper. Our algorithm for k-Cycle-Free Subgraph with $k = 2r$ consists of two steps. In the first step we identify in G a subgraph G' which is an *almost* regular bipartite graph with the property that G and G' have approximately the same optimal values. The construction of G' can be viewed as a preprocessing step of our algorithm and may be of independent interest for other optimization problems as well. In the second step of our algorithm, we use the special structure of G' to analyze the simple procedure that first removes edges at random from G' until only few k-cycles remain in G', and then continues to remove edges from G' deterministically (one edge per cycle) until G' becomes k-cycle free.

The proof of Theorem 4(i) gives an approximation ratio preserving reduction from Vertex-Cover on triangle free graphs to 3-Cycle Transversal. It is well known that breaking the ratio of 2 for Vertex-Cover on triangle free graphs is as hard as breaking the ratio of 2 on general graphs. The proof of Theorem 4(ii) uses the same reduction on graphs G that on one hand are triangle free, but on the other have a minimum vertex-cover of size $(1 - o(1))n$. Such graphs exist, and appear in several places in the literature; see for example [7].

The proof of part (i) of Theorem 5 is a natural extension of the proof of Krivelevich [12] of Theorem 1. Part (ii) simply follows from part (i).

Theorems 2, 4, and 5, are proved in Sections 2, 3, and 4, respectively.

2 Proof of Theorem 2

In what follows let $\mathsf{opt}(G)$ be the optimal value of the k-Cycle-Free Subgraph problem on G. We start by a simple reduction which shows that we may assume that our input graph G is bipartite, at the price of loosing only a constant in the approximation ratio. Fix an optimal solution G^* to k-Cycle Free Subgraph. Partition the vertex set V of G randomly into two subsets, A and B, each of size $n/2$, and remove edges internal to A or B. In expectation, the fraction of edges in G^* that remain after this process is $1/2$. With probability at least $1/3$ the fraction of edges in G^* that remain is at least $1/4$; here we apply the Markov inequality on the fraction of edges inside A and B.

Assuming that the input graph G is bipartite, our algorithm has two steps. In the first step, we extract from G a family $\mathcal{G}$ of subgraphs $G_i = (A_i + B_i, E_i)$, so that either: one of these subgraphs has a "θ-semi-regularity" property (see Definition 1 below) and a k-cycle-free subgraph of size *close* to $\mathsf{opt}(G)$ or we conclude that $\mathsf{opt}(G)$ is *small*. In the latter case, we just return a spanning tree in G. In the former case, it will suffice to approximate k-Cycle-Free Subgraph on $G_i \in \mathcal{G}$, which is precisely what we do in the second step of the algorithm.

Definition 1. *A subset A of nodes in a graph is* θ-semi-regular *if $\Delta_A \leq \theta \cdot d_A$ where Δ_A and d_A denote the maximum and the average degree of a node in A, respectively. A bipartite graph with sides A, B is* θ-semi-regular *if each of A, B is θ-semi-regular.*

We will prove the following two statements that imply Theorem 2.

Lemma 1. *Let $k = 2r$. For any bipartite instance G of k-*Cycle-Free Subgraph *there exists an algorithm that in $\varepsilon^{-O(1/\varepsilon)}\mathsf{poly}(n)$ time finds a family $\mathcal{G}$ of at most $2\varepsilon^{-2/\varepsilon}$ subgraphs of G so that at least one of the following holds:*

(i) *$\mathcal{G}$ contains an $n^{2\varepsilon}$-semi-regular bipartite subgraph G_i of G so that $\mathsf{opt}(G_i) = \Omega(\varepsilon^{2/\varepsilon})\mathsf{opt}(G)$.*
(ii) *$\mathsf{opt}(G) = O\left(n\varepsilon^{-2/\varepsilon}\right)$.*

Lemma 2. *k-*Cycle-Free Subgraph *on bipartite θ-semi-regular instances $G = (A+B, E)$ and $k = 2r$ admits an $\Omega\left(\left(\theta r(|A||B|)^{\frac{r-1}{r(2r-1)}}\right)^{-1}\right)$-approximation ratio in (randomized) polynomial time.*

Let us show that Lemmas 1 and 2 imply Theorem 2 for bipartite graphs. We first compute the family $\mathcal{G}$ as in Lemma 1. Then, for each $G_i \in \mathcal{G}$ we compute a k-cycle-free subgraph H_i of G_i using the algorithm from Lemma 2, with $\theta = n^{2\varepsilon}$. Let H be the largest among the subgraphs H_i computed. If H has more than n edges, we output H. Else, we return a spanning tree in G.

2.1 Reduction to θ-semi-regular graphs (Proof of Lemma 1)

Let $G = (A+B, E)$ be a bipartite connected graph, let $\varepsilon > 0$ be a small constant, let $n = |A| + |B|$, and let $\theta = n^\varepsilon$. For simplicity of exposition we will assume that θ and $\ell = 1/\varepsilon$ are integers.

We define an iterative process which partitions a subgraph $G' = (A'+B', E')$ of G with $A' \subseteq A$ and $B' \subseteq B$ into at most $\ell = 1/\varepsilon$ subgraphs so that at least one of the sides in each subgraph is θ-semi-regular. Specifically, the family $\mathcal{F}(G', A)$ is defined as follows. Partition the nodes in A' into at most ℓ sets A_j, where A_j consists of nodes in A' of degree in the range $\left[\theta^j, \theta^{j+1}\right)$. The family $\mathcal{F}(G', A)$ consists of the graphs $G_j = G' - (A' - A_j)$ (namely, G_j is the induced subgraph of G' with sides A_j and B). Note that A_j is a θ-semi-regular node set in G_j, but G_j may not be θ-semi-regular. In a similar way, the family $\mathcal{F}(G', B)$ is defined. Since the the union of the subgraphs in $\mathcal{F}(G', A)$ is G', and since $|\mathcal{F}(G', A)| = 1/\varepsilon$, there exists $G'' \in \mathcal{F}(G', A)$ so that $\mathsf{opt}(G'') \geq \varepsilon \cdot \mathsf{opt}(G')$; a similar statement holds for $\mathcal{F}(G', B)$. For a family $\mathcal{G}$ of subgraphs of G let $\mathcal{F}(\mathcal{G}, A) = \bigcup\{\mathcal{F}(G', A) : G' \in \mathcal{G}\}$ and $\mathcal{F}(\mathcal{G}, B) = \bigcup\{\mathcal{F}(G', B) : G' \in \mathcal{G}\}$.

Define a sequence of families of subgraphs of G as follows. $\mathcal{G}_0 = \{G\}$, $\mathcal{G}_1 = \mathcal{F}(\mathcal{G}_0, A)$, $\mathcal{G}_2 = \mathcal{F}(\mathcal{G}_1, B)$, and so on. Namely, $\mathcal{G}_i = \mathcal{F}(\mathcal{G}_{i-1}, A)$ if i is odd and $\mathcal{G}_i = \mathcal{F}(\mathcal{G}_{i-1}, B)$ if i is even. The following statement is immediate.

Claim. There exists a sequence of graphs $\{G_i = (A_i + B_i, E_i)\}_{i=0}^{2\ell}$ so that for every i: $G_i \in \mathcal{G}_i$, $G_i \subseteq G_{i-1}$, and $\mathsf{opt}(G_i) \geq \varepsilon \cdot \mathsf{opt}(G_{i-1})$.

We now study the structure of the graphs G_i. We show that the average degree in G_i is rapidly decreasing when i is increasing, until one of the G_i's is θ^2-semi-regular.

Claim. For every i, either G_{i+2} is θ^2-semi-regular, or at least one of the following holds:
- if i is even then $d_{A_{i+2}} < d_{A_{i+1}}/\theta$, where d_{A_i} is the average degree of A_i in G_i;
- if i is odd then $d_{B_{i+2}} < d_{B_{i+1}}/\theta$, where d_{B_i} is the average degree of B_i in G_i.

Proof. Suppose that i is even; the proof of the case when i is odd is similar. In $G_{i+1} \in \mathcal{G}_{i+1}$, the maximum degree $\Delta_{A_{i+1}}$ of A_{i+1} is at most θ times the average degree $d_{A_{i+1}}$ of A_{i+1}. If G_{i+2} is not θ^2 regular, then $\Delta_{A_{i+2}} \geq \theta^2 \cdot d_{A_{i+2}}$. However, the maximum degree in A_{i+2} is $\Delta_{A_{i+2}} \leq \Delta_{A_{i+1}} \leq \theta d_{A_{i+1}}$. This implies that $d_{A_{i+2}} \leq d_{A_{i+1}}/\theta$.

All in all, we conclude that for some $i \leq 2/\varepsilon$, G_i is θ^2-semi-regular and satisfies $\mathsf{opt}(G_i) \geq \varepsilon^i \mathsf{opt}(G)$; or $G_{2/\varepsilon}$ has constant average degree and satisfies $\mathsf{opt}(G_{2/\varepsilon}) \geq \varepsilon^{2/\varepsilon}\mathsf{opt}(G)$. The latter implies that $\mathsf{opt}(G) = O(\varepsilon^{-2/\varepsilon} n)$.

2.2 Algorithm for θ-semi-regular graphs (Proof of Lemma 2)

Let $G = (A + B, E)$ be a bipartite θ-semi-regular graph. Let d_A be the average degree of nodes in A, and d_B be the average degree of nodes in B. Let $m = d_A|A| = d_B|B| = \sqrt{d_A d_B |A||B|}$ be the number of edges in G. Our algorithm builds on the following two results (the first is by A. Naor and Verstraëte [17]).

Theorem 6 ([17]). *The maximum number of edges in a bipartite graph $G = (A + B, E)$ without cycles of length $k = 2r$ is:*

$$(2r-3)\left[(|A||B|)^{\frac{r+1}{2r}} + |A| + |B|\right] \quad \text{if } r \text{ is odd}$$

$$(2r-3)\left[|A|^{\frac{1}{2}}|B|^{\frac{r+2}{2r}} + |A| + |B|\right] \quad \text{if } r \text{ is even}$$

Lemma 3. *The number of k-cycles in G is at most $m\theta^{2r-1}d_A^{r-1}d_B^{r-1}$.*

Proof. Consider picking $k = 2r$ distinct nodes in G, r from A and r from B, uniformly at random. Denote the nodes $a_1, a_2, \ldots, a_r \in A$ and $b_1, \ldots, b_r \in B$. We analyze the probability that $(a_1, b_1, a_2, b_2, \ldots, a_r, b_r, a_1)$ is a k cycle in G. In our analysis, our random choices are made according to the order of the cycle at hand, i.e., we first pick a_1, then b_1, then a_2, and so on. As a_1 has degree at most θd_A, the probability that b_1 is adjacent to a_1 is at most $\theta d_A/|B|$. Similarly, as b_1 has degree at most θd_B, the probability that a_2 is adjacent to b_1 is at most $\theta d_B/|A|$. Continuing this line of argument, it is not hard to verify that the probability that $(a_1, b_1, a_2, b_2, \ldots, a_r, b_r, a_1)$ is a k cycle in G is at most

$$\theta^{2r-1}\frac{d_A^r d_B^{r-1}}{|A|^{r-1}|B|^r}\ .$$

The number of k-tuples $(a_1, b_1, a_2, b_2, \ldots, a_r, b_r)$ in G is bounded by $|A|^r|B|^r$. Thus the number of k-cycles in G is at most $\theta^{2r-1}d_A^r d_B^{r-1}|A| = m\theta^{2r-1}d_A^{r-1}d_B^{r-1}$.

We now present our algorithm for k-Cycle Free Subgraph. In our analysis, we assume w.l.o.g. that $|A| \geq |B|$. We also assume that $|A|$ and $|B|$ are sufficiently large with respect to θ. Namely we assume that $|A||B| \geq (16\theta)^2$. Otherwise, the subgraph consisting of a single edge adjacent to v for each node $v \in A$, will suffice to yield an approximation ratio of $\Omega(1/\theta)$ which will equal $\Omega(n^{-2\varepsilon})$ in our final setting of parameters. Theorem 6 implies that

$$\mathsf{opt}(G) \leq 4r((|A||B|)^{\frac{r+1}{2r}} + |A|)$$

for any r. We now consider two cases: the case in which $(|A||B|)^{\frac{r+1}{2r}} \geq |A|$ and thus $\mathsf{opt}(G) \leq 8r(|A||B|)^{\frac{r+1}{2r}}$; and the case in which $(|A||B|)^{\frac{r+1}{2r}} \leq |A|$ and thus $\mathsf{opt}(G) \leq 8r|A|$. In the later case, the subgraph consisting of a single edge adjacent to v for each node $v \in A$ will suffice to yield an approximation ratio of $\Omega(1/r)$. We now continue to study the case in which $\mathsf{opt}(G) \leq 8r(|A||B|)^{\frac{r+1}{2r}}$.

Consider the following random process in which we remove edges from G. Each edge will be removed from G independently with probability p to be defined later. Denote the resulting graph by H. Denote by $q = 1-p$ the probability that an edge is not removed.

Claim. As long as $mq \geq 16$, with probability at least $\frac{1}{2}$ the subgraph H satisfies:

- The number of edges in H is at least $mq/2$.
- The number of k cycles in H is at most $4q^{2r}m\theta^{2r-1}d_A^{r-1}d_B^{r-1}$.

Proof. The expected number of edges in H is $mq \geq 16$. Thus, using the Chernoff bound, the number of edges in H is at least half the expected value with probability $\geq 3/4$. In expectation, the number of k-cycles in H is at most $q^{2r}m\theta^{2r-1}d_A^{r-1}d_B^{r-1}$. With probability at least $3/4$ (Markov) the number of k-cycles in H will not exceed 4 times this expected value.

We now set q such that the number of k-cycles in H is at most $\frac{1}{2}$ the number of edges in H. Namely, we set q to satisfy $4q^{2r}m\theta^{2r-1}d_A^{r-1}d_B^{r-1} \leq mq/4$. Then:

$$q^{-1} = 16^{\frac{1}{2r-1}}\theta(d_Ad_B)^{\frac{r-1}{2r-1}}.$$

With this setting of parameters and our assumption that $|A||B| \geq 16\theta^2$, we have that $mq \geq 16$ and Claim 2.2 holds. Thus, we may remove an additional single edge from each remaining k-cycle in H to obtain a k-cycle-free subgraph with at least $mq/4$ edges. This is the graph our algorithm will return. To conclude our proof, we now analyze the quality of our algorithm.

We consider 2 cases. Primarily, consider the case that $m \leq 8r(|A||B|)^{\frac{r+1}{2r}}$. This implies that $(|A||B|d_Ad_B)^{\frac{1}{2}} \leq 8r(|A||B|)^{\frac{r+1}{2r}}$, which in turn implies that $d_Ad_B \leq 64r^2(|A||B|)^{\frac{1}{r}}$. Using the fact that $\mathsf{opt}(G) \leq m$ we obtain in this case an approximation ratio of

$$\begin{aligned}\frac{mq}{4\mathsf{opt}(G)} \geq \frac{q}{4} = \Omega\left(\frac{1}{\theta(d_Ad_B)^{\frac{r-1}{2r-1}}}\right) &\geq \Omega\left(\frac{1}{\theta(64r^2|A||B|)^{\frac{r-1}{r(2r-1)}}}\right)\\ = \Omega\left(\frac{1}{\theta(|A||B|)^{\frac{r-1}{r(2r-1)}}}\right).\end{aligned}$$

The second case is analyzed similarly. Assuming $m \geq 8r(|A||B|)^{\frac{r+1}{2r}}$ we get that $d_A d_B \geq 64r^2(|A||B|)^{\frac{1}{r}}$. Using the fact that $\mathsf{opt}(G) \leq 8r(|A||B|)^{\frac{r+1}{2r}}$ we obtain in this case an approximation ratio of

$$\frac{mq}{4\mathsf{opt}(G)} \geq \frac{(|A||B|d_A d_B)^{\frac{1}{2}}}{32r(|A||B|)^{\frac{r+1}{2r}} \cdot 16^{\frac{1}{2r-1}}\theta(d_A d_B)^{\frac{r-1}{2r-1}}} = \Omega\left(\frac{(d_A d_B)^{\frac{1}{2(2r-1)}}}{\theta r(|A||B|)^{\frac{1}{2r}}}\right)$$
$$= \Omega\left(\frac{1}{\theta r(|A||B|)^{\frac{r-1}{r(2r-1)}}}\right).$$

3 Proof of Theorem 4

Given an instance $J = (V_J, E_J)$ of Vertex-Cover, construct a graph $G = (V, E)$ for the 3-Cycle Transversal instance by adding to J a new node s and the edges $\{sv : v \in V_J\}$. Clearly, every edge $uv \in E_J$ corresponds to the 3-cycle $C_{uv} = \{us, sv, uv\}$ in G.

Suppose that J is 3-cycle-free. Then the set of 3-cycles of G is exactly $\{C_{uv} : uv \in E_J\}$. The following statement implies that w.l.o.g. we may consider only 3-cycle transversals that consist from edges incident to s.

Claim. Suppose that J is 3-cycle-free. Let F be a 3-cycle transversal in G and let $uv \in F \cap E_J$. Then $F - uv + su$ is also a 3-cycle transversal in G. Thus there exists a 3-cycle transversal $F' \subseteq \{sv : v \in V_J\}$ in G with $|F'| \leq |F|$.

Proof. The only 3-cycle in G that is covered by the edge uv is C_{uv}. This cycle is also covered by the edge su.

Claim. Suppose that J is 3-cycle-free. Then $U \subseteq V_J$ is a vertex-cover in J if, and only if, the edge set $F_U = \{su : u \in U\}$ is a k-cycle transversal in G.

Proof. We show that if $U \subseteq V_J$ is a vertex-cover in J then F_U is a 3-cycle transversal in G. Let C_{uv} be a 3-cycle in G. As U is a vertex-cover, $u \in U$ or $v \in U$. Thus $su \in F_U$ or $sv \in F_U$. In both cases, $C_{uv} \cap F_U \neq \emptyset$.

We now show that if F_U is a 3-cycle transversal in G, then U is a vertex-cover in J. Let $uv \in E_J$. Then C_{uv} is a 3-cycle in G, and thus $su \in F_U$ or $sv \in F_U$. This implies that $u \in U$ or $v \in U$, namely, the edge uv is covered by U.

From the claims above it follows that an α-approximation for 3-Cycle Transversal on G implies an α-approximation for Vertex-Cover on 3-cycle-free graphs J. Now we prove (for completeness, as we did not find an appropriate reference):

Claim. Any approximation algorithm with ratio $\alpha \geq 3/2$ for Vertex-Cover on 3-cycle-free graphs implies an α-approximation algorithm for Vertex-Cover (on general graphs).

Proof. Suppose that there is an α-approximation algorithm for Vertex-Cover on 3-cycle-free graphs. Let J be a general graph, and let $\mathsf{opt}(J)$ be the size of its minimum vertex cover. Consider the following two phase algorithm. Phase 1 starts with an empty cover F_1, and repeatedly, for every 3-cycle C in J, adds the nodes of C to F_1 and deletes them from J. Note that any vertex-cover contains at least two nodes of C, which implies a "local ratio" of 2/3. Let J_2 be the triangle free graph obtained after Phase 1. In Phase 2 use the α-approximation algorithm (for 3-cycle-free graphs) to compute a vertex-cover F_2 of J_2. The statement follows since: $\mathsf{opt}(J) \geq \frac{2}{3}|F_1| + \mathsf{opt}(J_2) \geq \frac{2}{3}|F_1| + \frac{|F_2|}{\alpha} \geq \frac{|F_1|+|F_2|}{\alpha}$.

We now prove part (ii) of the theorem, namely, that for $k = 3$ the integrality gap of (1) is at least $2 - \varepsilon$. We will use the fact that for any $\varepsilon > 0$, there exist infinitely many graphs $J = (V_J, E_J)$ which are 3-cycle-free and have minimum vertex-cover of size at least $|V_J|(1 - \frac{\varepsilon}{2})$. Such graphs appear in various places in the literature. For example see Theorem 1.2 in [7] in which 3-cycle-free graphs J with independence number at most $\frac{\varepsilon}{2}|V_J|$ are presented. For such graph J, the minimum k-cycle cover in the corresponding graph G has size at least $|V_J|(1-\frac{\varepsilon}{2})$. On the other hand, the solution $x_e = 1/2$ if e is incident to s and $x_e = 0$ otherwise is a feasible solution to LP (1) on G with value $|V_J|/2$. Hence the integrality gap is at least $\frac{(1-\frac{\varepsilon}{2})}{1/2} = 2 - \varepsilon$.

Theorem 4 easily extends to arbitrary $k \geq 4$. We use the same construction as for the case $k = 3$, but in addition subdivide every edge of J by $k - 3$ nodes (and do not make any assumptions on J). Hence every edge $uv \in E_J$ is replaced by a path P_{uv} of the length $k - 2$, and $C_{uv} = P_{uv} + su + sv$ is a k-cycle in G. Since $k \geq 4$, G has no other k-cycles, namely, the set of k-cycles in G is $\{C_{uv} = P_{uv} + su + sv : uv \in E_J\}$. The rest of the proof of this case is identical to the case $k = 3$, and thus is omitted.

4 Proof of Theorem 5

To prove Theorem 5, we prove two theorems that consider a more general setting of a family $\mathcal{F}$ of subgraphs of G which are not necessarily k-cycles, nevertheless each subgraph $C \in \mathcal{F}$ is of size $\leq k$. We need some definitions. Let G be a graph and let $\mathcal{F}$ be a family of subgraphs (edge subsets) of G. For a subgraph H of G, let $\mathcal{F}(H)$ be the restriction of $\mathcal{F}$ to subgraphs of H; H is *$\mathcal{F}$-free* if $\mathcal{F}(H) = \emptyset$. An edge set F that intersects every member of $\mathcal{F}$ is an *$\mathcal{F}$-transversal.* We consider the following two problems, that generalize the problems k-Cycle-Free Subgraph and k-Cycle Transversal. The instance of the problems is a graph $G = (V, E)$ and a family $\mathcal{F}$ of subgraphs of G. The goal is:

$\mathcal{F}$-Transversal: Find a minimum size $\mathcal{F}$-transversal.

$\mathcal{F}$-Free Subgraph: Find a maximum size $\mathcal{F}$-free subgraph of G.

For $\mathcal{F} = \mathcal{C}_k(G)$, we get the problems k-Cycle Transversal and k-Cycle Free Subgraph, respectively. Let $\tau^*_{\mathcal{F}}(H)$ denote the optimal value of the following LP-relaxation for $\mathcal{F}$-Transversal on H:

$$
\begin{aligned}
&\min \textstyle\sum_{e\in E(H)} x_e && && (2)\\
&\text{s.t. } \textstyle\sum_{e\in C} x_e \geq 1 && \forall C \in \mathcal{F}(H)\\
&\qquad x_e \geq 0 && \forall e \in E(H)
\end{aligned}
$$

An edge of H is *$\mathcal{F}$-redundant* if no member of $\mathcal{F}(H)$ contains it; e.g., if $\mathcal{F} = \mathcal{C}_k(G)$, then an edge of H is $\mathcal{F}$-redundant if it is not contained in any k-cycle of H. We prove:

Theorem 7. *Suppose that any subgraph H of G admits a polynomial time algorithm that:* (i) *Solves LP* (2) *for H;* (ii) *Finds $\mathcal{F}$-redundant edges of H;* (iii) *Finds an $\mathcal{F}(H)$-transversal of size at most $|E(H)| \cdot (k-1)/k$. Then there exist a polynomial time algorithm that finds an $\mathcal{F}(G)$-transversal of size $\leq (k-1) \cdot \tau^*_{\mathcal{F}}(G)$.*

To prove Theorem 5(ii) we connect the approximation of $\mathcal{F}$-Free Subgraph and $\mathcal{F}$-Transversal by the following theorem:

Theorem 8. *Suppose that for any graph G with m edges there exist a polynomial algorithm that finds an $\mathcal{F}(G)$-free subgraph of size $\geq \beta m$, and that* $\mathcal{F}$-Transversal *admits an α-approximation algorithm. Then* k-Cycle-Free Subgraph *admits an $\alpha\beta/(\alpha+\beta-1)$-approximation algorithm.*

Let us now show that Theorem 7 implies Theorem 5(i) and that Theorem 8 implies Theorem 5(ii). Let G be a graph with m edges. As was mentioned, it is not hard to find in G a subgraph with at least $m/2$ edges and without odd cycles. For Theorem 5(i), it is easy to see that this setting obeys the conditions of Theorem 7, hence we obtain a $(k-1)$-approximation for $\mathcal{F}$-Transversal in this case. For Theorem 5(ii), we apply Theorem 8 with $\beta = 1/2$ and $\alpha = k-1$. The ratio obtained is $\alpha\beta/(\alpha+\beta-1) = (k-1)/(2k-3) = \frac{1}{2} + \frac{1}{4k-6}$. We now prove Theorems 7 and 8 (in Sections 4.1 and 4.2, respectively).

4.1 Proof of Theorem 7

The algorithm is as follows:

Initialization: $H \leftarrow G$; $F_1 \leftarrow \emptyset$.
Phase 1:
While for an optimal solution x to (2) $x_e \geq 1/(k-1)$ for some $e \in E(H)$ do:
$\quad F_1 \leftarrow F_1 + e$; $H \leftarrow H - e$.
EndWhile
Phase 2:
- Remove all $\mathcal{F}(H)$-redundant edges from H. Denote the resulting graph by H_2.
- Compute an $\mathcal{F}(H_2)$-transversal F_2 of size at most $|E(H_2)| \cdot (k-1)/k$.
Return $F_1 \cup F_2$.

Under the assumptions of the Theorem, all steps can be implemented in polynomial time. It is also easy to see that the algorithm returns a feasible solution. We now analyze the approximation ratio. We start with a simple claim followed by our key Lemma.

Claim. Let H be the graph obtained after Phase 1 of our algorithm and let x_e be an optimal solution to *LP (2)* on H. Then $x_e = 0$ for every $\mathcal{F}(H)$-redundant edge e in H. Thus the restriction of x to H_2 is also an optimal solution to LP (2) on H_2.

Proof. Let e be an $\mathcal{F}(H)$-redundant edge. Assume for sake of contradiction that $x_e > 0$. We can now reduce the value of the LP solution by zeroing out x_e. The new solution is still valid, as e is $\mathcal{F}(H)$-redundant and thus does not appear in the first family of constraints of (2).

Let H_2 be obtained from H by removing all $\mathcal{F}(H)$-redundant edges. Then the restriction of x to H_2 is an optimal solution to (2) since any LP solution for H_2 can be extended to one for H by setting $x_e = 0$ for every $\mathcal{F}(H)$-redundant edge e.

Using the claim above, we may assume that the subgraph H_2 has an optimal solution x to (2) in which $x_e < 1/(k-1)$ (for all $e \in E(H_2)$).

Lemma 4. *Let H_2 be a subgraph of G without $\mathcal{F}$-redundant edges and let x be an optimal solution to LP (2). If $x_e < 1/(k-1)$ for every $e \in E(H_2)$ then $\tau^*_{\mathcal{F}}(H_2) \geq |E(H_2)|/k$.*

Proof. Let $\nu^*_{\mathcal{F}}(H_2) = \tau^*_{\mathcal{F}}(H_2)$ denote the optimal value of the dual LP:

$$\begin{aligned} &\max \textstyle\sum_{C \in \mathcal{F}} y_C \\ &\text{s.t.} \textstyle\sum_{C \ni e} y_C \leq 1 \quad \forall e \in E(H_2) \\ &\qquad y_C \geq 0 \qquad\quad \forall C \in \mathcal{F}(H_2) \end{aligned} \tag{3}$$

Let x and y be optimal solutions to (2) and to (3), respectively. Consider two cases, after noting that the primal complementary slackness condition is:

$$x_e > 0 \implies \sum_{C \ni e} y_C = 1 \tag{4}$$

Case 1: $x_e > 0$ for every $e \in E(H_2)$.
In this case $\tau^*_{\mathcal{F}}(H) \geq |E(H_2)|/k$, since from (4) we get:

$$|E(H_2)| = \sum_{e \in E(H_2)} 1 = \sum_{e \in E} \sum_{C \ni e} y_C = \sum_{C \in \mathcal{F}(H_2)} |C| y_C \leq \sum_{C \in \mathcal{F}(H_2)} k y_C = k\nu^*_{\mathcal{F}}(H_2) = k\tau^*_{\mathcal{F}}(H_2)\,.$$

Case 2: $x_f = 0$ for some $f \in E(H_2)$.
Since H_2 has no $\mathcal{F}$-redundant edges, there is $C \in \mathcal{F}(H_2)$ so that $f \in C$. Since $x_f = 0$, we have $\sum_{e \in C - f} x_e \geq 1$. Since $|C - f| \leq k - 1$, there exists $e \in C - f$ so that $x_e \geq 1/(k-1)$. A contradiction.

We now bound the value of $|F_1|$ and $|F_2|$ with respect to $\tau^*_{\mathcal{F}}(G)$. We start with some notation. Let $H^0 = G$ be the starting point of our algorithm. Let H^1 be graph obtained from H^0 by the removal of e_1 after the first round of Phase 1. Similarly, for the i'th round of Phase 1, let H^i be the graph obtained from H^{i-1} by the removal of e_i. Let $H = H^\ell$ be the graph obtained after Phase 1 of our algorithm (here ℓ denotes the number of rounds in Phase 1). It is not hard to verify that $\tau^*_{\mathcal{F}}(H^{i-1}) \geq \tau^*_{\mathcal{F}}(H^i) + x_{e_i}$. Here x_{e_i} is obtained from the optimal solution to H^{i-1}. This implies that $\tau^*_{\mathcal{F}}(G) \geq \tau^*_{\mathcal{F}}(H) + \sum_{i=1}^{\ell-1} x_{e_i}$.

Now to bound $|F_1|$ and $|F_2|$. First notice that $|F_1| \leq (k-1)\sum_{i=1}^{\ell-1} x_{e_i}$. Recall that H_2 is the graph obtained in Phase 2 from H by removing all $\mathcal{F}(H)$-redundant edges. It also holds that, $|F_2| \leq |E(H_2)| \cdot (k-1)/k$. By Lemma 4, $\tau^*_{\mathcal{F}}(H_2) \geq |E(H_2)|/k$. Hence

$$\frac{|F_2|}{\tau^*_{\mathcal{F}}(H_2)} \leq \frac{|E(H_2)| \cdot (k-1)/k}{|E(H_2)|/k} = k-1 \ .$$

As by Claim 4.1, $\tau^*_{\mathcal{F}}(H) = \tau^*_{\mathcal{F}}(H_2)$ we have that

$$|F_1| + |F_2| \leq (k-1)(\tau^*_{\mathcal{F}}(H) + \sum_{i=1}^{\ell-1} x_{e_i}) \leq (k-1)\tau^*_{\mathcal{F}}(G) \ ,$$

which concludes our proof.

4.2 Proof of Theorem 8

In what follows let opt be the optimal solution value of the $\mathcal{F}$-$\mathsf{Free\ Subgraph}$ problem on G. We choose the better result F from the following two algorithms:

Algorithm 1: Find an $\mathcal{F}(G)$-free subgraph of size $\geq \beta m$.
Algorithm 2: Find an $\mathcal{F}(G)$-transversal I of size $\leq \alpha$ times an optimal $\mathcal{F}(G)$-transversal, and return $G - I$.

Algorithm 1 computes a solution of size $\geq \beta m$. Algorithm 2 computes a solution of size $\geq m - \alpha(m - \mathsf{opt})$. The worse case is when these lower bounds coincide: $\beta m = m - \alpha(m - \mathsf{opt})$ which implies $\mathsf{opt} = m(\alpha + \beta - 1)/\alpha$. This gives the ratio $\frac{\beta m}{m(\alpha+\beta-1)/\alpha} = \frac{\alpha\beta}{\alpha+\beta-1}$. Formally, $|F| \geq \max\{\beta m, m - \alpha(m - \mathsf{opt})\}$. Consider two cases:

Case 1: $\beta m \geq m - \alpha(m - \mathsf{opt})$, so $\mathsf{opt} \leq m(\alpha + \beta - 1)/\alpha$. Then

$$\frac{|F|}{\mathsf{opt}} \geq \frac{\beta m}{\mathsf{opt}} \geq \frac{\beta m}{(\alpha+\beta-1)/\alpha} = \frac{\alpha\beta}{\alpha+\beta-1} \ .$$

Case 2: $m - \alpha(m - \mathsf{opt}) \geq \beta m$, so $m/\mathsf{opt} \leq \alpha/(\alpha + \beta - 1)$. Then

$$\frac{|F|}{\mathsf{opt}} \geq \frac{m - \alpha(m - \mathsf{opt})}{\mathsf{opt}} = \alpha - (\alpha - 1) \cdot \frac{m}{\mathsf{opt}} \geq \alpha - (\alpha - 1) \cdot \frac{\alpha}{\alpha+\beta-1} = \frac{\alpha\beta}{\alpha+\beta-1} \ .$$

In both cases the ratio is bounded by $\frac{\alpha\beta}{\alpha+\beta-1}$, which concludes our proof.

5 Open problems

For k-Cycle Transversal, we have ratios $k-1$ for odd values of k and k for even values of k. However, the best approximation threshold we have is 2. Closing this gap (even for $k=4,5$) is left open.

For k-Cycle-Free Subgraph, we have ratios $2/3$ for $k=3$ and $n^{-1/3-\varepsilon}$ for $k=4$. The best approximation threshold we have is APX-hardness. Hence, we do not even know if our ratio of $2/3$ for $k=3$ is tight. Our result for $k=3$ actually establishes a lower bound of $2/3$ on the integrality gap for the natural LP for 3-Cycle-Free Subgraph, but the best upper bound we have is only $3/4$. Finally, in our opinion, the most challenging open question is closing the huge gap for the case $k=4$.

References

1. N. Alon. Bipartite subgraphs. *Combinatorica*, 16:301–311, 1996.
2. N. Alon, B. Bollobás, M. Krivelevich, and B. Sudakov. Maximum cuts and judicious partitions in graphs without short cycles. *J. of Comb. Th. B*, 88(2):329–346, 2003.
3. N. Alon, R. Yuster, and U. Zwick. Finding and counting given length cycles. *Algorithmica*, 17(3):209–223, 1997.
4. J. A. Bondy and M. Simonovits. Cycles of even lengths in graphs. *J. Comb. Th. B*, 16:97105, 1974.
5. P. Erdös. Extremal problems in graph theory. In M.Fiedler, editor, *Theory of Graphs and Its Applications*. Academic Press, New York, 1965.
6. P. Erdös, T. Gallai, and Z. Tuza. Covering and independence in triangle structures. *Discrete Mathematics*, 150:89–101, 1996.
7. U. Feige, M. Langberg, and G. Schechtman. Graphs with tiny vector chromatic numbers and huge chromatic numbers. *SIAM J. Comput.*, 33(6):1338–1368, 2004.
8. Z. Furedi, A. Naor, and J. Verstraëte. On the Turan number of the hexagon. *Advances in Mathematics*, 203(2):476–496, 2006.
9. M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42(6):1115–1145, 1995.
10. S. Hoory. The size of bipartite graphs with a given girth. *Journal of Combinatorial Theory Series B*, 86(2):215–220, 2002.
11. J. Komlós. Covering odd cycles. *Combinatorica*, 17(3):393–400, 1997.
12. M. Krivelevich. On a conjecture of Tuza about packing and covering of triangles. *Discrete Mathematics*, 142(1-3):281–286, 1995.
13. D. Kühn and D. Osthus. Four-cycles in graphs without a given even cycle. *Journal of Graph Theory*, 48(2):147–156, 2005.
14. T. Lam. A result on $2k$-cycle free bipartite graphs. *Australasian Journal of Combinatorics*, 32:163–170, 2005.
15. T. Lam and J. Verstraëte. A note on graphs without short even cycles. *The Electronic Journal of Combinatorics*, 12(1,N5), 2005.
16. F. Lazebnik, V. A. Ustimenko, and A. J. Woldar. Polarities and $2k$-cycle free graphs. *Discrete Math.*, 197/198:503–513, 1999.
17. A. Naor and J. Verstraëte. A note on bipartite graphs without $2k$-cycles. *Probability, Combinatorics and Computing*, 14(5-6):845–849, 2005.
18. P. A. Pevzner, H. Tang, and G. Tesler. De novo repeat classification and fragment assembly. *Genome Research*, 14(9):1786–1796, 2004.