Approximation Algorithms for Cycle Packing Problems *

Michael Krivelevich † *Zeev Nutov* ‡ *Raphael Yuster* §

Abstract

The *cycle packing number* $\nu_c(G)$ of a graph G is the maximum number of pairwise edge-disjoint cycles in G. Computing $\nu_c(G)$ is an NP-hard problem. We present approximation algorithms for computing $\nu_c(G)$ in both the undirected and directed cases. In the undirected case we analyze a variant of the modified greedy algorithm suggested in [4] and show that it has approximation ratio $O(\sqrt{\log n})$ where $n = |V(G)|$, and this is tight. This improves upon the previous $O(\log n)$ upper bound for the approximation ratio of this algorithm. In the directed case we present a $\sqrt{n}$-approximation algorithm. Finally, we give an $O(n^{2/3})$-approximation algorithm for the problem of finding a maximum number of edge-disjoint cycles that intersect a specified subset S of vertices. Our approximation ratios are the currently best known ones and, in addition, provide bounds on the *integrality gap* of standard LP-relaxations to these problems.

Key words. cycle, packing, integrality gap, approximation algorithm

AMS subject classifications. 68W25, 68R10, 90C27, 90C35, 05C38

1 Introduction

We consider the following fundamental problem in Algorithmic Graph Theory. Given a graph (digraph) G, how many edge-disjoint (directed) cycles can be packed into G? Define the *cycle packing number* $\nu_c(G)$ of G to be the maximum size of a set of edge-disjoint cycles in G. The *maximum cycle packing problem* is to find a set of $\nu_c(G)$ edge-disjoint cycles in G. Problems concerning packing edge-disjoint or vertex-disjoint cycles in graphs and digraphs have been studied extensively (see, e.g., [1, 4, 9]).

It is well known that computing $\nu_c(G)$ (and hence finding a maximum cycle packing) is an NP-hard problem in both the directed and undirected cases. Even the very special case of deciding whether a graph (digraph) has a triangle decomposition is known to be NP-Complete (see, e.g. [6] for a more general theorem on the NP-Completeness of such decomposition problems). Thus, approximation algorithms are of interest. A *ρ-approximation algorithm* for a maximization problem is a polynomial time algorithm that produces a solution of value at least $1/\rho$ times the value of an optimal solution; ρ is called the *approximation ratio* of the algorithm.

*A preliminary version of this paper appeared in *Proceedings of the 16th Annual ACM-SIAM Syposium on Discrete Algorithms*, Vancouver, Canada, 2005.

†Department of Mathematics, Tel Aviv University, Tel Aviv, Israel. E–mail: krivelev@tau.ac.il

‡Department of Computer Science, The Open University of Israel, Tel Aviv, Israel. E–mail: nutov@openu.ac.il

§Department of Mathematics, University of Haifa, Haifa 31905, Israel. E–mail: raphy@research.haifa.ac.il

A recent result of Caprara, Panconesi, and Rizzi [4] shows that by slightly modifying the greedy algorithm one obtains an $O(\log n)$-approximation algorithm for the undirected maximum cycle packing problem. We get an $O(\sqrt{\log n})$-approximation algorithm by combining this modified greedy algorithm with an ordinary greedy algorithm. In particular, we obtain the following result.

Theorem 1.1 *There exists an $O(\sqrt{\log n})$ approximation algorithm for the undirected maximum cycle packing problem.*

We also prove that the approximation guarantee of the algorithm is $\Omega(\sqrt{\log n})$. The approximation ratio in Theorem 1.1 is currently the best known one for the maximum undirected cycle packing problem.

Our next two results are for directed graphs.

Theorem 1.2 *There exists a $\sqrt{n}$-approximation algorithm for the directed maximum cycle packing problem.*

The algorithms in Theorems 1.1 and 1.2 are easily adjusted to the capacitated version of the problems as well. For simplicity of exposition, we prove our results for the uncapacitated case, and then show how they extend to the capacitated case.

Finally, we consider the *maximum S-cycle packing* problem in directed graphs: given a directed graph G and a subset S of its vertices, find among the cycles that intersect S (henceforth, *S-cycles*) a maximum number $\nu_c(G,S)$ of edge-disjoint ones. We note that on directed simple graphs, the maximum S-cycle packing problem is a special case of the extensively studied *edge-disjoint paths* problem. See [5] for an $O(n^{4/5})$-approximation algorithm and [10] for an $O(n^{2/3}\log^{2/3} n)$-approximation algorithm for the edge-disjoint paths problem in directed graphs.

Theorem 1.3 *There exists an $O(n^{2/3})$-approximation algorithm for the directed maximum S-cycle packing problem on simple digraphs.*

Given a graph $G = (V,E)$, the *fractional cycle packing* in G is a function ψ from a subset $\mathcal{C}$ of cycles in G to $[0,1]$ satisfying $\sum_{e\in C\in\mathcal{C}}\psi(C) \le 1$ for each $e \in E$. Letting $|\psi| = \sum_{C\in\mathcal{C}}\psi(C)$, the *fractional cycle packing number* $\nu_c^*(G)$ of G is defined to be the maximum of $|\psi|$ taken over all fractional cycle packings ψ in G. The *cycle cover number* $\tau_c(G)$ of G is the minimum number of edges whose deletion makes G acyclic. Clearly, $\nu_c(G) \le \nu_c^*(G) \le \tau_c(G)$ for any graph/digraph G.

The approximation ratios in Theorems 1.1, 1.2, and 1.3 provide bounds on the integrality gap of the standard LP-relaxations to the problems. Specifically, each of the algorithms computes a packing $\mathcal{C}$ so that: $|\mathcal{C}|/\nu_c^*(G) = \Omega(1/\sqrt{\log n})$ in Theorem 1.1, $|\mathcal{C}|/\nu_c^*(G) \ge 1/\sqrt{n}$ in Theorem 1.2, and $|\mathcal{C}|/\tau_c(G,S) = \Omega(n^{-2/3})$ in Theorem 1.3, where $\tau_c(G,S)$ is the minimum number of edges needed to cover all S-cycles in G.

In the following three sections we prove Theorems 1.1, 1.2, and 1.3, respectively.

2 Proof of Theorem 1.1

As was mentioned in the introduction, we combine the Modified Greedy Algorithm suggested by Caprara et al. in [4] with the (ordinary) Greedy Algorithm. The algorithm of [4] starts with $\mathcal{C} = \emptyset$ and performs iteratively, until there are no edges left in G, the following three steps:

1. While G contains a vertex v of degree ≤ 1, delete v (and the edge incident to v, if exists).
2. While G contains a vertex v of degree 2 with neighbors v' and v'', delete v and edges vv', vv'' and replace them by a new edge $v'v''$.
3. Find a shortest cycle C in G, add C to the constructed solution $\mathcal{C}$ and remove its edges from G.

Our algorithm is as follows. At Phase 1, while $\text{girth}(G) \leq \sqrt{\log |G|}$ we apply the Modified Greedy Algorithm (where $|G|$ is the number of vertices in the *current* graph); the condition $\text{girth}(G) \leq \sqrt{\log |G|}$ is checked after Step 2. Phase 2 starts when $\text{girth}(G) > \sqrt{\log |G|}$ after Step 2 for the first time; then we repeatedly apply Step 3 only, which is the (ordinary) Greedy Algorithm.

2.1 The approximation ratio

Theorem 2.1 *The algorithm computes a cycle packing of size $\Omega(\nu_c^*(G)/\sqrt{\log n})$. Thus it is an $O(\sqrt{\log n})$-approximation algorithm.*

In the proof we use the following lemma, which gives an improved analysis of the performance of the greedy algorithm on graphs with large girth.

Lemma 2.2 *Let H be a graph with n nodes, $m > n$ edges, and girth g, and let $\mathcal{C}_H$ be the set of cycles found in H by the greedy algorithm. Then*

$$|\mathcal{C}_H| \geq \frac{(m-n)^2}{4m(m+n)} \cdot \frac{g}{\log(m-n)} \cdot \nu^*(H) \ .$$

In particular, if $m \geq (1+\epsilon)n$ for an $\epsilon > 0$ then

$$|\mathcal{C}_H| \geq \frac{\epsilon^2}{4(1+\epsilon)(2+\epsilon)} \cdot \frac{g}{\log(\epsilon n)} \cdot \nu^*(H) \ .$$

Proof: Consider the steps of the Greedy Algorithm when $|E(H)| \geq (m+n)/2 = n + (m-n)/2$. Bollobás and Thomason [3] proved that if a graph H sastisfies $|E(H)| \geq |H| + c$ for a $c > 0$, then $\text{girth}(J) \leq 2(1 + |H|/c)\log(2c)$. Thus during all these steps

$$\text{girth}(H) \leq \frac{2(m+n)}{m-n} \log(m-n) \ .$$

The number of edges deleted during these steps is at least $(m-n)/2$. Clearly, $\nu^*(H) \leq m/g$. Thus

$$|\mathcal{C}_H| \geq \frac{(m-n)^2}{4(m+n)\log(m-n)} \geq \frac{(m-n)^2}{4(m+n)\log(m-n)} \cdot \frac{g}{m} \nu^*(H) \ .$$

The second statement follows from the first by observing that the function $f(m) = \frac{(m-n)^2}{4m(m+n)}$ is increasing for $m \geq n$. $\square$

Completing the proof of Theorem 2.1: Note that Steps 1 and 2 of the Modified Greedy Algorithm do not change the value of an optimal solution. Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be the set of cycles added to the packing during Phase 1 and Phase 2, respectively. Recall that in Phase 1 we execute the Modified Greedy Algorithm, and the length of every added cycle does not exceed $\sqrt{\log |G|}$ (here $|G|$ is the number of vertices in the *current* graph). Phase 2 starts when a cycle added to an approximate packing has length more than $\sqrt{\log |G|}$, and executes the Greedy Algorithm. Fix an optimal fractional packing ψ^*, so $|\psi^*| = \nu_c^*(G)$. Let ψ_1^* be the restriction of ψ^* to cycles that intersect some cycle from $\mathcal{C}_1$, $\psi_2^* = \psi^* - \psi_1^*$. Since every cycle from $\mathcal{C}_1$ has length $\leq \sqrt{\log n}$

$$|\psi_1^*| \leq |\mathcal{C}_1|\sqrt{\log n}\ .$$

We claim that

$$|\psi_2^*| \leq 60|\mathcal{C}_2|\sqrt{\log n}\ .$$

Thus

$$|\psi^*| = |\psi_1^*| + |\psi_2^*| \leq \sqrt{\log n}(|\mathcal{C}_1| + 60|\mathcal{C}_2|) \leq 60\sqrt{\log n}|\mathcal{C}|\ .$$

We prove that $|\psi_2^*| \leq 60|\mathcal{C}_2|\sqrt{\log n}$ using Lemma 2.2. Let H be the graph the second phase starts with. Then $\mathrm{girth}(H) \geq \sqrt{\log |H|}$ (recall that $|H|$ is the number of vertices in H), and H has at least $3|H|/2$ edges (since H has minimum degree at least 3). Thus, by substituting $\epsilon = 1/2$ in the bound in Lemma 2.2 we get:

$$|\mathcal{C}_2| \geq \frac{1}{60} \cdot \frac{\mathrm{girth}(H)}{\log(|H|/2)} \cdot \nu^*(H) \geq \frac{1}{60} \cdot \frac{\sqrt{\log |H|}}{\log(|H|/2)} \cdot \nu^*(H) \geq \frac{1}{60} \cdot \frac{\sqrt{\log |H|}}{\log |H|} \cdot \nu^*(H) = \frac{\nu^*(H)}{60\sqrt{\log |H|}}\ .$$

Clearly, $|H| \leq n$. Also, $|\psi_2^*| \leq \nu^*(H)$, since ψ_2^* corresponds to a fractional packing in H. This implies $|\psi_2^*| \leq \nu^*(H) \leq 60\sqrt{\log |H|} \leq 60\sqrt{\log n}$, as claimed. $\square$

Remark: Our algorithm is easily adjusted to the capacitated version of the problem, where we are also given integral capacities $\{c_e : e \in E\}$ on the edges; the goal is to find a maximum weight family $\mathcal{C}$ of cycles so that for every edge $e \in E$ the capacity constraints $\sum\{w(C) : e \in C, C \in \mathcal{C}\} \leq c(e)$ are satisfied. We can immitate the capacitated case by the uncapacitated one if we replace every edge e by c_e parallel edges with the same ends as e, and "forbid" the arising cycles of the length 2. However, this will give only a pseudopolynomial time algorithm. To get a polynomial algorithm, let us show how to adjust steps 1, 2, and 3 in the modified greedy algorithm to handle this case. Each time an edge of capacity zero arises, it is deleted, so assumme that G has no zero capacity edges. Step 1 remains the same. In Step 2, the new edge $v'v''$ gets capacity $\min\{c_{vv'}, c_{vv''}\}$. In step 3, after the shortest cycle C is found, we also find the minimum capacity edge e in C, add C to the constructed packing, and assign it weight $w(C) = c_e$. Then, in G, we reduce by c_e the capacities along the edges of C, and remove the arising zero capacity edges. Each one of the steps can be performed in polynomial time, and leads to a graph with less edges. Thus the running time is polynomial. It is easy to see that our analysis of the approximation ratio is valid for the capacitated case as well. The algorithm in Section 3 for directed graphs (to follow) admits a similar adjustment.

2.2 A tight example

Theorem 2.3 *The approximation ratio of the Modified Greedy Algorithm is* $\Omega(\sqrt{\log n})$.

For the proof we will need the following technical lemma.

Lemma 2.4 *Let G be a graph on n vertices of maximal degree at most 7. Let $V_0 \subseteq V(G)$. If $|V_0| \geq n/2$ then there exists a subset $U \subset V_0$ of size $|U| = \lceil \log n \rceil$, such that all vertices of U are at distance more than $\frac{1}{3}\log n$ from each other.*

Proof: Note that every vertex $v \in G$ is at distance at most k from at most $7 \cdot 6^{k-1} < 7^k$ vertices from G. Define an auxiliary edge set E_0 on V_0 so that $(u,v) \in E_0$ if $dist_G(u,v) \leq \frac{1}{3}\log n$. Let $H = (V_0, E_0)$. Then H is a graph on at least $n/2$ vertices of maximal degree $\Delta(H) < 7^{\frac{1}{3}\log n} < n^{0.95}$, and has therefore an independent set U of size at least $|V(H)|/(1+\Delta(H)) > \log n$. Each such independent set gives a required set of vertices in G. □

The *k-sunflower* S^k is a cycle of length k (the *core cycle*) to each edge of which we attach a cycle of length $k+1$ (a *petal*), so that the petals are vertex-disjoint outside the core cycle. The number of vertices of S^k is k^2. Observe that the core is the shortest cycle in a k-sunflower, and removing its edges results in a cycle on k^2 vertices. We choose $k = \sqrt{\log n/3}$ and denote $t = k^2$ (we ignore floors and ceilings as they do not affect the asymptotic nature of our result).

Let now G_0 be a 3-regular graph on n vertices of girth more than $t = \frac{1}{3}\log n$. Such graphs exist for infinitely many values of n as proved by Erdős and Sachs [7]. We start with $G = G_0$, set $W = \emptyset$, $i = 1$, and repeat $n/(2t)$ times the following procedure:

1. Find a subset $U_i \subset V \setminus W$ such that $|U_i| = t$ and all vertices of U_i are at distance more than $\frac{1}{3}\log n$ from each other in G;
2. Insert a copy S_i of the k-sunflower in U_i, placing it arbitrarily within U_i;
   update $W \leftarrow W \cup U_i$; $i \leftarrow i+1$.

Since the sets U_i are disjoint and the maximum degree of S^k is 4, the graph G has maximum degree at most 7 during the execution of the above procedure. Also, $|W| \leq \frac{n}{2t} \cdot t = \frac{n}{2}$, and therefore finding a required U_i at each step is possible due to Lemma 2.4. Let us denote by G^* the final graph of the above procedure.

Claim 2.1 *Let C be a cycle of length at most $\frac{1}{3}\log n$ in G^*. Then C is a cycle in one of the inserted k-sunflowers S_i.*

Proof: Since girth$(G_0) > \frac{1}{3}\log n$, C contains an edge $e \in E(G^*) - E(G_0)$. Let $i^* = \max\{i : E(C) \cap E(S_i) \neq \emptyset\}$. We claim that C is a cycle in S_{i^*}. Let G_{i^*} be the graph created during the above described procedure after having inserted the sunflower S_{i^*}. Obviously, $C \subset G_{i^*}$. If $E(C) \subset E(S_{i^*})$ we are done. Assume otherwise. Since U_{i^*} spans only the edges of S_{i^*} in G_{i^*}, at some point C leaves U_{i^*} and then returns back. Let $u_1, u_2 \in U_{i^*}$ be the vertices of U_{i^*} where C leaves and reenters U_{i^*}. By our choice of U_{i^*}, $dist_{G_{i^*}}(u_1, u_2) > \frac{1}{3}\log n$, implying $|C| > \frac{1}{3}\log n$, a contradiction. $\square$

Completing the proof of Theorem 2.3: We analyze the performance of the modified greedy algorithm on G^*. By Claim 2.1, the shortest cycles in G^* are the $n/(2t) = O(n/\log n)$ core cycles of the inserted sunflowers, which are vertex-disjoint. Hence the algorithm starts by picking all of them. After all core cycles have been removed, none of the sunflowers contains a cycle of length at most $\frac{1}{3}\log n$, and applying Claim 2.1 again we infer that the modified greedy algorithm will be able to add at most $|E(G^*)|/(\log n/3) = O(n/\log n)$ cycles, altogether ending up with $O(n/\log n)$ cycles. On the other hand, a feasible solution can be obtained by taking all petals of all inserted sunflowers, whose total number is $(n/(2t)) \cdot k = \Theta(n/\sqrt{\log n})$. It follows that the approximation ratio of the modified greedy on G^* is

$$\Omega\left(\frac{\frac{n}{\sqrt{\log n}}}{\frac{n}{\log n}}\right) = \Omega(\sqrt{\log n}) \ .$$

$\square$

3 Proof of Theorem 1.2

It will be convenient to describe the algorithm with a certain parameter ℓ, which will be eventually set to $\ell = \sqrt{n}$. The algorithm starts with $\mathcal{C}_1, \mathcal{C}_2 = \emptyset$ and in the end outputs $\mathcal{C}_1 \cup \mathcal{C}_2$.

Phase 1:

As long as there is a directed cycle of length $\leq \ell$, find such a cycle, add it to $\mathcal{C}_1$, and delete its edges from the graph.

Phase 2:

For each $v \in V$, compute a maximum size set $\mathcal{C}_2(v)$ of edge-disjoint directed cycles that contain v. Among the packings computed, let $\mathcal{C}_2$ be one of maximal size.

Theorem 3.1 *For $\ell = \sqrt{n}$ the algorithm computes a packing $\mathcal{C}_1 \cup \mathcal{C}_2$ of size at least $\nu_c^*(G)/\sqrt{n}$.*

Proof: As in the proof of Theorem 2.1, let us fix an optimal fractional packing ψ^*, let ψ_1^* be the restriction of ψ^* to cycles that intersect some cycle from $\mathcal{C}_1$, $\psi_2^* = \psi^* - \psi_1^*$. Since every cycle from $\mathcal{C}_1$ has length $\leq \ell$ we have $|\psi_1^*| \leq \ell|\mathcal{C}_1|$. We claim that $|\mathcal{C}_2| \geq \ell|\psi_2^*|/n$. Thus by combining the bounds for $|\mathcal{C}_1|, |\mathcal{C}_2|$ and substituting $\ell = \sqrt{n}$ we get:

$$|\mathcal{C}_1| + |\mathcal{C}_2| \geq |\psi_1^*|/\ell + \ell|\psi_2^*|/n$$

$$= (|\psi_1^*| + |\psi_2^*|)/\sqrt{n} = |\psi^*|/\sqrt{n} \ .$$

To see that $|\mathcal{C}_2| \geq \ell|\psi_2^*|/n$, let G_2 be the graph at the beginning of Phase 2. For each $v \in V$ let $\psi_2^*(v)$ be the restriction of ψ_2^* to the cycles in G_2 containing v. Note that for every $v \in V$ we can compute $\mathcal{C}_2(v)$ using any max-flow algorithm and flow decomposition. By the integrality of an optimal flow from the Max-Flow Min-Cut Theorem, $|\mathcal{C}_2| \geq |\psi_2^*(v)|$ for every vertex v. Thus, since every cycle in G_2 has length $> \ell$, we have:

$$n|\mathcal{C}_2| \geq \sum_{v \in V} |\psi_2^*(v)| \geq \ell |\psi_2^*| \ .$$

$\square$

Although we are unable to prove that $\Theta(\sqrt{n})$ is also a lower bound for the integrality gap of directed cycle packing, we conjecture this is the case. This conjecture is supported by the following construction showing that $\Theta(\sqrt{n})$ is a lower bound for the odd directed cycle packing problem (namely, the maximum number of edge-disjoint directed cycles of odd length).

Proposition 3.2 *For infinitely many n, there exists a digraph G on n vertices, in which every pair of odd cycles has a common edge, and yet $\nu^*_{oddc}(G) = \Omega(\sqrt{n})$, where $\nu^*_{oddc}(G)$ is the fractional odd cycle packing number of G.*

Proof: Let N be an odd positive integer, and consider the digraph D_N whose vertices are (i,j) for $i,j = 1,\ldots,N$. The edges of D_N emanate from (i,j) to $(i+1,j)$ for $i = 1,\ldots,N-1$ and $j = 1,\ldots,N$ and from (i,j) to $(i,j+1)$ for $i = 1,\ldots,N$ and $j = 1,\ldots,N-1$. There are also edges from (i,N) to $(N+1-i,1)$.

We first show that D_N does not have two *vertex-disjoint* odd directed cycles. Clearly, every cycle of D_N is composed of segments, each segment starting in the first column, passes through every other column sequentially (and sometimes goes down in the rows) until it reaches the last column. Segments are separated by the edges connecting the vertices in the last column to the vertices in the first column. Thus, each segment has a unique start vertex from the first column, and a unique end vertex from the last column. The length of a segment is the number of vertices it contains. Thus, the length of a cycle is the sum of the lengths of its segments. We partition the vertices of D_N into two types, even and odd. Even vertices are those whose coordinates have the same parity. An even (odd) segment of a cycle is a segment of even (odd) length. Notice that since N is odd, the endpoints of even segments belong to different types, while the endpoints of odd segments belong to the same type. Also notice that the end vertex of a segment has the same type as the start vertex of the following segment. It follows that odd cycles must have an even number of even segments and, trivially, an odd number of odd segments. Thus, odd cycles have an odd number of segments. Notice that every cycle (whether even or odd) that does not contain a vertex from the middle row must have an even number of segments (as the segments alternate below and above the middle row). Thus, we have shown that every odd cycle must contain a vertex from the middle row. In particular, every odd cycle has a segment starting in $(i,1)$ and ending in (j,N) where $i \leq (N+1)/2$ and $j \geq (N+1)/2$. Now, let C and C' be two odd cycles. We may assume that C has a segment S starting in $(i,1)$ and ending in (j,N) where $i \leq (N+1)/2$ and $j \geq (N+1)/2$, and C' has a segment S' starting in $(k,1)$ and ending in (ℓ,N) where $k \leq (N+1)/2$ and $\ell \geq (N+1)/2$. Assume, w.l.o.g., that $i \leq k$. If $i = k$ or $\ell \leq j$ we are done since in this case both segments intersect. Thus, we may assume $i < k$ and $j < \ell$. But in this case we have, as before, that if C' does not contain any vertex of S then the segments of C' must alternate below and above the segment S, and hence C' must have an even number of segments, contradicting the fact that C' is an odd cycle.

To estimate from below the fractional odd cycle packing number of G, for each $1 \le i \le (N+1)/2$, define the cycle C_i as follows:

$$\begin{aligned} C_i &= ((i,1),(i,2),\ldots,(i,i),(i+1,i),(i+2,i),\ldots, \\ &\quad (N+1-i,i),(N+1-i,i+1), \\ &\quad (N+1-i,i+2),\ldots,(N+1-i,N),(i,1)) \end{aligned}$$

(i.e. C_i starts at $(i,1)$, goes horizontally till (i,i), then drops vertically to $(N+1-i,i)$ and then again goes horizontally till $(N+1-i,N)$ and finally returns to $(i,1)$). It is easy to see that each vertex of D_N belongs to at most two cycles C_i, and therefore, giving value $\psi(C_i) = 0.5$ to each cycle C_i, we obtain a fractional odd cycle packing of value $(N+1)/4$. Now, by replacing each vertex v of D_n with the path v_{in}, v_{mid}, v_{out} and replacing each edge (u,v) with the edge (u_{out}, v_{in}) we obtain a new graph D'_N with $3N^2$ vertices. Any set of edge-disjoint directed cycles in D'_N is also vertex-disjoint, and corresponds to a set of vertex-disjoint directed cycles in D_N. Furthermore, any odd (even) cycle in D_N corresponds to an odd (even) cycle in D'_N. Thus, by letting $n = 3N^2$ the desired construction follows. $\square$

4 Proof of Theorem 1.3

In this section we consider simple digraphs only. The greedy algorithm for the maximum S-cycle packing problem repeatedly chooses a shortest S-cycle and removes its edges from the graph.

Theorem 4.1 *Given a subset S of vertices of a simple digraph G, the greedy algorithm finds a set of at least $\tau_c(G,S)/(5n^{2/3})$ edge-disjoint directed S-cycles in G.*

Let $f(n,\ell)$ be the maximum of $\tau_c(G)$ taken over all simple digraphs G on n vertices with girth$(G) > \ell$. It is easy to see that if $\mathcal{C}$ is a cycle packing computed by the greedy algorithm on G, then $\tau_c(G) \le \ell|\mathcal{C}| + f(n,\ell)$ for any positive integer ℓ. A similar statement holds for the analogous definition of $f(n,l)$ in the undirected case. In fact, a similar statement holds for the analogous vertex-disjoint (directed or undirected) cycle packing and cycle cover problems. In the undirected vertex-disjoint case Komlós [8] showed that $f(n,\ell) = \Theta(\frac{n}{\ell}\ln(n/\ell))$. In the directed vertex-disjoint case, Seymour [9] showed that $f(n,\ell) \le 4\frac{n}{\ell}\ln(4n/\ell)\ln\log(4n/\ell)$. He also gave an example showing that $f(n,\ell) = \Omega(\frac{n}{\ell}\ln(n/\ell))$. In the edge-disjoint case, answering an earlier conjecture of Bollobás, Erdös, Simonovits, and Szemerédi [2], Komlós [8] established the asymptotically tight bound $f(n,\ell) = \Theta(\frac{n^2}{\ell^2})$ in undirected graphs.

We generalize this by defining $h(n,\ell)$ to be the maximum of $\tau_c(G,S)$ taken over all simple digraphs G on n vertices and $S \subseteq V(G)$ so that every S-cycle in G has length $> \ell$. Let $\tilde{\nu}(G,S)$ denote the size of an S-cycle packing computed by some run of the greedy algorithm.

Lemma 4.2 *For any positive integer ℓ,*

$$\tau_c(G,S) \le \ell\tilde{\nu}(G,S) + h(n,\ell) \le (\ell + h(n,\ell))\tilde{\nu}(G,S).$$

Proof: Fix an optimal cover F with $|F| = \tau_c(G,S)$, and partition it into two sets F_1 and F_2, where F_1 are the edges contained in S-cycles of length $\le \ell$ of the S-packing computed. Then $|F_1| \le \ell\tilde{\nu}(G,S)$, since every S-cycle of length $\le \ell$ in the packing computed contains at least one edge from F_1. On the other hand $|F_2| \le h(n,\ell)$, by the optimality of $|F|$ and by the definition of $h(n,\ell)$. The result follows. □

For digraphs, the bound $h(n,\ell) = O((n^2/\ell^2)\log^2(n/\ell))$ can be deduced from [10, Theorem 1.1] where a more general problem was considered. We will show that $h(n,\ell) = \Theta(n^2/\ell^2)$ using the following lemma of Komlós [8].

Lemma 4.3 ([8], Lemma 3) *Let $a_0, a_1, \ldots, a_t$ be a sequence of non-negative real numbers, and denote $s_k = \sum_{i=0}^{k} a_i$. Then there exist $k \in \{0,\ldots,t-1\}$ such that $a_k a_{k+1} < \frac{2e}{t^2} s_k s_t$.* □

Corollary 4.4 *Let $a_0, a_1, \ldots, a_t$ be a sequence of integers, and denote $s_k = \sum_{i=1}^{k} a_i$ and $p = \lceil t/2 \rceil$. Suppose that $s_p \le s_t/2$. Then there exists $k \in \{0,\ldots,p-1\}$ such that:*

$$a_k a_{k+1} < \frac{2e}{p^2} s_k s_p \le \frac{4e}{t^2} s_k s_t.$$

□

Lemma 4.5 *Let S be a subset of vertices of a simple digraph G on n vertices so that every S-cycle in G has length $> \ell$. Then there exists an S-cycle edge-cover F with $|F| \le 4e(n/\ell)^2$. Moreover, such F can be found in polynomial time.*

Proof: The proof is by induction on n. If G has no S-cycles, in particular if it has ℓ vertices or less, the statement is obvious. We can also assume that G is strongly connected; otherwise, validity of the result for every strongly connected component of G implies the result for G.

Since every S-cycle in G has length $> \ell$, there are vertices u, v with $u \in S$ and $v \in V(G)$ such that every (u,v)-dipath has length $\ge \ell$, and hence there is a partition of $V(G)$ into nonempty sets $X_0, \ldots, X_t$, where $t \ge \ell$, such that no edge of G has tail in X_i and head in X_j, for $j \ge i+2$. Let $a_i = |X_i|$ for $i = 0, \ldots, t$, and let s_k and p be as in Corollary 4.4. Notice that $s_t = n$. We may assume that $s_p \le n - s_p$, since otherwise we may consider the reversed sequence of $a_0, \ldots, a_t$. By Corollary 4.4, there exists $k \in \{0,\ldots,p-1\}$ such that:

$$a_k a_{k+1} < \frac{4e}{t^2} s_k n.$$

Let F' be the edge cut consisting of the set of edges going from X_k to X_{k+1} (if we consider the reversed sequence, then we take also the "reversed" cut). Then, since G is simple

$$|F'| \le a_k a_{k+1} < \frac{4e}{t^2} s_k n.$$

We delete F' and apply the inductive hypothesis to the subgraphs G_1 and G_2 of G induced by the corresponding parts $V_1 = X_1 \cup \cdots \cup X_k$ and $V_2 = X_{k+1} \cup \cdots \cup X_t$. Clearly, any S-cycle in $G - F'$ is entirely contained either in G_1 or in G_2.

To summarize, we can find a cut F' that divides G into two subgraphs G_1 and G_2, where G_i has n_i vertices, such that $n_1 + n_2 = n$ and $n_1 \leq n/2 \leq n_2$, and such that $|F'| \leq \frac{4e}{\ell^2} n_1 n$. We need to prove that:

$$|F'| + 4e\left(\frac{n_1^2}{\ell^2} + \frac{n_2^2}{\ell^2}\right) \leq 4e\frac{n^2}{\ell^2}.$$

Indeed,

$$\begin{aligned}|F'| + 4e\left(\frac{n_1^2}{\ell^2} + \frac{n_2^2}{\ell^2}\right) &\leq \frac{4e}{\ell^2}(n_1 n + n_1^2 + n_2^2) \\ &< \frac{4e}{\ell^2}(2n_1 n_2 + n_1^2 + n_2^2) = 4e\frac{n^2}{\ell^2}.\end{aligned}$$

□

The bound in Lemma 4.5 is tight up to a constant factor even for $S = V$, as can be seen by taking the blowup of a directed ℓ-cycle.

By Lemmas 4.2 and 4.5 we deduce:

Corollary 4.6 *Let S be a subset of vertices of a simple digraph G on n vertices. Then for any integer ℓ,*

$$\tau_c(G, S) \leq (\ell + 4e(n/\ell)^2)\tilde{\nu}(G, S).$$

In particular for $\ell = 2e^{1/3}n^{2/3}$ we have $\tau_c(G, S) \leq 3e^{1/3}n^{2/3}\tilde{\nu}(G, S) < 5n^{2/3}\tilde{\nu}(G, S)$ and this also completes the proofs of Theorems 4.1 and 1.3. □

Remark: In [5] it was shown that the greedy algorithm for the *undirected* Edge-Disjoint Paths problem has approximation ratio $O(n^{2/3})$. The method presented in this section can be used to provide a different proof for the same result.

Acknowledgment

The authors thank Noga Alon and Guy Kortsarz for useful discussions.

References

[1] P. Balister, *Packing digraphs with directed closed trails*, Combin. Probab. Comput. 12 (2003), 1–15.

[2] B. Bollobás, P. Erdős, M. Simonovits and E. Szemerédi, *Extremal graphs without large forbidden subgraphs*, Ann. Discrete Math. 3 (1978), 29–41.

[3] B. Bollobás and A. Thomason, *On the girth of Hamiltonian weakly pancyclic graphs*, J. Graph Theory 26 (1997), 165–173.

[4] A. Caprara, A. Panconesi and R. Rizzi, *Packing cycles in undirected graphs*, J. Algorithms 48 (2003), 239–256.

[5] C. Chekuri and S. Khanna, *Edge disjoint paths revisited*, Proc. 14th ACM-SIAM SODA (2003), 628–637.

[6] D. Dor and M. Tarsi, *Graph decomposition is NPC – A complete proof of Holyer's conjecture*, Proc. 20th ACM STOC, ACM Press (1992), 252–263.

[7] P. Erdős and H. Sachs, *Reguläre Graphen gegebener Taillenweite mit minimaler Knotenzahl*, Wiss. Z. Martin-Luther Univ. Halle-Wittenberg Math.-Natur. Reihe 12 (1963), 251.

[8] J. Komlós, *Covering odd cycles*, Combinatorica 17 (1997), 393–400.

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

[10] K. Varadarajan and G. Venkataraman, *Graph decomposition and a greedy algorithm for edge-disjoint paths*, Proc. 15th ACM-SIAM SODA (2004), 372–373.