

Local improvement algorithms for a path packing problem: A performance analysis based on linear programming

Koen M J de Bontridder, Bjarni V Halldórsson, Magnús M Halldórsson, Cor a J Hurkens, Jan Karel Lenstra, Ramamoorthi Ravi, Leen Stougie

▶ **To cite this version:**

Koen M J de Bontridder, Bjarni V Halldórsson, Magnús M Halldórsson, Cor a J Hurkens, Jan Karel Lenstra, et al.. Local improvement algorithms for a path packing problem: A performance analysis based on linear programming. Operations Research Letters, 2021, 49 (1), pp.62-68. 10.1016/j.orl.2020.11.005 . hal-03498424

HAL Id: hal-03498424
https://inria.hal.science/hal-03498424

Submitted on 21 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Operations Research Letters 49 (2021) 62–68

Contents lists available at ScienceDirect

Operations Research Letters

journal homepage: www.elsevier.com/locate/orl

Local improvement algorithms for a path packing problem: A performance analysis based on linear programming☆

K.M.J. De Bontridder [a], B.V. Halldórsson [b,c], M.M. Halldórsson [c], C.A.J. Hurkens [d,*], J.K. Lenstra [e], R. Ravi [f], L. Stougie [e]

[a] Mapscape b.v., Eindhoven, The Netherlands
[b] deCODE Genetics, Reykjavik, Iceland
[c] Reykjavik University, Iceland
[d] Eindhoven University of Technology, The Netherlands
[e] Centrum Wiskunde & Informatica, Amsterdam, The Netherlands
[f] Carnegie Mellon University, Pittsburgh, PA, USA

ARTICLE INFO

Article history:
Received 14 July 2020
Received in revised form 23 October 2020
Accepted 6 November 2020
Available online 13 November 2020

Keywords:
Path packing
Greedy algorithm
Local search
Performance guarantee
Linear programming

ABSTRACT

Given a graph, we wish to find a maximum number of vertex-disjoint paths of length 2. We propose a series of local improvement algorithms for this problem, and present a linear-programming based method for analyzing their performance.

© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Given a graph $G = (V, E)$, we wish to find a maximum number of vertex-disjoint paths of length 2. For this NP-hard problem [7], we propose a series of local improvement algorithms. The basic heuristic, denoted H_0, applies the greedy algorithm to obtain a maximal path packing. For $k \geq 1$, the kth heuristic in the series, denoted H_k, starts from a maximal path packing and attempts to improve it by replacing any k paths of length 2 by $k + 1$ paths of length 2; when no further improvements are found, H_k terminates. For fixed k, H_k runs in polynomial time, but H_k is unlikely to run in time polynomial in k, unless P = NP [2].

In this paper we are interested in the worst-case performance ratios ρ_k of the heuristics H_k ($k \geq 0$) to the optimum. We will show that $\rho_0 = 1/3$, $\rho_1 = 1/2$, $\rho_2 = 5/9$, $\rho_3 = 7/11$, and $\rho_4 = 2/3$, and that these ratios are tight.

In Section 2 we review related work. In Section 3 we present classes of instances for which $H_0, \ldots, H_4$ meet their claimed performance ratios. In Sections 4–6 we establish lower bounds on $\rho_0, \ldots, \rho_4$ that match the upper bounds provided by these examples. The lower bounds are obtained by solving linear programs, where the inverse of the performance ratio is the objective and an analysis of certain configurations that can or cannot be locally improved yields the constraints.

Our research was motivated by the relation between the path packing problem and the test cover problem with tests of size at most 2; see [2].

2. Earlier work

Hurkens and Schrijver [6] consider a series of analogous local improvement algorithms for the more general problem of packing vertex-disjoint subgraphs on t vertices in a given graph. They derive a lower bound ϕ_k on the performance ratio of their kth heuristic, and prove that it is tight if the subgraph is a clique. In particular, for $t = 3$,

$$\phi_k = \begin{cases} \frac{2 \cdot 2^{(k+2)/2} - 3}{3 \cdot 2^{(k+2)/2} - 3} & \text{if } k \text{ is even,} \\ \frac{2 \cdot 2^{(k+1)/2} - 2}{3 \cdot 2^{(k+1)/2} - 2} & \text{if } k \text{ is odd.} \end{cases}$$

Since a path of length 2 is a subgraph on three vertices, we know that $\rho_k \geq \phi_k$.

Table 1 lists the values of ϕ_k and ρ_k for $k = 0, \ldots, 4$. Note that $\lim_{k\to\infty} \phi_k = \rho_4$. The limiting value of ρ_k is unknown, but it

☆ This is a companion paper to De Bontridder et al. (2003) containing details of the local improvement algorithm described therein.

* Corresponding author.
E-mail addresses: koen.debontridder@navinfo.eu (K.M.J. De Bontridder), bjarnivh@ru.is (B.V. Halldórsson), mmh@ru.is (M.M. Halldórsson), c.a.j.hurkens@tue.nl (C.A.J. Hurkens), jkl@cwi.nl (J.K. Lenstra), ravi@andrew.cmu.edu (R. Ravi), leen.stougie@cwi.nl (L. Stougie).

https://doi.org/10.1016/j.orl.2020.11.005
0167-6377/© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Table 1
Performance ratios for local improvement algorithms.

Problem	k	0	1	2	3	4	⋯	∞
Triangle packing	ϕ_k	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{5}{9}$	$\frac{3}{5}$	$\frac{13}{21}$	⋯	$\frac{2}{3}$
Path packing	ρ_k	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{5}{9}$	$\frac{7}{11}$	$\frac{2}{3}$		

is likely to be strictly smaller than 1, since the problem of packing paths of length 2 is APX-hard [2].

For $k = 0, 1, 2$, the lower bound ϕ_k on ρ_k matches the upper bound provided by the examples given in Section 3, which proves part of our result. For $k = 3, 4$, the proof is more involved.

Hassin and Rubinstein [4] study packing 3-edge paths of maximum total weight and give a $\frac{3}{4}$-approximation algorithm. They present a $(\frac{89}{169} - \epsilon)$-approximation for the weighted triangle packing problem in [5], and show that this algorithm also gives a performance ratio of $(\frac{35}{67} - \epsilon)$ for the maximum weight packing of 2-edge paths. Bafna et al. [1] analyze the k-local search algorithm for computing maximum weighted independent sets in $(k+1)$-claw free graphs to show a performance ratio of $(k - 1 + \frac{1}{k})$. Fernau and Raible [3] give a parameterized algorithm for deciding whether a 2-edge path packing of size k exists in time $O^*(2.448^{3k})$ where the O^* hides polynomial bounds on the size of the instance (like n^c for some constant c).

From now on, any *path* consists of three vertices and two edges.

3. Upper bounds: worst-case examples

Fig. 1 shows instances for which the heuristics $H_0, \ldots, H_4$ may perform badly and meet the performance ratios claimed above. For each k, the solid edges form an optimal solution, and the dashed edges form a locally optimal solution, produced by H_k. The latter solution can be improved by sacrificing $k+1$ paths in the top-left, in return for $k+2$ or more alternative paths, indicated by solid lines. Sacrificing only k paths does not give room for improvement. We leave it to the reader to verify this assertion. Note that we obtain infinite classes of such examples by creating multiple copies of the graphs.

4. Lower bounds: outline of approach

Before getting into details, we will outline our approach of proving lower bounds on the performance of the heuristics $H_0, \ldots, H_4$.

For a given value of k, $0 \le k \le 4$, let P denote the path packing found by H_k, and let Q denote an arbitrary maximal packing. To investigate the relation between P and Q, we will give each vertex in P a label that quantifies the interaction between the path in Q to which it belongs and the paths in P. When the middle vertex of a path in P has label β and its end vertices have labels α and γ with $\alpha \le \gamma$, then the path is said to be of type $\alpha\beta\gamma$. A combinatorial analysis enables us to reduce the number of different path types to 49.

Next, for each path type $\alpha\beta\gamma$, we define a variable $x(\alpha\beta\gamma)$, whose value is equal to the fraction of paths in P of that type. We then formulate a linear program in these variables. The objective will be given by $|Q|/|P|$. The constraints capture conditions that are implied by the local optimality of the solutions. These conditions are rather straightforward for H_0, H_1 and H_2 but ask for some case analysis for H_3 and H_4. It will be seen that the optimum solution values match the inverses of the performance ratios $\rho_0, \ldots, \rho_4$ given in Table 1.

5. Lower bounds: vertex labels

For a given $k \ge 0$, let P denote the path packing found by H_k, and let Q denote an arbitrary maximal packing. A *P-path* is a path in P, a *Q-path* is a path in Q. For a vertex v, $p(v)$ denotes the P-path containing v, if it exists, and $q(v)$ denotes the Q-path containing v, if it exists.

Without loss of generality, we make the following assumptions.

(1) *G contains no other edges than those in P and Q.* Otherwise, delete the edges that are neither in P nor in Q from G. In the new graph, P is still locally optimal and Q remains maximal.

(2) $|P| < |Q|$. Otherwise, $|P|/|Q| \ge 1$, and we are done.

(3) *Each P-path intersects a Q-path.* Otherwise, Q is not maximal.

(4) *Each Q-path intersects a P-path.* Otherwise, P is not locally optimal.

(5) *No P-path and Q-path cover the same three vertices.* Otherwise, delete those three vertices. In the new graph, P is locally optimal, Q is maximal, and $|P|/|Q|$ is smaller.

(6) *Each node of degree > 1 is in both P and Q.* For a vertex v that is an endpoint of $p(v)$ or $q(v)$, the statement is evidently true, using (1). The case remains where $v = b$ is the midpoint of a path a–b–c, that is either in P or in Q. Then a or c or both must be covered twice, as otherwise P or Q is not maximal. Assume that c is covered twice. By (1), b has no neighbors besides a and c. Modify the graph by changing the path a–b–c to a–c–b (removing the edge ab and adding if needed the edge ac). In the new setting b has only one neighbor, c. Let P' and Q' denote the new packings and suppose P' can be improved to P''. P'' must involve the edge a–c, as it would otherwise also exist in the original graph. If P'' contained the path a–c–d, we can change it to a–c–b, since then b is not covered in P''. But then, by changing a–c–b to a–b–c, the improvement is also possible in the original graph.

We give each vertex v in P a *label*, which expresses the interaction of its Q-path $q(v)$ with the P-paths. Vertices not covered by P do not get a label. Vertex v in P receives label

0 if v is not on any Q-path;
1 if $q(v)$ intersects exactly one P-path $p(v)$, in one vertex v;
2 if $q(v)$ intersects exactly two P-paths $p(v)$ and $p(w)$ in the two vertices v and w only, with v being the middle vertex of $q(v)$; (w receives label 3;) note that, by assumption (6), either v or w is the middle vertex of $q(v)$;
3 if $q(v)$ intersects exactly two P-paths $p(v)$ and $p(w)$ in vertices v and w, where w receives label 2 or 5;
4 if $q(v)$ intersects three P-paths;
5 if $q(v)$ and $p(v)$ intersect in two vertices v and w, and the third vertex z on $q(v)$ is on another P-path $p(z)$; w receives label 5 as well; z receives label 3;
6 if $q(v)$ and $p(v)$ intersect in two vertices v and w, and the third vertex on $q(v)$ is not on another P-path; w receives label 6 as well.

The case that $q(v)$ and $p(v)$ intersect in three vertices is excluded by assumption (5). Note that the labeling depends on the triples of vertices that form a P-path. It is irrelevant which edges are used in a P-path. Further note that label 5 comes in pairs, and so does label 6. Fig. 2 illustrates some typical labelings.

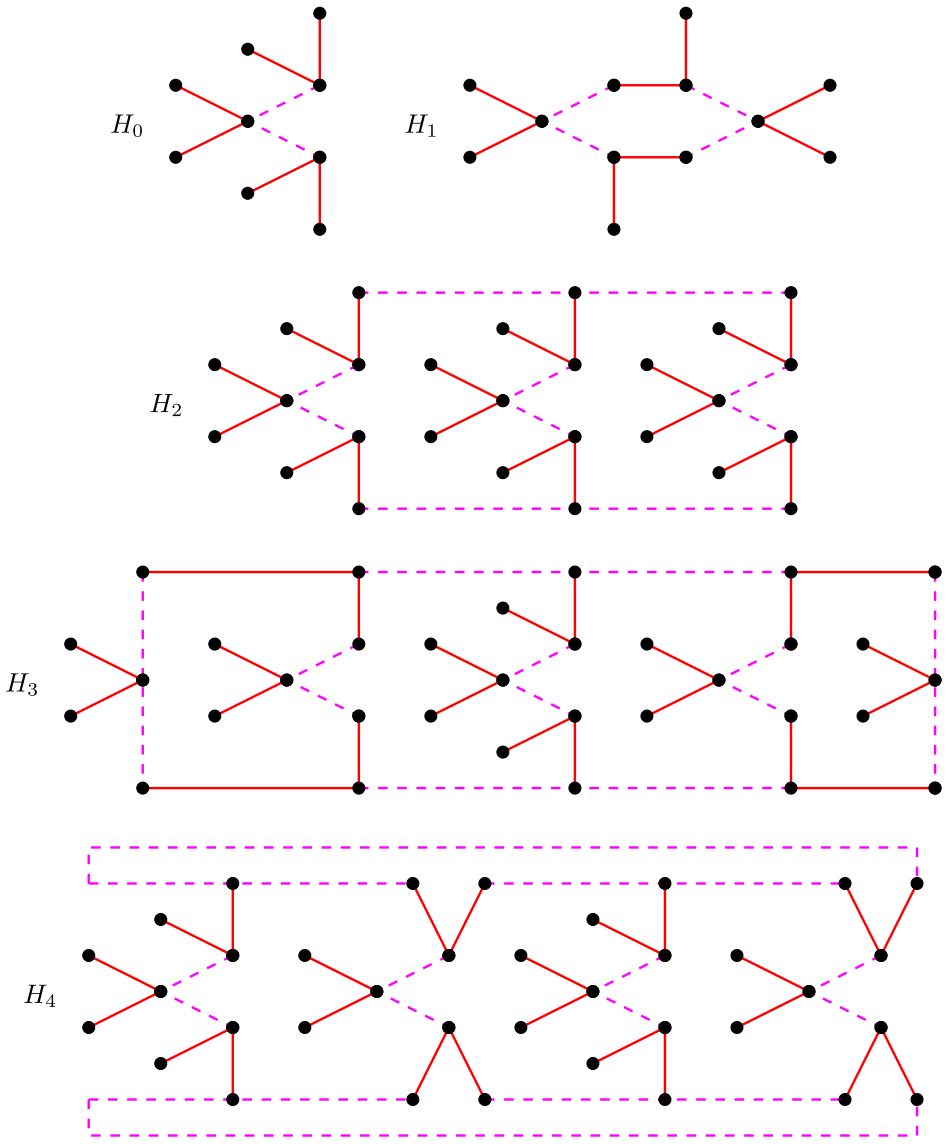


Fig. 1. Worst-case instances for heuristics $H_0, \ldots, H_4$. Solid edges form an optimal solution, dashed edges form a locally optimal solution.

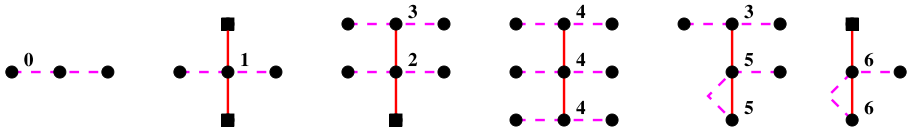


Fig. 2. Typical labelings. Solid edges form Q-paths, dashed edges form P-paths, squares denote vertices not covered by P.

We make some further assumptions.

(7) *Vertex pairs labeled* 5 *or* 6 *have a* Q*-path and a* P*-path with coinciding middle vertices.* Consider a P-path on a, b, c and a Q-path on a, c, d, with a and c labeled 5 or 6. Change G, replacing the paths by b–a–c and d–a–c, respectively. Any improvement in the new graph must sacrifice b–a–c and use, w.l.o.g., d–a–c. This improvement applies to the old graph too, replacing the P-path on a, b, c by the Q-path on a, c, d.

(8) *No vertex has label* 6. Consider a P-path a–b–c with labels 6 on b and c, intersecting a Q-path d–b–c. Change G, introducing a new vertex c' and replacing the P-path by a–b–c', with c' labeled 0 and b labeled 1. Any improvement of P in the new graph implies an improvement in the old graph.

(9) *No* P*-path is labeled* 155. Consider a P-path a–b–c with labels 1, 5, 5, intersecting a Q-path d–b–c, with d labeled 3. Change G, introducing a new vertex a' and replacing the P-path by a'–a–b, which is labeled 012. Vertex c becomes unlabeled. If the new graph allows for an improvement of P, the new solution uses w.l.o.g. the paths $q(a)$ and $q(b)$, and hence exists in the old graph as well.

(10) *If a P-path has a vertex labeled* 1*, its middle vertex has label* 1. Otherwise, consider a P-path a–b–c with c labeled 1. Change G as under assumption (6), and replace a–b–c in P by a–c–b. Any improvement of P in the new graph must sacrifice this path and use, w.l.o.g., the path $q(c)$, and hence not the edges $\{a, c\}$ and $\{c, b\}$. This improvement could have been made in the old graph as well. Note that by assumption (9), vertices a and b are not labeled 5. Hence, assumption (7) is maintained.

A P-path is said to be of *type* $\alpha\beta\gamma$ if its middle vertex has label β and its end vertices have labels α and γ, with $\alpha \le \gamma$. A P-path is said to be of *form* $\{\alpha\beta\gamma\}$ if its type is a permutation of $\alpha\beta\gamma$. A wildcard $*$ may be used to match any label. As a consequence of the assumptions made, there exist 49 different labelings.

A vertex v that is labeled 1, 2, 3, or 4 is called **black** if $p(v) \setminus v$ contains a vertex with label 1, or if v is an endpoint of $p(v)$ and $p(v)\setminus v$ contains a vertex labeled 2. In terms of local improvement, a black vertex v can be 'set free' by replacing $p(v)$ by a path not covering v. The replacing path uses one or two edges from a Q-path, the middle vertex of which lies on $p(v) \setminus v$, and one or two endpoints not on any P-path. We will mark a black vertex by an overscore.

It follows from the labeling of $p(v)$ whether vertex v is black or not. We may, for instance, have paths labeled $\bar{1}\bar{1}\bar{1}$, $\bar{2}1\bar{4}$, $23\bar{3}$, $\bar{4}2\bar{4}$, 434 and 255. From assumptions (6) and (10) we have that a vertex labeled 0, $\bar{2}$, $\bar{3}$, or $\bar{4}$ is never the middle vertex of a P-path.

6. Lower bounds: LP formulations

We define $x(\alpha\beta\gamma)$ as the fraction of P-paths of type $\alpha\beta\gamma$, $n(\sigma)$ as the number of occurrences of label σ, and $|\alpha\beta\gamma|$ as the number of paths of type $\alpha\beta\gamma$. Then

$$n(\sigma) = |P| \sum_{(\alpha\beta\gamma)} x(\alpha\beta\gamma)[\chi(\sigma,\alpha) + \chi(\sigma,\beta) + \chi(\sigma,\gamma)] \qquad (*)$$

where $\chi(\sigma,\tau) = 1$ if $\sigma = \tau$, and 0 otherwise. If σ has an overscore, we restrict the count to occurrences of a black σ.

A *short path* is a Q-path that intersects exactly two P-paths (and hence is labeled -23 or 355, where - denotes an unlabeled node). A **black short path** is a short path with at least one black vertex ($\bar{2}$ or $\bar{3}$).

Heuristic H_0

$$|Q| = n(1) + \frac{1}{2}n(2) + \frac{1}{2}n(3) + \frac{1}{3}n(4) + \frac{1}{4}n(5) \qquad (0a)$$

$$n(3) = n(2) + \frac{1}{2}n(5) \qquad (0b)$$

$$\sum_{(\alpha\beta\gamma)} x(\alpha\beta\gamma) = 1 \qquad (0c)$$

Dividing Eqs. (0a) and (0b) by $|P|$ and coupling the number of occurrences $n(\sigma)$ to the fractions x yields an LP-model. Its optimum solution value is $|Q|/|P| = 3$, for $x(111) = 1$. (Note that a 111 P-path is colored black: $\bar{1}\bar{1}\bar{1}$.)

The full version of the LP is given below. It has 49 variables $x(\alpha\beta\gamma)$, a variable $q = |Q|/|P|$ (which is the objective), and variables $\nu(k) = n(k)/|P|$. The first three constraints are equivalent to (0a), (0b), (0c); the six constraints defining $\nu(0), \ldots, \nu(5)$ follow from (*).

max q

subject to

$$q = \nu(1) + \tfrac{1}{2}\nu(2) + \tfrac{1}{2}\nu(3) + \tfrac{1}{3}\nu(4) + \tfrac{1}{4}\nu(5)$$

$$\nu(3) = \nu(2) + \tfrac{1}{2}\nu(5)$$

$$\begin{aligned}&x(010)+x(011)+x(012)+x(013)+x(014)+x(020)+x(022)\\&+x(023)+x(024)+x(030)+x(032)+x(033)+x(034)+x(040)\\&+x(042)+x(043)+x(044)+x(055)+x(111)+x(112)+x(113)\\&+x(114)+x(212)+x(213)+x(214)+x(222)+x(223)+x(224)\\&+x(232)+x(233)+x(234)+x(242)+x(243)+x(244)+x(255)\\&+x(313)+x(314)+x(323)+x(324)+x(333)+x(334)+x(343)\\&+x(344)+x(355)+x(414)+x(424)+x(434)+x(444)+x(455)\\&=1\end{aligned}$$

$$\begin{aligned}\nu(0) = {}&2x(010)+x(011)+x(012)+x(013)+x(014)+2x(020)\\&+x(022)+x(023)+x(024)+2x(030)+x(032)+x(033)\\&+x(034)+2x(040)+x(042)+x(043)+x(044)+x(055)\end{aligned}$$

$$\begin{aligned}\nu(1) = {}&x(010)+2x(011)+x(012)+x(013)+x(014)+3x(111)\\&+2x(112)+2x(113)+2x(114)+x(212)+x(213)+x(214)\\&+x(313)+x(314)+x(414)\end{aligned}$$

$$\begin{aligned}\nu(2) = {}&x(012)+x(020)+2x(022)+x(023)+x(024)+x(032)\\&+x(042)+x(112)+2x(212)+x(213)+x(214)+3x(222)\\&+2x(223)+2x(224)+2x(232)+x(233)+x(234)+2x(242)\\&+x(243)+x(244)+x(255)+x(323)+x(324)+x(424)\end{aligned}$$

$$\begin{aligned}\nu(3) = {}&x(013)+x(023)+x(030)+x(032)+2x(033)+x(034)\\&+x(043)+x(113)+x(213)+x(223)+x(232)+2x(233)\\&+x(234)+x(243)+2x(313)+x(314)+2x(323)+x(324)\\&+3x(333)+2x(334)+2x(343)+x(344)+x(355)+x(434)\end{aligned}$$

$$\begin{aligned}\nu(4) = {}&x(014)+x(024)+x(034)+x(040)+x(042)+x(043)\\&+2x(044)+x(114)+x(214)+x(224)+x(234)+x(242)\\&+x(243)+2x(244)+x(314)+x(324)+x(334)+x(343)\\&+2x(344)+2x(414)+2x(424)+2x(434)+3x(444)+x(455)\end{aligned}$$

$$\nu(5) = 2x(055)+2x(255)+2x(355)+2x(455)$$

$$x(\alpha\beta\gamma) \ge 0 \text{ for all } \alpha\beta\gamma, \quad \nu(k) \ge 0 \text{ for } k = 0, \ldots, 5$$

For the analysis of subsequent heuristics, with the focus on the number of black vertices, we extend the model with the constraints

$$\nu(\bar{1}) = 2x(011)+3x(111)+2x(112)+2x(113)+2x(114)$$

$$\begin{aligned}\nu(\bar{2}) = {}&x(012)+x(022)+x(112)+2x(212)+x(213)+x(214)\\&+2x(222)+x(223)+x(224)+2x(232)+2x(242)\end{aligned}$$

$$\begin{aligned}\nu(\bar{3}) = {}&x(013)+x(023)+x(113)+x(213)+x(223)+x(233)\\&+x(243)+2x(313)+x(314)+2x(323)+x(324)\end{aligned}$$

$$\begin{aligned}\nu(\bar{4}) = {}&x(014)+x(024)+x(114)+x(214)+x(224)+x(234)\\&+x(244)+x(314)+x(324)+2x(414)+2x(424)\end{aligned}$$

Heuristic H_1

No black node is labeled 1:

$$n(\bar{1}) \le 0 \qquad (1)$$

The LP extended by this additional constraint has an optimum solution value $|Q|/|P| = 2$ for $x(\bar{2}1\bar{3}) = 1$ or $x(\bar{2}1\bar{2}) = x(\bar{3}1\bar{3}) = 0.5$.

Heuristic H_2

No two black nodes are on a short path:

$$n(\bar{2}) + n(\bar{3}) \le n(3) \qquad (2)$$

To justify this constraint, consider black vertices u, v on a black short path $q(u) = q(v)$. Heuristic H_2 would replace the two paths $p(u)$ and $p(v)$ by disjoint paths only incident to $p(u) \setminus u$ and

$p(v)\setminus v$, and then add $q(u)$ as an additional P-path. The LP further extended by these constraints achieves its optimum $|Q|/|P| = \frac{9}{5}$ for $x(\bar{2}1\bar{2}) = \frac{3}{5}$ and $x(333) = \frac{2}{5}$.

Heuristics H_3 and H_4

As is clear from the model, the larger the fraction of nodes labeled 1 or 2, the larger is the ratio $|Q|/|P|$. This makes sense, as these nodes are neighbors to vertices covered by Q and not by P. Black nodes can transfer this degree of freedom, at the cost of rearranging one existing P-path. The fact that H_2 fails to improve the ratio indicates that one cannot have too many black nodes or nodes with label - or 0 close together. To see how this works out for heuristics H_3 and H_4, we now investigate how black short paths are involved in transferring the degree of freedom.

We need a more detailed description of how short paths interact. For each $\alpha\beta\gamma$-path with at least one non-black vertex labeled 2, 3 or 5, we distinguish between types $\alpha\beta\gamma^+$ and $\alpha\beta\gamma^-$. A P-path has a $^+$label if it intersects a black short path in a non-black vertex, and a $^-$label otherwise. It is called a P^+-path or P^--path, respectively. The 30 types $\alpha\beta\gamma$ that get a $+$ and $-$ variant are: 020, 02$\bar{2}$, 02$\bar{3}$, 02$\bar{4}$, 032, 042, $\bar{2}2\bar{2}$, $\bar{2}2\bar{3}$, $\bar{2}2\bar{4}$, 23$\bar{3}$, 23$\bar{4}$, 24$\bar{3}$, 24$\bar{4}$, 255, $\bar{3}2\bar{3}$, $\bar{3}2\bar{4}$, $\bar{4}2\bar{4}$, 030, 033, 034, 043, $\bar{2}3\bar{2}$, 333, 334, 343, 344, 355, 434, 055, and 455.

Non-black vertices labeled 2, 3 or 5 on a P^--path are colored *white* and denoted $\underline{2}$, $\underline{3}$, and $\underline{5}$, respectively. Vertices labeled 1, 2, 3, 4 or 5 that are neither black nor white are colored *gray* and denoted $\dot{1}, \dot{2}, \dot{3}, \dot{4}, \dot{5}$. We sometimes use the label $\dot{5}\dot{5}$ to denote the pair of vertices labeled 5 on a P^+-path. Notice that, from the (extended) path-type, we can deduce the color of each vertex. Vertices labeled 0 remain blank. Informally speaking, a vertex v is colored gray, if it lies on a black short path or if it might be 'freed up' by sacrificing two P-paths; one path to set free a nearby black vertex, and the path $p(v)$ to create a P-path without v. For a vertex v labeled 2, 3 or 5, such a nearby black vertex is available only if $p(v)$ is a P^+-path.

We observe that, by definition of P^--paths, no short path has a white and a black vertex.

Replace constraint (2) for heuristic H_2 by

$$n(\bar{2}) \le n(\dot{3}) \tag{2a}$$

$$n(\bar{3}) \le n(\dot{2}) + \frac{1}{2}n(\dot{5}) \tag{2b}$$

We justify these constraints as follows:

(2a) A black vertex v labeled 2 lies on $q(v)$ containing a gray vertex labeled 3.

(2b) A black vertex v labeled 3 lies on $q(v)$ that either contains a gray vertex w labeled $\dot{2}$ or two gray vertices labeled $\dot{5}$.

Heuristic H_3

The following additional constraints hold:

$$n(\bar{2}) \le |\{\dot{3}**\}^+| + |222^-| + |232^-| + |242| \tag{3a}$$

$$n(\bar{2}) + n(\bar{3}) \le |P^+| + |242| + |222^-| + |232^-| + |223^-| + |323^-| \tag{3b}$$

$$n(\bar{4}) \le 2n(\dot{4}) \tag{3c}$$

The LP defined by the preceding constraints achieves its optimum $|Q|/|P| = \frac{11}{7}$ for $x(\bar{2}1\bar{2}) = \frac{1}{7}$, $x(\bar{4}1\bar{4}) = \frac{4}{7}$, and $x(434^+) = \frac{2}{7}$. We justify the additional constraints as follows: Note that by construction, each black vertex labeled $\bar{2}$ or $\bar{3}$ has a corresponding P^+-path, but multiple $\bar{2}$ and $\bar{3}$ may share the same P^+-path.

(3a) A black vertex v labeled 2 lies on $q(v)$ containing a gray vertex w labeled 3. If another black vertex v' labeled 2 lies on $q(v')$ containing a gray vertex w' labeled 3, with $p(w) = p(w')$, heuristic H_3 will improve this configuration *unless* the vertices v and v' lie on the same P-path labeled $\bar{2}1\bar{2}$, $\bar{2}2\bar{2}$, $\bar{2}3\bar{2}$, or $\bar{2}4\bar{2}$. The path cannot have labels $\bar{2}1\bar{2}$ since heuristic H_3 would give an improvement. For the same reason, paths of type $\bar{2}2\bar{2}$ and $\bar{2}3\bar{2}$ must have a $^-$label.

(3b) A black vertex v labeled 2 or 3 lies on $q(v)$ containing a gray vertex w labeled 3, 2 or 5. If another black vertex v' labeled 2 or 3 lies on $q(v')$ containing a gray vertex w' labeled 3, 2 or 5, with $p(w) = p(w')$, heuristic H_3 will improve this configuration *unless* the vertices v and v' lie on the same P-path labeled $\bar{2}1\bar{2}$, $\bar{2}2\bar{2}$, $\bar{2}3\bar{2}$, $\bar{2}4\bar{2}$, $\bar{2}1\bar{3}$, $\bar{2}2\bar{3}$, $\bar{3}1\bar{3}$ or $\bar{3}2\bar{3}$. The path cannot have labels $\bar{2}1\bar{2}$, $\bar{2}1\bar{3}$ or $\bar{3}1\bar{3}$, since heuristic H_3 would give an improvement. For the same reason, the other five paths cannot have a $^+$label.

(3c) Each Q-path with three vertices labeled 4 can have at most two black vertices. If all three were black, heuristic H_3 would improve the packing.

Heuristic H_4

To derive additional constraints for H_4, we impose more stringent bounds on black vertices labeled $\bar{2}$ or $\bar{3}$ by starting with the number of corresponding paths labeled $+$, and carefully avoiding overcounts. Similarly, to bound the number of vertices labeled $\bar{2}$, we consider $q(v)$ as in (3a) containing a gray vertex w labeled $\dot{3}$ and carefully analyze the case whether $p(w)$ contains a gray vertex $\dot{4}$ or not to arrive at a second constraint.

The following additional constraints hold:

$$2n(\bar{2}) + 2n(\bar{3}) - 3(|222^-| + |232^-| + |323^-| + |223^-| + |242|) \le \tag{4a}$$
$$2|P^+| - (|\dot{2}\dot{5}\dot{5}^+| + |\dot{3}\dot{5}\dot{5}^+| + |0\dot{3}\dot{2}^+| + |\dot{2}\dot{3}4^+| + |0\dot{3}\dot{3}^+| + |\{\dot{3}\dot{3}4\}^+|) - 2|333^+| + n(\underline{2}) + n(\underline{3}) + \frac{1}{2}n(\underline{5})$$

$$2n(\bar{2}) + 2n(\bar{4}) - 2(|222^-| + |232^-| + |242|) - 3|224| \le \tag{4b}$$
$$4n(\dot{4}) + 2|\{\dot{3}*\}^+| - 2|\{\dot{3}\dot{4}*\}^+| - 4|344^+| - 4|434^+| - |334^+| - |343^+|$$

The LP defined by the preceding constraints achieves its optimum $|Q|/|P| = \frac{3}{2}$ for $x(\bar{2}1\bar{2}) = \frac{1}{4}$, $x(\bar{4}1\bar{4}) = \frac{1}{4}$, and $x(434^+) = \frac{1}{2}$. We justify the additional constraints as follows:

(4a) Like in the argument for (3b), black vertex v labeled 2 or 3 lies on $q(v)$ together with a vertex w labeled $\dot{2}$, $\dot{3}$, or $\dot{5}$. Now, either $p(w)$ is a P^+-path with one label $\dot{2}$, $\dot{3}$, or $\dot{5}\dot{5}$, or $p(w)$ has more than one such label. In the former case, we find a contribution to $2n(\bar{2}) + 2n(\bar{3})$ of $+2$ on the left-hand side of (4a), a contribution to $2|P^+|$ of $+2$ on the right-hand side, and no further negative contribution to the right-hand side.

Note that if another short path exists with black vertex v' and gray vertex w', with $p(w) = p(w')$, vertices v and v' lie on the same P-path labeled $\bar{2}2\bar{2}^-$, $\bar{2}3\bar{2}^-$, $\bar{2}4\bar{2}$, $\bar{2}2\bar{3}^-$ or $\bar{3}2\bar{3}^-$, as in the configuration left by heuristic H_3. Hence this can explain the second occurrence of a gray node labeled $\dot{2}$, $\dot{3}$, or $\dot{5}\dot{5}$ on $p(w)$, but not a third occurrence.

If $p(w)$ has two gray vertices labeled $\dot{2}$, $\dot{3}$ or $\dot{5}\dot{5}$, both stemming from a black short path, the above argument shows that we have a contribution to $2n(\bar{2})+2n(\bar{3})$ of $+4$ on the left-hand side of (4a), and a contribution of -3 for the path connecting v and v'. On the right-hand side we have a contribution to $2|P^+|$ of $+2$ and also a contribution of -1 for the path connecting w and w'.

If $p(w)$ has more than one gray vertex labeled $\dot{2}$, $\dot{3}$ or $\dot{5}\dot{5}$, and the black short path with vertex v is the only one to intersect the path $p(w)$, then there must be a vertex z on $p(w)$, with z having label $\dot{2}$, $\dot{3}$ or $\dot{5}$, so that $q(z)$ is not a black short path.

Now $q(z)$ either contains a white vertex 2 or 3, or a white pair of 5's, or it contains a gray vertex 2 or 3, or a gray pair of 5's, and therefore connects $p(w)$ with another path $p(w')$ with a $^+$label.

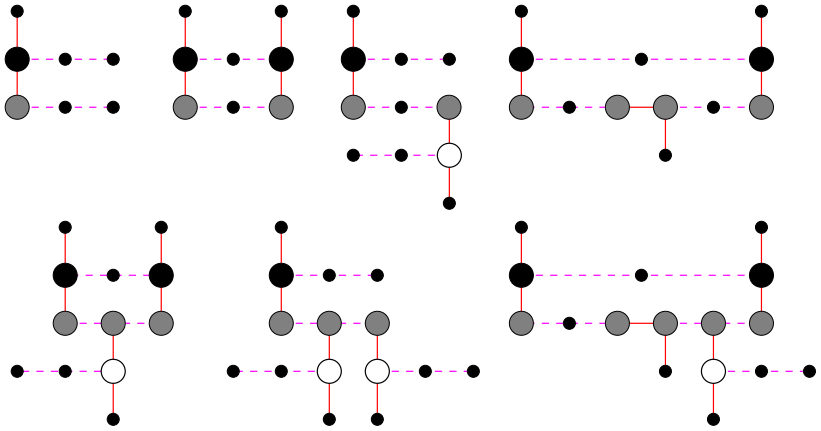


Fig. 3. Configurations around a black short path for H_4.

In the first case we have an extra contribution of $+1$ to the term $n(\underline{2}) + n(\underline{3}) + \frac{1}{2}n(\underline{5})$, so the total contribution to the right-hand side becomes $+2 = 2 - 1 + 1$.

In the latter case, the corresponding black vertices v and v' must be connected by a mutual P-path $p(v) = p(v')$ labeled $\bar{2}\bar{2}\bar{2}^-$, $\bar{2}\bar{3}\bar{2}^-$, $\bar{2}\bar{4}\bar{2}$, $\bar{2}\bar{2}\bar{3}^-$ or $\bar{3}\bar{2}\bar{3}^-$, otherwise heuristic H_4 can improve upon this solution. We then count on the left-hand side a contribution to $2n(\bar{2}) + 2n(\bar{3})$ of $+4$ and a contribution of -3 for the path connecting v and v'. On the right-hand side we have a contribution to $2|P^+|$ of $+4$ and also a contribution of -1 for the paths connecting w and w'.

If the $^+$labeled path involved contains more than two gray vertices $\dot{2}$, $\dot{3}$, or $\dot{5}\dot{5}$, and hence is labeled 333^+, it contains an extra gray vertex z' not lying on a black short path. It follows that the counterpart of z' on $q(z')$ is a white vertex. This adds a contribution of $+1$ to the term $n(\underline{2}) + n(\underline{3}) + \frac{1}{2}n(\underline{5})$, and of -2 to $-2|333^+|$, thus compensating for the lost contribution in $-(|\dot{2}\dot{5}\dot{5}^+| + |\dot{3}\dot{5}\dot{5}^+| + |0\dot{3}\dot{2}^+| + |\dot{2}\dot{3}\dot{4}^+| + |0\dot{3}\dot{3}^+| + |\{\dot{3}\dot{3}\dot{4}\}^+|)$.

Fig. 3 gives an overview of the possible configurations around a black short path. Each configuration satisfies the constraint (with inequality).

(4b) Note that adding inequalities (3a) and (3c) and multiplying the sum by 2 yields $2n(\bar{2})+2n(\bar{4})-2(|222^-|+|232^-|+|242|) \leq 4n(\dot{4})+2|\{\dot{3}**\}^+|)$, which is stronger than inequality (4b), unless we take into account P^+-paths containing both a $\dot{3}$ and a $\dot{4}$ labeled vertex. Such a path cannot contain a vertex labeled 1, 2 or 5, and hence at least one vertex labeled $\dot{3}$ is part of a black short path.

Consider a black vertex v labeled 2. As in the configuration left by heuristic H_3 it lies on a short path $q(w)$ with w labeled $\dot{3}$ and with $p(w)$ having a $^+$label. Possibly there is another vertex v' labeled $\bar{2}$ on a short path with w', with $p(w') = p(w)$, in which case v and v' have a common P-path labeled 222^-, 232^- or 242.

If $p(w)$ contains a single $\dot{3}$ and no $\dot{4}$, we neglect the labels on $p(w)\backslash w$ and have v contribute 2 to $n(\bar{2})$ on the left-hand side, and $p(w)$ contribute 2 to $2|\{\dot{3}**\}^+|$ on the right-hand side of (4b).

If $p(w)$ contains a single $\dot{3}$ lying on a black short path and a single $\dot{4}$, that is, $p(w)$ has type 034^+, 043^+, 334^+ or 343^+, then vertex v contributes 2 to the left-hand side, and $p(w)$ with type 334^+ or 343^+ contributes -1 to the right-hand side.

If the vertex z labeled $\dot{4}$ lies on $q(z)$ with three gray vertices, we let z contribute 1 to $n(\dot{4})$ on the right-hand side, in which case the inequality becomes $2 \leq 4 \cdot 1 - 1$.

If $q(z)$ contains one black vertex z' labeled $\bar{4}$, we have z' contribute $\frac{1}{2}$ to $n(\bar{4})$ on the left-hand side, and let z contribute 1 to $n(\dot{4})$ on the right-hand side, by which the inequality becomes $2 + 2 \cdot \frac{1}{2} \leq 4 - 1$.

If z is the only gray vertex of $q(z)$, with other nodes z' and z'', then $p(v) = p(z')$ or $p(v) = p(z'')$, otherwise H_4 would have improved this solution, namely, by dropping $p(w)$, rearranging $p(v)$, $p(z')$, and $p(z'')$, and adding $q(v)$ and $q(z)$. Assume that $p(v) = p(z')$ contains a third vertex v' labeled 1. Then H_3 would have improved the solution, by rearranging $p(z'')$, dropping $p(v)$ and $p(w)$, and adding $q(v), q(z), q(v')$. It follows that $p(v)$ must be v–v'–z' of type 224 (with either a $^+$label or a $^-$label). In this case the contributions of the vertices labeled 4 cancel out in the inequality, whereas v contributes 2 to the left-hand side, $p(v)$ contributes -3 to the left-hand side, and $p(w)$ contributes -1 or 0 to the right-hand side of (4b).

What remains is the case (33*) where $p(w)$ contains two gray vertices labeled 3, both lying on a black short path, and the case (344), where $p(w)$ is of type 344^+ or 434^+. In the case (33*), as described before, the black short paths are connected by a path $p(v)$ of type 222^-, 232^- or 242, which contributes -2 to the left-hand side. Should the path $p(w)$ contain a gray vertex z labeled 4, we let the vertex contribute to (4b). We do this in the same way as above, with a contribution of 4 to the right-hand side, and potentially a contribution of 1 to $2n(\bar{4})$ on the left-hand side. Note that z cannot be the only gray vertex on $q(z)$. If that were the case, the above reasoning would force $p(v)$ to intersect $q(z)$ and have type 224.

In the case (344) we would have $p(w)$ contain gray vertices z and z', say, both labeled $\dot{4}$. It follows from the above reasoning that at most one of z, z' can be the single gray vertex on $q(z)$ or $q(z')$ respectively, as $p(v)$ cannot intersect both $q(z)$ and $q(z')$. Note that $p(w)$ contributes -4 to the right-hand side, which is balanced by the gray nodes labeled 4, that contribute $+8$.

Hence in all cases, a vertex z labeled $\dot{4}$ contributes 4 to the right-hand side if $q(z)$ consists of three gray vertices; it contributes 1 to the left-hand side and 4 to the right-hand side if $q(z)$ has one black vertex; and it contributes 4 to both left-hand side and right-hand side if $q(z)$ contains two black vertices. We have to do so to prevent double counting. For instance, a Q-path labeled $\dot{4}\bar{4}\dot{4}$ can be adjacent to two distinct black short paths.

Fig. 4 depicts in principle each possible configuration around a black vertex labeled 2. In order to avoid double counting, for the inequality we only count gray and black nodes as far as they lie within the given boundaries. Gray and black nodes labeled 4 that are not accounted for in the picture must lie on (part of) a

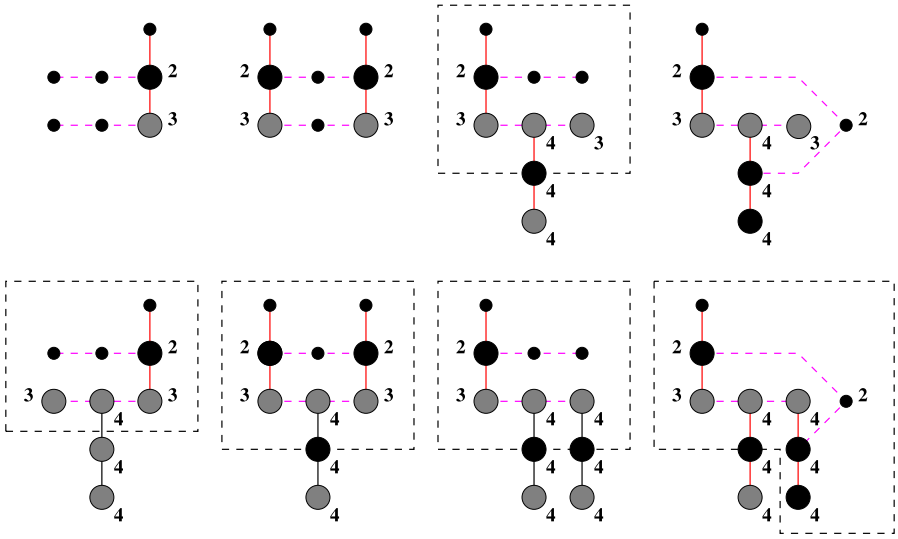


Fig. 4. More configurations around a vertex labeled $\bar{2}$ for H_4.

Q-path with three vertices labeled 4, in which at least one third of the nodes is gray. Each configuration satisfies the constraint, and almost all do so with equality.

Acknowledgments

We are grateful for the comments of the area editor and the referees, which helped us to improve the presentation.

References

[1] V. Bafna, B. Narayanan, R. Ravi, Nonoverlapping local alignments (weighted independent sets of axis-parallel rectangles), Discrete Appl. Math. 71 (1–3) (1996) 41–53.

[2] K.M.J. De Bontridder, B.V. Halldórsson, M.M. Halldórsson, C.A.J. Hurkens, J.K. Lenstra, R. Ravi, L. Stougie, Approximation algorithms for the test cover problem, Math. Program. B 98 (2003) 477–492.

[3] H. Fernau, D. Raible, A parameterized perspective on packing paths of length two, J. Comb. Optim. 18 (4) (2009) 319.

[4] Refael Hassin, Shlomi Rubinstein, An approximation algorithm for maximum packing of 3-edge paths, Inform. Process. Lett. 63.2 (1997) 63–67.

[5] Refael Hassin, Shlomi Rubinstein, An approximation algorithm for maximum triangle packing, Discrete Appl. Math. 154.6 (2006) 971–979.

[6] C.A.J. Hurkens, A. Schrijver, On the size of systems of sets every t of which have an SDR, with an application to the worst-case ratio of heuristics for packing problems, SIAM J. Discrete Math. 2 (1989) 68–72.

[7] D.G. Kirkpatrick, P. Hell, On the complexity of a generalized matching problem, in: Proceedings of the Tenth Annual ACM Symposium on Theory of Computing, 1978, pp. 240–245.