UNIVERSITY
OF TRENTO

DEPARTMENT OF INFORMATION AND COMMUNICATION TECHNOLOGY

38050 Povo – Trento (Italy), Via Sommarive 14
http://www.dit.unitn.it

IMPROVED APPROXIMATION FOR BREAKPOINT GRAPH DECOMPOSITION AND SORTING BY REVERSALS

Alberto Caprara and Romeo Rizzi

2002

Technical Report # DIT-02-0082

Improved Breakpoint Graph Decompositions

Alberto Caprara[*] Romeo Rizzi[o]

[*] DEIS
University of Bologna
Viale Risorgimento 2
I-40136 Bologna, Italy
e-mail: `acaprara@deis.unibo.it`

[o] BRICS[*]
Department of Computer Science
University of Aarhus
Ny Munkegade
DK-8000 Aarhus C, Denmark
e-mail: `rrizzi@science.unitn.it`

Abstract

Sorting by Reversals (SBR) is one of the most widely studied models of genome rearrangements in computational molecular biology. At present, $\frac{3}{2}$ is the best known approximation ratio achievable in polynomial time for SBR. A very closely related problem, called *Breakpoint Graph Decomposition* (BGD), calls for a largest collection of edge disjoint cycles in a suitably-defined graph. It has been shown that for almost all instances SBR is equivalent to BGD, in the sense that any solution of the latter corresponds to a solution of the former having the same value. In this paper, we show how to improve the approximation ratio achievable in polynomial time for BGD, from the previously known $\frac{3}{2}$ to $\frac{33}{23} + \epsilon$ for any $\epsilon > 0$. Our result uses the best known approximation algorithms for Stable Set on graphs with maximum degree 4 as well as for Set Packing where the maximum size of a set is 6. Any improvement in the ratio achieved by these approximation algorithms will yield an automatic improvement of our result.

Key words: sorting by reversals, breakpoint graph, alternating cycle decomposition, set packing, stable set, approximation algorithm.

1 Introduction

Sorting by Reversals (SBR) is one of the most widely studied models of genome rearrangements in computational molecular biology, and is defined as follows. Let $\pi = (\pi_1 \ \dots \ \pi_n)$ be a permutation of $\{1, \dots, n\}$. A *reversal* of the interval (i, j) is an inversion of the subsequence $\pi_i \ \dots \ \pi_j$ of π, yielding permutation $(\pi_1 \ \dots \ \pi_{i-1} \ \pi_j \ \pi_{j-1} \ \dots \ \pi_{i+1} \ \pi_i \ \pi_{j+1} \dots \ \pi_n)$. SBR calls for a shortest sequence of reversals transforming π into the identity permutation $(1 \ 2 \ \dots \ n-1 \ n)$. The length of such a sequence is denoted by d_π.

[*]Basic Research in Computer Science,
Centre of the Danish National Research Foundation.

A problem very closely related to SBR is the following *Breakpoint Graph Decomposition* (BGD). A *breakpoint graph* [1] $G = (V, E)$ is one in which:

- the edge set E is partitioned into subsets B of *black* edges and Y of *grey* edges (G is "bicolored");
- there are no parallel edges (G is "simple");
- each node of G is either isolated, or incident with one black and one grey edge, or incident with two black and two grey edges (G is "balanced" and $\Delta(G) \leq 4$);
- there is no *monochromatic cycle*, i.e. no cycle is fully contained in B or Y;

where $\Delta(G)$ denotes the maximum degree of a node of G. Let $b_G := |B|(= |Y|)$. An *alternating cycle* of G is a cycle whose edges are alternately black and grey, possibly visiting some nodes twice, but visiting each edge at most once. BGD calls for the maximum number of edge-disjoint alternating cycles of G, denoted by c_G. More precisely, the objective is to *minimize* $b_G - c^*$, where c^* is the number of alternating cycles in the solution, and the optimal solution value is $b_G - c_G$. In the following, we will refer to alternating cycles by calling them simply *cycles*.

In [1] it is shown that, given a permutation π, one can define a breakpoint graph $G(\pi)$ such that $d_\pi \geq b_{G(\pi)} - c_{G(\pi)}$. (On the other hand, [5] shows that any breakpoint graph is isomorphic to $G(\pi)$ for some permutation π.) For short, we will omit subscripts in the following and simply use the notation d, b, and c. Even if in the worst case d may be as large as $\frac{3}{2}(b - c)$ (but not more) [5], the extensive computational results carried out in [13, 7, 8] as well as the probabilistic analysis of [6] showed that $d = b - c$ in almost all cases, namely with probability $1 - \Theta(\frac{1}{n^5})$ for a permutation of n elements. More precisely, given a BGD solution of value $b - c$, in almost all cases one can immediately derive an SBR solution of the same value. This motivates the study of BGD itself, which has the advantage of being simpler than SBR in many respects. In particular, in the study of BGD one does not have to deal with complex combinatorial objects called *hurdles* [11], that typically make results for SBR much harder to prove than their counterparts for BGD.

At present, the best known approximation ratio achievable for both SBR and BGD is $\frac{3}{2}$, due to Christie [9]. One may wonder whether this ratio is the best possible. In [4], Berman and Karpinski showed that the two problems are APX-hard, namely they cannot be approximated within a ratio better than $1,0008$ in polynomial time unless P=NP, and posed as a challenging question the improvement of either the $1,0008$ lower bound or the $\frac{3}{2}$ upper bound. In this paper, we improve the approximation achievable for BGD, showing how to get a $\frac{33}{23} + \epsilon$ approximation for any $\epsilon > 0$. This proves that the $\frac{3}{2}$ ratio is not the best possible, at least for BGD. Moreover, our result makes use of the best known approximation algorithms for Stable Set on graphs with maximum degree 4 as well as for Set Packing where the maximum size of a set is 6, and any improvement in the ratio achieved by these approximation algorithms will yield an automatic improvement of our result.

2 The main scheme

Consider an optimal BGD solution and let c_{2k} denote the number of corresponding cycles of length $2k$ for $k = 2, 3, \ldots$. Note that $b = 2c_4 + 3c_6 + 4c_8 + \ldots \geq 2c$, and assume without loss of generality $b \geq 1$, as BGD is trivial when $E = \emptyset$ (this happens if and only if the input π to SBR is the identity permutation). The results in [4] imply that finding a largest collection of cycles of length 4 (and also of length $\leq 2k$ for any given $k \geq 2$) is NP-hard.

Our approximation algorithm is based on efficiently finding two collections of edge-disjoint cycles, one containing at least αc_4 cycles (of length 4) and the other containing at least $\beta(c_4 + c_6)$ cycles (of length ≤ 6). Therefore, the final objective value for BGD is $b - c^*$, where $c^* \geq \max\{\alpha c_4, \beta(c_4 + c_6)\}$. Before our work, the best known guarantees achievable in polynomial time for α and β were $\frac{1}{2}$ (see [9]) and $\frac{1}{3} - \epsilon$ for any $\epsilon > 0$ (see [12]), respectively. It is known and it will be clear from the discussion below that the bottleneck in order to improve on the $\frac{3}{2}$ approximation for BGD is the value $\frac{1}{2}$ for α. Accordingly, most of the paper will be devoted to the illustration of an improvement on this value. In particular, we will show that the problem of finding a largest collection of cycles of length 4 in G can be stated as the problem of finding a largest stable set in a suitable graph G^* with $\Delta(G^*) \leq 4$. Hence, we will be able to push α up to $\frac{5}{7} - \epsilon$ for any $\epsilon > 0$, which is the best known approximation guarantee for this version of Stable Set [3]. In particular, this guarantee is $\frac{5}{\Delta(G^*)+3} - \epsilon$. Here is a formal statement of the result that we will prove in the next section.

Lemma 2.1 *The problem of finding a largest collection of edge disjoint cycles of length 4 in a breakpoint graph G can be reduced to a Stable Set problem on a graph G^* with $\Delta(G^*) \leq 4$, for which the currently best known ratio achievable in polynomial time is $\frac{5}{7} - \epsilon$ for any $\epsilon > 0$.*

We did not succeed in improving the $\frac{1}{3} - \epsilon$ value for β. The same approach used to improve the value of α seems useless for this purpose. In particular, this approach considers only cycles of length 4 along with the fact that not many (at most 6, as shown in the next section) such cycles may share an edge with another given cycle of length 4. When also cycles of length 6 are considered, it is easy to realize that the number of such cycles sharing an edge with a given cycle of length 6 can be much larger, e.g. equal to 18.

The approximation ratio of $\frac{1}{3} - \epsilon$ is achieved by using a general technique to approximate the following problem, called *p-Set Packing*. The well known *Set Packing* problem is defined by a ground set F and a collection $S_1, \ldots, S_n$ of subsets of F. Two subsets S_i and S_j are called *independent* if $S_i \cap S_j = \emptyset$, and the objective is to find a largest subcollection of pairwise independent subsets. If the cardinality of each subset in the collection is bounded by a constant p, the problem is called p-Set Packing. Hurkens and Schrijver [12] described a local search scheme for p-Set Packing that achieves an approximation ratio of $\frac{2}{p} + \epsilon$ for any $\epsilon > 0$. Clearly, the problem of finding a largest collection of cycles of G of length at most 6 can be formulated as a 6-Set Packing problem where $F = E$ and the collection of subsets corresponds to all cycles of length ≤ 6. To formalize this discussion, we state the following

Lemma 2.2 *The problem of finding a largest collection of edge disjoint cycles of length ≤ 6 in G can be formulated as a 6-Set Packing problem, for which the currently best known ratio achievable in polynomial time is $\frac{1}{3} - \epsilon$ for any $\epsilon > 0$.*

The next result illustrates the approximation ratio that is achieved by the BGD solution depending on the values of α and β. In particular, one should compare the heuristic solution value $b - c^*$, where $c^* \geq \max\{\alpha c_4, \beta(c_4 + c_6)\}$, and the optimal solution value $b - c$. We note that, generalizing Lemma 2.2 in a straightforward way, we may also obtain a number of cycles at least equal to $\frac{1}{k}(c_4 + c_6 + \ldots + c_{2k}) - \epsilon$ for any $\epsilon > 0$ and $k = 4, 6, \ldots$, but this does not help in improving the approximation guarantee.

Lemma 2.3 *Let* $c^* \geq \max\{\alpha c_4, \beta(c_4 + c_6)\}$*, where* $0 \leq \beta < \alpha \leq 1$*. Then,*

$$\frac{b - c^*}{b - c} \leq \max\left\{\frac{4}{3}, 2 - \alpha, \frac{3 - \beta}{2}, \frac{3\alpha - \beta - \alpha\beta}{2\alpha - \beta}\right\}. \tag{1}$$

Proof: We prove the claim by solving the optimization problem

$$\max \frac{b - c^*}{b - c} \tag{2}$$

subject to

$$c \geq c_4 + c_6, \tag{3}$$

$$c \leq c_4 + c_6 + \frac{b - 2c_4 - 3c_6}{4}, \tag{4}$$

$$c^* \geq \max\{\alpha c_4, \beta(c_4 + c_6)\}, \tag{5}$$

$$b \geq 1, \tag{6}$$

$$c_4, c_6 \geq 0. \tag{7}$$

Constraint (4) follows from the fact that every cycle of length ≥ 8 contains at least 4 black (and grey) edges. The integrality of the variables does not have to be imposed explicitly, as any rational solution can be scaled by a suitable factor so as to obtain an integer solution of the same value (below we will show that we can restrict our attention to rational solutions).

Note first that c appears at the denominator of the objective function (2) with negative coefficient and is bounded by (3) and (4), therefore the maximum is attained when c takes its maximum value, i.e. when (4) is satisfied at equality. This allows us to remove variable c along with (4), replace (3) by

$$b \geq 2c_4 + 3c_6, \tag{8}$$

and write the new objective function

$$\max \frac{b - c^*}{\frac{3}{4}b - \frac{1}{2}c_4 - \frac{1}{4}c_6}. \tag{9}$$

Of course, the maximum is attained when (5) is satisfied at equality. We consider separately the two cases $c^* = \alpha c_4$ and $c^* = \beta(c_4 + c_6)$.

In the first case, $\alpha c_4 \geq \beta(c_4 + c_6)$. The problem can therefore be rewritten as (9) subject to

$$\alpha c_4 \geq \beta(c_4 + c_6) \tag{10}$$

$$b \geq 2c_4 + 3c_6 \tag{11}$$

$$b \geq 1 \tag{12}$$

$$c_6 \geq 0. \tag{13}$$

In particular, the non-negativity of c_4 is implied by (10) and the fact that $\alpha \geq \beta$. This is a *fractional linear programming* problem, which is the generalization of a linear programming problem in which the objective function is the ratio of two linear functions. It is well known [10] **WARNING: Alberto, mi sbagliavo quando avevo fornito questa referenza** that, provided the objective function is bounded in the feasible region F, the maximum is attained in an extreme point of F. Note that in our case the objective function is bounded both from below and from above.

The extreme points are found by imposing equality in three out of the four inequality constraints. We consider separately the 4 cases, indicating the inequality that is not tight for each of them.

(10) is not tight: We have $c_6 = 0$, $b = 1$ and $c_4 = \frac{1}{2}$, and the objective value is

$$2 - \alpha. \tag{14}$$

(11) is not tight: We have $c_4 = c_6 = 0$ and $b = 1$, and the objective value is

$$\frac{3}{4}. \tag{15}$$

(12) is not tight: We would have $b = c_4 = c_6 = 0$, which is clearly infeasible.

(13) is not tight: We have $b = 1$, $c_4 = \frac{\beta}{\alpha-\beta}c_6$ and $c_6 = \frac{\alpha-\beta}{3\alpha-\beta}$, i.e. $c_4 = \frac{\beta}{3\alpha-\beta}$, and the objective value is

$$\frac{1 - \frac{\alpha\beta}{3\alpha-\beta}}{\frac{3}{4} - \frac{\beta}{2(3\alpha-\beta)} - \frac{\alpha-\beta}{4(3\alpha-\beta)}} = \frac{3\alpha - \beta - \alpha\beta}{2\alpha - \beta}. \tag{16}$$

We now consider the case $c^* = \beta(c_4 + c_6)$, implying $\alpha c_4 \leq \beta(c_4 + c_6)$. The problem can be rewritten as (9) subject to

$$\alpha c_4 \leq \beta(c_4 + c_6) \tag{17}$$

$$b \geq 2c_4 + 3c_6 \tag{18}$$

$$b \geq 1 \tag{19}$$

$$c_4 \geq 0. \tag{20}$$

In particular, the non-negativity of c_4 is implied by (17) and the fact that $\alpha \geq \beta$. In this case, the extreme points correspond to the following cases:

(17) is not tight: We have $c_4 = 0$, $b = 1$ and $c_6 = \frac{1}{3}$, and the objective value is

$$\frac{3 - \beta}{2}. \tag{21}$$

(18) is not tight: We have $c_4 = c_6 = 0$ and $b = 1$, and the objective value is as in (15).

(19) is not tight: We would have $b = c_4 = c_6 = 0$, which is clearly infeasible.

(20) is not tight: We have $b=1$, $c_4 = \frac{\beta}{\alpha-\beta}c_6$ and $c_6 = \frac{\alpha-\beta}{3\alpha-\beta}$, and the objective value is as in (16).

The proof then follows from (14), (15), (16), and (21). □

As a consequence of Lemmas 2.1, 2.2 and 2.3, we have the improved approximation for BGD, obtained by plugging in the values of α and β in (1).

Theorem 2.4 *The approximation ratio achieved for BGD is* $\frac{33}{23}+\epsilon$ *for any* $\epsilon > 0$.

3 Cycles of length 4 and stable sets: Proof of Lemma 2.1

In this section we prove Lemma 2.1. We will only consider (alternating) cycles of length 4, called C4's for short. In many points in our proofs we will exclude the presence of monochromatic cycles, also called *black* or *grey cycles* depending on the color of their edges.

Let G^* be the graph having one node for each C4 of G and one edge connecting each pair of C4's that share an edge in G. The problem of finding a largest collection of edge disjoint C4's in G is clearly equivalent to the problem of finding a stable set of maximum cardinality in G^*. We will propose simple reductions for this second problem, in case G^* has a node of degree ≥ 5. The effect will be to transform the problem of finding a largest collection of edge disjoint C4's into a stable set problem in a graph G^* with $\Delta(G^*) \leq 4$, proving the lemma.

We say that two edges of G are *independent* if they have no common endpoint. The fact that G is simple implies

Fact 3.1 *Let e and f be two edges contained in a same C4. Then e and f are independent if and only if they have the same color.*

Fact 3.2 *Two C4's can share at most two edges. Moreover, if they share two edges then these two edges have different colors.*

Proof: Let C_1 and C_2 be two C4's. If C_1 and C_2 have at least three edges in common then $C_1 = C_2$ since G is simple. Let e and f be two edges contained both in C_1 and in C_2. By Fact 3.1, if e and f have the same color then they are independent. Here, $C_1 = C_2$ follows again since G has no monochromatic cycle. □

Fact 3.3 *Each edge belongs to at most three C4's.*

Proof: Let C_0, C_1, C_2 and C_3 be four distinct C4's using edge uv. We can assume that uv is black, and that xu and yv are the two grey edges of C_0. By Fact 3.2, there must exist two further grey edges $\bar{x}u$ and $\bar{y}v$ adjacent to uv and we can assume w.l.o.g. the following scenario: $xu, uv, v\bar{y} \in C_1$, $\bar{x}u, uv, vy \in C_2$, and $\bar{x}u, uv, v\bar{y} \in C_3$. But then G would contain a black cycle, made up by the following 4 edges: $x\bar{y}$ from C_1, $\bar{y}\bar{x}$ from C_3, $\bar{x}y$ from C_2, and yx from C_0. □

The next lemma shows that $\Delta(G^*) \leq 6$ and identifies those configurations in G that lead to a node of degree 5 or 6 in G^*. Since Stable Set on graphs with maximum degree 6 can be approximated within $\frac{5}{9}-\epsilon$ for any $\epsilon > 0$, by Lemma 2.3 this would already imply an approximation of $\frac{31}{21}+\epsilon$ for BGD.

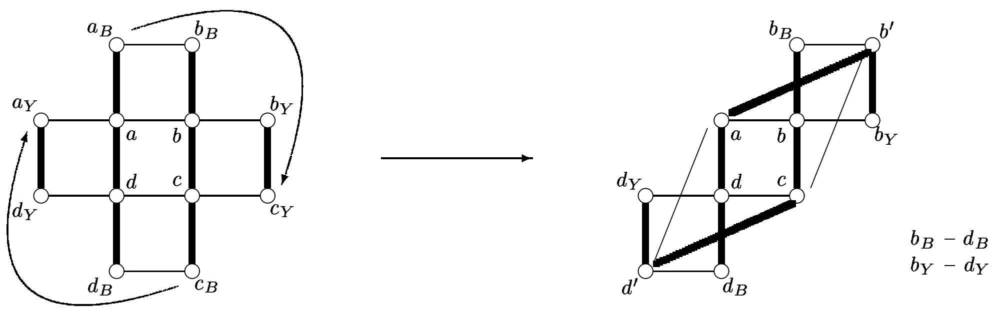

Figure 1: Configuration corresponding to a node of degree 6 in G^*.

Lemma 3.4 *No node of G^* has degree more than 6. Moreover, to each node of degree 6 there corresponds the configuration given in Fig. 1. Excluding cases which are equivalent by symmetry, to each node of degree 5 there corresponds one of the configurations given in Figs. 2, 3, 4, and 5, called* Type A*,* Type B*,* Type C *and* Type D configuration*, respectively.*

Proof: Let C be a node of G^*, i.e. a C4 of G. Let ab and cd be the two grey edges of C and bc and da be the two black edges of C. Let x be the number C4's of G sharing precisely one edge with C. Let y be the number C4's of G sharing precisely two edges with C.
Claim 1: $x \leq 4$. Indeed, assume $x > 4$. Let C_1 and C_2 be two C4's containing a same edge (w.l.o.g. ab) of C and such that C_1 and C_2 share only edge ab with C. Since at most two black edges are incident with every node of G, it follows that C_1 and C_2 must have both black edges in common. This is in contradiction with Fact 3.2. $\diamond$

Assume $\Delta(G^*) > 6$, i.e. $x + y \geq 7$. By Fact 3.3, each edge of C belongs to at most two C4's other than C. Since C has four edges $x + 2y \leq 4 \cdot 2 = 8$. Combining the two inequalities we get $x \geq 6$, a contradiction.

Next, let us consider the case in which C has degree 6, i.e. $x + y = 6$. Combining again with $x + 2y \leq 8$ we get $y \leq 2$ and $x \geq 4$. By the claim above, this implies $x = 4$ and $y = 2$. Let C_{ab} (resp. C_{bc}, C_{cd} and C_{da}) be the C4 sharing precisely edge ab with C (resp. edges bc, cd and da). Let aa_B and bb_B be the two black edges of C_{ab}. Even if in this way a same node can be referred to by more than one name, we call bb_Y and cc_Y the two grey edges of C_{bc}, cc_B and dd_B the two black edges of C_{cd}, dd_Y and aa_Y the two grey edges of C_{da}. Since $y = 2$, we must now exhibit the two C4's, say C_1 and C_2, having precisely two edges in common with C. By Fact 3.2, we can assume w.l.o.g. that C_1 contains edges ab and bc. Again, by Fact 3.2, this forces the edges of C_1 to be ab, bc, cc_Y and aa_B. Therefore, c_Y and a_B are actually the same node.

If C_2 contains edges cd and da, then the remaining two edges of C_2 are cc_B and aa_Y and $c_B = a_Y$. This case corresponds to the configuration given in Fig. 1, as stated by the lemma. Note that nodes b_B and d_B can still coincide. The same holds for nodes b_Y and d_Y, even if the two pairs cannot coincide at the same time, otherwise G would contain a monochromatic cycle. No two other nodes can coincide, since G is simple, with $\Delta(G) \leq 4$ and no monochromatic cycle.

Otherwise, we can assume by symmetry that C_2 contains edges bc and cd. In this case,

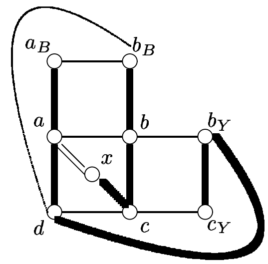

Figure 2: Type A configuration corresponding to a node of degree 5 in G^*.

the remaining two edges of C_2 are bb_Y and dd_B and $b_Y = d_B$. But then G contains the black cycle d_Bd, da, aa_B, c_Yb_Y.

Finally, let us consider the case in which C has degree 5, i.e. $x + y = 5$. Combining again with $x + 2y \leq 8$, we get $y \leq 3$ and $x \geq 2$.

Case 1: $x = 2$ and $y = 3$. (Type A configuration, see Fig. 2.)

Let C_1 and C_2 be the two C4's with precisely one edge in common with C.

Assume first C_1 contains edge ab and C_2 contains edge cd. We will show that this leads to a contradiction. Even if a same node can receive several names, call aa_B and bb_B the two black edges of C_1, and cc_B and dd_B the two black edges of C_2. Since $y = 3$, we must now exhibit the three C4's having precisely two edges in common with C. By symmetry, we can assume to have one which contains edges ab and bc and another which contains edges ab and ad. By Fact 3.2, the first one contains edge aa_B and a grey edge with one endpoint in a_B and the other in c. The second one contains edge bb_B and a grey edge with one endpoint in b_B and the other in d. But then G contains the grey cycle a_Bc, cd, db_B, b_Ba_B.

Assume now, by symmetry, that C_1 contains edge ab and C_2 contains edge bc. Even if a same node can receive several names, call aa_B and bb_B the two black edges of C_1, and bb_Y and cc_Y the two grey edges of C_2. Since $y = 3$, we must now exhibit the three C4's having precisely two edges in common with C. By symmetry, we can assume that one of these C4's contains edges ab and ad, containing also edge bb_B as well as a grey edge with one endpoint in b_B and the other in d. Note that none of these C4's can contain both ab and bc, since otherwise it would also contain edges aa_B and cc_Y, i.e. a_B and c_Y would coincide, with the consequence that G would contain the grey cycle db_B, b_Ba_B, c_Yc, cd. It follows that we must also have a C4 containing edges bc and cd, and a C4 containing edges ad and dc. The first one contains edge bb_Y as well as a black edge with one endpoint in b_Y and the other in d. The second one contains a grey edge with an endpoint in a and a black edge with an endpoint in c, and these two edges must have their other endpoint in common. Call x this common endpoint. Hence, this case corresponds to the configuration given in Fig. 2. Note that x cannot coincide with any of the other nodes seen so far. In fact, no two nodes of the configuration in Fig. 2 can coincide, since G is simple, with $\Delta(G) \leq 4$ and no monochromatic cycle. ◇

Case 2: $x = 3$ and $y = 2$. (Type B and Type C configurations, see Figs. 3 and 4.)

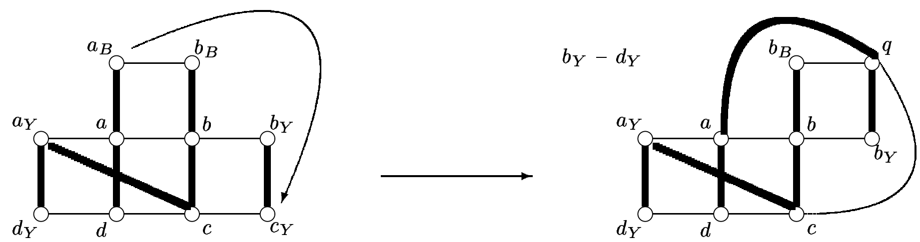

Figure 3: Type B configuration corresponding to a node of degree 5 in G^*.

By symmetry, we can assume that the three C4's containing exactly one edge of C are C_{da}, C_{ab} and C_{bc}, sharing with C edges da, ab and bc respectively. Even if a same node can receive several names, call dd_Y and aa_Y the two grey edges of C_{da}, aa_B and bb_B the two black edges of C_{ab}, bb_Y and cc_Y the two grey edges of C_{bc}. Since $y = 2$, we must now exhibit the two C4's, say C_1 and C_2, having precisely two edges in common with C.

Assume first C_1 contains the edges ab and bc. Then, by Fact 3.2, C_1 contains edges aa_B and cc_Y, and nodes a_B and c_Y must coincide. Now, C_2 cannot contain edges ab and ad since otherwise b_B and d_Y would coincide, and G would contain the grey cycle c_Yc, cd, dd_Y, b_Ba_B. Moreover, C_2 cannot contain edges bc and cd since otherwise C_2 would contain edge bb_Y as well as a black edge with one endpoint in b_Y and the other in d. Again G would contain the black cycle da, aa_B, c_Yb_Y, b_Yd. Hence, C_2 contains the edges ad and dc. So, C_2 contains edge aa_Y as well as a black edge with one endpoint in a_Y and the other in c. This case corresponds to the configuration given in Fig. 3. Note that nodes b_Y and d_Y can still coincide. No two other nodes can coincide, since G is simple, with $\Delta(G) \leq 4$ and no monochromatic cycle.

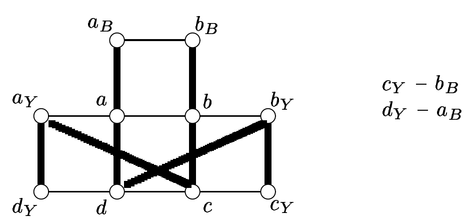

Figure 4: Type C configuration corresponding to a node of degree 5 in G^*.

Assume now, by symmetry, that C_1 contains edges ad and dc and C_2 contains edges bc and cd. So, C_1 contains edge aa_Y as well as a black edge with one endpoint in a_Y and the other in c. Moreover, C_2 contains edge bb_Y as well as a black edge with one endpoint in b_Y and the other in d. This case corresponds to the configuration given in Fig. 4. Note that nodes c_Y and b_B can still coincide, as well as nodes d_Y and a_B, even if the two pairs cannot coincide at the same time. No two other nodes can coincide, since G is simple, with $\Delta(G) \leq 4$ and no monochromatic cycle. ◇

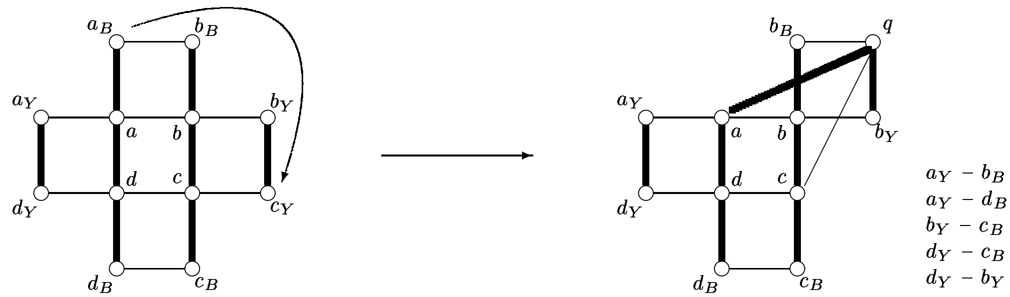

Figure 5: Type D configuration corresponding to a node of degree 5 in G^*.

Case 3: $x = 4$ and $y = 1$. (Type D configuration, see Fig. 5.)
Let C_{ab} (resp. C_{bc}, C_{cd} and C_{da}) be the C4's sharing precisely edge ab with C (resp. edge bc, cd and da). Even if a same node can receive several names, call aa_B and bb_B the two black edges of C_{ab}, bb_Y and cc_Y the two grey edges of C_{bc}, cc_B and dd_B the two black edges of C_{cd}, dd_Y and aa_Y the two grey edges of C_{da}. Since $y = 1$, we must now exhibit a C4, say $\tilde{C}$, having precisely two edges in common with C. We can assume w.l.o.g. that $\tilde{C}$ contains edges ab and bc. This forces the edges of $\tilde{C}$ to be ab, bc, cc_Y and aa_B, i.e. c_Y and a_B are actually the same node. Hence, this corresponds to the configuration given in Fig. 5. We can also identify one of the following pairs: a_Y and b_B; a_Y and d_B; b_Y and c_B; d_Y and c_B; d_Y and b_Y. Note that we cannot identify a_Y and c_B as we would get the degree 6 configuration in Fig. 1. □

3.1 Degree 6 configurations

In this subsection, we show how to get rid of the nodes of degree 6 in G^* by proving that a certain set of neighbors of a degree 6 node in G^* is contained in some optimal stable set of G^*. This allows one to remove from G^* this set of nodes, and address a reduced problem on a graph $\tilde{G}^*$ with $\Delta(\tilde{G}^*) \leq 5$. Note that

Fact 3.5 *The graph obtained from a breakpoint graph by removing the edges in a C4 is a breakpoint graph as well.*

Accordingly, the above reduction on G^* has an immediate counterpart on G, and one can operate on a reduced breakpoint graph in which each C4 intersects at most five C4's.

Due to Lemma 2.3 and the results in [2], limiting the degree of G^* to 5 already yields an approximation of $\frac{32}{22} - \epsilon$ for BGD.

Let H be the graph given in Fig. 6. Let $\bar{H}$ be the subgraph of H induced by the nodes in $V(H) \setminus \{C_b, C_d\}$. One can easily check that all nodes in $V(\bar{H}) = V(H) \setminus \{C_b, C_d\}$ correspond to C4's actually present in the configuration given in Fig. 1 and that G^* contains $\bar{H}$ as an induced subgraph. We have the following.

Lemma 3.6 *Assume G^* contains a node of degree 6. Correspondingly, G^* contains $\bar{H}$ as an induced subgraph. Then there exists a maximum stable set of G^* which includes the nodes $C_{d'd_Y dd_B}, C_{b'b_Y bb_B}, C_{d'adc}$ and $C_{b'cba}$.*

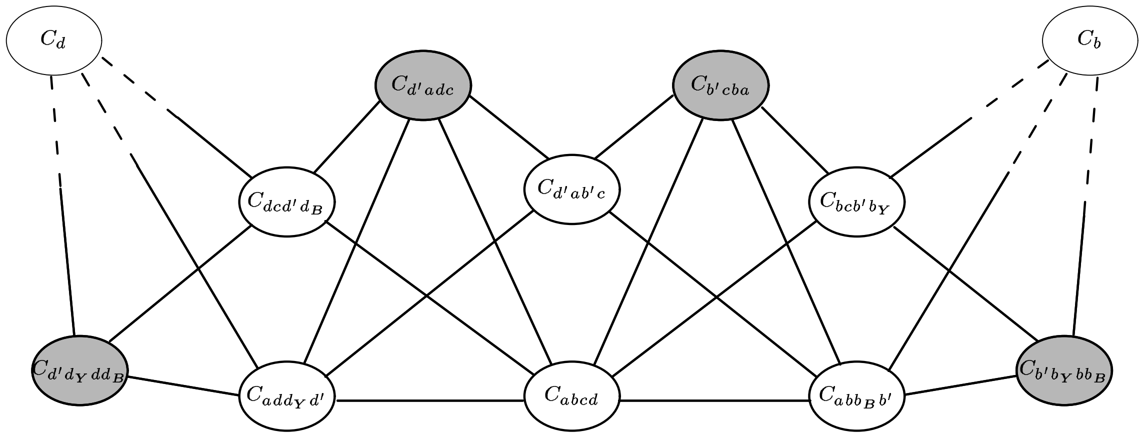

Figure 6: Degree 6 configuration in G^*.

Proof: Note first that the four nodes $C_{d'd_Y dd_B}, C_{b'b_Y bb_B}, C_{d'adc}$ and $C_{b'cba}$ form a stable set in $\bar{H}$ and hence in G^*. To prove the lemma, we will show that four is the size of a largest stable set in the graph $[\bar{H}]$, which is defined as the subgraph of G^* induced by the nodes of $\bar{H}$ and their neighbors. To this end, we will first examine these possible neighbors.

Let $\tilde{C}$ be a C4 of G which is not a node of $\bar{H}$. Since the degree of C_{abcd} is 6, $\tilde{C}$ cannot contain any of the edges ab, bc, cd or da. Moreover, $\tilde{C}$ cannot contain any of the edges ad', cd', ab' or cb' either. Indeed, by symmetry, assume $\tilde{C}$ contains edge ad'. Since $\tilde{C}$ does not contain ad, then it contains ab'. If the other black edge of $\tilde{C}$ is $d'c$, then $\tilde{C} = C_{d'ab'c}$, i.e. $\tilde{C}$ is a node of $\bar{H}$. If the other black edge of $\tilde{C}$ is $d'd_Y$, then $\tilde{C}$ contains a grey edge with one endpoint in b' and the other in d_Y. But then G contains the grey cycle $d_Y d, dc, cb', b'd_Y$ and we have a contradiction.

Consider now the case in which $\tilde{C}$ contains one of the edges dd_Y, dd_B, bb_Y or bb_B. By symmetry, assume that $\tilde{C}$ contains edge dd_Y. Since $da \notin \tilde{C}$ then $dd_B \in \tilde{C}$. If $d_Y d' \in \tilde{C}$ then $\tilde{C} = C_{dd_Y d'd_B}$, i.e. $\tilde{C}$ is a node of $\bar{H}$. Hence, $\tilde{C}$ contains a black edge with an endpoint in d_Y and a grey edge with an endpoint in d_B. These two edges must have their other endpoint in common. Note that $\tilde{C}$ is adjacent to the following nodes in G^*: $C_{dcd'd_B}, C_{add_Y d'}$ and $C_{d'd_Y dd_B}$. Hence, $\tilde{C}$ corresponds to node C_d in Fig. 6. Analogously, if $\tilde{C}$ contains edge bb_Y, then it corresponds to node C_b in Fig. 6.

Consider now the case in which $\tilde{C}$ contains one of the edges $d'd_Y, d'd_B, b'b_Y$ or $b'b_B$. By symmetry, assume that $\tilde{C}$ contains edge $b'b_Y$. As already seen above, $\tilde{C}$ cannot contain edge $b'c$. So, $b'b_B \in \tilde{C}$. Now, if $\tilde{C} \neq C_{b'b_Y bb_B}$, then $\tilde{C}$ does not contain $b_B b$ or $b_Y b$. Hence, $\tilde{C}$ contains a black edge with an endpoint in b_B and a grey edge with an endpoint in b_Y. These two edges must have their other endpoint in common. Note that $\tilde{C}$ is adjacent to the following nodes in G^*: $C_{bcb'b_Y}, C_{abb_B b'}$ and $C_{b'b_Y bb_B}$. Hence, $\tilde{C}$ corresponds to node C_b in Fig. 6. Analogously, if $\tilde{C}$ contains edge $d'd_Y$, then it corresponds to node C_d in Fig. 6.

By the discussion above, the possible neighbors in $V(G) \setminus V(\bar{H})$ for the nodes in $\bar{H}$ are C_b and C_d depicted in Fig. 6. It is well known that the size of a stable set in a graph is at most k if there exists a set of cliques $Q_1, \ldots, Q_k$ such that each node is contained in one of

these cliques. Let X be a stable set of $[\bar{H}]$ with $|X| = 5$.
If $C_{abcd} \in X$, then $X \setminus \{C_{abcd}\}$ is a size 4 stable set in the graph obtained from $[\bar{H}]$ by removing C_{abcd} and all of its neighbors. However, the nodes of this graph are all covered by the three cliques: $Q_1 = \{C_{d'ab'c}\}$; $Q_2 = \{C_{d'd_Y dd_B}, C_d\}$ (or simply $Q_2 = \{C_{d'd_Y dd_B}\}$ if C_d is not present in G^*); $Q_5 = \{C_{b'b_Y bb_B}, C_b\}$ (or simply $Q_5 = \{C_b\}$ if C_b is not present in G^*).
Otherwise, if $C_{abcd} \notin X$, then X is a size 5 stable set in the graph obtained from $[\bar{H}]$ by removing C_{abcd}. However, the nodes of this graph are all covered by the four cliques: $Q_1 = \{C_{d'adc}, C_{add_Y d'}, C_{d'ab'c}\}$; $Q_2 = \{C_{b'cba}, C_{abb_B b'}\}$; $Q_3 = \{C_{dcd'd_B}, C_{d'd_Y dd_B}, C_d\}$ (simply $Q_3 = \{C_{dcd'd_B}, C_{d'd_Y dd_B}\}$ if C_d is not present in G^*); $Q_4 = \{C_{bcb'b_Y}, C_{b'b_Y bb_B}, C_b\}$ (simply $Q_4 = \{C_{bcb'b_Y}, C_{b'b_Y bb_B}\}$ if C_b is not present in G^*). □

3.2 Degree 5 configurations

In the previous subsection, we saw how to get rid of degree 6 nodes in G^*. Here we will do the same for the nodes of degree 5. In the previous subsection, this was based on showing that a certain set of nodes S of G^* can simply be assumed to be contained in a maximum stable of G^*. A key reason why this actually leads to a reduction is that, after removal of the nodes in S and their neighbors, G^* still represents a breakpoint graph, namely the breakpoint graph obtained from G by deleting all edges contained in cycles corresponding to nodes in S. With degree 5 nodes we will most often only show the existence of a maximum stable of G^* which does not take certain nodes. We will hence have to exhibit also some operations for G, which on one side transform G into a new breakpoint graph, and on the other are counterparts for the reductions shown in G^* for the maximum stable set problem. In almost all cases this will be based on showing that removing some avoidable nodes in G^* corresponds to splitting a degree 4 node in the original breakpoint graph G, as illustrated in the following. Given a breakpoint graph, the *splitting* of a node w incident with black edges wu_B, wv_B and grey edges wu_Y, wv_Y corresponds to replacing w by two nodes w' and w'' and the associated edges by $w'u_B, w''v_B$ and either $w'u_Y, w''v_Y$ or $w''u_Y, w'v_Y$. We will say that two edges are *separated* by the splitting if their counterparts after the splitting are independent. We have the following

Fact 3.7 *The graph obtained from a breakpoint graph by splitting a node is a breakpoint graph as well.*

Consider a node of degree 5 in G^*. By Lemma 3.4, G contains one of the configurations given in Figs. 2, 3, 4, and 5. In G^*, these correspond to the configurations given in Figs. 7, 8, 9, 10, 11. In the next subsections, we will consider these configurations one by one.

3.2.1 Type A configuration

Fig. 2 illustrates a degree 5 configuration of type A in G, while Fig. 7 illustrates the same configuration in G^*. Let H be the graph given in Fig. 7. Let $\bar{H}$ be the subgraph of H induced by the nodes in $V(H) \setminus \{\tilde{C}\}$. One can easily check that all nodes in $V(\bar{H})$ correspond to C4's actually present in the configurations given in Fig. 2 and that G^* contains graph $\bar{H}$ as an induced subgraph.

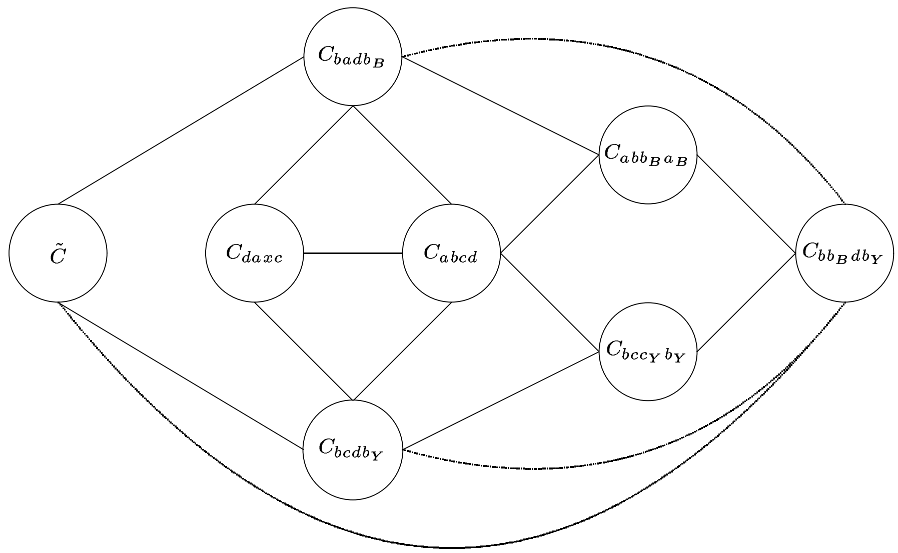

Figure 7: Type A degree 5 configuration in G^*.

Lemma 3.8 *Assume G contains the configuration shown in Fig. 2. Correspondingly, G^* contains the graph $\bar{H}$ as an induced subgraph. Then, there exists a maximum stable set X of G^* with $C_{abcd}, C_{bb_Bdb_Y} \notin X$.*

Proof: We first show that if a node $\tilde{C}$ in $V(G^*) \setminus V(\bar{H})$ is adjacent to C_{badb_B} or C_{bcdb_Y}, then the following happens:

(i) $\tilde{C}$ is adjacent to C_{badb_B}, C_{bcdb_Y} and $C_{bb_Bdb_Y}$;

(ii) no node in $V(G^*) \setminus V(\bar{H}) \setminus \tilde{C}$ is adjacent to C_{badb_B} or C_{bcdb_Y}.

Indeed, assume by symmetry that $\tilde{C}$ is adjacent to C_{badb_B}. Since C_{abcd} is a node of degree 5 of G^*, $\tilde{C}$ cannot contain any of the edges ab, bc, cd or da.

Assume $bb_B \in \tilde{C}$. Since $ab \notin \tilde{C}$, then $bb_Y \in C$. So, if $b_Bd \in \tilde{C}$, then $\tilde{C} = C_{bb_Bdb_Y}$. Otherwise, if $b_Ba_B \in \tilde{C}$, then $\tilde{C}$ contains a black edge with one endpoint in a_B and the other in b_Y. But then G contains the black cycle a_Ba, ad, db_Y, b_Ya_B.

Assume therefore $bb_B \notin \tilde{C}$ and hence $db_B \in \tilde{C}$. Since $da \notin \tilde{C}$, then $db_Y \in \tilde{C}$ and $\tilde{C}$ contains a grey edge with an endpoint in b_Y and a black edge with an endpoint in b_B. These two edges must have their other endpoint in common. In this case, $\tilde{C}$ is adjacent to C_{badb_B}, C_{bcdb_Y}, and $C_{bb_Bdb_Y}$. Since this was the only remaining possibility for $\tilde{C}$, we have proved the claim above.

Consider a maximum stable set X of G^*. The following arguments apply both if G^* contains a node $\tilde{C}$ as considered above or not. If $C_{abcd}, C_{bb_Bdb_Y} \in X$, then $X \cup \{C_{badb_B}, C_{bcdb_Y}\} \setminus \{C_{abcd}, C_{bb_Bdb_Y}\}$ is a maximum stable set of G^* with $C_{abcd}, C_{bb_Bdb_Y} \notin X$. Assume $C_{abcd} \in X$ and $C_{bb_Bdb_Y} \notin X$. In this case, $\tilde{C} \in X$ since otherwise $X \cup \{C_{badb_B}, C_{bcdb_Y}\} \setminus \{C_{abcd}\}$ would be a larger stable set. Therefore, $X \cup \{C_{badb_B}, C_{bcdb_Y}\} \setminus \{C_{abcd}, \tilde{C}\}$ is a maximum stable set of G^* with $C_{abcd}, C_{bb_Bdb_Y} \notin X$. Finally, assume $C_{abcd} \notin X$ and $C_{bb_Bdb_Y} \in X$. In this

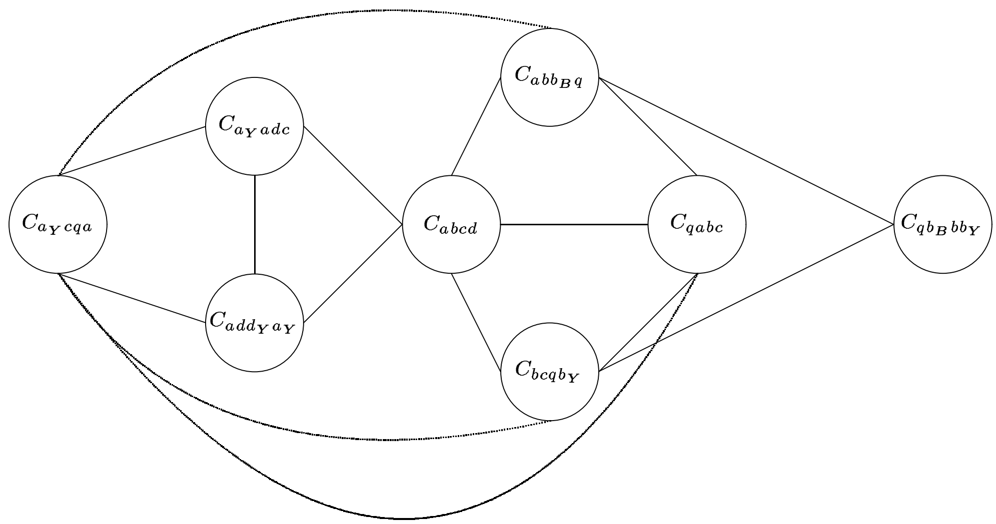

Figure 8: Type B degree 5 configuration in G^*.

case, $C_{daxc} \in X$ since otherwise $X \cup \{C_{badb_B}, C_{bcdb_Y}\} \setminus \{C_{bb_Bdb_Y}\}$ would be a larger stable set. Therefore, $X \cup \{C_{badb_B}, C_{bcdb_Y}\} \setminus \{C_{bb_Bdb_Y}, C_{daxc}\}$ is a maximum stable set of G^* with $C_{abcd}, C_{bb_Bdb_Y} \notin X$. □

The following observation shows how to modify G according to Lemma 3.8.

Observation 3.9 *Let H be the breakpoint graph obtained from G by splitting node b as to separate ab from bc and bb_Y. Then H^* is the graph obtained from G^* by removing the two nodes C_{abcd} and $C_{bb_Bdb_Y}$.*

Proof: Clearly, H^* is an induced subgraph of G^*. The C4's that are removed by splitting b as above are either of the form $C_{\widehat{abc}}$ with $ab, bc \in C_{\widehat{abc}}$, or of the form $C_{\widehat{b_Bbb_Y}}$ with $b_Bb, bb_Y \in C_{\widehat{b_Bbb_Y}}$.

If $C_{\widehat{abc}}$ contains the edge ad, then $C_{\widehat{abc}} = C_{abcd}$, whereas, if $C_{\widehat{abc}}$ contains the edge aa_B, then G contains a grey edge a_Bc and hence the grey cycle a_Bc, cd, db_B, b_Ba_B. Moreover, if $C_{\widehat{b_Bbb_Y}}$ contains the edge b_Bd, then $C_{\widehat{abc}} = C_{bb_Bdb_Y}$, whereas, if $C_{\widehat{abc}}$ contains the edge b_Ba_B, then G contains a black edge a_Bb_Y and hence the black cycle a_Bb_Y, b_Yd, da, aa_B. □

3.2.2 Type B configuration

Fig. 3 illustrates a degree 5 configuration of type B in G, while Fig. 8 illustrates the same configuration in G^*. Let $\bar{H}$ be the graph given in Fig. 8. One can easily check that all nodes in $V(\bar{H})$ correspond to C4's actually present in the configuration given in Fig. 3 and that G^* contains graph $\bar{H}$ as an induced subgraph.

Lemma 3.10 *Assume G contains the configuration shown in Fig. 8. Correspondingly, G^* contains graph $\bar{H}$ as an induced subgraph. Then, there exists a maximum stable set of G^* containing $C_{a_Y adc}$.*

Proof: We first show that no node in $V(G^*) \setminus V(\bar{H})$ is adjacent to $C_{a_Y adc}$ or C_{qabc}. Let $\tilde{C}$ be a node in $V(G^*) \setminus V(\bar{H})$ adjacent to $C_{a_Y adc}$. Since C_{abcd} is a node of degree 5 of G^*, then $\tilde{C}$ cannot contain any of the edges ab, bc, cd or da. Assume $aa_Y \in \tilde{C}$. Since $ad \notin \tilde{C}$, then $aq \in \tilde{C}$. So, if $qc \in \tilde{C}$, then $\tilde{C} = C_{a_Y cqa}$. Otherwise, $qb_B \in \tilde{C}$ and $\tilde{C}$ contains a black edge with one endpoint in b_B and the other in a_Y. This is not possible as there are already two black edges incident with a_Y and neither d_Y nor c can coincide with b_B. Now, assume $a_Y c \in \tilde{C}$, implying $qc \in \tilde{C}$. If $qa \in \tilde{C}$, again we have $\tilde{C} = C_{a_Y cqa}$. Otherwise, $qb_Y \in \tilde{C}$ and $\tilde{C}$ contains a grey edge with one endpoint in b_Y and the other in a_Y. But then G contains the grey cycle $a_Y a, ab, bb_Y, b_Y a_Y$. Hence, no node in $V(G^*) \setminus V(\bar{H})$ is adjacent to $C_{a_Y adc}$.

Now, let $\tilde{C}$ be a node in $V(G^*) \setminus V(\bar{H})$ adjacent to C_{qabc}. Note that $\tilde{C}$ is neither adjacent to C_{abcd} nor to $C_{a_Y adc}$. Assume $aq \in \tilde{C}$. Since $ab \notin \tilde{C}$ and $aa_Y \notin \tilde{C}$, we immediately have a contradiction. Similarly, assuming $qc \in \tilde{C}$, we have a contradiction as $bc \notin \tilde{C}$ and $a_Y c \notin \tilde{C}$. Hence, no node in $V(G^*) \setminus V(\bar{H})$ is adjacent to C_{qabc}.

Consider a maximum stable set X of G^* with $C_{a_Y adc} \notin X$. If $C_{bcqb_Y} \in X$, then $C_{abcd}, C_{a_Y cqa} \notin X$. Hence, $X \cup \{C_{a_Y adc}\} \setminus \{C_{add_Y a_Y}\}$ is a maximum stable set of G^* containing $C_{a_Y adc}$. Assume therefore $C_{bcqb_Y} \notin X$. Note that X contains at most one node out of $C_{a_Y cqa}, C_{a_Y adc}, C_{add_Y a_Y}$ and at most one node out of $C_{abcd}, C_{abb_B q}, C_{qabc}$, since these induce triangles. Therefore, $X \cup \{C_{a_Y adc}, C_{qabc}\} \setminus \{C_{a_Y cqa}, C_{add_Y a_Y}, C_{abcd}, C_{abb_B q}\}$ is a maximum stable set of G^* containing $C_{a_Y adc}$. □

3.2.3 Type C configuration

Fig. 4 illustrates a degree 5 configuration of type C in G, while Figs. 9 and 10 illustrate the same configuration in G^*. Let H be the graph given in Fig. 9. Let $\bar{H}$ be the subgraph of H induced by the nodes in $V(H) \setminus \{C_{b_Y c_Y tx}, C_{b_Y dd_Y x}, C_{a_Y cc_Y y}, C_{a_Y d_Y zy}\}$. This is also a subgraph of the graph in Fig. 10. One can easily check that all nodes in $V(\bar{H})$ correspond to C4's actually present in the configuration given in Fig. 4 and that G^* contains graph $\bar{H}$ as an induced subgraph.

Lemma 3.11 *Assume G contains the configuration shown in Fig. 4. Correspondingly, G^* contains graph $\bar{H}$ as an induced subgraph. If no neighbor of C_{abcd} in G^* has degree 5, then there exists a maximum stable set X of G^* with $C_{abcd} \notin X$. In case a neighbor $\overline{C}$ of C_{abcd} has degree 5, then no other neighbor of C_{abcd} or $\overline{C}$ has degree 5, and, given any stable set X in the graph obtained from G^* by deleting the edge $\overline{C}C_{abcd}$, a stable set X' of G^* with $|X'| = |X|$ can be derived from X in constant time.* **WARNING: Alberto, tutto lo statement del lemma va adeguato al fatto che invece diamo una riduzione completa. Tuttavia é meglio questo lo faccia tu parché ho visto che abbiamo gusti radicalmente diversi su come strutturare l'esposizione e tu hai piú autoritá su quale sia il gusto ufficiale. (Tu tendi ad accorpare evidenziando solo i risultati espressivi per l'applicazione, io cerco di scomporre ed organizzare in maniera modulare per**

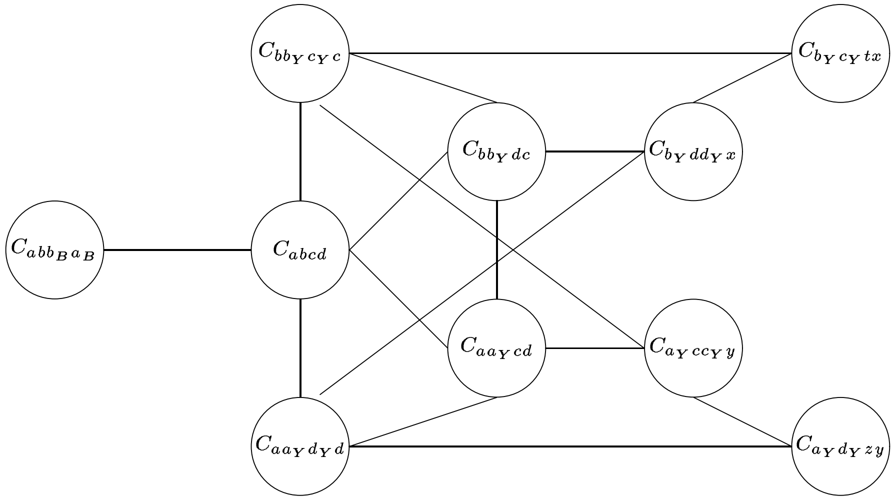

Figure 9: Type C degree 5 configuration in G^* (first case).

facilitare la comprensione). Io qui scomporrei anche questo lemma in almeno due blocchi.

Proof: Excluding cases which are equivalent by symmetry, one of the cases considered in the following must occur.
Case 1: either no node in $V(G^*) \setminus V(\bar{H})$ is adjacent to $C_{aa_Y cd}$, or no node in $V(G^*) \setminus V(\bar{H})$ is adjacent to $C_{bb_Y dc}$. Let X be a maximum stable set of G^* containing C_{abcd} and assume no node in $V(G^*) \setminus V(\bar{H})$ is adjacent to $C_{aa_Y cd}$. Then, $X \cup \{C_{aa_Y cd}\} \setminus \{C_{abcd}\}$ is a maximum stable set of G^* which does not contain C_{abcd}. The case in which no node in $V(G^*) \setminus V(\bar{H})$ is adjacent to $C_{bb_Y dc}$ is identical. This completes Case 1. $\diamond$

From now on, we will assume that Case 1 does not occur and hence both $C_{aa_Y cd}$ and $C_{bb_Y dc}$ have neighbors in $V(G^*) \setminus V(\bar{H})$. Let $\tilde{C}$ be a node in $V(G^*) \setminus V(\bar{H})$ adjacent to $C_{aa_Y cd}$. Since C_{abcd} is a node of degree 5 of G^*, $\tilde{C}$ cannot contain any of the edges ab, bc, cd or da.

Assume $aa_Y \in \tilde{C}$. Since $ad \notin \tilde{C}$, then $aa_B \in \tilde{C}$. So, if $a_Y d_Y \in \tilde{C}$, then $\tilde{C}$ contains a grey edge with one endpoint in d_Y and the other in a_B, i.e. $\tilde{C} = C_{aa_Y d_Y a_B}$. Otherwise, $a_Y c \in \tilde{C}$, and $\tilde{C}$ contains a grey edge with one endpoint in c and the other in a_B. This is not possible as there are already two grey edges incident with c and neither c_Y nor d can coincide with a_B.

Now assume $a_Y c \in \tilde{C}$ and $aa_Y \notin \tilde{C}$. Then, $cc_Y \in \tilde{C}$. If $c_Y b_Y \in \tilde{C}$, then $\tilde{C}$ contains a grey edge $a_Y b_Y$: This is not possible because then G would contain the grey cycle $a_Y a, ab, bb_Y, b_Y a_Y$. Hence, $\tilde{C}$ must contain a black edge $c_Y y$ and a grey edge ya_Y for some node y, which may be a new node or it may coincide with b_B (coincidence with other nodes is easily excluded). In this case, $\tilde{C} = C_{a_Y cc_Y y}$.

Summarizing, the two possible nodes in $V(G^*) \setminus V(\bar{H})$ adjacent to $C_{aa_Y cd}$ are $C_{aa_Y d_Y a_B}$ and $C_{a_Y cc_Y y}$. Symmetrically, the two possible nodes adjacent to $C_{bb_Y dc}$ are $C_{bb_Y c_Y b_B}$ and $C_{b_Y dd_Y x}$ for some node x which may be a new node or it may coincide with a_B. Note that x and y cannot coincide because otherwise G would contain the grey cycle $a_Y a, ab, bb_Y, b_Y x, xa_Y$.

If G contains neither a grey edge a_Bd_Y nor a grey edge b_Bc_Y, we get Case 2 below.
Case 2: G does not contain grey edge a_Bd_Y or grey edge b_Bc_Y, and there are two nodes in $V(G^*)\setminus V(\bar{H})$, one adjacent to C_{aa_Ycd} and $C_{bb_Yc_Yc}$ and the other with C_{bb_Ydc} and $C_{aa_Yd_Yd}$. In this case, we next show that either no other node in $V(G^*)\setminus V(\bar{H})$ besides $C_{b_Ydd_Yx}$ is adjacent to $C_{aa_Yd_Yd}$, or a cycle $C_{a_Yd_Yzy}$ is present, where node z maybe new or coincide with b_B or x. Indeed, a new cycle $\tilde{C}$ adjacent to $C_{aa_Yd_Yd}$ cannot contain edge aa_Y (otherwise it would be adjacent to C_{aa_Ycd}) nor edge ad (C_{abcd} has degree 5) nor edge d_Yd (otherwise it would be adjacent to C_{bb_Ydc}). Hence, $\tilde{C}$ must contain both a_Yd_Y and a_Yy. Symmetrically, either there is no other node in $V(G^*)\setminus V(\bar{H})$ besides $C_{a_Ycc_Yy}$ adjacent to $C_{bb_Yc_Yc}$, or cycle $C_{b_Yc_Ytx}$ is present, where node t maybe new or coincide with a_B or y. This situation, considering the possible presence of nodes $C_{a_Yd_Yzy}$ and $C_{b_Yc_Ytx}$ is illustrated in Fig. 9. Let X be a maximum stable set of G^* containing C_{abcd}. Then, X contains at most one of the two nodes $C_{b_Ydd_Yx}$ and $C_{b_Yc_Ytx}$ and at most one of the two nodes $C_{a_Yd_Yzy}$ and $C_{a_Ycc_Yy}$. Moreover, X contains no neighbor of C_{abcd}. Hence, X contains at most three of the nodes displayed in Fig. 9. Note that, however chosen a node $C_1 \in \{C_{b_Ydd_Yx}, C_{b_Yc_Ytx}\}$ and a node $C_2 \in \{C_{a_Yd_Yzy}, C_{a_Ycc_Yy}\}$, there exists a node in $\{C_{bb_Yc_Yc}, C_{bb_Ydc}, C_{aa_Ycd}, C_{aa_Yd_Yd}\}$ which is neither adjacent to C_1 nor to C_2. Hence, there exists a maximum stable set of G^* not containing C_{abcd}. This completes Case 2. ◇

From now on, we will assume that Case 2 does not occur and hence that G contains either a grey edge a_Bd_Y or a grey edge b_Bc_Y. Note that these two edges cannot be present at the same time, for otherwise G would contain the grey cycle $d_Yd, dc, cc_Y, c_Yb_B, b_Ba_B, a_Bd_Y$. Assume therefore, by symmetry, that G contains edge a_Bd_Y. This implies the presence of cycle $C_{aa_Yd_Ya_B}$, adjacent to C_{aa_Ycd}. Since we are assuming that also C_{bb_Ydc} has a neighbor in $V(G^*)\setminus V(\bar{H})$ (otherwise we would be in Case 1), G must contain edges d_Yx and xb_Y, yielding cycle $C_{b_Ydd_Yx}$. Note that edges ya_Y and c_Yy may or may not be present, yielding in the first case Case 4 and in the second Case 3. The situation is illustrated in Fig. 10, where node $C_{a_Ycc_Yy}$ is present only in Case 4.
Case 3: G contains a grey edge a_Bd_Y and the only node in $V(G^*)\setminus V(\bar{H})$ adjacent to C_{aa_Ycd} is $C_{aa_Yd_Ya_B}$. Let X be a maximum stable set of G^* containing C_{abcd}. If $C_{bb_Yc_Yc}$ has no neighbor in $V(G^*)\setminus V(\bar{H})$, then the set of neighbors of $C_{bb_Yc_Yc}$ is a subset of the neighbors of C_{abcd}. Therefore, $X \cup \{C_{bb_Yc_Yc}\} \setminus \{C_{abcd}\}$ is a maximum stable set of G^* not containing C_{abcd}. Otherwise, let $\tilde{C}$ be a node in $V(G^*)\setminus V(\bar{H})$ adjacent to $C_{bb_Yc_Yc}$. The arguments above exclude the presence in $\tilde{C}$ of edges bc, bb_Y ($\tilde{C}$ would contain edge b_Bc_Y), and cc_Y ($\tilde{C}$ would contain either edge bc or a_Yc, being adjacent to C_{aa_Ycd} in the latter case). Hence $\tilde{C} = C_{b_Yc_Ytx}$ for some node t. Note that $\tilde{C}$ is also adjacent to $C_{b_Ydd_Yx}$. If X does not contain $C_{b_Yc_Ytx}$, then the above considerations apply. Otherwise, X does not contain $C_{b_Ydd_Yx}$ and $X \cup \{C_{bb_Ydc}\} \setminus \{C_{abcd}\}$ is a maximum stable set of G^* not containing C_{abcd}. This completes Case 3. ◇
Case 4: G contains grey edge a_Bd_Y as well as black edges d_Yx, c_Yy and grey edges xb_Y, ya_Y, where x may be a new vertex or it may coincide with b_B and y may be a new vertex or it may coincide with a_B. We have the situation depicted in Fig. 10. We first show that the only neighbors of $C_{aa_Yd_Yd}$ in $V(G^*)\setminus V(\bar{H})$ are $C_{aa_Yd_Ya_B}$ and $C_{b_Ydd_Yx}$. Indeed, reasoning as in the previous paragraph for the cycle incident with $C_{bb_Yc_Yc}$, namely $C_{b_Yc_Ytx}$, the only possibility for a cycle $\tilde{C}$ incident with $C_{aa_Yd_Yd}$ would be $\tilde{C} = C_{a_Yd_Yzy}$ for some node z, but z cannot coincide with a_B nor with d, which are the two nodes connected to d_Y by a grey edge.

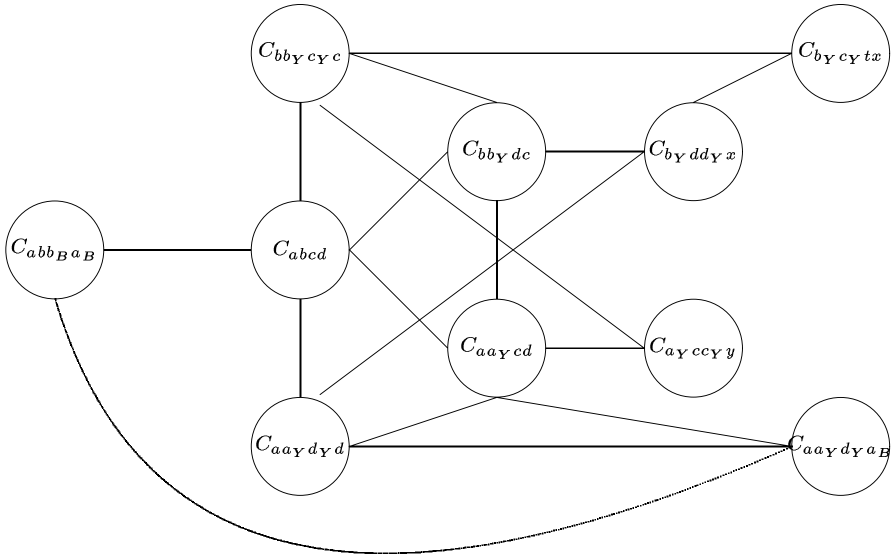

Figure 10: Type C degree 5 configuration in G^* (second case).

Hence, all the neighbors in G^* for the nodes in $V(\bar{H}) \setminus \{C_{abb_B a_B}\}$ are depicted in Fig. 10.

Consider the duality between the two nodes C_{abcd} and $C_{aa_Y cd}$, which are both centers of Type C degree 5 configurations in Case 4, under switching black and grey and substituting $c_Y \leftrightarrow a_B$, $b \leftrightarrow a_Y$, $b_y \leftrightarrow d_Y$, $b_B \leftrightarrow y$, etc..

If node $C_{aa_Y d_Y a_B}$ has in G^* only the two neighbors depicted in Fig. 10, then there exists a maximum stable set X of G^* with $C_{abcd} \notin X$. Thus we assume $C_{aa_Y d_Y a_B}$ to have a further neighbor, but it is easily verified that this can only happen as follows: G contains a grey edge hk and black edges hb_B and ka_B. Dually, if node $C_{abb_B a_B}$ has in G^* only the two neighbors depicted in Fig. 10, then there exists a maximum stable set X of G^* with $C_{aa_Y cd} \notin X$. Thus we assume $C_{abb_B a_B}$ to have a further neighbor, but it is easily verified that this can only happen as follows: G contains a black edge pq and grey edges py and qc_Y. If $x = p$ (or dually, $x = h$), then it is easily shown that there exists a maximum stable set of G^* containing the following nodes: $C_{abb_B a_B}$, $C_{aa_Y d_Y a_B}$, $C_{bb_Y dc}$, $C_{aa_Y d_Y d}$. Assume therefore $x \neq p, h$. It is easily verified that, among the C4's already present in the configuration now, only $C_{hb_B a_B k}$ and $C_{pyc_Y q}$ can be adjacent to further C4's. So, $C_{hb_B a_B k}$ and $C_{pyc_Y q}$ give a 2-node cut in G^*, and, where $\tilde{C}$ is the set of C4's already present in the configuration now and X is a maximum stable set in G^*, then either $|X \cap \tilde{C}| = 5$ and X contains both $C_{hb_B a_B k}$ and $C_{pyc_Y q}$ or $|X \cap \tilde{C}| = 4$ and we can assume $C_{hb_B a_B k}, C_{pyc_Y q} \notin X$. But then we have a reduction which does not spoil the approximation guarantee if we substitute the whole configuration in G with the configuration shown in picture. It can be verified that this can not introduce monochromatic cycles, hence we obtain a new breakpoint graph.

```
h          b_B
*=========*==========*
```

```
  |         |         |
  |         |         |
  |         |         |
k *=========*=========* y                *-------------*-------------*
     c_Y=a_B|         |             C_{hb_Ba_Bk}                  C_{pyc_Yq}
            |         |
            |         |
         q *==========* p

  new configuration in G                 new configuration in G^*

                           PICTURE
```

The following observation shows how to modify G according to Lemma 3.11 in case there exists a maximum stable set X of G^* with $C_{abcd} \notin X$.

Observation 3.12 *There exists a breakpoint graph $\tilde{G}$, obtained from G by suitably splitting node a or node b, such that $\tilde{G}^*$ is the graph obtained from G^* by removing the node C_{abcd}.*

Proof: Consider a C4 $\tilde{C}$ which is removed by splitting node a as to separate ab from ad and aa_Y. Then, either $\tilde{C}$ contains both ab and ad, or $\tilde{C}$ contains both aa_B and aa_Y. If $ab, ad \in \tilde{C}$, then either $dc \in \tilde{C}$ and $\tilde{C} = C_{abcd}$, or $dd_Y \in \tilde{C}$ and $\tilde{C}$ contains a black edge $d_Y b$, implying the existence of a black cycle $bc, ca_Y, a_Y d_Y, d_Y b$. If $aa_B, aa_Y \in \tilde{C}$, then either $a_Y c \in \tilde{C}$ and $\tilde{C}$ contains a grey edge $a_B c$, or $a_Y d_Y \in \tilde{C}$ and $\tilde{C}$ contains a grey edge $a_B d_Y$. To summarize, if there does not exist a splitting of node a with the properties stated in the lemma, then there exists in G a grey edge with an endpoint in a_B and the other in c or d_Y. Symmetrically, if there exists not a splitting of node b with the properties stated in the lemma, then there exists in G a grey edge with an endpoint in b_B and the other in d or c_Y. Note however that if these two grey edges were present at the same time, then G would contain a grey cycle. □

3.2.4 Type D configuration

Fig. 5 illustrates a degree 5 configuration of type D in G, while Fig. 11 illustrates the same configuration in G^*. Let H be the graph given in Fig. 11. Let $\bar{H}$ be the subgraph of H induced by the nodes in $V(H) \setminus \{C_{bb_Bxb_Y}, C_{aqb_Ba_Y}, C_{cqb_Yc_B}\}$. One can easily check that all nodes in $V(\bar{H})$ correspond to C4's actually present in the configuration given in Fig. 5 and that G^* contains graph $\bar{H}$ as an induced subgraph.

Lemma 3.13 *Let G be a breakpoint graph containing no Type A configuration and such that $\Delta(G^*) = 5$. Assume G contains the configuration shown in Fig. 5. Correspondingly, G^* contains the graph $\bar{H}$ as an induced subgraph. Then, there exists a maximum stable set X of G^* with $C_{abcd} \notin X$.*

Proof: Excluding cases which are equivalent by symmetry, one of the cases considered in the following must occur.

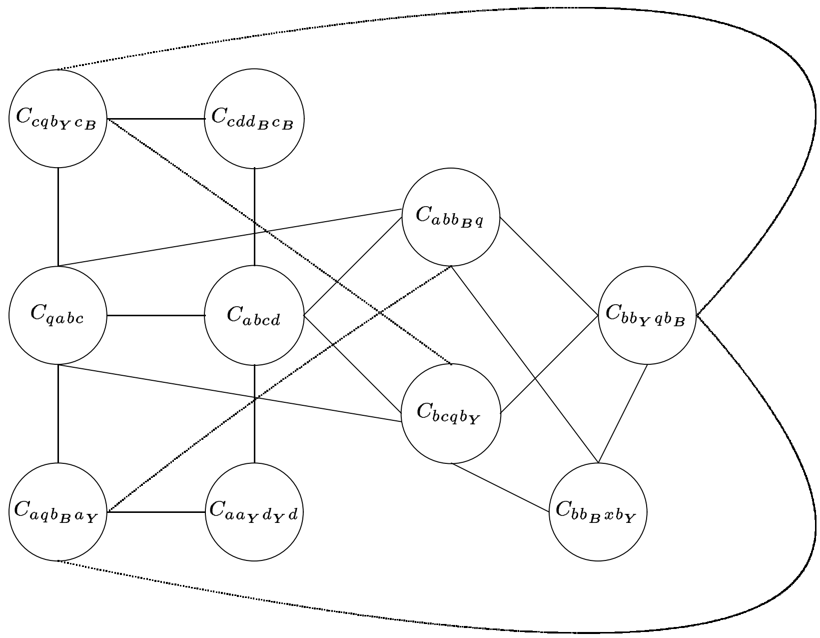

Figure 11: Type D degree 5 configuration in G^*.

Case 1: no node in $V(G^*) \setminus V(\bar{H})$ is adjacent to C_{qabc}. Let X be a maximum stable set of G^* containing C_{abcd}. Then, $X \cup \{C_{qabc}\} \setminus \{C_{abcd}\}$ is a maximum stable set of G^* which does not contain C_{abcd}. $\diamond$

From now on, we will assume that Case 1 does not occur and hence there exists a node $\tilde{C}$ in $V(G^*) \setminus V(\bar{H})$ adjacent to C_{qabc}. Since C_{abcd} is a node of degree 5 of G^*, then $\tilde{C}$ cannot contain any of the edges ab, bc, cd or da. Assume $qa \in \tilde{C}$. Since $ab \notin \tilde{C}$, then $aa_Y \in \tilde{C}$. So, if $qc \in \tilde{C}$, then $\tilde{C}$ contains a black edge with one endpoint in c and the other in a_Y. This is not possible, since there are already two black edges incident with c_B and a_Y and c_B cannot coincide (otherwise we would have a node of degree 6). Therefore, $qb_B \in \tilde{C}$ and then $\tilde{C}$ contains a black edge with one endpoint in b_B and the other in a_Y, namely $\tilde{C} = C_{aqb_Ba_Y}$. Symmetrically, assuming $qc \in \tilde{C}$, then one has that $\tilde{C} = C_{cqb_Yc_B}$.

By the above discussion, if there are two nodes in $V(G^*) \setminus V(\bar{H})$ adjacent to C_{qabc}, then these two nodes are $C_{aqb_Ba_Y}$ and $C_{cqb_Yc_B}$ and the following case occurs.

Case 2: two nodes in $V(G^*) \setminus V(\bar{H})$, namely $C_{aqb_Ba_Y}$ and $C_{cqb_Yc_B}$, are adjacent to C_{qabc}. Here, one can verify that node C_{qabc} is a node of degree 5 in a type A configuration as in Fig. 2 after renaming the nodes as follows: $c \to a$ (node c in Fig. 2 corresponds to node a in Fig. 5), $a \to c, x \to d, b \to q, d \to b, a_B \to c_B, b_B \to b_Y, b_Y \to b_B, c_Y \to a_Y$. This contradicts the first assumption in the lemma. $\diamond$

From now on, we will assume that *Case 2* does not occur and hence only one node in $V(G^*) \setminus V(\bar{H})$ is adjacent to C_{qabc}. By symmetry, we can assume that this node is $C_{aqb_Ba_Y}$. This node is also adjacent to C_{abb_Bq} and implies the presence of a black edge b_Ba_Y in G. Suppose that no other node of $V(G^*) \setminus V(\bar{H})$ is adjacent to C_{abb_Bq} or to C_{bcqb_Y}, then we have the following case.

Case 3: one node in $V(G^*) \setminus V(\bar{H})$, namely $C_{aqb_Ba_Y}$, is adjacent to C_{qabc} and C_{abb_Bq} and no other node in $V(G^*) \setminus V(\bar{H})$ is adjacent to C_{qabc} or C_{abb_Bq} or C_{bcqb_Y}. Let X be a maximum stable set of G^* containing C_{abcd}. Note that at most one out of $C_{bb_Y qb_B}$ and $C_{aqb_B a_Y}$ is in X. If $C_{bb_Y qb_B} \notin X$, then $X \cup \{C_{bcqb_Y}\} \setminus \{C_{abcd}\}$ is a maximum stable set of G^* and does not contain C_{abcd}. If $C_{aqb_B a_Y} \notin X$, then $X \cup \{C_{qabc}\} \setminus \{C_{abcd}\}$ is a maximum stable set of G^* and does not contain C_{abcd}. $\diamond$

We now consider, as last possibility, a node $\tilde{C} \neq C_{aqb_B a_Y}$ in $V(G^*) \setminus V(\bar{H})$ and adjacent to $C_{abb_B q}$ or to C_{bcqb_Y}. We first show that such a $\tilde{C}$ is actually adjacent to both $C_{abb_B q}$ and C_{bcqb_Y}. Indeed, assume by symmetry $\tilde{C}$ to be adjacent to C_{bcqb_Y}.

If $bb_Y \in \tilde{C}$, then $bb_B \in \tilde{C}$. Clearly, $b_B q \in \tilde{C}$, since otherwise $\tilde{C} = C_{bb_Y qb_B}$. Therefore, $\tilde{C}$ must also contain a black edge with one endpoint in b_Y and a grey edge with one endpoint in b_B having a common endpoint x. In this case, $\tilde{C} = C_{bb_B xb_Y}$ is also adjacent to $C_{abb_B q}$.

If $qc \in \tilde{C}$, then $cc_B \in \tilde{C}$. Now, if $qa \in \tilde{C}$, then G contains grey edge ac_B, which is not possible as already two grey edges are incident with a and a_Y and c_B cannot coincide, otherwise $qb_Y \in \tilde{C}$ and G contains edge $b_Y c_B$, and we would be in Case 2. Hence we can assume $qc \notin \tilde{C}$.

Finally, if $qb_Y \in \tilde{C}$, then $qb_B \in \tilde{C}$. If also $b_B b \in \tilde{C}$, then $\tilde{C} = C_{bb_Y qb_B}$, else $b_B a_Y \in \tilde{C}$ and G contains a grey edge with one endpoint in a_Y and the other in b_Y. This is a contradiction as G would contain the grey cycle $a_Y a, ab, bb_Y, b_Y a_Y$.

Summarizing, $C_{bb_B xb_Y}$ is the only possibility for $\tilde{C}$ and we are left with the following case.
Case 4: one node in $V(G^*) \setminus V(\bar{H})$, namely $C_{aqb_B a_Y}$, is adjacent to C_{qabc} and $C_{abb_B q}$ and another node in $V(G^*) \setminus V(\bar{H})$ is adjacent to $C_{abb_B q}$ or C_{bcqb_Y}. Here, one can verify that node $C_{abb_B q}$ is a node of degree 5 in a type A configuration as in Fig. 2 after renaming the nodes as follows: $c \to b_B, a \to a, x \to a_Y, b \to b, d \to q, a_B \to d, b_B \to c, b_Y \to b_Y, c_Y \to x$. This contradicts the first assumption in the lemma. □

The following observation shows how to modify G according to Lemma 3.13.

Observation 3.14 *There exists a breakpoint graph $\tilde{G}$, obtained from G by suitably splitting node a or node c, such that $\tilde{G}^*$ is the graph obtained from G^* by removing the node C_{abcd}.*

Proof: Qui non é affatto soddisfacente che ti riferisci ai casi interni ad un'altra dimostrazione. Tra il resto da qua confonde il fatto che esamini (giustamente) solo 2 di 4 casi If Case 1 in the proof occurs, let $\tilde{G}$ be the breakpoint graph obtained from G by splitting node a as to separate ab from ad and aa_Y. Then, $\tilde{G}^*$ is the graph obtained from G^* by removing the node C_{abcd}. Indeed, let $\tilde{C}$ be a C4 which is removed by the splitting. Then, either $\tilde{C}$ contains both ab and ad, or $\tilde{C}$ contains both aq and aa_Y. In both cases $\tilde{C}$ is adjacent to C_{qabc}. However, the only nodes adjacent to C_{qabc} are C_{abcd}, $C_{abb_B q}$ and C_{bcqb_Y}. Note that $C_{abb_B q}$ and C_{bcqb_Y} are not affected by the splitting.

Otherwise, Case 3 must occur. Let $\tilde{G}$ be the breakpoint graph obtained from G by splitting node c as to separate cb from cd. Then, $\tilde{G}^*$ is the graph obtained from G^* by removing the node C_{abcd}. Indeed, let $\tilde{C}$ be a C4 which is removed by the splitting. Then, either $\tilde{C}$ contains both cb and cd, or $\tilde{C}$ contains both cq and cc_B. In both cases $\tilde{C}$ is adjacent to C_{qabc}. However, the only nodes adjacent to C_{qabc} are C_{abcd}, $C_{abb_B q}$, C_{bcqb_Y} and $C_{aqb_B a_Y}$. Note that $C_{abb_B q}$, C_{bcqb_Y} and $C_{aqb_B a_Y}$ are not affected by the splitting. □

Acknowledgments

Part of this work was done while the first author was visiting BRICS. The first author was also partially supported by CNR and MURST, Italy.

References

[1] V. Bafna and P.A. Pevzner, "Genome Rearrangements and Sorting by Reversals", *SIAM Journal on Computing* 25 (1996) 272–289.

[2] P. Berman, T. Fujito, "On Approximation Properties of the Independent Set Problem for Low Degree Graphs", *Theory of Computing Systems* 32 (1999) 115–132.

[3] P. Berman, M. Fürer, "Approximating Maximum Independent Set in Bounded Degree Graphs", *Proceedings of the 5th Annual ACM-SIAM Symposium on Discrete Algorithms* (1994), ACM Press.

[4] P. Berman and M. Karpinski, "On Some Tighter Inapproximability Results", ECCC Report No. 29 (1998), University of Trier, 1998.

[5] A. Caprara, "Sorting Permutations by Reversals and Eulerian Cycle Decompositions", *SIAM Journal on Discrete Mathematics* 12 (1999) 91–110.

[6] A. Caprara, "On the Tightness of the Alternating-Cycle Lower Bound for Sorting by Reversals", *Journal of Combinatorial Optimization* 3 (1999) 149–182.

[7] A. Caprara, G. Lancia and S.K. Ng, "A Column-Generation Based Branch-and-Bound Algorithm for Sorting by Reversals", *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* 47 (1999) 213–226.

[8] A. Caprara, G. Lancia and S.K. Ng, "Fast Practical Solution of Sorting by Reversals", *Proceedings of the 11th Annual ACM-SIAM Symposium on Discrete Algorithms* (2000), ACM Press.

[9] D.A. Christie, "A 3/2 Approximation Algorithm for Sorting by Reversals", *Proceedings of the 9th Annual ACM-SIAM Symposium on Discrete Algorithms* (1998) 244–252, ACM Press.

[10] V. Chvátal, Linear programming. A Series of Books in the Mathematical Sciences. W.H. Freeman and Company, ISBN: 0-7167-1195-8 New York, 1983.

[11] S. Hannenhalli and P.A. Pevzner, "Transforming Cabbage into Turnip (Polynomial Algorithm for Sorting Signed Permutations by Reversals)", *Journal of the ACM* 48 (1999) 1–27.

[12] C.A.J. Hurkens and 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 Journal on Discrete Mathematics* 2 (1989) 68–72.

[13] J. Kececioglu and D. Sankoff, "Exact and Approximation Algorithms for Sorting by Reversals, with Application to Genome Rearrangement", *Algorithmica* 13 (1995) 180–210.