J Math Chem (2009) 45:941–945
DOI 10.1007/s10910-008-9444-x

ORIGINAL PAPER

A bound for the reversal distance of genome rearrangements

Kramer Alpar-Vajk

Received: 27 November 2007 / Accepted: 25 June 2008 / Published online: 7 August 2008

Abstract This paper answers the conjecture posed by Jianxiu Hao regarding an estimate of sorting by reversals. More precisely, for every permutation π the right-hand inequality of $d(\pi) \geq b(\pi) - c(\pi) \geq \frac{m(\pi)}{2} - \omega(G(\pi))$ will be proved.

Keywords Sorting by reversals · Breakpoint graph

The problem of sorting permutations by reversals is defined in the following way : given $n \in \mathbb{N}^* := \mathbb{N} \setminus \{0\}$ and given a permutation $\pi \in S_n$, the aim is to transform π into I_n (the identity permutation of order n) using the minimum number of reversals. A reversal on π is defined as the inversion of an arbitrary substring of π. The reversal distance of π is denoted with $d(\pi)$.

Example Take the permutation $41532 \in S_5$.

$$\underline{41}532 \rightarrow 1\underline{4532} \rightarrow 123\underline{54} \rightarrow 12345.$$

First we will recall some well-known notions. Therefore let be $n \in \mathbb{N}^*$ and let be $\pi \in S_n$,

$$\pi := \pi_1\pi_2 \ldots \pi_n$$

a permutation in one line notation. The extended permutation of π, denote by π_e, is defined by adding 0 at the beginning of the string and $n+1$ at the end of the string, thus

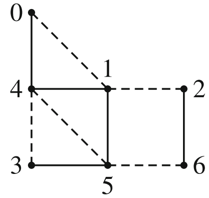
$$\pi_e := 0\ \pi_1\ \pi_2 \ldots \pi_n\ n+1.$$

K. Alpar-Vajk (✉)
Politecnico Milano, Milano, Italy
e-mail: vajk.kramer@polimi.it

Obviously these insertions mean that $\pi_0 := 0$ and $\pi_{n+1} := n + 1$.
Let be now $i \in \{0, 1, 2, \ldots, n\}$. The pair (π_i, π_{i+1}) is called a breakpoint of π if $|\pi_i - \pi_{i+1}| \neq 1$ in π_e.
Further, a two-colored graph $G(\pi)$, the so-called breakpoint graph of π, is defined:

- the vertex set of $G(\pi)$ is the set $\{0, 1, 2, \ldots, n + 1.\}$
- two type of edges are defined, black edges and dashed edges. The set of breakpoints will be the black edges.

For every $i \in \{0, 1, 2, \ldots, n\}$, the pair $(i, i + 1)$ will be a dashed edge if i and $i + 1$ are not on consecutive positions in π_e.
The graph $G(\pi)$ will be a balanced graph and thus will have at least one alternating cycle in every connected component. Thus the max. number of edge-disjoint alternating cycle of $G(\pi)$ is well-defined and is denoted with $c(\pi)$.



The breakpoint graph of 41532.

We can see that in the above example the max. number of edge-disjoint alternating cycles is 2. We have two possible decompositions one is (0, 1, 5, 4, 0) and (4, 1, 2, 6, 5, 3, 4). The other one is (0, 1, 4, 5, 3, 4, 0) and (1, 2, 6, 5, 1).

Definition *Let be $n \in \mathbb{N}^*$, let be $\pi \in S_n$ a permutation and let be $G(\pi)$ its breakpoint graph. We use following notations:*

- *$b(\pi)$:= the number of breakpoints of π.*
- *$m(\pi)$:= the number of vertices of degree 2 in $G(\pi)$.*
- *$k(\pi)$:= the number of vertices of degree 4 in $G(\pi)$.*
- *$\omega(G(\pi))$:= the number of components in $G(\pi)$.*
- *$c(\pi)$:= the max. number of edge-disjoint alternating cycles in $G(\pi)$.*
- *$d(\pi)$:= the reversal distance of π.*

Lemma 1 *Let be $n \in \mathbb{N}^*$ and let be $\pi \in S_n$. Following assertions hold true:*

1. $k(\pi) \geq c(\pi) - \omega(G(\pi))$.
2. $m(\pi) + 2k(\pi) = 2b(\pi)$.

Proof 1. Note that a vertex $v \in G(\pi)$ can have degree 2 or 4. Every vertex of degree 2 in $G(\pi)$ is incident to one black edge and one dashed edge, while every vertex of degree 4 in $G(\pi)$ is incident to two black and two dashed edges. Let K be a connected component of $G(\pi)$. Denote $k(K)$ the number of vertices of degree 4 in K and denote $c(K)$ the max. number of edge disjoint alternating cycles in K. Let K be further a decomposition of K in a max. number of edge-disjoint alternating cycles. A vertex of K belonging to two cycles in this decomposition will be called an intersecting vertex. In light of the above mentioned facts, every vertex of degree 2 in K belongs to exactly one cycle, while every vertex of degree 4 in K belongs to at most two cycles in this decomposition. Therefore, every intersecting vertex is a vertex of degree 4, but not necessarily vice versa. Denote with $i(K)$ the number of intersecting vertices of K respect to the decomposition of K in a max. number of edge-disjoint alternating cycles. Then obviously $k(K) \geq i(K)$. Moreover we claim that

$$i(K) \geq c(K) - 1.$$

If $c(K) = 1$ the claim is immediate.
If $c(K) = 2$ it is clear that there exist at least one intersecting vertex.
The general property follows inductively using the fact that K is a connected domain and the fact that every intersecting vertex belong to exactly two cycles. In fact let be $t \in \mathbb{N}^*$ and suppose that the property is true for all K such that $c(K) \in \{1, 2, \ldots, t\}$. We have to show that the property is true for all K with $c(K) = t + 1$ too. Therefore let K be such that $c(K) = t + 1$, and let there be a decomposition of K in $t + 1$ edge-disjoint alternating cycles. Suppose that $i(K) \leq t - 1$. Choose one arbitrary cycle in this decomposition, say C, and let $K \setminus C$ be the graph obtained by deleting the edges and the non-intersecting vertices of C. Clearly $c(K \setminus C) = t$ and by the induction hypothesis

$$i(K \setminus C) \geq t - 1. \tag{1}$$

On the other hand since K is connected, at least one intersecting vertex v of K will be destroyed when forming $K \setminus C$ thus the occurrence of

$$i(K \setminus C) \leq i(K) - 1 \leq t - 2$$

is a contradiction to (1). Thus $c(K)$ cycles implies the existence of at least $c(K) - 1$ intersecting vertices, that is there then are at least $c(K) - 1$ vertices of degree 4.
We transfer the situation now to $G(\pi)$. Clearly $c(\pi)$ results adding all $c(K)$, for all K connected components of $G(\pi)$ and the same holds for $k(\pi)$. Thus $k(\pi) \geq c(\pi) - \omega(G(\pi))$.
2. It is immediate that the breakpoint graph $G(\pi)$ has $2b(\pi)$ edges. Consider a decomposition of $G(\pi)$ in edge-disjoint alternating cycles. In this way every edge of $G(\pi)$ is distributed in one such cycle. It is well-known that a cycle has as many vertices as edges. Therefore a total of $2b(\pi)$ edges means a total of $2b(\pi)$ vertices such that some of them occur more than once. In $G(\pi)$ every vertex has degree 2 or degree 4. Clearly all vertices of degree 2 occur once, while all vertices with degree 4 occur twice. Therefore $m(\pi) + 2k(\pi) = 2b(\pi)$. □

Corollary *Let be $n \in N$ and let be $\pi \in S_n$. Then*

$$d(\pi) \geq b(\pi) - c(\pi) \geq \frac{m(\pi)}{2} - \omega(G(\pi)).$$

Proof The left-hand side inequality is well-known, for a proof consult [1]. We will prove the right-hand side inequality. From the Lemma we have

$$2b(\pi) = m(\pi) + 2k(\pi) \geq m(\pi) + 2c(\pi) - 2\omega(G(\pi)).$$

This means that

$$2b(\pi) - 2c(\pi) \geq m(\pi) - 2\omega(G(\pi)),$$

thus

$$b(\pi) - c(\pi) \geq \frac{m(\pi)}{2} - \omega(G(\pi)).$$ □

A further aim is a convenient characterisation of those permutations π for which

$$b(\pi) - c(\pi) = \frac{m(\pi)}{2} - \omega(G(\pi))$$

because the computational complexity is easier for the right-hand side expression of the above equality. However, an immediate equivalence follows.

Lemma 2 *Let be $n \in \mathbb{N}^*$ and let be $\pi \in S_n$ a permutation. Then*

$$b(\pi) - c(\pi) = \frac{m(\pi)}{2} - \omega(G(\pi)) \Leftrightarrow c(\pi) = k(\pi) + \omega(G(\pi)).$$

Proof The assertion follows using Lemma 1. □

Lemma 3 *Let be $n \in \mathbb{N}^*$ and let be $\pi \in S_n$ a permutation such that $k(\pi) = 0$. Then*

$$b(\pi) - c(\pi) = \frac{m(\pi)}{2} - \omega(G(\pi))$$

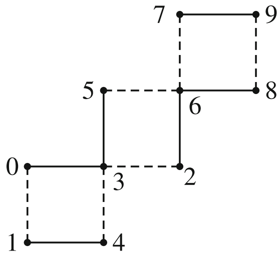
Proof From Lemma 1, we get that $\omega(G(\pi)) \geq c(\pi)$. On the other hand $c(\pi) \geq \omega(G(\pi))$ is always true therefore

$$c(\pi) = \omega(G(\pi)). \tag{2}$$

From Lemma 1, follows also that

$$b(\pi) = \frac{m(\pi)}{2}, \tag{3}$$

so from (2) and (3) we deduce the claim. □



The breakpoint graph of 35412687.

The example above is such that $k(\pi) > 0$ and

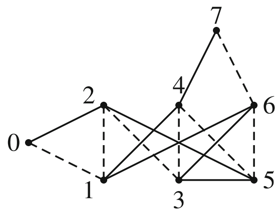
$$b(\pi) - c(\pi) = \frac{m(\pi)}{2} - \omega(G(\pi)).$$

An example of π such that $k(\pi) > 0$ but

$$b(\pi) - c(\pi) > \frac{m(\pi)}{2} - \omega(G(\pi))$$

is the permutation 41532 discussed earlier.

Remark Note that a breakpoint graph in general is not a planar graph. To see this it is enough to consider the breakpoint graph of the 253614 permutation.



The breakpoint graph of 253614.

Due to Kuratowski's theorem this is not a planar graph since it contains the complete bipartite graph $K_{3,3}$ on the vertices (1, 3, 5) and (2, 4, 6).

Acknowledgements I would like to thank to D. J. Klein and to the anonymous referee for their useful observations and suggestions.

Reference

1. V. Bafna, P. Pevzner, Genome rearrangements and sorting by reversals. SIAM J. Comput. **25**(2), 272–289 (1996)