



ELSEVIER

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Can a breakpoint graph be decomposed into none other than 2-cycles?

Lianrong Pu^a, Yu Lin^b, Daming Zhu^a, Haitao Jiang^{a,*}^a School of Computer Science and Technology, Shandong University, Jinan, China^b Research School of Computer Science, Australian National University, Canberra, Australia

ARTICLE INFO

Article history:

Received 26 September 2016

Received in revised form 11 July 2017

Accepted 21 September 2017

Available online xxxx

Keywords:

Algorithm

Complexity

Breakpoint graph

Genome rearrangement

Cycle decomposition

ABSTRACT

Breakpoint graph has been widely used as a key data structure in algorithm design for genome rearrangements. The problem of breakpoint graph cycle decomposition, which asks for a largest collection of edge-disjoint cycles, is crucial in computing rearrangement distances between genomes. This problem is NP-hard, and can be approximated to $1.4193 + \epsilon$. It is still open for deciding whether a breakpoint graph can admit a cycle decomposition with none other than 2-cycles. In this paper, we present a linear time algorithm to detect whether a breakpoint graph can be decomposed into none other than 2-cycles.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Bafna and Pevzner first introduced the so-called *breakpoint graph* as a data structure, to study the problem of sorting by reversals [1][2]. Breakpoint graph has also been used in sorting by translocations [5][6]. The problem of sorting unsigned genomes by reversals was first studied by Kececioğlu and Sankoff [8], and later proved NP-hard through the reduction from the problem of breakpoint graph cycle decomposition by Caprara [9].

The reversal distance between two signed permutations has been directly computed by a formula with some breakpoint graph parameters [10], since the breakpoint graph of two signed permutations admits a unique cycle decomposition. However, the breakpoint graph of two unsigned permutations contains many possible cycle decompositions, thus the algorithms for sorting unsigned genomes by reversals always depend on how the breakpoint graph is decomposed into cycles [11][12]. Moreover, to compute the translocation distance [5][13][14] as well as the double-cut-and-join distance [15][16], it has to take into account how many cycles a breakpoint graph can be decomposed into. Breakpoint graph cycle decomposition has also been learnt useful for ortholog assignment in comparative genomics [17][18] and approximation of exemplar breakpoint distance in the presence of duplicated genes [19].

The problem of decomposing a breakpoint graph with respect to unsigned genomes into a largest collection of cycles is NP-hard [9][21], and can be approximated to $1.4193 + \epsilon$ [7]. In those breakpoint graph cycle decomposition algorithms [7][3][4], a lot of effort has been put into finding as many 2-cycles as possible. However, it is still open for whether the breakpoint graph of two arbitrary unsigned genomes can be decomposed into none other than 2-cycles.

In this paper, we present a linear time algorithm to detect whether a breakpoint graph can be decomposed into none other than 2-cycles. We first simplify the breakpoint graph and construct a *matching graph* for the simplified breakpoint

* Corresponding author.

E-mail address: htjiang@sdu.edu.cn (H. Jiang).<https://doi.org/10.1016/j.tcs.2017.09.019>

0304-3975/© 2017 Elsevier B.V. All rights reserved.

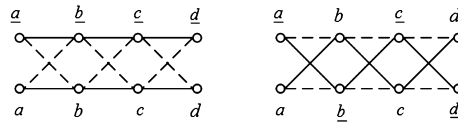


Fig. 1. A double tangle with vertex set $\{a, b, c, d, \underline{a}, \underline{b}, \underline{c}, \underline{d}\}$, where a solid line represents a black edge, a dotted line represents a gray edge. Here depicts two representations of the double tangle.

graph. Then we simplify the matching graph for which we can propose a linear time algorithm to detect whether it admits an *excellent match*. The algorithm can return an edge disjoint 2-cycle set which exactly cover all the edges of the breakpoint graph, if the simplified matching graph admits an excellent match.

Our algorithm can be used to test whether there exists a scenario between two genomes without reusing any breakpoints (e.g., under the infinite sites model of genome evolution [20]) and have potential applications in deriving better approximation algorithms for computing rearrangement distance and improving ortholog assignment in comparative genomics.

2. Preliminaries

Let $\pi = \pi_1, \dots, \pi_n$ be a permutation of $\{1, \dots, n\}$. The *breakpoint graph* with respect to π is constructed by: (1) adding 0 and $n + 1$ to the left side of π_1 and the right side of π_n respectively to form $\pi' = \pi_0, \pi_1, \dots, \pi_n, \pi_{n+1}$ where $\pi_0 = 0, \pi_{n+1} = n + 1$; (2) setting a vertex corresponding to each element in π' , where π_j with $0 \leq j \leq n + 1$ both represents an element in π' and a vertex corresponding to the element π_j in π' ; (3) setting a black edge (π_i, π_{i+1}) if $|\pi_i - \pi_{i+1}| \neq 1, 0 \leq i \leq n$, and a gray edge (π_i, π_j) if $|\pi_i - \pi_j| = 1$ and $|i - j| \neq 1$. We denote by $G(\pi)$ the breakpoint graph with respect to π . Let G be a subgraph of $G(\pi)$. Then respectively, we denote by $V(G), E(G), E_b(G), E_g(G)$ the vertex, edge, black and gray edge sets of G .

A cycle is *alternating*, if the cycle (as a path) is simple, and every two consecutive edges of the cycle have distinct colors. An alternating cycle with i black edges is abbreviated as an i -cycle. A cycle in $G(\pi)$ refers to a subgraph of $G(\pi)$, which is an alternating cycle. Since no black edge can have the same two ends as a gray edge has, no 1-cycle can occur in $G(\pi)$. That is,

Lemma 1. No 1-cycle occurs in $G(\pi)$.

In $G(\pi)$, a vertex is either *isolated* (i.e., zero degree), or incident with one black and one gray edge, or two black and two gray edges. Thus, a *cycle decomposition* of $G(\pi)$ refers to a set of alternating cycles in $G(\pi)$ which covers every edge of $G(\pi)$ exactly once. If we split each four degree vertex of $G(\pi)$ into two new ones, distribute one black and one gray edge of those four edges sharing the vertex for their ends to take one of the new ones for their ends, and distribute the other two edges to take the other of the two new ones for their ends, then we get a graph with each connected component as an alternating cycle. Given a cycle decomposition of $G(\pi)$, there must exist a way to split those four degree vertices of $G(\pi)$ to transform $G(\pi)$ into one in which every cycle in the cycle decomposition turns into a connected component.

3. Breakpoint graph to matching graph

A 2-cycle decomposition of $G(\pi)$ refers to a cycle decomposition of $G(\pi)$ with none other than 2-cycles. Let c, c_1 and c_2 all be 2-cycles. Then c_1 and c_2 cover c , if $E(c) \subseteq E(c_1) \cup E(c_2)$.

Lemma 2. Let $G_2(\pi)$ be a 2-cycle decomposition of $G(\pi)$. If a 2-cycle covered by two cycles in $G_2(\pi)$ does not occur in $G_2(\pi)$, then $G(\pi)$ admits a 2-cycle decomposition in which this 2-cycle occurs.

Proof. Let c_1 and c_2 be two 2-cycles in $G_2(\pi)$, and c a 2-cycle covered by c_1 and c_2 . Since c does not occur in $G_2(\pi)$, c must share a black and a gray edge with c_1 , as well as c_2 . It follows that those edges in $E(c_1) \cup E(c_2) - E(c)$ form a 2-cycle of $G(\pi)$, say c_3 . Thus, replacing c_1 and c_2 in $G_2(\pi)$ with c and c_3 transforms $G_2(\pi)$ into a 2-cycle decomposition of $G(\pi)$ in which c occurs. □

3.1. Double tangle removing

An induced subgraph of $G(\pi)$ is referred to as a *double tangle*, if it has eight vertices, where provided that its vertex set is $\{a, b, c, d, \underline{a}, \underline{b}, \underline{c}, \underline{d}\}$, its black and gray edge sets must be $\{(a, b), (b, c), (c, d), (\underline{a}, \underline{b}), (\underline{b}, \underline{c}), (\underline{c}, \underline{d})\}$ and $\{(a, \underline{b}), (\underline{a}, b), (b, \underline{c}), (\underline{b}, c), (c, \underline{d}), (\underline{c}, d)\}$. A double tangle example is depicted in Fig. 1 for intuitional observation.

Those edges in a double tangle can be excluded from consideration for deciding if $G(\pi)$ admits a 2-cycle decomposition. This is because,

Download English Version:

<https://daneshyari.com/en/article/6875457>

Download Persian Version:

<https://daneshyari.com/article/6875457>

[Daneshyari.com](https://daneshyari.com)