Linear Algebra and its Applications 490 (2016) 162–173

Contents lists available at ScienceDirect

Linear Algebra and its Applications

www.elsevier.com/locate/laa

Successful pressing sequences for a bicolored graph and binary matrices

Joshua Cooper, Jeffrey Davis *

Department of Mathematics, University of South Carolina, Columbia, SC 29208, United States

ARTICLE INFO

Article history:
Received 17 June 2015
Accepted 2 November 2015
Available online 18 November 2015
Submitted by R. Brualdi

MSC:
05C50
15B33
92D15

Keywords:
Binary matrix
Pressing sequence
Adjacency matrix
Matching
Bicolored graph

ABSTRACT

We apply matrix theory over $\mathbb{F}_2$ to understand the nature of so-called "successful pressing sequences" of black-and-white vertex-colored graphs. These sequences arise in computational phylogenetics, where, by a celebrated result of Hannenhalli and Pevzner, the space of sortings-by-reversal of a signed permutation can be described by pressing sequences. In particular, we offer several alternative linear-algebraic and graph-theoretic characterizations of successful pressing sequences, describe the relation between such sequences, and provide bounds on the number of them. We also offer several open problems that arose as a result of the present work.

1. Introduction

In a now classical paper in bioinformatics [5], Hannenhalli and Pevzner showed that there is a polynomial time algorithm to sort signed permutations by reversals, i.e., turn any signed permutation into the identity by reversing subwords (and flipping their signs).

* Corresponding author.
E-mail addresses: cooper@math.sc.edu (J. Cooper), davisj56@email.sc.edu (J. Davis).

http://dx.doi.org/10.1016/j.laa.2015.11.001

This has important implications for computational phylogenetics: when comparing the sequence of genes of two related species, the shortest length of a sequence of reversals that transforms one into the other is one prominent measure of the evolutionary distance of the associated organisms. The authors' strategy, and one that was improved upon in later work (for example, [8]), is to construct the so-called "breakpoint graph" for the permutation to be sorted, show that a certain operation on the breakpoint graph corresponds to reversals, and then use certain numerical invariants of subgraphs to guide the sequence of moves to the identity.

This framework is now a keystone of bioinformatics algorithms, but it leaves many questions unanswered. In particular, the proposed methodologies generate just one successful sorting of the signed permutation under consideration, and it is understood that there are often many such minimum-length sorting sequences. Since each is only representative of one *possible* evolutionary history, it would be valuable to be able to sample from all possible such sequences to obtain more sensitive statistical properties. As of yet, there is no full understanding of the space of possible histories, so Markov Chain Monte Carlo methods are valuable for approximately uniform sampling. Such approaches present their own problems, however: it is necessary to obtain a proof of connectivity of the underlying graph of the Markov Chain to know that it will eventually reach every vertex; and it is necessary to obtain bounds on the mixing time of the process to ensure that near-uniformity will be achieved in reasonable time. Indeed, some researchers have investigated these very kinds of questions: see, for example, [9].

In order to state our results and situate it in the above discussion, we need the following definitions.

Definition 1. A *bicolored* graph is a pair (G, c) where G is a simple graph, and $c : V(G) \to \{\text{black}, \text{white}\}$ is a coloring of its vertices. Write $\overline{\text{black}} = \text{white}$ and $\overline{\text{white}} = \text{black}$.

Denote by $V(G)$ the vertex set of a graph, $E(G)$ its edge set, and $G[S]$ the induced subgraph of a set $S \subset V(G)$; let $N(v) = N_G(v)$ denote the neighborhood of $v \in V(G)$, i.e., $\{w \in V(G) : \{v, w\} \in E(G)\}$, and $N^*(v) = N^*_G(v)$ the closed neighborhood of v, i.e., $N^*_G(v) = N_G(v) \cup \{v\}$.

Definition 2. Consider a bicolored graph, (G, c) with a black vertex $v \in V(G)$. "Pressing v" is the operation of transforming (G, c) into (G', c'), a new bicolored graph in which $G[N^*(v)]$ is complemented. That is, $V(G') = V(G)$,

$$E(G') = E(G) \triangle \binom{N^*(v)}{2}$$

(where "$\triangle$" denotes symmetric difference) and $c'(w) = c(w)$ for $w \notin N^*(v)$ and $c'(w) = \overline{c(w)}$ for $w \in N^*(v)$. Thus in pressing v, v becomes an isolated white vertex. See Fig. 1.