Discrete Mathematics 341 (2018) 2883–2893

Contents lists available at ScienceDirect

Discrete Mathematics

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

Aparallel digraphs and splicing machines

Andrew Vince

Department of Mathematics, University of Florida, USA

ARTICLE INFO

Article history:
Received 15 May 2017
Received in revised form 4 July 2018
Accepted 5 July 2018
Available online 21 July 2018

Keywords:
Digraph
Splicing

ABSTRACT

The concepts of a splicing machine and of an aparalled digraph are introduced. A splicing machine is basically a means to uniquely obtain all circular sequences on a finite alphabet by splicing together circular sequences from a small finite set of "generators". The existence and uniqueness of the central object related to an aparallel digraph, the *strong component*, is proved, and this strong component is shown to be the unique fixed point of a natural operator defined on sets of vertices of the digraph. A digraph is shown to be a splicing machine if and only if it is the strong component of an aparallel digraph. Motivation comes, on the applied side, from the splicing of circular sequences on a finite alphabet and, on the theoretical side, from the Banach fixed point theorem.

1. Introduction

Let G be a finite, directed graph on vertex set V. An edge in G, directed from vertex x to vertex y is denoted (x, y). With respect to vertex x, the edge (x, y) is an *out-edge*. A *walk* is always a directed walk. The *length* of a walk p, if finite, is the number of edges in p. A *circuit* is a closed walk. A *cycle* is a circuit with no repeated vertices (except the first and the last); i.e., a cycle does not cross itself.

The main objects in this paper are *aparallel digraphs* and *splicing machines*, whose definitions are given below. The concepts of aparallel digraph and splicing machine are closely connected; the exact relationship is discussed in Section 3. Motivation comes, on the applied side, from the splicing of circular sequences from a finite alphabet and, on the theoretical side, from the Banach fixed point theorem. Although we do not claim a direct application, circular RNAs (circRNAs) are abundant and are expressed in thousands of human genes. See [4] and references therein for an overview of the subject.

Modeling recombinant DNA behavior using formal language theory dates back at least to 1987 [8], and many subsequent papers have been written on the subject of such *splicing systems*, for example [5,6,10]. Although splicing of sequences is common to both, our splicing machine is not substantially related to these splicing systems. In particular, formal languages are not involved. From the other direction, fixed point theorems have been investigated via directed graphs; see for example [1] and references therein. These results also are largely independent of those in this paper.

Fig. 1 shows a 2-colored (black and red) digraph with the property that each vertex has exactly one outgoing edge colored black and exactly one outgoing edge colored red. (There are loops at vertices 1 and 8.) If the successive colors along a walk p are $(c_1, c_2, c_3, \dots, c_n)$, then we say that p has *type* $(c_1, c_2, c_3, \dots, c_n)$. Consider a sequence of colors, say $C = (1, 0, 0, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 0)$, where 0 stands for black and 1 for red. In the figure, the circuit with successive vertices 2, 5, 3, 2, 1, 1, 5, 7, 4, 6, 3, 6, 7, 8, 4, 2 is of type C. In fact, this particular digraph has the following property: (1) for any finite binary sequence C of colors, no matter how long, there is a circuit in the digraph of type C; (2) for any such sequence C of colors, the circuit in the digraph of type C is unique; and (3) there are no "extra" edges in the digraph in the sense that every edge appears in some circuit. Since every circuit in a digraph can be obtained by "splicing" cycles together, we will refer to such a digraph as a *splicing machine*, defined formally in Definition 3. Basically, in a splicing machine, any circular sequence of colors can be uniquely obtained by splicing together a subset of the finitely many cycle sequences.

E-mail address: avince@ufl.edu.

https://doi.org/10.1016/j.disc.2018.07.004

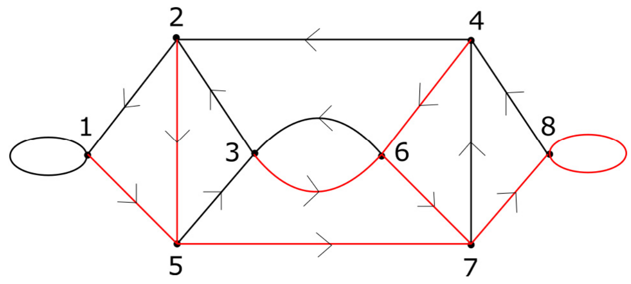

Fig. 1. A spicing machine. (The color red appears in the online version of this paper.)

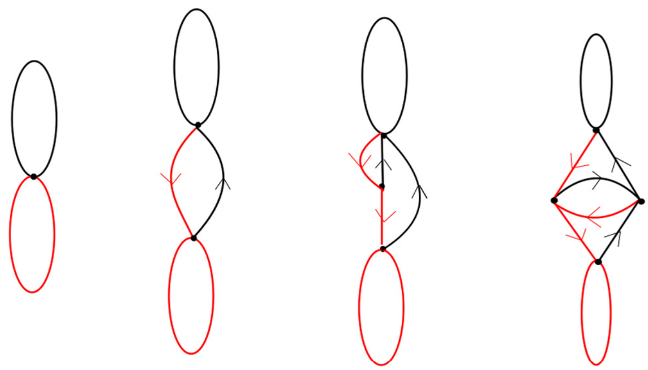

Fig. 2. Small aparallel digraphs with two colors. (The color red appears in the online version of this paper.)

1.1. Aparallel digraphs

Definition 1. Let $[N] = \{1, 2, \ldots, N\}$ for $N \geq 1$, and call $[N]$ the *set of colors*. A **colored-digraph** $G = (V, E, c)$ is a finite directed graph with vertex set V, edge set E, and edge coloring $c : E \to [N]$ such that every vertex has exactly N out-edges, one out-edge of each color. Multiple edges and loops are allowed.

For a colored-digraph G whose edges are colored in $[N]$ and a walk $p = x_0, x_1, \ldots$, finite or infinite, the *type* of p, denoted C_p, is defined as

$$C_p = (c(x_0, x_1),\ c(x_1, x_2),\ c(x_2, x_3)\ldots).$$

Given a sequence $C = (j_1, j_2, \ldots)$, finite or infinite, of colors, and a vertex $x_0 \in V$, there is a unique walk, denoted $p_C(x_0)$, of type C. The same vertex may, of course, appear many times in $p_C(x_0)$.

If an infinite walk p has successive vertices $x_0, x_1, x_2, \ldots$ and an infinite walk p' has successive vertices $x'_0, x'_1, x'_2, \ldots$, then we say that p and p' are *parallel* if $x_i \neq x'_i$ for all $i \geq 0$. Let $[N]^*$ denote the set of all finite sequences of colors and $[N]^\infty$ the set of all infinite sequences of colors. Given a sequence $C \in [N]^\infty$, parallel walks $p_C(x_0) = x_0, x_1, \ldots$ and $p_C(y_0) = y_0, y_1, \ldots$, with the same color sequence $C \in [N]^\infty$, will be called C-*parallel*.

Definition 2. A colored-digraph G is called **aparallel** if G has no pair of C-parallel walks for all $C \in [N]^\infty$. Such a colored-digraph will be referred to as an **aparallel digraph**.

Note that, if G is aparallel, then it must be connected as an undirected graph. Four small aparallel digraphs with $N = 2$ are shown in Fig. 2. Several infinite families of aparallel digraphs are provided in the examples below. The terminology "Cantor set" and "Sierpinski triangle" in Examples 2 and 3 will be explained in Example 6 of Section 2. The examples below are also revisited in Example 7 and Example 8.

Example 1 (*Discrete Interval*)**.** Consider the following infinite family $G(2k)$ for $k = 1, 2, \ldots,$ of 2-colored-digraphs. Let $V = \{0, 1, 2, \ldots, 2k-1\}$. The edges colored 1 are $\left(n, \lfloor \frac{n}{2} \rfloor\right)$ and the edges colored 2 are $\left(n, \lfloor \frac{n}{2} \rfloor + k\right)$ for $n = 0, 1, 2, \ldots 2k-1$. The colored-digraph $G(2k)$ is not, in general, aparallel. For example, it will follows from Lemma 1 in Section 2 that $G(6)$ is not an aparallel digraph because both $p_{12}(1)$ and $p_{12}(2)$ are cycles in $G(6)$. However, if k is a power of 2, then $G(2k)$ is aparallel. This will be proved in Example 8 of Section 5. The aparallel digraph $G(4)$ is the rightmost one in Fig. 2; digraph $G(8)$ is the one in Fig. 1.