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

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
0012-365X/

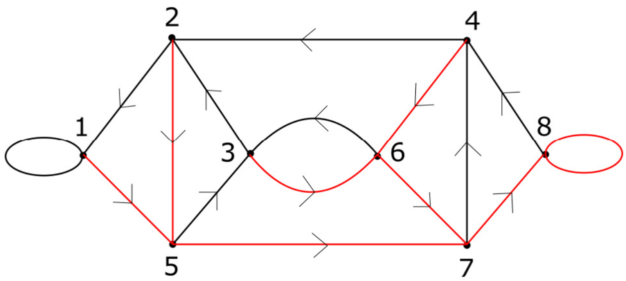

Fig. 1. A spicing machine. (The color red appears in the online version of this paper.)

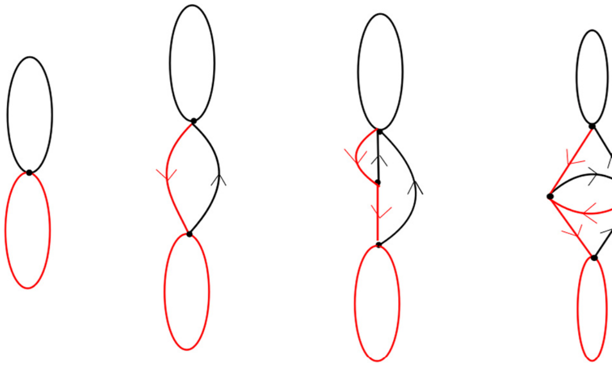

Fig. 2. Small aparallel digraphs with two colors. (The color red appears in the online version of this paper.)

1.1. Aparallel digraphs

Definition 1. Let $[N] = \{1, 2, \dots, N\}$ 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 $p = x_0, x_1, \dots$, finite or infinite, the *type* of p, denoted C_p, is defined as

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

Given a sequence $C = (j_1, j_2, \dots)$, 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 $x_0, x_1, x_2, \dots$ and an infinite walk p' has successive vertices $x'_0, x'_1, x'_2, \dots$, 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 $p_C(x_0) = x_0, x_1, \dots$ and $p_C(y_0) = y_0, y_1, \dots$, 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 $k = 1, 2, \dots,$ of 2-colored-digraphs. Let $V = \{0, 1, 2, \dots, 2k-1\}$. 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 $n = 0, 1, 2, \dots 2k-1$. 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.

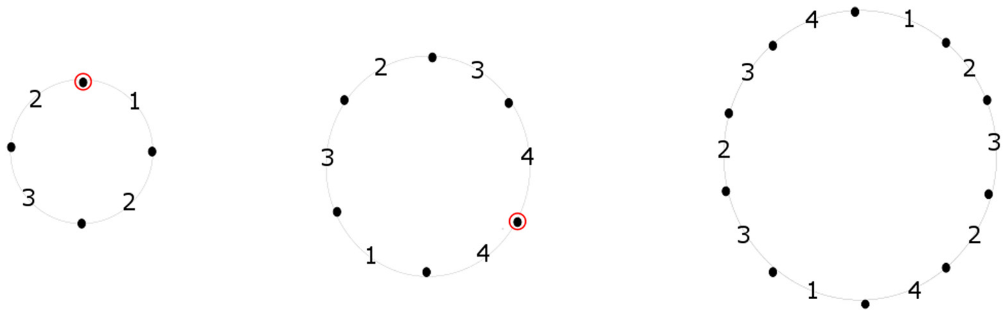

Fig. 3. The two circular sequences on the left are spliced at the points indicated by red circles. (The color red appears in the online version of this paper.)

Example 2 (*Discrete Cantor Set*)**.** Consider the following infinite family $H(3k)$ for $k = 1, 2, \ldots,$ of 2-colored-digraphs. Let $V = \{0, 1, 2, \ldots, 3k-1\}$. The edges colored 1 are $\big(n, \lfloor \frac{n}{3} \rfloor\big)$ and the edges colored 2 are $\big(n, \lfloor \frac{n}{3} \rfloor + 2k\big)$ for $n = 0, 1, 2, \ldots 2k-1$. That $H(3k)$ is aparallel if k is a power of 3 will be proved in Example 8 of Section 5.

Example 3 (*Discrete Sierpinski Triangle*)**.** Let k be a positive integer and $V = \{(x, y) \in \mathbb{N} \times \mathbb{N} : x + y < k\}$. The edges colored 1 of a family $S(k)$ of 3-colored-digraphs are $\big((m, n), (\lfloor \frac{m}{2} \rfloor, \lfloor \frac{m}{2} \rfloor)\big)$; the edges colored 2 are $\big((m, n), (\lfloor \frac{m}{2} \rfloor + k, \lfloor \frac{m}{2} \rfloor)\big)$; and the edges colored 3 are $\big((m, n), (\lfloor \frac{m}{2} \rfloor, \lfloor \frac{m}{2} \rfloor + k)\big)$. If k is a power of two, then $S(k)$ is an aparallel digraph.

Example 4 (*Colored-Digraphs that are not Aparallel*)**.** Cayley graphs [7] are colored-digraphs, the number N of colors being the cardinality of the generating set of the underlying group of the Cayley graph. Cayley graphs, however, are not in general aparallel digraphs because the necessary condition for being aparallel provided in Proposition 1 of Section 2 usually fails to hold.

1.2. Splicing machines

A *circular sequence in* $[N]$ is a sequence of the form

$$(j_1, j_2, \ldots, j_n) = (j_2, j_3, \ldots, j_n, j_1) = (j_3, j_4, \ldots, j_n, j_1, j_2) = \cdots = (j_n, j_1, \ldots, j_{n-2}, j_{n-1}),$$

where $j_1, \ldots, j_n \in [N]$. Let $[N]^o$ denote the set of all circular sequences. Two circular sequences $C = (j_1, j_2, \ldots, j_m)$ and $C' = (j'_1, j'_2, \ldots, j'_n)$ are said to be *spliced at position* (s, t) to obtain the circular sequence

$$C \bullet C' = C \bullet_{\{s,t\}} C' = (j_1, j_2 \ldots, j_s, j'_{t+1}, j'_{t+2}, \ldots, j'_t, j_{s+1}, j_{s+2}, \ldots, j_n).$$

Using clockwise orientation, Fig. 3 shows a splicing of two circular sequences. It is easy to verify that splicing operations are commutative and associative: $C \bullet C' = C' \bullet C$ and $(C \bullet C') \bullet C'' = C \bullet (C' \bullet C'')$. Therefore, the circular sequence obtained by multiple splicing does not depend on the order of splicing.

The basic idea behind a splicing machine is for the splicing of circular sequences to take place within a digraph. If $\gamma_x = z, x_1, \ldots, x_s, z$ and $\gamma_y = z, y_1, \ldots, y_t, z$ are two circuits of a digraph (in terms of their successive vertices) with common vertex z, then the circuit $\gamma = z, x_1, \ldots, x_s, z, y_1, y_2, \ldots, y_t, z$ is said to be obtained by *splicing* γ_x and γ_y at vertex z. Let G be a digraph whose edges have colors in the set $[N]$. If γ is a circuit of G, then the mapping $\gamma \mapsto C_\gamma$ assigns to each circuit of G a circular sequence. When no confusion arises, we may use the terminology "splicing two circuits in G" and "splicing the corresponding circular sequences in $[N]^o$" interchangeably. Of course, in this context the splicing positions of two circular sequences is restricted to be a vertex of G.

Let G be a digraph with edges colored in the set $[N]$. To insure that, for every vertex v and any sequence C of colors, there is a walk of type C starting at v, it will be assume that, for every vertex v and every color $c \in [N]$, there is an out-edge colored c. It is not assumed apriori, however, that there is at most one out-edge colored c. The number of out-edges from v colored c will be referred to as the *multiplicity of* c *at* v. Given such a digraph G with edges colored in the set $[N]$, we may ask: what circular sequences can be obtained by splicing together cycles of G? Since any circuit in a digraph can be obtained by splicing cycles, this is equivalent to asking about the set

$$\Gamma_G = \{C_p : p \text{ is a circuit in } G\}$$

of circuit types of G. For such a graph G to be a splicing machine, defined formally below, it is required that every circular sequence can be uniquely obtained by splicing a set of cycles of G.

Definition 3. A **splicing machine for** $[N]$ is a digraph G, with possible loops and multiple edges, such that, for every vertex, every color has multiplicity at least 1, and such that

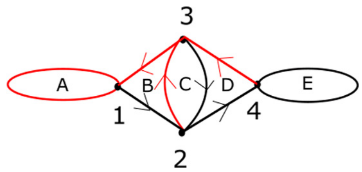

Fig. 4. Splicing machine; see Example 5. (The color red appears in the online version of this paper.)

(1) $\Gamma_G = [N]^o$,
(2) for any $C \in [N]^o$, the circuit in the digraph of type C is unique, and
(3) every edge of G appears in some circuit (hence some cycle) of G.

The set

$$\mathcal{C}_G = \{C_p : p \text{ is a cycle in } G\}$$

of cycle types is called the **set of generators**. Condition (3) in the definition is a minimality requirement: if G satisfies conditions (1) and (2) but has an edge e that does not appear in a circuit, then e can be removed without violating conditions (1) and (2).

Example 5. As will be proved in Section 3, the colored-digraph G in Fig. 4 is a splicing machine. The set of generators is $\mathcal{C}_G = \{0, 1, 10, 110, 100, 0011\}$, where 1 stands for red and 0 for black. The cycles in G corresponding to the circular sequences in $\mathcal{C}_G$ are denoted E, A, C, B, D, F. The cycles A, B, C, D, E are labeled; F is the cycle 1,2,4,3,1. The average length of the generators is $2\frac{1}{3}$. The circular sequence $(0, 0, 0, 1, 0, 1)$ of length 6, for example, is obtained by the splicing $D \bullet_4 E \bullet_3 C$. The corresponding circuit after splicing is 3, 2, 4, 4, 3, 2, 3. Although not required by Definition 3, the digraph G in this example is an aparallel digraph.

1.3. Organization and results

The central object associated with an aparallel digraph is the *strong component*. The proof of its existence and uniqueness (Theorem 1), and an investigation of its properties is the subject of Section 2. In particular, the strong component is the unique fixed point of a natural operator defined on the set of subsets of V (Theorem 2).

The objective of a splicing machine is to obtain, in an efficient way, from a finite set $\mathcal{C}$ of generators, all circular sequences. The main result of Section 3 is that G is a splicing machine if and only if it is the strong component of an aparallel digraph (Theorem 3). This section also addresses questions about the efficiency of splicing machines.

There is a natural way to address the vertices of a aparallel digraph. This is explained in Section 4, in particular Corollary 2.

Although the paper is completely graph theoretic, one motivation is the Banach fixed point theorem [2] from analysis, namely, a contraction on a complete metric space has a unique fixed point. This connection is explained in Section 5, in particular by Theorems 4 and 5.

Open areas of research related to the notions in this paper appear in Section 6.

2. The strong component

Recall that $[N]^*$ denotes the set of all finite sequences of colors. If $C \in [N]^*$, then $C^n := C\,C\,C \cdots C$ denotes the n times concatenation of C, and $\overline{C} = C\,C\,C \cdots$ denotes the infinite concatenation. For $C \in [N]^*$, the terminal vertex of the walk $p_C(x_0)$ is denoted $t_C(x_0)$.

Lemma 1. *If G is an aparallel digraph and $p_C(x)$ and $p_C(y)$ are both circuits in G, then $x = y$.*

Proof. Assume by way of contradiction that $p_C(x)$ and $p_C(y)$ are both circuits in G. If $x \neq y$, then $p_{\overline{C}}(x)$ and $p_{\overline{C}}(y)$ are C-parallel walks, contradicting the definition of aparallel. □

Proposition 1. *Let G be an aparallel digraph. For every $C \in [N]^*$ there is a unique circuit in G of type C. In particular, for every $j \in [N]$ there is a unique loop in G colored j.*

Proof. Uniqueness follows from Lemma 1. Concerning existence, let $x_n = t_{C^n}(x_0)$ for $n \geq 1$, where $x_0 \in V$. Since V is finite, it must be the case that $x_i = x_j$ for some $j > i$. If $j > i + 1$, then $x_{i+1} \neq x_i$ and $p_{\overline{C}}(x_i)$ and $p_{\overline{C}}(x_{i+1})$ are $\overline{C}$-parallel, contradicting the aparallel property. Therefore $x_{i+1} = x_i$, which implies that $p_C(x_i)$ is a circuit in G of type C. □

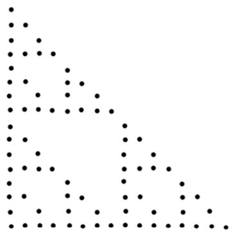

Fig. 5. The vertices of the strong component of $S(4)$; edges have been omitted.

It follows from Lemma 1 that, in an aparallel digraph, the uniqueness of a circuit γ of type C as a sequence in $[N]^*$ is equivalent to the uniqueness of γ of type C as a circular sequence in $[N]^o$.

The set of maximal strongly connected subgraphs of a digraph, each called a *strong component*, partitions the vertex set [3]. Call a strong component *trivial* if it consists of just one vertex with no loop. (A strong component that is a single vertex with loop(s) is non-trivial.) Given a subset $X \subseteq V$ of a colored-digraph $G = (V, E, c)$, by abuse of language, we often do not distinguish between X and the subgraph of G induced by X.

Theorem 1. *Given an aparallel digraph $G = (V, E, c)$, there exists a unique non-trivial strong component A of G. Moreover*

(1) *G has no edge (a, b) for which $a \in A$ and $b \in V \setminus A$;*
(2) *there is no circuit in G containing a vertex in $V \setminus A$; and*
(3) *for every $C \in [N]^\infty$ and $x \in V$, the walk $p_C(x)$ eventually enters and remains in A.*

Proof. By Proposition 1 there is a strong component that contains a loop colored $1 \in [N]$ at vertex, say a. Let A be the strong component containing a. This strong component is non-trivial since it contains a loop.

Assume, by way of contradiction, that A fails to satisfy property (1). Consider an edge (a, b) for which $a \in A$ and $b \in V \setminus A$. There can be no walk from b to a vertex of A; otherwise the maximality of A is contradicted. Therefore the walks $p_{\overline{1}}(a)$ (repeated loop) and $p_{\overline{1}}(b)$ are parallel, contradicting the assumption that G is aparallel.

We next prove that A satisfies property (2). If $C \in [N]^*$ is such that $p_C(x)$ is a circuit such that $x \in V \setminus A$, then $p_{\overline{C}}(x)$ and $p_{\overline{C}}(a)$ are parallel for any $a \in A$, contradicting that G is aparallel.

To prove the uniqueness of A, assume, by way of contradiction, that there is a non-trivial strong component $B \neq A$. Since B is strongly connected, there is a circuit in B, contradicting property (2) which was proved in the paragraph above.

Concerning property (3), it follows from property (1) that, once a walk enters A, it remains in A. Assume, therefore, that there is a $C \in [N]^\infty$ and an $x \in V \setminus A$ such that $p_C(x)$ is contained in $V \setminus A$. Then, for any $a \in A$, the walks $p_C(x)$ and $p_C(a)$ are parallel, contradicting that G is aparallel. $\square$

Properties (1–3) in Theorem 1 motivate the following terminology.

Definition 4. Given an aparallel digraph $G = (V, E, c)$, the unique non-trivial strong component A of G will be referred to as simply the **strong component** of G.

For each of the colored-digraphs in Fig. 2 and the colored-digraph of Example 1, the strong component is the digraph itself.

Example 6 (*Discrete Cantor Set and Discrete Sierpinski Triangle Revisited*)**.** In Example 2, if the strong component of $H(3^k)$ is denoted by A_k, then

$$A_k = \left\{ a \in \{0, 1, 2, \dots, 3^k - 1\} \,:\, \text{the base 3 representation of } a \text{ does not contain the digit } 1 \right\}.$$

The strong component is a discrete version of the Cantor set. If these points of A_k are scaled by $1/3^k$ and plotted on the real line, then, as $k \to \infty$, the sets A_k approach (in the Hausdorff metric) the classical Cantor set. This geometric description is by way of motivation; it is not intrinsic to the definition of A_k.

A similar situation holds for Example 3. If the points of the strong component A_k of $S(2^k)$ are scaled by $1/2^k$ and plotted in $\mathbb{R}^2$, then, as $k \to \infty$, the strong components A_k approach (in the Hausdorff metric) a Sierpinski triangle (see Fig. 5).

Corollary 1. *Let $G = (V, E, c)$ be an aparallel digraph with strong component A, and let $C \in [N]^o$. The unique circuit in G of type C, as insured by Proposition 1, is contained in A.*

Proof. Statement (2) of Theorem 1 implies that there is no circuit that contains a point of $V \setminus A$. $\square$

Let $G = (V, E, c)$ be a colored-digraph. Define a map $T : H(V) \to H(V)$ from the set $H(V)$ of all non-empty subsets of V to itself as follows. For $x \in V$ and $X \subseteq V$, define

$$T(x) = \{y \in V \ : \ (x, y) \in E\} \qquad \text{and} \qquad T(X) = \bigcup_{x \in X} T(x). \tag{1}$$

The set $T(x)$ is just the set of out-neighbors of x. Let $T^n(X) = T \circ T \circ \cdots \circ T(X)$, where it is an n-fold composition, and $T^0(X) = X$. The set $T^n(X)$ is the set of vertices of V reachable from a vertex of X by some walk of length n. Call a subset $X \in H(V)$ a *fixed point of* T if $T(X) = X$.

Theorem 2. *An aparallel digraph G with strong component A has the following properties:*

(1) *A is the unique fixed point of T, and*
(2) *there is an integer n_0 such that $T^n(X) = A$ for every $X \subseteq V$ and every $n \geq n_0$.*

Proof. Statement (1): By statement (1) of Theorem 1, there is no edge (a, b) with $a \in A$ and $b \in V \setminus b$. Therefore $T(A) \subseteq A$. Given any two vertices $a_1, a \in A$, strong connectivity implies there is a path from a_1 to a in A. If a_2 is the vertex on this path just before a, then $a \in T(a_2)$. Therefore $A \subseteq T(A)$ and hence $T(A) = A$.

Statement (2): If $n_1 = |V \setminus A|$ and $x \in V$, then it follows from Theorem 1 that any walk of length n_1 must terminate in A. Therefore

$$T^n(X) \subseteq A \quad \text{for every} \quad X \subseteq V \quad \text{and every} \quad n \geq n_1. \tag{2}$$

Let a be the vertex of the loop colored 1, whose existence is insured by Proposition 1. By Corollary 1, vertex a lies in A. From Eq. (2), the fact that $a \in T(a)$, and by the strong connectivity of A, there is an $n_2 \geq n_1$ such that

$$a \in T^n(X) \quad \text{for every} \quad X \subseteq V \quad \text{and every} \quad n \geq n_2. \tag{3}$$

From $a \in T(a)$, it also follows that $\{a\} \subseteq T(a) \subseteq T^2(a) \subseteq \cdots$. From this and from the strong connectivity of A it follows that there is an $n_0 \geq n_2$ such that

$$T^n(a) = A \quad \text{for all} \quad n \geq n_0. \tag{4}$$

From Eqs. (2), (3), and (4) it follows that $T^n(X) = A$ for every $X \subseteq V$ and every $n \geq n_0$.

To prove the uniqueness of the fixed point of T, assume that $T(A_1) = A_1$ and $T(A_2) = A_2$. By what was proved above, we have $A_1 = T^{n_0}(A_1) = A = T^{n_0}(A_2) = A_2$. □

3. Splicing machines

The subject of this section is the relationship between aparallel digraphs and splicing machines.

Lemma 2. *If G is a spicing machine on $[N]$, then G is a colored-digraph, i.e., each vertex has exactly one out-edge of each color.*

Proof. Assume, by way of contradiction, that there is a vertex v that has two out-edges colored, say 1. By condition (3) in Definition 3, there is a circuit, say γ, whose first edge (v, v') is colored 1 and where γ is of type, say C. Let x_n, $n \geq 0$, be vertices on a walk α such that (i) $x_0 = v$, (ii) there is a walk from x_i to x_{i+1} is of type C, and (iii) the first edge of α is not (v, v'). Since the vertex set of G is finite, it must be the case that $x_j = x_k$ for some $k > j$. Let γ' be the subcircuit of α from x_j to $x_k = v_j$. Unless $v_j = v$, this however, contradicts condition (2) in Definition 3 because γ' and γ spliced $k - j$ times are both of type C^{k-j}. If $v_j = v$, then let γ' be the subcircuit of α from v to $v_j = v$. Again γ' and γ contradict condition (2) in Definition 3. □

Theorem 3. *A directed graph G is a splicing machine if and only if G is the strong component of an aparallel digraph.*

Proof. Assume that G is the strong component of an aparallel digraph. Let C be any circular sequence. According to Proposition 1, there is a unique circuit in G of type C. Therefore, conditions (1) and (2) in Definition 3 hold. By the strong connectivity of G, every edge in G lies on a circuit in G. Therefore, condition (3) in Definition 3 holds.

Conversely, assume that G is a splicing machine. By Lemma 2, G is a colored-digraph. It is first shown, by contradiction, that G must be connected as an undirected graph. By condition (1) in Definition 3 there is a loop l in G colored 1 in undirected component, say G_1. Let γ be a walk in a different undirected component G_2 of type $111\cdots$. Since G_2 is finite, there must be a circuit in G_2 of type 1^k for some $k \geq 1$. This, however, contradicts condition (2) in Definition 3 because there would be two circuits in G of type 1^k, the first obtained by splicing the l loop k times.

It is next shown that G is strongly connected. By way of contradiction, assume two strong components G_1 and G_2 and p a path from G_1 to G_2. Since there can be no path from G_2 to G_1, no edge on p is contained in a circuit, contradiction condition (3) in Definition 3.

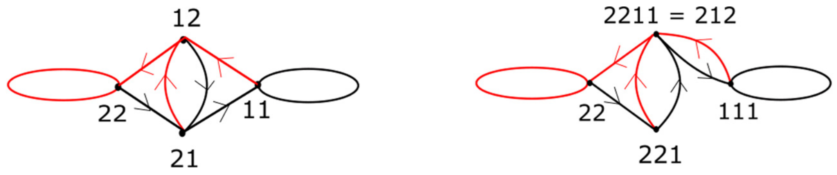

Fig. 6. Addresses of the vertices of G where *black* $= 1$, *red* $= 2$. (The color red appears in the online version of this paper.)

Finally it is shown that G is an aparallel digraph. Consider the graph G^2 whose vertex set is $\{(x, y) : x, y \in V, x \neq y\}$ and whose edge set is

$$\{((x, y), (t_i(x), t_i(y))) : x, y \in V,\ i \in [N]\}.$$

Assume, by way of contradiction, that p and p' are C-parallel walks in G, starting at vertices v_0 and v_0', respectively. If $C = (c_1, c_2, c_3, \ldots)$, let $v_i := t_{(c_1,c_2,\ldots,c_i)}(v_0)$ and $v_i' := t_{(c_1,c_2,\ldots,c_i)}(v_0')$ for $i \geq 1$, and let $u_0 := (v_0, v_0')$, $u_1 := (v_1, v_1')$, $u_2 := (v_2, v_2'), \ldots$. Now $u_0, u_1, u_2, \ldots$ is a walk in G^2. Since G^2 is finite, there are integers $k > j$ such that $u_j = u_k$. This implies that $v_j, v_{j+1}, \ldots, v_k = v_j$ and $v_j', v_{j+1}', \ldots, v_k' = v_j'$ are circuits of the same type, contradicting condition (2) in Definition 3. □

The digraphs in Figs. 1 and 4 are both strong components of aparallel digraphs; hence both are splicing machines. Using 0 for black and 1 for red, the set of generators of the splicing machine in Fig. 4 is $\mathcal{C}_1 = \{\{0, 1, 10, 110, 100, 0011\}\}$. The set of generators of the splicing machine in Fig. 1 is $\mathcal{C}_2 = \{\{0, 1, 10, abcd\}\}$, where a can take the value 0 or 10, b can take the value 0 or 100, c can take the value 1 or 01, and d can take the value 1 or 011; hence $\mathcal{C}_2$ has 19 elements. The average length of the generators in C_2 is about 6.1 in contrast to about 2.3 for the generators in C_1.

Factors relevant to the "efficiency" of a splicing machine G include the following:

(1) The number n of vertices in G should be small. Note that, if the splicing machine has n vertices with N, the number of colors, fixed, then the number of edges is nN, linear in n.
(2) The number of cycles in G should be small.
(3) The number of splices required to obtain a given circular sequence C should be small.
(4) The average length of the cycles in G should be large. In a splicing machine with a fixed number of vertices, the average number of splices required to obtain a circular sequence is likely inversely proportional to the average length of the cycles in the splicing machine.

Of course, these are conflicting goals. The splicing machine that consists of a single vertex and N loops at this vertex obviously has small order, namely $n = 1$. On the other hand, to obtain a circular sequence of length k with this splicing machine requires k splices, the largest possible. Question 3 in Section 6 pertains to the above parameters.

4. Address map

A scheme is developed in this section for addressing the vertices of the strong component of an aparallel digraph. An address is an element of $[N]^*$. Basically, a sequence $C \in [N]^*$ is an address of vertex a in the strong component if all walks of type C lead to a, independent of the initial vertex.

Definition 5. Given an aparallel digraph G with strong component A and a vertex $a \in A$, a sequence $C \in [N]^*$ is called an **address** of a if $t_C(x) = a$ for all $x \in V$.

Let C be an address of $a \in A$ and $C' \in [N]^*$ be an arbitrary finite color sequence. Clearly, the concatenation $C\,C'$ is an address of $t_{C'}(a)$; and clearly $C'\,C$ is another address of a. This motivates the following definition.

Definition 6. Given an aparallel digraph G with strong component A and a vertex $a \in A$, a sequence $C = (j_1, j_2, \ldots, j_k) \in [N]^*$ is called a **minimal address** of a if $t_C(x) = a$ for all $x \in V$, but $t_{j_2 j_3 \cdots j_k}(x) \neq a$ for some $x \in V$.

Fig. 6 shows minimal addresses of the vertices of the strong components of two aparallel digraphs. Edges labeled 1 are colored black; edges labeled 2 are colored red. Note that, for the strong component on the right, a minimal address does not have to be unique. We next show that every vertex in the strong component of an aparallel digraph has an address, hence a minimal address.

Let $[N]^k$ denote the set of sequences of length k of elements of $[N]$.

Proposition 2. *Let $G = (V, E, c)$ be an aparallel digraph with strong component A. There is an integer K such that for any $m \geq K$ and for any $C \in [N]^m$ we have $t_C(x) = t_C(y)$ for every $x, y \in V$.*

Proof. Let $k = |A|^2$; let $C = (j_1, j_2, \dots, j_k)$; and let $x, y \in A$. Consider the sequence of ordered pairs in $A \times A$ defined by

$$(x_0, y_0) = (x, y), \qquad (x_1, y_1) = (t_{j_1}(x), t_{j_1}(y)), \qquad (x_2, y_2) = (t_{j_1,j_2}(x), t_{j_1,j_2}(y)), \dots,$$
$$(x_k, y_k) = (t_C(x), t_C(y)).$$

Since A is finite, there must be two terms, say (x_s, y_s) and (x_t, y_t), $s < t$, in this sequence that are identical. If $x_s \neq y_s$, then letting $C' = (j_{s+1}, \dots, j_t)$, the walks $p_{\overline{C'}}(x_s)$ and $p_{\overline{C'}}(y_s)$ are $\overline{C}'$-parallel, contradicting that G is aparallel. Therefore $x_s = y_s$, which implies that $t_C(x) = t_C(y)$.

In the paragraph above, $x, y \in A$. Now consider any $x, y \in V$. By statement (2) of Theorem 2, there is an n such that $t_C(x) \in A$ for every $C \in [N]^n$ and every $x \in V$. Therefore, if $C \in [N]^m$ for any $m \geq K = n + k$, then $t_C(x) = t_C(y)$ for all $x, y \in V$. $\square$

Definition 7. Given an aparallel digraph G with strong component A and constant K as in Proposition 2, let $[N]^{\geq K} = \bigcup_{m \geq K} [N]^m$. Define a map $\pi : [N]^{\geq K} \to A$, called the **address map**, by

$$\pi(C) = t_C(x),$$

which is independent of $x \in V$ by Proposition 2.

Corollary 2. *Every vertex in the strong component of aparallel digraph has a minimal address.*

Proof. Since $T^m(A) = A$ for all m by Theorem 2, the map π is surjective. Every string in $\pi^{-1}(a)$ is an address of the vertex $a \in A$. Since every $a \in A$ has an address, every such a has at least one minimal address. $\square$

Define an *inverse shift map* $s_i : [N]^* \to [N]^*$ by $s_i(C) = C\,i$ for $i \in [N]$. Given K as in Proposition 2 and Definition 7, it is routine to check that the following diagram commutes.

$$\begin{array}{ccc} [N]^{\geq K} & \xrightarrow{s_i} & [N]^{\geq K} \\ \pi \downarrow & & \downarrow \pi \\ A & \xrightarrow[t_i]{} & A \end{array} \tag{5}$$

Interpreting the diagram: appending $i \in [N]$ to the end of an address of a point x is an address of a point y that is adjacent to x and such that $c(x, y) = i$.

Example 7 (*Discrete Interval and Discrete Cantor Set Revisited*)**.** For the colored-digraphs $G(2^k)$, $k \geq 0$, and $H(3^k)$, $k \geq 0$, of Examples 1 and 2, denote the respective strong components by $A_{G,k}$ and $A_{H,k}$, respectively. Addresses for the vertices of $A_{G,k}$ and $A_{H,k}$ are used to show that $A_{G,k}$ and $A_{H,k}$ are isomorphic as colored-digraphs. Here *isomorphism*, denoted by $\approx$, is a bijection of the vertex sets that preserves directed edges and colors.

The vertices of $A_{G,k}$ are $V_{G,k} := \{0, 1, 2, \dots, 2^{k-1}\}$, and the vertices of $A_{H,k}$ are $V_{H,k} := \{a \in \{0, 1, 2, \dots, 3^k - 1\} :$ the base 3 representation of a does not contain the digit $1\}$. Let $a \in A_{G,k}$ and let $a = \alpha_{k-1}\alpha_{k-2} \cdots \alpha_0$, the right hand side being the binary representation of a. If $\beta_i = \alpha_i + 1$ for $i \geq 0$, it is not hard to show that

$$\pi(\beta_0\, \beta_1 \cdots \beta_{k-1}) = a,$$

in other words $\beta_0\, \beta_1 \cdots \beta_{k-1}$ is an address of a.

Likewise, let $a \in A_{H,k}$ and let $a = \alpha_{k-1}\alpha_{k-2} \cdots \alpha_0$, the right hand side being the ternary representation of a. Let $\beta_i = 1$ if $\alpha_i = 2$ and otherwise $\beta_i = \alpha_i = 0$, for all $i \geq 0$. It is not hard to show that

$$\pi(\beta_0\, \beta_1 \cdots \beta_{k-1}) = a,$$

in other words $\beta_0\, \beta_1 \cdots \beta_{k-1}$ is an address of a.

Note that $|V_{G,k}| = |V_{H,k}| = 2^k$, and the set of all address of vertices in $V_{G,k}$ and in $V_{H,k}$ are both $[2]^k$. If, in terms of vertex addresses, the map $\phi : V_{G,k} \to V_{H,k}$ is defined by $\phi(C) = C$ for all $C \in [2]^k$, then the commutative diagram (5) implies that ϕ is an isomorphism and

$$A_{G,k} \approx A_{H,k}$$

for all $k \geq 1$. Although the geometric motivation is quite different for the two families $G(2^k)$ and $H(2^k)$ of colored-digraphs, their strong components are isomorphic.

5. Banach fixed point theorem

The connection between aparallel digraphs and the Banach fixed point theorem is the subject of this section. Let $d : V \times V \to \mathbb{R}$ be a metric on the space V. A function $f : V \to V$ is a *contraction* on the metric space (V, d) if there is a real number $0 \le r < 1$ such that $d(f(x), f(y)) \le r\, d(x, y)$ for all $x, y \in V$. According to the Banach fixed point theorem, a contraction on a complete metric space has a unique fixed point.

The situation becomes more interesting when there is more than one contraction on (V, d). Let $\mathcal{F} = \{f_1, f_2, \dots, f_N\}$ for $N \ge 1$, be a set of contractions on a metric space V. In the fractal literature, $\mathcal{F}$ is called an *iterated function system*. Taking a graph theoretic point of view, define $G(V, \mathcal{F})$ to be a digraph with vertex set V and, for every two points $x, y \in V$, there is an edge from x to y colored $i \in \{1, 2, \dots, N\}$ if and only if $v = f_i(u)$. The digraph $G(V, \mathcal{F})$ is a colored-digraph in the sense of Definition 1, except that V may be infinite. In particular, the definition of aparallel digraph carries over to this infinite setting.

Consider the special case of a discrete metric space. A metric space (V, d) is *discrete* if, for each point $x \in V$, there is a ball centered at x containing only x, i.e., if and only if the metric induces the discrete topology on V. The compact subsets of a discrete metric space are the finite subsets. A metric space (V, d) is *uniformly discrete* if there exists an $\epsilon > 0$ such that for each point $x \in V$ there is a ball of radius ϵ centered at x containing only x. For a finite space V, discrete and uniformly discrete are equivalent.

Remark 1. A standard example of a discrete metric on a set V is

$$d_0(x, y) = \begin{cases} 1 & \text{if } x \neq y \\ 0 & \text{if } x = y. \end{cases}$$

This metric, however, is not relevant in our context because any contraction on (V, d_0) must be a constant function.

Theorem 4. *If (V, d) is a uniformly discrete metric space for which each function in $\mathcal{F} = \{f_1, f_2, \dots, f_N\}$ is a contraction on V with respect to d, then $G(V, \mathcal{F})$ is an aparallel digraph.*

Proof. Assume that $G(V, \mathcal{F})$ is not an aparallel digraph. Then there exists $C = (j_1, j_2, \dots) \in [N]^\infty$ and C-parallel walks $P_C(x_0) = x_0, x_1, x_2, \dots$ and $P_C(y_0) = y_0, y_1, y_2, \dots$ and an $0 \le r < 1$ such that

$$\begin{aligned} d(x_0, y_0) &\ge \frac{1}{r}\, d(f_{j_1}(x_0), f_{j_1}(y_0)) = \frac{1}{r}\, d(x_1, y_1) \\ &\ge \frac{1}{r^2}\, d(f_{j_2}(x_1), f_{j_2}(y_1)) = \frac{1}{r^2}\, d(x_2, y_2) \\ &\ge \cdots \ge \frac{1}{r^n}\, d(x_n, y_n) \ge \cdots, \end{aligned}$$

which implies that the metric space is not uniformly discrete. $\square$

Note that a discrete metric space is complete. Let $\mathcal{F} = \{f_1, f_2, \dots, f_N\}$ be a set of contractions on a discrete metric space (V, d). The unique fixed point of the contraction $f_i \in \mathcal{F}$ is the unique vertex of $G(V, F)$ on which there is a loop colored i (as guaranteed by Proposition 1). The operator

$$T(X) = \bigcup_{f \in \mathcal{F}} f(X)$$

defined on finite non-empty subsets X of V, often called the *Hutchinson operator* [9], is exactly the operator $T : H(X) \to H(V)$ in Eq. (1) when V is finite. By Theorem 2, the unique fixed point of T is the strong component of $G(V, \mathcal{F})$.

For V finite, the converse of Theorem 4 holds. Recall that two colored-digraphs are *isomorphic*, denoted by $\approx$, if there is a bijection of the respective vertex sets that preserves directed edges and colors.

Theorem 5. *If $G = (V, E, c)$ is a (finite) aparallel digraph, then there exists a discrete metric $d : V \times V \to \mathbb{R}$ and a set $\mathcal{F}$ of contractions on V such that $G(V, \mathcal{F}) \approx G(V, E, c)$.*

Proof. Given a (finite) aparallel digraph $G = (V, E, c)$, define functions $f_j : V \to V$ for $j = 1, 2, \dots, N$ by setting $f_j(u) = v$ for each $u \in V$, where (u, v) is the unique edge colored c. Let $\mathcal{F} = \{f_1, f_2, \dots, f_N\}$. By construction $G(V, \mathcal{F}) \approx G(V, E, c)$.

To define the metric for which each $f \in \mathcal{F}$ is a contraction, consider the graph G_2 whose vertex set is $\binom{V}{2} = \{\{x, y\} : x \neq y \in V\}$ and whose edge set is

$$\{(\{x, y\}, \{f(x), f(y)\}) : x, y \in V, f \in \mathcal{F}\}.$$

Lemma 1 implies that G_2 is acyclic. Therefore the ordering on $\binom{V}{2}$, defined by $\{a, b\} \preceq \{c, d\}$ if and only if there exists a (directed) path in G_2 from $\{c, d\}$ to $\{a, b\}$, is a partial order. Consider a linear extension

$$\{x_1, y_1\} \prec^* \{x_2, y_2\} \prec^* \{x_3, y_3\} \prec^* \cdots \prec^* \{x_n, y_n\}$$

of this partial order, where $n = \left|\binom{V}{2}\right|$. It is clearly possible to define $d : V \times V \to \mathbb{R}$ so that, for all $x, y \in V$,

$$\begin{aligned}
&d(x, x) = 0\\
&d(x, y) = d(y, x)\\
&d(x_1, y_1) < d(x_2, y_2) < d(x_3, y_3) < \cdots < d(x_n, y_n)\\
&d(x_n, y_n) < 2\, d(x_1, y_1).
\end{aligned}$$

This function d is clearly a discrete metric on V. To show that each $f \in \mathcal{F}$ is a contraction, let $x, y \in V$. By the definition of the partial order, we have $\{f(x), f(y)\} \prec \{x, y\}$. (Note that it is not possible that $\{f(x), f(y)\} = \{x, y\}$ by Lemma 1.) By the definition of linear extension, we have $\{f(x), f(y)\} \prec^* \{x, y\}$, which implies that $d(f(x), f(y)) < d(x, y)$. Since this is true for every pair x, y and since V is finite, there is an $0 \le r < 1$ such that $d(f(x), f(y)) < r\, d(x, y)$ for all $x, y \in V$. □

Example 8 (*The Discrete Interval and the Discrete Cantor Set Revisited*)**.** We show that each colored-digraph in the family $G(2^k),\ k \geq 1$, of Example 1 (the discrete interval) and in the family $H(3^k),\ k \geq 1$, of Example 2 (the discrete Cantor set) is an aparallel digraph.

For $G(2^k)$, define a function $d : V \times V \to \mathbb{R}$ on its vertex set $V = \{0, 1, 2, \ldots, 2^k - 1\}$ as follows. Let $d(n, n) = 0$ for all $n \in V$; let $d(0, 1) = 1$; and for every $n \in V,\ n > 1$, let $d(n - 1, n) = s$, where s is the largest integer such that $n \equiv 0 \bmod 2^s$. In general, for $n > m$, set $d(m, n) = d(n, m) = \sum_{i=m+1}^{n} d(i - 1, i)$. It is not hard to check that d is a metric on V for which the functions f_1 and f_2 defined in Theorem 5 are contractions. By that theorem, $G(2^k)$ is an aparallel digraph for $k \geq 1$.

Likewise, for $H(3^k)$, define a function $d : V \times V \to \mathbb{R}$ on its vertex set $V = \{0, 1, 2, \ldots, 3^k - 1\}$ as follows. Let $d(n, n) = 0$ for all $n \in V$; let $d(0, 1) = 1$; and for every $n \in V,\ n > 1$, let $d(n - 1, n) = s$, where s is the largest integer such that $n \equiv 0 \bmod 3^s$. In general, for $n > m$, set $d(m, n) = d(n, m) = \sum_{i=m+1}^{n} d(i - 1, i)$. Again, it is not hard to check that d is a metric on V for which the functions f_1 and f_2 defined in Theorem 5 are contractions. By that theorem, $H(3^k)$ is an aparallel digraph for $k \geq 1$.

6. Open problems

Several questions on aperiodic digraphs and splicing machines naturally arise.

Question 1. Several families of splicing machines are provided in previous sections. Find additional constructions of splicing machines.

Question 2. Find an algorithm to determine whether or not a given set $\mathcal{C} \subset [N]^o$ is the set of generators of a splicing machine.

Question 3. For a splicing machine G, let $h(G, k)$ denote the average number of splicings required to obtain a circular sequence of length k. Some sequences, like $111 \cdots 1$ will take k splicings, some may take no splicing. What bounds can be obtained for

$$h(n, N, k) := \min h(G, k),$$

where the minimum is taken over all splicing machines on n vertices and N colors? What can be said about the extremal cases, the aparallel digraphs that attain the minimum?

The average number of splices may be inversely proportional to the average length of the cycles in a splicing machine. So one can ask about bounds on

$$C(n, N) := \max C(G),$$

where $C(G)$ is the average length of the cycles in splicing machine G, and the maximum is taken over all slicing machines on n vertices and N colors? For N fixed, does $\lim_{n\to\infty} C(n, N)/n$ exist?

Question 4. Let $G = (V, E, c)$ be an aparallel digraph. By Theorem 5, there exists a metric $d : V \times V \to \mathbb{R}$ on V and a set $\mathcal{F}$ of contractions on V such that $G(V, \mathcal{F}) \approx G(V, E, c)$. Define the *contractivity* of G by

$$r(G) = \min_d \max \left\{ \frac{d(f(x), f(y))}{d(x, y)} \;:\; x, y \in V,\ f \in \mathcal{F} \right\},$$

where the minimum is taken over all aparallel metrics d on V. What bounds can be obtained for $r(G)$?

Acknowledgment

This work was partially supported by a grant from the Simons Foundation (322515 to Andrew Vince).

References

[1] M. Abbas, M.R. Alfuraidan, A.R. Khan, T. Nazir, Fixed point results for set-contractions on metric spaces with a directed graph, Fixed Point Theory Appl. 2015 (2015) 14. http://dx.doi.org/10.1186/s13663-015-0263-z.
[2] S. Banach, Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales, Fund. Math. 3 (1922) 133–181.
[3] J. Bang-Jensen, G. Gutin, Digraphs - Theory, Algorithms and Applications, Springer-Verlag, New York, 2007.
[4] S.P. Barrett, J. Salzman, Circular RNAs: analysis, expression and potential functions, Development 11 (2016) 1838–1847.
[5] K. Culik, T. Harju, Splicing semigroups of dominoes and DNA, Discrete Appl. Math. 31 (1991) 261–277.
[6] R. Freund, L. Kari, G. Păum, DNA computing based on splicing: the existence ofd universal computers, Theory Comput. Syst. 32 (1999) 69–112.
[7] C. Godsil, G. Royle, Algebraic Graph Theory, Springer - Graduate Texts in Mathematics, New York, 2000.
[8] T. Head, Fractals and self-similarity, Formal language theory and DNA: an analysis of the generative capacity of specific recombinant behavoirs, Bull. Math. Biol. 49 (1987) 737–759.
[9] J. Hutchinson, Fractals and self-similarity, Indiana Univ. Math. J. 30 (1981) 713–747.
[10] D. Pixton, Regularity of splicing languages, Discrete Appl. Math. 69 (1996) 101–124.