DIRECTED 3-CYCLE ANCHORED DIGRAPHS AND THEIR APPLICATION IN THE UNIFORM SAMPLING OF REALIZATIONS FROM A FIXED DEGREE SEQUENCE

M. Drew LaMar

Department of Biology
2137 Integrated Science Center
The College of William and Mary
Williamsburg, VA 23187, USA

ABSTRACT

In this paper, we give structural and degree sequence characterizations for a new class of digraphs called directed 3-cycle anchored. A digraph in this class has the property that, for every realization of its degree sequence, there is a directed 3-cycle through each vertex of a labeled vertex set. We end by illustrating their use in the uniform sampling of simple directed graph realizations from a fixed degree sequence.

1 INTRODUCTION

Given a digraph property P, a degree sequence d is said to be forcibly P-digraphic (Rao 1981) if and only if every digraph with degree sequence d has property P. In the special case where the property P is the existence of an induced subgraph $\vec{H}$, we will instead say d is forcibly $\vec{H}$-digraphic. In this extended abstract we consider the subgraph $\vec{H} = \vec{C}_3$, where $\vec{C}_3$ is a directed 3-cycle with vertex set $\{v_1, v_2, v_3\}$ and arc set $\{(v_1, v_2), (v_2, v_3), (v_3, v_1)\}$, and show how to identify what we call $\vec{C}_3$-anchored degree sequences. $\vec{C}_3$-anchored degree sequences are forcibly $\vec{C}_3$-digraphic degree sequences that include the local constraint that vertices induced by a specific set of indices of the degree sequence are a part of a directed 3-cycle for all realizations.

We then illustrate the use of directed 3-cycle anchored digraphs in the approximate uniform sampling of simple directed graph realizations from a fixed degree sequence, with application to problems in epidemiology and ecology (Liljeros et al. 2001, Cobb and Chen 2003), among others. In the example in epidemiology, researchers obtain sexual contact networks through anonymous surveys which lead to degree sequences with no connectivity information (Liljeros et al. 2001). Many of the more generally applicable sampling algorithms fall into two categories: Markov chain Monte Carlo random walks (Cobb and Chen 2003, Rao, Jana, and Bandyopadhyay 1996) and sequential importance sampling (Blitzstein and Diaconis 2011, del Genio et al. 2010). We focus on the Markov chain Monte Carlo algorithms, whereby a sequence of moves from a move-set are used to go from one realization to another. This results in a random walk on a *meta-graph*, where each vertex corresponds to a realization and the arcs connecting these vertices correspond to moves from the move-set. If the meta-graph is strongly connected with appropriate probability weights for the arcs (see (Cobb and Chen 2003)), we will obtain an approximate uniformly sampled realization from the fixed degree sequence.

To sample *simple* directed realizations (i.e., no self-loops or multi-arcs), there are two types of moves in our move-set (Rao, Jana, and Bandyopadhyay 1996), each of which preserves degree sequences: a 2-switch and the reorientation of a directed 3-cycle $\vec{C}_3$. A 2-switch modifies the arc set A by removing the arcs $\{(v_1, w_1), (v_2, w_2)\} \subset A$ and replacing them with $\{(v_1, w_2), \{v_2, w_1\} \subset A^C$. This process is illustrated in the following diagram:
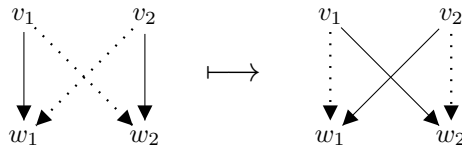
where dotted lines denote no arcs. A directed 3-cycle reorientation modifies the arc set A by removing the arcs $\{(v_1, v_2), (v_2, v_3), (v_3, v_1)\} \subset A$ and replacing them with $\{(v_2, v_1), (v_3, v_2), (v_1, v_3)\} \subset A^C$. This process is illustrated as follows:
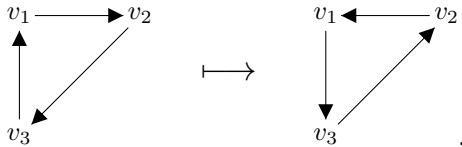
It was shown in (Rao, Jana, and Bandyopadhyay 1996), and rediscovered by Berger and Mller-Hannemann (2010), that the meta-graph is strongly connected under 2-switches and 3-cycle reorientations, with the diameter of the meta-graph on the order of $O(m)$, where m is the number of arcs given by the degree sequence. It is mentioned without rigorous explanation in (Roberts 2000) and (Rao, Jana, and Bandyopadhyay 1996) that for most degree sequences, only 2-switches are necessary. In this paper we identify the cases where $\vec{C}_3$ reorientations are necessary, which turn out to be a subset of the $\vec{C}_3$-anchored degree sequences which we call $\vec{C}_3^*$-anchored, and show that we can reduce our move-set to only 2-switches. Using a smaller move set, however, generally leads to longer mixing times for the Markov chain, as discussed in general in (Cobb and Chen 2003) and more specifically for the case of our move set in (Berger and Mller-Hannemann 2010).

Recently, Berger and Müller-Hannemann posted a paper with similar results. As stated above, in (Berger and Mller-Hannemann 2010) they rediscover the result by Rao, Jana, and Bandyopadhyay (1996) proving strong connectivity of the meta-graph using 2-switches and directed 3-cycle reorientations, as well as the fact that the diameter of the meta-graph can be bounded by $O(m)$. They also implement a similar Monte Carlo algorithm which achieves approximately uniform samples of simple realizations from a directed degree sequence. They show, as we do, that the special cases where directed 3-cycle reorientations are required are precisely the $\vec{C}_3^*$-anchored degree sequences. Our paper differs in that we characterize a larger class of degree sequences, i.e. directed 3-cycle anchored, which contain $\vec{C}_3^*$-anchored degree sequences as a special case. Also, our proof relies on the structural characterization of $\vec{C}_3$-anchored digraphs and uses the degree sequence characterization of the $\vec{C}_3$-anchored digraphs to identify the $\vec{C}_3$-anchors, as opposed to a more computationally intensive algorithm which finds all induced 3-cycles for a given realization using matching techniques. Using the degree sequence characterization is much faster, more specifically quadratic in the number of vertices, as opposed to quadratic in the number of arcs as in (Berger and Mller-Hannemann 2010). We both, however, arrive at the result that the meta-graph consists of 2^k isomorphic subgraphs, where k is the number of anchored 3-cycles.

Many proofs in this paper were omitted due to space constraints and can be found in (LaMar 2010a).

1.1 Notation

All graphs in this paper will be simple and directed, i.e., with no self loops or multiple arcs. We denote directed graphs by $\vec{G}$, with $V(\vec{G})$ the vertex set and $A(\vec{G})$ the arc set. We will drop the reference to $\vec{G}$ when the digraph is understood through the notation $\vec{G} = (V, A)$, for example. An arc between vertices a and b will be denoted by (a, b), with the orientation given by the ordering. If $X, Y \subseteq V$, then $(X, Y) \subseteq A$ is defined as the set of arcs $(x, y) \in A$ such that $x \in X$ and $y \in Y$. For disjoint vertex sets $X, Y \subseteq V$, we define the subgraph $\vec{G}[X, Y] = (X \cup Y, A[X, Y])$, where $A[X, Y] = \{(x, y) \in A : x \in X \text{ and } y \in Y\}$. When $X = Y$, we have the usual definition of an induced subgraph and will denote this by $\vec{G}[X]$. We will denote an isomorphism between two digraphs $\vec{G}$ and $\vec{G}'$ by $\vec{G} \simeq \vec{G}'$.

In figures and diagrams, we will frequently use arrows to denote relations between vertices in the following manner:

$$
\begin{aligned}
x \rightarrow y &\iff (x,y) \in A \\
x \leftarrow y &\iff (y,x) \in A \\
x \leftrightarrow y &\iff \{(x,y),(y,x)\} \subseteq A \\
x \cdots y &\iff \{(x,y),(y,x)\} \subseteq A^C.
\end{aligned}
$$

We will also use a dashed-dotted directional line between x and y to denote an allowable arc, i.e., $(x,y) \in A$ or $(x,y) \notin A$. When vertices are substituted by sets of vertices, we will use the same convention above, e.g. $X \cdots Y$ if and only if $\{(x,y),(y,x)\} \subseteq A^C$ for all $x \in X$ and $y \in Y$.

All integer and integer-pair sequences will be non-negative, unless explicitly stated otherwise. We will often refer to an integer-pair sequence $d = \{(d_i^+, d_i^-)\}_{i=1}^N$ in matrix form as

$$
d = \begin{pmatrix} d_1^+ & \cdots & d_N^+ \\ d_1^- & \cdots & d_N^- \end{pmatrix}
$$

with the first and second rows corresponding to d^+ and d^-, respectively. Given a degree sequence d, we will denote the set of all digraphs with degree sequence d by $R(d)$.

2 DIRECTED 3-CYCLE ANCHORED DIGRAPHS

Given a degree sequence d and digraph $\vec{H}$, we say d is **forcibly $\vec{H}$-digraphic** if and only if for all $\vec{G} \in R(d)$, there is a subgraph $\vec{H}' \subseteq \vec{G}$ such that $\vec{H}' \simeq \vec{H}$. We define a *local* generalization of this definition by specifying a particular set of labeled vertices that must be in the vertex set of one of these subgraphs for all realizations.

Definition 1 We call a degree sequence d $\vec{H}$**-anchored** if it is forcibly $\vec{H}$-digraphic and there exists a nonempty set of indices $\Upsilon(\vec{H})$, called an $\vec{H}$**-anchor set**, such that for every index $i \in \Upsilon(\vec{H})$ and every $\vec{G} \in R(d)$, there is an induced subgraph $\vec{H}' \subseteq \vec{G}$ with $\vec{H}' \simeq \vec{H}$ and $x_i \in V(\vec{H}')$. All realizations $\vec{G} \in R(d)$ are called $\vec{H}$**-anchored digraphs**.
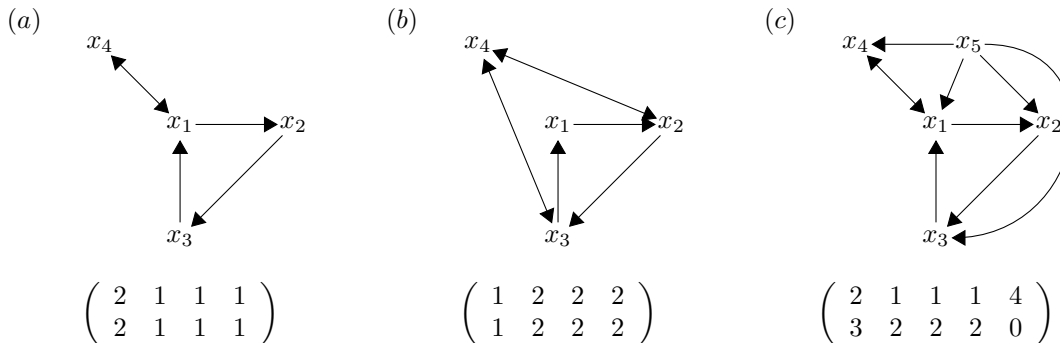
We are particularly interested in the case $\vec{H} = \vec{C}_3$. We first give a few examples to illustrate the non-trivial and specialized structure of $\vec{C}_3$-anchored digraphs. We will need one more definition used to describe their structure.

Definition 2 Let d be $\vec{C}_3$-anchored with Υ the $\vec{C}_3$-anchor set. For every $j \in \Upsilon$, define the $\vec{C}_3$**-scaffold set** S_j by

$$
S_j = \bigcup_{\vec{G} \in R(d)} \{l(C) : x_j \in C \subseteq V(\vec{G}) \text{ and } \vec{G}[C] \simeq \vec{C}_3\},
$$

where $l : V \longrightarrow \{1, \ldots, |V|\}$ is the vertex labeling function. □

Example 1 Consider the following three $\vec{C}_3$-anchored digraphs, listed with their degree sequences:
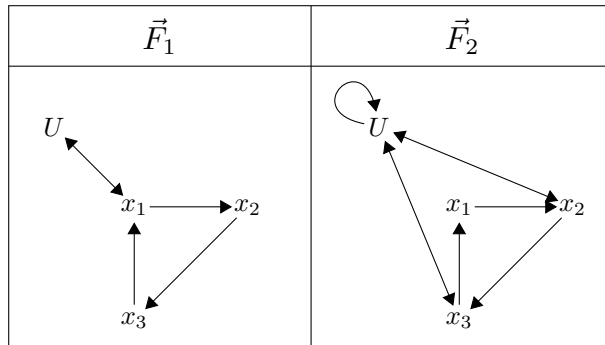

$$
(a) \quad \begin{pmatrix} 2 & 1 & 1 & 1 \\ 2 & 1 & 1 & 1 \end{pmatrix}
\qquad
(b) \quad \begin{pmatrix} 1 & 2 & 2 & 2 \\ 1 & 2 & 2 & 2 \end{pmatrix}
\qquad
(c) \quad \begin{pmatrix} 2 & 1 & 1 & 1 & 4 \\ 3 & 2 & 2 & 2 & 0 \end{pmatrix}
$$

Figure 1: $\vec{F}_1$ and $\vec{F}_2 \simeq \vec{F}_1^C$ are the two basic classes of $\vec{C}_3$-anchored digraphs. In $\vec{F}_1$, U is an independent set, while in $\vec{F}_2$, U is a clique (denoted by a self-loop).

In all three digraphs, the $\vec{C}_3$-anchor set $\Upsilon = \{1\}$, while the $\vec{C}_3$-scaffold set $S_1 = \{1,2,3,4\}$.

It is important to note that there can be multiple disjoint induced $\vec{C}_3$ subgraphs, as the following example shows:

$$\begin{pmatrix} 5 & 5 & 5 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 5 & 4 & 4 & 4 \end{pmatrix}$$
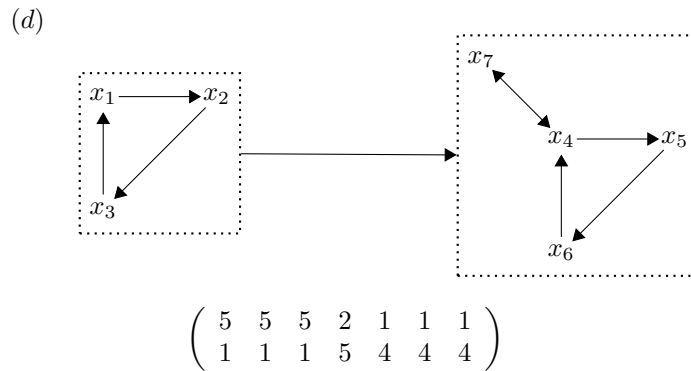
This graphic is meant to illustrate a *complete join* from the digraph in the box on the left to the one on the right, i.e. there is an arc $(x_i, x_j) \in A$ for all $i \in \{1,2,3\}$ and $j \in \{4,5,6,7\}$. In this case, the $\vec{C}_3$-anchor set $\Upsilon = \{1,2,3,4\}$, with $\vec{C}_3$-scaffold sets $S_1 = S_2 = S_3 = \{1,2,3\}$ and $S_4 = \{4,5,6,7\}$.

Besides the trivial example $\vec{C}_3$, examples 1(a) and (b) are specific examples of two fundamental building blocks of $\vec{C}_3$-anchored digraphs. These two fundamental building blocks are illustrated in Figure 1, with $\vec{F}_1$ and $\vec{F}_2$ generalizations of example 1(a) and (b), respectively. In both cases there is a vertex set U, which is an independent set in $\vec{F}_1$ and a clique in $\vec{F}_2$. Note that $\vec{F}_2 \simeq \vec{F}_1^C$.

2.1 Structural Characterization

The fundamental structure of $\vec{C}_3$-anchored digraphs was already discussed in the previous section, i.e. the base structure is given by either $\vec{F}_1$ or $\vec{F}_2$ as illustrated in Figure 1. Denote the vertex set for these graphs by F. The general structural characterization of $\vec{C}_3$-anchored digraphs consists of a composition of the base case $\vec{F}_i$ with a splitted digraph (LaMar 2010b). This structure is illustrated in Figure 2, with four additional vertex sets $S^\pm$, S^+, S^- and S^0 that form a *split partition*.

Split partitions consist of a disjoint union of at most four sets $\Pi = \{S^\pm, S^+, S^-, S^0\}$. Two of the sets, namely $S^\pm$ and S^0, are a clique and an independent set, respectively, with arbitrary connections between them. For the other two vertex sets S^+ and S^-, $\vec{G}[S^+]$ and $\vec{G}[S^-]$ are arbitrary induced subgraphs. A portion of the diagram in the left of Figure 2 shows the relations within and between the four different classes, with solid and dashed-dotted arrows denoting forced and allowable arcs, respectively, and the absence of an arc denoting no connections. The dashed self-loop on the set F denotes that the connections between

The corresponding adjacency matrix:

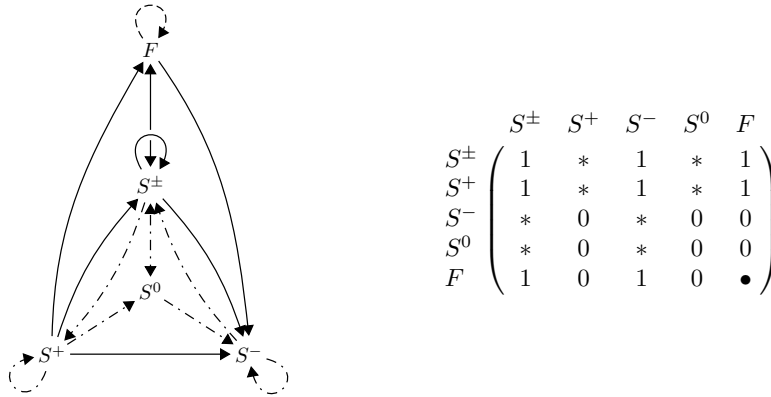|	$S^\pm$	S^+	S^-	S^0	F
$S^\pm$	1	*	1	*	1
S^+	1	*	1	*	1
S^-	*	0	*	0	0
S^0	*	0	*	0	0
F	1	0	1	0	•

Figure 2: Structural characterization of $\vec{C}_3$-anchored digraphs. The diagram on the left shows the relations within and between the 4 possible vertex sets of a split partition $\Pi = \{S^\pm, S^+, S^-, S^0\}$, as well as their relations via the split composition operator with a vertex set F. $\vec{G}[F]$ is isomorphic to one of the two basic classes of $\vec{C}_3$-anchored digraphs denoted in Figure 1, with this specific structure on F denoted by the dashed self-loop in this diagram. Solid and dashed-dotted arrows denote forced and allowable arcs, respectively, while the absence of an arrow denotes no arcs. $S^\pm$ is a clique, S^0 an independent set, while $\vec{G}[S^+]$ and $\vec{G}[S^-]$ are arbitrary subgraphs. The corresponding adjacency matrix in M-partition block form is displayed on the right.

vertices in F have the specified form in $\vec{F}_i$. Note that a number of vertex sets in Π can be empty. A digraph $\vec{G}$ together with a split partition Π of its vertex set is called a *splitted digraph* and is denoted by $(\vec{G}, \Pi)$.

Another useful illustration of a split partition $\Pi = \{S^\pm, S^+, S^-, S^0\}$ is via an adjacency matrix, which is denoted in the right of Figure 2, excluding the row and columns associated with the vertex set F, using the notation of M-partitions (Feder, Hell, Klein, and Motwani 2003). In general, an M-partition of a digraph $\vec{G}$ is a partition of the vertex-set $V(\vec{G})$ into k disjoint sets $\{X_1, \ldots, X_k\}$, where the arc constraints within and between sets are given by a $k \times k$ matrix M with elements in $\{0, 1, *\}$. M_{ii} equals 0 or 1 when X_i is an independent set or clique, respectively, and is set to $*$ when $\vec{G}[X_i]$ is an arbitrary subgraph. Similarly, for $i \neq j$, M_{ij} equal to 0, 1, or $*$ corresponds to $\vec{G}[X_i, X_j]$ having no arcs from X_i to X_j, all arcs from X_i to X_j, and no constraints on arcs from X_i to X_j, respectively. For the vertex set F, we have extended this definition by adding the symbol • to denote an induced subgraph $\vec{G}[F]$ with a specified structure (in this case, $\vec{G}[F] \simeq \vec{F}_i$).

The connections between F and the split partition Π in Figure 2 can be formalized by the use of a composition operator $\circ$ as follows. Letting $\vec{\Sigma}$ be the set of splitted digraphs and $\vec{\Gamma}$ the set of simple digraphs, we define the composition operator

$$\circ : \vec{\Sigma} \times \vec{\Gamma} \to \vec{\Gamma}$$

such that if $(\vec{H}, \Pi) \in \vec{\Sigma}$ and $\vec{F} = (F, A) \in \vec{\Gamma}$, the digraph $(\vec{H}, \Pi) \circ \vec{F}$ has adjacency matrix (in M-partition form) given on the right of Figure 2, with the corresponding connectivity diagram on the left. We thus have the following theorem.

Theorem 2 $\vec{G}$ is a $\vec{C}_3$-anchored digraph if and only if $\vec{G} = (\vec{H}, \Pi) \circ \vec{F}$, where $(\vec{H}, \Pi) \in \vec{\Sigma}$ and $\vec{F}$ is isomorphic to one of the two digraphs in Figure 1. $\square$

An alternate definition of $\vec{C}_3$-anchored digraphs is $\vec{G}$ is $\vec{C}_3$-anchored if and only if there is a splitted digraph $(\vec{H}, \Pi)$ such that $\vec{G}$ or $\vec{G}^C$ is isomorphic to $(\vec{H}, \Pi) \circ \vec{F}_1$. The following illustrates these concepts using Example 1.

Example 3 (1a) $F = \{x_1, x_2, x_3, x_4\}$, $\vec{G}[F]$ isomorphic to left graph in Figure 1 with $U = \{x_4\}$.
(1b) $F = \{x_1, x_2, x_3, x_4\}$, $\vec{G}[F]$ isomorphic to right graph in Figure 1 with $U = \{x_4\}$.

(1c)   Same as (a) with the addition of $S^+ = \{x_5\}$.

(1d)   (*i*) $F = \{x_1, x_2, x_3\}$, $\vec{G}[F] \simeq \vec{C}_3$ with $S^- = \{x_4, x_5, x_6, x_7\}$

(*ii*) $F = \{x_4, x_5, x_6, x_7\}$, $\vec{G}[F]$ isomorphic to left graph in Figure 1 with $U = \{x_7\}$ and $S^+ = \{x_1, x_2, x_3\}$.

2.2 Degree Sequence Characterization

For the degree sequence characterization, we will need integer-pair sequences to be non-increasing, in particular under the lexicographical ordering of either the first or second coordinate, which we define as follows:

Definition 3 (Positive/Negative ordering) An integer-pair sequence $d = \{(d_i^+, d_i^-)\}_{i=1}^N$ is non-increasing relative to the **positive lexicographical ordering** if and only if $d_i^+ \geq d_{i+1}^+$, with $d_i^- \geq d_{i+1}^-$ when $d_i^+ = d_{i+1}^+$. We say d is non-increasing relative to the **negative lexicographical ordering** by giving preference to the second coordinate. □

For a given integer-pair sequence $d = \{(d_i^+, d_i^-)\}_{i=1}^N$, define the sequences $\bar{d} = \{(\bar{d}_i^+, \bar{d}_i^-)\}_{i=1}^N$ and $\underline{d} = \{(\underline{d}_i^+, \underline{d}_i^-)\}_{i=1}^N$ to be positive and negative lexicographical orderings of d, respectively.

The characterization of $\vec{C}_3$-anchored degree sequences is closely linked to the inequalities that determine whether an integer-pair sequence is digraphic, which was first given by Fulkerson in the following theorem:

Theorem 4 (Fulkerson 2004) An integer-pair sequence d is digraphic if and only if $\sum_{i=1}^N d_i^+ = \sum_{i=1}^N d_i^-$ and for $k = 1, \ldots, N$,

$$\sum_{i=1}^k \min[\bar{d}_i^-, k-1] + \sum_{i=k+1}^N \min[\bar{d}_i^-, k] \geq \sum_{i=1}^k \bar{d}_i^+$$

$$(and)$$

$$\sum_{i=1}^k \min[\underline{d}_i^+, k-1] + \sum_{i=k+1}^N \min[\underline{d}_i^+, k] \geq \sum_{i=1}^k \underline{d}_i^-.$$

□

For an integer-pair sequence d, define the *slack* sequences $\bar{s}$ and $\underline{s}$ by

$$\bar{s}_k = \sum_{i=1}^k \min[\bar{d}_i^-, k-1] + \sum_{i=k+1}^N \min[\bar{d}_i^-, k] - \sum_{i=1}^k \bar{d}_i^+ \quad \text{with} \quad \bar{s}_0 \equiv 0,$$

$$\underline{s}_k = \sum_{i=1}^k \min[\underline{d}_i^+, k-1] + \sum_{i=k+1}^N \min[\underline{d}_i^+, k] - \sum_{i=1}^k \underline{d}_i^- \quad \text{with} \quad \underline{s}_0 \equiv 0.$$

Note that it is possible for the slack sequences to be negative. However, if the integer-pair sequence is digraphic, then they will both be non-negative. In other words, Theorem 4 can be restated as d is digraphic if and only if $\sum_{i=1}^N d_i^+ = \sum_{i=1}^N d_i^-$, with $\underline{s}$ and $\bar{s}$ both non-negative sequences. Note that by definition, we have $\underline{s}_0 = \bar{s}_0 = \underline{s}_N = \bar{s}_N = 0$. It is the structure of these slack sequences that indicates when a degree sequence is $\vec{C}_3$-anchored, as illustrated in the following Theorem:

Theorem 5 The degree sequence d is $\vec{C}_3$-anchored if and only if for $J = \{j_1, \ldots, j_n\}$ a set of indices with $3 \leq n \leq N-1$ and $(k, l) \geq (0, 0)$ an index pair, d satisfies one of the three cases

(*i*) $n = 3$ and $d_{j_1} = d_{j_2} = d_{j_3} = (l+1, k+1)$
(*ii*) $n > 3$ and $d_{j_1} = (l+n-2, k+n-2)$, $d_{j_2} = \cdots = d_{j_n} = (l+1, k+1)$
(*iii*) $n > 3$ and $d_{j_1} = \cdots = d_{j_{n-1}} = (l+n-2, k+n-2)$, $d_{j_n} = (l+1, k+1)$

with

$$(d_{j_1}, \ldots, d_{j_n}) = (\bar{d}_{k+1}, \ldots, \bar{d}_{k+n}) = (\underline{d}_{l+1}, \ldots, \underline{d}_{l+n})$$

and the slack sequences satisfying

$$(0,1,\ldots,1,0) = (\bar{s}_k, \bar{s}_{k+1}, \ldots, \bar{s}_{k+n-1}, \bar{s}_{k+n}) = (\underline{s}_l, \underline{s}_{l+1}, \ldots, \underline{s}_{l+n-1}, \underline{s}_{l+n}).$$

□

Before giving examples, we will illustrate a computationally more tractable technique to compute the slack sequences. Given an integer sequence a, define the *corrected conjugate sequence* a'' (Berge 1973) by

$$a_k'' = |\{i \mid i < k \text{ and } a_i \geq k-1\}| + |\{i \mid i > k \text{ and } a_i \geq k\}|.$$

The numbers a'' can be represented by what is known as the *corrected Ferrers diagram* (Berge 1973), shown with the example sequence $a = (4\ 3\ 4\ 2\ 1)$ below.

	a_1''	a_2''	a_3''	a_4''	a_5''	
a_1	○	●	●	●	●	4
a_2	●	○	●	●	○	3
a_3	●	●	○	●	●	4
a_4	●	●	○	○	○	2
a_5	●	○	○	○	○	1
	4	3	2	3	2	

If in the i-th row a_i solid dots are filled in from left to right, making sure we skip the i-th column, then the value a_k'' is found by simply counting the number of solid dots in the k-th column, giving $a'' = (4\ 3\ 2\ 3\ 2)$. The partial sums of the corrected conjugate sequence are given by

$$\sum_{i=1}^{k} a_i'' = \sum_{i=1}^{k} \min[a_i, k-1] + \sum_{i=k+1}^{N} \min[a_i, k],$$

corresponding to the sums on the left of the inequalities in Theorem 4. Thus, we have

$$\bar{s}_k = \sum_{i=1}^{k} [\bar{d}^-]_i'' - \sum_{i=1}^{k} \bar{d}_i^+ \qquad \text{and} \qquad \underline{s}_k = \sum_{i=1}^{k} [\underline{d}^+]_i'' - \sum_{i=1}^{k} \underline{d}_i^-.$$

Example 6 We will now illustrate Theorem 5 on the degree sequence from Example 1(d), namely

$$\begin{pmatrix} 5 & 5 & 5 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 5 & 4 & 4 & 4 \end{pmatrix}$$

with the Ferrer diagrams for $[\bar{d}^-]''$ on the left and $[\underline{d}^+]''$ on the right given by

```
○  ●  ○  ○  ○  ○  ○          ○  ●  ●  ○  ○  ○  ○
●  ○  ○  ○  ○  ○  ○          ●  ○  ○  ○  ○  ○  ○
●  ○  ○  ○  ○  ○  ○          ●  ○  ○  ○  ○  ○  ○
●  ●  ●  ○  ●  ●  ○          ●  ○  ○  ○  ○  ○  ○
●  ●  ●  ●  ○  ○  ○          ●  ●  ●  ●  ○  ●  ○
●  ●  ●  ●  ○  ○  ○          ●  ●  ●  ●  ●  ○  ○
●  ●  ●  ●  ○  ○  ○          ●  ●  ●  ●  ●  ○  ○
```

$[\bar{d}^-]''$	6	5	4	3	1	1	0		$[\underline{d}^+]''$	6	4	4	3	2	1	0
$\bar{d}^+$	5	5	5	2	1	1	1		$\underline{d}^-$	5	4	4	4	1	1	1
$\bar{s}$	1	1	0	1	1	1	0		$\underline{s}$	1	1	1	0	1	1	0

Since there are two anchored 3-cycles in this example, it satisfies the following two specific cases of Theorem 5: (*i*) with $(k,l) = (0,4)$ and $J = \{1,2,3\}$, and (*ii*) with $(k,l) = (3,0)$ and $J = \{4,5,6,7\}$. $\square$

The following theorem illustrates how to identify the $\vec{C}_3$-anchor set from the degree sequence, as well as which indices of the degree sequence belong to the vertex set F and the split partition $\Pi = \{S^{\pm}, S^+, S^-, S^0\}$.

Theorem 7 Let d be $\vec{C}_3$-anchored with Υ the $\vec{C}_3$-anchor set. For each $j \in \Upsilon$, the $\vec{C}_3$-scaffold set S_j is given by the index set $\{j_1, \ldots, j_n\}$ from Theorem 5 with $n \geq 3$ and $(k,l) \geq (0,0)$. We can also identify all $j \in \Upsilon$ in reference to the 3 cases in Theorem 5 as follows:

(*i*) $j \in \{j_1, j_2, j_3\}$ and $d_j = (l+1, k+1)$
(*ii*) $j = j_1$ and $d_j = (l+n-2, k+n-2)$.
(*iii*) $j = j_n$ and $d_j = (l+1, k+1)$.

For the structural characterization, letting $F = V_{S_j}$, we have (*i*) $\vec{G}[F] \simeq \vec{C}_3$, (*ii*) $\vec{G}[F] \simeq \vec{F}_1$, and (*iii*) $\vec{G}[F] \simeq \vec{F}_2$. Finally, to identify the split partition, if $A_k = \{\bar{\pi}(i)\}_{i=1}^k$ and $B_l = \{\underline{\pi}(i)\}_{i=1}^l$, then we have

$$
\begin{aligned}
S^{\pm} &= V_{A_k \cap B_l} \\
S^+ &= V_{A_k - B_l} \\
S^- &= V_{B_l - A_k} \\
S^0 &= V_{A_k^C \cap B_l^C - S_j}.
\end{aligned}
$$

$\square$

Example 8 Continuing again with Example 1(d), we have $\Upsilon = \{1,2,3,4\}$. For $j \in \{1,2,3\} \subset \Upsilon$, we have $S_j = \{1,2,3\}$ and $(k,l) = (0,4)$. In this case, $F = V_{S_j}$ with $\vec{G}[F] \simeq \vec{C}_3$. For the split partition, we first note that $\bar{\pi} = (1,2,3,4,5,6,7)$ and $\underline{\pi} = (4,5,6,7,1,2,3)$. This gives $A_k = A_0 = \emptyset$ and $B_l = B_4 = \{4,5,6,7\}$, which leads to $S^{\pm} = S^+ = S^0 = \emptyset$ and $S^- = \{x_4, x_5, x_6, x_7\}$.

For the next case with $j = 4 \in \Upsilon$, we have $S_4 = \{4,5,6,7\}$ and $(k,l) = (3,0)$. Now we have $F = V_{S_4}$ with $\vec{G}[F] \simeq \vec{F}_1$. For the split partition in this case, $A_k = A_3 = \{1,2,3\}$ and $B_l = B_0 = \emptyset$, leading to $S^{\pm} = S^- = S^0 = \emptyset$ and $S^+ = \{x_1, x_2, x_3\}$.

3 UNIFORM SAMPLING REALIZATIONS FROM A FIXED DEGREE SEQUENCE

As mentioned in the introduction, the uniform sampling of realizations from a fixed degree sequence has a long history. In this section we will use the characterization of $\vec{C}_3$-anchored degree sequences to shed some light into the structure of Markov chain Monte Carlo arc-switching techniques commonly used to uniformly sample *simple directed* realizations from a fixed degree sequence.

Given a degree sequence d, we define the meta-graph $\vec{\Omega}_d = (\mathcal{V}, \mathcal{A})$, where $\mathcal{V}$ is in one-to-one correspondence with $R(d)$. We will denote $V_{\vec{G}} \in \mathcal{V}$ to be the vertex corresponding to $\vec{G} \in R(d)$. There are two types of arcs $\mathcal{A} = \mathcal{A}_2 \cup \mathcal{A}_3$: $(V_{\vec{G}}, V_{\vec{G}'}) \in \mathcal{A}_2$ if there is a 2-switch from $\vec{G}$ to $\vec{G}'$. Similarly, $(V_{\vec{G}}, V_{\vec{G}'}) \in \mathcal{A}_3$ if there is a $\vec{C}_3$ reorientation from $\vec{G}$ to $\vec{G}'$. Note that the meta-graph at this point is symmetric, and therefore it may seem a little strange to be representing it as a digraph. As we will see below, we will put probability weights on the arcs which will represent the probability of making a particular move. In general, the resulting weighted digraph will not be symmetric.

Example 9 The meta-graph for the degree sequence from Example 1(d) is given by (directed arcs suppressed for clarity)
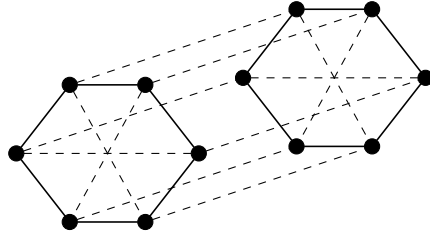
with each vertex corresponding to a realization and a solid/dashed edge corresponding to a 2-switch/$\vec{C}_3$-reorientation connecting two realizations.

The following result by Rao, Jana, and Bandyopadhyay (1996) shows that we can go from one realization to any other realization by a series of 2-switches and 3-cycle reorientations.

Theorem 10 (Rao, Jana, and Bandyopadhyay 1996) The meta-graph $\vec{\Omega}_d$ is strongly connected. □

The simplest example illustrating that 3-cycle reorientations are necessary in general is the 3-cycle itself $\vec{C}_3$. In this case, there are no 2-switches that can reorient the cycle.

We can define a Markov chain random walk on $\vec{\Omega}_d$ by an appropriate choice of probability weights for each arc in $\mathscr{A}$. There are many choices for the weights, but for simplicity we will give as an example probability weights induced by a particularly simple random walk algorithm (Roberts 2000, Cobb and Chen 2003). Given a realization $\vec{G}^{(i)} \in R(d)$, with probability p attempt a 2-switch and with probability $1 - p$ a $\vec{C}_3$ reorientation. For a 2-switch, choose four vertices without replacement and, if possible, perform a 2-switch to arrive at $\vec{G}^{(i+1)}$. Otherwise, do nothing, i.e. $\vec{G}^{(i+1)} = \vec{G}^{(i)}$. Similarly, for a $\vec{C}_3$ reorientation choose three vertices without replacement and, if possible, perform a $\vec{C}_3$ reorientation to arrive at $\vec{G}^{(i+1)}$. Otherwise, do nothing. The resulting probabilities for this Markov chain are given by

$$P_{ij} = \begin{cases} p/\binom{n}{4} & \text{if } (V_{\vec{G}_i}, V_{\vec{G}_j}) \in \mathscr{A}_2, \\ (1-p)/\binom{n}{3} & \text{if } (V_{\vec{G}_i}, V_{\vec{G}_j}) \in \mathscr{A}_3, \end{cases}$$

where n represents the number of vertices, or the length of the degree sequence. By doing nothing with failed move attempts, we impose self-loops at each realization such that $\sum_{j=1}^{|R(d)|} P_{ij} = 1$. By Theorem 10, this Markov chain is irreducible, and it is easily seen to be symmetric and aperiodic. Thus, there is a unique limiting distribution which by symmetry must be the uniform distribution (Cobb and Chen 2003, Lovasz 1993). Note that while there are many other techniques that lead to larger probability weights, and therefore faster mixing times, it is unclear if this leads to shorter run-times as they usual require extra data structures whose maintenance may be costly.

It is mentioned in (Roberts 2000) and (Rao, Jana, and Bandyopadhyay 1996) that in most situations one need only use 2-switches, and thus we can choose p to be close to 1. The difficulty with this is there are degree sequences where this may lead to very long mixing times, due to the rare cases where there is not a path with arcs in $\mathscr{A}_2$ connecting two realizations (see the meta-graph in example 9 above). The rarity of these cases is also unknown, and so there is no way to know how close to 1 one should choose p. We now show that the only degree sequences that require directed 3-cycle reorientations are a subset of the $\vec{C}_3$-anchored degree sequences, which we call $\vec{C}_3^*$-anchored and defined as follows:

Definition 4 ($\vec{C}_3^*$-anchored)

(i) d is $\vec{C}_3^*$-anchored if and only if d satisfies case (i) of Theorem 5.

(ii) $\vec{G}$ is $\vec{C}_3^*$-anchored if and only if $\vec{G} \simeq (\vec{H}, \Pi) \circ \vec{C}_3$, where $(\vec{H}, \Pi)$ is a splitted digraph.

If we define the meta-graph $\vec{\Omega}'_d = (\mathscr{V}, \mathscr{A}_2)$, i.e. we only allow 2-switches in our move-set, then the following is the main theorem of this section.

Theorem 11 $\vec{\Omega}'_d$ is disconnected if and only if d is $\vec{C}_3^*$-anchored.

Proof Let d be $\vec{C}_3^*$-anchored and $\vec{G} \in R(d)$. It should be clear that $\vec{\Omega}'_d$ is disconnected, since every realization has a directed 3-cycle through the same three vertices, and thus no 2-switches will connect realizations with the opposite orientations for that 3-cycle without first removing the 3-cycle.

Suppose d is a degree sequence with $\vec{\Omega}'_d$ disconnected. By Theorem 10, there must be a realization $\vec{G} \in R(d)$ with an oriented directed 3-cycle C such that $V_{\vec{G}}$ is in one connected component of $\vec{\Omega}'_d$ and $V_{\vec{G}'}$ is in another connected component, where $\vec{G}'$ is found from $\vec{G}$ by reorienting C. Let $C = \{u, v, w\}$ with $\{(u,v), (v,w), (w,u)\} \subset A$. We want to show that for any vertex $x \in V - C$, x must be in one of the vertex classes S^0, S^-, S^+ or $S^{\pm}$. We will show that $(C, x) \subset A$ and/or A^C (by symmetry, we will also have $(x, C) \subset A$ and/or A^C). Suppose there is only one of the three arcs, and without loss of generality choose $(v, x) \in A$, with $\{(u,x), (w,x)\} \subset A^C$ (the case with two existing arcs follows by considering the graph complement). The left panel in Fig. 3 shows that we can perform a series of 2-switches to reorient the 3-cycle, which contradicts $\vec{G}$ and $\vec{G}'$ being in two separate connected components of $\vec{\Omega}'_d$. Thus, we must have $(C, x) \subset A$ and/or A^C, showing $x \in S^0$, S^-, S^+ or $S^{\pm}$.

Now we must show that the connections between the vertex classes are given as in Figure 2. Let $x \in S^0 \cup S^+$ and $y \in S^0 \cup S^-$, and suppose that $(x, y) \in A$. In the right panel of Fig. 3, we see again that there is a series of 2-switches which reorients the 3-cycle, showing $(x, y) \notin A$. By considering the graph complement, we can prove $(x, y) \in A$ for $x \in S^{\pm} \cup S^-$ and $y \in S^{\pm} \cup S^+$. This shows d is $\vec{C}_3^*$-anchored, thereby completing the proof.

□

For every anchored 3-cycle C, there are two isomorphic copies of connected components of $\vec{\Omega}'_d$ corresponding to each orientation of C. In general, $\vec{\Omega}_d$ thus has the form given by the following corollary.

Corollary 12

$$\vec{\Omega}_d \simeq \vec{\Omega}_d[\mathcal{V}(\vec{G}_2)] \times \left(\times_{i=1}^k \vec{K}_2 \right),$$

where $\vec{G}_2$ is one connected component of $\vec{\Omega}'_d$, k denotes the number of anchored 3-cycles and $\vec{K}_2$ is the complete digraph on 2 vertices. □

If we know where the anchored 3-cycles are, then we can simply choose an orientation for each anchored 3-cycle uniformly at random, and then perform a random walk on the graph $\vec{\Omega}'_d$ using only 2-switches. As mentioned previously, however, using only 2-switches can lead to longer mixing times, so we could implement yet another technique mentioned in (Berger and Mller-Hannemann 2010) whereby we reduce the degrees of all vertices in the anchor set by 1 and then perform a random walk using both 2-switches and 3-cycle reorientations. The efficiency of either of these techniques depends on whether the identification of the anchored 3-cycles can be done without too much work, of which Corollary 7 allows us to do. The computational complexity is easily seen to be at least quadratic in the number of vertices, as seen in the calculation of the slack sequences and Theorem 5. This is in contrast to (Berger and Mller-Hannemann 2010) where identification of the anchored 3-cycles is quadratic in the number of arcs.

Unfortunately, for degree sequences that are not $\vec{C}_3^*$-anchored, all of this extra work does not gain us anything. Given the apparent rarity of $\vec{C}_3^*$-anchored degree sequences (how rare is still an open question), it seems it might be better in general to just do a random walk using 2-switches and 3-cycle reorientations from the start. Thus, the usefulness of this result depends on how the complexity of $\vec{C}_3$-anchor identification compares with the mixing time, which is still an open question.

4 CONCLUSION

This paper gives structural and degree sequence characterizations of a class of digraphs called $\vec{C}_3$-anchored digraphs and discusses their application to the uniform sampling of digraphs from a fixed degree sequence. This digraph structure uses the important concept of splitted digraphs, whose undirected counterparts, split
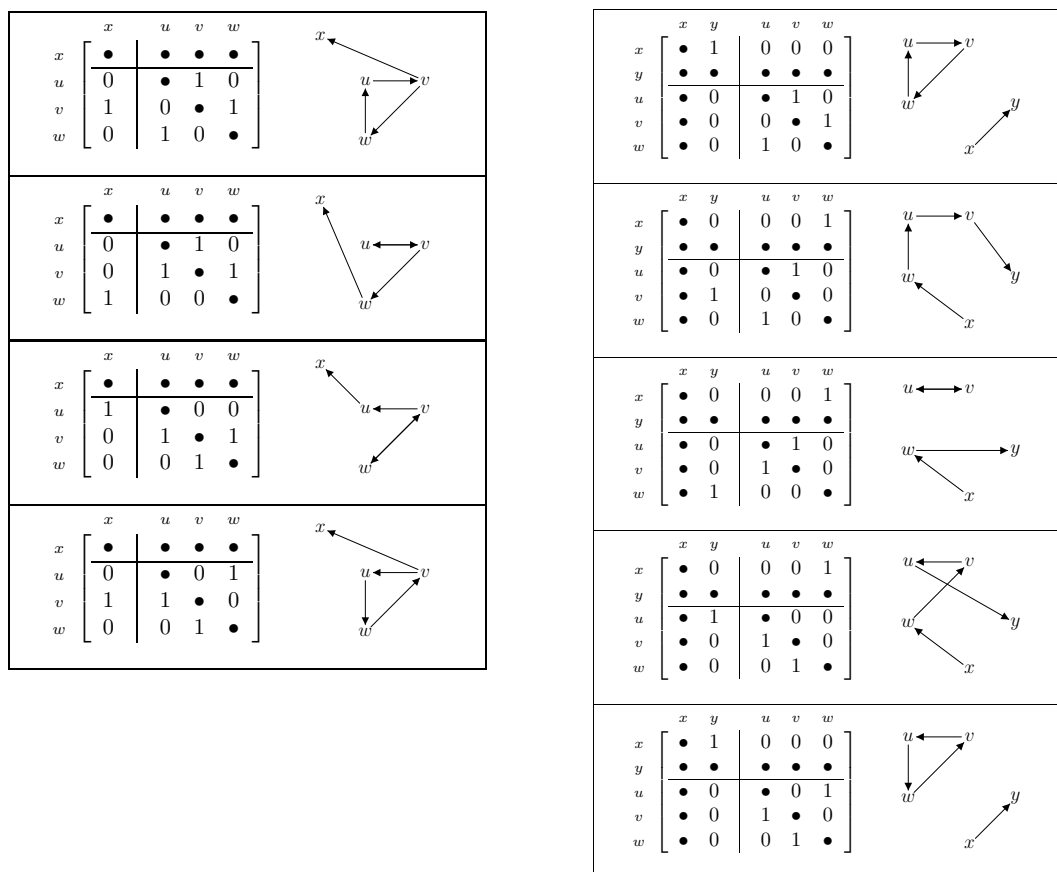
Figure 3: Both left and right panels show a series of 2-switches used to reorient a directed 3-cycle $\vec{C}_3$ (see Theorem 11). The solid dots denote entries of the adjacency matrix which are not used in the series of moves.

graphs, have proven to be very useful (Mahadev and Peled 1995). This paper hopefully serves as just a starting point in exploring the undoubtedly many parallels between the undirected and directed versions of split graphs.

An interesting avenue for future research would be an exploration into structural and degree sequence characterizations of $\vec{H}$-anchored digraphs, for arbitrary graphs $\vec{H}$. If the degree sequence for $\vec{H}$ is unidigraphic, i.e. all realizations are isomorphic, then by referring to Figure 2 it can be easily seen that for any splitted digraph $(\vec{G}, \Pi)$, the graph $(\vec{G}, \Pi) \circ \vec{H}$ will be $\vec{H}$-anchored. Whether this is a necessary condition is an open question.

Lastly, a direct comparison of Markov chain Monte Carlo arc-switching techniques and sequential importance sampling algorithms is of interest. This is complicated by the fact that general bounds on mixing times for Markov chain Monte Carlo techniques are either not known or far from tight, making run-time complexity difficult to compute. Related to this are the many possibilities for constructing the probability weights on the arcs of the meta-graph: in general, while some choices lead to larger switching probabilities, and therefore faster mixing times, these can be misleading as they tend to be associated with the required use of data structures which may lead to either poor run-time implementations and/or memory issues. Also, to the authors knowledge there exists no known sequential importance sampling algorithm for the case of simple directed graphs. Presumably, however, many of the ideas from (Blitzstein and Diaconis 2011) and (del Genio et al. 2010) can be extended to the case of directed graphs (see (Erdős, Miklós, and Toroczkai 2010) and (Kim et al. 2009) for beginning steps).

ACKNOWLEDGMENTS

REFERENCES

Berge, C. 1973. "Graphs and hypergraphs". North-Holland Publishing Company.

Berger, A., and M. Mller-Hannemann. 2010. "Uniform Sampling of Digraphs with a Fixed Degree Sequence". In *Graph Theoretic Concepts in Computer Science*, edited by D. Thilikos, Volume 6410 of *Lecture Notes in Computer Science*, 220–231. Springer Berlin / Heidelberg.

Blitzstein, J., and P. Diaconis. 2011. "A sequential importance sampling algorithm for generating random graphs with prescribed degrees". *Internet Math.* 6 (4): 489–522.

Cobb, G. W., and Y.-P. Chen. 2003. "An application of Markov Chain Monte Carlo to community ecology". *The American Mathematical Monthly* 110 (4): 265–288.

del Genio, C. I., H. Kim, Z. Toroczkai, and K. E. Bassler. 2010, March. "Efficient and exact sampling of simple graphs with given arbitrary degree sequence". *PLoS ONE* 5 (4): 1–7.

Erdős, P. L., I. Miklós, and Z. Toroczkai. 2010, April. "A simple Havel-Hakimi type algorithm to realize graphical degree sequences of directed graphs". *The Electronic Journal of Combinatorics* 17 (1): 1–10.

Feder, T., P. Hell, S. Klein, and R. Motwani. 2003, January. "List Partitions". *SIAM J. Discrete Math.* 16 (3): 449–478.

Fulkerson, D. R. 2004, March. "Zero-one matrices with zero trace". *Pacific Journal of Mathematics* 10 (3): 831–836.

Kim, H., Z. Toroczkai, P. L. Erdős, I. Miklós, and L. Székely. 2009, September. "Degree-based graph construction". *Journal of Physics A: Mathematical and Theoretical* 42:1–11.

LaMar, M. D. 2010a, June. "Algorithms for realizing degree sequences of directed graphs". arXiv:0906.0343.

LaMar, M. D. 2010b, May. "Split digraphs". arXiv:1005.2452.

Liljeros, F., C. R. Edling, L. A. Amaral, H. E. Stanley, and Y. Aberg. 2001, June. "The web of human sexual contacts". *Nature* 411 (6840): 907–908.

Lovasz, L. 1993, January. "Random walks on graphs: A survey". *Combinatorics* 2:1–46.

Mahadev, N. V. R., and U. N. Peled. 1995, September. *Threshold Graphs and Related Topics, Volume 56 (Annals of Discrete Mathematics)*. 1 ed. North Holland.

Rao, A., R. Jana, and S. Bandyopadhyay. 1996. "A Markov chain Monte Carlo method for generating random (0,1)-matrices with given marginals". *Sankhya Ser. A* 58 (2): 225–242.

Rao, S. 1981. "A survey of the theory of potentially P-graphic and forcibly P-graphic degree sequences". In *Combinatorics and Graph Theory*, edited by S. Rao, Volume 885 of *Lecture Notes in Mathematics*, 417–440. Springer Berlin / Heidelberg. 10.1007/BFb0092288.

Roberts, J. 2000. "Simple methods for simulating sociomatrices with given marginal totals". *Social Networks* 22 (3): 273–283.

AUTHOR BIOGRAPHIES

M. DREW LAMAR is an Assistant Professor in the Department of Biology at the The College of William and Mary. He holds a PhD degree in Mathematics from The University of Texas at Austin. His research interests are in mathematical biology, scientific computation, dynamical systems, stochastic processes and graph theory, with an emphasis in coupled oscillators, complex networks, directed graphs, Markov chains, neurobiology, cellular physiology and ecology. His email address is mdlama@wm.edu.