Discrete Applied Mathematics 158 (2010) 1977–1985

Contents lists available at ScienceDirect

Discrete Applied Mathematics

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

Maximal pivots on graphs with an application to gene assembly

Robert Brijder *, Hendrik Jan Hoogeboom

Leiden Institute of Advanced Computer Science, Leiden University, The Netherlands

ARTICLE INFO

Article history:
Received 15 July 2009
Received in revised form 4 August 2010
Accepted 31 August 2010
Available online 24 September 2010

Keywords:
Principal pivot transform
Algebraic graph theory
Overlap graph
Gene assembly in ciliates

ABSTRACT

We consider principal pivot transform (*pivot*) on graphs. We define a natural variant of this operation, called dual pivot, and show that both the kernel and the set of maximally applicable pivots of a graph are invariant under this operation. The result is motivated by and applicable to the theory of gene assembly in ciliates.

1. Introduction

The pivot operation, due to Tucker [18], partially (component-wise) inverts a given matrix. It appears naturally in many areas including mathematical programming and numerical analysis; see [17] for a survey. Over $\mathbb{F}_2$ (which is the natural setting to consider for graphs), the pivot operation has, in addition to matrix and graph interpretations [11], also an interpretation in terms of delta matroids [1].

In this paper we define the *dual pivot*, which has an identical effect on graphs as the (regular) pivot, however the condition for it to be applicable differs. The main result of the paper is that any two graphs in the same orbit under dual pivot have the same family of maximal pivots (cf. Theorem 16), i.e., the same family of maximally partial inverses of that matrix. This result is obtained by combining each of the aforementioned interpretations of pivot.

This research is motivated by the theory of gene assembly in ciliates [9], which is recalled in Section 7. Without the context of gene assembly this main result (Theorem 16) is surprising; it is not found in the extensive literature on pivots. It fits however with the intuition and results from the string based model of gene assembly [4], and in this paper we formulate it for the more general graph based model. It is understood and proven here using completely different techniques, algebraical rather than combinatorial.

2. Notation and terminology

The field with two elements is denoted by $\mathbb{F}_2$. Our matrix computations will be over $\mathbb{F}_2$. Hence addition is equal to the logical exclusive-or, also denoted by $\oplus$, and multiplication is equal to the logical conjunction, also denoted by $\wedge$. These operations carry over to sets, e.g., for sets $A, B \subseteq V$ and $x \in V$, $x \in A \oplus B$ iff $(x \in A) \oplus (x \in B)$.

A set system is a tuple $M = (V, D)$, where V is a finite set and $D \subseteq 2^V$ is a set of subsets of V. Let $\min(D)$ ($\max(D)$, resp.) be the family of minimal (maximal, resp.) sets in D w.r.t. set inclusion, and let $\min(M) = (V, \min(D))$ ($\max(M) = (V, \max(D))$, resp.) be the corresponding set systems.

* Corresponding author. Fax: +31 71 5276985.
E-mail address: rbrijder@liacs.nl (R. Brijder).

0166-218X/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.dam.2010.08.030

Let V be a finite set, and A be a $V \times V$-matrix (over an arbitrary field), i.e., A is a matrix where the rows and columns of A are identified by elements of V. Therefore, e.g., the following matrices with $V = \{p, q\}$ are equal: $\begin{array}{cc} & \begin{array}{cc} p & q \end{array} \\ \begin{array}{c} p \\ q \end{array} & \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \end{array}$ and $\begin{array}{cc} & \begin{array}{cc} q & p \end{array} \\ \begin{array}{c} q \\ p \end{array} & \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \end{array}$. For $X \subseteq V$, the principal submatrix of A w.r.t. X is denoted by $A[X]$, i.e., $A[X]$ is the $X \times X$-matrix obtained from A by restricting to rows and columns in X. Similarly, we define $A \setminus X = A[V \setminus X]$. Notions such as matrix inversion A^{-1} and determinant $\det(A)$ are well defined for $V \times V$-matrices. By convention, $\det(A[\varnothing]) = 1$.

A set $X \subseteq V$ is called dependent in A iff the columns of A corresponding to X are linearly dependent. We define $\mathcal{P}_A = (I, D)$ to be the partition of 2^V such that D (I, respectively) contains the dependent (independent, respectively) subsets of V in A. By convention, $\varnothing \in I$. The sets in $\max(I)$ are called the bases of A.

We have that $\mathcal{P}_A = (I, D)$ is uniquely determined by $\max(I)$ (and the set V). Similarly, $\mathcal{P}_A$ is uniquely determined by $\min(D)$ (and the set V). These properties are specifically used in matroid theory, where a matroid may be described by its independent sets (V, I), by its family of bases $(V, \max(I))$, or by its circuits $(V, \min(D))$. Moreover, for each basis $X \in \max(I)$, $|X|$ is equal to the rank r of A.

We consider undirected graphs without parallel edges, however we do allow loops. For a graph $G = (V, E)$ we use $V(G)$ and $E(G)$ to denote its set of vertices V and set of edges E, respectively, where for $x \in V$, $\{x\} \in E$ iff x has a loop. For $X \subseteq V$, we denote the subgraph of G induced by X as $G[X]$.

With a graph G one associates its adjacency matrix $A(G)$, which is a $V \times V$-matrix $(a_{u,v})$ over $\mathbb{F}_2$ with $a_{u,v} = 1$ iff $\{u, v\} \in E$. The matrices corresponding to graphs are precisely the symmetric $\mathbb{F}_2$-matrices; loops corresponding to diagonal 1's. Note that for $X \subseteq V$, $A(G[X]) = (A(G))[X]$.

Over $\mathbb{F}_2$, vectors indexed by V can be identified with subsets of V, and a $V \times V$-matrix defines a linear transformation on subsets of V. The kernel (also called null space) of a matrix A, denoted by $\ker(A)$ is determined by those linear combinations of column vectors of A that sum up to the zero vector 0. Working in $\mathbb{F}_2$, we regard the elements of $\ker(A)$ as subsets of V. Moreover, the kernel of A is the eigenspace $E_0(A)$ on value 0, and similar as $\ker(A)$, the elements of the (only other) eigenspace $E_1(A) = \{v \in V \mid Av = v\}$ on value 1 are also considered as sets.

We will often identify a graph with its adjacency matrix, so, e.g., by the determinant of graph G, denoted by $\det G$, we will mean the determinant $\det A(G)$ of its adjacency matrix computed over $\mathbb{F}_2$. In the same vein we will often simply write $\ker(G)$, $E_1(G)$, $\mathcal{P}_G$, etc.

Let $\mathcal{P}_G = (I, D)$ for some graph G. As G is a $V \times V$-matrix over $\mathbb{F}_2$, we have that $X \in D$ iff there is a $S \subseteq X$ with $S \in \ker(G) \setminus \{\varnothing\}$. Moreover, $\min(D) = \min(\ker(G) \setminus \{\varnothing\})$ and $\ker(G)$ is the closure of $\min(D)$ under $\oplus$ (i.e., $\min(D)$ spans $\ker(G)$). Consequently, $\min(D)$ uniquely determines $\ker(G)$ and vice versa. As $\min(D)$ in turn uniquely determines $\mathcal{P}_G$, the following holds.

Corollary 1. *For graphs G_1 and G_2, $\ker(G_1) = \ker(G_2)$ iff the families of bases of G_1 and of G_2 are equal.*

3. Pivots

In general the pivot operation can be studied for matrices over arbitrary fields, e.g., as done in [17]. In this paper we restrict ourselves to symmetric matrices over $\mathbb{F}_2$, which leads to a number of additional viewpoints to the same operation, and for each of them an equivalent definition for pivoting. Each of these definitions is known, but (to the best of our knowledge) they were not before collected in one text.

Matrices. Let A be a $V \times V$-matrix (over an arbitrary field), and let $X \subseteq V$ be such that $A[X]$ is nonsingular, i.e., $\det A[X] \neq 0$. The *pivot* of A on X, denoted by $A * X$, is defined as follows; see [18]. Let $A = \left(\begin{array}{c|c} P & Q \\ \hline R & S \end{array}\right)$ with $P = A[X]$. Then

$$A * X = \left(\begin{array}{c|c} P^{-1} & -P^{-1}Q \\ \hline RP^{-1} & S - RP^{-1}Q \end{array}\right).$$

Matrix $(A * X) \setminus X = S - RP^{-1}Q$ is called the *Schur complement* of X in A.

The pivot is sometimes considered a partial inverse, as A and $A * X$ are related by the following characteristic equality, where the vectors x_1 and y_1 correspond to the elements of X. In fact, this formula defines $A * X$ given A and X [17].

$$A \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \quad \text{iff} \quad A * X \begin{pmatrix} y_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ y_2 \end{pmatrix}. \tag{1}$$

Note that if $\det A \neq 0$, then $A * V = A^{-1}$. By Eq. (1) we see that a pivot operation is an involution (i.e., operation of order 2), and more generally, if $(A * X) * Y$ is defined, then $A * (X \oplus Y)$ is defined and they are equal.

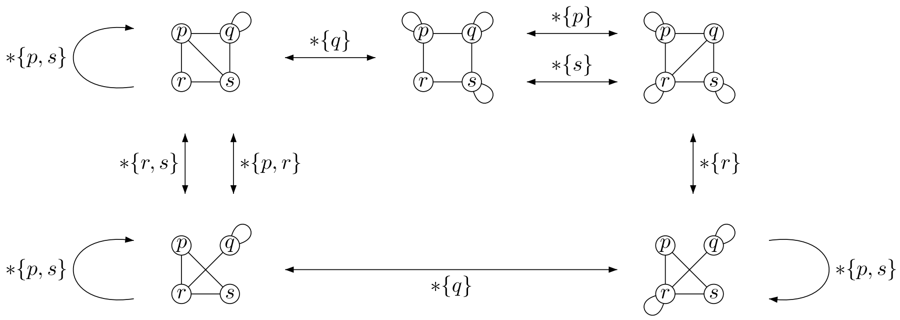

Fig. 1. The orbit of G under pivot. Only the elementary pivots are shown.

The following fundamental result on pivots is due to Tucker [18] (see also [7, Theorem 4.1.1]). It is used in [3] to study sequences of pivots.

Proposition 2 (*[18]*)**.** *Let A be a $V \times V$-matrix, and let $X \subseteq V$ be such that* $\det A[X] \neq 0$. *Then, for $Y \subseteq V$,* $\det(A * X)[Y] = \det A[X \oplus Y]/\det A[X]$.

It may be interesting to remark here that Proposition 2 for the case $Y = V \setminus X$ is called the Schur determinant formula and was shown already in 1917 by Issai Schur; see [16].

It is easy to verify from the definition of pivot that $A * X$ is skew-symmetric whenever A is. In particular, if G is a graph (i.e., a symmetric matrix over $\mathbb{F}_2$), then $G * X$ is also a graph. From now on we restrict our attention to graphs.

Delta Matroids. Consider now a set system $M = (V, D)$. We define, for $X \subseteq V$, the *twist* $M * X = (V, D * X)$, where $D * X = \{Y \oplus X \mid Y \in D\}$.

Let G be a graph and let $\mathcal{M}_G = (V(G), D_G)$ be the set system with $D_G = \{X \subseteq V(G) \mid \det G[X] = 1\}$. It is easy to verify that G can be (re)constructed given $\mathcal{M}_G$: $\{u\}$ is a loop in G iff $\{u\} \in D_G$, and $\{u, v\}$ is an edge in G iff $(\{u, v\} \in D_G) \oplus ((\{u\} \in D_G) \wedge (\{v\} \in D_G))$; see [2, Property 3.1]. In this way, the family of graphs (with set V of vertices) can be considered as a subset of the family of set systems (over set V).

Proposition 2 allows for another (equivalent) definition of pivot over $\mathbb{F}_2$. Indeed, over $\mathbb{F}_2$, we have by Proposition 2, $\det(A*X)[Y] = \det A[X \oplus Y]$ for all $Y \subseteq V$ assuming $A*X$ is defined. Therefore, for $\mathcal{M}_{G*X}$ we have $D_{G*X} = \{Z \mid \det((G*X)[Z]) = 1\} = \{Z \mid \det(G[X \oplus Z]) = 1\} = \{X \oplus Y \mid \det(G[Y]) = 1\} = D_G * X$; see [1]. Hence $\mathcal{M}_G * X = \mathcal{M}_{G*X}$ is an alternative definition of the pivot operation over $\mathbb{F}_2$.

It turns out that $\mathcal{M}_G$ has a special structure, that of a *delta matroid*, allowing a specific exchange of elements between any two sets of D_G; see [1]. However, not every delta matroid M has a graph representation, i.e., M may not be of the form $\mathcal{M}_G$ for any graph G (a characterization of such representable delta matroids over $\mathbb{F}_2$ is given in [2]).

Example 3. Let G be the graph depicted in the upper-left corner of Fig. 1. We have $A(G) = \begin{array}{c|c} & \begin{array}{cccc} p & q & r & s \end{array} \\ \begin{array}{c} p \\ q \\ r \\ s \end{array} & \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \end{array}$. This corresponds to $\mathcal{M}_G = (\{p, q, r, s\}, D_G)$, where

$$D_G = \{\varnothing, \{q\}, \{p, q\}, \{p, r\}, \{p, s\}, \{q, s\}, \{r, s\}, \{p, q, s\}, \{p, q, r\}, \{q, r, s\}\}.$$

For example, $\{p, q\} \in D_G$ since $\det(G[\{p, q\}]) = \det \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} = 1$. Then $D_G * \{p, q\} = \{\varnothing, \{p\}, \{r\}, \{s\}, \{p, q\}, \{p, s\}, \{q, r\}, \{q, s\}, \{p, r, s\}, \{p, q, r, s\}\}$, and the corresponding graph is depicted on the top-right in the same Fig. 1. Equivalently, this graph is obtained from G by pivot on $\{p, q\}$. Also note that we have $D_G * \{p, s\} = D_G$, and therefore the pivot of G on $\{p, s\}$ obtains G again. The set inclusion diagrams of $\mathcal{M}_G$ and $\mathcal{M}_{G*\{p,q\}}$ are given in Fig. 4.

Graphs. The pivots $G * X$ where X is a minimal element of $\mathcal{M}_G \setminus \{\varnothing\}$ w.r.t. inclusion are called *elementary*. It is noted in [11] that an elementary pivot X corresponds to either a loop, $X = \{u\} \in E(G)$, or to an edge, $X = \{u, v\} \in E(G)$, where both vertices u and v are non-loops. Moreover, each $Y \in \mathcal{M}_G$ can be partitioned $Y = X_1 \cup \cdots \cup X_n$ such that $G * Y = G * (X_1 \oplus \cdots \oplus X_n) = (\cdots (G * X_1) \cdots * X_n)$ is a composition of disjoint elementary pivots. Consequently, a direct definition of the elementary pivots on graphs G is sufficient to define the (general) pivot operation.

The elementary pivot $G * \{u\}$ on a loop $\{u\}$ is called *local complementation*. It is the graph obtained from G by complementing the edges in the neighbourhood $N_G(u) = \{v \in V \mid \{u, v\} \in E(G), u \neq v\}$ of u in G: for each $v, w \in N_G(u)$, $\{v, w\} \in E(G)$ iff $\{v, w\} \notin E(G*\{u\})$, and $\{v\} \in E(G)$ iff $\{v\} \notin E(G*\{u\})$ (the case $v = w$). The other edges are left unchanged.

The elementary pivot $G * \{u, v\}$ on an edge $\{u, v\}$ between distinct non-loop vertices u and v is called *edge complementation*. For a vertex x consider its closed neighbourhood $N'_G(x) = N_G(x) \cup \{x\}$. The edge $\{u, v\}$ partitions the

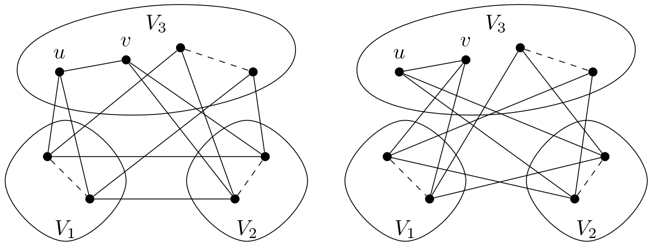

Fig. 2. Pivoting $\{u, v\}$ in a graph. Connection $\{x, y\}$ is toggled iff $x \in V_i$ and $y \in V_j$ with $i \neq j$. Note that u and v are connected to all vertices in V_3, these edges are omitted in the diagram. The operation does not affect edges adjacent to vertices outside the sets V_1, V_2, V_3, nor does it change any of the loops.

vertices of G connected to u or v into three sets $V_1 = N'_G(u) \setminus N'_G(v)$, $V_2 = N'_G(v) \setminus N'_G(u)$, $V_3 = N'_G(u) \cap N'_G(v)$. Note that $u, v \in V_3$.

The graph $G * \{u, v\}$ is constructed by "toggling" all edges between different V_i and V_j: for $\{x, y\}$ with $x \in V_i$ and $y \in V_j$ $(i \neq j)$: $\{x, y\} \in E(G)$ iff $\{x, y\} \notin E(G[\{u, v\}])$; see Fig. 2. The remaining edges remain unchanged. Note that, as a result of this operation, the neighbours of u and v are interchanged.

Example 4. The whole orbit of G of Example 3 under pivot is given in Fig. 1. It is obtained by iteratively applying elementary pivots to G. Note that $G * \{p, q\}$ *is* defined (top-right) but it is not an elementary pivot.

4. Dual pivots

In this section we introduce the dual pivot and show that it has some interesting properties.

First note that the next result follows directly from Eq. (1).

Lemma 5. *Let A be a $V \times V$-matrix (over some field) and let $X \subseteq V$ with $A[X]$ nonsingular. Then the eigenspaces of A and $A * X$ on value* 1 *are equal, i.e., $E_1(A) = E_1(A * X)$.*

Proof. We have $v \in E_1(A)$ iff $Av = v$ iff $(A * X)v = v$ iff $v \in E_1(A * X)$. $\square$

For a graph G, we denote $G + I$ to be the graph having adjacency matrix $A(G) + I$ where I is the identity matrix. Thus, $G + I$ is obtained from G by replacing each loop by a non-loop and vice versa.

Definition 6. Let G be a graph and let $X \subseteq V$ with $\det((G + I)[X]) = 1$. The *dual pivot* of G on X, denoted by $G\bar{*}X$, is $((G + I) * X) + I$.

Note that the condition $\det((G + I)[X]) = 1$ in the definition of dual pivot ensures that the expression $((G + I) * X) + I$ is defined. The dual pivot may be considered as the pivot operation *conjugated* by addition of the identity matrix I. As $I + I$ is the null matrix (over $\mathbb{F}_2$), we have, similar as for pivot, that dual pivot is an involution, and more generally $(G\bar{*}X)\bar{*}Y$, when defined, is equal to $G\bar{*}(X \oplus Y)$.

By Lemma 5, we have the following result.

Lemma 7. *Let G be a graph and let $X \subseteq V$ such that $G\bar{*}X$ is defined. Then* $\ker(G\bar{*}X) = \ker(G)$.

Proof. Note that $Ax = 0$ iff $(A + I)x = x$. Hence, $\ker(G) = E_1(G + I)$. Since $I + I$ is the null matrix (over $\mathbb{F}_2$), we have also $\ker(G + I) = E_1(G)$.

Therefore, we have $\ker(G\bar{*}X) = \ker(((G + I) * X) + I) = E_1((G + I) * X) = E_1(G + I)$, where we used Lemma 5 is the last equality. Finally, $E_1(G + I) = \ker(G)$ and therefore we obtain $\ker(G\bar{*}X) = \ker(G)$. $\square$

In particular, for the case $X = V$, we have that $\ker((G+I)^{-1} + I) = \ker(G)$ (the inverse is computed over $\mathbb{F}_2$) if the left-hand side is defined.

Remark 8. By Lemma 7 and Corollary 1 we have that the (column) matroids associated with G and $G\bar{*}X$ are equal. Note that here the matroids are obtained from the column vectors of the adjacency matrices of G and $G\bar{*}X$; this is not to be confused with graphic matroids which are obtained from the column vectors of the incidence matrices of graphs.

We call dual pivot $G\bar{*}X$ *elementary* if $*X$ is an elementary pivot for $G + I$. Equivalently, they are the dual pivots on X for which there is no non-empty $Y \subset X$ where $G\bar{*}Y$ is applicable. An elementary dual pivot $\bar{*}\{u\}$ is defined on a non-loop vertex u, and an elementary dual pivot $\bar{*}\{u, v\}$ is defined on an edge $\{u, v\}$ where both u and v have loops. This is the only difference between pivot and its dual: both the elementary dual pivot $\bar{*}\{u\}$ and the elementary pivot $*\{u\}$ have the same effect on the graph — both "take the complement" of the neighbourhood of u. Similarly, the effect of the elementary dual pivot $\bar{*}\{u, v\}$ and the elementary pivot $*\{u, v\}$ is the same, only the condition when they can be applied differs.

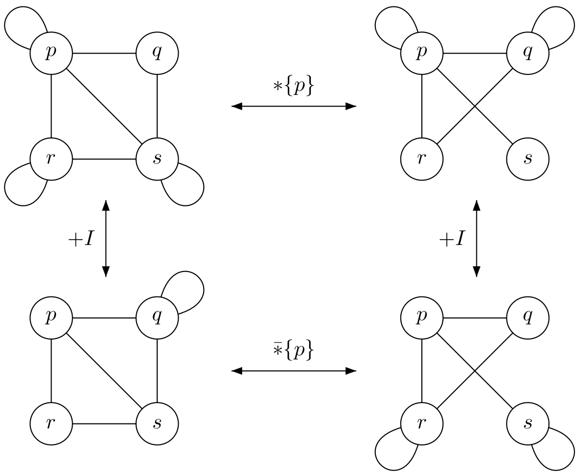

Fig. 3. Dual pivot of graph G from Example 3 (G is shown in the lower-left corner).

Note that the eigenspaces $E_0(G) = \ker(G)$ and $E_1(G)$ have a natural interpretation in graph terminology. For $X \subseteq V(G)$, $X \in E_0(G)$ iff every vertex in $V(G)$ is connected to an even number of vertices in X (loops do count). Also, $X \in E_1(G)$ iff every vertex in $V(G) \setminus X$ is connected to an even number of vertices in X and every vertex in X is connected to an odd number of vertices in X (again loops do count).

Example 9. Let G' be the graph depicted on the upper-left corner of Fig. 3. We have $A(G') = \begin{array}{c c} & \begin{array}{cccc} p & q & r & s \end{array} \\ \begin{array}{c} p \\ q \\ r \\ s \end{array} & \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \end{array}$. Note that $E_1(G') = \{\varnothing, \{p, r, s\}\}$ is of dimension 1. We can apply an elementary pivot over p on G'. The resulting graph $G' * \{p\}$ is depicted on the upper-right corner of Fig. 3, and we have $A(G' * \{p\}) = \begin{array}{c c} & \begin{array}{cccc} p & q & r & s \end{array} \\ \begin{array}{c} p \\ q \\ r \\ s \end{array} & \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \end{array}$. Note that the elements of $E_1(G')$ are precisely the eigenvectors (or eigensets) on 1 for $A(G' * \{p\})$; cf. Lemma 5. The graphs $G' + I$ (which is G in Example 3) and $G' * \{p\} + I$ are depicted in the lower-left and lower-right corner of Fig. 3, respectively. By definition of the dual pivot we have $G\bar{*}\{p\} = (G' + I)\bar{*}\{p\} = G' * \{p\} + I$.

It is a basic fact from linear algebra that elementary row operations retain the kernel of matrices. Lemma 7 suggests that the dual pivot may possibly be simulated by elementary row operations. We now show that this is indeed the case. Over $\mathbb{F}_2$ the elementary row operations are (1) row switching and (2) adding one row to another (row multiplication over $\mathbb{F}_2$ does not change the matrix). The elementary row operations corresponding to the dual pivot operation are easily deduced by restricting to elementary dual pivots. The dual pivot on a non-loop vertex u corresponds, in the adjacency matrix, to adding the row corresponding to u to each row corresponding to a vertex in the neighbourhood of u. Moreover, the dual pivot on edge $\{u, v\}$ (where both u and v have loops) corresponds to (1) adding the row corresponding to u to each row corresponding to a vertex in the neighbourhood of v except u, (2) adding the row corresponding to v to each row corresponding to a vertex in the neighbourhood of u except v, (3) switching the rows of u and v. Note that this procedure allows for another, equivalent, definition of the regular pivot: add I, apply the corresponding elementary row operations, and finally add I again.

Note that the dual pivot has the property that it transforms a symmetric matrix to another symmetric matrix with equal kernel. Applying elementary row operations however will in general not obtain symmetric matrices.

5. Maximal pivots

In Section 3 we recalled that the minimal elements of $\mathcal{M}_G$, corresponding to elementary pivots, form the building blocks of (general) pivots. In this section we show that the set of maximal elements of $\mathcal{M}_G$, corresponding to "maximal pivots", is invariant under dual pivot.

For $\mathcal{M}_G = (V, D_G)$, we define $\mathcal{F}_G = \max(D_G)$. Thus, for $X \subseteq V(G)$, $X \in \mathcal{F}_G$ iff $\det G[X] = 1$ while $\det G[Y] = 0$ for every $Y \supset X$.

Example 10. We continue Example 3. Let G be the graph on the lower-left corner of Fig. 3. Then from the set inclusion diagram of $\mathcal{M}_G$ in Fig. 4 we see that $\mathcal{F}_G = \{\{p, q, s\}, \{p, q, r\}, \{q, r, s\}\}$. Also we see from the figure that $\mathcal{F}_{G*\{p,q\}} = \{V\}$.

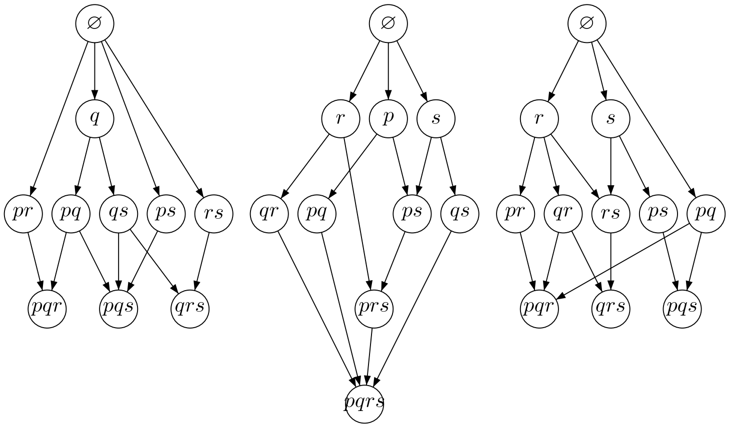

Fig. 4. Set inclusion diagram of $\mathcal{M}_G$, $\mathcal{M}_{G*\{p,q\}}$, and $\mathcal{M}_{G\bar{*}\{p\}}$ for G, $G*\{p,q\}$, and $G\bar{*}\{p\}$ as given in Examples 3 and 9.

Next we recall the Strong Principal Minor Theorem for (quasi-) symmetric matrices from [12] — it is stated here for graphs (i.e., symmetric matrices over $\mathbb{F}_2$). [1]

Proposition 11. *Let G be a graph such that $A(G)$ has rank r, and let $X \subseteq V(G)$ with $|X| = r$. Then X is independent for $A(G)$ iff* $\det G[X] = 1$.

Note that the independent sets X of cardinality equal to the rank are precisely the bases of a matrix A.

The following result is easy to see now from Proposition 11.

Lemma 12. *Let G be a graph such that $A(G)$ has rank r. Each element of $\mathcal{F}_G$ is of cardinality r.*

Proof. If there is an $X \in \mathcal{F}_G$ of cardinality $q > r$, then the columns of $A(G[X])$ are linearly independent, and thus so are the columns of $A(G)$ corresponding to X. This contradicts the rank of $A(G)$.

Finally, assume that there is an $X \in \mathcal{F}_G$ of cardinality $q < r$. Since the columns of $A(G[X])$ are linearly independent, so are the columns of $A(G)$ corresponding to X. Since $A(G)$ has rank r, X can be extended to a set X' with cardinality r. Hence by Proposition 11 $\det G[X'] = 1$ with $X' \supset X$ — a contradiction of $X \in \mathcal{F}_G$. □

Example 13. We continue Example 3. Let again G be the graph on the lower-left corner of Fig. 3. Then the elements $\mathcal{F}_G = \{\{p,q,s\},\{p,q,r\},\{q,r,s\}\}$ are all of cardinality 3 — the rank of $A(G)$. Moreover, $\mathcal{F}_{G*\{p,q\}} = \{V\}$ and $|V| = 4$ is equal to the rank of $G*\{p,q\}$.

Combining Proposition 11 and Lemma 12, we have the following result.

Corollary 14. *Let G be a graph, and let $X \subseteq V(G)$. Then X is a basis for $A(G)$ iff $X \in \mathcal{F}_G$.*

Equivalently, with $\mathcal{P}_G = (I, D)$ from Section 2, Corollary 14 states that $\max(I) = \mathcal{F}_G$.

By Corollaries 1 and 14 we have now the following.

Lemma 15. *Let G and G' be graphs. Then $\mathcal{F}_G = \mathcal{F}_{G'}$ iff* $\ker(G) = \ker(G')$.

Recall that Lemma 7 shows that the dual pivot retains the kernel. We may now conclude from Lemma 15 that also $\mathcal{F}_G$ is retained under dual pivot. It is the main result of this paper, and, as we will see in Section 7, has an important application.

Theorem 16. *Let G be a graph, and let $X \subseteq V$. Then $\mathcal{F}_G = \mathcal{F}_{G\bar{*}X}$ if the right-hand side is defined.*

In particular, the case $X = V$, we have $\mathcal{F}_{G+I} = \mathcal{F}_{G^{-1}+I}$ if G is invertible (over $\mathbb{F}_2$).

Let $\mathcal{O}_G = \{G\bar{*}X \mid X \subseteq V, \det(G+I)[X] = 1\}$ be the orbit of G under dual pivot, and note that $G \in \mathcal{O}_G$. By Theorem 16, if $G_1, G_2 \in \mathcal{O}_G$, then $\mathcal{F}_{G_1} = \mathcal{F}_{G_2}$. Note that the reverse implication does not hold: e.g. $\mathcal{O}_I = \{I\}$ and $\mathcal{F}_I = \{V\}$, while clearly there are many other graphs G with $\det G = 1$ (which means $\mathcal{F}_G = \{V\}$).

Example 17. We continue Example 9. Let again G be the graph on the lower-left corner of Fig. 3. Then $G\bar{*}\{p\}$ is depicted on the lower-right corner of Fig. 3. We have $\mathcal{F}_{G\bar{*}\{p\}} = \{\{p,q,s\},\{p,q,r\},\{q,r,s\}\}$, see Fig. 4, so indeed $\mathcal{F}_G = \mathcal{F}_{G\bar{*}\{p\}}$.

For symmetric $V \times V$-matrices A over $\mathbb{F}_2$, Theorem 16 states that if A can be partially inverted w.r.t. $Y \subseteq V$, where Y is maximal w.r.t. set inclusion, then this holds for every matrix obtained from A by dual pivot.

[1] Clearly, for a matrix A, $\det A[X] \neq 0$ implies that X is independent for A. The reverse implication is not valid in general.

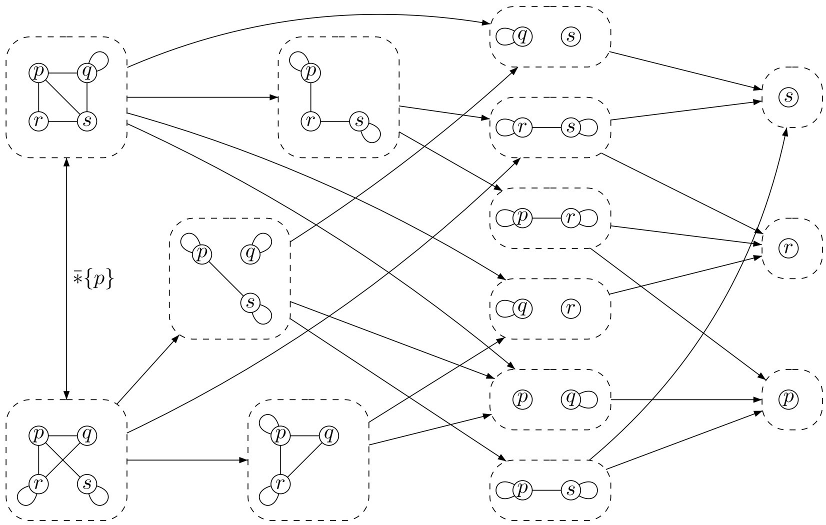

Fig. 5. Elementary contractions starting from G and $G\bar{*}\{p\}$.

6. Maximal contractions

For a graph G, we define the *contraction* of G on $X \subseteq V$ with $\det G[X] = 1$, denoted by $G * \backslash X$, to be the graph $(G * X) \setminus X$ — the pivot on X followed by the removal of the vertices of X. Equivalently, contraction is the Schur complement applied to graphs. A contraction of G on X is *maximal* if there is no $Y \supset X$ such that $\det G[Y] = 1$, hence if $X \in \mathcal{F}_G$. The graph obtained by a maximal contraction on X is a discrete graph G' (without loops). Indeed, if G' were to have a loop $e = \{u\}$ or an edge $e = \{u, v\}$ between two non-loop vertices, then, since $\det G[X \oplus e] = \det((G * X)[e]) = \det((G * \backslash X)[e]) = 1$, $X \oplus e \supset X$ would be a contradiction of the maximality of X. Moreover, by Lemma 12, the number of vertices of G' is equal to the nullity (dimension of the kernel, which equals the dimension of the matrix minus its rank) of G.

Remark 18. In fact, it is known that any Schur complement in a matrix A has the same nullity as A itself — it is a consequence of the Guttman rank additivity formula; see, e.g., [19, Section 6.0.1]. Therefore the nullity is invariant under contraction in general (not only maximal contraction).

By Theorem 16 we have the following.

Corollary 19. *The set of discrete graphs obtainable through contractions is equal for G and $G\bar{*}X$ for all $X \subseteq V$ with* $\det(G + I)[X] = 1$.

In this sense, all the elements of the orbit $\mathcal{O}_G$ have equal "behaviour" w.r.t. maximal contractions.

Example 20. We continue the example. Recall that, from Example 17, $\mathcal{F}_G = \mathcal{F}_{G\bar{*}\{p\}} = \{\{p, q, s\}, \{p, q, r\}, \{q, r, s\}\}$. The elementary contractions starting from G and $G\bar{*}\{p\}$ are given in Fig. 5. Notice that the maximal contractions of G and $G\bar{*}\{p\}$ obtain the same set of (discrete) graphs.

It is important to realize that while the *maximal* contractions (corresponding to $\mathcal{F}_G$) are the same for graphs G and $G\bar{*}X$, the whole set of contractions (corresponding to $\mathcal{M}_G$) may be spectacularly different. Indeed, e.g., in Example 20, the elementary pivots for G are $*\{q\}$, $*\{p, s\}$, $*\{p, r\}$, and $*\{r, s\}$, while the elementary pivots for $G\bar{*}\{p\}$ are $*\{r\}$, $*\{s\}$, and $*\{p, q\}$ (see Fig. 4).

7. Application: gene assembly

Gene assembly is a highly involved and parallel process occurring in one-cellular organisms called ciliates. During gene assembly a nucleus, called micronucleus (MIC), is transformed into another nucleus called macronucleus (MAC). Segments of the genes in the MAC occur in scrambled order in the MIC [9]. During gene assembly, recombination takes place to "sort" these gene segments in the MIC in the right orientation and order to obtain the MAC gene. The transformation of single genes from their MIC form to their MAC form is formally modelled, see [8,10,9], as both a string based model and a (almost equivalent) graph based model. It is observed in [3] that two of the three operations in the graph based model are exactly the two elementary principal pivot transform (PPT, or simply *pivot*) operations on the corresponding adjacency matrices considered over $\mathbb{F}_2$. The third operation simply removes isolated vertices.

Maximal contractions are especially important within the theory of gene assembly in ciliates — such a maximal sequence determines a complete transformation of the gene to its MAC form. We first recall the string rewriting system, and then recall the generalization to the graph rewriting system.

Let A be an arbitrary finite alphabet. The set of letters in a string u over A is denoted by $L(u)$. String u is called a *double occurrence string* if each $x \in L(u)$ occurs exactly twice in u. For example, $u = 41215425$ is a double occurrence string over $L(u) = \{1, \ldots, 5\}$. Let $\bar{A} = \{\bar{x} \mid x \in A\}$ with $A \cap \bar{A} = \varnothing$, and let $\tilde{A} = A \cup \bar{A}$. We use the "bar operator" to move from A to $\bar{A}$ and back from $\bar{A}$ to A. Hence, for $x \in \tilde{A}$, $\bar{\bar{x}} = x$. For a string $u = x_1x_2 \cdots x_n$ with $x_i \in A$, the *inverse* of u is the string $\bar{u} = \bar{x}_n\bar{x}_{n-1} \cdots \bar{x}_1$.

We define the morphism $\|\cdot\| : (\tilde{A})^* \to A^*$ as follows: for $x \in \tilde{A}$, $\|x\| = x$ if $x \in A$, and $\|x\| = \bar{x}$ if $x \in \bar{A}$, i.e., $\|x\|$ is the "unbarred" variant of x. Hence, e.g., $\|2\bar{5}\bar{3}\| = 253$. A *legal string* is a string $u \in (\tilde{A})^*$ where $\|u\|$ is a double occurrence string. We denote the empty string by λ.

Example 21. The string $u = qps\bar{q}rpsr$ over $\tilde{A}$ with $A = \{p, q, r, s\}$ is a legal string. As another example, the legal string $344567567898\bar{3}\bar{2}289$ over $\tilde{B}$ with $B = \{2, 3, \ldots, 9\}$ represents the micronuclear form of the gene corresponding to the actin protein in the stichotrichous ciliate *Sterkiella nova*; see [14,6].

It is postulated that gene assembly is performed by three types of elementary recombination operations, called loop, hairpin, and double-loop recombination on DNA; see [15]. These three recombination operations have been modelled as three types of string rewriting rules operating on legal strings [8,9] — together they form the string pointer reduction system. For all $x, y \in \tilde{A}$ with $\|x\| \neq \|y\|$ we define:

- the *string negative rule* for x by $\mathbf{snr}_x(u_1xxu_2) = u_1u_2$,
- the *string positive rule* for x by $\mathbf{spr}_x(u_1xu_2\bar{x}u_3) = u_1\bar{u}_2u_3$,
- the *string double rule* for x, y by $\mathbf{sdr}_{x,y}(u_1xu_2yu_3xu_4yu_5) = u_1u_4u_3u_2u_5$,

where $u_1, u_2, \ldots, u_5$ are arbitrary (possibly empty) strings over $\tilde{A}$.

Example 22. Let again $u = qps\bar{q}rpsr$ be a legal string. We have $\mathbf{spr}_q(u) = \bar{s}\bar{p}rpsr$. And moreover, $\mathbf{spr}_{\bar{r}}\ \mathbf{spr}_{\bar{p}}\ \mathbf{spr}_q(u) = \bar{s}\bar{s}$. Finally, $\mathbf{snr}_{\bar{s}}\ \mathbf{spr}_{\bar{r}}\ \mathbf{spr}_{\bar{p}}\ \mathbf{spr}_q(u) = \lambda$.

We now define a graph for a legal string representing whether or not intervals within the legal string "overlap". Let $u = x_1x_2 \cdots x_n$ be a legal string with $x_i \in \tilde{A}$ for $1 \leq i \leq n$. For letter $y \in L(\|u\|)$ let $1 \leq i < j \leq n$ be the positions of y in u, i.e., $\|x_i\| = \|x_j\| = y$. The *y-interval* of u, denoted by intv_y, is the substring $x_kx_{k+1} \cdots x_l$ where $k = i$ if $x_i = y$ and $k = i + 1$ if $x_i = \bar{y}$, and similarly, $l = j$ if $x_j = y$ and $l = j - 1$ if $x_j = \bar{y}$ (i.e., a border of the interval is included in the case of y and excluded in the case of $\bar{y}$). Now the *overlap graph* of u, denoted by $\mathcal{G}_u$, is the graph (V, E) with $V = L(\|u\|)$ and $E = \{\{x, y\} \mid x \text{ occurs exactly once in } \|\mathrm{intv}_y\|\}$. Note that E is well defined as x occurring exactly once in the y-interval of u is equivalent to y occurring exactly once in the x-interval of u. Note that we have a loop $\{x\} \in E$ iff both x and $\bar{x}$ occur in u. The overlap graph as defined here is an extension of the usual definition of overlap graph (also called circle graph) from simple graphs (without loops) to graphs (where loops are allowed). See [13, Section 7.4] for a brief overview of (simple) overlap graphs.

Example 23. The overlap graph $\mathcal{G}_u$ of $u = qps\bar{q}rpsr$ is exactly the graph G of Example 3.

It is shown in [8,10], see also [9], that the string rules $\mathbf{snr}_x$, $\mathbf{spr}_x$, and $\mathbf{sdr}_{x,y}$ on legal strings u can be simulated as graph rules $\mathbf{gnr}_x$, $\mathbf{gpr}_x$, and $\mathbf{gdr}_{x,y}$ on overlap graphs $\mathcal{G}_u$ in the sense that $\mathcal{G}_{\mathbf{spr}_x(u)} = \mathbf{gpr}_x(\mathcal{G}_u)$, where the left-hand side is defined iff the right-hand side is defined, and similarly for $\mathbf{gdr}_{x,y}$ and $\mathbf{gnr}_x$.[2] It was shown in [3] that $\mathbf{gpr}_x$ and $\mathbf{gdr}_{x,y}$ are exactly the two types of contractions of elementary pivots $* \setminus \{x\}$ and $* \setminus \{x, y\}$ on a loop $\{x\}$ and an edge $\{x, y\}$ without loops, respectively. The $\mathbf{gnr}_x$ rule is the removal of isolated vertex x.

Example 24. The sequence $\mathbf{spr}_{\bar{r}}\ \mathbf{spr}_{\bar{p}}\ \mathbf{spr}_q$ applicable to u given in Example 22 corresponds to a maximal contraction of graph $G = \mathcal{G}_u$ of Example 3 as can be seen in Fig. 5.

Within the theory of gene assembly one is interested in maximal recombination strategies of a gene. These strategies correspond to maximal contractions of a graph G (hence decomposable into a sequence φ_1 of contractions of elementary pivots $\mathbf{gpr}$ and $\mathbf{gdr}$ applicable to (defined on) G) followed by a sequence φ_2 of $\mathbf{gnr}$ rules, removing isolated vertices, until the empty graph is obtained. Here we call these sequences $\varphi = \varphi_2\varphi_1$ of graph rules *complete contractions*. If we define the set of vertices v of φ used in $\mathbf{gnr}_v$ rules by $\mathrm{gnrdom}(\varphi)$, then the following result holds by Corollary 19.

[2] There is an exception for $\mathbf{gnr}_x$: although $\mathcal{G}_{\mathbf{snr}_x(u)} = \mathbf{gnr}_x(\mathcal{G}_u)$ holds if the left-hand side is defined, there are cases where the right-hand side is defined (x is an isolated vertex in $\mathcal{G}_u$) while the left-hand side is not defined (u does not have substring xx). This is why the string and graph models are "almost" equivalent. This difference in models is not relevant for our purposes.

Theorem 25. *Let $G_1, G_2 \in \mathcal{O}_G$ for some graph G, and let φ be a complete contraction of graph G_1. Then there is a complete contraction φ' of G_2 such that* $\mathrm{gnrdom}(\varphi) = \mathrm{gnrdom}(\varphi')$.

Hence, Theorem 25 shows that all the elements of $\mathcal{O}_G$, for any graph G, have the same behaviour w.r.t. the applicability of the rule $\mathbf{gnr}_x$.

A similar result as Theorem 25 was shown for the string rewriting model; see [4, Theorem 34].[3] It should be stressed however that Theorem 25 is a real generalization of the result in [4] as not every graph has a string representation (i.e., not every graph is an overlap graph), and moreover it is obtained in a very different way: here the result is obtained using techniques from linear algebra.

8. Discussion

We introduced the concept of dual pivot and have shown that it has interesting properties: it has the same effect as the (regular) pivot *and* can be simulated by elementary row operations — consequently it keeps the kernel invariant. The dual pivot in this way allows for an alternative definition of the (regular) pivot operation. Furthermore, we have shown that two graphs have equal kernel precisely when they have the same set of maximal pivots. From this it follows that the set of maximal pivots is invariant under dual pivot.

This main result is motivated by the theory of gene assembly in ciliates in which maximal contractions correspond to complete transformations of a gene to its macronuclear form. However, as applying a maximal pivot corresponds to calculating a maximal partial inverse of the matrix, the result is also interesting from a purely theoretical point of view.

Acknowledgements

We thank Lorenzo Traldi and the two anonymous referees for their valuable comments on the paper. R.B. is supported by the Netherlands Organization for Scientific Research (NWO), project "Annotated graph mining".

References

[1] A. Bouchet, Representability of Δ-matroids, in: Proc. 6th Hungarian Colloquium of Combinatorics, Colloquia Mathematica Societatis János Bolyai, vol. 52, North-Holland, 1987, pp. 167–182.
[2] A. Bouchet, A. Duchamp, Representability of Δ-matroids over $GF(2)$, Linear Algebra and its Applications 146 (1991) 67–78.
[3] R. Brijder, T. Harju, H.J. Hoogeboom, Pivots, determinants, and perfect matchings of graphs (2008) (submitted for publication) [arXiv:0811.3500].
[4] R. Brijder, H.J. Hoogeboom, The fibers and range of reduction graphs in ciliates, Acta Informatica 45 (2008) 383–402.
[5] R. Brijder, H.J. Hoogeboom, M. Muskulus, Strategies of loop recombination in ciliates, Discrete Applied Mathematics 156 (2008) 1736–1753.
[6] A.R.O. Cavalcanti, T.H. Clarke, L.F. Landweber, MDS_IES_DB: a database of macronuclear and micronuclear genes in spirotrichous ciliates, Nucleic Acids Research 33 (2005) D396–D398.
[7] R.W. Cottle, J.-S. Pang, R.E. Stone, The Linear Complementarity Problem, Academic Press, San Diego, 1992.
[8] A. Ehrenfeucht, T. Harju, I. Petre, D.M. Prescott, G. Rozenberg, Formal systems for gene assembly in ciliates, Theoretical Computer Science 292 (2003) 199–219.
[9] A. Ehrenfeucht, T. Harju, I. Petre, D.M. Prescott, G. Rozenberg, Computation in Living Cells — Gene Assembly in Ciliates, Springer Verlag, 2004.
[10] A. Ehrenfeucht, I. Petre, D.M. Prescott, G. Rozenberg, String and graph reduction systems for gene assembly in ciliates, Mathematical Structures in Computer Science 12 (2002) 113–134.
[11] J.F. Geelen, A generalization of Tutte's characterization of totally unimodular matrices, Journal of Combinatorial Theory, Series B 70 (1997) 101–117.
[12] V. Kodiyalam, T.Y. Lam, R.G. Swan, Determinantal ideals, Pfaffian ideals, and the principal minor theorem, in: Noncommutative Rings, Group Rings, Diagram Algebras and Their Applications, American Mathematical Society, 2008, pp. 35–60.
[13] T.A. McKee, F.R. McMorris, Topics in Intersection Graph Theory, Society for Industrial Mathematics, 1999.
[14] D.M. Prescott, M. DuBois, Internal eliminated segments (IESs) of oxytrichidae, Journal of Eukaryotic Microbiology 43 (1996) 432–441.
[15] D.M. Prescott, A. Ehrenfeucht, G. Rozenberg, Molecular operations for DNA processing in hypotrichous ciliates, European Journal of Protistology 37 (2001) 241–260.
[16] J. Schur, Über Potenzreihen, die im Innern des Einheitskreises beschränkt sind, Journal für die reine und angewandte Mathematik 147 (1917) 205–232.
[17] M.J. Tsatsomeros, Principal pivot transforms: properties and applications, Linear Algebra and its Applications 307 (1–3) (2000) 151–165.
[18] A.W. Tucker, A combinatorial equivalence of matrices, in: Combinatorial Analysis, Proceedings of Symposia in Applied Mathematics, vol. X, American Mathematical Society, 1960, pp. 129–140.
[19] F. Zhang, The Schur Complement and Its Applications, Springer, 1992.

[3] This result states that two legal strings equivalent modulo "dual" string rules have the same reduction graph (up to isomorphism). It then follows from [5, Theorem 44] that these strings have complete contractions with equal snrdom, the string equivalent of gnrdom.