

## Graph Polynomials Motivated by Gene Rearrangements in Ciliates

Peer-reviewed author version

BRIJDER, Robert & Hoogeboom, Hendrik Jan (2014) Graph Polynomials Motivated by Gene Rearrangements in Ciliates. In: Beckmann, Arnold; Csuhaj-Varjú, Erzsébet; Meer, Klaus (Ed.). Language, Life, Limits, p. 63-72.

DOI: 10.1007/978-3-319-08019-2\_7

Handle: <http://hdl.handle.net/1942/18453>

# Graph Polynomials Motivated by Gene Rearrangements in Ciliates

Robert Brijder<sup>1\*</sup> and Hendrik Jan Hoogeboom<sup>2</sup>

<sup>1</sup> Hasselt University and Transnational University of Limburg, Belgium

robert.brijder@uhasselt.be

<sup>2</sup> Leiden Institute of Advanced Computer Science,

Leiden University, The Netherlands

h.j.hoogeboom@liacs.leidenuniv.nl

**Abstract.** Gene rearrangements within the process of gene assembly in ciliates can be represented using a 4-regular graph. Based on this observation, Burns et al. [Discrete Appl. Math., 2013] propose a graph polynomial abstracting basic features of the assembly process, like the number of segments excised. We show that this *assembly polynomial* is essentially (i) a single variable case of the *transition polynomial* by Jaeger and (ii) a special case of the *bracket polynomial* introduced for simple graphs by Traldi and Zulli.

## 1 Introduction

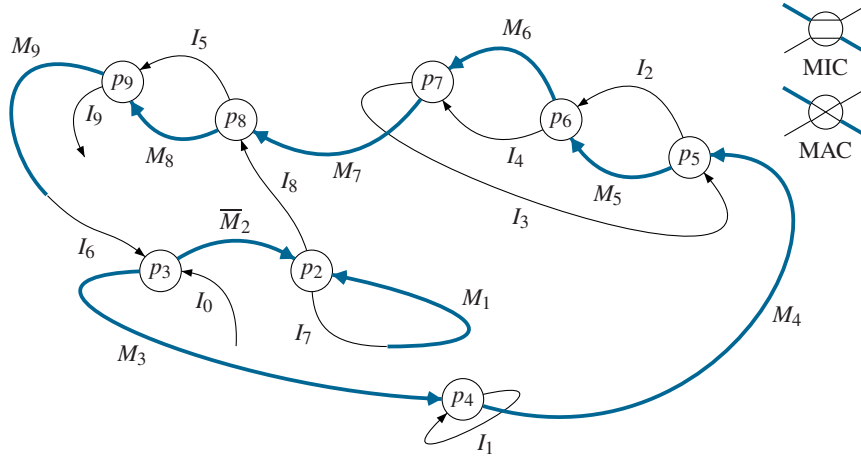
Ciliates are an ancient group of unicellular organisms. They have the remarkable property that their DNA is stored in two vastly different types of nuclei. During conjugation a germline nucleus called the *micronucleus* (MIC) is transformed into a somatic nucleus called the *macronucleus* (MAC). In this way, each MIC gene is transformed into its corresponding MAC gene, in a process that we call *gene assembly*. Various formal models for this gene transformation process are presented in [3].

One of these formal models is string based, with letters representing “pointers” (special DNA sequences in a MIC gene) together with their relative orientation in the corresponding MAC gene [10]. The model postulates that three operations called *loop excision*, *hairpin recombination*, and *double loop recombination* accomplish the transformation of a MIC gene to its corresponding MAC gene. This model has been significantly generalized using the notion of circle graph, see, e.g., [10,3].

This string-based formal model can be very naturally fitted within the well-developed theory of transformations of Eulerian circuits in 4-regular graphs [4]. An example using the Actin I gene of *Sterkiella nova*, taken from Prescott [18], is recalled below. It is known that these Eulerian circuit transformations may in turn be viewed both as a special case of a matrix operation called *principal pivot transform* (see [22]) and as a set system operation called *twist*. It turns out that the interplay of the successive operations can be much better understood in these more general settings compared to the string or graph settings. These generalizations for gene assembly are outlined in [4].

---

\* R.B. is a postdoctoral fellow of the Research Foundation – Flanders (FWO).



**Fig. 1.** Actin I gene of *Sterkiella nova*. Schematic diagram, based on [18].

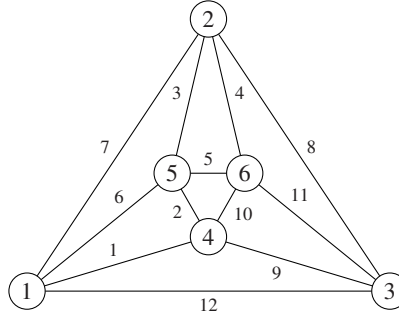
Here we recall the 4-regular graph representation that represents both the MIC and MAC form of a gene. We recall how the MIC and MAC form of a gene are two different sets of circuits in a 4-regular graph, and may be obtained from one another by making different decisions where to continue a path at each vertex. Possible intermediate results of this recombination can also be read from this 4-regular graph. Based on this observation, Burns et al. [7] have proposed the *assembly polynomial*, a graph polynomial intended to abstract basic features of the assembly process, like the number of molecules excised during this process.

In this paper we show that the assembly polynomial is closely related to the following graph polynomials: (i) the *transition polynomial* by Jaeger [15,13] and (ii) the *bracket polynomial* introduced for simple graphs by Traldi and Zulli [21]. We show how notions and results related to these polynomials may be carried over to the assembly polynomial. We note that the transfer of notions and results between other (related) graph polynomials have been accomplished, as can be seen in [11,12].

### Gene Assembly as 4-Regular Graphs

Prescott and Greslin [19] have unraveled the different representations of genetic material in the MIC and MAC form of genes in ciliates. We illustrate this using the example of the Actin I gene of *Sterkiella nova*, the presentation of which is based on [18]. In MIC form this gene consists of macronuclear destined sequences (MDSs for short), which end up in the MAC form of this gene, that are scrambled (their order may be permuted and some MDSs may be inverted). The segments in between the MDSs are called internal eliminated sequences (IESs for short) that are excised and do not appear in the MAC form of this gene.

The MIC of the Actin I gene can be written as  $I_0 M_3 I_1 M_4 I_2 M_6 I_3 M_5 I_4 M_7 I_5 M_9 I_6 \bar{M}_2 I_7 M_1 I_8 M_8 I_9$ , where  $M_i$  and  $I_j$  represent MDSs and IESs, respectively. The final MAC can be written as  $M_1 M_2 \dots M_8 M_9$ . Both the MIC and MAC form of this gene can



**Fig. 2.** Graph  $G_w$  defined by string  $w = 145265123463$ , see Example 1. The edges trace the Eulerian circuit  $C_w$  as defined by  $w$  according to the edge numbers given.

be read from the 4-regular graph  $G$  in Fig. 1. The vertices of  $G$  represent the *pointers*, i.e., the places where recombination takes place, and the edges represent the MDSs and IESs. Each vertex has two incoming and two outgoing edges. The MIC and MAC form both trace paths in the diagram. For the MIC form successive incoming and outgoing edges are chosen such that the two visits of a vertex do not cross, whereas for the MAC form the successive visits are connected in a crossing fashion (see top right illustration).

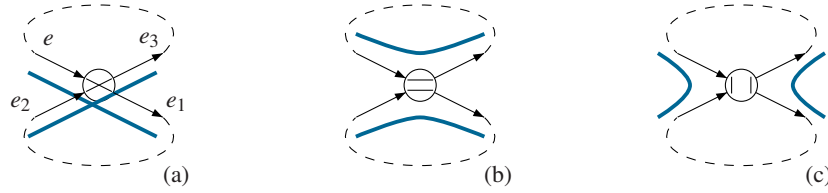
Without loss of generality, in the abstract context considered in the next sections, the initial edge and final edge are joined to form a single edge, in order to obtain an actual 4-regular graph.

## 2 Eulerian Circuits in 4-regular Graphs

The example in the introduction illustrates that 4-regular graphs are one of the key tools for describing the transformation of a MIC gene into its MAC gene. The MIC gene corresponds to an Eulerian circuit, while the MAC gene corresponds to a number of circuits (each edge belongs to one such circuit), such that one of those contains the MDSs in the proper order and orientation, while the other circuits contain only IESs that are excised during the gene assembly process. We abstract from genes and describe circuits in 4-regular graphs over an abstract alphabet (the elements of which, in effect, denote pointers).

A *double-occurrence string* over an alphabet  $V$  contains each letter from  $V$  exactly twice. Each double-occurrence string describes a 4-regular graph  $G_w$  together with an Eulerian circuit  $C_w$  for  $G_w$ , choosing vertex set  $V$  and edges that follow consecutive letters in  $w$  (in a cyclic fashion).

*Example 1.* Let  $V = \{1, \dots, 6\}$ . For  $w = 145265123463$ , the 4-regular graph  $G_w$  is given in Fig. 2, where the edges are numbered according to the Eulerian circuit  $C_w$ . Thus, first introduce a vertex for each element in  $V$ . Then continue by reading  $w$  and adding (undirected) edges between consecutive letters (in a cyclic fashion):  $1 \overset{1}{-} 4 \overset{2}{-} 5 \overset{3}{-} 2 \overset{4}{-} 6 \overset{5}{-} 5 \overset{6}{-} 1 \overset{7}{-} 2$ , etc.  $\square$



**Fig. 3.** Three ways to connect pairs of edges in a 4-regular graph relative to an (oriented) Eulerian circuit: **(a)**  $\xrightarrow{e} v \xrightarrow{e_1}$  following the circuit, **(b)**  $\xrightarrow{e} v \xrightarrow{e_3}$  in an orientation-consistent way, and **(c)**  $\xrightarrow{e} v \xrightarrow{e_2}$  in an orientation-inconsistent way.

We discuss here the basic theory of splitting and joining Eulerian circuits in 4-regular graphs, see [16] and also [14, Chapter 17]. Let  $G$  be a undirected 4-regular graph and  $C$  an Eulerian circuit of  $G$ . We assume here that circuits are not oriented.

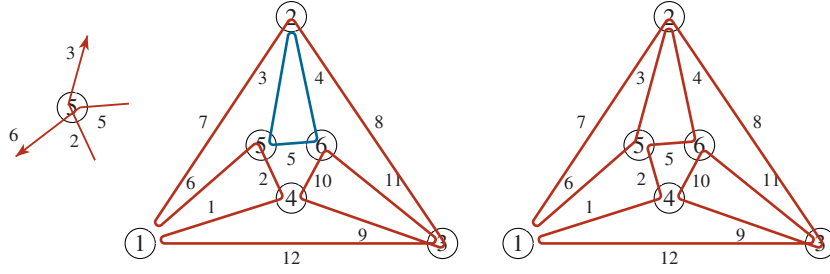
Consider a *circuit partition*  $P$  of  $G$ , i.e., a set of circuits of  $G$  that together contain all the edges exactly once. We can describe  $P$  relative to  $C$  by considering at each vertex how the circuits in  $P$  follow the edges relative to  $C$ . In fact, when during a walk of  $P$ ,  $P$  enters a vertex  $v$ , then  $P$  can leave in one of exactly three directions.<sup>1</sup> They can be classified as follows, cf. Fig. 3. Let  $\dots \xrightarrow{e} v \xrightarrow{e_1} \dots \xrightarrow{e_2} v \xrightarrow{e_3} \dots$  be an arbitrary orientation of  $C$ . (i) If, by entering  $v$  via  $e$  in  $P$ ,  $P$  leaves  $v$  via  $e_1$ , then we say that  $P$  follows  $C$  at  $v$ . (ii) If  $P$  leaves  $v$  via  $e_3$ , then we say that  $P$  is *orientation-consistent* with  $C$  at  $v$ , and finally (iii) if  $P$  leaves  $v$  via  $e_2$ , then we say that  $P$  is *orientation-inconsistent* with  $C$  at  $v$ . In [15], the first two are called *coherent* and the last one is called *anticoherent*. Moreover, in [7] the last two are called (parallel) *p-smoothing* and *n-smoothing*, respectively.

We let  $D(C, P) = (D_1, D_2, D_3)$  be the ordered partition (we allow some  $D_i$  to be empty) of  $V$  such that the vertices of  $P$  that (i) follow  $C$  are in  $D_1$ , (ii) are orientation-consistent with  $C$  are in  $D_2$ , and (iii) are orientation-inconsistent with  $C$  are in  $D_3$ . (We return to this notion later, see Example 6 and Theorem 5.)

A *transition* at a vertex  $v$  is a partition in (unordered) pairs of the edges incident to  $v$ . A *transition system* of  $G$  is a set of transitions, one for each vertex of  $G$ . Note that a circuit partition uniquely determines a transition system. Indeed, if a circuit of  $P$  visits vertex  $v$ , entering and leaving at  $v$  via a pair of edges, then another visit of  $v$  (which may occur either in the same circuit or in another circuit of  $P$ ) uses the remaining pair of edges. In the same way, a transition system uniquely determines a circuit partition.

**Remark 2.** In the process of gene assembly in ciliates, the MIC form of a gene traces an Eulerian circuit, while its corresponding MAC form has a different transition at each vertex (i.e., pointer). For each vertex  $v$ , this transition is uniquely determined by the relative orientation in the MIC form of the two MDSs with edges incident to  $v$ . When the MDSs have the same orientation, the orientation-consistent transition is taken; otherwise, the orientation-inconsistent transition is taken. In this way, an intermediate result

<sup>1</sup> In case  $G$  has loops one has to consider “half edges” to obtain three directions, but these technicalities are left to the reader.



**Fig. 4.** Changing the Eulerian circuit of Fig. 2 at vertex 5 in orientation-consistent and orientation-inconsistent ways, respectively. The transition that follows  $C_w$  at vertex 5 is indicated on the left.

of the gene assembly process either follows the MIC or the MAC transition at each vertex. Thus in modelling the rearrangements in gene assembly in ciliates one has to keep track of the relative orientation of the MDS-segments.

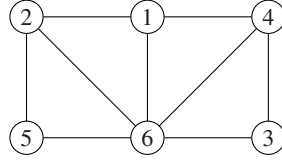
In Fig. 3, we see that changing an Eulerian circuit at a single vertex by choosing the transition in an orientation-consistent manner, splits the circuit into two circuits. Choosing the orientation-inconsistent transition instead “inverts” part of the circuit.

*Example 3.* The Eulerian circuit  $C_w$  is given by  $w = 14\mathbf{5}26\mathbf{5}123463$ , cf. Example 1. By changing  $C_w$  at vertex 5 in an orientation-consistent way, we have the transitions  $\frac{2}{\mathbf{5}}\mathbf{5}\frac{6}{\mathbf{3}}$  and  $\frac{5}{\mathbf{3}}\mathbf{5}\frac{2}{\mathbf{6}}$ , so we obtain two circuits  $14\mathbf{5}123463$  and  $\mathbf{5}26$ . By recombining in an orientation-inconsistent way we have the transitions  $\frac{2}{\mathbf{5}}\mathbf{5}\frac{5}{\mathbf{6}}$  and  $\frac{3}{\mathbf{5}}\mathbf{5}\frac{6}{\mathbf{2}}$ , so we obtain a single circuit  $14\mathbf{5}62\mathbf{5}123463$  with segment inverted, see Fig. 4.  $\square$

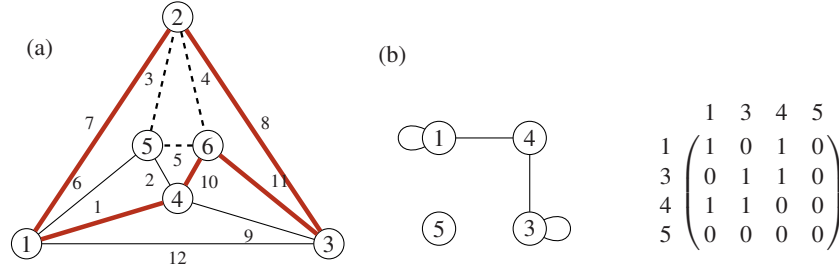
We noted that changing the transition of vertex  $v$  in an Eulerian circuit  $C$  to an orientation-consistent way splits the circuit in two. However, when we in parallel change transitions in an Eulerian circuit at distinct vertices  $u$  and  $v$ , both in an orientation-consistent way, then again an Eulerian circuit is formed, provided that  $u$  and  $v$  are “interlaced” in  $C$ , i.e., they occur in the order  $\dots u \dots v \dots u \dots v \dots$  (again, strings are regarded cyclic). In that case, two segments between the occurrences  $u$  and  $v$  are swapped. This is an important observation in the context of intra-molecular models of gene assembly. The interlacement of pairs of vertices is captured in an *interlace graph* (also called a *circle graph* as it can be defined using a collection of chords in a circle with edges denoting their intersection, see, e.g., [14, Chapter 17]).

The *interlace matrix*  $I(C)$  of  $C$  is the  $V \times V$ -matrix (i.e., the rows and columns are not ordered, but indexed by  $V$ ) over  $GF(2)$  that has 1 at position  $(u, v)$  if vertices  $u$  and  $v$  are interlaced in  $C$ , i.e., occur in the order  $\dots u \dots v \dots u \dots v \dots$ , and 0 otherwise. Since  $I(C)$  is symmetric,  $I(C)$  may be viewed as the adjacency matrix of a simple graph  $G$ , called the *interlace graph* of  $C$ . We are sloppy, and also use  $I(C)$  as a notation for  $G$ .

*Example 4.* For  $C_w$  defined by  $w = 145265123463$ , the interlace graph  $I(C_w)$  is given in Fig. 5.  $\square$



**Fig. 5.** The interlace graph  $I(C_w)$  for  $w = 145265123463$ , cf. Example 4.



**Fig. 6.** Cf. Example 6. (a) Partition  $P$  of the edges of the 4-regular graph  $G_w$  into three closed walks; edge numbers according to  $C_w$ . (b) Both the graphical and matrix representation of the graph  $(I(C_w) + \Delta(\{1,3\})) \setminus \{2,6\}$ .

Cohn and Lempel [8] proved a surprisingly simple formula for the number of circuits that result from an Eulerian circuit  $C$  in a 4-regular graph by changing  $C$  in an orientation-consistent way at several vertices. The formula is in terms of the nullity (dimension of the null space) of the interlace matrix. A more general treatment of this result is given by Traldi [20], allowing one to change transitions in both orientation-consistent and orientation-inconsistent ways. Above we have borrowed the terminology of [20].

Traldi has shown that the number of circuits in  $P$  can be elegantly expressed in terms of the interlace matrix  $I(C)$  and the partition  $D(C, P)$ . Let  $A$  be a  $V \times V$ -matrix. If  $D \subseteq V$ , then  $A \setminus D$  denotes the restriction of  $A$  to the principal submatrix of  $A$  indexed by  $V \setminus D$ , and  $\Delta(D)$  is the  $V \times V$ -matrix that has 1 only on diagonal elements  $(u, u)$  with  $u \in D$ , and 0 elsewhere. Also,  $\nu$  and  $\rho$  denote the nullity and rank of matrices computed over  $\text{GF}(2)$ , respectively.

**Theorem 5 (Traldi-Cohn-Lempel).** *Let  $G$  be an undirected, connected 4-regular graph with Euler cycle  $C$ , and let  $P$  be a circuit partition of  $E(G)$ , such that  $D(C, P) = (D_1, D_2, D_3)$ . Then  $|P| = \nu((I(C) + \Delta(D_3)) \setminus D_1) + 1$ .*

*Example 6.* We continue the running example. Fig. 6(a) shows a circuit partition  $P$  of  $G_w$  in three parts, with circuits 14632, 1345, and 265. At vertex 5, the circuits trace the transitions  $\frac{2}{5} \frac{5}{6}$  and  $\frac{5}{5} \frac{3}{3}$ , which is orientation-consistent w.r.t. the circuit  $C_w$ . Hence 5 belongs to the second component of the partition  $D(C_w, P)$ . Considering all vertices, we find  $D(C_w, P) = (\{2, 6\}, \{4, 5\}, \{1, 3\})$ . Graph  $(I(C_w) + \Delta(\{1, 3\})) \setminus \{2, 6\}$  is obtained from the interlace graph  $I(C_w)$  by deleting vertices 2, 6 and adding loops

to 1, 3. It is given in Fig. 6(b). The corresponding matrix has (dimension  $4 \times 4$ , rank 2, and) nullity 2. Indeed  $|P| = 2 + 1$  satisfies Theorem 5.  $\square$

### 3 Graph Polynomials Motivated by Gene Rearrangement

One of the first graph polynomials is defined by Martin [17] for 4-regular graphs. Given such a graph  $G$ , the coefficient of  $y^k$  in its polynomial  $M_G(y)$  equals the number of circuit partitions  $P$  of  $G$  with  $|P| = k$ . Currently, there is an impressive body of results on graph polynomials, see, e.g., the overview papers by Ellis-Monaghan and Merino [11,12]. Typical topics that are studied are algebraic and combinatorial in nature, and include recursive formulations of the polynomials, and the interpretation of evaluations at specific values. Generalizations have been obtained for structures like knots and matroids [15]. Also, a multimatroid polynomial is proposed as a unified framework to several polynomials for graphs and matroids [5].

The study of gene rearrangements has motivated the introduction of new polynomials, most notably the *interlace* [2,1] and *assembly* [7] polynomials, as, e.g., a feature that could measure and compare the complexity of the rearrangement process. Both these polynomials fit in the corpus of existing graph polynomials, and thus techniques and results can be carried over. For the interlace polynomial it has been shown that it is tied to the well-known Tutte polynomial [1,5]. Here we discuss how the assembly polynomial is related to other known polynomials.

#### The Transition Polynomial and its Relatives

It was observed by Jaeger [15] that several polynomials for 4-regular graphs are special cases of the *transition polynomial*, which is a multivariate/weighted polynomial. Similar as the Martin polynomial, the transition polynomial counts circuit partitions  $P$  w.r.t. an arbitrary Eulerian circuit  $C$ . However, the circuit partitions  $P$  have weights in the transition polynomial that depend on  $D(C, P)$ . Unfortunately, it seems that [15] is not widely distributed.

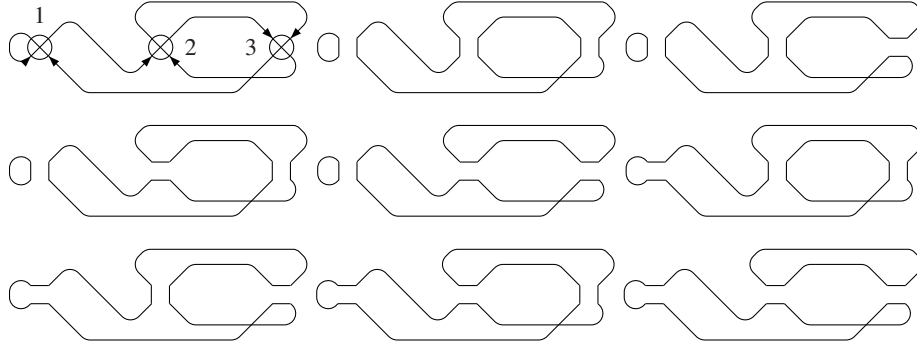
We show that the assembly polynomial is closely related to the transition polynomial, essentially by embedding its second variable into the weights. Secondly, we observe a less obvious relation to the bracket polynomial for graphs in terms of the nullity of the adjacency matrix of the circle graph using Theorem 5.

Burns et al. [7, Section 6] (see also [9]) define the assembly polynomial for a 4-regular graph  $G_w$  together with an Eulerian circuit  $C_w$  belonging to a double-occurrence string  $w$  over alphabet  $V$ . The *assembly polynomial* of  $G_w$  w.r.t. Eulerian circuit  $C_w$  is

$$S(G_w)(p, t) = \sum_s p^{\pi(s)} t^{c(s)-1},$$

where the sum is taken over all  $2^{|V|}$  transition systems  $s$  that differ at each vertex from the transition system corresponding to  $C_w$  (a transition that differs from the transition of  $C_w$  is called a *smoothing* in [7]),  $\pi(s)$  equals the number of orientation-consistent transitions of  $s$  w.r.t.  $C_w$  (called *p-smoothings* in [7]), and  $c(s)$  equals the number of circuits in the circuit partition corresponding to  $s$ .





**Fig. 7.** Construction of  $G_w$  for  $w = 1\ 1\ 2\ 3\ 2\ 3$ , and the eight smoothings that determine its assembly polynomial, see Example 7.

*Example 7.* Consider the double-occurrence string  $w = 1\ 1\ 2\ 3\ 2\ 3$ . In Fig. 7 we show that its corresponding Eulerian circuit  $C_w$  in the 4-regular graph  $G_w$  (top-left) and its eight possible smoothings (at each vertex, vertical connections correspond to a consistent-orientation change, i.e., a p-smoothing). Note that  $G_w$  has both a loop and parallel edges. The assembly polynomial equals  $S(G_w) = p^3t + 2p^2t + p^2 + pt^2 + 2p + t$  (as can be verified using the Assembly Words online tool [6]).

We discuss the transition polynomial from [15]. Let  $P(G)$  be the set of all transition systems of  $G$ . Note that  $|P(G)| = 3^{|V|}$ , where  $V$  is the set of vertices of  $G$ . A weight function  $W$  assigns a weight to each of the three possible transitions at each vertex. The weight  $\omega(s)$  of the transition system  $s$  is the product of the weights of  $s$  at each vertex. The (weighted) *transition polynomial* of  $G$  is now defined as

$$q(G, W; x) = \sum_{s \in P(G)} \omega(s) x^{c(s)-1}.$$

We obtain the assembly polynomial as a special case of the transition polynomial by fixing the weights  $W(C)$  relative to the Eulerian circuit  $C$ . Transitions that follow  $C$  have weight 0 (so are not counted at all), orientation-consistent and orientation-inconsistent transitions have weight  $p$  and 1 respectively. We have  $\omega(s) = p^{\pi(s)}$  if  $s$  contains no transitions that follow  $C$ , and  $\omega(s) = 0$  otherwise. Hence  $S(G_w)(p, t) = q(G_w, W(C_w); t)$ .

The interlace polynomial on the other hand, is equal to a suitable generalisation of the transition polynomial where zero weight is assigned to the orientation-inconsistent transitions.

Various results are known for the transition polynomial. For example, let  $P$  be a circuit partition of  $G$  and  $W(P)$  be such that the transitions belonging to  $P$  have weight 0, and the other transitions have weight 1. Then  $q(G, W(P); -2) = (-1)^{|V|}(-2)^{c(P)-1}$ , see [15, Proposition 11]. Consequently,  $S(G_w)(1, -2) = (-1)^{|V|}$  since  $P$  is the Euler circuit  $C_w$  and  $c(C_w) = 1$ .

We now move to a second interpretation of the assembly polynomial. Consider a transition system  $s$  that never follows circuit  $C$ , as is relevant in this context. It determines a set of circuits  $P_s$ , and the number of circuits  $c(s) = |P_s|$  is given by the Traldi-Cohn-Lempel formula, and equals  $v(I(C_w) + \Delta(D_3)) + 1$ , where  $D(C, P_s) = (\emptyset, D_2, D_3)$ . Recall that  $s$  determines  $D(C, P_s)$  but also vice versa. Thus we reformulate the assembly polynomial of  $G_w$  w.r.t.  $C_w$  replacing summation over transition systems  $s$  by summation over partitions  $(\emptyset, D_2, D_3)$ . The notation  $\dot{\cup}$  represents disjoint set union.

$$S(G_w)(p, t) = \sum_{D_2 \dot{\cup} D_3 = V} p^{|D_2|} t^{v(I(C_w) + \Delta(D_3))}.$$

In this formulation we recognize a related graph polynomial defined by Traldi and Zulli [21]. Let  $\mathcal{A}(G)$  be the adjacency matrix of graph  $G$ . The *bracket polynomial* of a graph  $G$  is

$$[G](A, B, d) = \sum_{\Delta} A^{v(\Delta)} B^{\rho(\Delta)} d^{v(\mathcal{A}(G) + \Delta)}$$

with a summand for each  $n \times n$  diagonal matrix  $\Delta = \Delta(D)$  for some  $D \subseteq V$ .

Now let  $D \subseteq V$ , and set  $D_3 = D$  and  $D_2 = V \setminus D$ . Then  $v(\Delta(D)) = |V| - |D_3| = |D_2|$ . Taking  $A = p$ ,  $B = 1$ , and  $d = t$  we see that the terms for the bracket and assembly polynomials match. To be precise, the assembly polynomial  $S(G_w)$  of the 4-regular graph  $G_w$  defined by the double-occurrence string  $w$  is equal to the bracket polynomial  $[I(C_w)](p, 1, t)$  of the interlace graph  $I(C_w)$ . As a consequence, when two double-occurrence strings have the same interlace graph, their 4-regular graphs have the same assembly polynomial, cf. [9, Proposition 3].

The transition polynomial allows for a straightforward recursive relation [15, Proposition 4], cf. [7, Lemma 6.4] for the case of the assembly polynomial. In fact, this recursive relation *characterizes* the transition polynomial. In a similar way, the bracket polynomial allows for a characteristic recursive relation [21], using a generalization of Euler circuit transformations called local complementation and edge complementation (which in turn is a special case of the general matrix operation of principal pivot transform [22]).

We hope these connections may be a starting point for transferring notions and results from one of these fields to another.

## References

1. Aigner, M., van der Holst, H.: Interlace polynomials. *Linear Algebra and its Applications* **377** (2004) 11–30
2. Arratia, R., Bollobás, B., Sorkin, G.B.: The interlace polynomial of a graph. *Journal of Combinatorial Theory, Series B* **92**(2) (2004) 199–233
3. Brijder, R., Daley, M., Harju, T., Jonoska, N., Petre, I., Rozenberg, G.: Computational nature of gene assembly in ciliates. In Rozenberg, G., Bäck, T., Kok, J., eds.: *Handbook of Natural Computing. Volume 3*. Springer (2012) 1233–1280
4. Brijder, R., Hoogeboom, H.J.: The algebra of gene assembly in ciliates. In Jonoska, N., Saito, M., eds.: *Discrete and Topological Models in Molecular Biology. Natural Computing Series*. Springer Berlin Heidelberg (2014) 289–307

5. Brijder, R., Hooeboom, H.J.: Interlace polynomials for multimatroids and delta-matroids. *European Journal of Combinatorics* (2014 to appear) [arXiv:1010.4678].
6. Burns, J., Dolzhenko, E.: Assembly words (properties) (visited march 2014) <http://knot.math.usf.edu/assembly/properties.html>.
7. Burns, J., Dolzhenko, E., Jonoska, N., Muche, T., Saito, M.: Four-regular graphs with rigid vertices associated to DNA recombination. *Discrete Applied Mathematics* **161**(10-11) (2013) 1378–1394
8. Cohn, M., Lempel, A.: Cycle decomposition by disjoint transpositions. *Journal of Combinatorial Theory, Series A* **13**(1) (1972) 83–89
9. Dolzhenko, E., Valencia, K.: Invariants of graphs modeling nucleotide rearrangements. In Jonoska, N., Saito, M., eds.: *Discrete and Topological Models in Molecular Biology*. Natural Computing Series. Springer Berlin Heidelberg (2014) 309–323
10. Ehrenfeucht, A., Harju, T., Petre, I., Prescott, D.M., Rozenberg, G.: *Computation in Living Cells – Gene Assembly in Ciliates*. Springer Verlag (2004)
11. Ellis-Monaghan, J.A., Merino, C.: Graph polynomials and their applications I: The Tutte polynomial. In Dehmer, M., ed.: *Structural Analysis of Complex Networks*. Birkhäuser Boston (2011) 219–255
12. Ellis-Monaghan, J.A., Merino, C.: Graph polynomials and their applications II: Interrelations and interpretations. In Dehmer, M., ed.: *Structural Analysis of Complex Networks*. Birkhäuser Boston (2011) 257–292
13. Ellis-Monaghan, J.A., Sarmiento, I.: Generalized transition polynomials. *Congressus Numerantium* **155** (2002) 57–69
14. Godsil, C., Royle, G.: *Algebraic Graph Theory*. Springer (2001)
15. Jaeger, F.: On transition polynomials of 4-regular graphs. In Hahn, G., Sabidussi, G., Woodrow, R., eds.: *Cycles and Rays*. Volume 301 of NATO ASI Series. Kluwer (1990) 123–150
16. Kotzig, A.: Eulerian lines in finite 4-valent graphs and their transformations. In: *Theory of graphs, Proceedings of the Colloquium, Tihany, Hungary, 1966*, Academic Press, New York (1968) 219–230
17. Martin, P.: *Énumérations eulériennes dans les multigraphes et invariants de Tutte-Grothendieck*. PhD thesis, Institut d’Informatique et de Mathématiques Appliquées de Grenoble (IMAG) (1977) Available online: <http://tel.archives-ouvertes.fr/tel-00287330.v1/>.
18. Prescott, D.M.: Genome gymnastics: Unique modes of DNA evolution and processing in ciliates. *Nature Reviews* **1** (2000) 191–199
19. Prescott, D.M., Greslin, A.F.: Scrambled actin I gene in the micronucleus of *Oxytricha nova*. *Developmental Genetics* **13** (1992) 66–74
20. Traldi, L.: Binary nullity, Euler circuits and interlace polynomials. *European Journal of Combinatorics* **32**(6) (2011) 944–950
21. Traldi, L., Zulli, L.: A bracket polynomial for graphs, I. *Journal of Knot Theory and Its Ramifications* **18**(12) (2009) 1681–1709
22. Tsatsomeros, M.J.: Principal pivot transforms: properties and applications. *Linear Algebra and its Applications* **307**(1-3) (2000) 151–165