Split graphs: combinatorial species and asymptotics

Justin M. Troyka[*]

Department of Mathematics and Statistics
York University
Toronto, Canada

jmtroyka@yorku.ca

Abstract

A split graph is a graph whose vertices can be partitioned into a clique and a stable set. We investigate the combinatorial species of split graphs, providing species-theoretic generalizations of enumerative results due to Bína and Přibil (2015), Cheng, Collins, and Trenk (2016), and Collins and Trenk (2018). In both the labeled and unlabeled cases, we give asymptotic results on the number of split graphs, of unbalanced split graphs, and of bicolored graphs, including proving the conjecture of Cheng, Collins, and Trenk (2016) that almost all split graphs are balanced.

Mathematics Subject Classifications: 05C30, 05A15, 05A16

A *split graph* is a graph whose vertices can be partitioned into two sets, K and S, such that the vertices in K form a clique (complete subgraph) and the vertices in S form a stable set (independent set). A partition of a graph in this way is called a *KS-partition*.

Split graphs are a well-known class of perfect graphs, and they are precisely the graphs G such that both G and $\overline{G}$ are triangulated. A summary of split graphs from that perspective is found in Golumbic [7, Sec. 6]. Moreover, Hammer and Simeone [8] characterize split graphs in terms of their degree sequences and thereby provide an efficient algorithm for determining whether a graph is a split graph.

For a given class of graphs, two enumerative problems are to count the unlabeled graphs (the isomorphism classes of graphs) on n vertices and to count the labeled graphs on n vertices (in which the n vertices have distinct labels 1 through n). In recent years, there has been interest in counting both unlabeled and labeled split graphs. Royle [12] gives a bijection between unlabeled split graphs and minimal set covers, of which the latter

---

[*]Work done during graduate fellowship, Department of Mathematics, Dartmouth College, Hanover, NH, United States.

had previously been counted. Cheng, Collins, and Trenk [4] explore connections between split graphs and Nordhaus–Gaddum graphs, and Collins and Trenk [5] give bijections between unlabeled split graphs, XY-graphs (called bicolored graphs in this paper), and bipartite posets. Collins and Trenk also characterize the minimal set covers, XY-graphs and bipartite posets that correspond to unbalanced split graphs (defined in Section 2). As for counting labeled split graphs, the exact enumeration has been done by Bína and Přibil [3]; but our results yield a formula (Corollary 3.4) somewhat simpler than theirs. The asymptotic enumeration of labeled split graphs was done much earlier, by Bender, Richmond, and Wormald [1], who show the asymptotic number is the same as for a few other classes of labeled graphs, including what our paper calls bicolored graphs.

In this paper, we extend some of these enumerative results to the setting of combinatorial species. The theory of combinatorial species, introduced by Joyal [11], is a powerful conceptual framework for thinking about combinatorial structures that can be labeled or unlabeled, such as graphs. Our results on the level of species are valuable not only because we can instantly recover from them the known labeled and unlabeled enumerations, but because they say something more than enumeration alone: for two types of structures to have the same (or isomorphic) species means that they are combinatorially equivalent in some sense. For a comprehensive treatment of species theory, see [2]; for a summary of the theory and some applications to graph enumeration, see [10].

Section 1 is an exposition of the parts of species theory we will use. In Section 2 we provide more background on split graphs and prove several identities about the species of split graphs and related species. This culminates in Theorem 2.11, a species version of the result from [4, 5] that the number of unbalanced split graphs on n vertices equals the number of split graphs on $\leqslant n - 1$ vertices. In Section 3 we relate the species of split graphs to the species of bicolored graphs, and we prove several asymptotic results, including Theorem 3.14, that almost all split graphs are balanced, which was a conjecture of Cheng, Collins, and Trenk [4].

1   Combinatorial species

This section is an all-too-brief statement of some of the ideas originating from Joyal [11]; for many more details, see [11, 2, 10]. We describe two equivalent ways of viewing a combinatorial species: one in terms of categories and functors, and one in terms of permutation group actions.

A *(combinatorial) species* F is a functor from the category of finite sets with bijections to itself. That is, F is a rule that does the following:

- To each finite set I, assigns a finite set of *structures*, denoted $F[I]$;

- To each bijection φ between finite sets I and J, assigns a bijection $F[\varphi]$ between the sets $F[I]$ and $F[J]$.

The set I is thought of as a set of labels, in which case $F[I]$ is the set of F-structures in which each label in I occurs exactly once, and the bijection $F[\varphi]$ maps each F-structure

on label set I to an F-structure on label set J obtained by replacing label i with label $\varphi(i)$ for all $i \in I$.

A basic example of a species is the species of sets, denoted E (for *ensemble*, the French word for "set"). This is defined by $E[I] = \{I\}$: there is one structure with label set I, namely the set I itself. For a bijection $\varphi \colon I \to J$, we must have $E[\varphi]$ map the one element of $\{I\}$ to the one element of $\{J\}$.

For the species in this paper, the structures are graphs with a certain property. In this case, $F[I]$ is the set of graphs with that property whose vertices are labeled with the elements of I, and $F[\varphi] \colon F[I] \to F[J]$ maps each graph in $F[I]$ to the corresponding isomorphic graph in $F[J]$.

In the special case of $I = J = [n]$, we write $F[n]$ instead of $F[[n]]$, and $F[n]$ is considered as the set of labeled F-structures of size n. A bijection $[n] \to [n]$ is a permutation in the symmetric group $\mathfrak{S}_n$. Given $\sigma \in \mathfrak{S}_n$, the function $F[\sigma]$ is a permutation of the set $F[n]$. Thus F induces an action of $\mathfrak{S}_n$ on $F[n]$ for each n.

If F is a species of graphs with a certain property, then $\mathfrak{S}_n$ acts on $F[n]$ by graph isomorphisms. Given a graph $g \in F[n]$, each permutation induces an isomorphism from g to some graph in $F[n]$, and the permutations that map g to itself are the automorphisms of g.

In the example of E (the species of sets), since there is only one E-structure on label set $[n]$, the symmetric group must act trivially on it. Thus the automorphism group of an E-structure (a set) of size n is all of $\mathfrak{S}_n$: permuting the elements of a set does not change what the set is.

1.1 Labeled and unlabeled structures, generating functions, and species isomorphism

Let F be a species. The elements of $F[n]$ are the (labeled) F-structures of size n. The orbits of $F[n]$ under the action of $\mathfrak{S}_n$ are isomorphism classes of F-structures, and we write $F[n]/\mathfrak{S}_n$ to denote the set of these orbits. We think of each orbit as an *unlabeled* F-structure. For example, in the case where F is a species of graphs, the orbits are the isomorphism classes of graphs, which are unlabeled graphs.

A species F has three kinds of generating functions associated with it:

- $F(x)$ denotes the exponential generating function for the labeled F-structures:

$$F(x) = \sum_{n \geqslant 0} |F[n]| \frac{x^n}{n!}.$$

- $\widetilde{F}(x)$ denotes the ordinary generating function for the unlabeled F-structures:

$$\widetilde{F}(x) = \sum_{n \geqslant 0} |F[n]/\mathfrak{S}_n| \, x^n.$$

- $Z_F(p_1, p_2, \ldots)$ denotes the cycle index series of F, a generating function in infinitely many variables. We will not use the cycle index series in this paper, but it is

important because it generalizes $F(x)$ and $\widetilde{F}(x)$, in the sense that $Z_F(x, 0, 0, \ldots) = F(x)$ and $Z_F(x, x^2, x^3, \ldots) = \widetilde{F}(x)$.

Note that for two species F and G we can have $F(x) = G(x)$ without having $\widetilde{F}(x) = \widetilde{G}(x)$, or the other way around.

For the set species E, since $|E[n]| = 1$ for all n, we have $E(x) = \displaystyle\sum_{n \geqslant 0} \frac{x^n}{n!} = e^x$ and $\widetilde{E}(x) = \displaystyle\sum_{n \geqslant 0} x^n = \frac{1}{1-x}$.

A *species isomorphism* α from F to G is a natural equivalence from F to G as functors. That is, α is a family of bijections $\alpha_I \colon F[I] \to G[I]$ for each label set I that commutes with bijections between label sets: for any bijection $\varphi \colon I \to J$, we have $\alpha_J \circ F[\varphi] = G[\varphi] \circ \alpha_I$. Viewing species in terms of group actions, this is equivalent to a bijection $\alpha_n \colon F[n] \to G[n]$ for each n that preserves the action of $\mathfrak{S}_n$, meaning that $\alpha_n(\sigma \cdot f) = \sigma \cdot \alpha_n(f)$ for all $\sigma \in \mathfrak{S}_n$ and $f \in F[n]$. Thus, F and G are isomorphic if as functors they are naturally equivalent, or if $F[n]$ and $G[n]$ are isomorphic $\mathfrak{S}_n$-sets for every n; in this case we simply write $F = G$.

The species E is isomorphic to the species of complete graphs (cliques), and it is also isomorphic to the species of edgeless graphs (stable sets). It is useful to think of E in this way when we think about building certain species of graphs from other species of graphs.

If F and G are isomorphic, then $Z_F = Z_G$; and, as already discussed, if $Z_F = Z_G$, then $F(x) = G(x)$ and $\widetilde{F}(x) = \widetilde{G}(x)$. The converse does not hold for any of these implications. Thus, species isomorphism is the finest notion of equality between combinatorial structures that we have discussed here. We will view isomorphic species as being equal, simply writing $F = G$ if F and G are isomorphic.

1.2 Addition and multiplication of species

Given species F and G, the sum $F + G$ is a species whose structures are F-structures or G-structures — that is, $(F+G)[I] = F[I] \sqcup G[I]$ (the disjoint union). From the perspective of group actions, $(F + G)[n]$ is the disjoint union of $F[n]$ and $G[n]$ as $\mathfrak{S}_n$-sets.

If F and G are species of two different types of graphs, then $F + G$ is the species of graphs of one type or the other type, provided that no graph is of both types. If the two types of graphs do overlap, then $F + G$ double-counts the intersection.

The product $F \cdot G$ is defined as the species whose structures are ordered pairs of an F-structure and a G-structure. That is, the structures in $(F \cdot G)[I]$ are obtained by partitioning the labels as $I = U \cup V$ and forming ordered pairs (f, g) with $f \in F[U]$ and $g \in G[V]$. The size of such an ordered pair as an $(F \cdot G)$-structure is the sum of the sizes of its two components.

The sum and product of species correspond to the sum and product of their generating functions: that is, $(F + G)(x) = F(x) + G(x)$ and $\widetilde{(F+G)}(x) = \widetilde{F}(x) + \widetilde{G}(x)$ and $Z_{F+G} = Z_F + Z_G$, and likewise for multiplication. Furthermore, sum and product respect species isomorphism, in the sense that the isomorphism class of $F+G$ or $F \cdot G$ is determined by the isomorphism classes of F and G.

We can use these operations to build up complicated species from simpler ones. For instance, $E \cdot E$ is the species of partitions of a set into an ordered pair of two sets, which are equivalent to subsets of a set. The labeled generating function is $E(x) E(x) = e^x e^x = e^{2x}$, in which the coefficient of $x^n/n!$ is 2^n; hence a set of size n has 2^n subsets. The unlabeled generating function is $\widetilde{E}(x) \widetilde{E}(x) = \dfrac{1}{(1-x)^2}$, in which the coefficient of x^n is $n+1$; hence, if the elements of a set of size n are unlabeled, then the set has $n + 1$ distinguishable subsets (one of each size).

1.3 Other species notation

The zero species, denoted 0, is defined as the species with no structures: $0[I] = \varnothing$ for every set I. The zero species is the additive identity: $F + 0 = F$ for all F. It also satisfies $0 \cdot F = 0$ for all F. The one species, denoted 1, is defined as the species with one structure of size 0 and no other structures: $1[\varnothing] = \{\varepsilon\}$ (a null structure) and $1[I] = \varnothing$ for every non-empty I. The one species is the multiplicative identity: $1 \cdot F = F$ for any species F.

Given a species F, the species F_k is the species of F-structures of size k; that is, $F_k[I] = \varnothing$ if $|I| \neq k$, and $F_k[I] = F[I]$ if $|I| = k$. This means that $F = \sum_{n \geqslant 0} F_n$. We also write $F_{\leqslant k}$ to denote the species of F-structures of size $\leqslant k$, so $F_{\leqslant k} = F_0 + \cdots + F_k$.

1.4 The ring of virtual species

We have defined addition and multiplication of species. By writing formal differences of species, such as $F - G$, we can extend the set of species to a ring, called *the ring of virtual species*, in which the species 0 and 1 are the additive and multiplicative identities. The elements of this ring are called *virtual species*. Every virtual species F has the form $F^+ - F^-$, where F^+ and F^- are ordinary species. If $F^+ - F^-$ and $G^+ - G^-$ are two virtual species in this form, then we identify them as the same virtual species if the two ordinary species $F^+ + G^-$ and $G^+ + F^-$ are isomorphic, i.e. $F^+ + G^- = G^+ + F^-$. The fact that virtual species form a ring makes algebraic manipulation much easier with virtual species than with combinatorial species alone, as we will see in our computations with the species of split graphs.

A virtual species F is a unit if and only if $F_0 = 1$ or $F_0 = -1$, i.e. the "constant term" of F is equivalent as a virtual species to the species 1 or its negative -1. When this is the case, we will write $1/F$ to denote the multiplicative inverse of F, and we will write fractions accordingly.

2 Split graphs

We begin with a characterization of the sizes of K and S in a KS-partition of a split graph.

Proposition 2.1 ([7, Thm. 6.2]). *Let G be a split graph and fix a KS-partition of G. Let $\omega(G)$ (resp. $\alpha(G)$) denote the maximum size of a clique (resp. stable set) in G. Exactly one of the following holds:*

(i) $|K| = \omega(G)$ *and* $|S| = \alpha(G)$ *and* G *has a unique* KS-*partition;*

(ii) $|K| = \omega(G) - 1$ *and* $|S| = \alpha(G)$ *and there is* $x \in S$ *such that* $K \cup \{x\}$ *is a clique;*

(iii) $|K| = \omega(G)$ *and* $|S| = \alpha(G) - 1$ *and there is* $x \in K$ *such that* $S \cup \{x\}$ *is a stable set.*

This proposition prompts the following definition, following [4, 5]:

Definition 2.2. Let G be a split graph. We say G is *balanced* if it has a unique KS-partition, and *unbalanced* otherwise. A given KS-partition of G is *S-max* if S is as large as possible, and *K-max* if K is as large as possible. Furthermore, a vertex x as in case (ii) or (iii) of Proposition 2.1 is called a *swing vertex* of G.

In this language, Proposition 2.1 gives us these facts: every KS-partition of G is S-max or K-max; if G is unbalanced then the K in a K-max partition has one vertex more than the K in an S-max partition, and similarly for S; and a split graph is unbalanced if and only if it has a swing vertex.

In this section, we define four types of split graphs and define colored split graphs (Section 2.1), we describe the species of split graphs and various related species (Section 2.2), and we prove our main theorem and other identities on these species (Sections 2.3 and 2.4).

2.1 Classifying split graphs

A result due to Cheng, Collins, and Trenk [4] describes the structure of the swing vertices of a split graph and lists all of the KS-partitions.

Proposition 2.3 ([4, Thm. 10]). *Let G be an unbalanced split graph, and let A be the set of swing vertices of G. Then A is either a clique or a stable set, and the non-swing vertices admit a partition into sets Y and Z such that every vertex in A is adjacent to every vertex in Y and no vertex in Z. Furthermore:*

- *If A is a clique, then there is a unique K-max partition, namely $K = A \cup Y$ and $S = Z$; and the S-max partitions are given by $K = (A \smallsetminus \{a\}) \cup Y$ and $S = Z \cup \{a\}$ for $a \in A$.*

- *If A is a stable set, then there is a unique S-max partition, namely $K = Y$ and $S = A \cup Z$; and the K-max partitions are given by $K = Y \cup \{a\}$ and $S = (A \smallsetminus \{a\}) \cup Z$ for $a \in A$.*

In particular, every vertex that is not a swing vertex is either in K for all KS-partitions or in S for all KS-partitions. Proposition 2.3 allows us to classify split graphs according to whether their swing vertices form a clique or a stable set:

Definition 2.4. Let G be a split graph.

- G is *K-canonical* if the set of swing vertices forms a clique of size $\geqslant 2$;

- G is *S-canonical* if the set of swing vertices forms a stable set of size $\geqslant 2$;

- G is *ambiguous* if there is exactly one swing vertex;

- G is *balanced* if there are no swing vertices.

By Proposition 2.3, every split graph is exactly one of those four types. In [4], a split graph that is K-canonical or ambiguous is an *NG-1 graph*, and a split graph that is S-canonical or ambiguous is an *NG-2 graph*. Our choice of names is because we will use the following "canonical" KS-partition of a split graph G: if G is K-canonical, the canonical partition is the unique K-max partition; if G is S-canonical, the canonical partition is the unique S-max partition; if G is ambiguous, there is no canonical partition; and if G is balanced, the canonical partition is the unique KS-partition.

Definition 2.5. A *colored split graph* is a split graph with a chosen S-max partition. Equivalently, it is a split graph with vertices colored green (Kelly green) and red (Scarlet) such that the green set and the red set are respectively K and S in an S-max partition.

The four types of split graphs extend to colored split graphs. If a split graph is S-canonical, ambiguous, or balanced, then it has a unique S-max partition, so there is only one way to color the vertices to obtain a colored split graph. However, if a split graph is K-canonical, then there is more than one S-max partition, and each one gives rise to a different colored split graph.

2.2 The species of split graphs and related species

Let S be the species of split graphs. What this means is that, for I a finite set, $S[I]$ is the set of split graphs on vertex set I, and for any bijection φ between finite sets I and J, $S[\varphi]$ is the bijection between $S[I]$ and $S[J]$ that maps each split graph on I to the isomorphic copy obtained by replacing label i with $\varphi(i)$.

Also let B be the species of balanced split graphs, and let U be the species of unbalanced split graphs; note that $S = B + U$. Recall that, for a species F, $F(x)$ is the exponential generating function that counts labeled F-structures, and $\widetilde{F}(x)$ is the ordinary generating function that counts unlabeled F-structures. In [4, 5], it is proved that the number of unlabeled unbalanced split graphs on n vertices equals the number of unlabeled split graphs on $\leqslant n-1$ vertices: in the language of generating functions,

$$\widetilde{U}(x) = \frac{x}{1-x}\widetilde{S}(x). \tag{2.1}$$

The idea behind this is as follows: every *unlabeled* unbalanced split graph has a unique K-max partition, and from this partition we can obtain a split graph by removing all the swing vertices from K. However, this does not work for *labeled* graphs, because a labeled

split graph can have more than one K-max partition. So we will need to be more careful in order to obtain an identity like (2.1) for the labeled generating functions or on the level of species.

Let U^K be the species of K-canonical split graphs. Let U^S be the species of S-canonical split graphs. Let U^{amb} be the species of ambiguous split graphs. Then $U = U^K + U^S + U^{\mathrm{amb}}$. Taking the graph complement gives a bijection between K-canonical split graphs and S-canonical split graphs, and this bijection commutes with graph isomorphisms, so it is a species isomorphism between U^K and U^S; thus, $U^K = U^S$.

Let $\mathfrak{c}S$ denote the species of colored split graphs. The isomorphisms between colored graphs are the graph isomorphisms that preserve color. It is a consequence of Proposition 2.3 that a split graph has a unique S-max partition up to relabeling of the vertices. This means that colored split graphs with the same underlying graph are isomorphic. Therefore:

Proposition 2.6. $\widetilde{\mathfrak{c}S}(x) = \widetilde{S}(x)$; *in words, the number of unlabeled colored split graphs on n vertices equals the number of unlabeled split graphs on n vertices.* $\qquad\square$

Let $\mathfrak{c}U^K$, $\mathfrak{c}U^S$, $\mathfrak{c}U^{\mathrm{amb}}$, and $\mathfrak{c}B$ denote the species of colored split graphs that are respectively K-canonical, S-canonical, ambiguous, and balanced. Then $\mathfrak{c}S = \mathfrak{c}U^K + \mathfrak{c}U^S + \mathfrak{c}U^{\mathrm{amb}} + \mathfrak{c}B$. Since a split graph that is not K-canonical has only one colored split graph associated with it, we have the following equalities of species:

Proposition 2.7. $\mathfrak{c}U^S = U^S$ *and* $\mathfrak{c}U^{\mathrm{amb}} = U^{\mathrm{amb}}$ *and* $\mathfrak{c}B = B$, *and* $\mathfrak{c}S - \mathfrak{c}U^K = S - U^K$.

Proof. The first three equalities are immediate from the remark preceding the proposition. The last one follows from the first three because $S = U^K + U^S + U^{\mathrm{amb}} + B$ and $\mathfrak{c}S = \mathfrak{c}U^K + \mathfrak{c}U^S + \mathfrak{c}U^{\mathrm{amb}} + \mathfrak{c}B$. $\qquad\square$

Thus we will make no further use of the symbols $\mathfrak{c}U^S$, $\mathfrak{c}U^{\mathrm{amb}}$, and $\mathfrak{c}B$.

We summarize the species we have defined in this section and the relations between them that we have seen so far:

S	split graphs	$\mathfrak{c}S$	colored split graphs
U	unbalanced split graphs		
U^K	K-canonical split graphs	$\mathfrak{c}U^K$	K-canonical colored split graphs
U^S	S-canonical split graphs		
U^{amb}	ambiguous split graphs		
B	balanced split graphs		

$$S = U + B$$
$$U = U^K + U^S + U^{\mathrm{amb}}$$
$$U^S = U^K$$
$$\mathfrak{c}S = \mathfrak{c}U^K + U^S + U^{\mathrm{amb}} + B$$
$$\mathfrak{c}S - \mathfrak{c}U^K = S - U^K$$

2.3 Species identities involving colored split graphs

This section includes three very bijective proofs of species identities involving colored split graphs. The colored split graphs are needed so that we can get the results on split graphs that we really want, in the next section.

The first theorem is a partial analog of (2.1) on the level of species:

Theorem 2.8. $U^K = E_{\geqslant 2} \cdot \mathfrak{c}S$.

Proof. We find a bijection between the labeled structures that commutes with isomorphisms, i.e. that is invariant under permuting the labels. The left side, U^K, counts K-canonical split graphs. The right side, $E_{\geqslant 2} \cdot \mathfrak{c}S$, counts ordered pairs of a set of size $\geqslant 2$ and a colored split graph.

Let G be a labeled K-canonical split graph, with its canonical K-max partition. Color the vertices in K green and the vertices in S red. Coloring the structures of U^K in this way does not change U^K, because isomorphisms between K-canonical split graphs preserve the K-max partition.

We now define the bijection by mapping G to $(A, G - A)$, where A is the set of swing vertices of G and $G - A$ is the colored graph obtained by removing A from G. The swing vertices form a clique of size $\geqslant 2$, so A is a structure in $E_{\geqslant 2}$. The green set and the red set in $G - A$ (inheriting the colors from G) are respectively K and S in a KS-partition of $G - A$; so, to show that $G - A$ is a colored split graph, we show that this partition is S-max.

In the canonical K-max partition of G, all the swing vertices of G are in K, so no vertices of S are removed from G to form $G - A$. Suppose $G - A$ is not S-max. Then, by Proposition 2.3, $G - A$ has a vertex $v \in K$ adjacent to none of the vertices in S. But G and $G - A$ have the same S, so v can be moved to S to form a new KS-partition of G, making v a swing vertex of G. Thus $v \in A$, a contradiction. Therefore the green set and the red set form an S-max partition of $G - A$.

The mapping $G \mapsto (A, G - A)$ does not depend on how the labels of the vertices in G are permuted, precisely because the K-max partition of G is unique.

To show this is a bijection, we describe its inverse. Let A be a set of size $\geqslant 2$ and let H be a colored split graph with chosen partition $K \cup S$, with the elements of A and the vertices of H given distinct labels from a shared label set. We obtain a split graph G from the ordered pair (A, H) by adding the elements of A into K as follows: put an edge between every element of A and every vertex in K, and also put an edge between every pair of elements of A. Every element of A is now a swing vertex, so the swing vertices of G form a clique of size $\geqslant 2$, and so G is K-canonical.

The set A in the mapping $(A, H) \mapsto G$ becomes the set of swing vertices of G; and conversely, in the canonical partition of a K-canonical graph, the swing vertices are all in K and are adjacent to no vertices in S (by Proposition 2.3). Hence these two functions are inverses, proving that $U^K = E_{\geqslant 2} \cdot \mathfrak{c}S$. $\qquad \square$

Now that we have done a detailed proof of species equality, the next ones will be somewhat abbreviated, as they use the same idea.

Theorem 2.9. $U^{\mathrm{amb}} = X \cdot B$.

Proof. The left side, U^{amb}, counts ambiguous split graphs. The right side, $X \cdot B$, counts ordered pairs of a single element and a balanced split graph. We go from the left side to the right side as follows: given an ambiguous split graph G with swing vertex a, map G to $(a, G - a)$. It turns out that $G - a$ is balanced, which we prove below.

The inverse, going from the right side to the left side, is as follows: given a single element a and a balanced split graph H, append a as a new swing vertex in H, adding an edge between it and every vertex in K. This is well-defined precisely because a balanced split graph has a unique KS-partition.

We now need to show that, if G is an ambiguous split graph with swing vertex a, then $G - a$ is balanced. By Proposition 2.3, the vertices of $G - a$ can be partitioned into a clique Y and a stable set Z such that a is adjacent to everything in Y and nothing in Z. Suppose $G - a$ is unbalanced. Without loss of generality, the partition $Y \cup Z$ is a K-max partition of $G - a$. Then by Proposition 2.3 there is $y \in Y$ that is adjacent to nothing in Z. This makes y a swing vertex of G as well, contradicting that a is the only swing vertex of G. Therefore, $G - a$ is a balanced split graph, as claimed. $\square$

Theorem 2.10. $\mathfrak{c}U^K = X \cdot E_{\geqslant 1} \cdot \mathfrak{c}S$.

Proof. We can think of $X \cdot E_{\geqslant 1}$ as the species of "pointed sets" of size $\geqslant 2$, a pointed set being a set with a chosen distinguished element. The left side, $\mathfrak{c}U^K$, counts colored K-canonical split graphs. The right side, $X \cdot E_{\geqslant 1} \cdot \mathfrak{c}S$, counts ordered pairs of a pointed set of size $\geqslant 2$ and a colored split graph.

A colored K-canonical split graph can be obtained by taking a K-canonical split graph and choosing one of its swing vertices to be in S. From Theorem 2.8 we have

$$U^K = E_{\geqslant 2} \cdot \mathfrak{c}S,$$

where $E_{\geqslant 2}$ represents the set of swing vertices in a K-canonical split graph; choosing one of the swing vertices to be in S can be accomplished by replacing $E_{\geqslant 2}$ with the species of pointed sets $X \cdot E_{\geqslant 1}$, which yields the desired result. $\square$

We summarize the species equalities we obtained in this section:

$$U^K = E_{\geqslant 2} \cdot \mathfrak{c}S$$
$$U^{\mathrm{amb}} = X \cdot B$$
$$\mathfrak{c}U^K = X \cdot E_{\geqslant 1} \cdot \mathfrak{c}S$$

2.4 Main theorem on the species of split graphs

In this section, we will manipulate the species equalities found in the previous two sections, obtaining an equation relating the species of unbalanced split graphs and the species of split graphs. This is where the ring of virtual species finally pays off.

Theorem 2.11. $U = \dfrac{(2 - X) \cdot E - 2}{(1 - X) \cdot E} \cdot S$.

Proof.

$$\mathfrak{c}S = S - U^K + \mathfrak{c}U^K \qquad\qquad \text{(Prop. 2.7)}$$

$$\mathfrak{c}S = S - U^K + X \cdot E_{\geqslant 1} \cdot \mathfrak{c}S \qquad\qquad \text{(Thm. 2.10)},$$

and we can solve for $\mathfrak{c}S$ (in the ring of virtual species) to get

$$\mathfrak{c}S = \frac{S - U^K}{1 - X \cdot E_{\geqslant 1}}. \qquad\qquad (2.2)$$

The virtual species $1 - X \cdot E_{\geqslant 1}$ has constant term 1, so it is a unit and we can divide by it.

Now we can find U^K in terms of S alone:

$$U^K = E_{\geqslant 2} \cdot \mathfrak{c}S \qquad\qquad \text{(Thm. 2.8)}$$

$$U^K = E_{\geqslant 2} \cdot \frac{S - U^K}{1 - X \cdot E_{\geqslant 1}} \qquad\qquad \text{(Eqn. (2.2))},$$

and we can solve for U^K to get

$$U^K = \frac{E_{\geqslant 2}}{1 - X \cdot E_{\geqslant 1} + E_{\geqslant 2}} \cdot S.$$

By substituting $E_{\geqslant 1} = E - 1$ and $E_{\geqslant 2} = E - 1 - X$, we get

$$U^K = \frac{E - 1 - X}{(1 - X) \cdot E} \cdot S. \qquad\qquad (2.3)$$

The species $(1 + X) \cdot E$ has constant term 1, so it is a unit and we can divide by it.

Now we can find U in terms of S alone:

$$\begin{aligned}
U &= U^K + U^S + U^{\mathrm{amb}} \\
&= 2\,U^K + U^{\mathrm{amb}} \\
&= 2\,U^K + X \cdot B \qquad\qquad \text{(Thm. 2.9)} \\
&= 2\,U^K + X \cdot (S - U) \\
U &= 2\,\frac{E - 1 - X}{(1 - X) \cdot E} \cdot S + X \cdot (S - U) \qquad\qquad \text{(Eqn. (2.3))},
\end{aligned}$$

and solving for U gives the desired theorem. $\qquad\square$

Theorem 2.11 is the true generalization of (2.1). It expresses the species of unbalanced split graphs as the product of split graphs with a species involving sets. In fact, (2.1) is recovered from Theorem 2.11 by passing to the unlabeled generating functions, using the fact that $\widetilde{E}(x) = \dfrac{1}{1 - x}$. In the same way, we can pass to the labeled generating functions, using the fact that $E(x) = e^x$. This process yields a new result:

Theorem 2.12. $U(x) = \dfrac{(2 - x)e^x - 2}{(1 - x)e^x} S(x) = \dfrac{2 - x - 2e^{-x}}{1 - x} S(x).$ $\qquad\square$

3   Bicolored graphs and asymptotics

This section concerns asymptotic enumeration. Given non-negative sequences x_n and y_n, we say x_n asymptotically equals y_n if $\lim_{n \to \infty} \dfrac{x_n}{y_n} = 1$, and we write this as $x_n \sim y_n$. If x_n counts certain objects of size n, then we say almost all of the objects have a certain property if the fraction of size-n objects with that property goes to 1 as $n \to \infty$.

A *bicolored graph* is a graph in which each vertex is colored green or red such that no two adjacent vertices are the same color. In other words, it is a bipartite graph with a chosen bipartition. Green and red are not interchangeable, meaning that swapping the color of every vertex will generally result in a different bicolored graph; that is, the two parts in the chosen bipartition are an *ordered* pair. Bicolored graphs are often studied as a step towards bipartite graphs, as in [9, 6].

Let BC be the species of bicolored graphs, where the isomorphisms between structures are the graph isomorphisms that preserve color. The species BC is fundamental, in the sense that its cycle index series and other associated generating functions have explicit formulas that can be derived from scratch rather than built up from those of simpler species (see [6]). In particular, the number of labeled bicolored graphs is simple to express and routine to derive:

$$|\mathrm{BC}[n]| = \sum_{k=0}^{n} \binom{n}{k} 2^{k(n-k)}, \tag{3.1}$$

because we choose a subset of k vertices to color green, and then for each of the $k(n-k)$ pairs of opposite-color vertices we choose whether to make them adjacent.

In this section, we prove some identities about the species of split graphs and the species of bicolored graphs (Section 3.1); and we prove some asymptotic results on the number of split graphs and unbalanced split graphs (Sections 3.2 and 3.3).

3.1   The species of bicolored graphs

Let BC^* be the species of bicolored graphs in which no green vertex is isolated. Then clearly $\mathrm{BC} = E \cdot \mathrm{BC}^*$, because E provides the set of isolated green vertices. Collins and Trenk [5] find a bijection between unlabeled split graphs and unlabeled BC^*-graphs (they say "XY-graph" where we say "bicolored graph"), and their bijection easily proves this species identity:

Theorem 3.1. $\mathfrak{c}S = \mathrm{BC}^* = \dfrac{\mathrm{BC}}{E}$.

Proof. The equality $\mathrm{BC}^* = \dfrac{\mathrm{BC}}{E}$ is clear (see the paragraph preceding the theorem), and note that E is a unit in the ring of virtual species.

Now we define a bijection between labeled colored split graphs and bicolored graphs with no isolated green vertex: given a colored split graph, remove the edges in the green clique. The inverse is: given a bicolored graph with no isolated green vertex, add an edge between every pair of green vertices, making the green vertices a clique. A swing vertex

in K in the colored split graph becomes an isolated green vertex in the bicolored graph, and the converse is also true; consequently the bijection does take colored split graphs to BC*-graphs, and likewise for the inverse map. Since this bijection respects isomorphism of colored graphs, the equality of species is proved. $\qquad\square$

Using Theorem 3.1 and some of the species identities from Sections 2.3 and 2.4, we obtain a relationship between S and BC:

Theorem 3.2. $\mathrm{BC} = \dfrac{1}{1 - X} \cdot S.$

Proof. By Theorem 2.8, $U^K = E_{\geqslant 2} \cdot \mathfrak{c}S$. By Theorem 3.1, $\mathfrak{c}S = \dfrac{\mathrm{BC}}{E}$. Therefore,

$$U^K = \frac{E_{\geqslant 2}}{E} \cdot \mathrm{BC}. \tag{3.2}$$

Equation (2.3) is that

$$U^K = \frac{E_{\geqslant 2}}{(1 - X) \cdot E} \cdot S;$$

setting this equal to (3.2) and canceling the factor of $\dfrac{E_{\geqslant 2}}{E}$ yields the desired result. $\qquad\square$

Note that $\dfrac{1}{1 - X}$ is the species of sequences, or linear orders. Is there a direct combinatorial way in which a bicolored graph is a split graph and a sequence?

By passing to the labeled and unlabeled generating functions, we obtain:

Corollary 3.3. $\mathrm{BC}(x) = \dfrac{1}{1 - x} S(x)$ *and* $\widetilde{\mathrm{BC}}(x) = \dfrac{1}{1 - x} \widetilde{S}(x).$ $\qquad\square$

The latter equation was already known, as it comes from $\widetilde{U}(x) = x\,\widetilde{\mathrm{BC}}(x)$ [5, Thm. 15] and $\widetilde{U}(x) = \dfrac{x}{1 - x} \widetilde{S}(x)$ (which is (2.1)). The former equation, when rearranged, implies that the number of labeled split graphs on n vertices is $|\mathrm{BC}[n]| - n\,|\mathrm{BC}[n-1]|$; as we have an explicit formula for $|\mathrm{BC}[n]|$ in (3.1), we obtain:

Corollary 3.4. *The number of labeled split graphs on n vertices is*

$$\sum_{k=0}^{n} \binom{n}{k} 2^{k(n-k)} - n \sum_{k=0}^{n-1} \binom{n-1}{k} 2^{k(n-k-1)}. \qquad\square$$

This is a different formula than the one obtained by Bína and Přibil [3]; theirs gives the number of labeled split graphs as

$$1 + \sum_{k=2}^{n} \binom{n}{k} \left((2^k - 1)^{n-k} - \sum_{j=1}^{n-k} \frac{jk}{j+1} \binom{n-k}{j} (2^{k-1} - 1)^{n-k-j} \right).$$

Apparently this gives the same numbers as Corollary 3.4; this of course is assured by the validity of both their proof and ours, and we have also checked by computer that the two agree up to $n = 318$ (the n that had been reached by the time we finished our sandwich). It would probably not be too hard to find an elementary proof that they are equal, by rearranging the sums and using properties of binomial coefficients.

3.2 Almost all labeled split graphs are balanced

Let b_n be the number of labeled bicolored graphs on n vertices, let s_n be the number of labeled split graphs on n vertices, and let u_n be the number of labeled unbalanced split graphs on n vertices. In this subsection we prove that almost all labeled split graphs are balanced: $\lim_{n \to \infty} \dfrac{s_n - u_n}{s_n} = 1$, or equivalently $\lim_{n \to \infty} \dfrac{u_n}{s_n} = 0$.

Bender, Richmond, and Wormald [1] show that

$$s_n \sim b_n = \sum_{k=0}^{n} \binom{n}{k} 2^{k(n-k)} \sim c(n) \binom{n}{\lfloor n/2 \rfloor} 2^{n^2/4} \qquad (3.3)$$

where $c(n) > 0$ depends only on whether n is even or odd — in particular,

$$c(n) = \begin{cases} \displaystyle\sum_{k \in \mathbb{Z}} 2^{-k^2} \approx 2.128937 & \text{if } n \text{ is even;} \\ \displaystyle\sum_{k \in \mathbb{Z}} 2^{-(k+\frac{1}{2})^2} \approx 2.128931 & \text{if } n \text{ is odd.} \end{cases}$$

Also note that the equality $b_n = \sum_{k=0}^{n} \binom{n}{k} 2^{k(n-k)}$ in (3.3) is just (3.1).

From (3.3) we immediately obtain this lemma:

Lemma 3.5. $\dfrac{s_n}{s_{n-1}} \geqslant 2^{(n+1)/2}$ and $\dfrac{b_n}{b_{n-1}} \geqslant 2^{(n+1)/2}$, for large enough n. $\qquad \square$

Theorem 3.6. *Almost all labeled split graphs are balanced:* $\lim_{n \to \infty} \dfrac{u_n}{s_n} = 0$.

Proof. Let $A(x) = \dfrac{2 - x - 2e^{-x}}{1 - x}$. From Theorem 2.12, we have $U(x) = A(x) S(x)$. Let a_i be the coefficient of x^i in $A(x)$; then $0 \leqslant a_i \leqslant 1$ for all i, and $a_0 = 0$. Then

$$\frac{u_n}{n!} = \sum_{k=0}^{n-1} a_{n-k} \frac{s_k}{k!} \leqslant \sum_{k=0}^{n-1} \frac{s_k}{k!}.$$

By Lemma 3.5, for large enough n, the highest term in this sum is the last one, $\dfrac{s_{n-1}}{(n-1)!}$.

Thus, for large enough n,

$$\frac{u_n}{n!} \leqslant n\frac{s_{n-1}}{(n-1)!}$$

$$u_n \leqslant n^2 s_{n-1}$$

$$\frac{u_n}{s_n} \leqslant \frac{n^2 s_{n-1}}{s_n} \leqslant \frac{n^2}{2^{(n+1)/2}}$$

(this last inequality uses Lemma 3.5), and $\dfrac{n^2}{2^{(n+1)/2}}$ goes to 0 as $n \to \infty$. $\square$

3.3 Almost all unlabeled split graphs are balanced

Let $\widetilde{b}_n$ be the number of unlabeled bicolored graphs on n vertices, let $\widetilde{s}_n$ be the number of unlabeled split graphs on n vertices, and let $\widetilde{u}_n$ be the number of unlabeled unbalanced split graphs on n vertices. We will use the following "folklore" result about bicolored graphs (proved e.g. in [13]):

Theorem 3.7 (see [13]). *The number of unlabeled bicolored graphs on n vertices asymptotically equals the number of labeled bicolored graphs on n vertices divided by n!. That is, $\widetilde{b}_n \sim b_n/n!$.*

We will now use this to show that $\lim\limits_{n\to\infty} \dfrac{\widetilde{u}_n}{\widetilde{s}_n} = 0$. Along the way, we also find that $\widetilde{s}_n \sim \widetilde{b}_n$.

Theorem 3.7 allows us to convert our results on labeled graphs into results on unlabeled graphs. From (3.3), we obtain:

Lemma 3.8. $\widetilde{b}_n \sim \dfrac{c(n)}{\lfloor n/2\rfloor! \lceil n/2\rceil!} 2^{n^2/4}$, *where $c(n)$ is defined as in Section 3.2.* $\square$

And from Lemma 3.5, we obtain:

Lemma 3.9. $\dfrac{\widetilde{b}_n}{\widetilde{b}_{n-1}} \geqslant 2^{(n+1)/2}/n$, *for large enough n.* $\square$

We now give a result on $\widetilde{s}_n$ analogous to (3.3):

Theorem 3.10. *The number of unlabeled split graphs on n vertices asymptotically equals the number of unlabeled bicolored graphs on n vertices: $\widetilde{s}_n \sim \widetilde{b}_n$.*

Proof. By Corollary 3.3, $\widetilde{s}_n = \widetilde{b}_n - \widetilde{b}_{n-1}$. Then $\lim\limits_{n\to\infty} \dfrac{\widetilde{s}_n}{\widetilde{b}_n} = 1 - \lim\limits_{n\to\infty} \dfrac{\widetilde{b}_{n-1}}{\widetilde{b}_n}$, and Lemma 3.9 implies that $\lim\limits_{n\to\infty} \dfrac{\widetilde{b}_{n-1}}{\widetilde{b}_n} = 0$. $\square$

Theorem 3.10 has three immediate corollaries:

Corollary 3.11. *The number of unlabeled split graphs on n vertices asymptotically equals the number of labeled split graphs on n vertices divided by $n!$. That is, $\widetilde{s}_n \sim s_n/n!$.* □

Corollary 3.12. *$\widetilde{s}_n \sim \dfrac{c(n)}{\lfloor n/2 \rfloor! \lceil n/2 \rceil!} \, 2^{n^2/4}$, where $c(n)$ is defined as in Section 3.2.* □

Corollary 3.13. *$\dfrac{\widetilde{s}_n}{\widetilde{s}_{n-1}} \geqslant 2^{(n+1)/2}/n$, for large enough n.* □

We now finally reach the proof of the conjecture of Cheng, Collins, and Trenk [4]:

Theorem 3.14. *Almost all unlabeled split graphs are balanced:* $\lim\limits_{n \to \infty} \dfrac{\widetilde{u}_n}{\widetilde{s}_n} = 0$.

Proof. By (2.1), $\widetilde{u}_n = \sum\limits_{k=0}^{n-1} \widetilde{s}_k$. By Corollary 3.13, for large enough n, the highest term in this sum is the last one, $\widetilde{s}_{n-1}$. Thus $\widetilde{u}_n \leqslant n\,\widetilde{s}_{n-1}$, for large enough n, and so

$$\frac{\widetilde{u}_n}{\widetilde{s}_n} \leqslant \frac{n\widetilde{s}_{n-1}}{\widetilde{s}_n} \leqslant \frac{n^2}{2^{(n+1)/2}}$$

(this last inequality uses Corollary 3.13), and $\dfrac{n^2}{2^{(n+1)/2}}$ goes to 0 as $n \to \infty$. □

References

[1] E. A. Bender, L. B. Richmond, and N. C. Wormald, Almost all chordal graphs split, *J. Austral. Math. Soc. (Series A)* **38** (1985), 214–221.

[2] F. Bergeron, G. Labelle, and P. Leroux, *Combinatorial Species and Tree-like Structures*, trans. M. Readdy, Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge, 1998.

[3] V. Bína and J. Přibil, Note on enumeration of labeled split graphs, *Comment. Math. Univ. Carolin.* **56** (2015), 133–137.

[4] C. Cheng, K. L. Collins, and A. N. Trenk, Split graphs and Nordhaus–Gaddum graphs, *Discrete Math.* **339** (2016), 2345–2356.

[5] K. L. Collins and A. N. Trenk, Finding balance: Split graphs and related classes, *Electron. J. Combin.* **25** (2018), #P1.73.

[6] A. Gainer-Dewar and I. M. Gessel, Enumeration of bipartite graphs and bipartite blocks, *Electron. J. Combin.* **21** (2014), #P2.40.

[7] M. C. Golumbic, *Algorithmic Graph Theory and Perfect Graphs*, Academic Press, New York, 1980.

[8] P. L. Hammer and B. Simeone, The splittance of a graph, *Combinatorica* **1** (1981), 275–284.

[9] P. Hanlon, The enumeration of bipartite graphs, *Discrete Math.* **28** (1979), 49–57.

[10] A. Hardt, P. McNeely, T. Phan, and J. M. Troyka, Combinatorial species and graph enumeration, undergraduate thesis (2013), `arXiv:1312.0542`.

[11] A. Joyal, Une théorie combinatoire des séries formelles, *Adv. Math.* **42** (1981), 1–82.

[12] G. F. Royle, Counting set covers and split graphs, *J. Integer Seq.* **3** (2000), Article 00.2.6.

[13] E. M. Wright, The *k*-connectedness of bipartite graphs, *J. Lond. Math. Soc. (2)* **25** (1982), 7–12.