Connectedness, classes, and cycle index

E. A. Bender, P. J. Cameron, A. M. Odlyzko and B. L. Richmond

Center for Communications Research
4320 Westerra Court
San Diego, CA 92121
U.S.A.

School of Mathematical Sciences
Queen Mary and Westfield College
Mile End Road
London E1 4NS
U.K.

AT&T Labs
180 Park Avenue
Florham Park, NJ 07932-0971
U.S.A.

Department of Combinatorics and Optimization
Faculty of Mathematics
University of Waterloo
Waterloo, Ontario N2L 3G1
Canada

Abstract

This paper begins with the observation that half of all graphs containing no induced path of length 3 are disconnected. We generalize this in several directions. First, we give necessary and sufficient conditions (in terms of generating functions) for the probability of connectedness in a suitable class of graphs to tend to a limit strictly between zero and one. Next we give a general framework in which this and related questions can be posed, involving operations on classes of finite structures. Finally, we discuss briefly an algebra associated with such a class of structures, and give a conjecture about its structure.

1 Introduction

The class of graphs containing no induced path of length 3 has many remarkable properties, stemming from the following well-known observation. Recall that an *induced subgraph* of a graph consists of a subset S of the vertex set together with all edges contained in S.

Proposition 1.1 *Let G be a finite graph with more than one vertex, containing no induced path of length 3. Then G is connected if and only if its complement is disconnected.*

Proof: It is trivial that the complement of a disconnected graph is connected. Moreover, since P_3 is self-complementary, the property of containing no induced P_3 is self-complementary. So let G be a minimal counterexample: thus, G and $\overline{G}$ are connected but, for any vertex v, either $G - v$ or $\overline{G} - v$ is disconnected. Choose a vertex v and assume, without loss, that $G - v$ is disconnected. Then v is joined to a vertex in each component of $G - v$. Since $\overline{G}$ is connected, there is a vertex x' not joined to v (in G). Let w' be a neighbour of v in the component C of $G - v$ containing x'. Then there is a path from w' to x' in C, and hence an edge wx such that $w \sim v$, $x \not\sim v$. If u is a neighbour of v in a different component of $G - v$, then $\{u, v, w, x\}$ induces P_3, contrary to assumption.

The particular view of this result we will take here is that a random P_3-free graph on more than one vertex is connected with probability $\frac{1}{2}$. This leads to the general question: *in which classes of graphs, having good notions of "connectedness" and "induced substructure", does it hold that the probability of connectedness of a random n-vertex graph in the class tends to a limit p, with $0 < p < 1$, as $n \to \infty$?* There are two questions here, since we could take either labelled or unlabelled structures.

One example is the class of forests of rooted trees, where the limiting probability of connectedness is $1/e = 0.3679\ldots$ for the labelled structures, and $1/2.997\ldots = 0.3367\ldots$ for the unlabelled structures. (The latter holds because there is a natural bijection between forests of rooted trees on n vertices, and rooted trees on $n + 1$ vertices; and $2.997\ldots$ is the exponential constant in the asymptotic formula for the number of unlabelled trees: see Otter [11]. For the labelled case, see Rényi [15].

There is no need to restrict ourselves to graphs. The probability of connectedness of a random N-free poset tends to $(\sqrt{5} - 1)/2$, in both the labelled and the unlabelled case (El-Zahar [4]; see also Stanley [18]). (Incidentally, there is no simple explanation for why the golden ratio appears here, nor for why the answer is the same in the labelled and unlabelled case.)

Contrary to what Proposition 1.1 might suggest, this question turns out to have little to do with the detailed structure of the class, but involves the rate of growth of the number of n-element structures. This will be analysed in

Section 2, where we show that a sufficient condition can be expressed in terms of convergence and smoothness properties of the generating function.

The relation between connected and arbitrary structures has several analogues, such as that between partial and total structures, or between reduced and arbitrary structures in a class whose members have a natural "congruence" relation (where a structure is reduced if the congruence is equality). These are best described in terms of two kinds of composition of classes of finite structures, "multiplication" and "substitution", defined in Section 3. The behaviour of generating functions under these operations is expressed in terms of a cycle index function, described in Section 4. The final section defines a graded algebra based on a class of structures, and gives a conjecture on this algebra and a structure theorem under additional hypotheses.

2 Convergence and smoothness

Let $\mathcal{A}$ be a class of graphs or other structures which has a notion of "connectedness"; let $\mathcal{C}$ denote the class of connected structures in $\mathcal{A}$. We assume that every member of $\mathcal{A}$ can be expressed uniquely as a disjoint union of members of $\mathcal{C}$, and that any disjoint union of members of $\mathcal{C}$ is in $\mathcal{A}$. Let c_n and C_n be the numbers of unlabelled and labelled structures in $\mathcal{C}$, and a_n and A_n the corresponding numbers for $\mathcal{A}$. (We assume that $c_0 = C_0 = 0$, $a_0 = A_0 = 1$.) As is usual in enumeration theory, we use exponential generating functions $C(z) = \sum_{n=0}^{\infty} C_n z^n/n!$ and $A(z) = \sum_{n=0}^{\infty} A_n z^n/n!$ for labelled structures, and ordinary generating functions $c(z) = \sum_{n=0}^{\infty} c_n z^n$ and $a(z) = \sum_{n=0}^{\infty} a_n z^n$ for unlabelled structures. (This notation will be used throughout this section.) With our assumptions, we have

$$A(z) = \exp(C(z)), \tag{1}$$

$$a(z) = \exp\left(\sum_{n=1}^{\infty} \frac{c(z^n)}{n}\right) = \prod_{n=1}^{\infty} (1 - z^n)^{-c_n}. \tag{2}$$

(See Wright [19]; we will derive these well-known equations in Section 5.) Now the probability of connectedness of a random n-element structure in $\mathcal{A}$ is c_n/a_n in the unlabelled case, or C_n/A_n in the labelled case. So the general question is: *What conditions on the sequence (C_n) or (c_n) guarantee that C_n/A_n or c_n/a_n tends to a limit strictly between zero and one as $n \to \infty$, where A_n and a_n are defined by the formulae above?*

Similar questions were first considered by Wright [19], who proved the following.

Theorem 2.1 *If $c_n \geq 0$ for all n, then $c_n/a_n \to 1$ if and only if*

(a) $c(x)$ has radius of convergence 0, and

(b) $\sum_{s=1}^{n-1} h_s h_{n-s} = o(h_n)$, *where* h_n *may be either* c_n *or* a_n.

The same holds for C_n *and* A_n.

In Cameron [2], it was conjectured that a necessary and sufficient condition for the probability of connectedness to tend to a limit strictly between zero and one is that the appropriate generating function has finite radius of convergence R and converges at $z = R$, and that its coefficients satisfy some "smoothness" condition. In this section, we prove a result of this form. First, we observe that the convergence condition is necessary.

Theorem 2.2 *Suppose that, with the above notation,* $C(z)$ *has finite non-zero radius of convergence* R*, and that* $C(z)$ *is unbounded on its circle of convergence. Then* $\liminf_{n\to\infty} C_n/A_n = 0$. *The analogous result holds also for* $c(z)$ *and* $a(z)$.

Proof: Consider the labelled case, and suppose that $C_n > \delta A_n$ for all n, where $\delta > 0$. Then $C(z) > \delta(A(z) - 1)$ for $0 \leq z \leq R$, and so $C(z) > \delta(\exp(C(z) - 1)$ as $z \to R$. This is clearly impossible if $C(R)$ is divergent. If $C(R)$ is convergent, then $C(z)$ is uniformly convergent for $|z| = R$.

The argument in the unlabelled case is similar.

Here is an example. Take a finite alphabet Q, with $|Q| = q$, and let $\mathcal{A}$ consist of all finite words in Q. An induced substructure is taken to be a (not necessarily consecutive) subword. (This example can be recast as a relational structure, where a word on n letters is regarded as n-set which is totally ordered and is partitioned into q subsets corresponding to the elements of Q.) A *Lyndon word* is one which is lexicographically smaller than any proper cyclic shift of itself. Now it can be shown that any word can be expressed uniquely as the concatenation, in lexicographically decreasing order, of Lyndon words. Thus, taking "disjoint union" to mean concatenation in decreasing order, we are in the general situation of this section. In this case, a_n (the number of words of length n) is equal to q^n, while c_n (the number of Lyndon words of length n) is given by the well-known formula

$$c_n = \frac{1}{n} \sum_{d|n} \mu(d) q^{n/d} \sim \frac{q^n}{n}.$$

We see directly that the radii of convergence of $a(z)$ and $c(z)$ are equal to $1/q$, and that both series diverge at $1/q$; also

$$c_n/a_n \sim 1/n \to 0.$$

Furthermore, because words are totally ordered, we have $C_n = n!c_n$ and $A_n = n!a_n$, so also $C_n/A_n \to 0$.

In addition, some smoothness condition on the coefficients is required to ensure that $\liminf(C_n/A_n) = \limsup(C_n/A_n)$ (or the analogous condition for c_n/a_n). An example to show this was given in Cameron [2]. We give two alternative definitions of "smoothness" which will work. In the analysis that follows, the arguments in the unlabelled and labelled cases are virtually identical. So we speak of the functions $c(z)$ and $a(z)$, but $C(z)$ and $A(z)$ may be substituted.

The first smoothness condition we consider is *Hayman admissibility*, defined in Hayman [8] or Odlyzko [10]. Rather than give the definition here, we quote a theorem of Hayman which frequently allows an easy proof of Hayman admissibility for generating functions in combinatorial situations.

Theorem 2.3 *(Hayman) Let $f(z)$ and $g(z)$ be admissible for $|z| < R$, $R \leq \infty$. Let $h(z)$ be analytic in $|z| < R$ and real for real z. Let $p(z)$ be a polynomial with real coefficients.*

(i) If the coefficients a_n of the Taylor series of $\exp(p(z))$ are positive for all sufficiently large n, then $\exp(p(z))$ is Hayman admissible for all z.

(ii) $\exp(f(z))$ and $f(z)g(z)$ are admissible for in $|z| < R$.

(iii) If, for some $\eta > 0$ and $R_1 < r < R$,

$$\max_{|z|=r} |h(z)| = O(f(r)^{1-\eta}),$$

then $f(z) + h(z)$ is H-admissible in $|z| < R$. In particular, $f(z) + p(z)$ is H-admissible in $|z| < R$ and, if the leading coefficient of $p(z)$ is positive, then $p(f(z))$ is H-admissible in $|z| < R$.

Hayman [8] proved the following theorem:

Theorem 2.4 *Let $f(z) = \sum f_n z^n$ be Hayman-admissible in $|z| < R$. Let*

$$\begin{aligned} a(r) &= \frac{rf'(r)}{f(r)}, \\ b(r) = ra'(r) &= \frac{rf'(r)}{f(r)} + \frac{r^2 f''(r)}{f(r)} - \left(r\frac{f'(r)}{f(r)} \right)^2. \end{aligned}$$

Then

$$f_n \sim (2\pi b(r_n))^{-1/2} f(r_n) r_n^{-n} \text{ as } n \to \infty,$$

where r_n is defined uniquely for sufficiently large n by $a(r_n) = n$. Furthermore, for all $\epsilon > 0$, we have $r_n \to R$, $f(r_n) \to \infty$, and $b(r_n) = o(f(r_n)^\epsilon)$ as $n \to \infty$.

The first result of this section is:

Theorem 2.5 *Let $c(x)$ be the generating function for the unlabelled (or labelled) connected structures in a class and let $0 < R \leq \infty$ be the radius of convergence of $c(x)$. If $\exp(c(x))$ is admissible, then $c(x) \to \infty$ diverges as $x \to R$ and the probability of connectedness of an unlabelled (or labeled) structure goes to 0 as $n \to \infty$.*

Proof: In the unlabeled case, a_n is at least as large as the coefficient of x^n in $\exp(c(x))$. Hence it suffices to work with $f(x) = \exp(c(x))$ in both the labeled and unlabeled cases and show that $c_n/a_n \to 0$.

The divergence of $f(x)$, and hence $c(x)$, at $x = R$ follows from $f(r_n) \to \infty$ in Theorem 2.4.

We have

$$\begin{aligned} f_n &\sim (2\pi b(r_n))^{-1/2} \exp(c(r_n)) r_n^{-n} \\ &> \exp((1-\epsilon)c(r_n)) r_n^{-n} \\ &> M c(r_n) c(r_n) r_n^{-n} \\ &\geq M \log(f(r_n)) c_n, \end{aligned}$$

where the last inequality follows from the fact that a sum of nonnegative terms is at least as large as a single term. Since $f(r_n) \to \infty$, the proof is complete.

The other smoothness condition we impose on $c(x)$ is satisfaction of the Flajolet–Odlyzko singularity analysis [6]. See [10], Section 11, for the definition of a function of slow variation at ∞, and for the definition of the region Δ. Flajolet and Odlyzko showed:

Theorem 2.6 *Suppose that $c(x)$ has a unique singularity at R on its circle of convergence, that the radius of convergence of $h(x)$ exceeds R, that $L(u)$ is a function of slow variation at ∞, and that*

$$c(x) - h(x) \sim (R-x)^\alpha L(1/(R-x))$$

as $x \to R$ in Δ, where α is not a non-negative integer. Then

$$c_n \sim \frac{R^{-n} n^{-\alpha-1} L(n)}{\Gamma(\alpha)}$$

as $n \to \infty$.

Remark. A similar result is proved in [6] when α is a non-negative integer: $\Gamma(\alpha)$ must be replaced by a suitable constant.

Remark. If $\alpha < 0$ and

$$L(u) = (\log u)^{\beta_1} (\log\log u)^{\beta_2} \cdots,$$

then $\exp(c(x))$ is admissible (Hayman [8]), and we can use Theorem 2.5.

We now come to the other result of this section:

Theorem 2.7 *Suppose that $C(x)$ satisfies the hypotheses of Theorem 2.6. A necessary and sufficient condition for the probability of connectedness of labelled structures in the class to have a limit strictly between 0 and 1 is that $C(x)$ converge at R.*

Suppose that $c(x)$ satisfies the hypotheses of Theorem 2.6 and has radius of convergence $R < 1$. A necessary and sufficient condition for the probability of connectedness of unlabelled structures in the class to have a limit strictly between 0 and 1 is that $C(x)$ converge at R.

Proof: ¿From Theorem 2.1, we cannot have $c_n/a_n \to 1$ if $R > 0$.

Suppose $c(R)$ diverges (that is, if $\alpha \leq 0$). As in the proof of Theorem 2.5, it suffices to consider $a(x) = \exp(c(x))$. We have

$$\begin{aligned} a_n &= [x^n]\exp(c(x)) \\ &\geq [x^n]c(x)^2/2 \\ &\sim r^{-n}n^{-2\alpha-1}L(n)^2/\Gamma(2\alpha), \end{aligned}$$

and the result follows.

Suppose now that $c(x)$ is convergent at $x = R$. Then, in Δ, we have

$$c(x) - h(x) \sim (R-x)^\alpha L(1/(R-x)),$$

where $\alpha > 0$ and $h(x)$ is analytic in $|x| < R + \delta$ for some $\delta > 0$. Note that $h(R) = c(R)$. In the labelled case, as $x \to R$ in Δ,

$$a(x) - \mathrm{e}^{h(x)} \sim \mathrm{e}^{h(R)}(R-x)^\alpha L(1/(R-x)),$$

and we conclude that

$$a_n \sim \mathrm{e}^{h(R)}c_n,$$

so that $c_n/a_n \to \mathrm{e}^{-h(R)} = \mathrm{e}^{-c(R)}$ as $n \to \infty$. The unlabelled case is similar provided $R < 1$.

For example, Cayley's Theorem for the number T_n of labelled trees on n vertices, combined with Stirling's approximation, shows that

$$T_n/n! \sim Cn^{-5/2}\mathrm{e}^n.$$

It is known that $T(z) = \sum T_n z^n/n!$ satisfies the hypotheses of Theorem 2.7, and we conclude that the probability of connectedness of a random forest tends to a limit strictly between zero and one. In fact the limit is $1/\sqrt{\mathrm{e}}$ (Rényi [15]).

It remains to prove that various natural classes of graphs satisfy "smoothness" conditions of the types described above. We pose the following problems. If $\mathcal{C}$ is a class of finite graphs, let $\mathcal{X}(\mathcal{C})$ denote the class of finite graphs containing no induced subgraph isomorphic to a member of $\mathcal{C}$. Note that, if every member of $\mathcal{C}$ is connected, then a graph lies in $\mathcal{X}(\mathcal{C})$ if and only if all its connected components do, so the analysis of this section applies (if the appropriate growth and smoothness conditions can be shown).

(a) Is it true that, if $\mathcal{C}$ is finite, then the probability of connectedness of labelled or unlabelled graphs in $\mathcal{X}(\mathcal{C})$ tends to a limit?

(b) Is it true that P_3 is the only finite connected graph H such that, in $\mathcal{X}(\{H\})$, the limiting probability of connectedness is strictly between 0 and 1?

3 Operations on classes

In this section, we propose a general framwork in which a number of questions like the probability of connectedness can be posed and studied. We work in the context of a class $\mathcal{A}$ of finite structures, which is closed under isomorphism and closed under taking induced substructures. Thus $\mathcal{A}$ may be the set of models of a universal theory in a first-order relational language. (We allow the language to have infinitely many relation symbols, but require that there are only finitely many n-element structures in $\mathcal{A}$ (up to isomorphism), for each n. As in the preceding section, we let a_n and A_n be the numbers of unlabelled and labelled n-element structures in $\mathcal{A}$, and use the ordinary generating function $\sum_{n=0}^{\infty} a_n z^n$ for the sequence (a_n), and the exponential generating function $\sum_{n=0}^{\infty} A_n z^n/n!$ for (A_n). We assume that $a_0 = A_0 = 1$.

Following Fraïssé [7], the *age* of a countable structure M is the class of all finite structures embeddable in M as induced substructures. Among structures satisfying our assumptions, ages are characterised by the *joint embedding property*: given $A, B \in \mathcal{A}$, there exists $C \in \mathcal{A}$ containing both A and B as induced substructures.

We will frequently make use of the class $\mathcal{S}$ of *sets* (without any structure). We have $S_n = s_n = 1$ for all n; so $S(z) = \exp(z)$, $s(z) = 1/(1-z)$. Other simple classes are $\mathcal{T}$, the *total orders*, with $t_n = 1$, $T_n = n!$, $t(z) = T(z) = 1/(1-z)$; and the class $\mathcal{P}$ of *permutations*, with $P_n = n!$, $p_n = p(n)$ (the partition function), $P(z) = 1/(1-z)$, $p(z) = \prod_{n=1}^{\infty} 1/(1-z^n)$. (Two permutations are isomorphic if and only if they are conjugate in the symmetric group, so the number of unlabelled permutations of an n-set is the number of partitions of n.)

Now we define two operations on classes of structures as follows. Let $\mathcal{A}$, $\mathcal{B}$ be classes. Then the operation of *multiplication* produces the class $\mathcal{A}\times\mathcal{B}$, defined as follows: a structure in the class with point set X consists of a partition of X into two parts Y, Z (possibly empty), with an $\mathcal{A}$-structure on Y and a $\mathcal{B}$-structure on Z.

Proposition 3.1 *If $\mathcal{C} = \mathcal{A} \times \mathcal{B}$, then $C(z) = A(z)B(z)$ and $c(z) = a(z)b(z)$.*

For example, if $\mathcal{D}$ is the class of *derangements* (permutations with no fixed points), then $\mathcal{P} = \mathcal{D} \times \mathcal{S}$, from which we obtain the exponential generating function for derangement numbers: $D(z) = 1/((1-z)\exp(z))$.

The operation of *substitution* of $\mathcal{B}$ into $\mathcal{A}$ produces the class $\mathcal{A}[\mathcal{B}-1]$, defined as follows: a structure in the class on the point set X consists of a partition of X into an arbitrary number of *non-empty* parts, a $\mathcal{B}$-structure on each part, and an $\mathcal{A}$-structure on the set of parts. The -1 in the notation is intended to suggest that we remove the empty structure from $\mathcal{B}$ before performing the substitution.

Proposition 3.2 *If $\mathcal{C} = \mathcal{A}[\mathcal{B} - 1]$, then $C(z) = A(B(z) - 1)$.*

The function $c(z)$ cannot be determined from $a(z)$ and $b(z)$ alone, as we see in the next section.

For example, if $\mathcal{C}$ denotes the class of *cyclic orders*, then the cycle decomposition of a permutation can be expressed as $\mathcal{P} = \mathcal{S}[\mathcal{C} - 1]$, so that $P(z) = S(C(z)-1)$, in agreement with the values calculated above for $P(z)$ and $S(z)$ and the fact that $C(z) = 1 - \log(1 - z)$ (from $C_n = (n-1)!$ for $n \geq 1$).

Now let $\mathcal{A}$ be a class of graphs closed under disjoint unions, and let $\mathcal{C}$ be the class of connected graphs in $\mathcal{A}$. We have $\mathcal{A} = \mathcal{S}[\mathcal{C} - 1]$. So the problem of the probability of connectedness is an instance of the following more general problem:

Problem: Suppose that $\mathcal{A}$ is obtained from $\mathcal{B}$ and $\mathcal{C}$ by some operation such as multiplication or substitution. Under what conditions does it hold that C_n/A_n or c_n/a_n tends to a limit strictly between zero and one?

We give two further examples.

If $\mathcal{A} = \mathcal{S} \times \mathcal{C}$, then $\mathcal{A}$-structures can be regarded as *partial $\mathcal{C}$-structures*, consisting of a set with a $\mathcal{C}$-structure on a subset. For example, if $\mathcal{A}$ is a class of graphs closed under adding isolated vertices, and $\mathcal{C}$ is the class of members of $\mathcal{A}$ with no isolated vertices, then this relation holds. So our question would be: When does it occur that the proportion of partial structures which are total tends to a limit between zero and one? This is discussed in Cameron [2].

Suppose that $\mathcal{A} = \mathcal{C}[\mathcal{S}-1]$. Then an $\mathcal{A}$-structure carries a natural equivalence relation or *congruence* $\equiv$, such that if a relation holds for an n-tuple of points, then it remains true if some or all of the points are replaced by equivalent ones. With this interpretation, $\mathcal{C}$ is the class of *reduced* structures, those in which the relation $\equiv$ is just equality. For example, in a suitable class $\mathcal{A}$ of graphs, set $v \equiv w$ if v and w have the same neighbour sets; a graph is reduced if different vertices have different neighbour sets. Now our question is: When does it occur that the proportion of structures which are reduced tends to a limit between zero and one?

4 Cycle index

It is possible, following Joyal [9], to define a *cycle index* of a class of structures such that the generating functions for labelled and unlabelled structures defined above are specialisations of it. Its behaviour under multiplication and substitution can also be described.

Recall that the *cycle index* $z(g)$ of a permutation g on n letters is the monomial in indeterminates $s_1, \ldots, s_n$ given by

$$z(g) = s_1^{c_1(g)} s_2^{c_2(g)} \cdots,$$

where $c_i(g)$ is the number of cycles of length i in the cycle decomposition of g. If G is a permutation group on n letters, the *cycle index* of G is the average of the cycle indices of its elements:

$$Z(G) = \frac{1}{|G|} \sum_{g \in G} z(g).$$

Now if $\mathcal{G}$ is a class of finite permutation groups, containing only finitely many members of degree n for each n, we define the *cycle index* of $\mathcal{G}$ by $\mathcal{Z}(\mathcal{G}) = \sum_{G \in \mathcal{G}} Z(G)$. (This is a formal power series in infinitely many indeterminates; but the assumption guarantees that each monomial occurs only finitely often in the sum.)

This definition can be extended in two ways. First, let $\mathcal{A}$ be a class of structures, as in the preceding section. We define its *cycle index* to be $\mathcal{Z}(\mathcal{A}) = \mathcal{Z}(\{\mathrm{Aut}(A) : A \in \mathcal{A}\})$, where each unlabelled structure in $\mathcal{A}$ is used once in the sum. Now the generating functions for $\mathcal{A}$ are specializations of the cycle index. If Φ is a formal power series, we let $\Phi(s_i \leftarrow t_i)$ be the result of making the substitution t_i for s_i for all i. Some conditions are required in general in order that this is well-defined. For example, if the t_i are formal power series in one indeterminate z, it suffices that t_i has no term of degree less than i.

Proposition 4.1 *(i)* $A(z) = \mathcal{Z}(\mathcal{A})(s_i \leftarrow z^i)$.

(ii) $a(z) = \mathcal{Z}(\mathcal{A})(s_1 \leftarrow z, s_i \leftarrow 0 \textit{ for } i > 1)$.

The cycle index behaves as follows under multiplication and substitution:

Proposition 4.2 *(i)* $\mathcal{Z}(\mathcal{A} \times \mathcal{B}) = \mathcal{Z}(\mathcal{A})\mathcal{Z}(\mathcal{B})$.

(ii) $\mathcal{Z}(\mathcal{A}[\mathcal{B} - 1]) = \mathcal{Z}(\mathcal{A})(s_i \leftarrow t_i - 1)$, *where* $t_i = \mathcal{Z}(\mathcal{B})(s_j \leftarrow s_{ij})$.

¿From the second part of this proposition, we obtain the missing formula for the generating function for unlabelled structures in $\mathcal{A}[\mathcal{B} - 1]$:

Proposition 4.3 *If* $\mathcal{C} = \mathcal{A}[\mathcal{B} - 1]$, *then* $c(z) = \mathcal{Z}(\mathcal{A})(s_i \leftarrow b(z^i) - 1)$.

For the class $\mathcal{S}$ we have

$$\mathcal{Z}(\mathcal{S}) = \exp\left(\sum_{n=1}^{\infty} \frac{s_n}{n}\right).$$

¿From this, the formula of Section 2 relating connected and arbitrary graphs in a class closed under taking disjoint unions follows.

The other extension is to a class of infinite permutation groups. A permutation group G on a set X is called *oligomorphic* if the number of orbits of G on the set of n-tuples of points of X is finite for all n. See Cameron [1] for an account of these groups.

Clearly, the previous definition of the cycle index of an infinite permutation group makes no sense. However, an oligomorphic permutation group G has a so-called *modified cycle index* $\tilde{Z}(G)$, defined as follows. Choose representatives for the orbits of G on finite subsets of X. (By assumption, there are only finitely many of each size.) For each representative Y, let $G(Y)$ be the permutation group induced on Y by its setwise stabiliser in G. Then $\tilde{Z}(G)$ is defined to be the cycle index of this collection of finite permutation groups.

The relationship with the preceding is as follows. A structure M is said to be *homogeneous* if any isomorphism between finite substructures of M extends to an automorphism of M. Examples of homogeneous structures include the pentagon, the rational numbers $\mathbb{Q}$ (as ordered set), and the random graph or Rado's graph [5], [14]. A theorem of Fraïssé [7] characterizes the ages of countable homogeneous structures. In particular, a class $\mathcal{A}$ of structures satisfying our conditions (that is, closed under isomorphism and under induced substructures and containing only finitely many n-element structures up to isomorphism) is the age of a countable homogeneous structure if and only if it satisfies the *amalgamation property:* if $B, C \in \mathcal{A}$ have isomorphic substructures A, A' respectively, then there is a structure $D \in \mathcal{A}$ in which B and C can be embedded in such a way that the substructures are identified or "glued together" according to the isomorphism.

Proposition 4.4 *Let M be a countable homogeneous structure with automorphism group G and age $\mathcal{A}$. Then*

(i) the number of orbits of G on n-element subsets is equal to the number of unlabelled n-element structures in $\mathcal{A}$;

(ii) the number of orbits of G on n-tuples of distinct elements is equal to the number of labelled n-element structures in $\mathcal{A}$;

(iii) $\tilde{Z}(G) = \mathcal{Z}(\mathcal{A})$.

This result gives a covenient translation between ages satisfying the amalgamation property and oligomorphic permutation groups. Under this translation, multiplication and substitution of ages correspond to the direct product (in its intransitive action) and the wreath product (in its imprimitive action) of permutation groups.

5 Algebras

In this section, a graded algebra is associated with a class of finite structures. We make a conjecture about its structure, and give a structure theorem under additional hypotheses. See Cameron [3] for more details of the latter.

Let $\mathcal{A}$ be a class of finite structures satisfying our usual conditions (closed under isomorphism and under induced substructures, and containing only finitely many n-element structures up to isomorphism). For each n, let V_n denote the $\mathbb{Q}$-vector space of all isomorphism-invariant rational functions on the set of n-element structures in $\mathcal{A}$. Thus, $\dim(V_n) = a_n$; a basis for V_n consists of the characteristic functions of the isomorphism classes of n-element structures.

We define

$$\mathrm{Alg}(\mathcal{A}) = \bigoplus_{n=0}^{\infty} V_n,$$

and define a product as follows. Take $f \in V_n$, $g \in V_m$. Then fg is the function in V_{n+m} whose value on the $(n+m)$-element structure $X \in \mathcal{A}$ is given by

$$(fg)(X) = \sum_{\substack{Y \subseteq X \\ |Y|=n}} f(Y)g(X \setminus Y).$$

This multiplication is extended linearly to the whole of $\mathrm{Alg}(\mathcal{A})$. The algebra is easily seen to be commutative and associative. An element of $\mathrm{Alg}(\mathcal{A})$ is said to be *homogeneous of degree n* if it is contained in V_n.

The construction behaves well with respect to multiplication of classes:

Proposition 5.1

$$\mathrm{Alg}(\mathcal{A} \times \mathcal{B}) = \mathrm{Alg}(\mathcal{A}) \otimes_{\mathbb{Q}} \mathrm{Alg}(\mathcal{B}).$$

For $\mathcal{A} = \mathcal{S}$, we have $\dim(V_n) = 1$ for all n, and in fact $\mathrm{Alg}(\mathcal{S})$ is a polynomial algebra in one variable, the generator being the function in V_1 taking the value 1 on all singleton sets.

Conjecture: Suppose that $\mathcal{A}$ has the following property: for any $A, B \in \mathcal{A}$, there exists $C \in \mathcal{A}$ in which A and B can be embedded as disjoint substructures. Then $\mathrm{Alg}(\mathcal{A})$ is an integral domain (i.e., has no divisors of zero).

We make some remarks about this conjecture. First, the condition is clearly necessary. For, if A and B cannot be embedded disjointly in any $\mathcal{A}$-structure, then $f_A f_B = 0$, where f_A is the characteristic function of the isomorphism class of A. Also, the condition is a strengthening of the joint embedding property; so a class $\mathcal{A}$ satisfying it is the age of a countably infinite structure M. In this case, $\mathrm{Alg}(\mathcal{A})$ is a subalgebra of the *reduced incidence algebra* of the poset of finite subsets of M (Rota [17]).

If a graded algebra is a polynomial algebra generated by homogeneous elements, then the relation between the sequence enumerating the polynomial generators by degree and the sequence of dimensions of the homogeneous components is identical to the relation between the sequences enumerating unlabelled conneced and arbitrary structures, met with at the start of Section 2. This observation motivates the following result, taken from Cameron [3].

Theorem 5.2 *Let $\mathcal{A}$ be a class of structures. Suppose that $\mathcal{A}$ possesses*

(i) a subclass of "connected" structures;

(ii) a partial order $\leq$ of "involvement" on the set of n-element structures for each n;

(iii) a commutative and associative "composition" $\circ$ such that $|A \circ B| = |A| + |B|$.

Assume:

(i) Any structure in $\mathcal{A}$ is uniquely the composition of connected structures.

(ii) If $A \in \mathcal{A}$ is partitioned into substructures $A_1, A_2, \ldots$, then $A_1 \circ A_2 \circ \cdots \leq A$.

Then $\mathrm{Alg}(\mathcal{A})$ *is a polynomial algebra generated by the characteristic functions of the isomorphism classes of connected structures in $\mathcal{A}$.*

The conditions of the theorem are satisfied when $\mathcal{A}$ is the class of all graphs; we take "connected" to have its usual meaning, "involvement" to mean "spanning subgraph", and "composition" to be "disjoint union". A more unusual example involves the words considered in Section 2. As there, a word is "connected" if it is a Lyndon word (smaller than any proper cyclic shift of itself); "involvement" is lexicographic order, reversed; and "composition" is concatenation in lexicographically decreasing order. Now the hypotheses of the Theorem are satisfied. The algebra $\mathrm{Alg}(\mathcal{A})$ is the *shuffle algebra*, which occurs in the theory of free Lie algebras (see Reutenauer [16]), and which was shown to be a polynomial algebra by Radford [13].

References

[1] P. J. Cameron, *Oligomorphic Permutation Groups*, London Math. Soc. Lecture Notes **152**, Cambridge University Press, Cambridge, 1990.

[2] P. J. Cameron, On the probability of connectedness, *Discrete Math.*, submitted.

[3] P. J. Cameron, The algebra of an age, in preparation.

[4] M. H. El-Zahar, Enumeration of ordered sets, pp. 327–352 in *Algorithms and Order* (ed. I. Rival), Kluwer, Dordrecht, 1989.

[5] P. Erdős and A. Rényi, Asymmetric graphs, *Acta Math. Acad. Sci. Hungar.* **14** (1963), 295–315.

[6] P. Flajolet and A. M. Odlyzko, Singularity analysis of generating function, *SIAM J. Discrete Math.* **3** (1990), 216–240.

[7] R. Fraïssé, Sur certains relations qui généralisent l'ordre des nombres rationnels, *C. R. Acad. Sci. Paris* **237** (1953), 540–542.

[8] W. K. Hayman, A generalisation of Stirling's formula, *J. Reine Angew. Math.* **196** (1956), 67–95.

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

[10] A. M. Odlyzko, Asymptotic enumeration methods, pp. 1063–1229 in *Handbook of Combinatorics* (ed. R. Graham, M. Grötschel and L. Lovász), Elsevier, Amsterdam, 1995.

[11] R. Otter, The number of trees, *Ann. Math.* (2) **49** (1948), 583–599.

[12] G. Pólya, Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen, *Acta Math.* **68** (1937), 145–254.

[13] D. E. Radford, A natural ring basis for the shuffle algebra and an application to group schemes, *J. Algebra* **58** (1979), 432–454.

[14] R. Rado, Universal graphs and universal functions, *Acta Arith.* **9** (1964), 331–340.

[15] A. Rényi, Some remarks on the theory of trees, *Publ. Math. Inst. Hungar. Acad. Sci.* **4** (1959), 73–85.

[16] C. Reutenauer, *Free Lie Algebras*, Oxford University Press, Oxford, 1994.

[17] G.-C. Rota, On the foundations of combinatorial theory, I: Theory of Möbius functions, *Z. Wahrscheinlichkeitstheorie* **2** (1964), 340–368.

[18] R. P. Stanley, Enumeration of posets generated by disjoint unions and ordinal sums, *Proc. Amer. Math. Soc.* **45** (1974), 295–299.

[19] E. M. Wright, A relationship between two sequences, I, *Proc. London Math. Soc.* (3) **17** (1967), 296–304; II, *ibid.* (3) **17** (1967), 547–552; III, *J. London Math. Soc.* (1) **43** (1968), 720–724.