Asymptotic Enumeration of Predicate-Junction Flowgraphs

C. Cooper
School of Mathematical Sciences
University of North London
London N7 8DB

October 16, 2005

Abstract

We consider unlabelled flowgraphs for a model of binary logic without the constraints of structured programming. The number of such flowgraphs is asymptotic to $(3.4n)^{n/2}$, where n is the number of nodes in the flowgraph. This is to be compared with bounds of between $(8.8)^{n/2}$ and of $(9.8)^{n/2}$ for unlabelled structured flowgraphs of the Böhm and Jacopini type.
Of the space of flowgraphs we study, 41% are prime, that is contain no proper subflowgraphs. The main obstructions to primality being the Dijkstra-structures, which are based on *If_Then_Else* and *Do_While* constructs.

1 Introduction

Flowgraphs are an abstraction of program flowcharts and model the logical structure of computer programs. The vertices of a flowgraph represent points of branching or reconvergence in the program logic, the arcs, transfer of control. The topic of flowgraphs is of importance in software engineering and is extensively treated in [Fe], [FH],[He] and [Zu].

The structure of the flowgraph depends on the type of conditional statements provided by the programming language, or allowed by the style of programming imposed. The simplest type of branching used in a computer program consists of conditional statements of the *If* type, which evaluate to either *True* or *False*. Typically, the flowgraph models this binary logic by binary branching, and reconvergence.

The unrestricted combining of conditional and *Go_To* statements results in a style of programming which is referred to as unstructured. The debate on the virtues of structured as opposed to unstructured programming, was strongly influenced by Dijkstra in the paper

Go to statement considered harmful, [Di]. This paper advocates the replacement of *Go_To* statements by *Do_While* and *If_Then_Else* constructs which are known as *D-structures*, and the flowgraphs arising from this style of programming are known as *D-charts*. The generalization of this type of structured programming gives rise to *Böhm and Jacopini charts of type m* or BJ_m-charts [BJ]. The BJ_1-charts are the *D-charts*, and the higher level BJ_m-charts use *Do_While_With_(at most)_m_Exits*, as advocated by Knuth [Kn].

Enumeration of unlabelled flowgraphs is of interest in determining the number of intrinsically different program structures arising from the model of programming logic under consideration. In the paper *Enumeration of structured flowcharts*, Bender and Butler [BB] studied the number of unlabelled BJ_m flowgraphs. They found the number of unlabelled BJ_1-charts was asymptotic to 0.3888 $k^{-3/2}(8.849)^k$ increasing to $0.3808k^{-3/2}(9.742)^k$ for BJ_∞-charts. Here k is the number of *decision nodes* (binary branch points) in the program. In our paper, results are expressed in terms of the *total number n*, of nodes in the flowgraph, and thus $n = 2k$. Our results, like those in [BB] are asymptotic in the number n.

Using the methodology of random graphs, we enumerate the number of unlabelled flowgraphs for the simple model of binary logic considered in [BB] but without the restrictions of structuring. No branching compatible with the logic, including *Go_To* statements, is excluded. As a result, the number of flowgraphs on n nodes increases from about $(8.9)^{n/2}$ to about $(3.4n)^{n/2}$; the precise value being stated in Theorem 1. Thus the structured $BJ-$charts are found to represent a vanishingly small proportion of the totality of these flowgraphs.

Many flowgraphs in our model do however have some structuring, mainly in the form of isolated D-structures. The concept of *prime flowgraphs* [FH] is used to indicate the absence of local structure; prime flowgraphs being essentially those with no stand-alone subprograms, a precise definition being given in Section 2. For our model, a large proportion of flowgraphs (41%) are prime, where this result follows from Theorem 1. The main obstructions to primality are the D-structures, the proportion of other obstructions tending to zero. In the limit, 59% of the flowgraphs have some D-structures, whilst all others are prime.

There is a wide range of algorithms for the analysis or manipulation of flowgraphs in the literature of static analysis. Typical applications include complexity measurement [Fe], optimization [He] and restructuring [LMW]. It is of obvious benefit to have an effective method of producing random flowgraphs for algorithm testing. For the model we study here, the flowgraphs may be generated by rejection sampling, with a success probability of about 1/9 at each sampling.

2 Flowgraphs

It is worth noting that in general, flowgraphs may have loops and parallel edges. The flowgraph models only the branching structure of the program, so that an edge may represent many sequentially executed statements in the actual program. Subdividing the edges with vertices of degree two, to model these sequential statements, is not considered here, although the definition of flowgraphs does not prevent it.

Let $G = (V, E)$ be a digraph, and let a and z be vertices of G such that

F1) z has out-degree 0,

F2) Every $v \in V$ is on some directed (a, z)-walk,

then $F = (G, a, z)$ is a *labelled flowgraph.* The distinguished vertices a and z represent the beginning and end of the program, respectively. The addition of a *back edge* $e = (z, a)$ to a flowgraph F, gives a *pre-flowgraph* $D = F + e$. It follows from property F2) that D is strongly connected.

Let G be a digraph, and let G' be a sub-digraph of G. Then $F' = (G', a', z')$ is a *sub-flowgraph* of $F = (G, a, z)$ if

S1) F' is a flowgraph,

S2) Only a' and z' may have greater in-degree in G than G',

S3) If a is in G' then $a = a'$ or $a = z'$,

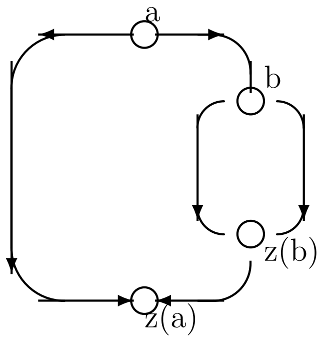
S4) Only z' may have greater out-degree in G than G'.

A sub-flowgraph is *trivial* is it has at most one edge. A *prime flowgraph*, is a flowgraph with no non-trivial proper sub-flowgraphs. We may view a sub-flowgraph as an integral piece of stand-alone logic (a subprogram), and a prime flowgraph as one which has no non-trivial sub-logics (subprograms).

As an example, if we consider *If_Then_Else* code implemented in the form *if...then...else... endif*, so that each decision node x (*if*) has a reconvergent node $z(x)$ (*endif*), then the program

if a then (if b then p else q endif) else r endif

has flowgraph $F = (G, a, z(a))$ shown in the figure below.



In this example F is not prime because of the sub-digraph induced by vertices $b, z(b)$. The pre-flowgraph of F is $G + (z(a), a)$.

We restrict our attention to a simple model of flowgraphs, the *predicate-junction flowgraphs,* discussed in general in [LMW]. All vertices other than the distinguished vertices are either *predicate nodes* with in-degree 1 and out-degree 2 or *junction nodes* with in-degree 2 and out-degree 1. The distinguished nodes a and z both have degree 2. The addition of a back edge $e = (z, a)$ to such a flowgraph F gives a pre-flowgraph D in which *every vertex* is either a predicate or a junction node. The node z is always a junction node in the pre-flowgraph by F1), but the node a may be of either type. Loops cannot occur in this model, as the branching and reconvergent aspects of any conditional statement have been separated.

We consider all possible flowgraphs with n vertices, conforming to this model. We will call these flowgraphs *unstructured* as no restrictions of structured programming are imposed. The model used by [BB] is, with minor modifications, a subset of this unstructured predicate-junction model, consisting of those flowgraphs which are BJ_m-charts.

Let D be any labelled (multi-)digraph with vertex set $V = [n]$, n even, where all vertices of D have either $d^-(v) = 1, d^+(v) = 2$ or $d^-(v) = 2, d^+(v) = 1$. The degree sum shows there are $n/2$ vertices of each type. We shall call such a digraph a *predicate-junction digraph* (PJD), and denote by $\mathcal{D}$ the set of all such PJD's.

Let $\mathcal{D}_F$ denote the strongly connected elements of $\mathcal{D}$. Let $\mathcal{F}$ be the set of flowgraphs $(D - e, a, z)$ obtained by the deletion of any edge $e = (z, a)$ from any junction node z of any element $D \in \mathcal{D}_F$. Then $\mathcal{D}_F$ and $\mathcal{F}$ are exactly the pre-flowgraphs and flowgraphs of the unstructured predicate-junction model.

Denote by $\mathcal{P}$ the subset of $\mathcal{F}$ whose elements are prime flowgraphs, and by $\mathcal{D}_P$ their PJD's. Let $\mathcal{U}_F$ (resp. $\mathcal{U}_P$) be the set unlabelled flowgraphs (resp. prime flowgraphs) arising from $\mathcal{F}$ (resp. $\mathcal{P}$).

Theorem 1 *For sufficiently large n, the number of unlabelled, unstructured predicate-junction flowgraphs is given by*

$$|\mathcal{U}_F| = \frac{(\frac{3}{2}n)!}{2^n(\frac{n}{2}!)^2} n \frac{e^{2/9}}{18}(1 + o(1)),$$

$$|\mathcal{U}_P| = \frac{(\frac{3n}{2})!}{2^n(\frac{n}{2}!)^2} n \frac{e^{-2/3}}{18}(1+o(1)).$$

The primality of predicate-junction flowgraphs has been studied by Prather [Pr]. The following lemma gives a sufficient condition for a predicate-junction flowgraph to be prime.

Lemma 2 (Pr) *Let* $D \in \mathcal{D}_F$ *and let* $F = D - (z, a)$*, where* z *is a junction node of* D*. If the underlying graph of* D *is 3-connected then* F *is a prime flowgraph.*

Proof. Suppose that D is $3-$connected and $F = (G, a, z)$ is not prime. Let $F' = (G', a', z')$ be a non-trivial sub-flowgraph of F. By S2) and S4) G' is a subgraph of G with *at most two* vertices of attachment, namely a' and z'. If z is in G' then $z = z'$ as $d^+(z) = 0$ by F1). Thus by the 3-connectivity of D, a' and z' are the only vertices in G'.

If G' has two vertices, but is non-trivial, then by S1) and F1), G' is an *If_Then_Else* construct with two edges directed from a' to z'. Let (z', w) be the out-edge of z' in D, then the removal of a' and w disconnects z', which is a contradiction.

□

3 Predicate-junction digraphs

We restrict our set $\mathcal{D}$ of PJD's in the following manner. Given the vertex set $[n]$, we consider vertices $B = \{i : i = 1...n/2\}$ to be predicate or *black* vertices and $W = \{i : i = n/2+1, ..., n\}$ to be junction or *white* vertices. This restriction on the labelling is unimportant, as our final goal is the unlabelled flowgraphs.

We generate the PJD's indirectly using a *configuration model* (see Bollobás [Bo1]). With each vertex $i \in [n]$, we associate two sets U_i and V_i, where $|U_i| = d^+(i)$ and $|V_i| = d^-(i)$. Thus for $i = 1, ..., n/2$, the predicate (black) nodes, $U_i = \{a_{i,1}, a_{i,2}\}$ and $V_i = \{b_{i,1}\}$ whereas for $i = n/2+1, ..., n$, the junction (white) nodes, we have $U_i = \{a_{i,1}\}$ and $V_i = \{b_{i,1}, b_{i,2}\}$.

A *configuration* C is a matching of $V = \cup V_i$ in random order against $U = \cup U_i$ arranged in the order $a_{1,1}, a_{1,2}, ..., a_{n,1}$. If $b \in V$ is matched against $a \in U$ then (a, b) is regarded as an edge of the bipartite graph $B(C)$ with bipartition (U, V). The underlying digraph $D(C)$ (resp. bipartite digraph $B(D(C))$), of the configuration C is obtained by identifying U_i, V_i with the vertex i (resp. separate copies of the vertex i), the direction of the edges being from U to V.

Let $\mathcal{C}$ be the set of all such configurations. The set $\mathcal{D}$ of PJD's is the set of underlying (multi-)digraphs of $\mathcal{C}$. We work with the configuration space $\mathcal{C}$ for most of the proofs, and say $C \in \mathcal{C}$ has property Q if the underlying digraph $D(C)$ has Q. When we refer to the probability of Q the counting is done in $\mathcal{C}$. The subset of $\mathcal{C}$ corresponding to $\mathcal{D}_Q$ (the digraphs with property Q) is denoted by $\mathcal{C}_Q$.

Say a cycle in a PJD is *monochromatic* if all vertices on the cycle have the same out-degree (colour). Prather [Pr] noted that monochromatic cycles were an obstruction to strong connectivity, as (eg) all black cycles M have no arcs directed into them from $V - M$. Note that any loops present in a PJD are monochromatic. The following theorem is proved in the next section.

Theorem 3 *Conditional on the absence of monochromatic cycles, a.e.*[1] $C \in \mathcal{C}$ *(resp.* $D \in \mathcal{D}$*) is strongly connected.*

We must now explain how to prove Theorem 1.

(a) It follows from Theorem 3 that if $\mathcal{C}_N$ are the configurations with *no monochromatic cycles* and $\mathcal{C}_F$ are the *strongly connected* configurations, then $\mathcal{C}_F \subset \mathcal{C}_N$ and $|\mathcal{C}_F| = |\mathcal{C}_N|(1 - o(1))$. Thus we can a.a. identify those configurations whose digraph is a pre-flowgraph.

(b) Let $i \in B, j \in W$, then define the following

(T_I) *Type I obstruction.* Two edges from i to j,

(T_{II}) *Type II obstruction.* Bichromatic cycle length 2 on i and j.

The derived measure of the digraphs $\mathcal{D}_F$ is uniform conditional on the number of Type I obstructions. Thus we estimate $|\mathcal{D}_F|$ using Lemma 4.

(c) In the strongly connected configurations $\mathcal{C}_F$, there are two major obstructions to primality. These are simple *If_Then_Else* (ITE) and *Do_While* (DW) subgraphs.

A Type I obstruction with edges $e_1 = (i, j), e_2 = (i, j)$ defines an ITE sub-flowgraph $F' = (\{e_1, e_2\}, i, j)$. A Type II with edges $e_1 = (i, j), e_2 = (j, i)$ and the additional out-edge $e_3 = (i, k)$ of the black vertex i, defines a DW sub-flowgraph $F' = (\{e_1, e_2, e_3\}, j, k)$.

If $\mathcal{C}_T$ are the configurations with Type I or II obstructions, $\mathcal{C}_T \cap \mathcal{C}_F$ gives digraphs which are *non-prime flowgraphs* on the removal of *the out-edge from any white vertex.*

(d) By Lemma 8, a.e. configuration in $\mathcal{C}_F \setminus \mathcal{C}_T$ is 3-connected in the underlying graph, defining a property $P3$. Thus $\mathcal{C}_{P3} \subset \mathcal{C}_F \setminus \mathcal{C}_T$ and $|\mathcal{C}_{P3}| = |\mathcal{C}_F \setminus \mathcal{C}_T|(1 - o(1))$. By Lemma 2, all flowgraphs derived from $\mathcal{C}_{P3}$ are prime. The derived measure is uniform, due to the absence of Type I obstructions. Thus we estimate $|\mathcal{D}_P| = |\mathcal{D}_{N \setminus T}|(1 - o(1))$ from Lemma 4.

(e) By Lemma 9, the digraphs of $\mathcal{D}_F$ a.a. have trivial automorphism group and this condition persists on removal of any back edge $(z, a), z \in W$ to give the flowgraphs $\mathcal{F}$.

---

[1] Throughout this paper, the terms *almost every* (a.e.) and *almost always* (a.a.) refer to events in a probability space of graphs, digraphs or configurations, which occur with probability tending to 1 as n, the number of vertices of the graphs (etc) tends to infinity.

(f) Combining Lemma 4 and the factor $|\mathcal{U}_X| = \frac{n}{2(\frac{n}{2}!)^2}|\mathcal{D}_X|$, $X = F, P3$ from Lemma 9 we estimate $|\mathcal{U}_F|$, $|\mathcal{U}_P|$ as required. This completes the proof of Theorem 1.

We now give the asymptotic formulae for what is essentially the number of labelled pre-flowgraphs and prime pre-flowgraphs derived from our configuration model.

Lemma 4 *Let $\mathcal{D}_N$ denote those digraphs of $\mathcal{D}$ with no monochromatic cycles, and $\mathcal{D}_{N\setminus T}$ those with no monochromatic cycles or Type I or II obstructions, then*

$$|\mathcal{D}_N| \sim \frac{(\frac{3n}{2})!}{2^n}\frac{e^{2/9}}{9}, \qquad |\mathcal{D}_{N\setminus T}| \sim \frac{(\frac{3n}{2})!}{2^n}\frac{e^{-2/3}}{9}.$$

Proof. Let $C \in \mathcal{C}$, and let M, T_1, T_2 be the number of monochromatic cycles, and obstructions of Type I, II (resp.) in C. Then M, T_1, T_2 are asymptotically independently Poisson distributed with parameters, $m = \log 9$, $t_i = 2^i/9, \quad i = 1, 2$. (We note that all logarithms in this paper are to the base e.) The expected number of monochromatic cycles length c is

$$2\binom{n/2}{c}(c-1)!2^c\frac{(\frac{3n}{2}-c)!}{(\frac{3n}{2})!} = \frac{2}{c}\left(\frac{2}{3}\right)^c(1+o(1))$$

thus

$$m \sim 2\sum_{c=1}^{\infty}\frac{1}{c}\left(\frac{2}{3}\right)^c = \log 9.$$

Similar calculations can be made for t_1, t_2. As configurations are random permutations, techniques in the Poisson paradigm for establishing the convergence to independent Poisson distributions of the small cycle structure, can be applied to the random variables M, T_1, T_2. The reader is referred to [BHJ] for a general discussion, and to [AGG] and [BHJ] for specific relevant examples.

Thus $|\mathcal{C}_N| \sim |\mathcal{C}|/9$, and $|\mathcal{C}_{N\setminus T}| \sim |\mathcal{C}|e^{-2/3}/9$, where $|\mathcal{C}| = \left(\frac{3n}{2}\right)!$.

The number of Type I obstructions conditions the number of distinct configurations which give rise to a given D. These obstructions appear as cycles of length 2 in $B(D)$ the bipartite graph of D.

Let $D \in \mathcal{D}_{N\setminus T}$, then $B(D)$ consists of paths, and cycles of length at least 4. There are n vertices of degree 2 in $B(D)$. The configuration vertices can be assigned unambiguously to each such vertex in 2 ways. Thus D has 2^n configurations $C(D)$ in $\mathcal{C}$.

If $D \in \mathcal{D}_N$, then $B(D)$ may have j cycles of length 2 corresponding to Type I obstructions. There are only two ways to assign the four configuration vertices to such a cycle, so D then has 2^{n-j} configurations $C(D)$ in $\mathcal{C}$. If $\mathcal{C}_{N,j}$ denotes configurations with no monochromatic cycles and exactly j Type I obstructions then

$$|\mathcal{C}_{N,j}| \sim \frac{e^{-2/9}}{9}\frac{(3n/2)!}{2^n}\frac{(2/9)^j}{j!}.$$

Thus

$$|\mathcal{D}_N| \sim \frac{e^{-2/9}}{9}\frac{(3n/2)!}{2^n}\sum_{j\geq 0}\frac{2^j(2/9)^j}{j!}.$$

□

4 Strong Connectivity

Let $d_X^-(v)$ (resp. $d_X^+(v)$) be the number of arcs from (resp. to) the set of vertices X to (resp. from) the vertex v. If X is a graph we refer to the vertex set $V(X)$.

Among the strongly connected components of any digraph there will be at least one, H say, with out-degree zero. (see [BM] Ex 10.1.9). Denote the vertex set of H by S, so now $N^+(S) = \emptyset$. Let $T = V \setminus S$. If $A \subset S \cap W$ say A is an *interface (from T to S)* if $d_T^-(v) = 1$ for all $v \in A$. Similar definitions apply to components of in-degree zero.

Lemma 5 *Let $D \in \mathcal{D}$, and let H be a strongly connected component of D, with vertex set S, such that $N^+(S) = \emptyset$. Then H consists of an interface A, and two (possibly empty) sets $B_1 \subset B, W_1 \subset W$ of the same size.*

Proof. Let $T = V - S$ and $A = \{v \in S : d_T^-(v) > 0\}$, and let $|A| = a$. As H is strongly connected, we know that $d_H^-(v), d_H^+(v) > 0$ so $A \subset W$ and $d_T^-(v) = 1$ if $v \in A$. Thus A is an interface from T to S, and moreover $|N^-(S)| = a$.

Let $k = |B \cap S|, l = |W \cap S|$ and $i = |E(H)|$.

Then $d^-(H) = i + a = k + 2l$, $d^+(H) = i = 2k + l$, so $l - a = k$ as required.

□

Such a strongly connected component H is characterized by two parameters, $a = |A|$ and $b = |B_1| = |W_1|$. H has $a + 2b$ vertices and $a + 3b$ internal edges. We will say H is a strongly connected component with parameters a, b. If D is not strongly connected then at least one such component must have at most $n/2$ vertices. We note that monochromatic cycles correspond to such components in the special case where $b = 0$.

Lemma 6 *Let $C \in \mathcal{C}$*

(i) the expected number $E(a, b)$ of strongly connected H in C with parameters a, b is given by

$$E(a,b) = \binom{n/2}{a+b}\binom{a+b}{a}\binom{n/2}{b}\frac{2^a(a+3b)!(3n/2-(a+3b))!}{(3n/2)!}$$

(ii)

$$\sum_{\substack{b\geq 1\\ a+2b\leq n/2}} E(a,b) = O\left(\frac{1}{n}\right).$$

Proof. i) The term 2^a represents the choice of the configuration vertex of v, for the in-edge from $V - S$ to v for each $v \in A$.

ii) If $a = 0$, then $E(0, b) = O\left((\frac{2b}{n})^b\right)$ where $b \le n/4$.

Let $c = a + 2b,\ h = n/2, \nu = 3n/2$. For $c \le h$, $a \ge 1$, $b \ge 1$ and sufficiently large n,

$$\begin{aligned} E(a, b) &\le \left[(1 + o(1))\frac{h^2}{(2\pi)^2\nu}\frac{(c + b)}{(c - 2b)b^2}\frac{\nu - (c + b)}{(h - c + b)(h - b)}\right]^{1/2} \\ &\times \left(\frac{h^{2h}2^{c-2b}}{\nu^\nu}\frac{(c + b)^{c+b}}{(c - 2b)^{c-2b}b^{2b}}\frac{(\nu - (c + b))^{\nu-(c+b)}}{(h - c + b)^{h-c+b}(h - b)^{h-b}}\right) \\ &= [A]^{1/2} \times (B) \quad \text{say.} \end{aligned}$$

Case of $c \le n/100$

The term A is $O(1)$ and

$$B \le (\tfrac{2}{3})^c \left(\frac{1}{6enb^2}\right)^b \frac{(c + b)^{c+b}}{(c - 2b)^{c-2b}} \exp\left(\frac{(c + b)^2}{\nu}\right).$$

We examine the maxima of $y = \log g$, where

$$g = \frac{(c + b)^{c+b}}{(6en)^b b^{2b}(c - 2b)^{c-2b}}.$$

For fixed c,

$$\frac{\partial y}{\partial b} = 0 \iff 4b^3 - 6nb^2 - 3c^2b + c^3 = 0 \tag{1}$$

and $\frac{\partial^2 y}{\partial b^2} < 0$ for $0 \le b \le c/2$. Let $n = dc$ $(d \ge 100)$ and $b = kc$, then (1) becomes

$$4k^3 - 6dk^2 - 3k + 1 = 0. \tag{2}$$

Now (2) has three real roots; of which one is negative and one occurs above $k = 3d/2$. The relevant root, r, is bounded by

$$\frac{1}{\sqrt{7d}} \le r \le \frac{1}{\sqrt{6d}}.$$

Thus, for fixed c the maximum occurs at

$$b^* = \frac{c}{\sqrt{\theta 6d}} = \beta c,$$

where $1 \le \theta \le (1 + \frac{1}{6})$. Hence

$$\begin{aligned} g &\le \left[\left(\frac{\theta}{e}\right)^\beta \frac{(1 + \beta)^{1+\beta}}{(1 - 2\beta)^{1-2\beta}}\right]^c \\ &\le [\exp(\tfrac{13}{6} + \beta)]^{\beta c} \\ &\le \exp\left(\frac{7c}{3\sqrt{6d}}\right), \end{aligned}$$

as $\beta < 1/6$. Thus

$$B \leq \left[\tfrac{2}{3} \exp\left(\frac{7}{3\sqrt{6d}}\right) \exp \tfrac{3}{2d}\right]^c \leq \left(\frac{3}{4}\right)^c.$$

Note that if $c \leq n^{1/3}$, then $d \geq n^{2/3}$ and $b^* < 1$. If so we may consider $b = 1$ as the maximum, as the case $b = 0$ is not under discussion here. Hence

$$\sum_{\substack{a \geq 1, b \geq 1 \\ c \leq n^{1/3}}} E(a,b) \leq \frac{1}{n} \sum_{c \leq n^{1/3}} \left(\frac{2}{3}\right)^c c^3 = O(\frac{1}{n}).$$

Case of $n/100 \leq c \leq n/2$

The term A is $O(\sqrt{n})$. We may transform the term B by writing $a = a'n, b = b'n, ..., \nu = \nu' n$ where $a', b', ..., \nu'$ are constants. For convenience of notation we will write a' as a, b' as b etc. Then $B = [f(b,c)]^n$ where $f(b,c)$ is a constant given by

$$f(b,c) = \frac{1}{2}\left(\frac{2}{3}\right)^{3/2} \frac{2^{c-2b}(c+b)^{c+b}(\nu-(c+b))^{\nu-(c+b)}}{(c-2b)^{c-2b} b^{2b} (h-c+b)^{h-c+b}(h-b)^{h-b}}.$$

Let R denote the region $\{(b,c) : 1/100 \leq c < 1/2,\ b < c/2\}$. We will now prove that $f(b,c) < 1$ for $(b,c) \in R$.

Let $y = \log(f/(\frac{1}{2})(\frac{2}{3})^{3/2})$. The function y is well defined, continous and twice differentiable for the given values of b and c, and

$$\frac{\partial y}{\partial b} = \log\left(\frac{(c+b)(c-2b)^2(h-b)}{4b^2(h-c+b)(\nu-b-c)}\right), \tag{3}$$

$$\frac{\partial y}{\partial c} = \log\left(\frac{2(c+b)(h-c+b)}{(c-2b)(\nu-b-c)}\right). \tag{4}$$

The second derivatives $\partial^2 y/\partial b^2$, $\partial^2 y/\partial c^2$ are negative in R, so the function is concave in b (resp. c) for constant c (resp. b). From (4) we see that if $\frac{\partial y}{\partial c} = 0$ then

$$c^2 + \frac{c}{2} + bc - 4b = 0. \tag{5}$$

Only the positive root

$$c^*(b) = \frac{1}{2}\left(\sqrt{\tfrac{1}{4} + 17b + b^2} - b - \tfrac{1}{2}\right)$$

is feasible here. Thus y has a unique maximum at $(b, c^*(b))$ for fixed b. As $0 < c^* < 1/2$ we find that $0 < b < 1/7$. Within this range, b is determined from (5) by $b = c(c+\frac{1}{2})/(4-c)$ when $c = c^*(b)$. Substituting this value into (3) when $\partial y/\partial b = 0$ we find,

$$\frac{(c+b)(c-2b)^2(h-b)}{4b^2(h-c+b)(\nu-b-c)} = 1 \Longleftrightarrow \frac{27}{32}\frac{c(c+2)}{(c+\frac{1}{2})^2} = 1,$$

and hence $c = \frac{2}{5}$ or 4. Thus there is a point of inflection at $(b,c) = (1/10, 2/5)$ giving a value for f bounded by 0.9625. This is a minimum, as $f(b,c) \to 1$ as $c \to 0$.

When $c^* = 1/100$, b is given by $\beta = 51/39900$. For $b > 1/7$, $c^* > 1/2$ is outside R and f is a monotone function of c for given b. Thus we find that

$$\max\{f(b,c) : (b,c) \in R\} = \max\left\{f(\beta, 1/100),\ f(1/7,1/2),\ \max_{1/7<b<1/4} f(b,2b),\ \max_{1/7<b<1/4} f(b,\tfrac{1}{2})\right\},$$

where $f(b,2b) = \sqrt{2}\, b^b (h-b)^{(h-b)}$ is monotone decreasing for $1/7 < b < 1/4$.

We now evaluate $f(b,1/2) = g(b)$ say, where

$$g(b) = \frac{2}{3^{3/2}} \frac{(h+b)^{h+b}(1-b)^{1-b}}{2^{2b}b^{3b}(h-2b)^{h-2b}(h-b)^{h-b}}.$$

Let $y = \log(g \times 3^{3/2}/2)$. Then

$$\frac{\partial y}{\partial b} = \log\left(\frac{(h+b)(h-2b)^2(h-b)}{4b^3(1-b)}\right)$$

$$\frac{\partial y}{\partial b} = 0 \iff 32b^3 - 12b^2 + 8b - 1 = 0.$$

This cubic has one real root. This is (inconveniently) located in $[1/7, 1/4]$, at approximately $b' = 0.1442$, which is just above $b = 1/7$ (the value of b given by $c^*(b) = 1/2$, for which $f(1/7, 1/2) = 0.96348$). On evaluation we find $f(b', 1/2) \le 0.9635$. Thus

$$\max\{f(b,c) : (b,c) \in R\} \le \max\left\{(\tfrac{3}{4})^{1/100},\ 0.96348,\ 0.7415,\ 0.9635\right\} \le 0.9972.$$

□

The minutiae associated with bounding the function f seem endless, and the reader may be reassured to know that we have computed f over the relevant range, and the answers are indeed correct.

Corollary 7 *a.e. configuration with no monochromatic cycle is strongly connected, and thus* $|\mathcal{C}_F| = |\mathcal{C}_N|(1 - o(1))$.

This completes the proof of Theorem 3.

5 Other Considerations

The work by Bollobás [Bo1,2] established the connectivity of random 3-regular graphs, and the properties of their automorphism groups. We must briefly return to these topics here, as the underlying graphs of $\mathcal{D}$ do not correspond directly to the space of 3-regular graphs with the uniform measure.

Lemma 8 *a.e* $D \in \mathcal{D}_{N\setminus T}$ *has an underlying graph which is vertex 3-connected.*

Proof. Suppose S is a separator of size at most 2, separating non-empty vertex sets R, T. Then $|S| = 2$ by strong connectivity. We consider the case where $\min\{|R|, |T|\}$is a constant here, larger values being covered by edge density arguments.

If G is some small unlabelled digraph, the expected number of copies of G occurring in a separated set T of D is,

$$cn^{b+w+s-(e+i+j+h)} = cn^{\lambda},$$

where $b = |B \cap T|, w = |W \cap T|, s = |S|, e = |E(T)|, i = |E(S \times T)|, j = |E(T \times S)|$ and $h = |E(S)|$. We consider the case where $\lambda \geq 0$, where $\lambda = 2 - (f + i + j + h)$.

We may assume that the underlying graph of T is connected, otherwise we can treat each component separately. The following constraints pertain.

$e = b + 2w - i$	in-degree of T
$e = 2b + w - j$	out-degree of T
$e = b + w + f,\ f \geq -1$	connectivity in underlying graph
$i \geq 1,\ j \geq 1$	strong connectivity between S and T
$i + j \leq 4 - 2h$	strong connectivity between S and R

Note that $w - i = b - j = f$. Writing $i + j = 2 + t$, we have $t + 2h \leq 2$, and $\lambda = -(f + t + h)$.

If $f = -1$ and $t = h = 0$ then $T = \emptyset$. $f = -1, t = 0, h = 1$ implies $f = 0$ in $S \cup T$, which now also has a separator of size 2. If $f = -1, t = 1, h = 0$ we have a T_I or T_{II} obstruction, as $w = 0, b = 1$ or vice versa.

We now consider $f = t = h = 0$, and thus $b = w = 1, e = 2$ which also implies a T_I or T_{II} obstruction.

□

We now consider the triviality of the automorphism group for a.e. $D \in \mathcal{D}_F$. The trivial automorphism property persists in $F \in \mathcal{F}$ after the removal of any edge $(z, a), z \in W$ from the pre-flowgraphs $\mathcal{D}_F$.

Lemma 9 *a.e* $D \in \mathcal{D}_F$ *and* $F \in \mathcal{F}$ *has trivial automorphism group.*

Proof. Let $D \in \mathcal{D}_F,\ e = (z, a),\ z \in W$. Then $F = D - e$ is the labelled flowgraph (F, a, z). There are $n/2$ choices for e.

Let $M_n = \{\sigma \in S_n : \{1, ..., n/2\} \xrightarrow{\sigma} \{1, ..., n/2\}\}$ be the subset of S_n which permutes black (resp. white) vertices among black (resp. white) vertices. $|M_n| = m = (\frac{n}{2}!)^2$.

For $\sigma \in M_n$, let

$F(\sigma) = \{F \in \mathcal{F} : F$ is invariant under $\sigma\}$, and $\phi(\sigma) = |F(\sigma)|$,

$D(\sigma) = \{D \in \mathcal{D}_F : D$ is invariant under $\sigma\}$, and $\psi(\sigma) = |D(\sigma)|$.

Then

$$|\mathcal{U}_F| = \sum_{F\in\mathcal{F}} \frac{|Aut(F)|}{m} = \frac{1}{m}\sum_{\sigma\in M_n}\phi(\sigma) = \frac{\phi(1)}{m}\left(1+\sum_{\sigma\in M_n\setminus\{1\}}\frac{\phi(\sigma)}{\phi(1)}\right).$$

But $\phi(1) = \frac{n}{2}|\mathcal{D}_F| = \frac{n}{2}\psi(1)$, as it is clear from the flowgraph F which vertex is a and which is z, for $d^+(z) = 0$ whereas $d^+(a) > 0$. Thus there is a *unique* replacement edge direction (z, a) irrespective of the colour of a. Now,

$$\sum_{\sigma\in M_n\setminus\{1\}}\frac{\phi(\sigma)}{\phi(1)} \le \frac{1}{\phi(1)}\sum_{\sigma\in M_n\setminus\{1\}}\frac{n}{2}\psi(\sigma) \le \sum_{\sigma\in M_n\setminus\{1\}}\frac{\psi(\sigma)}{\psi(1)}. \tag{6}$$

For if σ is an automorphism of $F = D - e$, then $\sigma(a) = a, \sigma(z) = z$ as the vertices are distinguished by their out-degree. Thus σ is an automorphism of D. Thus $F \in F(\sigma)$ implies $D \in D(\sigma)$ although the converse is not necessarily true.

We now explain how the original proof of Bollobás [Bo2] for 3-regular graphs can be used to infer the final term in (6) is $o(1)$. The probability of a set of u fixed edges in the $\mathcal{C}$ model is $(\frac{c}{n})^u$ as it is in the 3-regular case, but for a different constant c. The other calculations in [Bo2] go through unchanged, except now, some of the orbits of the undirected case, eg type $(2, 0), (xy \to xy$ by $\sigma(x) = y, \sigma(y) = x)$ no longer occur in a directed context. The reader is referred to the paper for the details. ☐

6 Acknowledgement

Many thanks to John Bainbridge and Robin Whitty for ideas and background material, and to Robin Whitty for considerable help with the definitional aspects of the problem.

7 References

[AGG] Arratia, R., Goldstein, L. & Gordon, L. *Poisson approximation and the Chen-Stein method.* Statistical Science 5.4 (1990) 403-434.

[BHJ] Barbour, A.D., Holst, L. and Janson, S. *Poisson Approximation.* Oxford (1992).

[BB] Bender, E.A., & Butler, J.T. *Enumeration of structured flowcharts.* JACM 32.3 (1985) 537-548.

[BJ] Böhm, C., & Jacopini, G. *Flow-diagrams, Turing machines, and languages with only two formation rules.* Commun. ACM 9 (May 1966) 366-371.

[Bo1] Bollobás, B. *Random Graphs.* Academic Press, London (1985).

[Bo2] Bollobás, B. *The asymptotic number of unlabelled regular graphs* J. London Math. Soc 2.26 (1982) 201-206

[BM] Bondy, J.A. & Murty, U.S.R. *Graph Theory With Applications.* North Holland (1976).

[Dj] Dijkstra, E.W. *Go to statement considered harmful.* Commun. ACM 11 (Mar 1968) 147-148.

[Fe] Fenton, N.E. *Software Metrics: A Rigorous Approach.* Chapman & Hall (1991).

[FH] Fenton, N.E., & Hill, G. *Systems Construction and Analysis, A Mathematical and Logical Framework.* McGraw-Hill (1992).

[He] Hecht, M.S. *Flow Analysis of Computer Programs.* North-Holland, New York (1977).

[Kn] Knuth, D.E. *Structured programming with GOTO statement.* Comput. Surv. 6, 4 (Dec 1974) 261-301.

[LMW] Linger, R.C., Mills, H.D., & Witt, B.I. *Structured Programming: Theory and Practice.* Addison-Wesley, New York (1979).

[Pr] Prather, R.E. *Hierarchical metrics and the prime generation problem.* Software Engineering Journal (Sept. 1993) 246-252.

[Zu] Zuse, H. *Software Metrics: Tools and Techniques.* de Gruyter, Berlin (1991).