Discrete Mathematics 19 (1977) 17–21.

THE NUMBER OF FINITE RELATIONAL STRUCTURES*

Ronald FAGIN
IBM Research Laboratory, San Jose, California 95193, USA

Received 19 December 1975
Revised 23 August 1976

An elementary proof is presented of an asymptotic estimate for the number (up to isomorphism) of finite relational structures, under a quite general definition of "relational structure".

1. Introduction

It seems to be well-known in combinatorial folklore that "almost all" structures with many nodes are *rigid*, that is, have no nontrivial automorphisms. In other words, the fraction of n-node structures (of a specified type) which are rigid goes to 1 as n goes to infinity. Equivalently [5, 8], if a_n is the number of n-node "labeled structures" and b_n is the number of n-node "unlabeled structures", then $b_n \sim a_n/(n!)$, where "$\sim$" is read "is asymptotic to". For example, in 1958, Harary [6] noted that the number of unlabeled directed graphs on n-nodes (that is, the number of distinct directed graphs, *up to isomorphism*, on n-nodes), is asymptotic to $2^{n^2}/(n!)$. (Harary's result is a straightforward extension of results in Ford and Uhlenbeck [5], which in turn are based on unpublished work of Polya.) In 1966, Oberschelp [8] generalized this result as follows. Instead of considering n-node labeled directed graphs (each such graph can be thought of as a distinguished binary relation over $\{1, \ldots, n\}$), he considers n-node labeled structures with k distingushed r-ary relations (each such structure can be thought of as a k-tuple of distinguished r-ary relations over $\{1, \ldots, n\}$). The special case of directed graphs corresponds to $k = 1$ and $r = 2$. Oberschelp shows that if k and r are held fixed (with $r \geq 2$), then the number of unlabeled such structures with n nodes (that is, the number of isomorphism classes) is asymptotic to the number of labeled such structures on n nodes divided by $n!$ A natural final generalization (which does not seem to appear in the literature)[1] is to labeled structures with u_i i-ary relations

* This work was carried out while the author was at the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, U.S.A.

[1] This result was apparently obtained by Oberschelp in 1967 (unpublished) and was recently obtained, independently of Oberschelp and the author, by A. Ehrenfeucht (unpublished).

($i = 1,\ldots,t$): for example, if $t = 3$, $u_1 = 1$, $u_2 = 5$, and $u_3 = 2$, then we are dealing with labeled structures that have 1 distinguished unary relation, 5 distinguished binary relations, and 2 distingushed ternary relations, all over $\{1,\ldots,n\}$. Such structures correspond exactly to what Tarski [9] calls "finite relational systems of finite order in a fixed similarity class". In this paper, we present an elementary proof that as long as $u_i > 0$ for some $i > 1$ (that is, as long as we are dealing with structures in which at least one distinguished relation is not unary), and if $t, u_1,\ldots,u_t$ are each held fixed, then once again, the number of unlabeled such n-node structures is asymptotic to the number of labeled such n-node structures, divided by $n!$ (The statement is easily seen to be false if $u_i = 0$ for each $i > 1$.)

The paper arose when the author needed this result, in its full generality, as a lemma to prove a result in mathematical logic [4].

2. Definitions

Let S be a *type*, that is a t-tuple of nonnegative integers for some positive integer t. If $S = (u_1,\ldots,u_t)$, then by a (*labeled n-node*) *S-structure*, we mean a $(\sum u_i)$-tuple of u_1 distinguished unary relations, u_2 distinguished binary relations, $\ldots, u_t$ distinguished t-ary relations, all over $\{1,\ldots,n\}$ (we assume for convenience that $u_t > 0$). Thus, if

$$\mathcal{A} = (R_1^1,\ldots,R_{u_1}^1; R_1^2,\ldots,R_{u_2}^2;\ldots; R_1^t,\ldots,R_{u_t}^t) \tag{1}$$

is a labeled n-node S-structure, then R_j^i is an i-ary relation over $\{1,\ldots,n\}$, that is, a set of i-tuples of integers between 1 and n ($i = 1,\ldots,t$; $j = 1,\ldots,u_i$). A (0,1) structure is a labeled directed graph; Oberschelp dealt with $(0,\ldots,0,k)$-structures in [8]. If

$$\mathcal{B} = (Q_1^1,\ldots,Q_{u_1}^1; Q_1^2,\ldots,Q_{u_2}^2;\ldots; Q_1^t,\ldots,Q_{u_t}^t)$$

is another n-node labeled S-structure, then we say that $\mathcal{A}$ and $\mathcal{B}$ are isomorphic if there is a bijection $\pi:\{1,\ldots,n\}\to\{1,\ldots,n\}$ such that for each $i, j, a_1,\ldots,a_i$, ($1 \le i \le t$, $1 \le j \le u_i$, $1 \le a_1 \le n,\ldots, 1 \le a_i \le n$),

$$(a_1,\ldots,a_i) \in R_j^i \quad \text{iff } (\pi a_1,\ldots,\pi a_i) \in Q_j^i.$$

Let $a_n(S)$ be the number of distinct n-node labeled S-structures. Clearly,

$$a_n(S) = 2^{\sum_{i=1}^t u_i n^i}. \tag{2}$$

Let $b_n(S)$ be the number of isomorphism classes of n-node labeled S-structures. We will show that as long as S is not of type (u_1), that is, as long as we are dealing with structures with at least one distinguished relation which is not unary, then $b_n(S) \sim a_n(S)/(n!)$.

3. Proof of Theorem 3.1

We will prove the following result.

Theorem 3.1. *Assume that* $S = (u_1, \ldots, u_t)$, $t > 1$ *and* $u_t > 0$. *Then* $b_n(S) \sim a_n(S)/(n!)$.

Proof. If π is a permutation of $\{1, \ldots, n\}$, then π induces a natural permutation π_i on the set of i-tuples of $\{1, \ldots, n\}$, via

$$\pi_i(a_1, \ldots, a_i) = (\pi a_1, \ldots, \pi a_i).$$

Define $C_i(\pi)$ to be the total number of cycles (including singletons) in the cyclic decomposition of π_i $(i = 1, \ldots, t)$.

Lemma 3.2 (McKenzie [7]). *Let* π *be a permutation of* $\{1, \ldots, n\}$. *The number of* n*-node labeled* S*-structures for which* π *is an automorphism is exactly* $2^{\sum_{i=1}^t u_i C_i(\pi)}$.

Proof of Lemma 3.2. Assume that $(\underline{v}^{(1)}, \ldots, \underline{v}^{(k)})$ is one of the $C_i(\pi)$ cycles in the cyclic decomposition of π_i (here $\underline{v}^{(1)}, \ldots, \underline{v}^{(k)}$ are each i-tuples.) Let $\mathscr{A}$ be as in (1). If π is an automorphism of $\mathscr{A}$, then

$$\underline{v}^{(1)} \in R_j^i \quad \text{iff } \underline{v}^{(2)} \in R_j^i \quad \text{iff} \cdots \quad \text{iff } \underline{v}^{(k)} \in R_j^i.$$

If $d = \sum_{i=1}^t u_i C_i(\pi)$, then it follows that π is an automorphism of exactly 2^d n-node labeled S-structures; intuitively, there are d "degrees of freedom". This proves the lemma.

We return to the proof of the theorem. Denote by $N(\pi)$ the number of n-node labeled S-structures for which π is an automorphism. By the lemma, $N(\pi) = 2^{\sum_i u_i C_i(\pi)}$. By Burnside's Theorem ([1], Sec. 145, Theorem VII; see also [2], p. 150),

$$b_n(S) = \frac{1}{n!} \sum_\pi N(\pi).$$

So by Lemma 3.2,

$$b_n(S) = \frac{1}{n!} \sum_\pi 2^{\sum_i u_i C_i(\pi)} \tag{3}$$

The asymptotic relationship between $a_n(S)$ and $b_n(S)$ in the statement of the theorem is equivalent to

$$\frac{n! b_n(S)}{a_n(S)} \to 1 \quad \text{as } n \to \infty.$$

By (2) and (3), it is sufficient to show that

$$\sum_\pi 2^{\sum_i u_i [C_i(\pi) - n^i]} \to 1 \quad \text{as } n \to \infty. \tag{4}$$

If π is the identity permutation I [where $I(j)=j$, for $j=1,\ldots,n$], then $C_i(\pi)=n^i$ for each i. So, if we split the sum in (4) into the term with $\pi=I$ and those terms for which $\pi\neq I$, we find that (4) is equivalent to

$$\sum_{\pi\neq I} 2^{\Sigma_i u_i[C_i(\pi)-n^i]} \to 0 \quad \text{as } n\to\infty. \tag{5}$$

Define a permutation to be of *kind* m if exactly m points are not fixed (i.e., are not mapped onto themselves by the permutation.) If π is of kind m, then π_i is of kind at least mn^{i-1}, since if $1\leq a_1\leq n$ and a_1 is not fixed by π [i.e., $\pi(a_1)\neq a_1$], then the n^{i-1} i-tuples $(a_1,a_2,\ldots,a_i)$, where $a_2,\ldots,a_i$ each run through $\{1,\ldots,n\}$, are not fixed by π_i. (Remark: In fact, π_i is exactly of kind $n^i-(n-m)^i$.)

It is simple to see that if π_i is of kind x, then $C_i(\pi)\leq n^i-(x/2)$. So by the above, if π is of kind m, then π_i is of kind at least mn^{i-1}, and so

$$C_i(\pi)\leq n^i-(mn^{i-1}/2). \tag{6}$$

If we substitute the right-hand side of (6) for each occurrence of $C_i(\pi)$ in (5), and if we write $m(\pi)$ for m, then we find that the expression on the left-hand side of the first "$\to$" in (5) is dominated by

$$\sum_{\pi\neq I} 2^{-\frac{1}{2}m(\pi)n^{i-1}}. \tag{7}$$

Our goal is to show that expression (7) goes to 0 as $n\to\infty$. We will show even more, that (7) is $o(\theta^{n^{i-1}})$ as long as $\frac{1}{2}<\theta<1$.

The number of permutations π of $\{1,\ldots,n\}$ which are of kind m is obviously at most $\binom{n}{m}m!$, which is dominated by n^m. So (7) is dominated by

$$\sum_{m=2}^{n} n^m 2^{-\frac{1}{2}mn^{i-1}}, \tag{8}$$

where the sum starts from $m=2$ and not from $m=1$ since there are no permutations of kind $m=1$. Now expression (8) equals

$$\sum_{m=2}^{n} 2^{-\frac{1}{2}m(n^{i-1}-2\log n)}, \tag{9}$$

where the logarithm is to the base 2.

Assume that n is large enough so that $n^{i-1}-2\log n$ is positive. The largest summand of (9) occurs when $m=2$, and so (9) is dominated by $n-1$ (i.e., the number of summands) times the summand for $m=2$, that is, (9) is dominated by

$$(n-1)2^{-\frac{1}{2}(n^{i-1}-2\log n)},$$

which equals

$$2^{-\frac{1}{2}(n^{i-1}-2\log n-\log(n-1))}.$$

This expression is clearly $o(\theta^{n^{i-1}})$, if $\frac{1}{2}<\theta<1$, which completes the proof.

4. Further remarks

(a) If S is not of type (u_1), then $b_n(S)$ gets close to $a_n(S)/(n!)$ "quickly". Specifically, if S is of type $(u_1, \ldots, u_t)$, with $t > 1$ and $u_t > 0$, then we have shown that

$$\left|\frac{(n!)b_n(S)}{a_n(S)}\right| - 1 = o(\theta^{n^{t-1}}) \quad \text{if } \tfrac{1}{2} < \theta < 1.$$

In particular, since $t \geq 2$, we see that $(n!)b_n(S)/a_n(S)$ converges to 1 at least "geometrically fast".

(b) It is possible to generalize further our notion of "type". Thus, instead of a type "saying" that there must be a certain number of distinguished binary relations, and so on, a type might also say for example, that, in addition, there must be a certain number of distinguished "*undirected* graph relations" (by an "undirected graph relation", we mean a set of *unordered* pairs of $\{1, \ldots, n\}$). Similarly, for arbitrary k, the type can say that there must be a certain number of "undirected k-ary relations", each of which is a set of unordered k-tuples of $\{1, \ldots, n\}$ (cf. the "linear graphs" of [3]). It is straightforward to check that with minor modifications, the proof of the theorem still goes through, as long as the type includes at least one kind of relation, undirected or not, which is not unary.

Acknowledgements

The author is grateful to Ralph McKenzie for Lemma 3.2, to George Markowsky for helpful suggestions which improved readability, and to the referee for suggesting certain simplifications in the proof and for pointing out the existence of Oberschelp's unpublished manuscript.

References

[1] W. Burnside, Theory of Groups of Finite Order, 2nd edition (Cambridge University Press, Cambridge, 1911; Dover Publications, NY, 1955).
[2] N.G. de Bruijn, Polya's theory of counting, in: E.F. Beckenbach, ed., Applied Combinatorial Mathematics (Wiley, NY, 1964) 144-184.
[3] N.G. de Bruijn and D.A. Klarner, Enumeration of generalized graphs, Indag. Math. 31 (1969) 1-9.
[4] R. Fagin, Probabilities on finite models, J. Symbol. Logic 41 (1976) 50–58.
[5] G.W. Ford and G.E. Uhlenbeck, Combinatorial problems in the theory of graphs, IV, Proc. Nat. Acad. Sci. 43 (1957) 163–167.
[6] F. Harary, Note on Carnap's relational asymptotic relative frequencies, J. Symbol. Logic 23 (1958) 257–260.
[7] R. McKenzie, private communication.
[8] W. Oberschelp, Strukturzahlen in endlichen Relationssystemen, in: H.A. Schmidt, et al., eds., Contributions to Mathematical Logic, Proc. of 1966 Logic Colloquium (North-Holland, Amsterdam, 1968) 199–213.
[9] A. Tarski, Contributions to the theory of models I, II, Indag. Math. 16 (1954) 572-588.