

# Fraïssé-Ehrenfeucht Games and Asymptotics

Alan Woods

University of Western Australia

March 23, 1998

[summary by Julien Clément and Jean-Marie Le Bars]

## Abstract

Fraïssé-Ehrenfeucht games are played on two structures, where a structure might, for example, consist of a unary function mapping a finite set into itself. Via generating series and a Tauberian theorem, it is possible to investigate the asymptotic probability of having a winning strategy for such a game, when it is played using a fixed structure, and a random structure of size  $n$ , with  $n$  going to infinity. Actually for unary functions this gives a convergence law for all properties of the structure which are definable in monadic second order logic.

## 1. Introduction

We consider here structures  $\mathcal{A}$  based upon a set  $A$  and finitely many relations  $E_j$  of finite arity

$$\mathcal{A} = \langle A, E_1(x, y), E_2(x), E_3(x, y, z), \dots \rangle.$$

A classical example is a set of vertices  $V$  and an edge relation  $E(x, y)$  so that  $\mathcal{V} = \langle V, E \rangle$  describes a graph. We can also think of simple structures  $\mathcal{A} = \langle A, f \rangle$  consisting of a finite set  $A$  and a unary function mapping this set into itself (see fig. 1). This unary function induces a binary relation  $F(x, y) \Leftrightarrow f(x) = y$ .

In order to use generating functions (see the last section) we need to translate a decomposition property of structures to the generating functions: this will be done through the *disjoint union*. Let us consider two structures

$$\mathcal{A} = \langle A, E_1^A, \dots \rangle \text{ and } \mathcal{B} = \langle B, E_1^B, \dots \rangle.$$

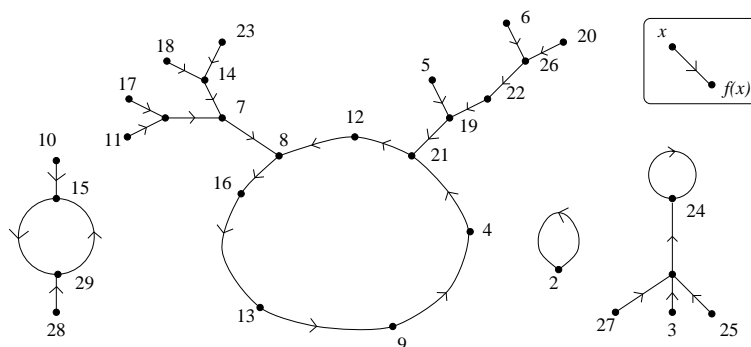


FIGURE 1. Graphical representation a structure  $\mathcal{A} = \langle [29], f \rangle$  (where the unary function  $f$  maps  $\{1, 2, \dots, 29\}$  on itself).

If  $A \cap B = \emptyset$  and each  $E_i^A$  has the same arity as  $E_i^B$ , the *disjoint union* is defined as the structure whose domain is the union of the domains and whose relations are the unions of the corresponding relations

$$\mathcal{A} \sqcup \mathcal{B} = \langle A \cup B, E_1^A \cup E_1^B, \dots \rangle.$$

A class of structures has *components* if each structure can be *uniquely* decomposed into disjoint unions of structures (called *component structures*) from some *components classes*. For structures  $\mathcal{A} = \langle [n], f \rangle$ , where  $[n]$  denotes  $\{1, \dots, n\}$  and  $f$  is a unary function, one can define component classes relative to the size of the unique loop present in each connected component of the graph of  $f$ . From this point of view, for the structure  $\mathcal{A}$  of figure 1, we see three components. The first component of  $\mathcal{A}$  consists of two component structures in the first component class (the class corresponding to loops of size one i.e. fixed elements of  $f$ ). The two other components consist in two single component structures and are respectively in the component classes 2 and 7 (relatively to the size of the loop).

Let us define the *rank*  $r(\varphi)$  of a formula  $\varphi$  in the context of the second order logic (or MSO logic for short) inductively by:

1. If  $\varphi$  has no quantifiers, then  $r(\varphi) = 0$ ;
2. If  $\varphi$  is  $\neg\sigma$ , then  $r(\varphi) = r(\sigma)$ ;
3. If  $\varphi$  is obtained from  $\sigma_1, \sigma_2$  by the application of a binary propositional connective (e.g., if  $\varphi$  is  $\sigma_1 \wedge \sigma_2$ ,  $\sigma_1 \leftrightarrow \sigma_2$ , etc.) then  $r(\varphi) = \max\{r(\sigma_1), r(\sigma_2)\}$ ;
4. If  $\varphi$  is of the form  $\forall v\sigma$ ,  $\exists v\sigma$ ,  $\forall V\sigma$  or  $\exists V\sigma$  for some variable  $v, V$ , then  $r(\varphi) = r(\sigma) + 1$ .

A *sentence* is a formula that has *no free variables* and is a *property* of a structure.

The key observation is that *there are only finitely many inequivalent sentences*  $\xi_1, \dots, \xi_m$  of rank  $r$ . Hence every structure  $\mathcal{A}$  satisfies *exactly* one of the sentences (also of rank  $r$ )

$$\psi_1 = \xi_1 \wedge \dots \wedge \xi_m, \psi_2 = \neg\xi_1 \wedge \dots \wedge \xi_m, \dots, \psi_{2^m} = \neg\xi_1 \wedge \dots \wedge \neg\xi_m.$$

Given a rank  $r$  (and implicitly the sentences  $\psi_1, \dots, \psi_{2^m}$ ), for each  $i \in \{1, \dots, 2^m\}$  we define the class of structures which satisfies  $\psi_i$ . These classes can be viewed as equivalence classes of Fraïssé-Ehrenfeucht games.

## 2. Fraïssé-Ehrenfeucht Games

The goal is to see whether or not we can distinguish two structures in a  $r$  moves game. The game is played with two structures  $\mathcal{A} = \langle A, E_1^A, \dots \rangle$  and  $\mathcal{B} = \langle B, E_1^B, \dots \rangle$ .

- At move  $i$ , SPOIL chooses  $\mathcal{A}$  or  $\mathcal{B}$  (let's say  $\mathcal{B}$ ) and one of the following is satisfied
  1. an element  $b_i \in B$  or
  2. a subset  $B_i \subseteq B$ .
- DUPE responds on the other structure ( $\mathcal{A}$  here) choosing one of the following
  1. an element  $a_i \in A$  or
  2. a subset  $A_i \subseteq A$ .

DUPE wins if after  $r$  moves the map  $\{a_i, \dots\} \rightarrow \{b_i, \dots\}$  taking  $a_i \mapsto b_i$  is an isomorphism of the induced substructures of  $\langle \mathcal{A}, A_j, \dots \rangle, \langle \mathcal{B}, B_j, \dots \rangle$  on these sets. We write

$$\mathcal{A} \equiv_r \mathcal{B} \Leftrightarrow \text{DUPE has a winning strategy.}$$

Note that there is no *ex æquo* (either SPOIL or DUPE has a winning strategy). These games are the main tools for proving the following theorems:

**Theorem 1.** *Let us consider some structures  $\mathcal{A}_1, \mathcal{A}_2, \mathcal{B}_1, \mathcal{B}_2$ , one has*

$$\mathcal{A}_1 \equiv_r \mathcal{B}_1, \mathcal{A}_2 \equiv_r \mathcal{B}_2 \Rightarrow \mathcal{A}_1 \sqcup \mathcal{A}_2 \equiv_r \mathcal{B}_1 \sqcup \mathcal{B}_2.$$

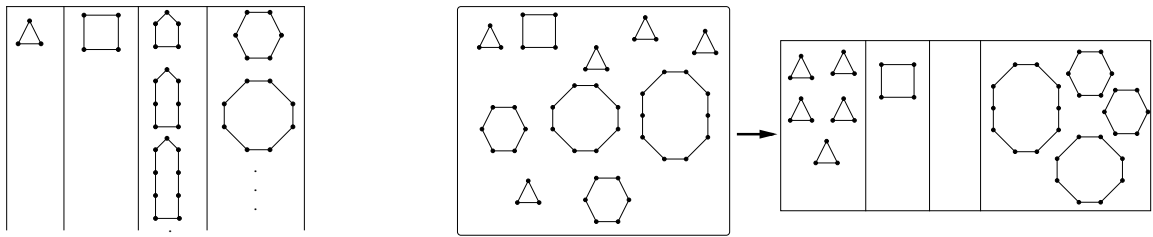


FIGURE 2. The components classes  $\mathcal{C}_1, \dots, \mathcal{C}_4$  relative to  $\equiv_3$  (left), the structure  $\mathcal{A}$  and its four components (right).

**Theorem 2.** For every structures  $\mathcal{A}$  and  $\mathcal{B}$ , one has

$$\mathcal{A} \equiv_r \mathcal{B} \text{ iff there exists } i \text{ such that } \mathcal{A} \models \psi_i \text{ and } \mathcal{B} \models \psi_i,$$

where the sentences  $\psi_i$ 's are defined in the first section.

**Corollary 1.** There are only finitely many  $\equiv_r$  classes.

Another problem consists in determining the  $\equiv_r$  class of a given structure  $\mathcal{A}$ . It is solved if we know the number of component structures lying in each  $\equiv_r$  component class (or color if we think of  $\equiv_r$  as a colouring). On figure 2, we have 5 component classes  $\mathcal{C}_1, \dots, \mathcal{C}_5$  relative to the  $\equiv_3$  relation (namely triangles, squares, cycles of odd length strictly greater than 3, cycles of even length strictly greater than 4). The numbers of component structures in each component of the structure  $\mathcal{A}$  are respectively  $m_1 = 5, m_2 = 1, m_3 = 0, m_4 = 4$ .

### 3. Counting Structures with Components

We count either

1. the number  $a_n$  of *labelled structures* with  $n$  elements, or
2. the number  $b_n$  of *unlabelled structures* with  $n$  elements (which is, also, the number of nonisomorphic structures with  $n$  elements).

Here we focus on counting labelled structures. So the exponential generating series

$$a(x) = \sum_{n=0}^{\infty} \frac{a_n}{n!} x^n$$

will prove highly useful. Indeed, for a structure  $\mathcal{A} = \mathcal{G} \sqcup \mathcal{H}$ , letting  $a(x)$ ,  $h(x)$  and  $g(x)$  be the corresponding exponential generating series, we write

$$a(x) = g(x)h(x) \quad \text{or} \quad a(x) = \frac{g(x)^2}{2},$$

whether  $\mathcal{G}$  and  $\mathcal{H}$  are in different classes or not. By induction the exponential generating series associated to  $\mathcal{A} = \mathcal{G}^{(1)} \sqcup \dots \sqcup \mathcal{G}^{(m)}$  the disjoint union of  $m$  structures  $\mathcal{G}^{(1)}, \dots, \mathcal{G}^{(m)}$ , is

$$a(x) = g^{(1)}(x) \dots g^{(m)}(x) \quad \text{or} \quad a(x) = \frac{g(x)^m}{m!},$$

respectively if  $\mathcal{G}^{(1)}, \dots, \mathcal{G}^{(m)}$  are all from different classes or all in the same class. Hence the generating series  $a(x)$  for structures with components in the component class  $\mathcal{C}$  is

$$a(x) = 1 + c(x) + \frac{c(x)^2}{2!} + \dots + \frac{c(x)^m}{m!} + \dots = e^{c(x)},$$

where  $c(x) = \sum_n \frac{c_n}{n!} x^n$  ( $c_n$  is the number of labelled structures in the component class  $\mathcal{C}$  with  $n$  elements).

There is a connection with monadic second order logic due to Compton [2]. Let us consider the component classes (relatively to  $\equiv_r$ )  $\mathcal{C}_1, \dots, \mathcal{C}_k$  (so that the generating series for whole component class is  $c(x) = \sum_{i=1}^k c_i(x)$ ). There is a unique  $k$ -tuple  $(m_1, \dots, m_k)$  associated to each structure  $\mathcal{A}$ , where  $m_i$  is the number of component structures of  $\mathcal{A}$  lying in the  $i$ -th component class  $\mathcal{C}_i$ . Moreover for two structures  $\mathcal{A}$  and  $\mathcal{B}$  (with  $k$ -tuples  $(m_1, \dots, m_k)$  and  $(n_1, \dots, n_k)$ ), there is an integer  $R = R(r)$  such that if  $\forall i \in \{1, \dots, k\}$  either  $m_i = n_i$  or  $m_i, n_i \geq R$ , then  $\mathcal{A} \equiv_r \mathcal{B}$  (plainly speaking, too many component structures of the same component class prevent to distinguish structures). Hence for a sentence  $\varphi$  of rank  $r$ , the number of labelled structures  $\mathcal{A}$  such that  $\mathcal{A} \models \varphi$  depends only on  $m_1, \dots, m_k$  where  $m_i \in \{0, 1, \dots, R-1, \infty\}$  is the number of components in  $\mathcal{C}_i$  ( $\infty$  means anything equal to at least  $R = R(r)$ ). Considering the exponential generating series  $a_\varphi(x) = \sum a_n^\varphi/n!$  where  $a_n^\varphi$  the number of labelled structures with  $n$  elements satisfying  $\varphi$ , we can write

$$a_\varphi(x) = \sum_{(m_1, \dots, m_k) \in S} \frac{c_1(x)^{m_1}}{m_1!} \dots \frac{c_k(x)^{m_k}}{m_k!},$$

where  $S$  is finite and  $c_i(x)^\infty/\infty!$  denotes  $\sum_{m=R}^\infty c_i(x)^m/m! = e^{c_i(x)} - \sum_{m=0}^{R-1} c_i(x)^m/m!$ . The series  $a_\varphi(x)$  is a finite sum of very similar terms. It is enough just to consider a series of the form

$$a_\varphi(x) = \frac{c_1(x)^{m_1}}{m_1!} \dots \frac{c_t(x)^{m_t}}{m_t!} e^{c_{t+1}(x)} \dots e^{c_k(x)}.$$

This formula means that a structure  $\mathcal{A}$  satisfying  $\varphi$  has exactly  $m_i$  components in the class  $i$  for  $i \in \{1, \dots, t\}$  and any number of components in the other classes. We want to know  $a_n^\varphi$  or equivalently  $\mu_n(\varphi) = a_n^\varphi/a_n$ , the fraction of structures of size  $n$  satisfying  $\varphi$ . We are also interested in the *asymptotic probability*  $\mu_\varphi = \lim_{n \rightarrow \infty} \mu_n(\varphi)$ , when this limit exists.

It is Compton's idea to use partial converses *Tauberian lemmas* to get limit laws for  $\mu_n$ . Here is a sample theorem whose proof is based on such lemmas.

**Theorem 3.** *For any class with components, if  $a_n/n! \sim C\tau^n/n^\alpha$  and  $c_n/n! = O(\tau^n/n)$  (with  $\alpha > -1$ ) then  $\mu(\varphi) = \lim_{n \rightarrow \infty} \mu_n(\varphi)$  exists for all MSO sentences  $\varphi$  and is equal to  $a_\varphi(\rho)/a(\rho)$ .*

Due to known results about  $a_n$  and  $c_n$  for structures with one unary function, we have also

**Corollary 2.** *The asymptotic probability  $\mu_\varphi$  always exists with one unary function.*

## References

- [1] Compton (Kevin J.). – Application of a Tauberian theorem to finite model theory. *Archiv für Mathematische Logik und Grundlagenforschung*, vol. 25, n° 1-2, 1985, pp. 91–98.
- [2] Compton (Kevin J.). – A logical approach to asymptotic combinatorics. II. Monadic second-order properties. *Journal of Combinatorial Theory. Series A*, vol. 50, n° 1, 1989, pp. 110–131.
- [3] Fagin (Ronald). – Probabilities on finite models. *Journal of Symbolic Logic*, vol. 41, n° 1, 1976, pp. 50–58.
- [4] Glebskiĭ (Ju. V.), Kogan (D. I.), Liogon'kiĭ (M. I.), and Talanov (V. A.). – Volume and fraction of satisfiability of formulas of the lower predicate calculus. *Kibernetika (Kiev)*, vol. 5, n° 2, 1969, pp. 17–27. – English translation *Cybernetics*, vol. 5 (1972), pp. 142–154.
- [5] Woods (Alan). – Counting finite models. *The Journal of Symbolic Logic*, vol. 62, n° 3, September 1997, pp. 925–949.