UNIVERSITEIT UTRECHT

MASTER THESIS

Logical laws on caterpillars

Author:
Roel LAMBERS

Supervisor:
Dr. Tobias MÜLLER

Mathematisch Instituut

December 22, 2016

UNIVERSITEIT UTRECHT

Abstract

Logical laws on caterpillars

by Roel LAMBERS

Glebskii et al. (1969) and Fagin (1975) first proved a Zero-One Law with respect to First Order-logic on the Erdos-Renyi random graph $G(n, p)$ with $p = \frac{1}{2}$. Since then, multiple classes of graphs have been proven to obey a Zero-One Law or not, on different kinds of logical languages. The class of labeled caterpillars does not obey a Zero-One law in First Order Logic; however, there's a convergence law, as is proven in this thesis. Subsequently, it's also proven for the class of unlabeled caterpillars.

While proving these convergence laws, we use the fact that only the 'outer parts' of a caterpillar can be different, while the 'inside' tends to be the same for large enough caterpillars. In the final chapter we therefore introduce the group of caterpillars, having their outer vertices attached to each other to create bracelets. The class of bracelets is proven to obey a Zero-One Law.

Acknowledgements

First of all, I would like to thank Dr. Tobias Müller, for being my supervisor the past year. In all the talks we had, he remained optimistic and he helped me to go in the right direction, pointing out where I made mistakes and where I was too careless. I would also like to thank Dr. Karma Dajani for being the second reader of this thesis.

Let me also thank everyone that supported me last year, in particular my parents, who also pointed out several linguistic issues when pre-reading this text.

Contents

Chapter 1

Introduction

We start with an introduction to notation used in this thesis. As usual, a graph $G = (V, E)$ exists of vertices V and edges between vertices $v, w \in V$ if $(v, w) \in E$, also denoted as $v \sim w$. We assume all graphs to be loop-free and without multi-edges during this thesis, that is, there can only be one edge between two points $v, w \in V$ and $(v, v) \notin E \ \forall v \in V$. For a vertex $v \in V$, the degree $\deg(v) = d(v)$ is defined as $d(v) = |\{w \in V : (v, w) \in E\}|$. Further, K_n will denoted the complete graph with n vertices and $K_{i_1,\ldots,i_r}$ the r-partite graph between the r sets $V_1, \ldots, V_r$ of vertices, with $|V_j| = i_j$.

Two graphs $G = (V, E), H = (V', E')$ are said to be *isomorphic* if there exists a bijection $\psi : V \to V'$ s.t.:

$$(v, w) \in E \iff (\psi(v), \psi(w)) \in E' \quad \forall v, w \in V$$

We denote isomorphism of G, H by $G \cong H$. Clearly, when $G \cong H$ and $H \cong F$, $G \cong F$ and $G \cong G$, so isomorphism is an equivalence relation.

Notice that every vertex in V of $G = (V, E)$ is distinguishable, we say there's some labeling to distinguish them. We can also define unlabeled graphs.

Definition 1 *We define the unlabeled graph $[UG]$ of a labeled graph G as the equivalence class over isomorphisms.*

In Figure 1.1, there are two different graphs shown on vertex set $[3]$; both graphs are clearly isomorphic, they only differ in labeling, so they both belong to the same equivalence class and only represent one unlabeled graph.

We usually denote a family of labeled graphs with $\mathcal{A}$ and it's corresponding unlabeled family of graphs with $\mathcal{UA}$. The set $\mathcal{A}_n \subset \mathcal{A}$ consists of all different graphs in $\mathcal{A}$ of order n, with vertex set $[n]$, and $\mathcal{UA}_n \subset \mathcal{UA}$ are all unlabeled graphs of order n.

Definition 2 *A minor H of G is a subgraph obtained by deleting and/or contracting, which is deleting the edge and merging the vertices, edges of G, deleting isolated vertices and deleting any multi-edges or loops created this way. A family*

FIGURE 1.1: Different labels, different graphs

of graphs $\mathcal{A}$ is called a minor-closed class if for a graph $A \in \mathcal{A}$ and all it's minors H of A, we have $H \in \mathcal{A}$.

This means for instance that the graph with 3 vertices and edges between all of them, K_3, which can be drawn as a triangle, is a minor of a square-shaped graph, or any graph containing a cyclic path; just remove all edges and vertices not in the cyclic path, and contract every edge in the cyclic path until you have 3 edges and vertices left and obtain K_3. A minor-closed class of graphs not containing K_3 may therefore not contain any cyclic path at all, in any graph.

Theorem 1 ***(Robertson-Seymour)*** *[6] Every minor-closed class $\mathcal{A}$ is determined by a finite number of excluded (forbidden) minors $A_1, \dots, A_r$.*

Example 1

Let $\mathcal{A}$ be the set of planar graphs, graphs that can be drawn in $\mathbb{R}^2$ without edges crossing. Every minor of a planar graph is obviously a planar graph. Furthermore, $\mathcal{A}$ is given by all the graphs not containing $K_{3,3}$ and K_5 as minors. The class containing all trees, isn't minor-closed, as we can delete an edge in a tree G, to obtain two disjoint trees. The class of forests of trees, in which every graph is the disjoint union of trees, is however minor-closed, and the excluded minor is K_3 (the triangle), as a trees can't have a cyclic path. ■

Let $\mathcal{A}$ be a class of graphs. We define $A_n \in_n \mathcal{A}_n$ as the random uniformly chosen graph of order n with vertex set $[n]$, that is:

$$\forall A \in \mathcal{A}_n \;\; \mathbb{P}(A_n = A) = \frac{1}{|\mathcal{A}_n|}$$

Example 2

Let $\mathcal{T}, \mathcal{UT}$ be the set of all labeled and unlabeled trees respectively. There are 2 different unlabeled trees of order 4, T_A being a path of length 4 and T_B being a central vertex, adjacent to the other 3. Let T_4, UT_4 be the random labeled and unlabeld tree respectively of order 4. Then:

$$\mathbb{P}(UT_4 \cong T_A) = \mathbb{P}(UT_4 \cong T_B) = \frac{1}{2} \tag{1.1}$$

However, T_A can be labeled in 12 different ways, while T_B can only be labeled in 4 different ways. Therefore, if we would look to isomorphism of T_4 without labels:

$$\mathbb{P}(T_4 \cong T_A) = \frac{3}{4} \qquad\qquad \mathbb{P}(T_4 \cong T_B) = \frac{1}{4}$$

As only T_B contains a vertex of degree 3, we already see a different stochastic distribution between T_4 and UT_4:

$$\mathbb{P}(\exists_{x \in T_4} \deg(x) = 3) = \frac{1}{4} \qquad\qquad \mathbb{P}(\exists_{x \in UT_4} \deg(x) = 3) = \frac{1}{2}$$

■

In this thesis, a lot of factorials occur in calculations. To work with these factorials in an easy way, we'll often use the Stirling approximation:

Theorem 2 ***(Stirling's approximation)*** *For $n \in \mathbb{N}$ it holds that:*

$$n! \sim \left(\frac{n}{e}\right)^n \sqrt{2\pi n} \tag{1.2}$$

And there's the equivalent variant with logarithms:

$$\log n! \sim n \log n - n + \frac{1}{2} \log 2\pi n$$

Proof We'll prove this using the Laplace method, which is given in the following lemma:

Lemma 1 *[5] Let $f(z)$ be a twice differentiable function, with a unique maximum z_0 on $[a,b]$, with a, b either finite or infinite, and $f''(z_0) < 0$. Then:*

$$\lim_{n\to\infty} \frac{\int_a^b e^{nf(z)} dz}{e^{nf(z_0)}\sqrt{\frac{2\pi}{-nf''(z_0)}}} = 1$$

Which can also be formulated as:

$$\int_a^b e^{nf(z)} dz \sim e^{nf(z_0)} \sqrt{\frac{2\pi}{-nf''(z_0)}}$$

For any $n \in \mathbb{N}$, we have:

$$n! = \Gamma(n+1) := \int_0^\infty e^{-x} x^n dx$$

Where $\Gamma(n+1)$ is the Gamma function. Changing variables $x = nz$ in the integral, gives:

$$n! = \int_0^\infty e^{-nz}(nz)^n n \;\, dz = n^{n+1} \int_0^\infty e^{n(\log z - z)} dz$$

The function $f(z) = \log z - z$ has a single maximum on $(0, \infty)$, which is at $z = 1$. So, we may apply the Laplace Method with this $f(z)$ and $z_0 = 1$, $f(1) = -1$ as maximum and $f''(z) = -\frac{1}{z^2}$ as second derivative. Thus, we get:

$$n! = n^{n+1} \int_0^\infty e^{n(\log z - z)} dz \sim n^{n+1} e^{-n} \sqrt{\frac{2\pi}{n}} = \left(\frac{n}{e}\right)^n \sqrt{2\pi n} \tag{1.3}$$

□

Corollary 1 *For $n \in \mathbb{N}$, we have $\log n! = n \log n - n + \mathcal{O}(\log n)$.*

Proof A direct result from Theorem 2. □

Chapter 2

Caterpillars

2.1 Basic properties

A caterpillar [3, 4] $C = (V, E)$ is a graph consisting of a path and vertices of degree one attached to the path. The vertices on the path are called *the spine* and the vertices of degree one are called *legs*. To avoid ambiguity, the spine of a caterpillar is defined to consist of all elements of degree greater than 1, thus the (at most) two outer vertices of the path have at least 1 leg attached to them. We order the caterpillar by size N, which is the total number of vertices in the caterpillar. Labeled caterpillars are caterpillars in which every vertex is labeled. In Figure 2.1 the caterpillars of size $N = 3, 4, 5, 6$ are shown, together with the number of different labels (LBL) they can obtain. The spine vertices in each caterpillar are solid black.

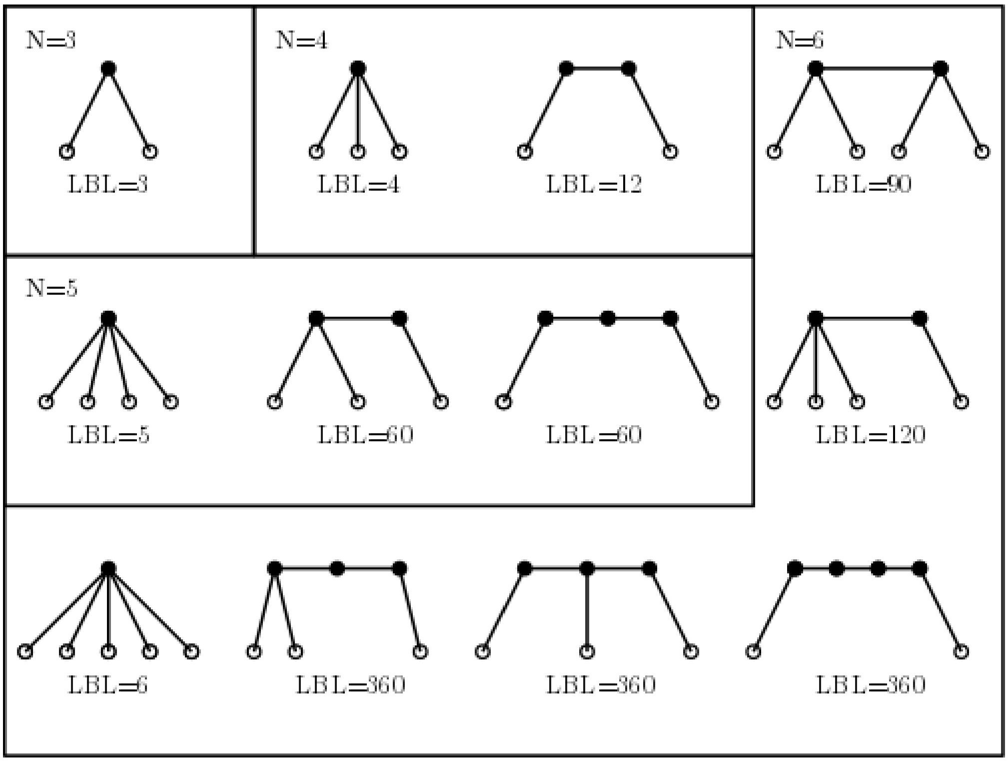

FIGURE 2.1: Some caterpillars of small order

The class of unlabeled caterpillars will be denoted by $\mathcal{UC}$ and the class of labeled caterpillars by $\mathcal{C}$. The set of all caterpillars of order n is denoted by $\mathcal{UC}_n$ in the unlabeled class and $\mathcal{C}_n$ in the labeled class and $UC_n \in_n \mathcal{UC}_n$,

$C_n \in_n \mathcal{C}_n$ represent the random uniformly chosen caterpillar of order n, unlabeled and labeled respectively.

Regarding Robertson-Seymour Theorem, the class of caterpillars $\mathcal{C}$ is nót minor-closed, as deleting an edge in the spine vertex could result in two disjoint caterpillars. However, the class of forest of caterpillars ís minor-closed, with excluded minors being the triangle K_3 and a particular tree containing a vertex attached to three vertices of degree 2, as seen in figure 2.2.

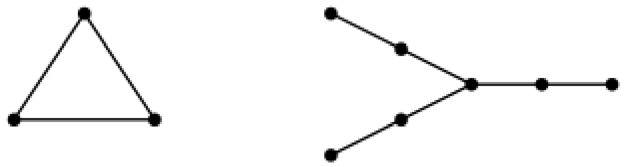

FIGURE 2.2: Excluded minors of $\mathcal{C}$

The random caterpillar UC_n differs from C_n. Let $\ell : \mathcal{C} \to \mathbb{N}$ be the function such that $\ell(C)$ is the spine length of C. Then, $\ell(C_n)$ is the random variable distributed as the spine length of $C_n \in_n \mathcal{C}_n$ is distributed. In Table 2.1 we see that $\mathbb{E}\ell(C_n)$ differs from $\mathbb{E}\ell(UC_n)$. So, the two classes and their associated random caterpillars clearly have different properties. From here, we'll mostly discuss the class of labeled caterpillars.

n	$\lvert\mathcal{UC}_n\rvert$	$\lvert\mathcal{C}_n\rvert$	$\mathbb{E}\ell(UC_n)$	$\mathbb{E}\ell(C_n)$
3	1	3	1	1
4	2	16	$\frac{3}{2}$	$\frac{7}{4}$
5	3	125	2	$\frac{61}{25}$
6	6	1296	$\frac{5}{2}$	$\frac{671}{216}$

TABLE 2.1: Labeled vs. unlabeled caterpillars

2.1.1 Spine length

As we have the class $\mathcal{C}$ of labeled caterpillars and the corresponding C_n caterpillars, we want to know the distribution of the spine length ℓ of C_n, as n goes to infinity. To ease calculations, we orient the caterpillars, having a left and a right side. This orientation can be easily reversed, since all caterpillars with a spine length bigger than 1, are isomorphic to exactly two oriented caterpillars. Let $S_{n,\ell}$ denote the total number of oriented, labeled caterpillars of order n with spine length $\ell \leq n-2$. When $\ell = 1$, we have $S_{n,\ell} = n$; all different caterpillars with spine length 1 are distinguished by the label assigned to the spine vertex.

For $S_{n,\ell}$, with $\ell \neq 1$, we have the exact expression:

$$
\begin{aligned}
S_{n,\ell} &= \binom{n}{\ell} \ell! \left(\ell^{n-\ell} - 2(\ell-1)^{n-\ell} + (\ell-2)^{n-\ell} \right) \\
&= \binom{n}{\ell} \ell! \ell^{(n-\ell)} \left(1 - 2\left(1 - \frac{1}{\ell}\right)^{n-\ell} + \left(1 - \frac{2}{\ell}\right)^{n-\ell} \right)
\end{aligned}
$$

First, ℓ labels out of n are picked for the spine, this gives the binomial $\binom{n}{\ell}$, then there are $\ell!$ different ways to label the spine with these ℓ labels. Further, the remaining $n-\ell$ vertices can be distributed freely over the ℓ spine vertices as legs, this accounts for the $\ell^{n-\ell}$ term. However, the two outer vertices should contain at least one vertex, which is accounted for between the brackets.

Having an exact number $S_{n,\ell}$ for all n, ℓ, we can see for fixed n, for which spine length ℓ there are the most caterpillars, that is, $S_{n,\ell}$ is largest. We can use this to determine the distribution of the spine length of the random caterpillar C_n. Clearly, since C_n is chosen randomly, it holds that:

$$\mathbb{P}(\ell(C_n) = \ell) = \frac{S_{n,\ell}}{|\mathcal{C}_n|} \tag{2.1}$$

As noted, a caterpillar of order n can have a spinelength of $1, \ldots, n-2$. Define $\mu_n = \{\frac{1}{n}, \ldots, \frac{n-2}{n}\}$, then $\mathbb{P}(\exists_{m \in \mu_n} \text{ s.t. } \ell(C_n) = mn) = 1$.

When $n \to \infty$, it turns out that:

Theorem 3 *For $c = \frac{1}{1+\rho}$, with ρ the unique solution of $xe^x = 1$:*

$$\lim_{n\to\infty} \sum_{m \in \mu_n \cap [c-\alpha, c+\alpha]} \mathbb{P}(\ell(C_n) = mn) = 1 \qquad \forall \alpha > 0$$

In other words, when n goes to infinity, the probability that a random caterpillar C_n has spine length arbitrary close to cn goes to 1 almost surely.

Proof We'll separate in two cases of the spine length ℓ first, being of a certain size and see how big $S_{n,\ell}$ is.

1. $\ell \leq \epsilon n$,
2. $\epsilon n < \ell \leq n-2$

where ϵ is fixed with $\epsilon < \frac{1}{e+1}$.

We'll first look at the first case. Basic calculation gives:

$$\begin{aligned}
\sum_{\ell \leq \epsilon n} S_{n,\ell} &= \sum_{\ell \leq \epsilon n} \binom{n}{\ell} \ell! \ell^{(n-\ell)} \left(1 - 2\left(1 - \frac{1}{\ell}\right)^{n-\ell} + \left(1 - \frac{2}{\ell}\right)^{n-\ell}\right) \\
&\leq \sum_{\ell \leq \epsilon n} \binom{n}{\ell} \ell! \ell^{(n-\ell)} \\
&= n! \sum_{\ell \leq \epsilon n} \frac{\ell^{n-\ell}}{(n-\ell)!}
\end{aligned}$$

And by Stirlings approximation in Theorem 2, we know that for large enough $(n-\ell)! \geq (\frac{n-\ell}{e})^{(n-\ell)}$, so we get:

$$\sum_{\ell \leq \epsilon n} S_{n,\ell} \leq n! \sum_{\ell \leq \epsilon n} \left(\frac{e\ell}{n-\ell}\right)^{(n-\ell)}$$

Since $\ell \leq \epsilon n$ and $n - \ell \geq (1-\epsilon)n$, we get:

$$\sum_{\ell \leq \epsilon n} S_{n,\ell} \leq n! \sum_{\ell \leq \epsilon n} \left(\frac{e\epsilon}{1-\epsilon}\right)^{n-\ell}$$

By choice of ϵ, $\frac{e\epsilon}{1-\epsilon} \leq 1$, and then we find that:

$$\sum_{\ell \leq \epsilon n} S_{n,\ell} \leq n!(\epsilon n)$$

Now, for the second case, we take ℓ s.t. $\epsilon n < \ell \leq n$ and for any such ℓ we can write $\ell = mn$, with $m \in \mu_n$ and m bounded away from zero, since it's bigger than ϵ. Then:

$$\begin{aligned}
S_{n,\ell} &= \binom{n}{\ell} \ell! \ell^{(n-\ell)} \left(1 - 2\left(1 - \frac{1}{\ell}\right)^{n-\ell} + \left(1 - \frac{2}{\ell}\right)^{n-\ell}\right) \\
&= \frac{n!}{(n-\ell)!} \ell^{n-\ell} \left(1 - 2\left(1 - \frac{1}{\ell}\right)^{n-\ell} + \left(1 - \frac{2}{\ell}\right)^{n-\ell}\right) \\
&= n! \frac{(mn)^{(1-m)n}}{((1-m)n)!} \left(1 - 2\left(1 - \frac{1}{mn}\right)^{(1-m)n} + \left(1 - \frac{2}{mn}\right)^{(1-m)n}\right)
\end{aligned}$$

For any fixed n and $m \in \mu_n, m > \epsilon$, we can use Stirling's approximation to find:

$$((1-m)n)! \sim \left(\frac{(1-m)n}{e}\right)^{(1-m)n} \sqrt{2\pi(1-m)n}$$

So we can write:

$$\begin{aligned}
\frac{(mn)^{(1-m)n}}{((1-m)n)!} &\sim \frac{1}{\sqrt{2\pi(1-m)n}} \left(\frac{e}{(1-m)n}\right)^{(1-m)n} (mn)^{(1-m)n} \\
&\sim \frac{1}{\sqrt{2\pi(1-m)n}} \left(\frac{em}{1-m}\right)^{(1-m)n}
\end{aligned}$$

This leads to:

$$S_{n,mn} \sim n! \frac{1}{\sqrt{2\pi(1-m)n}} \left(\frac{em}{1-m}\right)^{(1-m)n} \left(1 - 2\left(1 - \frac{1}{mn}\right)^{(1-m)n} + \left(1 - \frac{2}{mn}\right)^{(1-m)n}\right)$$

For logarithms $\log(1-x)$, a series expansion around 1 results in $\log(1-x) = -x + \mathcal{O}(x^2)$. And, we can write:

$$\left(1 - \frac{1}{mn}\right)^{(1-m)n} = e^{(1-m)n \log(1 - \frac{1}{mn})} \quad \left(1 - \frac{2}{mn}\right)^{(1-m)n} = e^{(1-m)n \log(1 - \frac{2}{mn})}$$

Applying the expansion, using that $mn > \epsilon n$, thus $\frac{1}{mn} < \frac{1}{\epsilon n} \ll 1$, to get:

$$\left(1 - \frac{1}{mn}\right)^{(1-m)n} = e^{(1-m)n\left(-\frac{1}{mn} + \mathcal{O}(\frac{1}{(mn)^2})\right)} \qquad = e^{-\frac{1-m}{m} + \mathcal{O}(\frac{1}{n})}$$
$$\left(1 - \frac{2}{mn}\right)^{(1-m)n} = e^{(1-m)n\left(-\frac{2}{mn} + \mathcal{O}(\frac{1}{(mn)^2})\right)} \qquad = e^{-\frac{2(1-m)}{m} + \mathcal{O}(\frac{1}{n})}$$

Using this to find:

$$1 - 2\left(1 - \frac{1}{mn}\right)^{(1-m)n} + \left(1 - \frac{2}{mn}\right)^{(1-m)n} = 1 - 2e^{-\frac{1-m}{m} + \mathcal{O}(\frac{1}{n})} + e^{-\frac{2(1-m)}{m} + \mathcal{O}(\frac{1}{n})}$$
$$= \left(1 - e^{-\frac{1-m}{m} + \mathcal{O}(\frac{1}{n})}\right)^2$$

So we can write:

$$S_{n,mn} \sim n! \frac{1}{\sqrt{2\pi(1-m)n}} \left(\frac{em}{1-m}\right)^{(1-m)n} \left(1 - e^{-\frac{1-m}{m} + \mathcal{O}(\frac{1}{n})}\right)^2$$

We now have estimates of the number of caterpillars of certain spine length. That is, we know that:

$$\ell \le \epsilon n \qquad \sum_{\ell \le \epsilon n} S_{n,\ell} \le n!(\epsilon n)$$
$$\epsilon n < \ell (= mn) \quad S_{n,\ell} \sim \frac{n!}{\sqrt{2\pi(1-m)n}} \left(\frac{em}{1-m}\right)^{(1-m)n} \left(1 - e^{-\frac{1-m}{m} + \mathcal{O}(\frac{1}{n})}\right)^2$$

When we look at $S_{n,\ell n}$ for larger n, the caterpillars with spine length less than ϵn are outnumbered by those with bigger spine length. There are only $n!\mathcal{O}(n)$ caterpillars in total with spine length $\ell \le \epsilon n$. On the other hand, for $\ell > \epsilon n$, $S_{n,\ell} \sim n!p(n)A^n$, with $A = (\frac{em}{1-m})^{1-m} > 1$. This exponential term will be dominating for large n, since exponential growth is faster than polynomial growth. Since $|C_n| = \sum_{\ell \le \epsilon n} S_{n,\ell} + \sum_{\ell > \epsilon n} S_{n,\ell}$, the first summation becomes negligible for large n.

Recall equation 2.1 and rewrite this to get:

$$\mathbb{P}_n(\ell(C_n) = mn) = \frac{S_{n,mn}}{|\mathcal{C}_n|} = \frac{S_{n,mn}}{\sum_k S_{n,kn}} = \frac{1}{\sum_k \frac{S_{n,kn}}{S_{n,mn}}} \tag{2.2}$$

For any n, we can look at the the ratio between $S_{n,mn}$ and $S_{n,\tilde{m}n}$, with $m, \tilde{m} \in \mu_n$ and $m, \tilde{m} > \epsilon$, which is given by:

$$\frac{S_{n,mn}}{S_{n,\tilde{m}n}} \sim \left(\left(\frac{em}{1-m}\right)^{1-m} \left(\frac{1-\tilde{m}}{e\tilde{m}}\right)^{1-\tilde{m}}\right)^n \sqrt{\frac{1-\tilde{m}}{1-m}} \left(\frac{1 - e^{-\frac{1-m}{m} + \mathcal{O}(\frac{1}{n})}}{1 - e^{-\frac{1-\tilde{m}}{\tilde{m}} + \mathcal{O}(\frac{1}{n})}}\right)^2$$
$$= A^n B$$

Where

$$A = \left(\frac{em}{1-m}\right)^{1-m}\left(\frac{1-\tilde{m}}{e\tilde{m}}\right)^{1-\tilde{m}} \qquad B = \sqrt{\frac{1-\tilde{m}}{1-m}}\left(\frac{1-e^{-\frac{1-m}{m}+\mathcal{O}(\frac{1}{n})}}{1-e^{-\frac{1-\tilde{m}}{\tilde{m}}+\mathcal{O}(\frac{1}{n})}}\right)^2$$

In this ratio, as n becomes large, A^n will be the dominant term, either becoming arbitrarily large or getting arbitrarily close to 0. For A, we know that:

$$A > 1 \iff \left(\frac{em}{1-m}\right)^{1-m} > \left(\frac{e\tilde{m}}{1-\tilde{m}}\right)^{1-\tilde{m}} \tag{2.3}$$

$$A < 1 \iff \left(\frac{em}{1-m}\right)^{1-m} < \left(\frac{e\tilde{m}}{1-\tilde{m}}\right)^{1-\tilde{m}} \tag{2.4}$$

Define the function f to be:

$$f : (0,1) \to \mathbb{R} \qquad f(x) = \left(\frac{ex}{1-x}\right)^{1-x}$$

When $f(x)$ attains a maximum in x_0, the function $\log f(x) = (1-x)(\log x + 1 - \log(1-x))$ will also attain a maximum at x_0. A basic calculation yields:

$$0 = \frac{d}{dx}\log f(x) = -\log\frac{x}{1-x} - 1 + (1-x)\left(\frac{1}{x}+\frac{1}{1-x}\right) \implies \log\frac{1-x}{x} + \frac{1-x}{x} = 0$$

This is only true when $\frac{1-x}{x} = \rho$, the solution of $ze^z = 1$. This solution is unique, as ze^z is increasing on $\mathbb{R}_{\geq 0}$. As $\frac{1-x}{x} = \rho$, we get $x = \frac{1}{1+\rho}$ as the unique point on $(0,1)$ where $f(x)$ attains its maximum value - to be called c from this point on. That this is indeed a maximal value, is imminent as $f(\epsilon) = 1$ and $f(\frac{1}{2}) = \sqrt{e} > 1$, while $c > \frac{1}{2}$. This is also the only local maximum, thus f is increasing on $(0,c)$ and decreasing on $(c,1)$, meaning that $f(x) < f(y)$, when $0 < x < y < c$ and $f(x) > f(y)$ when $c < x < y < 1$.

We can use this in relation to the ratio $\frac{S_{n,mn}}{S_{n,\tilde{m}n}}$. Pick $\alpha > 0$. By construction of μ_n, there's an N s.t. for every $n \geq N$, $[c-\frac{\alpha}{2}, c] \cap \mu_n \neq \emptyset$ and $[c, c+\frac{\alpha}{2}] \cap \mu_n \neq \emptyset$. When we have an $m \in \mu_n$ with $m \in (0, c-\alpha)$, and a $\tilde{m} \in \mu_n$ with $\tilde{m} \in (c-\frac{\alpha}{2}, c]$, we can find an upper bound for the corresponding A in the ratio $\frac{S_{n,mn}}{S_{n,\tilde{m}n}}$:

$$A = \left(\frac{em}{1-m}\right)^{1-m}\left(\frac{1-\tilde{m}}{e\tilde{m}}\right)^{1-\tilde{m}} \leq \left(\frac{e(c-\alpha)}{1-(c-\alpha)}\right)^{1-(c-\alpha)}\left(\frac{1-(c-\frac{\alpha}{2})}{e(c-\frac{\alpha}{2})}\right)^{1-(c-\frac{\alpha}{2})}$$

And define this upper bound as A_α. Recall (2.14) and rewrite this, for $m, \tilde{m} \in \mu_n$ and $m \leq c-\alpha$ and $\tilde{m} \in [c-\frac{\alpha}{2}, c]$:

$$\mathbb{P}_n(\ell(C_n)) = \frac{S_{n,mn}}{\sum_k S_{n,kn}} \leq \frac{S_{n,mn}}{S_{n,\tilde{m}n}} \leq A_\alpha^n B(\frac{1}{n})$$

Then, the probability that $\ell(C_n) < (c-\alpha)n$, when $n \to \infty$, is given by:

$$\lim_{n\to\infty} \mathbb{P}_n(\ell(C_n) < (c-\alpha)n) = \lim_{n\to\infty} \sum_{m\in\mu_n\cap[0,c-\alpha]} \mathbb{P}_n(\ell(C_n) = mn)$$
$$\leq \lim_{n\to\infty} (c-\alpha)nA^nB = 0$$

As $A < 1$ and exponential decline dominates the behavior of $B(n)$. Analogously, it can be shown that $\lim_{n\to\infty} \mathbb{P}_n(\ell(C_n) > (c-\alpha)n) = 0$, thus resulting in:

$$\lim_{n\to\infty} \sum_{m\in\mu_n\cap[c-\alpha,c+\alpha]} \mathbb{P}(\ell(C_n) = mn) = 1$$

Importantly, this result holds for all $\alpha > 0$.

$\square$

2.2 Distribution of the legs

Knowing the distribution of the spine length of a random caterpillar, we can also look at the distribution of its legs. If we would pick a random spine vertex v of the random caterpillar C_n, we want to determine the probability that $d(v) = k+2$, i.e., that v has exactly k legs attached to it.

2.2.1 Single vertex

We want to determine for a spine vertex v and a fixed $k \in \mathbb{N}$, the existence and the value of:

$$\lim_{n\to\infty} \mathbb{P}_n(d(v) = k)$$

However, this probability depends on the choice of v. As we take limit over the caterpillars, two different caterpillars obviously have different spine vertices. We can chose to pick the first vertex in every oriented caterpillar C_n, or we chose a regular spine vertex of C_n randomly, and this gives different results. If we enumerate all spine vertices of an oriented random caterpillar C_n as $v_1, \ldots, v_\ell$, with $\ell = \ell(C_n)$, where v_1 will be the left-most vertex, then, for the distribution of $d(v_1)$ in the random caterpillar C_n, we have, for $n \geq 3$:

$$\mathbb{P}_n(d(v_1) = k+1) = \sum_{\ell=1}^{n-2} \mathbb{P}_n(d(v_1) = k+1 | \ell(C_n) = \ell)\mathbb{P}_n(\ell(C_n) = \ell) \quad (2.5)$$

Notice that we sum from 1 to $n-2$, as the outer vertices must both have at least 1 leg, and those leg vertices can't be part of the spine. Caterpillars of order 1 or 2 have no spine vertices at all.

Lemma 2 *Let v_1 be the vertex of the spine of C_n (often called an 'outer vertex' in this thesis) and fix k. Then $\lim_{n\to\infty} \mathbb{P}_n(d(v_1) = k+1)$ exists and is equal to* $\frac{\rho^{k+1}}{k!}(1-\rho)^{-1}$

Proof As we saw in Theorem 3, there's a constant value c such that the expected spine length $\mathbb{E}[\ell(C_n)]$ get's arbitrarily close to cn, as $n \to \infty$ and $\lim_{n\to\infty} \frac{\ell(C_n)}{n} = c$ with high probability.

Thus, $\lim_{n\to\infty} \mathbb{P}_n(\ell(C_n) \notin ((c-\alpha)n, (c+\alpha)n) = 0, \forall \alpha > 0$. Applying this to (2.5), we get:

$$\lim_{n\to\infty} \mathbb{P}_n(d(v_1) = k+1) \tag{2.6}$$

$$= \lim_{n\to\infty} \sum_{\ell=1}^{n-2} \mathbb{P}_n(d(v_1) = k+1 | \ell(C_n) = \ell)\mathbb{P}_n(\ell(C_n) = \ell) \tag{2.7}$$

$$= \lim_{n\to\infty} \sum_{m \in \mu_n \cap ((c-\alpha),(c+\alpha))} \mathbb{P}_n(d(v_1) = k+1 | \ell(C_n) = mn)\mathbb{P}_n(\ell(C_n) = mn) \tag{2.8}$$

To further determine this limit, we need to calculate $\mathbb{P}(d(v_1) = k+1|\ell(C_n) = \ell)$, which can be done by an analysis similar to finding $S_{n,\ell}$. Assume that $\ell > \epsilon n$, with $\epsilon > 0$, which we can do, since the number of caterpillars with small spine length is neglectable. The ℓ vertices in the spine can be labeled in $\binom{n}{\ell}\ell!$ ways, the k legs in $\binom{n-\ell}{k}$ ways and the remaining $n-\ell-k$ vertices can be distributed over $\ell - 1$ spine vertices. Thus, for $k \geq 1$, we have:

$$\begin{aligned}
&\mathbb{P}_n(d(v_1) = k+1 | \ell(C_n) = \ell) \\
&= \frac{1}{S_{n,\ell}} \binom{n}{\ell} \ell! \binom{n-\ell}{k} \left((\ell-1)^{n-\ell-k} - (\ell-2)^{n-\ell-k} \right) \\
&= \frac{1}{S_{n,\ell}} \frac{n!}{(n-\ell-k)!k!} (\ell-1)^{n-\ell-k} \left(1 - \left(1 - \frac{1}{\ell-1} \right)^{n-\ell-k} \right)
\end{aligned}$$

Applying the expression found for $S_{n,\ell}$ earlier, we get:

$$\mathbb{P}_n(d(v_1) = k+1 | \ell(C_n) = \ell) \tag{2.9}$$

$$= \frac{(n-\ell)!}{(n-\ell-k)!k!} \frac{(\ell-1)^{n-\ell-k}}{\ell^{n-\ell}} \frac{1 - \left(1 - \frac{1}{\ell-1}\right)^{n-\ell-k}}{1 - 2\left(1 - \frac{1}{\ell}\right)^{n-\ell} + \left(1 - \frac{2}{\ell}\right)^{n-\ell}} \tag{2.10}$$

$$= \binom{n-\ell}{k} \left(\frac{\ell-1}{\ell} \right)^{n-\ell} \frac{1}{(\ell-1)^k} \frac{1 - \left(1 - \frac{1}{\ell-1}\right)^{n-\ell-k}}{1 - 2\left(1 - \frac{1}{\ell}\right)^{n-\ell} + \left(1 - \frac{2}{\ell}\right)^{n-\ell}} \tag{2.11}$$

We can rewrite a part of this, using that $\exists m \in \mu_n$ s.t. $\ell = mn$:

$$\begin{aligned}
&\frac{(n-\ell)!}{(n-\ell-k)!k!} \frac{(\ell-1)^{n-\ell-k}}{\ell^{n-\ell}} \frac{1}{(\ell-1)^k} \\
&= \frac{1}{k!} \frac{n-\ell}{\ell-1} \cdots \frac{n-\ell-(k-1)}{\ell-1} \left(\frac{\ell-1}{\ell} \right)^{n-\ell} \\
&= \frac{1}{k!} \frac{(1-m)n}{mn-1} \cdots \frac{(1-m)n-(k-1)}{mn-1} \left(1 - \frac{1}{mn} \right)^{(1-m)n}
\end{aligned}$$

We can write:

$$\frac{(1-m)n}{mn-1}\cdots\frac{(1-m)n-(k-1)}{mn-1} = \prod_{i=0}^{k-1}\frac{(1-m)n-i}{mn-1}$$
$$= \prod_{i=0}^{k-1}\left(\frac{1-m}{m}\left(1+\frac{1}{mn-1}\right)-\frac{i}{mn-1}\right)$$
$$= \left(\frac{1-m}{m}\right)^k\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right)$$

Using that, for $x \ll 1$, we have the series expansion $\log(1-x) = -x+\mathcal{O}(x^2)$, we can write:

$$\left(1-\frac{1}{mn}\right)^{(1-m)n} = \exp(1-m)n\log\left(1-\frac{1}{mn}\right) = e^{-\frac{1-m}{m}+\mathcal{O}(\frac{1}{n})}$$

And in a similar way, we can rewrite the big fraction in (2.11):

$$\frac{1-\left(1-\frac{1}{\ell-1}\right)^{n-\ell-k}}{1-2\left(1-\frac{1}{\ell}\right)^{n-\ell}+\left(1-\frac{2}{\ell}\right)^{n-\ell}} = \frac{1-\left(1-\frac{1}{mn-1}\right)^{(1-m)n-k}}{1-2\left(1-\frac{1}{mn}\right)^{(1-m)n}+\left(1-\frac{2}{\ell}\right)^{(1-m)n}}$$
$$= \frac{1-e^{-\frac{1-m}{m}+\mathcal{O}(\frac{1}{n})}}{1-2e^{-\frac{1-m}{m}+\mathcal{O}(\frac{1}{n})}+e^{-2\frac{1-m}{m}+\mathcal{O}(\frac{1}{n})}}$$
$$= \frac{1}{1-e^{-\frac{1-m}{m}+\mathcal{O}(\frac{1}{n})}}$$

Combining all this, gives:

$$\mathbb{P}_n(d(v_1)=k+1|\ell(C_n)=\ell) = \frac{1}{k!}\left(\frac{1-m}{m}\right)^k e^{-\frac{1-m}{m}+\mathcal{O}(\frac{1}{n})}\frac{1+\mathcal{O}\left(\frac{1}{n}\right)}{1-e^{-\frac{1-m}{m}+\mathcal{O}(\frac{1}{n})}}$$

Applying this to (2.8), and combining all the $\mathcal{O}(\frac{1}{n})$ terms, gives:

$$\lim_{n\to\infty}\mathbb{P}_n(d(v_1)=k+1) \tag{2.12}$$
$$= \lim_{n\to\infty}\sum_{m\in\mu_n\cap(c-\alpha,c+\alpha)}\frac{1}{k!}\left(\frac{1-m}{m}\right)^k\frac{e^{-\frac{1-m}{m}}}{1-e^{-\frac{1-m}{m}}}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right)\mathbb{P}_n(\ell(C_n)=mn) \tag{2.13}$$

We only sum over m in $(c-\alpha, c+\alpha)$, so we can find upper and lower values of every term in the summation. So:

$$\forall m \in (c-\alpha, c+\alpha)\quad \frac{1-(c+\alpha)}{c+\alpha} \leq \frac{1-m}{m} \leq \frac{1-(c-\alpha)}{c-\alpha}$$

Hence:

$$\left(\frac{1-(c+\alpha)}{c+\alpha}\right)^k\frac{e^{-\frac{1-(c-\alpha)}{c-\alpha}}}{1-e^{-\frac{1-(c-\alpha)}{c-\alpha}}} \leq \left(\frac{1-m}{m}\right)^k\frac{e^{-\frac{1-m}{m}}}{1-e^{-\frac{1-m}{m}}} \leq \left(\frac{1-(c-\alpha)}{c-\alpha}\right)^k\frac{e^{-\frac{1-(c+\alpha)}{c+\alpha}}}{1-e^{-\frac{1-(c+\alpha)}{c+\alpha}}}$$

Which means that:

$$\lim_{n\to\infty}\left(\frac{1-(c+\alpha)}{c+\alpha}\right)^k \frac{e^{-\frac{1-(c-\alpha)}{c-\alpha}}}{1-e^{-\frac{1-(c-\alpha)}{c-\alpha}}} \sum_{m\in\mu_n\cap(c-\alpha,c+\alpha)} \mathbb{P}_n(\ell(C_n)=mn)\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right)$$
$$\leq \lim_{n\to\infty} \sum_{m\in\mu_n\cap(c-\alpha,c+\alpha)} \mathbb{P}_n(d(v_1)=k|\ell(C_n)=mn)\mathbb{P}(\ell(C_n)=mn)$$
$$\leq \lim_{n\to\infty}\left(\frac{1-(c-\alpha)}{c-\alpha}\right)^k \frac{e^{-\frac{1-(c+\alpha)}{c+\alpha}}}{1-e^{-\frac{1-(c+\alpha)}{c+\alpha}}} \sum_{m\in\mu_n\cap(c-\alpha,c+\alpha)} \mathbb{P}_n(\ell(C_n)=mn)\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right)$$

As we may choose α arbitrarily close to zero, we have:

$$\lim_{\alpha\downarrow 0}\frac{1}{k!}\left(\frac{1-(c+\alpha)}{c+\alpha}\right)^k \frac{e^{-\frac{1-(c-\alpha)}{c-\alpha}}}{1-e^{-\frac{1-(c-\alpha)}{c-\alpha}}} \leq \lim_{n\to\infty}\mathbb{P}_n(d(v_1)=k+1|\ell(C_n)=\ell)$$
$$\leq \lim_{\alpha\downarrow 0}\frac{1}{k!}\left(\frac{1-(c-\alpha)}{c-\alpha}\right)^k \frac{e^{-\frac{1-(c+\alpha)}{c+\alpha}}}{1-e^{-\frac{1-(c+\alpha)}{c+\alpha}}}$$

Evaluating the limits yields:

$$\frac{1}{k!}\left(\frac{1-c}{c}\right)^k \frac{e^{-\frac{1-c}{c}}}{1-e^{-\frac{1-c}{c}}} \leq \lim_{n\to\infty}\mathbb{P}_n(d(v_1)=k+1|\ell(C_n)=\ell) \quad \leq \frac{1}{k!}\left(\frac{1-c}{c}\right)^k \frac{e^{-\frac{1-c}{c}}}{1-e^{-\frac{1-c}{c}}}$$

So the inequalities are actually equalities. When we use that $\frac{1-c}{c}=\rho$, with $\rho = e^{-\rho}$, we get:

$$\lim_{n\to\infty}\mathbb{P}_n(d(v_1)=k+1|\ell(C_n)=\ell) = \frac{1}{k!}\rho^k\frac{e^{-\rho}}{1-e^{-\rho}} = \frac{1}{1-\rho}\frac{\rho^{k+1}}{k!}$$

Thus, we've proven that:

$$\lim_{n\to\infty}\mathbb{P}_n(d(v_1)=k+1|\ell(C_n)=\ell) = \frac{1}{1-\rho}\frac{\rho^{k+1}}{k!}$$

□

We now want to establish a similar result, but for spine vertices that aren't outer vertices. First, we'll define how to pick a random spine vertex on C_n.

Definition 3 *A random picked spine vertex v, not an outer one, on a caterpillar $C_n \in_n \mathcal{C}_n$, is uniformly random picked out of the not-outer spine vertices in the random picked caterpillar C_n. Thus, the distribution of the degree $d(v)$ of vertex v is given by:*

$$\mathbb{P}_n(d(v)=k+2) = \sum_{\ell=3}^{n-k-5}\mathbb{P}_n(d(v)=k+2|\ell(C_n)=\ell)\mathbb{P}_n(\ell(C_n)=\ell) \quad (2.14)$$

This leads to the following lemma:

Lemma 3 *For a random, non outer spine vertex v on C_n, we have:*

$$\lim_{n\to\infty}\mathbb{P}_n(d(v)=k+2) = \frac{\rho^{k+1}}{k!} \quad \forall k\in\mathbb{N}$$

where ρ is the solution of $xe^x = 1$. This also holds for spine vertex v_i, the fixed i-th spine vertex from the left in an oriented caterpillar.

Proof We'll first prove the statement with the randomly chosen spine vertex. The proof is similar to that of Lemma 2. Fix $k \in \mathbb{N}$. We can rewrite, when n approaches infinity, (2.14), using Theorem 3, for any $\alpha > 0$ and c the limit spine length of random caterpillars:

$$\lim_{n\to\infty} \mathbb{P}_n(d(v) = k+2) = \lim_{n\to\infty} \sum_{m\in\mu_n\cap[c-\alpha,c+\alpha]} \mathbb{P}(d(v) = k+2|\ell(C_n) = mn)\mathbb{P}(\ell(C_n = mn))$$

When a caterpillar has spine length mn, then a non outer spine vertex v has degree $d(v) = k+2$ with probability:

$$\mathbb{P}_n(d(v) = k+2|\ell(C_n) = mn) \tag{2.15}$$

$$= \frac{n}{S_{n,mn}}\binom{n-1}{k}\binom{n-k-1}{mn-1}(mn-1)!(mn-1)^{(1-m)n-k} \tag{2.16}$$

$$\cdot \left(1-2\left(1-\frac{1}{mn-1}\right)^{(1-m)n-k} + \left(1-\frac{2}{mn-1}\right)^{(1-m)n-k}\right) \tag{2.17}$$

$$= \frac{1}{S_{n,mn}} \frac{n!(mn-1)^{(1-m)n-k}}{((1-m)n-k)!k!}\left(1-2\left(1-\frac{1}{mn-1}\right)^{(1-m)n-k} + \left(1-\frac{2}{mn-1}\right)^{(1-m)n-k}\right) \tag{2.18}$$

For a fixed non-outer vertex, we can label it in n different ways, its k legs in $\binom{n-1}{k}$ ways, the remaining $mn-1$ spine vertices in $\binom{n-k-1}{mn-1}(mn-1)!$, since we work with oriented caterpillars, and the remaining $(1-m)n-k$ vertices are legs of the other spine vertices, where the outer vertices must have at least one leg each. This is the amount of caterpillars of order n with spine length mn where v is of degree $d(v) = k+2$. Divided by the total number of caterpillars with spine length mn, gives the probability that $d(v) = k+2$ given spine length mn.

We have an exact formula for $S_{n,mn}$, and we can apply this to get:

$$\mathbb{P}_n(d(v) = k+2|\ell(C_n) = mn)$$

$$= \binom{(1-m)n}{k}\left(\frac{mn-1}{mn}\right)^{(1-m)n}\frac{1}{(mn-1)^k}\frac{1-2\left(1-\frac{1}{mn-1}\right)^{(1-m)n-k}+\left(1-\frac{2}{mn-1}\right)^{(1-m)n-k}}{1-2\left(1-\frac{1}{mn}\right)^{(1-m)n}+\left(1-\frac{2}{mn}\right)^{(1-m)n}}$$

By the same analysis as in the proof of Lemma 2, we can write:

$$\begin{aligned}
\binom{(1-m)n}{k}\frac{1}{(mn-1)^k} &= \frac{1}{k!}\frac{(1-m)n}{mn-1}\cdots\frac{(1-m)n-k+1}{mn-1} \\
&= \frac{1}{k!}\left(\frac{1-m}{m}\right)^k\left(1+\mathcal{O}(\frac{1}{n})\right) \\
\left(\frac{mn-1}{mn}\right)^{(1-m)n} &= \left(1-\frac{1}{mn}\right)^{(1-m)n} \\
&= e^{-\frac{1-m}{m}+\mathcal{O}(\frac{1}{n})} \\
\frac{1-2\left(1-\frac{1}{mn-1}\right)^{(1-m)n-k}+\left(1-\frac{2}{mn-1}\right)^{(1-m)n-k}}{1-2\left(1-\frac{1}{mn}\right)^{(1-m)n}+\left(1-\frac{2}{mn}\right)^{(1-m)n}} & \\
&= \frac{1-2e^{-\frac{1-m}{m}+\mathcal{O}(\frac{1}{n})}+e^{-2\frac{1-m}{m}+\mathcal{O}(\frac{1}{n})}}{1-2e^{-\frac{1-m}{m}+\mathcal{O}(\frac{1}{n})}+e^{-2\frac{1-m}{m}+\mathcal{O}(\frac{1}{n})}} \\
&= 1+\mathcal{O}\left(\frac{1}{n}\right)
\end{aligned}$$

Combining all this, including the $\mathcal{O}(\frac{1}{n})$ terms, results in:

$$\mathbb{P}_n(d(v)=k+2|\ell(C_n)=mn) = \frac{1}{k!}\left(\frac{1-m}{m}\right)^k e^{-\frac{1-m}{m}}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{2.19}$$

Hence, we get:

$$\begin{aligned}
\lim_{n\to\infty}\mathbb{P}_n(d(v)=k+2) &= \lim_{n\to\infty}\sum_{m\in\mu_n\cap[c-\alpha,c+\alpha]}\mathbb{P}(d(v)=k+2|\ell(C_n)=mn)\mathbb{P}(\ell(C_n=mn)) \\
&= \lim_{n\to\infty}\sum_{m\in\mu_n\cap[c-\alpha,c+\alpha]}\frac{1}{k!}\left(\frac{1-m}{m}\right)^k e^{-\frac{1-m}{m}}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right)
\end{aligned}$$

As $m\in[c-\alpha,c+\alpha]$, we can find bounds:

$$\frac{1}{k!}\left(\frac{1-(c+\alpha)}{c+\alpha}\right)^k e^{-\frac{1-(c-\alpha)}{c-\alpha}} \le \frac{1}{k!}\left(\frac{1-m}{m}\right)^k e^{-\frac{1-m}{m}} \le \frac{1}{k!}\left(\frac{1-(c-\alpha)}{c-\alpha}\right)^k e^{-\frac{1-(c+\alpha)}{c+\alpha}}$$

And as this holds for any $\alpha>0$, we can use the same argument as in the proof of the previous Lemma, taking the limit of $\alpha\downarrow 0$ on both sides, to find:

$$\lim_{n\to\infty}\mathbb{P}_n(d(v)=k+2) = \frac{1}{k!}\left(\frac{1-c}{c}\right)^k e^{-\frac{1-c}{c}} = \frac{\rho^{k+1}}{k!}$$

As:

$$\frac{1-c}{c}=\rho \qquad\qquad e^{-\frac{1-c}{c}}=e^{-\rho}=\rho$$

To see that this also holds for spine vertex v_i, for fixed i, notice that the formula to determine the probability of being of degree $k+2$, given the spine length $\ell=mn$, and with this spine length bigger than i, is equal to the probability given for a randomly chosen spine vertex, also with spine length ℓ, which is given in (2.15). As the spine vertex will be bigger than i with probability going to 1, as $n \to \infty$, we can apply the same arguments to complete the proof for a fixed spine vertex v_i.

$\square$

For an oriented caterpillar with spine length ℓ, let $v_1, \ldots, v_\ell$ be the spine vertices from left to right.

Lemma 4 *Fix $i, r \in \mathbb{N}$ and let C_n be a uniformly random caterpillar with sufficiently large (larger than $i + r$) spine length. Fix $k_0, \ldots, k_{r-1} \in \mathbb{N}$ and take $K = \sum_i k_i$. When $i > 1$, thus the sequence doesn't start at the first vertex of the spine, we get:*

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}_n(d(v_i) = k_0 + 2, \ldots, d(v_{i+r-1}) = k_{r-1} + 2) &= \lim_{n\to\infty} \prod_{j=0}^{r-1} \mathbb{P}(d(v_{i+j}) = k_j + 2) \\ &= \frac{\rho^{r+K}}{k_0! \cdots k_{r-1}!}\end{aligned}$$

When $i = 1$, we would get:

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}_n(d(v_i) = k_0 + 2, \ldots, d(v_{i+r-1}) = k_{r-1} + 2) &= \lim_{n\to\infty} \prod_{j=0}^{r-1} \mathbb{P}(d(v_{i+j}) = k_j + 2) \\ &= \frac{1}{1-\rho} \frac{\rho^{r+K+1}}{(k_0+1)! \cdots k_{r-1}!}\end{aligned}$$

Proof As the proof of this Lemma is similar to that of Lemma 2, we'll only prove the statement for $i > 1$. The statement for $i = 1$ is completely analogous, only adjusting with v_1 being an outer vertex Again, we use that $\ell = mn$, with $m \in \mu_n$ and that when $n \to \infty$, all caterpillars will have a spine length close to cn. We also condition on C_n of having a spine length $\ell = i + r$, but as this is true with high probability when $n \to \infty$, it's not incorporated into the formula for reading purposes.

$$\begin{aligned}&\lim_{n\to\infty} \mathbb{P}_n(d(v_i) = k_0 + 2, \ldots, d(v_{i+r-1}) = k_{r-1} + 2) \\ &= \lim_{n\to\infty} \sum_{m\in\mu_n\cap[c-\alpha,c+\alpha]} \mathbb{P}_n(d(v_{i+j}) = k_j + 2, \ \forall j \le r-1 | \ell(C_n) = mn) \mathbb{P}_n(\ell(C_n) = mn)\end{aligned}$$

The r segments with $k_0, \ldots, k_{r-1}$ legs in each segment respectively, can be labeled in L different ways, where:

$$L = \binom{n}{r} r! \binom{n-r}{k_0} \cdots \binom{n-r-K+k_{r-1}}{k_{r-1}} = \frac{n!}{k_0! \cdots k_{r-1}!(n-r-K)!}$$

and the remaining vertices can be distributed over the remaining spine vertices. So:

$$\begin{aligned}
&\mathbb{P}_n(d(v_{i+j}) = k_j + 2,\ \ \forall j \le r-1 | \ell(C_n) = mn) \\
&= \frac{L}{S_{n,mn}} \binom{n-K-r}{mn-r} (mn-r)!(mn-r)^{n-K-mn} \\
&\cdot \left(1 - 2\left(1 - \frac{1}{mn-r}\right)^{n-K-mn} + \left(1 - \frac{2}{mn-r}\right)^{n-K-mn}\right) \\
&= \frac{1}{\prod_{i=2}^{r+1} k_i!} \frac{(n-mn)!}{(n-K-mn)!} \frac{(mn-r)^{n-K-mn}}{mn^{n-mn}} \\
&\cdot \frac{1 - 2\left(1 - \frac{1}{mn-r}\right)^{n-K-mn} + \left(1 - \frac{2}{mn-r}\right)^{n-K-mn}}{1 - 2(1-\frac{1}{mn})^{n-mn} + (1-\frac{2}{mn})^{n-mn}}
\end{aligned}$$

Similar to Lemma 2, we can write:

$$\frac{(n-mn)!}{(n-K-mn)!}(mn-r)^{-K} = \frac{(n-mn)}{mn-r} \cdots \frac{(n-mn-K+1)}{mn-r} \tag{2.20}$$

$$= \left(\frac{1-m}{m}\right)^K \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) \tag{2.21}$$

$$\left(\frac{mn-r}{mn}\right)^{n-mn} = \left(1 - \frac{r}{mn}\right)^{(1-m)n} = e^{-r\frac{1-m}{m} + \mathcal{O}(\frac{1}{n})} \tag{2.22}$$

And, as $(1 - \frac{1}{mn})^{n-mn} = e^{-\frac{1-m}{m} + \mathcal{O}(\frac{1}{n})}$, and similar for $(1 - \frac{1}{mn-r})^{n-mn-K}$, we have:

$$\frac{1 - 2\left(1 - \frac{1}{mn-r}\right)^{n-K-mn} + \left(1 - \frac{2}{mn-r}\right)^{n-K-mn}}{1 - 2(1-\frac{1}{mn})^{n-mn} + (1-\frac{2}{mn})^{n-mn}} = 1 + \mathcal{O}\left(\frac{1}{n}\right) \tag{2.23}$$

When we combine all this, we see that:

$$\mathbb{P}_n(d(v_{i+j}) = k_j + 2,\ \ \forall j \le r-1 | \ell(C_n) = mn) = \frac{1}{\prod_j k_j!} \left(\frac{1-m}{m}\right)^K e^{-r\frac{1-m}{m}} \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right)$$

As we sum over $m \in [c-\alpha, c+\alpha]$, we get boundaries:

$$\begin{aligned}
\frac{1}{\prod_j k_j!} \left(\frac{1-(c+\alpha)}{c+\alpha}\right)^K e^{-r\frac{1-(c-\alpha)}{c-\alpha}} &\le \frac{1}{\prod_j k_j!} \left(\frac{1-m}{m}\right)^K e^{-r\frac{1-m}{m}} \\
&\le \frac{1}{\prod_j k_j!} \left(\frac{1-(c-\alpha)}{c-\alpha}\right)^K e^{-r\frac{1-(c+\alpha)}{c+\alpha}}
\end{aligned}$$

And, as this holds for all $\alpha > 0$, we find:

$$\lim_{\alpha \downarrow 0} \lim_{n \to \infty} \frac{1}{\prod_j k_j!} \left(\frac{1-(c+\alpha)}{c+\alpha} \right)^K e^{-r\frac{1-(c-\alpha)}{c-\alpha}} \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) \tag{2.24}$$

$$\leq \lim_{n \to \infty} \mathbb{P}_n(d(v_{i+j}) = k_j + 2, \;\; \forall j \leq r-1) \tag{2.25}$$

$$\leq \lim_{\alpha \downarrow 0} \lim_{n \to \infty} \frac{1}{\prod_j k_j!} \left(\frac{1-(c-\alpha)}{c-\alpha} \right)^K e^{-r\frac{1-(c+\alpha)}{c+\alpha}} \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) \tag{2.26}$$

Evaluating the limits gives:

$$\lim_{n \to \infty} \mathbb{P}_n(d(v_{i+j}) = k_j + 2, \;\; \forall j \leq r-1) = \frac{1}{\prod_j k_j!} \left(\frac{1-c}{c} \right)^K e^{-r\frac{1-c}{c}} \tag{2.27}$$

And, as $\frac{1-c}{c} = \rho$, with $e^{-\rho} = \rho$, we get:

$$\lim_{n \to \infty} \mathbb{P}_n(d(v_{i+j}) = k_j + 2, \;\; \forall j \leq r-1) = \frac{\rho^{K+r}}{\prod_j k_j!}$$

$\square$

As we have converging limiting distribution for the legs attached to a spine vertex in a random caterpillar, we can now prove:

Theorem 4 *Let $v_1, \ldots, v_\ell$ denote the spine vertices of an oriented random caterpillar C_n with spine length ℓ. Then, for fixed $k_1, \ldots, k_r \in \mathbb{N}$, we have:*

$$\lim_{n \to \infty} \mathbb{P}_n(\exists_i \;\; d(v_{i+1}) = k_1 + 2, \ldots, d(v_{i+r}) = k_r + 2) = 1$$

This is a crucial result, stating that every finite sequence of segments in a caterpillar will eventually occur in C_n.

Proof Take $R = g \cdot r$, $g \in \mathbb{N}$. Let $p > 0$ be the limit probability $p = \lim_{n\to\infty} \mathbb{P}_n(d(v_2) = k_1 + 2, \ldots, d(v_{r+1}) = k_r + 2)$. This is the same probability as $\mathbb{P}(d(v_{r(j-1)+2}) = k_1 + 2, \ldots, d(v_{rj+1}) = k_r + 2)$, for all $j \leq g$ by Lemma 4. So, the probability that in the first $R + 1$ vertices, there is no leg sequence k_i, is smaller than $(1-p)^g$. As the size of the caterpillar, and the spine length, goes to infinity, we can take $g \to \infty$ and see that:

$$\mathbb{P}(\exists_i \;\; d(v_{i+1}) = k_1 + 2, \ldots, d(v_{i+r}) = k_r + 2)) \geq \lim_{g \to \infty} 1 - (1-p)^g = 1$$

$\square$

Corollary 2 *Every fixed finite sequence of segments will occur arbitrarily often in C_n, when $n \to \infty$.*

Proof Every finite sequence of degrees $k_1, \ldots, k_r$ occurs with probability 1 when $n \to \infty$. Thus, any finite repetition of this sequence will also occur with probability 1, so every finite sequence will occur arbitrarily often in C_n.

$\square$

Chapter 3

Ehrenfeucht-Fraisse Games

3.1 General idea

In order to determine wether or not two graphs are the same, we can check their properties. Do they both contain a triangle? Do they both contain two disjoint vertices? These questions can be formalized, using logical operators AND, OR, NOT ($\wedge, \vee, \neg$), variables (usually denoted as $x, y, z, \ldots$) and quantifiers EXISTS and FORALL, ($\exists, \forall$). This is applied as $\forall x$ and $\exists y$. Further, we have two relations, $=, \sim$, where $=$ is equality ($x = y$ means that x is the same vertex as y) and $\sim$ is adjacency, $x \sim y$ means there's an edge between x and y. The existence of a triangle would then become the sentence:

$$\Delta = \exists x \exists y \exists z \quad x \sim y \wedge x \sim z \wedge y \sim z$$

That is, there exist three different vertices, that have edges between them. For any graph G containing a triangle, we would write $G \models \Delta$, and if G doesn't contain a triangle, we would write $G \models \neg\Delta$. Since φ is either true or not true, exactly one of these must hold.

Questions like this, containing quantifiers, variables and logical operators are called *formulas*. When all variables in φ are quantified, φ is called a *sentence*. A sentence, like the previous one with the triangle, is always true or false for any graph G. An example of a formula which isn't a sentence would be:

$$\psi = \forall_x (x \neq y) \implies x \sim y$$

This can be true for a graph G, depending on the choice of y. We'll usually work with sentences, denoted with φ. The set of all finite sentences that can be constructed with said operators and quantifiers, is called First-Order logic. For every sentence, the *quantifier depth* of a sentence is the depth of the nesting of the quantifiers in φ, denoted $qd(\varphi)$, and is defined by the following axioms:

- $qd(\neg\varphi) = qd(\varphi)$
- $qd(\varphi \vee \psi) = qd(\varphi \wedge \psi) = qd(\varphi \implies \psi) = \max(qd(\varphi), qd(\psi))$
- $qd(\exists_x \varphi) = qd(\forall_x \varphi) = 1 + qd(\varphi)$
- $qd(x = y) = qd(x \sim y) = 0$

3.1.1 Equivalence

When $\mathcal{L}$ is the set of all logical FO-sentences φ, we can define $\mathcal{L}^k \subset \mathcal{L}$ as the subset of sentences which quantifier depth at most k. Two graphs G, H are said to be k-equivalent, $G \equiv_k H$, when $G \models \varphi$ iff $H \models \varphi$, $\forall \varphi \in \mathcal{L}^k$. That is, $G \equiv_k H$ when G is true for expression φ, which is of quantifier depth at most k, only if it's also true for H and vice versa.

A way to check if two graphs are the same, is playing the (First Order) *Ehrenfeucht-Fraisse* game [1]. In this game, there are two players, Spoiler and Duplicator. They alternate making moves, Spoiler goes first, then Duplicator, then Spoiler, etc.. A move consists of Spoiler picking a vertex in either graph G or H, and Duplicator picking a vertex in the other graph. This way, after k moves, there are induced graphs $G_k = ([k], E_k)$ and $H_k = ([k], E'_k))$, where vertices can be chosen multiple times. Both graphs consist of the labeled vertices (labels $1, \dots, k$) and there are edges between two vertices if and only if there's and edge between the two corresponding vertices in the original graphs G, H. Duplicator has as objective to make sure that $G_k = H_k$ and Spoiler has as objective to pick it's vertices x_i in such a way that Duplicator can't pick it's y_i s.t. $G_I = H_I$. If Duplicator succeeds, thus $G_i = H_i$, for all $i \leq k$, irrespective of Spoilers moves, Duplicator is said to have a winning strategy.

Theorem 5 *[1] Fix $k > 0$ and let G, H be graphs. Then the following are equivalent:*

- *$G \equiv_k H$*
- *Duplicator has a winning strategy in the k-move Ehrenfeucht-Fraisse game.*

Example 3

Let $G = (V, E)$ be a path of length 3, so $V = \{x, y, z\}$ and $E = \{(x, y), (y, z)\}$ and let H be a path of length 4. Then, Spoiler has a winning strategy in a 2-move game. This strategy is to pick $y \in G$, the middle vertex in move 1, labeling y as 1. Then, Duplicator has to pick a vertex $x_1 \in H$ in his first move. In the second move, Spoiler chooses vertex $x_2 \in H$ s.t. $x_2 \not\sim x_1$. Such a vertex always exists, independent of the choice of x_1. Now, Duplicator has to pick a vertex $y_2 \in G$ s.t. $y_2 \not\sim y$, but this vertex doesn't exist. So Spoiler always wins the game, and $G \not\equiv_2 H$. See Figure 3.1 for a visualization of how the game could play out. ■

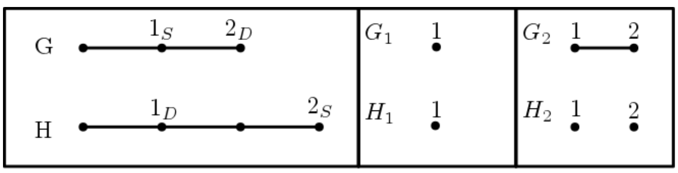

FIGURE 3.1: How Spoiler wins in 2 moves

Lemma 5 *When $F \equiv_k G$ and $G \equiv_k H$, then $F \equiv_k H$.*

Proof Suppose that Spoiler and Duplicator play a game on F, H. Duplicator also keeps a copy of G next to it. When Spoiler plays the j-th move on F, Duplicator first applies his winning strategy on the F vs. G game, for his pick of G. Then, his move on H would be following his winning strategy if it were a G vs H game and Spoiler played a move on G. Clearly, the same can be done when Spoiler plays a move on H. Since Duplicator has a winning strategy in both games, he can always proceed this way, thus he has a winning strategy comparing F and H. $\square$

3.1.2 Zero-One (convergence) laws

Let $\mathcal{A}$ be a class of graphs and let $\mathcal{A}_n \subset \mathcal{A}$ be the graphs of size n. If $A_n \in_n \mathcal{A}_n$ is a graph, chosen randomly via some probability measure on $\mathcal{A}_n$, and φ is a sentence, we have $\mathbb{P}(A_n \models \varphi)$ as the probability that a randomly chosen graph in $\mathcal{A}_n$ satisfies φ. We say that a class $\mathcal{A}$ obeys the *Zero-One law* on a logical language $\mathcal{L}$, if:

$$\forall \varphi \in \mathcal{L} \quad \lim_{n\to\infty} \mathbb{P}(A_n \models \varphi) \in \{0,1\}$$

This means that the probability that a sentence is true for a randomly picked graph, is either always 1 or always 0 when the size of the graph goes to infinity.

In terms of the Ehrenfreucht-Fraisse Game, we can see this in the following way. Suppose that $\mathcal{A}$ obeys a Zero-One law. Thus, for two random graphs $G \in \mathcal{A}_n$, $H \in \mathcal{A}_m$, for any finite sentence φ, the probability that $G, H \models \varphi$ or $G, H\neg \models \varphi$ tends to 1 as $n, m \to \infty$. In particular, this holds for all $\varphi \in \mathcal{L}^k$, thus $G \equiv_k H$ with probability tending to 1 as n, m go infinity, for all $k \in \mathbb{N}$. This implies that Duplicator has a winning strategy for all k-move games, for all k, as n, m go to infinity.

Example 4

The Erdos-Renyi random graph $G(n,p)$ is a graph on the vertices $[n]$ in which, for every distinct duo $x, y \in V$, $(x,y) \in E$ with probability p. Together, these random graphs form the class $\mathcal{G}$ with sub-classes $\mathcal{G}_n$. We can choose a graph $G_n \in \mathcal{G}_n$ randomly, where a graph $G = ([n], E)$ has probability $\mathbb{P}(G_n = G) = p^{|E|}(1-p)^{\binom{n}{2}-|E|}$ to be chosen.

For $p = n^{-\alpha}$, $\alpha \in (0,1]$, due to Shelah and Spencer [7] we know that $\mathcal{G}$ satisfies a Zero-One law if and only if $\alpha \notin \mathbb{Q}$. When $p = \frac{1}{2}$ is fixed, there's the so-called *Rado Graph* G, for which:

$$G \models \varphi \iff \lim_{n\to\infty} \mathbb{P}(G(n,p) \models \varphi) = 1$$

The class corresponding to this $p = \frac{1}{2}$, $\mathcal{G}$, also obeys a Zero-One law [2]. We'll prove this last statement.

We'll first prove that two graphs having the *extension property* are equal, and then that the random graph $\lim_{n\to\infty} G(n,p)$ has indeed the extension property. A graph G is said to have the extension property P when:

$$\forall_{n,m\in\mathbb{N}} \forall_{u_1,\ldots,u_n,v_1,\ldots,v_m \in G} \exists_{z\in G} (z \neq u_i, v_j) \wedge (u_i \sim z) \wedge (v_j \not\sim z) \tag{3.1}$$

That is, for any two finite subsets of vertices $U, V \in G$, there exists a vertex $z \in G$ such that z is adjacent to all vertices in U and not to any vertex in V.

Lemma 6 *If graphs G, H both have property P, then $G \equiv H$.*

Proof Let G, H be graphs having property P and enumerate the vertices in G, $x_1, \ldots$ and the vertices in H, $y_1, \ldots$. We'll play the E-F game, to show equivalence between G and H. Suppose k moves are already played and $G_k \cong H_k$. Spoiler moves next, playing $w \in G$. When $w \in G_k$, w is played before, as j-th move, Duplicator can play the opposite j-th move and $G_{k+1} \cong H_{k+1}$. Otherwise, there's a partition U, V of the vertices in G_k, such that:

$$w \sim u \quad \forall u \in U \qquad\qquad w \not\sim v \quad \forall v \in V$$

As $G_k \cong H_k$, there's a partition U_h, V_h of H_k as well, s.t. $U_h \cong U$ and $V_h \cong V$. By property P, there also exists a $z \in H$ s.t.

$$z \sim u \quad \forall u \in U_h \qquad\qquad z \not\sim v \quad \forall v \in V_h \tag{3.2}$$

As $k+1$-th move, Duplicator will play $z = y_j \in H$ satisfying (3.2), where $j \in \mathbb{N}$ is the lowest number that has not been played. Clearly, by construction, $G_{k+1} \cong H_{k+1}$. As, after zero moves, $G_0 \cong H_0$, this is a winning strategy for Duplicator. □

Next, we'll prove that any countable random graph (with $p = \frac{1}{2}$) has this property.

Lemma 7 *When G is a graph with countably many vertices and $\mathbb{P}(u \sim v) = \frac{1}{2}$ for all $u, v \in G$ independently, G has property P.*

Proof Let G be such a random graph. Then, for two disjoint finite sets U, V, the probability that a vertex $z \in G \setminus U, V$ is adjacent to all vertices in U is $2^{-|U|}$ and that it's not adjacent to any vertex in $|V|$ is $2^{-|V|}$. Thus, the probability that z isn't the vertex needed, is $1 - 2^{-|U|+|V|}$. However, by independence of all edges, the probability that none of the vertices in $G \setminus U \cup V$ satisfy, is equal to $(1 - 2^{-|U|+|V|})^{|G|-|U|-|V|}$. As G is of countable size, this probability is zero.

There are only countably many different selections of U, V, and the union of countably many sets of measure zero is still of measure zero. So for all finite disjoint sets U, V, there's a $z \in G$ adjacent to all vertices in U and to none in V. Thus, G has property P. □

As we've seen that any countable random graph G has the extension property, and all graphs with the extension property are equivalent to each other, we can conclude that there's a Zero-One law on the class $\mathcal{G}$. ■

Example 5

As we'll see, the class of caterpillars doesn't obey a Zero-One law with respect to First Order logic. However, the class of *forests of caterpillars*, $\mathcal{F}$, where each graph consists of disjoint caterpillars, does obey a Zero-One law. [4] ■

Let $\mathcal{A}$ again be a class of graphs and suppose there is a sentence $\varphi \in \mathcal{L}$ for which, with $A_n \in \mathcal{A}_n$ uniform randomly chosen, $\lim_{n\to\infty} \mathbb{P}(A_n \models \varphi) \notin \{0,1\}$, but the limit does exist. Then, obviously, there is no Zero-One law. However, $\mathcal{A}$ can still satisfy a *convergence law* with respect to $\mathcal{L}$, that is to say that:

$$\forall \varphi \in \mathcal{L} \quad \lim_{n\to\infty} \mathbb{P}(A_n \models \varphi) \text{ exists.}$$

So $\mathcal{A}$ obeys a convergence law when the limit with n going to infinity of the probability that $\varphi \in \mathcal{L}$ holds for a random graph $A_n \in \mathcal{A}_n$ exists for all φ.

3.2 E-F games on paths

Let $\mathcal{P}$ denote the class of labeled paths, where a path is a connected graph with vertices of degree at most 2. There's obviously only one unlabeled path of length n, so we have $|\mathcal{P}_n| = \frac{n!}{2}$ for $n \geq 2$, as the only difference between paths of length n is their labeling. For the random path $P_n \in_n \mathcal{P}_n$, this will also be just a path of length n with probability 1. The class of paths obeys a Zero-One Law. In order to prove this, we'll need to prove a few lemma's, using the Ehrenfeucht-Fraisse game extensively.

We'll first introduce some notation. In general, x represents a vertex on G and y a vertex on H. In a k-move game on G, H, x_i, y_i are the vertices played (labeled) in move i, with $x_i \in G$ and $y_i \in H$. We'll denote a vertex adjacent to x as $x-1$ or $x+1$ and a vertex a away from vertex x in a path as $x-a$ or $x+a$. This way, $[x-a, x+a]$ will denote the path created taking all vertices within a distance a of x, preserving any labels given to any vertices. When x is within distance a of the end of the path, that is, $x-a$ or $x+a$ doesn't exist, $[x-a, x+a]$ will consist of less vertices than the expected $2a+1$. The only way to recognize an outer vertex x_0 in $[x-a, x+a]$ as an outer vertex, is when $x_0 \neq x \pm a$, since that will leave blanks in the interval.

When comparing the two intervals $[x_l, x_r] \subset G$, $[y_l, y_r] \subset H$ between $x_l, x_r \in G$, $y_l, y_r \in H$, we write $[x_l, x_r] = [y_l, y_r]$ if all of the following is true:

- $|[x_l, x_r]| = |[y_l, y_r]|$
- For all labels $x_i \in [x_l, x_r]$, there's a corresponding label $y_i \in [y_l, y_r]$ and $\{|[x_l, x_i]|, |[x_i, x_r]|\} = \{|[y_l, y_i]|, |[y_i, y_r]|\}$.
- For all labels $x_i, x_j \in [x_l, x_r]$, there are corresponding labels $y_i, y_j \in [y_l, y_r]$ and $|[x_i, x_j]| = |[y_i, y_j]|$.

When we want to compare $[x-a, x+a]$ on G with $[y-a, y+a]$ on H, with $a \in \mathbb{N}$ and $x \in G$, $y \in H$, it's possible that any of these intervals is not of size $2a+1$, since they can be around the outer end of the paths G, H. Let x_l, x_r and y_l, y_r be the outer most vertices in $[x-a, x+a]$, $[y-a, y+a]$ respectively. Then, $[x-a, x+a] = [y-a, y+a]$ if $[x_l, x_r] = [y_l, y_r]$ and $\{|[x_l, x]|, |[x, x_r]|\} = \{|[y_l, y]|, |[y, y_r]|\}$.

On intervals of different sizes, we can prove the following useful lemma:

Lemma 8 *Suppose the E-F game is played on paths G, H and $x_1, x_2 \in G$ $y_1, y_2 \in H$ are already played with j moves left and $|[x_1, x_2]| < |[y_1, y_2]|$. If $|[x_1, x_2]| \leq 2^j + 1$, Spoiler has a winning strategy.*

Proof We prove this by induction on j. If $j = 0$, $|[x_1, x_2]| \in \{1, 2\}$ and either $x_1 = x_2$, while $y_1 \neq y_2$ or $x_1 \sim x_2$ while $y_1 \not\sim y_2$, thus Spoiler has won.

Suppose it's true for all $j' < j$. Spoiler plays x_3 s.t. $|[x_1, x_3]|, |[x_2, x_3]| \leq 2^{j-1} + 1$. In every move y_3 Duplicator plays, at least one of $|[y_1, y_3]| > |[x_1, x_3]|$ and $|[x_2, x_3]| > |[y_2, y_3]|$ has to hold. Then, by induction hypothesis, Spoiler has a winning strategy. □

An example of this, with i_S, i_D representing the i-th move made by Spoiler and Duplicator respectively, can be seen in Figure 3.2. Duplicator's third move is absent, since he'll lose the game whatever he plays.

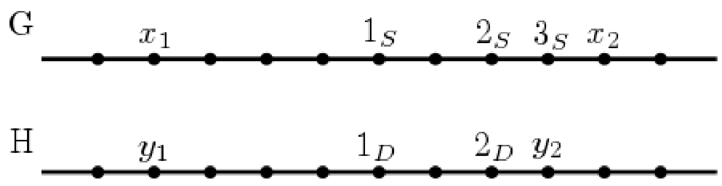

FIGURE 3.2: Spoiler wins 3 moves after x_1, x_2, y_1, y_2 are played

Lemma 9 *Suppose the E-F game is played on paths G, H and $j \geq 1$ moves have already been played. When $|G|, |H| \geq (j+1)(2^{k-j}+1)$ and:*

$$\forall_{i \leq j}[x_i - 2^{k-j}, x_i + 2^{k-j}] = [y_i - 2^{k-j}, y_i + 2^{k-j}]$$

Duplicator has a winning strategy. In particular this implies that when $|G|, |H| \geq 2^k+2$, if Duplicator can match the first move by Spoiler, he'll win the k-move game.

Proof. We'll prove it by backward induction on j. Suppose the conditions hold after $j = k$. Then, for $m, n \leq k$, $x_n \sim x_m$ iff $x_n \in [x_m - 1, x_m + 1]$ and since, by assumption, $[y_m - 1, y_m + 1] = [x_m - 1, x_m + 1]$, we get $y_n \sim y_m$. Thus, for the induced subgraphs G_k, H_k we have $G_k \cong H_k$, and we have an induction basis.

Suppose the Lemma is true for $k \geq j' > j$ and suppose j moves have been played. Let $x_{j+1} \in G$ be the $(j+1)$-th move played by Spoiler (the assumption that Spoiler plays in G is wlog) and let $I = \{i : x_i \in [x_{j+1} - 2^{k-j-1}, x_{j+1} + 2^{k-j-1}]\}$ be the indices of the moves already played near x_{j+1}.

If $I = \emptyset$, then, Duplicator needs to play $y_{j+1} = y$ such that $\forall i \leq j$, $y_i \notin [y - 2^{k-j-1}, y + 2^{k-j-1}]$. The previously played j moves y_i divide the path into $j+1$-parts, and by the hypothesis $|H| \geq (j+1)(2^{k-j}+1)$, there has to be a path γ of length $|\gamma| = 2^{k-j}+1$ in H containing no y_i and in this path, and for the center point $y \in \gamma$ it holds that $[y - 2^{k-j-1}, y + 2^{k-j-1}] = \gamma$. Then, Duplicator plays this $y_{j+1} = y$.

Suppose that $I \neq \emptyset$. Then:

$$[x_{j+1} - 2^{k-j-1}, x_{j+1} + 2^{k-j-1}] \subset [x_i - 2^{k-j}, x_i + 2^{k-j}] \;\; \forall i \in I$$

And since $[x_i - 2^{k-j}, x_i + 2^{k-j}] = [y_i - 2^{k-j}, y_i + 2^{k-j}]\ \forall i$, there's a point $y \in H$ s.t.:

$$[x_{j+1} - 2^{k-j-1}, x_{j+1} + 2^{k-j-1}] = [y - 2^{k-j-1}, y + 2^{k-j-1}]$$

And this point y will be the move played as y_{j+1}.

Then, by choice of y_{j+1}, for all $i \in I$ we have that:

$$[x_i - 2^{k-j-1}, x_i + 2^{k-j-1}] = [y_i - 2^{k-j-1}, y_i + 2^{k-j-1}]$$

And this will obviously hold for all $m \notin I$. Further, we chose y_{j+1} s.t. it matches x_{j+1}. Notice that for the length $|G|, |H|$, we see:

$$\begin{aligned}(j+1)\left(2^{k-j}+1\right) - ((j+1)+1)\left(2^{k-(j+1)}-1\right) &= (j+1)2^{k-j-1} - \left(2^{k-j-1}-1\right) \\ &= j2^{k-j-1} - 1 \geq 0\end{aligned}$$

Where the inequality is a result from taking $j \geq 1$. Thus, if $|G|, |H|$ matches the length criterion after j moves, they will satisfy the criterion also after $j+1$ moves. Thus, we can apply the induction hypothesis that for $j+1$ moves, the Lemma is true. Therefore, Duplicator has a winning strategy.

When $|G|, |H| \geq 2^k + 2$, we can apply this lemma. For every first move x_1 played by Spoiler, Duplicator can play y_1 s.t. $[x_1 - 2^{k-1}, x_1 + 2^{k-1}] = [y_1 - 2^{k-1}, y_1 + 2^{k-1}]$ and the lemma can be applied to guarantee a win by Duplicator. So $G \equiv_k H$ when $|G|, |H| \geq 2^k + 2$.

□

This is already enough to prove the Zero-One Law on paths. However, a further specification in equivalence classes of paths is given in the following lemma:

Lemma 10 *When G and H are paths then*

- $k = 1$, $G \equiv_k H$ *iff* $|G| = |H|$ *or* $|G|, |H| \geq 1$.
- $k = 2$, $G \equiv_k H$ *iff* $|G| = |H|$ *or* $|G|, |H| \geq 4$
- $k = 3$, $G \equiv_k H$ *iff* $|G| = |H|$ *or* $|G|, |H| \geq 7$
- $\forall k \geq 4$ *it holds that* $G \equiv_k H$ *iff* $|G| = |H|$ *or* $|G|, |H| \geq 2^k$.

Proof

Notice that when $|G| = |H|$, Duplicator has a winning strategy; he can just copy every move made by Spoiler. In the rest of the proof, we'll assume that $|G| \neq |H|$.

When $k = 1$, Spoiler picks a vertex, wlog in G. This is possible as G, H can't be both of order 0. Then, if $|H| \geq 1$, Duplicator can respond by picking a vertex in H and win the game. So if both $|G|, |H| \geq 1$, $G \equiv_1 H$.

When $k = 2$, If $|G| = 2$, $|H| \geq 3$, Spoiler plays both outer vertices of H, these are non-adjacent, this isn't possible on G. By Example 3, Spoiler wins when $|G| = 3$, $|H| = 4$. When $|G|, |H| \geq 4$, Duplicator wins as, in the second move, he can always play a vertex adjacent to the earlier played vertex when Spoiler did this, as well as non-adjacent, since no vertex in either G or H is adjacent to all others.

When $k = 3$, Spoiler wins when $|G| = 6, |H| = 7$, Spoiler wins as G contains two vertices x, y such that all other vertices are adjacent to either x or y, while H doesn't contain this vertex. Spoiler picks $x, y \in G$ as first two moves and then plays a vertex non-adjacent to the ones played by Duplicator.

Suppose $|G|, |H| \geq 7$. When Spoiler plays $x_1 \in G$, splitting G in G_1, G_2 with $G_1 \cap G_2$, where we assume wlog that $|G_1| \leq |G_2|$, Duplicator plays $y_1 \in H$, such that $|H_1| = \min |G_1|, 4$. Spoiler plays a second move, for instance x_2, and Duplicator plays y_2 and makes sure that the distances $|[x_1, x_2]| = |[y_1, y_2]|$, if $|[x_1, x_2]| \leq 3$, and $|[y_1, y_2]| > 3$, when $|[x_1, x_2]| > 3$. This is always possible, and also works when Spoiler played y_2 instead of x_2. In the final move, Duplicator can always play a vertex such that $x_3 \sim x_i$ if and only if $y_3 \sim y_i$, and he won the game.

That Spoiler wins when $|G|, |H| \leq 6$ isn't shown here, but requires little work.

Suppose $k \geq 4$. If Spoiler wants to win a k-move game, he either has to play a vertex not-adjacent to earlier vertices, while Duplicator can't copy this, or he plays a vertex adjacent to earlier played vertex/vertices, and Duplicator can't do the same. However, in a path of size more than $3(k-1)$, after $k-1$ moves, there's always a vertex not adjacent to earlier played vertices. So, in a k-move game, if $|G|, |H| > 3(k-1)$, Spoiler can't beat Duplicator by only playing non-adjacent vertices, since Duplicator can always copy this. And, as $2^k > 3(k-1)\ \forall k \geq 4$, to prove the Lemma we have to find a strategy in which Spoiler plays a vertex adjacent to 1 or 2 earlier played vertices and Duplicator can't copy this.
For the rest of the proof of the theorem, we'll separate cases. We already have that $|G|, |H| \geq 2^k + 2$ results in k-equivalence.

1. $|G| = 2^k + 1, |H| = 2^k + 2$, **D wins**
2. $|G| = 2^k - 1, |H| \leq 2^k - 2$, **S wins**
3. $|G| = 2^k - 1, |H| = 2^k$, **S wins**
4. $|G| = 2^k, |H| = 2^k + 1$, **D wins**

Case 1 Notice that at the start, that is, after 0 moves, for every $x \in G$, there is a y in H s.t. $[x - 2^{k-1}, x + 2^{k-1}] = [y - 2^{k-1}, y + 2^{k-1}]$ and vice versa. Thus, Duplicator can always respond playing a matching $y_1 \in H$ (or $x_1 \in G$).

Next, every move made by Spoiler can be copied by Duplicator, unless x_1 was the center vertex of G and Spoiler plays $y_2 \in H$ s.t. $[y_2 - 2^{k-2}, y_2 + 2^{k-2}]$ is a path, not containing y_1 and not being cut-off by some outer vertex, since there's no such path of length $2^{k-1}+1$ in G left. In order to win, Spoiler must create this situation, otherwise the conditions of Lemma 9 are satisfied after 2 moves and Duplicator would have a winning strategy. Thus, as Spoiler plays perfect, he has to play as described in the first two moves. The response by Duplicator will be to pick $x_2 \in G$, such that $[x_2 - 2^{k-2}, x_2 + 2^{k-2}]$ will be a path of length 2^{k-1}, not containing x_1 (in fact, x_1 has to be adjacent to one of the outer vertices of this path) and containing an outer vertex.

For the following moves, Spoiler has to play $x_j \in G$ s.t. $[x_j - 2^{k-j}, x_j + 2^{k-j}]$ contains x_{j-1} and an outer vertex (there's only one possible vertex to pick every time) and Duplicator can respond to this by playing y_j s.t.

$$y_j \in [y_{j-1} - 2^{k-j-1}, y_{j-1} + 2^{k-j-1}] = [x_{j-1} - 2^{k-j-1}, y_{j-1} + 2^{k-j-1}]$$

If Spoiler wouldn't play this way, Duplicator can play y_j s.t. the conditions of Lemma 9 are satisfied and he has a winning strategy.

Playing this way for $k-1$ moves, Spoiler still can't win the game. In the last round, x_{k-1} will be a vertex adjacent to an outer vertex. Thus, x_{k-1} has a neighboring vertex x s.t. $x_{k-1} \sim x \nsim x_{k-2}$ and another vertex $x_{k-2} \sim x' \sim x_{k-1}$. However, there are corresponding y and y' that satisfy these properties on H with respect to y_{k-1}, y_{k-2}. Since $[y_m - 1, y_m + 1] = [x_m - 1, x_m + 1]$ $\forall m \leq k-2$, every move made by Spoiler can be copied by Duplicator, thus **Duplicator has a winning strategy.**

Case 2 Spoiler starts with $x_1 \in G$, splitting G in two paths G_1, G_2 both of length 2^{k-1}. Duplicator has to play $y_1 \in H$, splitting H in two paths H_1, H_2, where one of the two (wlog we pick H_1) has to have a length $|H_1| < 2^{k-1}$. To simplify notation, we introduce y_0, which is the outer vertex of H_1 and $H_1 = [y_0, y_1]$ and similarly, we introduce $x_0 \in G$, also the outer vertex of G_1.

Spoiler continues from here and plays $y_2 \in H_1$ such that $|[y_1, y_2]| = 2^r + 1$, with r the largest possible integer. For this r, we know that $r \leq k-2$, as $|H_1| < 2^{k-1}$. Notice that $|[y_0, y_2]| \leq |[y_2, y_1]|$. Duplicator has to play $x_2 \in G_1$ (wlog) and he plays s.t. either $|[x_1, x_2]| \neq |[y_1, y_2]|$ or $|[x_0, x_2]| \neq |[y_0, y_2]|$.

If the former is the case, by the first Lemma, Spoiler has a winning strategy. If only the latter is the case, thus, $|[x_0, x_2]| \neq |[y_0, y_2]|$, then $|[x_0, x_2]| = 2^{k-2}$ while $|[y_0, y_2]| \leq 2^{k-2} - 1$. As third move, Spoiler can play y_3 s.t. $|[y_3, y_2]| = 2^{r'} + 1$, with $r' < r$, and again, whatever Duplicator responds, there has to be a path of non matching length, just like after move 2. If $|[y_3, y_2]| \neq |[x_3, x_2]|$, by the same lemma as before, Spoiler has a winning strategy. So, in perfect play, Duplicator has to keep matching $|[x_j, x_{j-1}]| = |[y_j, y_{j-1}]|$ in the next moves. After at most $j = k-1$ moves, however, $|[x_0, x_{k-1}]| = 2$, while $[y_0, y_{k-1}] \leq 2$, thus y_{k-1} is an outer vertex of the path H. This means that if $y \sim y_{k-1}$, $y \sim y_{k-2}$ as well so if Spoiler plays $x_k = x_0$, Duplicator can't respond to this. Therefore, **Spoiler has a winning strategy**.

NB: The same can be done when both $|G|, |H| < 2^k - 2$, since Spoiler can always play such that there's a two corresponding parts of different length, both smaller than $2^{k-j} + 1$ after j moves and win the game.

Case 3 We label the outer vertices of G, H preliminary x_0, x_∞ and y_0, y_∞. These labels are not part of the game.

In the first move, Spoiler plays $x_1 \in G$ the middle vertex of G, s.t. $|[x_0, x_1]| = |[x_1, x_\infty]| = 2^{k-1}$. Next, Duplicator plays $y_1 \in H$, s.t. $|[y_0, y_1]| \neq 2^{k-1}$ or $|[y_1, y_\infty]| \neq 2^{k-1}$. If any of the two paths has a length less than 2^{k-1}, by analysis done in Case 2, Spoiler has a winning strategy. So suppose that $|[y_0, y_1]| = 2^{k-1} + 1$ and $|[y_1, y_\infty]| = 2^{k-1}$.

In the second move, Spoiler plays y_2 s.t. $|[y_1, y_2]| = 2^{k-2}+2$ and $|[y_2, y_0]| = 2^{k-2}$. If Duplicator plays x_2 s.t. $|[x_1, x_2]| \leq 2^{k-2} + 1$, by the first Lemma, Spoiler has a winning strategy. Thus, Duplicator has to play x_2 s.t. $|[x_1, x_2]| \geq 2^{k-2} + 2$, thus leaving $|[x_2, x_0]| \leq 2^{k-2} - 1$. From here, Spoiler continues playing x_j s.t. $|[x_j, x_{j-1}]| = 2^{k-j} + 1$ and $|[x_0, x_j]| \leq 2^{k-j} - 1$ and Duplicator has to follow with $|[y_j, y_{j-1}]| = 2^{k-j} + 1$ and $|[y_j, y_0]| \geq 2^{k-j}$, otherwise he'll lose. After $j = k-1$ moves, there's no vertex $x \sim x_{k-1} = x_0$ not adjacent to x_{k-2}, and there is a $y \sim y_{k-1}$ not adjacent to y_{k-2}. To win, Spoiler plays this y in it's k-th move. So **Spoiler has a winning strategy.**

Case 4 We label the outer vertices of G, H preliminary x_0, x_∞ and y_0, y_∞. These labels are not part of the game.

Spoiler plays first, either x_1 or y_1 and Duplicator responds, s.t. he matches the smallest value of $|[x_0, x_1]|, |[x_1, x_\infty]|$ with it's equivalent on H with his move y_1 (or vice versa if Spoiler plays on H in his first move). This way, after this move, there are two matching parts $[x_1, x_\infty]$,$[y_1, y_\infty]$, two odd paths s.t. $|[x_0, x_1]| \geq 2^{k-1}$, $|[y_0, y_1]| \geq 2^{k-1} + 1$. If $|[y_0, y_1]| > 2^{k-1} + 1$, $|[x_0, x_1]| \geq 2^{k-1}+1$ and by analysis done in Case 1, Duplicator can win, thus $|[y_0, y_1]| = 2^{k-1} + 1$, which means that $|[y_0, y_1]| = |[y_1, y_\infty]| = |[x_1, x_\infty]|$. So, in his second move, Spoiler has to play a vertex $x_2 \in [x_0, x_1]$. If $|[x_1, x_2]| \leq 2^{k-2} + 1$, Duplicator plays y_2 s.t. $|[y_1, y_2]| = |[x_1, x_2]|$, otherwise he matches $|[x_0, x_2]| = |[y_0, y_2]|$ and wins by the previous lemma. The next moves go similar and after $j = k-1$ moves, $|[x_0, x_{k-1}]| = 2$ and $|[y_0, y_{k-1}]| = 3$, which is won by Duplicator. So, **Duplicator has a winning strategy.** □

From here, the Zero-One law on $\mathcal{P}$ follows immediately:

Theorem 6 *The class $\mathcal{P}$ of paths obeys a Zero-One law.*

Proof By Lemma 10 we know that $P_n \equiv_k P_m$ when $m, n \geq 2^k$. Thus:

$$\forall k \in \mathbb{N} \ \forall \varphi \in \mathcal{L}^k \quad P_n \models \varphi \iff P_m \models \varphi \quad \text{if } m, n \geq 2^k$$

And:

$$\forall \varphi \in \mathcal{L} \ \lim_{n \to \infty} \mathbb{P}(P_n \models \varphi) = \mathbb{P}\left(P_{2^{qd(\varphi)}} \models \varphi\right) \in \{0, 1\} \tag{3.3}$$

□

Chapter 4

Convergence law on Caterpillars

4.1 Basic games

Example 6

We can apply the Ehrenfeucht-Fraisse game on caterpillars, to see if they are in some sense the same. We could ask the question, whether or not a caterpillar G has an outervertex of degree 2, i.e. it only has 1 leg. The logical question would be:

$$\varphi = \exists_{xyz}((y \neq z) \wedge x \sim y \wedge x \sim z) \wedge (\exists_w((w \neq x) \wedge y \sim w) \tag{4.1}$$
$$\wedge (\forall_w(z \sim w) \implies w = x) \wedge (\forall_w(w \sim x) \implies (w = y \vee w = z))) \tag{4.2}$$

This can be read in the following way. There exists x, connected to 2 distinct vertices y, z. There is another vertex w connected to y and there is no other vertex connected to either x or z. This means that x is of degree $d(x) = 2$ and it has only one neighbor which is also of degree at least 2, namely y. The only vertices in a caterpillar of degree at least 2 with only one neighbor of degree at least 2 are the outer vertices. Thus, x is an outer vertex and since it's of degree 2, it has only 1 leg.

A strategy for Spoiler to defeat Duplicator if we have two caterpillars, G and H, where G has outer vertex v_1 of degree 2 and H has higher degrees on it's outer vertices (so both outer vertices are of degree at least 3), would be the following. In it's first move, Spoiler picks $v_1 \in G$. Then, Duplicator has multiple options:

- if Duplicator decides not to pick a vertex on the spine (so he picks a leg) in H, Spoiler wins in two moves, by picking two vertices connected to v_1 in G.
- if Duplicator doesn't pick an outer vertex in H, but just a vertex w on the spine of H for which $d(w) = 2$, Spoiler wins in 2 more moves. He picks the vertex $y \in G$, which is a leg of v_1. Now, Duplicator has to choose a point z adjacent to w in H, but this point has to be in the spine. Then, Spoiler picks another point, not w, in H adjacent to z and Duplicator can't counter this move.
- if Duplicator picks a vertex w on the spine of H of degree $d(w) \geq 3$, Spoiler wins in three moves by choosing three neighbors of w in H. Since x has only two neighbors, Duplicator can't win this game.

- if Duplicator picks an outer vertex, say $w \in G$, Spoiler can win in 3 moves. He first picks the two neighbors of v_1 in G, then Duplicator has to choose two neighbors of w in H. Then, Spoiler chooses another neighbor of w in H, which exists by assumption. Duplicator can't respond to this.

All four situations are drawn in Figure 4.1.

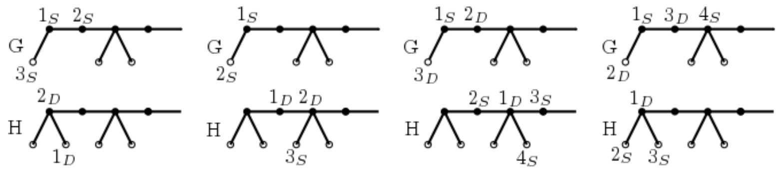

FIGURE 4.1: Spoiler needs at most 4 moves

So, Spoiler needs at most 4 moves to prove that the two caterpillars aren't the same. Obviously, this can be extended to show the difference between any two different sets of outer vertices in a certain number of moves.

In the example, a winning strategy for Spoiler is described, when comparing two outer vertices of caterpillars with degree 2 and 3 (or more). This can be generalized, to comparing regular vertices of different degree.

Proposition 1 *Let G, H be caterpillars on which Spoiler and Duplicator play the Ehrenfeucht-Fraisse game. If Spoiler and Duplicator picked vertices $x \in G$, $y \in H$ in round 1, then:*

- *If Spoiler and Duplicator both picked a spine vertex, with degrees d_1, d_2, for which $d_1 > d_2$, it takes at most $d_2 - 1$ additional steps for Spoiler to win the game.*

- *If Spoiler and Duplicator both picked legs connected to vertices of degree d_1, d_2, for which $d_1 > d_2$, it takes Spoiler at most d_2 additional steps to win the game.*

- *If Spoiler picked a spine vertex and Duplicator a leg, or vice versa, it takes Spoiler at most 2 additional steps to win the game.*

Proof For the first statement, Spoiler chooses to pick a leg of x, where Duplicator has to respond by picking legs of y, otherwise, Spoiler picks a vertex in H adjacent to the one chosen by Duplicator but not y and wins the game. After $d_2 - 2$ moves - a spine vertex of degree d_2 has $d_2 - 2$ legs if it's not an outer vertex - all legs of y have been chosen and some legs of x haven't been chosen yet. Spoiler picks one of these. Duplicator can only respond by picking a spine vertex adjacent to y. Now, Spoiler chooses a point adjacent to the last choice by Duplicator (not y) and wins the game.

For the second statement, assume x is a leg of X and y of Y. Spoiler again chooses to pick legs of X and Duplicator again has to respond by choosing legs of Y. Again, when all legs of Y are chosen, Spoiler picks a leg of X that isn't chosen yet. When Duplicator doesn't respond by a spine vertex z adjacent to Y, Spoiler wins by picking X. If Duplicator does respond in this way, Spoiler picks a vertex u adjacent to Duplicators choice

z, but not Y - as z is another spine vertex, z is of degree at least 2 and there will exist such a vertex u. Duplicator can't respond, as he has to play a vertex adjacent to a leg of X, and it can't be the spine vertex X. Notice that Spoiler doesn't have to pick X itself, the idea of doing so is enough.

For the third statement, Spoiler picks two vertices adjacent to the spine vertex and wins the game. □

Spoiler doesn't have a strategy that guarantees him a faster win. Suppose a game is played on G, H, where G and H are copies, except one spine vertex z has an extra leg in H, compared to its counterpart w in G. When $x = w, y = z$ are played in the first moves, the only way for Spoiler to win, is to show that z has indeed an extra leg; anything otherwise can be copied by Duplicator, as both graphs are completely the same. To show z is of higher degree, Spoiler needs to create a situation in which Duplicator can't select a new leg of w, while Spoiler selected a new leg of z. In order to do this, Spoiler has to select all the legs of w first, which is exactly the described strategy.

In the same way it can be argued that the other two options are the fastest methods that guarantee a win by Spoiler.

4.2 Convergence law

Recall $\mathcal{C}$, the class of labeled caterpillars. This class does not obey a Zero-One law. Take $\varphi \in \mathcal{L}$ as in (4.2), the probability of having an outer vertex of degree two. By Lemma 2, we have for an outer vertex $v_1 \in C_n$

$$\lim_{n\to\infty} \mathbb{P}_n(d(v_1) = 2) = \frac{\rho^2}{1-\rho}$$

So we find:

$$\lim_{n\to\infty} \mathbb{P}_n(C_n \models \varphi) = 1 - \left(1 - \frac{\rho^2}{1-\rho}\right)^2 \notin \{0, 1\}$$

As this probability isn't either 0 or 1, the class $\mathcal{C}$ doesn't obey a Zero-One law. However, we can prove that there's a convergence law on caterpillars. To do this, we'll introduce some notation first.

Let C denote a caterpillar. We define a *segment* in C as a spine vertex together with its legs. Thus, if C has spine length ℓ, it has exactly ℓ segments. With $[X, Y] \subset C$ we denote the subgraph of C containing and including all segments between X and Y, so $[X, X] = X$ and if X, Y are the outer segments of C, $[X, Y] = C$. Further, $|[X, Y]|$ is equal to the number of segments in $[X, Y]$. When $v_1, v_2 \in C$ are spine vertices of segments V_1, V_2, we can also write $[v_1, v_2]$ instead of $[V_1, V_2]$. As was the case with the paths, for $a \in \mathbb{N}$ and segment $V \subset C$, we can define $[V - a, V + a]$ as all the segments $W \in C$ s.t. $|[W, V]| \leq a + 1$. When V is an outer segment of C, $[V, V + a]$ the first $a + 1$ segments of C, starting from V.

4.2.1 Strategy

We want to show that there's a convergence law on $\mathcal{C}$. To do this, we want to show that for every k, there's a converging probability that Duplicator can win the k-move E-F game.

First, notice that we can see the segments in a caterpillar $\mathcal{C}$ as single vertices, connected if the segments are connected. This way, we can create a path of length ℓ. For paths, we already found important properties, regarding strategies for Spoiler and Duplicator. Looking at Lemma 10, we see that when Spoiler and Duplicator play a k-move game on two caterpillars G, H, with $\ell(G), \ell(H) \geq 2^k$, there's no winning strategy for Spoiler which consists of showing that $\ell(G) \neq \ell(H)$, as the paths $P_{\ell(G)}$ and $P_{\ell(H)}$ of length $\ell(G), \ell(H)$ respectively, are k-equivalent.

For a connected series of segments of length r, we have the following proposition:

Proposition 2 *There's a finite set $\mathcal{A}_{k,r}$ of graphs, where each graph consists of r oriented connected segments, such that for every oriented graph G of r connected segments, $\exists_{A \in \mathcal{A}_{k,r}} A \equiv_k G$.*

Proof By Proposition 1, we know that Duplicator has a winning strategy comparing two segments with k legs or more in a k-move game, thus it's k equivalent if a segment has k or more than k legs. There are r segments, and every segment has $0, \ldots, k-1$ or k or more legs. So, up to k-equivalence, there are $k+1$ options per segment, thus we can conclude that there need to only be at most $(k+1)^r$ different graphs in $\mathcal{A}_{k,r}$. $\square$

As a caterpillar consists of connected segments with outer segments containing at least 1 leg, for an oriented caterpillar C with spine length r and a $k \in \mathbb{N}$, $\exists A \in \mathcal{A}_{k,r}$ s.t. $A \equiv_k C$.

Corollary 3 *Let $C_n \in_n \mathcal{C}_n$ be a uniformly chosen random caterpillar of order n. Let A be a sequence of r connected segments and fix $k \in \mathbb{N}$. Let $V_1, \ldots, V_j, \ldots$ denote the segments of C_n. Then:*

$$\lim_{n \to \infty} \mathbb{P}_n(\exists_i [V_i, V_{i+r-1}] \ \equiv_k A) = 1$$

And not only will the sequence of segments occur, it will occur arbitrarily often almost surely in C_n, when $n \to \infty$.

Proof By Proposition 2 we know that $\exists B \in \mathcal{A}_{k,r}$ s.t. $A \equiv_k B$, and B has segments with at most k legs. By Theorem 4, B will eventually occur in C_n with probability 1. Applying Corollary 2 directly gives that it will occur arbitrarily often. $\square$

Next, we're going to describe a winning strategy for Duplicator in a k-move game. If G, H are caterpillars, we can name their outer spine vertices v_1, v_∞ and w_1, w_∞ respectively, being the spine vertices in the segments $V_1, V_\infty, W_1, W_\infty$. For a fixed parameter k, we define G_l, G_r, H_l, H_r on G, H as the $2^{k+1} + 1$ segments on both outer ends:

$$G_l = [V_1, V_1 + 2^{k+1}] \qquad G_r = [V_\infty - 2^{k+1}, V_\infty] \tag{4.3}$$

$$H_l = [W_1, W_1 + 2^{k+1}] \qquad H_r = [W_\infty - 2^{k+1}, W_\infty] \tag{4.4}$$

These subgraphs are obtained by starting on an outer segment and picking the segments following in the caterpillar, clearly having an orientation from the outside to the inside. We define G_l, G_r, H_l, H_r such that this orientation is preserved, so for instance V_1, V_∞ are recognizable as outer vertices. However, as the original caterpillar has no 'left' or 'right' orientation, V_1, V_∞ are not distinguishable.

Lemma 11 *Let $k \in \mathbb{N}$. Let $G_n \in_n \mathcal{C}_n$, $H_m \in_m \mathcal{C}_m$ be uniform randomly chosen caterpillars, conditioned on having G_l, G_r and H_l, H_r with recognizable outer vertices $V_1, V_\infty, W_1, W_\infty$ to be equivalent, so that the following holds:*

$$(G_l \equiv_k H_l \wedge G_r \equiv_k H_r) \vee (G_r \equiv_k H_l \wedge G_l \equiv_k H_r) \tag{4.5}$$

Then, the probability that Duplicator has a winning strategy goes to 1 almost surely, when $n, m \to \infty$.

The proof will consist of describing a winning strategy for Duplicator, where we use that by only looking at caterpillars G, H for which (4.5) holds, we are assured that the outer ends of the caterpillars are k-equivalent to each other. Next, we use that every finite sequence of caterpillars will occur eventually in a caterpillar of order n, when taking the limit $n \to \infty$.

Proof During this proof, we'll use capitals like X to denote segments, and lowercase x to denote the spine vertex in X. Further, we assume that $G_l \equiv_k H_l$ and $G_r \equiv_k H_r$. As basic rule, when Spoiler plays a leg vertex, so will Duplicator, and when Spoiler plays a spine vertex, Duplicator will also play a spine vertex. Suppose that as j-th move, Spoiler plays $x_j \in G$, with x_j in segment X_j. That Spoiler plays in G can be assumed without loss of generality. As response, Duplicator is going to play $y_j \in H$ such that:

$$[X_j - 2^{k-j}, X_j + 2^{k-j}] \equiv_{k-j} [Y_j - 2^{k-j}, Y_j + 2^{k-j}]$$

We'll prove by induction that this is a winning strategy for Duplicator.

Suppose that k moves are played, with $x_1, \ldots, x_k \in G$ and $y_1, \ldots, y_k \in H$ and for all x_i, y_i in segments X_i, Y_i, we have:

$$[X_i - 1, X_i + 1] \equiv_0 [Y_i - 1, Y_i + 1]$$

This directly implies that the induced graphs G_k, H_k are exactly the same, as any vertex x_m adjacent to x_i will be in $[X_i - 1, X_i + 1]$, and therefore, y_m will be in $[Y_i - 1, Y_i + 1]$, and adjacent to y_i, and vice versa.

As induction hypothesis, assume that it's true for all $j' > j$ and that we have a game where $x_1, \ldots, x_{j-1} \in G$ and $y_1, \ldots, y_{j-1}$ are already played, such that $\forall i \leq j-1$:

$$[X_i - 2^{k-j+1}, X_i + 2^{k-j+1}] \equiv_{k-j+1} [Y_i - 2^{k-j+1}, Y_i + 2^{k-j+1}] \tag{4.6}$$

As j-th move, we may assume that Spoiler plays $x_j \in G$. We now separate cases.

If $\exists_{m<j} x_m \in [X_j - 2^{k-j}, X + 2^{k-j}]$, by (4.6), we know there's a $y \in [Y_m - 2^{k+1}, Y_m + 2^{k-j+1}]$ such that:

$$[Y - 2^{k-j}, Y + 2^{k-j}] \equiv_{k-j} [X_j - 2^{k-j}, X_j + 2^{k-j}] \tag{4.7}$$

And this y will be the move y_j.

If there's no such m, but either $v_0, v_\infty \in [X_j - 2^{k-j}, X_j + 2^{k-j}]$, as we know that $G_l \equiv_k H_l$ and $G_r \equiv_k H_r$, there exists a y s.t. (4.7) holds, and this will be Duplicators y_j.

And, if x_j isn't played close to v_0, v_∞, $[X_j - 2^{k-j}, X_j + 2^{k-j}]$ is just a series of segments without any labeling. Then, we just need to find a y such that (4.7) holds. This isn't necessarily possible and depends of G_n, H_m, obviously. However, we may apply Corollary 3 to find that the probability that such a y exists, approaches 1 almost surely when $m, n \to \infty$. So, this y and this will be Duplicators y_j.

As we've a response for Duplicator in all situations, we've found a winning strategy by applying the induction hypothesis, as (4.6) is clearly satisfied for the $j+1$-th move, by choice of y_j.

So, we have an almost surely winning strategy for Duplicator as $m, n \to \infty$.

□

4.2.2 Convergence

Recall $A_{k,r}$, the set of all different k-equivalent oriented graphs consisting of r connected segments, as seen in Proposition 2. Define $\mathcal{E}_k \subset A_{k,2^{k+1}+1}$ as the set containing all different k-equivalent outer $2^{k+1}+1$ segments a caterpillar can have. This means that for any caterpillar G with a sufficiently large spine and G_l as defined in (4.4), $\exists! E \in \mathcal{E}_k$ s.t. $E \equiv_k G_l$. For every two $E, E' \in \mathcal{E}_k$, not necessarily distinct, we can define:

$$\mathcal{C}^k_{EE'} = \{C \in \mathcal{C} | (C_l \equiv_k E \wedge C_r \equiv_k E') \vee (C_l \equiv_k E' \wedge C_r \equiv_k E)\}$$

So $\mathcal{C}^k_{EE'}$ consists of all caterpillars where of the outer $2^{k+1}+1$ segments on both sides, one is k-equivalent to E and the other to E'. In any caterpillar, there's no difference between left and right orientation, thus $\mathcal{C}^k_{EE'} = \mathcal{C}^k_{E'E}$. The random caterpillar of order n in $\mathcal{C}^k_{EE'}$ is denoted with $C^{k,n}_{EE'}$.

Proposition 3 *Fix $k \in \mathbb{N}$ and let $C_n \in_n \mathcal{C}_n$ be as usual. Then*

$$\forall_{E,E' \in \mathcal{E}_k} \lim_{n\to\infty} \mathbb{P}_n(C_n \in \mathcal{C}^k_{EE'}) = p^k_{EE'} \qquad \sum_{E,E' \in \mathcal{E}_k} p^k_{EE'} = 1 \tag{4.8}$$

Where we have taken every combination E, E' only once. Furthermore, the class $\mathcal{C}^k_{EE'}$ obeys a Zero-One Law on $\mathcal{L}^k$.

Proof The existence of $p^k_{EE'}$ is a consequence of Lemma 4.

For every element $E \in \mathcal{E}_k$, there are at most countably many $2^{k+1}+1$ oriented connected segments k-equivalent to E. For all such possible sequences of segments, the probability that the first $2^{k+1}+1$ outer segments of C_n are of exactly equal degree, converges when $n \to \infty$ by Lemma 4. For any two $E, E' \in \mathcal{E}_k$, there are also countably many different $2^{k+1}+1$ segments on both sides of C_n, that are k-equivalent to E, E'. Thus, the probability that C_n has its outer $2^{k+1}+1$ segments such that they are k-equivalent to E, E' is a sum over the probability of countably many disjoint events, all having converging when $n \to \infty$. The probability that any of countably many, converging events hold, converges as well in a probability space, thus $p^k_{EE'}$ exists.

The second statement, $\sum_{E,E' \in \mathcal{E}_k} p^k_{EE'} = 1$, is the same as saying

$$\lim_{n\to\infty} \mathbb{P}(\exists_{E,E'} C_n \in C^k_{EE'}) = 1$$

Which is true, as the spine length of C_n goes to infinity when $n \to \infty$ and $\mathcal{E}_k$ is chosen to contain all possible different k-equivalent outer $2^{k+1}+1$ connected segments. That $C^k_{EE'}$ obeys a Zero-One Law on $\mathcal{L}^k$, is a direct consequence of Lemma 11, as Duplicator has a winning strategy for k moves.

$\square$

Theorem 7 *There's a convergence law on the class of caterpillars $\mathcal{C}$.*

Proof Let $\varphi \in \mathcal{L}$ be a logical question with $qd(\varphi) = k$. Then:

$$\begin{aligned}\mathbb{P}_n(C_n \models \varphi) = &\sum_{E,E' \in \mathcal{E}_k} \mathbb{P}(C_n \in \mathcal{C}^{k,n}_{EE'})\mathbb{P}_n(C_n \models \varphi | C_n \in \mathcal{C}^{k,n}_{EE'}) + \\ &\mathbb{P}_n\left(\nexists_{E,E' \in \mathcal{E}_k} C_n \in \mathcal{C}_{EE'}\right)\mathbb{P}(C_n \models \varphi | \nexists_{E,E' \in \mathcal{E}_k} C_n \in \mathcal{C}_{EE'})\end{aligned}$$

When $n \to \infty$, the spine length of C_n goes to infinity almost surely and by choice of $\mathcal{E}_k$, we know that:

$$\lim_{n\to\infty} \mathbb{P}_n\left(\nexists_{E,E' \in \mathcal{E}_k} C_n \in \mathcal{C}_{EE'}\right) = 0$$

By Proposition 3, we know that:

$$\lim_{n\to\infty} \mathbb{P}_n(C_n \in C^k_{EE'}) = p^k_{EE'} \qquad \lim_{n\to\infty} \mathbb{P}_n\left(C^{k,n}_{EE'} \models \varphi\right) \in \{0,1\} \tag{4.9}$$

Combining all this gives:

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}_n(C_n \models \varphi) &= \lim_{n\to\infty} \sum_{E,E' \in \mathcal{E}_k} \mathbb{P}(C_n \in \mathcal{C}^{k,n}_{EE'})\mathbb{P}_n(C_n \models \varphi | C_n \in \mathcal{C}^{k,n}_{EE'}) \\ &= \lim_{n\to\infty} \sum_{E,E' \in \mathcal{E}_k} p^k_{EE'}\mathbb{P}(C^{k,n}_{EE'} \models \varphi)\end{aligned}$$

Thus, $\lim_{n\to\infty} \mathbb{P}_n(C_n \models \varphi)$ converges for all $\varphi \in \mathcal{L}$ and $\mathcal{C}$ obeys a convergence law. $\square$

We now have a convergence law on the class of caterpillars, which guarantees the existence of c in $\lim_{n\to\infty} \mathbb{P}(C_n \models \varphi) = c$, but it doesn't say anything about the possible values of c. For this values of c, we can prove the following:

Proposition 4 *Define the set $\mathcal{P}(\mathcal{L})$ over the class of caterpillars $\mathcal{C}$ as:*

$$\mathcal{P}(\mathcal{L}) = \{c \in [0,1] | \exists_{\varphi \in \mathcal{L}} \lim_{n\to\infty} \mathbb{P}(C_n \models \varphi) = c\}.$$

This set $\mathcal{P}(\mathcal{L})$ is dense in $[0,1]$.

Proof If we would randomly pick a spine vertex $v \in C_n$, not an outer spine vertex, in the way described in Definition 3, we have:

$$\lim_{n\to\infty} \mathbb{P}_n(d(v) = k+2) = \frac{\rho^{k+1}}{k!}$$

In particular, for such a spine vertex, we have $\lim_{n\to\infty} \mathbb{P}_n(d(v) = 2) = \rho$. So, the probability that both the segments next to both outer vertices in C_n contain no legs, which means that $d(v) = 2$ for both spine vertices, is equal to $p = \rho^2$ when n approaches infinity. The same holds for the two segments next to these, and next to those, and so on.

Let $x \in [0,1]$ and $\epsilon > 0$. We'll show that there's a sentence $\varphi \in \mathcal{L}$ s.t.:

$$\lim_{n\to\infty} |\mathbb{P}_n(C_n \models \varphi) - x| < \epsilon$$

Which would prove that $\mathcal{P}(\mathcal{L})$ is indeed dense in $[0,1]$.

The probability that the first r non-outer segments on both sides have no legs, has a limit equal to p^r. The logical sentence, checking if these segments indeed have no legs, φ_r, can be written down as:

$$\varphi_r = \exists_{x_0\neq x_1\neq\cdots\neq x_{r+2}}(x_0 \sim x_1) \wedge (x_1 \sim x_2) \wedge \cdots \wedge (x_{r+1} \sim x_{r+2}) \tag{4.10}$$

$$\forall_{w\sim x_1}((w \neq x_2) \implies (\neg\exists_{v\neq x_1} w \sim v)) \tag{4.11}$$

$$\forall_{y\sim x_i, 2\leq i\leq r+1}(y = x_{i-1}) \vee (y = x_{i+1}) \tag{4.12}$$

As explanation, the first part picks $r+3$ vertices $x_0, \ldots, x_{r+2}$, all different, forming a path $x_0 \sim x_1$, $x_1 \sim x_2$, etcetera. Then, for all vertices w adjacent to x_1, when w is not x_2, it is of degree 1 (only adjacent to x_1). In a caterpillar, this would mean that w is a leg. That means that x_1 is only adjacent to 1 vertex of degree 2 or more, namely x_1, but it's also adjacent to x_0, so x_1 is of degree 2 or more, thus x_1 must be an outer spine vertex. Also, for all vertices y adjacent to x_i, with $2 \leq i \leq r+1$, y is either equal to x_{i-1} or x_{i+1}. Thus, as $x_2, \ldots, x_{r+1}$ are all of degree 2, all are spine vertices but they all are exactly of degree 2, so they all have no legs.

Thus, φ_r isn't true when somewhere on both sides of the caterpillar, one of the first r non-outer segments has a leg, so:

$$\lim_{n\to\infty} \mathbb{P}_n(C_n \models \varphi_r) = 1 - (\rho^r)^2 = 1 - p^r$$

As $p \neq 1$, $\exists r$ such that $p^r < \epsilon$ and $\lim_{n\to\infty} \mathbb{P}(C_n \models \neg\varphi_r) = 1 - p^r > 1 - \epsilon$. For all $a \in [0,1]$, we have $a - a(1-\epsilon) = a\epsilon < \epsilon$, thus the set $\{(1-p^r)^i\}$, with $i \in \mathbb{N}$ forms a mesh on $[0,1]$, partitioning $[0,1]$ in parts in which two elements in the same part are within a distance of ϵ. As $x \in [0,1]$, x is also in one of these parts, thus there must be a finite k such that $|x - (1-p^r)^k| < \epsilon$. We can now construct a sentence $\tilde{\varphi}$, checking if the first k blocks of r non-outer segments all have at least 1 leg, which has limiting probability:

$$|x - \lim_{n\to\infty} \mathbb{P}(C_n \models \tilde{\varphi})| = |x - (1-p^r)^k| < \epsilon$$

□

Chapter 5

Similar results

5.1 Unlabeled caterpillars

So far, we've only talked about labeled caterpillars. However, we can also look at the class of unoriented unlabeled caterpillars $\mathcal{UC}$. As before, $\mathcal{UC}_n \subset \mathcal{UC}$ are the caterpillars C of order n, and $|C| = n$ if C contains n vertices.

5.1.1 Properties

To derive the basic properties of the class of unlabeled caterpillars, it's easier to first look at oriented, unlabeled caterpillars. For this set of caterpillars, we have the following useful lemma:

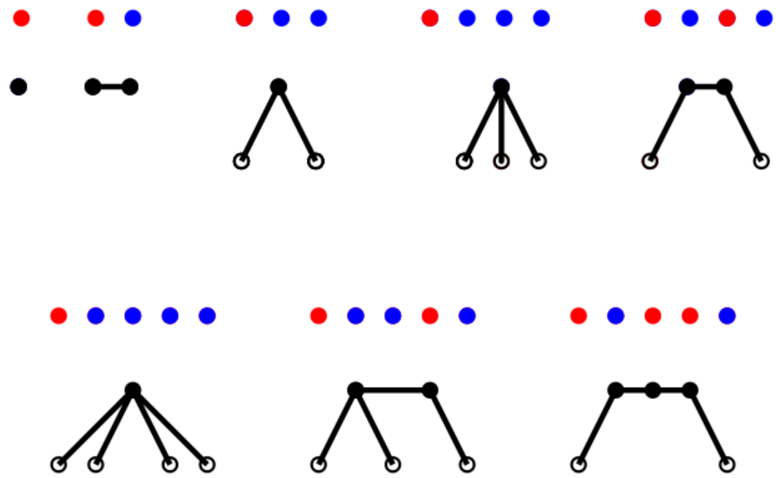

FIGURE 5.1: Every RB-coloring encodes an unlabeled caterpillar

Lemma 12 *There is a one-to-one correspondence between oriented caterpillars of order $n \geq 3$ and the $R \setminus B$-coloring of a path of $n-3$ vertices.*

Proof Let there be a path consisting of nodes $v_1, \ldots, v_n$, where every vertex is colored red or blue, with v_1 red and v_2, v_n are blue. Then, we can construct a caterpillar in the following way. Then, we start moving from v_1 up to v_n. When we see a red vertex, it's a spine vertex and the next blue vertices are it's legs, up until the next red vertex. Every spine vertex in the caterpillar is connected to the spine vertices corresponding to the surrounding red vertices on the path. This way, we can describe every oriented unlabelled caterpillar. By coloring v_1, v_2, v_n in advance, we made sure that the outer vertices have at least one leg, thus always creating a caterpillar in this way.

The other way around, having an oriented caterpillar, also gives a coloring of n vertices in a path. Start at the left spine vertex, this gives a red node on v_1. Then, for all legs the spine vertex had, we color $v_2, \dots$ blue. Then, we continue with the next spine vertex on the caterpillar, thus coloring a vertex red on the path. This way, all oriented caterpillars are translated to a coloring on the path. Thus, this procedure constructs all possible oriented caterpillars of size n exactly once. Since there are $n-3$ vertices to color freely on a path of length n, there's a one-to-one correspondence between unlabeled caterpillars of order n and $R \setminus B$-coloring of a path of $n-3$ vertices. □

Corollary 4 *The set $\mathcal{UC}_n \subset \mathcal{UC}$ has size $|\mathcal{UC}_n| \sim 2^{n-4}$*

Proof For $n \geq 3$, with Lemma 12, we can determine the number of oriented unlabeled caterpillars. With the $R \setminus B$-coloring correspondence, we immediately see that there are 2^{n-3} such oriented caterpillars. Roughly, for every unoriented caterpillar, there are two oriented caterpillars. This is true for all but the symmetric caterpillars. Thus:

$$|\mathcal{UC}_n| = \frac{1}{2}\left(2^{n-3} + |\{\text{Symmetric caterpillars of order } n\}|\right) \tag{5.1}$$

Say $n = 2k$ or $n = 2k+1$, when n is even or odd respectively. To create a symmetric caterpillar, the first $k+1$ vertices in the $R \setminus B$ representation can be colored freely, besides v_1, v_2. If n is odd, any symmetric caterpillar will have to have an odd spine length, restricting the remaining k vertices to 1 particular coloring. When n is even, the color of the $k+1$-th vertex determines whether the spine needs to have an even length (red) or odd (blue). The colors of the remaining $k-1$ are restricted for the caterpillar to be symmetric. Thus:

$$|\{\text{Symmetric caterpillars of order } n\}| = 2^{\lfloor \frac{n}{2} \rfloor - 1}$$

This is clearly negligible to the total size of $|\mathcal{UC}_n|$, when $n \to \infty$. So $\mathcal{UC}_n \sim 2^{n-4}$. □

This corollary also shows that, for large enough n, we can approach some characteristics of the class of random caterpillars as if it were an oriented one, as for almost every random caterpillar, there are precisely two oriented random caterpillars.

Proposition 5 *Let $C_n \in \mathcal{UC}_n$ denote a uniformly random chosen unlabeled caterpillar of size n, let $\ell(C) = \ell$ be the function with output the spine length of the caterpillar C. Then:*

1.
$$\mathbb{E}[\ell(C_n)] = \frac{1}{2}(n-1) + o(1) \quad \mathbb{P}[\ell(C_n) = \ell] = \binom{n-3}{\ell-1} 2^{-n+3} + o(1)$$

2. *Let v be a random spine vertex. Then:*

$$\lim_{n\to\infty} \mathbb{P}(d(v) = k) = \left(\frac{1}{2}\right)^{k-1} \forall k \geq 2 \qquad \lim_{n\to\infty} \mathbb{E}[d(v)] = 3$$

Proof We'll use the $R \setminus B$-coloring of Lemma 12 sometimes, which uses a way to describe every oriented caterpillar of order n as the result of the coloring of $n-3$ vertices.

For the first statement, with the $R \setminus B$-coloring, the spine length of oriented caterpillars C_n of order n is given by $\ell(C_n) = 1 + \#\{\text{chosen red vertices}\}$. As there are $n-3$ vertices to be colored red or blue with equal probability, we expect $\frac{1}{2}(n-3)$ to be colored red and $\ell(C_n) = 1 + \frac{1}{2}(n-3) = \frac{1}{2}n - 1$.

As we saw, there are also symmetric oriented caterpillars, so $\mathbb{E}(\ell(C_n))$ is not necessarily the same for unoriented caterpillars. However, in the proof of Corollary 4, it's argued that we can still use the $R \setminus B$-coloring as the total number of symmetric caterpillars is negligible for large n.

For the distribution of the spine length $\ell(C_n)$, we have $\ell(C_n) = \ell$ if $\ell - 1$ of the $n-3$ vertices are colored red. All colorings occur with equal probability 2^{-n+3} and there are $\binom{n-3}{\ell-1}$ ways to color $\ell - 1$ out of $n-3$ vertices red. Again, we can write $o(1)$ to deal with the symmetric cases.

The second statement. Again, we analyse the oriented caterpillars and when $n \to \infty$, a random caterpillar is asymmetric almost surely, we can directly use the results. Every spine vertex is already adjacent to 2 vertices, namely the neighboring spine vertices or, if it's an outer vertex, also the obligatory leg vertex. The degree of a spine vertex being k corresponds to the next $k-2$ vertices being blue in the $R \setminus B$-coloring, and the one after that being red. In the $R \setminus B$-coloring, we can pick a red vertex, not an outer one, randomly and the probability that the next $k-2$ vertices are indeed blue and the following red, is $2^{-(k-2)-1} = 2^{-(k-1)}$ (given that there are that many vertices to color left). When it is an outer vertex, the next $k-1$ vertices should be blue, following with a red vertex. However, the first blue vertex is already colored, so again, there's just need to be exactly $k-2$ blue vertices following the first blue vertex. Again, this happens with probability $2^{-(k-1)}$. When the order of the caterpillar goes to infinity, we get:

$$\lim_{n\to\infty} \mathbb{P}(d(v) = k) = 2^{-k+1}$$

Finally, for any red vertex v, when $n \to \infty$, the expected value of consecutive blue vertices is given by:

$$\mathbb{E}[d(v)] = \lim_{n\to\infty} \sum_{i=2}^{n-1} i2^{-i+1} = \lim_{n\to\infty} 2 + \sum_{i=0}^{n-3} i2^{-i-1} = 2 + 1 = 3 \tag{5.2}$$

Where we've used that:

$$\lim_{n\to\infty} \sum_{i=0}^{n-3} i2^{-i-1} = \frac{1}{2} \sum_{i=0}^{\infty} i2^{i} = \frac{1}{2} \cdot 2 = 1$$

$\square$

We can also state a Lemma similar to Lemma 4 and Theorem 4, only this time for unlabeled caterpillars.

Lemma 13 *Fix $i, r \in \mathbb{N}$ and let C_n be a uniformly random oriented unlabeled caterpillar, with spine length $\ell > r + i$. Let $v_1, \ldots, v_\ell$ denote the spine vertices of C_n. Fix $k_0, \ldots, k_{r-1} \in \mathbb{N}_{\geq 2}$ and take $K = \sum_j k_j$. Then:*

$$\lim_{n\to\infty} \mathbb{P}(d(v_i) = k_0, \ldots, d(v_{i+r-1}) = k_{r-1}) = \lim_{n\to\infty} \prod_{j=0}^{r-1} \mathbb{P}(d(v_{i+j}) = k_j)$$
$$= 2^{-K+r}$$

Furthermore, the probability that there is a sequence of r segments with degrees $k_0, \ldots, k_{r-1}$, goes to 1 almost surely as $n \to \infty$:

$$\lim_{n\to\infty} \mathbb{P}_n(\exists_i \ \ d(v_i) = k_0, \ldots, d(v_{i+r-1}) = k_{r-1}) = 1$$

Proof Starting from a red vertex v_i in an oriented caterpillar, the probability that the following $K - r$-vertices are correctly colored red or blue to satisfy $d(v_{i+j}) = k_j$ for all $j < r$, is given by 2^{-K+r}. When $n \to \infty$, the random unoriented caterpillar will have an equal distribution the to the random oriented caterpillar. That any finite sequence of segments will occur, can be proven in the same way as done in Theorem 4, once the positive, converging limit probability is proven. □

Now that we've proven the existence of any finite sequence in C_n, when $n \to \infty$, we can formulate and prove the FO-convergence law on $\mathcal{UC}$

Theorem 8 *The class $\mathcal{UC}$ of unlabeled caterpillars obeys a convergence law on $\mathcal{L}$.*

Proof The structure of an unlabeled caterpillar is the same as for labeled caterpillars, they are isomorphic when disregarding the labeling, so the strategy of Duplicator as described in Lemma 11 can still be applied.

We can create unlabeled versions $\mathcal{UC}^k_{EE'}$ of the labeled classes $\mathcal{C}^k_{EE'}$ as seen in Propostion 3, and by Lemma 13 together with Proposition 2, we can prove that there's a Zero-One law on these $\mathcal{UC}^k_{EE'}$ with regard to $\mathcal{L}^k$. Then, the theorem can be proven in the same way as Theorem 7. □

5.2 Self-eating caterpillars

The class of caterpillars obeys a convergence law, and no Zero-One law, on First Order logic, since there's an F-O question φ checking the degree of the outer ends in a caterpillar. Even when the size of the caterpillar goes to infinity, the distribution of the legs on the outer ends converges to different values with different positive probabilities, so $\lim_{n\to\infty} \mathbb{P}(C \models \varphi) \notin \{0, 1\}$. However, there's a Lemma that says that every finite series of segments will occur in every caterpillar. This leads to the thought that, when Spoiler isn't allowed to play in the outer ends, or, when there are no outer ends, Duplicator would always have a winning strategy.

Definition 4 *A graph $B = (V, E)$ is a bracelet when it consists of a polygon, the spine, with attached to each spine vertex some leg vertices, which are of degree 1. The order of a bracelet is as usual the number of vertices $|V|$ and $\mathcal{B}$ denotes the class of all labeled bracelets, and $\mathcal{B}_n \subset \mathcal{B}$ all labeled bracelets of order n.*

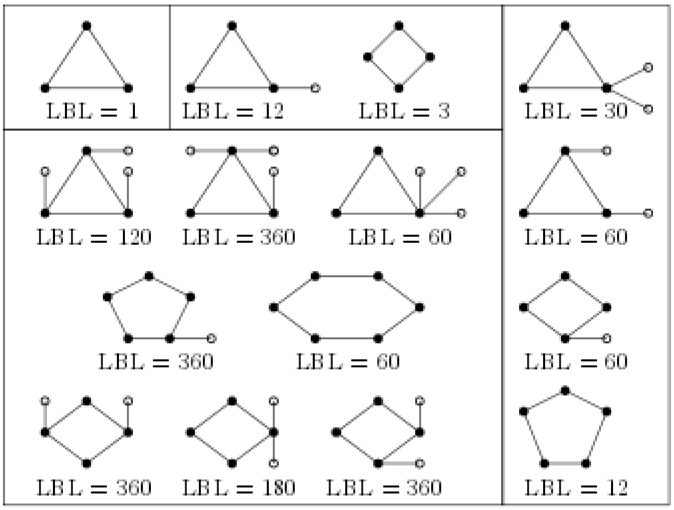

FIGURE 5.2: Some bracelets of small order

In Figure 5.2, the first couple of bracelets are shown, together with the number of different labelings they have.

5.2.1 Zero-One Law

To prove a Zero-One Law on this bracelet, we first need to determine the distributions of the spine length and the legs of the random bracelet B_n.

Proposition 6

$$|\mathcal{B}_n| = \sum_{k=3}^{n} B_{n,k} = \frac{1}{2}\sum_{k=3}^{n} \frac{n!}{(n-k)!} k^{n-k-1}$$

where $B_{n,k}$ is the number of bracelets of size n with spine length k.

Proof When labeling a bracelet B of order n, one uses the labels $1, \ldots, n$. To some vertex $v \in B$, the label 1 will be assigned and this vertex is either a spine vertex or a leg.

When v is a spine vertex and B a bracelet of spine length k, then there are $\frac{1}{2}\binom{n-1}{k-1}(k-1)!$ different ways to label the other spine vertices, since it's basically an unoriented path of length $k-1$, and k^{n-k} ways to distribute the left over $n-k$ vertices with their labels as legs over the k spine vertices. So:

$$|B_{n,k} \cap \{1 \text{ is a spine vertex}\}| = \frac{1}{2}\binom{n-1}{k-1}(k-1)!k^{n-k} = \frac{1}{2}\frac{(n-1)!}{(n-k)!}k^{n-k}$$

When the vertex v, assigned to label 1, is a leg, however, there are $n-1$ options for its corresponding spine vertex, $\frac{1}{2}\binom{n-2}{k-1}(k-1)!$ to label the other spine vertices and k^{n-k-1} options to distribute the other $n-k-1$ leg vertices over k spine vertices. This gives:

$$|B_{n,k} \cap \{1 \text{ is a leg vertex}\}| = \frac{1}{2}(n-1)\binom{n-2}{k-1}(k-1)!k^{n-k-1} = \frac{1}{2}\frac{(n-1)!}{(n-k-1)!}k^{n-k-1}$$

Combining these two terms gives:

$$\begin{aligned} B_{n,k} &= \frac{1}{2}\frac{(n-1)!}{(n-k)!}k^{n-k} + \frac{1}{2}\frac{(n-1)!}{(n-k-1)!}k^{n-k-1} \\ &= \frac{1}{2}\frac{(n-1)!}{(n-k)!}k^{n-k-1}\left(k + (n-k)\right) \\ &= \frac{1}{2}\frac{n!}{(n-k)!}k^{n-k-1} \end{aligned}$$

Obviously, $|\mathcal{B}_n| = \sum_k B_{n,k}$, so we see that:

$$|\mathcal{B}_n| = \frac{1}{2}\sum_{k=3}^{n}\frac{n!}{(n-k)!}k^{n-k-1}$$

□

Having an expression for the number of bracelets of order n, and an expression for the number of bracelets of order n with spine length k, we can formulate a Lemma similar to Theorem 3, determining the limiting distribution of $\ell(B_n)$ when $n \to \infty$. Recall $\mu_n = \{\frac{1}{n}, \ldots, 1\}$ and the constant $c = \frac{1}{1+\rho}$, with ρ the solution of $xe^x = 1$.

Lemma 14 *For a random bracelet $B_n \in_n \mathcal{B}_n$ of order n:*

$$\lim_{n\to\infty} \sum_{m \in \mu_n \cap [c-\alpha, c+\alpha]} \mathbb{P}_n(\ell(B_n) = mn) = 1 \quad \forall \alpha > 0$$

Proof The probability that a uniformly random bracelet of order n has spine length ℓ, is given by:

$$\mathbb{P}_n(\ell(B_n) = \ell) = \frac{\mathcal{B}_{n,\ell}}{|\mathcal{B}_n|}$$

As $|\mathcal{B}_n| = \sum_{k\in\mu_n} |\mathcal{B}_{n,kn}|$, and we can write $\ell = mn$, with $m \in \mu_n$, we find that:

$$\mathbb{P}_n(\ell(B_n) = mn) = \frac{1}{\sum_{k\in\mu_n} \frac{|\mathcal{B}_{n,mn}|}{|\mathcal{B}_{n,kn}|}}$$

The number of bracelets with spine length ℓ is equal to $\mathcal{B}_{n,\ell} = \frac{n!}{2(n-\ell)!}\ell^{n-\ell-1}$. We can write, using Stirlings approximation and $\ell = mn$:

$$\begin{aligned}
B_{n,mn} &= \frac{n!}{2((1-m)n)!}(mn)^{(1-m)n-1} \\
&\sim \frac{n!}{2}\frac{1}{\sqrt{2\pi(1-m)n}}\left(\frac{e}{(1-m)n}\right)^{(1-m)n}(mn)^{(1-m)n-1} \\
&= \frac{n!}{2(mn)}\frac{1}{\sqrt{2\pi(1-m)n}}\left(\frac{em}{1-m}\right)^{(1-m)n}
\end{aligned}$$

For $\tilde{m} \in \mu_n$, with $m \neq \tilde{m}$, we can calculate $\frac{|\mathcal{B}_{n,mn}|}{|\mathcal{B}_{n,\tilde{m}n}|}$:

$$\begin{aligned}
\frac{|\mathcal{B}_{n,mn}|}{|\mathcal{B}_{n,\tilde{m}n}|} &\sim \frac{\tilde{m}}{m}\sqrt{\frac{1-\tilde{m}}{1-m}}\left(\left(\frac{em}{1-m}\right)^{1-m}\left(\frac{1-\tilde{m}}{e\tilde{m}}\right)^{1-\tilde{m}}\right)^{n} = b\cdot A^n \\
b &= \frac{\tilde{m}}{m}\sqrt{\frac{1-\tilde{m}}{1-m}} \qquad A = \left(\frac{em}{1-m}\right)^{1-m}\left(\frac{1-\tilde{m}}{e\tilde{m}}\right)^{1-\tilde{m}}
\end{aligned}$$

Recall the analysis of A as done in Theorem 3, around equation (2.4), and that $B_{n,mn}$ takes its largest values around when m is around c, Fix $\alpha > 0$, for large enough n, we know that $\mu_n \cap [c-\frac{1}{2}\alpha, c] \neq \emptyset$ and $\mu_n \cap [c, c+\frac{1}{2}\alpha] \neq \emptyset$. For any $\ell = mn$, $m \in \mu_n$, with $m \leq c - \alpha$, we can find the bound:

$$\mathbb{P}(\ell(B_n) = mn) = \frac{|\mathcal{B}_{n,mn}|}{\sum_{k\in\mu_n}|\mathcal{B}_{n,kn}|} \leq \frac{|\mathcal{B}_{n,mn}|}{|\mathcal{B}_{n,\tilde{m}n}|} \quad \tilde{m} \in \mu_n \cap \left[c - \frac{\alpha}{2}, c\right]$$

Since $m \leq (c-\alpha)n$, we have an upper bound for $|B_{n,mn}|$ and since $cn \geq \tilde{m} \geq (c - \frac{\alpha}{2})n$, we have a lower bound for $B_{n,\tilde{m}n}$. Combining this gives:

$$\mathbb{P}_n(\ell(B_n) = mn) \leq bA_l^n \quad A_l = \left(\frac{e(c-\alpha)}{1-(c-\alpha)}\right)^{1-(c-\alpha)}\left(\frac{1-(c-\frac{\alpha}{2})}{e(c-\frac{\alpha}{2})}\right)^{1-(c-\frac{\alpha}{2})}$$

Where $A < 1$. Then:

$$\mathbb{P}_n(\ell(B_n) \leq (c-\alpha)n) = \sum_{m\in\mu_n, m\leq(c-\alpha)n} \mathbb{P}_n(\ell(B_n) = mn) \leq n \cdot bA^n$$

As A^n dominates the polynomial n when $n \to \infty$, we get:

$$\lim_{n\to\infty} \mathbb{P}_n(\ell(B_n) \leq (c-\alpha)n) \leq \lim_{n\to\infty} nbA^n = 0$$

This line of reasoning can also be applied to find an upper bound of $\mathbb{P}_n(\ell(B_n) = mn)$, with $m \geq (c+\alpha)n$ and to find that $\lim_{n\to\infty} \mathbb{P}(\ell(B_n) \geq (c+\alpha)n) = 0$. So, we see that:

$$\lim_{n\to\infty} \sum_{m\in\mu_n\cap[c-\alpha,c+\alpha]} \mathbb{P}_n(\ell(B_n) = mn) = 1$$

And this holds for all $\alpha > 0$. $\square$

To find the distribution of the legs of a random spine vertex on a random bracelet, in the sense of Definition 3, when $n \to \infty$, we use an approach which is again similar to the caterpillar case and Lemma 2. So if v is a randomly chosen spine vertex on a random bracelet B_n with spine length $\ell = mn$, the probability that it has k legs is equal to:

$$\mathbb{P}_n(d(v) = k+2) = \sum_{\ell=3}^{n-k} \mathbb{P}_n(d(v) = k+2 | \ell(B_n) = \ell)\mathbb{P}_n(\ell(B_n) = \ell)$$

We have an exact expression for $\mathbb{P}_n(d(v) = k+2|\ell(B_n) = mn)$, given by:

$$\mathbb{P}_n(d(v) = k+2|\ell(B_n) = mn) = \frac{1}{2(mn)|\mathcal{B}_{n,mn}|}\binom{n}{k}\binom{n-k}{mn}(mn)!(mn-1)^{n-mn-k}$$

As we pick k labels for the vertices in the leg of v, mn labels for the vertices in the spine of B_n, which can be in $\frac{1}{2}(mn)!$ different orders on the spine and distribute the remaining $n - mn - k$ vertices as legs over the remaining $mn - 1$ spine vertices. Then, we divided by $2(mn)$, since the bracelet can be rotated and it's not oriented, and by $|\mathcal{B}_{n,mn}|$. We use that $|\mathcal{B}_{n,mn}| = \frac{n!}{2(n-mn)!}(mn)^{n-mn-1}$ to find:

$$\begin{aligned}
\mathbb{P}_n(d(v) = k+2|\ell(B_n) = mn) &= \frac{1}{(mn)k!}\frac{(n-mn)!}{(n-mn-k)!}\frac{(mn-1)^{n-mn-k}}{(mn)^{n-mn-1}} \\
&= \frac{1}{k!}\frac{(n-mn)!}{(n-mn-k)!}\frac{(mn-1)^{n-mn-k}}{(mn)^{n-mn}} \\
&= \frac{1}{k!}\left(\frac{mn-1}{mn}\right)^{n-mn}\frac{\prod_{j=0}^{k-1}(n-mn-j)}{(mn-1)^k}
\end{aligned}$$

Like we did in Lemma 2, we can write:

$$\left(1-\frac{1}{mn}\right)^{(1-m)n} = \exp\left(-(1-m)n\log 1 - \frac{1}{mn}\right) = e^{-\frac{1-m}{m}+\mathcal{O}(\frac{1}{n})}$$

$$\prod_{j=0}^{k-1}\frac{(1-m)n-j}{mn-1} = \left(\frac{1-m}{m}\right)^k\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right)$$

$$\mathbb{P}_n(d(v) = k+2|\ell(B_n) = mn) = \frac{1}{k!}\left(\frac{1-m}{m}\right)^k e^{\frac{1-m}{m}}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right)$$

By Lemma 14, we may assume $m \in [c-\alpha, c+\alpha]$ for $\alpha > 0$. For these m, we can find upper and lower bounds for $\mathbb{P}_n(d(v) = k+2|\ell(B_n) = mn)$, as:

$$\mathbb{P}_n(d(v) = k+2|\ell(B_n) = mn) \geq \frac{1}{k!}\left(\frac{1-(c+\alpha)}{c+\alpha}\right)^k e^{-\frac{1-(c-\alpha)}{c-\alpha}}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right)$$

$$\mathbb{P}_n(d(v) = k+2|\ell(B_n) = mn) \leq \frac{1}{k!}\left(\frac{1-(c-\alpha)}{c-\alpha}\right)^k e^{-\frac{1-(c+\alpha)}{c+\alpha}}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right)$$

For any $\alpha > 0$, we now have an upper and lower bound for:

$$\lim_{n\to\infty}\mathbb{P}_n(d(v) = k+2) = \lim_{n\to\infty}\sum_{m\in\mu_n\cap[c-\alpha,c+\alpha]}\mathbb{P}_n(d(v) = k+2|\ell(B_n) = mn)\mathbb{P}_n(\ell(B_n) = mn)$$

And as the lower bound equals the upper bound when $\alpha \downarrow 0$, we get, using that $\frac{1-c}{c} = \rho$, with $\rho e^{\rho} = 1$:

$$\lim_{n\to\infty} \mathbb{P}_n(d(v) = k+2) = \lim_{\alpha\downarrow 0} \frac{1}{k!}\left(\frac{1-(c+\alpha)}{c+\alpha}\right)^k e^{-\frac{1-(c-\alpha)}{c-\alpha}} = \frac{\rho^{k+1}}{k!} \tag{5.3}$$

We can extrapolate this approach to any sequence of legs $k_1, \ldots, k_r$ on connected spine vertices $v_1, \ldots, v_r$ to get, with $K = \sum_{i=1}^r k_i$:

$$\lim_{n\to\infty} \mathbb{P}(d(v_1) = k_1 + 2, \ldots, d(v_r) = k_r + 2) = \frac{\rho^{K+r}}{k_1! \cdots k_r!} \tag{5.4}$$

From here, we immediately get that any sequence of legs $k_1, \ldots, k_r$ will eventually occur on B_n, when $n \to \infty$. After all, if any sequence $v_1, \ldots, v_r$ has legs $k_1, \ldots, k_r$ with probability p, it won't with probability $1-p$ and the same holds for the next r spine vertices, and the next, etc, so the probability of not having a sequence $v_1, \ldots, v_r$ with legs $k_1, \ldots, k_r$ is less than $(1-p)^{\lfloor \frac{\ell}{r} \rfloor}$. Since $\ell(B_n) \to \infty$, $(1-p)^{\lfloor \frac{\ell}{r} \rfloor} \to 0$ and the probability that B_n has such a connected sequence of segments goes to 1, when $n \to \infty$. So, we can combine this all to:

Lemma 15 *The probability that a connected sequence of spine vertices, with v_1 as randomly chosen starting point, $v_1, \ldots, v_r$ in B_n has legs $k_1, \ldots, k_r$, with $K = \sum_i k_i$, is in the limit of $n \to \infty$ given by $\frac{\rho^{K+r}}{k_1! \cdots k_r!}$. The probability that there is no such sequence of connected spine vertices in B_n approaches 0 almost surely when $n \to \infty$.*

□

Theorem 9 *The class $\mathcal{B}$ obeys a Zero-One Law on $\mathcal{L}$.*

Proof We'll describe a strategy for Duplicator. Let a spine vertex together with its legs be a segment, as it was with caterpillars, where the spine vertex v is in segment V. For $m \in \mathbb{N}$, we define $[V-m, V+m]$ the sequence of m segments on both sides of V.

Spoiler and Duplicator play a k-move game on uniform, randomly chosen bracelets G, H, where G is of order m and H or order n and we'll describe a stratey that will be winning with high probability for Duplicator when $m, n \to \infty$.

The strategy Suppose $x_1, \ldots, x_{j-1} \in G$, $y_1, \ldots, y_{j-1} \in H$ are already played and Spoiler plays its j-th move, wlog we assume $x_j \in G$, in segment X_j. Duplicator will respond by playing $y_j \in H$, in segment Y_j, such that $[Y_j - 2^{k-j}, Y_j + 2^{k-j}] \equiv_{k-j} [X_j - 2^{k-j}, X_j + 2^{k-j}]$, playing a spine vertex when Spoiler played a spine vertex and a leg vertex when Spoiler played a leg vertex.

Why is this a winning strategy for Duplicator almost surely, when $n \to \infty$? We'll first prove the following claim:

Claim Suppose $x_1, \dots, x_j \in G, y_1, \dots, y_j \in H$ are already played, x_i a spine vertex iff y_i is a spine vertex. When

$$[X_i - 2^{k-j}, X_i + 2^{k-j}] \equiv_{k-j} [Y_i - 2^{k-j}, Y_i + 2^{k-j}] \quad \forall_{i \le j}, \tag{5.5}$$

Duplicator has a winning strategy almost surely.
Proof of Claim We'll prove it by backward-induction. When $j = k$, the game is over and $[X_i - 1, X_i + 1] = [Y_i - 1, Y_i + 1]\ \forall i$. Thus, all adjacent x_p, x_q are also adjacent as y_p, y_q, thus the induced subgraphs G_k, H_k are equal and Duplicator has won.

Suppose that for all $j' < j$, the claim is true, and j moves are already played, and (5.5) holds. Suppose Duplicator plays $x_{j+1} \in G$ as $j+1$-th move.

When $\forall_{i \le j} x_i \notin [X_{j+1} - 2^{k-j-1}, X_{j+1} + 2^{k-j-1}]$, Duplicator only has to find y such that

$$\forall_{i \le j} y_i \notin [Y - 2^{k-j-1}, Y + 2^{k-j-1}]$$
$$[X_{j+1} - 2^{k-j-1}, X_{j+1} + 2^{k-j-1}] \equiv_{k-j-1} [Y - 2^{k-j-1}, Y + 2^{k-j-1}]$$

We can apply Corollary 3, together with the existence of all finite sequence of segments as proven in Lemma 15, to see that this y will exists almost surely when $n \to \infty$ and this y will be the $j+1$-th move.

Otherwise, when $\exists_{i \le j} x_i \in [X_{j+1} - 2^{k-j-1}, X_{j+1} + 2^{k-j-1}]$, as equation (5.5) holds, we know that:

$$[X_{j+1} - 2^{k-j-1}, X_{j+1} + 2^{k-j-1}] \subset [X_i - 2^{k-j}, X_i + 2^{k-j}] \equiv_{k-j} [Y_i - 2^{k-j}, Y_i + 2^{k-j}]$$

Thus, there will be a $y \in [Y_i - 2^{k-i}, Y_i + 2^{k-i}]$ such that $[X_{j+1} - 2^{k-j-1}, X_{j+1} + 2^{k-j-1}] \equiv_{k-i} [Y - 2^{k-j-1}, Y + 2^{k-j-1}]$, and Duplicator will play this y as y_{j+1}. By playing this, Duplicator assures that during the game, after $j+1$ moves:

$$\forall_{i \le j+1}\ [X_i - 2^{k-j-1}, X_i + 2^{k-j-1}] \equiv_{k-j-1} [Y_i - 2^{k-j-1}, Y_i + 2^{k-j-1}]$$

And, by the induction hypothesis, Duplicator has an almost surely winning strategy from here and we've proven the Claim.

When Spoiler and Duplicator start the game over k moves on two bracelets $G \in_m \mathcal{B}_m$, $H \in_n \mathcal{B}_n$, Duplicator will follow his strategy. By his choice of y_j, such that:

$$[X_j - 2^{k-j}, X_j + 2^{k-j}] \equiv_{k-j} [Y_j - 2^{k-j}, Y_j + 2^{k-j}]$$

After j moves, it's guaranteed that (5.5) holds. And this is true for all j, so when $j = k$, it's also true, thus Duplicator has won. The probability that he can apply this strategy, is 1 as $|G|, |H| \to \infty$. Thus, there's a Zero-One Law on $\mathcal{B}$.

□

Chapter 6

Discussion

In this master thesis, we've mostly discussed the class of caterpillars $\mathcal{C}$ and the limiting behavior of its uniform randomly chosen caterpillar $C_n \in \mathcal{C}_n$ of order n. When $n \to \infty$, the spine length $\ell(C_n)$ of the caterpillar will be $(c + o(1))n$ with high probability. Furthermore, when picking a random spine vertex, we know how the number of legs attached to it, is distributed, whether the spine vertex was an outer vertex or not. Moreover, this distribution is independent of the legs of other spine vertices.

Also, every finite sequence of connected spine vertices with fixed degree will eventually occur in the random caterpillar C_n almost surely, as n approaches infinity. This is an important step in showing the convergence law on $\mathcal{C}_n$, as it shows that Duplicator can find any desired sequence of segments almost surely, thus giving him a winning strategy under certain conditions, leading to the convergence law on the class $\mathcal{C}$.

On the unlabeled class, the expected spine length of UC_n, differs. However, the unlabeled caterpillar still has spine length converging to $\frac{1}{2}$ almost surely, and the distribution of the legs per spine vertex converges as well. So any finite sequence of segments will also occur on the unlabeled caterpillar UC_n almost surely, so on UC_n there's also a convergence law.

There's no Zero-One law on these classes, as the outer ends can be very different and are recognizable as outer ends. However, when the two outer ends of the caterpillars are attached, creating bracelets and the class of bracelets $\mathcal{B}_n$, this disappears. As any finite sequence of segments still occurs almost surely on $B_n \in_n \mathcal{B}_n$, when $n \to \infty$, Duplicator will have a winning strategy with high probability, On the class $\mathcal{B}_n$, we did find a Zero-One law.

The strategy applied by Duplicator to win the Ehrenfeucht-Fraissé game in all these classes, is based on a few things. First of all, there's the spine. The spine consists of all vertices of degree 2 or more, and importantly, is just a path to which other vertices are attached. Via this construction, the caterpillar can be seen as a series of disjoint graphs, all connected via one particular vertex in every graph. In the case of the caterpillar, these disjoint graphs are stars, graphs with a single vertex that is attached to all other vertices.

This approach could perhaps be extended, to for instance graphs where a spine connects all types of circular graphs $D_3, D_4, D_5, \ldots$, or a graph where a spine connects graphs chosen out of a finite (or even infinite) set M of possible graphs. As long as the construction of the spine can be done in a non-ambiguous way, the spine length goes to infinity almost surely, and the distribution of the graphs attached to the spine converge when $n \to \infty$, there will be a possibility for Duplicator to win the game, perhaps under certain conditions around the ends of the spine.

The main reason why Duplicator can hope for a winning strategy, is that if the spine grows, most of the vertices will be ’far’ away from each other. In K_n, every vertex is adjacent to every other vertices, so there’s only a distance of 1 between two vertices. When a graph consists of a spine connecting disjoint graphs, most vertices will be far away from each other, the shortest path of adjacent vertices between v and w will most likely still be long. This allows Duplicator to play his move y in a way that, locally around y, so within a certain distance, everything seems equal how it looks locally around x, the move made by Spoiler.

Overall, it seems that the convergence law on the caterpillar class could be extended to some kind general law on graphs with a ’spine’ that satisfy some requirements.

Bibliography

[1] Noga Alon and Joel Spencer. *The Probabilistic Method*. New York: John Wiley & Sons, 1990, 254p.

[2] Peter Cameron. *The mathematics of Paul Erdos II*. New York: Springer, 1997, pp. 353–378.

[3] Frank Harary and Allen Schwenk. “The number of caterpillars”. In: *Discrete Mathematics* 6.4 (1973), pp. 359–365.

[4] Peter Heinig et al. “Logical limit laws for minor closed graphs”. 2014.

[5] J.Gajjar. *Laplace’s method*. *Available at* http://www.maths.manchester.ac.uk/~gajjar/MATH44011/notes/44011_note3.pdf.

[6] Laszlo Lovasz. “Graph minor theory”. In: *Bulletin of AMS* 43.1 (2005), pp. 75–86.

[7] Saharon Shelah and Joel Spencer. “Zero-one laws for sparce random graphs”. In: *Journal of American Mathematical Society* 1.1 (1988), pp. 97–115.