1. Abstract

We are concerned with exploring the probabilities of first order statements for Galton-Watson trees with Poisson $c$ offspring distribution. Fixing a positive integer $k$, we exploit the $k$-move Ehrenfeucht game on rooted trees for this purpose. Let $\Sigma$, indexed by $1 \leq j \leq m$, denote the finite set of equivalence classes arising out of this game, and $D$ the set of all probability distributions over $\Sigma$. Let $x_j(c)$ denote the true probability of the class $j \in \Sigma$ under Poisson $c$ regime, and $\vec{x}(c)$ the true probability vector over all the equivalence classes. Then we are able to define a natural recursion function $\Gamma$, and a map $\Psi = \Psi_c : D \to D$ such that $\vec{x}(c)$ is a fixed point of $\Psi_c$, and starting with any distribution $\vec{x} \in D$, we converge to this fixed point via $\Psi$ because it is a contraction. We show this both for $c \leq 1$ and $c > 1$, though the techniques for these two ranges are quite different.

2. Introduction

The Galton-Watson tree (henceforth, GW tree) $T = T_c$ with parameter $c > 0$ is a much studied object. It is a random rooted tree. Each node, independently, has $Z$ children where $Z$ has the Poisson distribution with mean $c$. We let $P_c$ denote the probability under $T_c$.

We shall examine the first order language on rooted trees. This consists of a constant symbol $R$ (the root), equality $v = w$, and a parent function $\pi[v]$ defined for all vertices $v \neq R$. (Purists may prefer a binary relation $\pi^*[v,w]$, that $w$ is the parent of $w$.) Sentences must be finite and made up of the usual Boolean connectives ($\neg, \lor, \land, \ldots$) and existential $\exists_v$ and universal $\forall_v$ quantification over vertices. The quantifier depth of a sentence $A$ is the depth of the nesting of the existential and universal quantifiers.

Example 2.1. No node has precisely one child.

$$\neg[\exists_u \pi[x] = u \land \forall_z ([z \neq x] \Rightarrow \neg \pi[z] = u)]. \quad (2.1)$$

We outline some of our results. For any first order sentence $A$ set $f_A(c) = \Pr[T_c \models A]$.

the probability that $T = T_c$ has property $A$. Except in examples we will work with the quantifier depth $k$ of $A$. The value $k$ shall be arbitrary but fixed throughout this presentation. With $3.3$ below we decompose any $f_A(c)$ into its “atomic” $x_j(c)$. In Section $3.3$ we show that the $x_j(c)$ are solutions to a finite system of equations involving polynomials and exponentials. The solution is described as the fixed point of a map $\Psi_c$ over the space of distributions $D$ defined by $(3.1)$. In Theorem $7.1$ we show that this system has a unique solution. In Sections $4.3$ (for the subcritical case) and $8$ (for the general case) we show that $\Psi_c$ is a contraction. Employing the Implicit Function Theorem in Section $9$ we then achieve one of our main results:

Theorem 2.2. Let $A$ be first order. Then $f_A(c)$ is a $C^\infty(0, \infty)$ function. That is, all derivatives of $f_A(c)$ exist and are continuous at all $c > 0$. 

Remark 2.3. Let $y = g(c)$ be the probability that $T = T_c$ is infinite. It is well known that $g(c) = 0$ for $c \leq 1$ while for $c > 1$, $y = g(c)$ is the unique positive real satisfying $e^{-cy} = 1 - y$. The value $c = 1$ is often referred to as a critical, or percolation, point for GW-trees. The function $g(c)$ is not differentiable at $c = 1$. The right sided derivative $\lim_{c \to 1^+} (g(c) - g(1))/(c - 1)$ is 2 while the left sided derivative is zero. An interpretation of Theorem 2.2 that we favor is that the critical point $c = 1$ cannot be seen through a First Order lens. Theorem 2.2 thus yields that the property of $T$ being infinite is not expressible in the first language – though this can be shown with a much weaker hammer!

The plot clearly shows how the function is not differentiable at $c = 1$, and how the solution is non-unique for $c > 1$.

But the plot corresponding to the property in Example 2.1 shows that the probability is a smooth function of $c$, which is in keeping with Theorem 2.2.

Definition 2.4. With $v \in T$, $T(v)$ denotes the rooted tree consisting of $v$ and all of its descendents, with $v$ regarded as the root. For $s$ a nonnegative integer, $T|_s$ denotes the rooted tree consisting of its vertices up to generation at most $s$. We call $T|_s$ the $s$-cutoff of $T$. (This is defined even if no vertices are at generation $s$.) $T(v)|_s$ denotes the $s$-cutoff of $T(v)$. 
3. The Ehrenfeucht Game

3.1. Equivalence Classes. Let $k$ denote an arbitrary positive integer. We may then define an equivalence relation $\equiv_k$ (we often omit the subscript) on all trees $T$, as follows.

**Definition 3.1.** $T \equiv_k T'$ if they have the same truth value for all $A$ of quantifier depth at most $k$. Equivalently $T \equiv_k T'$ if Duplicator wins the $k$-move Ehrenfeucht game $EHR[T, T', k]$. $\Sigma = \Sigma_k$ denotes the set of all equivalence classes.

Critically, $\Sigma_k$ is a finite set. As a function of $k$ we note that $|\Sigma_k|$ grows like a tower function. We give [2] as a general reference to these basic results.

**Definition 3.2.** For any rooted tree $T$, the Ehrenfeucht value of $T$, denoted by $EV[T]$, is that equivalence class $\sigma \in \Sigma$ to which $T$ belongs.

For convenience we denote the elements of $\Sigma$ by $\Sigma = \{1, \ldots, m\}$. We let $D \subset \mathbb{R}^m$ denote the set of all possible probability distributions over $\Sigma$. That is,

$$D = \{(x_1, \ldots, x_m) : \sum_{j=1}^m x_j = 1 \text{ and all } x_j \geq 0\}. \quad (3.1)$$

(We let $\vec{x}(c)$ denote the probability distribution for the equivalence class of $T = T_c$.) Recall that $P_c$ denotes the probability under $T_c$, the GW process with Poisson offspring. Let

$$x_j(c) = P_c(j) = P[EV[T_c] = j], \quad j \in \Sigma. \quad (3.2)$$

Then $\vec{x}(c) = \{x_j(c) : 1 \leq j \leq m\}$ denotes the probability vector in $D$ under $P_c$.

**Theorem 3.3.** $\vec{x}(c)$ has derivatives of all orders. In particular, each $x_i(c)$ has derivatives of all orders.

The proof of Theorem 3.3 is a goal of this paper, accomplished only in Section 9 after many preliminaries. Any first order sentence of quantifier depth $A$ is determined, tautologically, by the set $S(A)$ of those $j \in \Sigma$ such that all $T$ with $EV[T] = j$ have property $A$. For any $j \in \Sigma$ either all $T$ with $EV[T] = j$ or no $T$ with $EV[T] = j$ have property $A$. We may therefore decompose the $f_A(c)$ of (2.2) into

$$f_A(c) = \sum_{j \in S(A)} x_j(c). \quad (3.3)$$

Theorem 3.3 will therefore follow from Theorem 3.3.

3.2. Recursive States. In the $k$-move Ehrenfeucht game, values $\geq k$ are roughly all “the same.” This will be made precise in the subsequent Theorem 3.4. We define

$$C = \{0, 1, \ldots, k - 1, \omega\} \quad (3.4)$$

The phrase “there are $\omega$ copies” is to be interpreted as “there are $\geq k$ copies.” We call $v \in T$ a rootchild if its parent is the root $R$. For $w \neq R$ we say $v$ is the rootancestor of $w$ if $v$ is that unique rootchild with $w \in T(v)$. Of course, a rootchild is its own rootancestor.

**Theorem 3.4.** Let $\vec{n} = (n_1, \ldots, n_m)$ with all $n_j \in C$. Let $T$ have the property that for all $1 \leq j \leq m$ there are $n_j$ rootchildren $v$ with $EV[T(v)] = j$. Then $\sigma = EV[T]$ is uniquely determined.

**Definition 3.5.** Let

$$\Gamma : \{(n_1, \ldots, n_m) : n_i \in \{0, 1, \ldots, k - 1, \omega\}\} \to \Sigma \quad (3.5)$$

be given by $\sigma = \Gamma(\vec{n})$ with $\vec{n}, \sigma$ satisfying the conditions of Theorem 3.4. Then $\Gamma$ is called the recursion function.
Proof of Theorem 3.4. Let $T, T'$ have the same $\vec{n}$. We give a strategy for Duplicator in the Ehrenfeucht game $EHR(T, T', k)$. Duplicator will create a partial matching between the rootchildren $v \in T$ and the rootchildren $v' \in T'$. When $v, v'$ are matched, $EV[T(v)] = EV[T'(v')]$. At the end of any round of the game call a rootchild $v \in T$ (similarly $v' \in T'$) free if no $\omega \in T(v)$ has yet been selected.

Suppose Spoiler plays $w \in T$ (similarly $w' \in T'$) with rootancestor $v$. Suppose $v$ is free. Duplicator finds a free $v' \in T'$ with $EV(T(v)) = EV(T'(v'))$. (As the number for each class is the same for $T, T'$ this may be done when $n_j \neq \omega$.) When $EV[T(v)] = j \in \Sigma$ and $n_j \neq \omega$, then as the number of rootchildren of $T$ with Ehrenfeucht value $j$ is exactly the same as that in $T'$, hence this can be done. In the special case where $n_j = \omega$ the vertex $v'$ may be found as there have been at most $k - 1$ moves prior to this move and so there are at most $k - 1$ rootchildren $v'$ with $EV[T'(v')] = j$ that are not free. Duplicator then matches $v, v'$. Duplicator can win $EHR(T(v), T(v'), k)$ as $EV[T(v)] = EV[T'(v')]$. He employs that strategy to find a response $w' \in T(v')$ corresponding to $w \in T(v)$. Once $v, v'$ have been matched any move $z \in T(v)$ (similarly $z' \in T'(v')$) is responded to with a move in $z' \in T'(v')$ using the strategy for $EHR(T(v), T'(v'), k)$. \hfill \qed

Remark 3.6. Tree automata consist of a finite state space $\Sigma$, an integer $k \geq 1$, a map $\Gamma$ as in (3.3) and a notion of accepted states. While first order sentences yield tree automata, the notion of tree automata is broader. Tree automata roughly correspond to second order monadic sentences, a topic we hope to explore in future work.

3.3. Solution as Fixed Point. We come now to the central idea. We define, for $c > 0$, a map $\Psi_c : D \rightarrow D$. Let $\vec{x} = (x_1, \ldots, x_m) \in D$, a probability distribution over $\Sigma$. Imagine root $R$ has Poisson mean $c$ children. To each child we assign, independently, a $j \in \Sigma$ with distribution $\vec{x}$. Let $n_j \in C$ be the number of children assigned $j$. Let $\vec{n} = (n_1, \ldots, n_m)$. Apply the recursion function (equation 3.5) $\Gamma$ to get $\sigma = \Gamma(\vec{n})$, the Ehrenfeucht value of the root $R$. Then $\Psi_c(\vec{x}) \in D$ is the distribution thus induced on the Ehrenfeucht value of $R$.

The special nature of the Poisson distribution allows a concise expression. When the initial distribution is $\vec{x}$, the number of children assigned $j$ will have a Poisson distribution with mean $cx_j$ and these numbers are mutually independent over $j \in \Sigma$. Thus

\[
\Pr[n_j = u] = e^{-cx_j} \frac{(cx_j)^u}{u!} \quad \text{for } u \in C, u \neq \omega, \tag{3.6}
\]

and

\[
\Pr[n_j = \omega] = 1 - \sum_{u=0}^{k-1} \Pr[n_j = u]. \tag{3.7}
\]

From the independence, for any $\vec{a} = (a_1, \ldots, a_m)$ with $a_1, \ldots, a_m \in C$,

\[
\Pr[\vec{n} = \vec{a}] = \prod_{j=1}^{m} \Pr[n_j = a_j]. \tag{3.8}
\]

Thus, writing $\Psi_c(x_1, \ldots, x_m) = (y_1, \ldots, y_m)$,

\[
y_j = \sum \Pr[\vec{n} = \vec{a}] \tag{3.9}
\]

where the summation is over all $\vec{a}$ with $\Gamma(\vec{a}) = j$.

We place all $\Psi_c$ into a single map $\Delta$:

\[
\Delta : D \times (0, \infty) \rightarrow D \text{ by } \Delta(\vec{x}, c) = \Psi_c(\vec{x}) \tag{3.10}
\]

Setting $\Delta(x_1, \ldots, x_m, c) = (y_1, \ldots, y_m)$, the $y_j$ are finite sums of products of polynomials and exponentials in the variables $x_1, \ldots, x_m, c$. In particular, all partial derivatives of all orders exist everywhere.

Recall (3.2). $\vec{x}(c)$ denotes the probability distribution for the equivalence classes under the probability measure $P_c T = T_c$.

**Theorem 3.7.** $\vec{x}(c)$ is a fixed point for $\Psi_c : D \rightarrow D$. That is, $\Psi_c(\vec{x}(c)) = \vec{x}(c)$. 
Proof. To show that $\vec{x}(c)$ is a fixed point of iteration $\Psi_c$, we start with the initial probability vector $\vec{x} = \vec{x}(c)$. Once we perform the iteration, using the definition of the recursive function $\Gamma$ from (3.5), we know, for any $j \in \Sigma$,

$$
\Psi_c(\vec{x}(c))(j) = \sum_{\vec{a} \Gamma(\vec{a}) = j} \prod_{i=1}^{m} P[n_i = a_i | \vec{x} = \vec{x}(c)], \quad \text{from (3.8)},
$$

$$
= \sum_{\vec{a} \Gamma(\vec{a}) = j} \prod_{i=1}^{m} P[\text{Poisson}(c \cdot x_i(c)) = a_i]
$$

$$
= \sum_{\vec{a} \Gamma(\vec{a}) = j} P_c[\vec{n} = \vec{a}]
$$

$$
= P_c[j] = x_j(c),
$$

where recall that $P_c$ is the probability induced under the Poisson$(c)$ offspring distribution. \hfill \Box

**Example 3.8.** For many particular $A$ the size of $\Sigma$, which may be thought of as the state space, may be reduced considerably. Let $A$ be the property given in (2.1) that no node has precisely one child. We define state 1, that $A$ is true and state 2, that $A$ is false. We set $C = \{0, 1, \omega\}$ with $\omega$ meaning “at least two.” Let $n_1, n_2 \in C$ be the number of rootchildren $v$ with $T(v)$ having state 1,2 respectively. Then $T$ is in state 1 if and only if $\vec{n} = (n_1, n_2)$ has one of the values $(0, 0), (\omega, 0)$. Let $D$ be the set of distributions on the two states, $D = \{(x, y) : 0 \leq x, y \leq 1, y = 1 - x\}$. Then $\Psi_c(x, y) = (z, w)$ with $w = 1 - z$ and

$$
z = e^{-cx}e^{-cy} + e^{-cy}(1 - e^{-cx} - (cx)e^{-cx}) = e^{-cy}[1 - (cx)e^{-cx}]. \quad (3.11)
$$

The fixed point $(x, y)$ then has $x = Pr[A]$ satisfying the equation

$$
x = e^{-c(1-x)}[1 - (cx)e^{-cx}]. \quad (3.12)
$$

**Example 3.9.** Let $A$ be that there is a vertex $x$ with precisely one child who has precisely one child. Let state 1 be that $A$ is true. Let state 2 be that $A$ is false but that the root has precisely one child. Let state 3 be all else. Set $C = \{0, 1, \omega\}$. Set $D = \{x, y, z : x + y + z = 1, x \geq 0, y \geq 0, z \geq 0\}$. $T$ is in state 1 if and only if $\vec{n} = (n_1, n_2, n_3)$ has either $n_1 \neq 0$ or $n_1 = 0, n_2 = 1, n_3 = 0$. $T$ is in state 2 if and only if $\vec{n} = (0, 0, 1)$. Then $\vec{x}(c) = (x, y, z)$ must satisfy the system (noting $z = 1 - x - y$ is dependent)

$$
x = (1 - e^{-cx}) + e^{-cy}(cxye^{-cy}) - 1 - e^{-cx} + cye^{-c} \quad (3.13)
$$

$$
y = e^{-cx}e^{-cy}(cze^{-cz}) = cze^{-c} \quad (3.14)
$$

Here $x = Pr[A]$. In general, however, $Pr[A]$ will be the sum (3.3).

4. The Contraction Formulation

4.1. The Total Variation Metric. On $D$ we let $\rho(\vec{x}, \vec{y})$ denote the usual Euclidean metric, and $||\vec{x} - \vec{y}||_1$ the $L^1$ distance. We let $TV(\vec{x}, \vec{y})$ denote the total variation distance. With $\vec{x} = (x_1, \ldots, x_m)$ and $\vec{y} = (y_1, \ldots, y_m)$ this standard metric is given by

$$
TV(\vec{x}, \vec{y}) = \frac{1}{2}||\vec{x} - \vec{y}||_1 = \frac{1}{2} \sum_{j=1}^{m} |x_j - y_j|. \quad (4.1)
$$

Total variation distance between any two probability distributions $\mu, \nu$ on the same probability space has a natural interpretation in terms of coupling $\mu$ and $\nu$. Let $p = \sum x \mu(x) \land \nu(x)$. Flip a coin that has probability of heads $p$.

- If it lands heads, then choose a value $Z$ according to the probability distribution

$$
F_{min}(x) = \frac{\mu(x) \land \nu(x)}{p}, \quad (4.2)
$$

and set $X = Y = Z$. 

• If the coin lands tails, choose $X$ according to the probability distribution

$$F_1(x) = \begin{cases} \frac{\mu(x) - \nu(x)}{TV(\mu, \nu)} : & \text{if } \mu(x) > \nu(x); \\ 0 : & \text{otherwise.} \end{cases}$$

Independently choose $Y$ according to the probability distribution

$$F_2(x) = \begin{cases} \frac{\nu(x) - \mu(x)}{TV(\mu, \nu)} : & \text{if } \nu(x) > \mu(x); \\ 0 : & \text{otherwise.} \end{cases}$$

Then $X \sim \mu, Y \sim \nu$ and $(X, Y)$ are coupled in such a way that

$$TV(\mu, \nu) = P[X \neq Y].$$

(4.3)

We call such a coupling **optimal**.

As such, we have the general well-known result:

**Theorem 4.1.** For any two probability distributions $\mu, \nu$ on a common probability space,

$$TV(\mu, \nu) = \inf\{P[X \neq Y] : (X, Y) \text{ any coupling of } \mu, \nu\}.$$ 

(4.4)

We refer the reader to [3] for further reading.

### 4.2. The Contraction Theorem.

**Theorem 4.2.** For all $c > 0$ there exists a positive integer $s$ and an $\alpha < 1$ such that for all $\vec{x}, \vec{y} \in D$

$$\rho(\Psi^s(\vec{x}), \Psi^s(\vec{y})) \leq \alpha \rho(\vec{x}, \vec{y}).$$

(4.5)

The map $\Psi^s : D \to D$ has a natural interpretation. Let $\vec{x} = (x_1, \ldots, x_m) \in D$. Generate a random GW tree $T = T_c$ but stop at generation $s$. (The root is at generation 0.) To each node (there may not be any) $v$ at generation $s$ assign independently $j \in \Sigma$ from distribution $\vec{x}$. Now we work up the tree towards the root. Suppose, formally by induction, that all $v$ at generation $i$ have been assigned some $j \in \Sigma$. A $v$ at generation $i - 1$ will then have $n_j$ children assigned $j$ (allowing $n_j = \omega$). The value at $v$, which is now determined by Theorem 4.4, is given by the recursion function $\Gamma(\vec{n})$ of Definition 3.5. $\Psi^s(\vec{x})$ will then be the distribution of the Ehrenfeucht value assigned to the root.

**Remark 4.3.** The non-first order property $A$ that $T$ is infinite may be similarly examined. Set $C = \{0, \omega\}$ ($\omega$ denoting $\geq 1$) and let state 1 be that $T$ is infinite, state 2 that it is not. Then $D = \{(x, 1 - x) : 0 \leq x < 1\}$ and

$$\Psi_c(x, 1 - x) = (1 - e^{-cx}, e^{-cx}).$$

(4.6)

However, $\Psi_c$ has two fixed points: $(0, 1)$ and the “correct” $(x(c), 1 - x(c))$ when $c > 1$. The contraction property [4,5] will not hold. With $c$ small, $1 - e^{-cx} \sim cx$ and so $\vec{x} = (0, 1)$, and $\vec{y} = (c, 1 - c)$ become further apart on application of $\Psi_c$, when $c > 1$.

### 4.3. The Subcritical Case.** Here we prove Theorem 4.2 under the additional assumption that $c < 1$. The proof in this case is considerably simpler. Further, it may shed light on the general proof.

**Theorem 4.4.** For any $c < 1$ and any $\vec{x}, \vec{y} \in D$

$$TV(\Psi_c(\vec{x}), \Psi_c(\vec{y})) \leq c \cdot TV(\vec{x}, \vec{y}).$$

(4.7)

**Proof.** The main idea is to use suitable coupling of $\vec{x}$ and $\vec{y}$. First we fix $s \in \mathbb{N}$. We then create two pictures. In both pictures, let $v$ have $s$ many children $v_1, \ldots, v_s$. In picture 1, we assign, mutually independently, labels $X_1, \ldots, X_s \in \Sigma$ to $v_1, \ldots, v_s$ respectively, with $X_i \sim \vec{x}, 1 \leq i \leq s$. In picture 2, we assign, again mutually independently, labels $Z_1, \ldots, Z_s \in \Sigma$ to $v_1, \ldots, v_s$, with $Z_i \sim \vec{y}, 1 \leq i \leq s$. The pairs $(X_i, Z_i), 1 \leq i \leq s$ are mutually independent, but for every $i$, $(X_i, Z_i)$ is coupled so that

$$P[X_i \neq Z_i] = TV(\vec{x}, \vec{y}).$$

(4.8)
Suppose $X_v$ is the label of the root $v$ in picture 1 that we get from the recursion function $\Gamma$ (from (3.5)), and $Z_v$ that in picture 2. Then $X_v \sim \Psi_v(\vec{x})$, $Z_v \sim \Psi_v(\vec{y})$.

\[ TV(\Psi_v(\vec{x}), \Psi_v(\vec{y})) \leq P[X_v \neq Z_v] \]

\[ \leq \sum_{s=0}^{\infty} P[\text{Poisson}(c) = s] \sum_{i=1}^{s} P[X_i \neq Z_i] \]

\[ = \sum_{s=0}^{\infty} P[\text{Poisson}(c) = s] \cdot s \cdot TV(\vec{x}, \vec{y}) \]

\[ = \lambda \cdot TV(\vec{x}, \vec{y}). \]

\[ \square \]

**Theorem 4.5.** Theorem 4.2 holds when $c < 1$.

**Proof.** The inequalities

\[ |\vec{z}|_1 \geq |\vec{z}|_2 \geq \alpha^{-1/2} |\vec{z}|_1 \]

bound the $L^1$ and $L^2$ norms on $\mathbb{R}^m$ by multiples of each other. As $TV(\vec{x}, \vec{y}) = \frac{1}{2} |\vec{x} - \vec{y}|_1$,

\[ TV(\vec{x}, \vec{y}) \geq \frac{1}{2} \rho(\vec{x}, \vec{y}) \geq \alpha^{-1/2} \cdot TV(\vec{x}, \vec{y}). \]

Applying Theorem 4.4 repeatedly,

\[ TV(\Psi^c_\epsilon(\vec{x}), \Psi^c_\epsilon(\vec{y})) \leq c^s \cdot TV(\vec{x}, \vec{y}). \]

Combining (4.10) and (4.11)

\[ \rho(\Psi^c_\epsilon(\vec{x}), \Psi^c_\epsilon(\vec{y})) \leq 2 \cdot TV(\Psi^c_\epsilon(\vec{x}), \Psi^c_\epsilon(\vec{y})) \leq 2c^s \cdot TV(\vec{x}, \vec{y}) \leq 2c^s \sqrt{m} \rho(\vec{x}, \vec{y}). \]

We select $s$ so that $2c^s \sqrt{m} < 1$ and set $\alpha = 2c^s \sqrt{m}$. \[ \square \]

5. **Universality**

We define a function $Rad[i]$ on the nonnegative integers by the recursion

\[ Rad(0) = 0 \text{ and } Rad(i+1) = 3R(i) + 1 \text{ for } i \geq 0. \]

\[ (5.1) \]

**Definition 5.1.** In $T$ we define a distance $\rho(v, w)$ to be the minimal $r$ for which there is a sequence $v = z_0, z_1, \ldots, z_r = w$ where each $z_{i+1}$ is either the parent or a child of $z_i$. We set $\rho(v, v) = 0$.

As an example, cousins would be at distance four.

**Definition 5.2.** For $r$ a nonnegative integer, $v \in T$, the ball of radius $r$ around $v$, denoted $B(v, r)$ is the set of $w \in T$ with $\rho(v, w) \leq r$. We consider $v$ a designated vertex of $B(v, r)$.

We define an equivalence relation, depending on $k$, on such balls.

**Definition 5.3.** $B(v, r) \equiv_k B(v', r)$ if the two sets satisfy the same first order sentences of quantifier depth at most $k - 1$ with $v, v'$ as designated vertices, allowing $\pi, =,$ and $\rho$.

**Remark 5.4.** Note that the $(k-1)$-round Ehrenfeucht game with $v, v'$ designated is identical to the $k$-round Ehrenfeucht game in which the first round is mandated to select $v, v'$.

Equivalently, $B(v, r) \equiv_k B(v', r)$ if Duplicator wins the $k$-move Ehrenfeucht game on these sets in which the first round is mandated to be $v, v'$ and Duplicator must preserve $\pi, =,$ and $\rho$. The distance function $\rho$ could be replaced by the binary predicates $\rho_i(w_1, w_2) : \rho(w_1, w_2) = i, 0 \leq i \leq 2r$. As this is a finite number of predicates the number of equivalence classes is finite. Let $\Sigma^\text{BALL}_k$ denote the set of equivalence classes.

**Definition 5.5.** We say $S_1, S_2 \subset T$ are strongly disjoint if there are no $v_1 \in S_1, v_2 \in S_2$ with $\rho(v_1, v_2) \leq 1$.

**Definition 5.6.** We say $T$ is $k$-full if for any $v_1, \ldots, v_{k-1} \in T$ and any $\sigma \in \Sigma_k^\text{BALL}$ there exists a vertex $v$ such that
Let example, \( \sigma \).

Proof. Let \( B \) satisfy the condition of Definition 3.2. Suppose, as per Definition 5.3, \( B(R, \text{Rad}(k)) \equiv_{k+1} B(R', \text{Rad}(k)) \). Then \( T, T' \) have the same \( k \)-Ehrenfeucht value as given by Definition 3.2.

**Theorem 5.7.** Let \( T, T' \) with roots \( R, R' \) both be \( k \)-full. Suppose, as per Definition 5.3, \( B(R, \text{Rad}(k)) \equiv_{k+1} B(R', \text{Rad}(k)) \). Then \( T, T' \) have the same \( k \)-Ehrenfeucht value as given by Definition 3.2.

**Proof.** Let \( T, T' \) satisfy the condition of Theorem 5.7. We give a strategy for Duplicator to win the \( k \)-move Ehrenfeucht game. For convenience we add a move zero in which the roots \( R, R' \) are selected. Suppose \( i \) moves remain. Consider the union of the balls of radius \( \text{Rad}(i) \) about the chosen vertices (including the root) in each tree. These split into components. Duplicator shall insure that the corresponding chosen vertices in the same components and that the components are equivalent. At the start, with \( i = k \), this is true by assumption, because \( B(R, \text{Rad}(k)) \equiv_{k} B(R', \text{Rad}(k)) \). Suppose this holds with \( i \) moves remaining and Spoiler selects (the other case being symmetric) \( v \in T \). There are two cases.

**Inside:** \( v \) is at distance at most \( 2\text{Rad}(i-1) + 1 \) from a previously selected \( v_s \). Then its ball of radius \( \text{Rad}(i-1) \) lies entirely inside (from the recursion (5.1)) the ball of radius \( \text{Rad}(i) \) around \( v_s \). Duplicator then considers the equivalent component in \( T' \) and moves the corresponding \( v' \).

**Outside:** Now the ball of radius \( \text{Rad}(i-1) \) about \( v \) lies in a separate component from the balls of radius \( \text{Rad}(i-1) \) about the previously chosen vertices. As \( T' \) is \( k \)-full, Duplicator selects \( v' \in T' \) satisfying the conditions of Definition 5.7.

In either case Duplicator continues the property. At the end of the game there are zero moves left. The union of the balls of radius zero, the vertices selected, are equivalent in \( T, T' \) and Duplicator has won. □

**Definition 5.8. Christmas Tree:** We replace the complex notion of \( k \)-full by a simpler sufficient condition. For each \( \sigma \in \Sigma^B_{k} \) create \( k \) copies of a ball in that class. Take a root vertex \( v \) and on it place \( k \cdot |\Sigma^B_{k}| \) disjoint paths (parent to child) of length \( \text{Rad}(k) + \text{Rad}(k-1) + 1 \). Make each endpoint the top of one of these copies.

**Definition 5.9.** The \( k \)-universal tree, denoted \( \text{UNIV}_k \), is the Christmas Tree defined above.

**Definition 5.10.** \( T \) is called \( s \)-universal (given a fixed positive integer \( k \)) if all \( T' \) with \( |T'|_s \equiv |T'|_s \) have the same \( k \)-Ehrenfeucht value. Thus \( EV[T] \) is determined by \( T|_s \) completely.

**Theorem 5.11.** If for some \( v \), \( T(v) \equiv \text{UNIV}_k \) then \( T \) is \( k \)-full. Thus, by Theorem 5.7, the \( k \)-Ehrenfeucht value of \( T \) is determined by \( T|_{\text{Rad}(k)} \), or in other words, \( T \) is \( \text{Rad}(k) \)-universal.

**Remark 5.12.** Many other trees could be used in place of \( \text{UNIV}_k \), we use this particular one only for specificity.

**Remark 5.13.** A subtree \( T(v) \), where \( v \) is not the root, cannot determine the Ehrenfeucht value of \( T \) as, for example, it cannot tell us if the root has, say, precisely two children. Containing this universal tree \( \text{UNIV}_k \) tells us everything about the Ehrenfeucht value of \( T \) except properties relating to the local neighborhood of the root.

### 6. Rapidly Determined Properties

We consider the underlying probability space for the GW tree \( T = T_c \) to be an infinite sequence \( X_1, X_2, \ldots \) of independent variables, each Poisson with mean \( c \). These naturally create a tree. Let the root have \( X_1 \) children. Now we go through the nodes in a breadth first manner. Let the \( i \)-th node (the root is the first node) have \( X_i \) children. This creates a unique rooted tree. Note, however, that when the tree is finite with, say, \( n \) nodes, then the values \( X_j \) with \( j > n \) are irrelevant. In that case we say that the process aborts at time \( n \).

We employ a useful notation of Donald Knuth.
We say an event occurs quite surely if the probability that it does not occur drops exponentially in the given parameter.

Definition 6.2. Let $A$ be any property or function of rooted trees. We say that $A$ is rapidly determined if quite surely (in $s$, with $T = T_s$, and $c$ given) $X_1, \ldots, X_s$ tautologically determine $A$.

Remark 6.3. Consider the property that $T$ is infinite and suppose $c > 1$. Given $X_1, \ldots, X_s$ if the tree has stopped then we know it is finite. Suppose however (as holds with positive limiting probability) after $X_1, \ldots, X_s$ the tree is continuing. If at that stage there are many nodes we can be reasonably certain that $T$ will be infinite, but we cannot be tautologically sure. This property is not rapidly determined.

Remark 6.4. In this work we restrict the language in which $A$ is expressed. It has been suggested that another approach would be to restrict $A$ to rapidly determined properties.

Theorem 6.5. Let $T_0$ be an arbitrary finite tree. Let $A$ be the (non first order) property that either the process has aborted by time $s$ or there exists $v \in T$ with $T(v) \cong T_0$. Then $A$ is rapidly determined in parameter $s$.

The proof is given in [1]. Let $T_0$ have depth $d$. Roughly speaking, when we examine $X_1, \ldots, X_s$ either the process has aborted or it has not. If not, quite surely some $i \leq sc$ has $T(i) \cong T_0$. Here $c$ is chosen small enough (dependent on $c, d$) so that quite surely the children of all $i \leq sc$ down $d$ generations.

Theorem 6.6. Every first order property $A$ is rapidly determined.

Proof. Let $A$ have quantifier depth $k$. Let $T_0$ be the universal tree $UNIV_k$ as given by Theorem 5.11. From Theorem 6.5 if $T$ has not aborted by time $s$ then quite surely some $T(i) \cong T_0$. But then $T$ is already $k$-full and already has depth at least $Rad(k)$. By Theorem 5.7 the $k$-Ehrenfeucht value of $T$, hence the truth value of $A$, is determined solely by $T|_{Rad(k)}$, and hence tautologically by $X_1, \ldots, X_s$. □

Theorem 6.7. Fix a positive integer $k$. Let $T \sim T_c$. Then quite surely (in $s$), $T$ is $s$-universal.

Theorem 6.6 gives that the $k$-Ehrenfeucht value of $T$ is quite surely determined by $X_1, \ldots, X_s$. When this is so it is tautologically determined by $T|_s$, which has more information.

7. Unique Fixed Point

Theorem 7.1. The map $\Psi_c : D \to D$ has a unique fixed point.

Proof. Let $f(s)$ be the probability that $T_c$ is not $s$-universal. For any $\vec{y}, \vec{z} \in D$ we couple $\Psi_c^\vec{y}(\vec{y})$, $\Psi_c^\vec{z}(\vec{z})$. Create $T_c$ down to generation $s$ and then give each node at generation $s$ a $\sigma \in \Sigma$ with independent distribution $\vec{y}$, respectively $\vec{z}$. Then $\Psi_c^\vec{y} : \Psi_c^\vec{z}$ will be the induced state of the root. But when $T_c$ is $s$-universal this will be the same for any $\vec{y}, \vec{z}$. Hence $TV[\Psi_c^\vec{y}], \Psi_c^\vec{z}] \leq f(s)$. When $\vec{y}, \vec{z}$ are fixed points of $\Psi$, $TV[\vec{y}, \vec{z}] \leq f(s)$. As $f(s) \to 0$, $\vec{y} = \vec{z}$. □

Remark 7.2. Theorem 7.1 will also follow from the more powerful Theorem 6.2.

Remark 7.3. It is a challenging exercise to show directly that the solution $x$ to (3.12) or the solution $x, y$ to the system (3.13 and 3.14) are unique.

8. A Proof of Contraction

8.1. A Two Stage Process. Here we prove Theorem 1.2 for arbitrary $c$. Let $D_0$ be the depth of $UNIV_k$, as given by Definition 5.9. We shall set

$$s = s_0 + D_0$$

and think of $T|_s$ as being generated in two stages. In Stage 1 we generate $T|_{s_0}$. From Theorem 6.7 by taking $s_0$ large, this will be $s_0$-universal with probability near one. In Stage 2 we begin with an arbitrary but fixed $T_0$ of depth at most $s_0$. (We say “at most” because it includes the possibility that $T_0$ has no vertices at depth $s_0$.) From each node at depth $s_0$, mutually independently, we generate a GW-tree down to depth $D_0$. We denote by $Ext(T_0)$ this random tree, now of depth (at most) $s$. 

**Definition 8.1.** For any $T_0$ of depth at most $s_0$, $BAD[T_0]$ is the event that $Ext(T_0)$ is not $s_0$-universal.

**Theorem 8.2.** There exists positive $\beta$ such that for any $T_0$ of depth at most $s_0$

$$\Pr(BAD[T_0]) \leq e^{-t\beta} \tag{8.2}$$

where $t$ denotes the number of nodes of $T_0$ at depth $s_0$.

**Proof.** Let $v_1, \ldots, v_t$ denote the nodes of $T_0$ at generation $s_0$. Each of them independently generates a GW tree. Let $1 - e^{-\beta}$ denote the probability that $T(v_i) \cong UNIV_k$. With probability $e^{-t\beta}$ no $T(v_i) \cong UNIV_k$.

But otherwise $Ext(T_0)$ is $s_0$-universal. \hfill \Box

### 8.2. Splitting the Extension

**Definition 8.3.** $\Psi(T_0, \vec{x})$ denotes the induced distribution of the Ehrenfeucht value for the root $R$ as derived in the description above.

**Theorem 8.4.**

$$TV(\Psi(T_0, \vec{x}), \Psi(T_0, \vec{y})) \leq \sum \Pr(T|s_0 = T_0) \cdot TV(\Psi(T_0, \vec{x}), \Psi(T_0, \vec{y})) \tag{8.3}$$

where the sum is over all $T_0$ of depth (at most) $s_0$.

**Proof.** We split the distribution of $T$ into the distribution of $Ext(T_0)$, with probability $Pr[T|s_0 = T_0]$, over each $T_0$ of depth (at most) $s_0$. \hfill \Box

### 8.3. Some Technical Lemmas

**Lemma 8.5.** There exists $\beta > 0$ and $y_0$ such that for $y \geq y_0$

$$\Pr[Y \geq yt] \leq e^{-yt\beta}. \tag{8.4}$$

**Proof.** Set $f(\lambda) = \ln[E[e^{\lambda Y}]]$. We employ Chernoff bounds suboptimally, taking simply $\lambda = 1$. (We require here a standard argument that $E[e^{\lambda Y}]$ is finite.) Then $E[e^{\lambda Y}] = e^{f(1)}$ and

$$\Pr[Y \geq yt] \leq E[e^{\lambda Y}]e^{-yt} \leq e^{f(1) - yt}. \tag{8.5}$$

For $y \geq 2f(1)$, $f(1) - y \leq -y/2$ and we may take $\beta = \frac{1}{2}$.

**Lemma 8.6.** Let $K, \gamma > 0$. Let $BAD$ be an event with $Pr[BAD] \leq Ke^{-\gamma}$. Then, for a positive constants $k, \kappa$,

$$E[Y 1_{BAD}] \leq kTe^{-\kappa}, \tag{8.6}$$

where $Y$ is as defined above.

**Remark 8.7.** In the worst case the event $BAD$ would coincide with the top probability $Ke^{-\gamma}$ in the distribution of $Y$.

**Proof of Lemma 8.6.** We split $Y$ into $Y < y_1 t$ and $Y \geq y_1 t$, where $y_1$ needs to be chosen suitably.

$$E[Y 1_{BAD}] = E[Y 1_{BAD} 1_{Y < y_1 t}] + E[Y 1_{BAD} 1_{Y \geq y_1 t}], \tag{8.7}$$

where the first term is bounded by $y_1 tP[BAD] \leq Ky_1 t e^{-\gamma}$. The second term is bounded above by $E[Y 1_{Y \geq y_1 t}]$, which we use Chernoff type argument to bound. First, recall that $Y$ is the sum of $t$ i.i.d. copies of $X = X(c, s)$ which is the number of nodes at generation $s$ of $T_c$. Suppose $\varphi_x$ denotes the cumulant-generating function of $X$, defined as

$$\varphi_x(\lambda) = \log E[e^{\lambda X}]. \tag{8.8}$$

(Here $X$ is the number of children at generation $s$ of $T_c$.)
We fix some $\lambda > 1$ and choose
\[
y_1 = \max \left\{ y_0, \frac{\gamma + \varphi_x(\lambda)}{\lambda} \right\},
\] (8.9)
where $\gamma$ is as in the bound of $P[BAD]$. From (8.4) and (8.9), we then have
\[
E[Y \, 1_{Y \geq y_1}] \leq y_1 e^{-y_1 t} + \sum_{k=\lfloor y_1 t \rfloor + 1}^{\infty} P[Y \geq k]
\]
\[
\leq y_1 e^{-y_1 t} + e^{\lambda y_1 t} \int_{y_1 t}^{\infty} e^{-\lambda x} dx
\]
\[
= y_1 e^{-y_1 t} + e^{\varphi_x(\lambda)} t e^{-\lambda y_1 t}
\]
\[
\leq y_1 e^{-y_1 t} + \frac{1}{\lambda} e^{-\gamma t}.
\]
The desired bound now follows easily, by choosing $\lambda = \min \{y_1, \beta, \gamma\}$ and $k = Ky_1 + y_1 + 1$. \hfill $\square$

8.4. Bounding Expansion.

**Theorem 8.8.** There exists $K_0$ (dependent only on $s_0, k$) such that for any $T_0$ and any $\bar{x}, \bar{y} \in D$
\[
TV(\Psi_c^e(T_0, \bar{x}), \Psi_c^e(T_0, \bar{y})) \leq K_0 \cdot TV(\bar{x}, \bar{y}).
\] (8.10)

**Remark 8.9.** As $K_0$ may be large, Theorem 8.3, by itself, does not give a contracting mapping. It does limit how expanding $\Psi_c^e(T_0, \cdot)$ can be.

**Remark 8.10.** Let $t$ be the number of nodes of $T_0$ at depth $s_0$. Let $TV(\bar{x}, \bar{y}) = \epsilon$. The expected number of nodes in $Ext(T_0)$ at level $s = s_0 + D_0$ is then $tK_1$ with $K_1 = e^{\rho_0}$. The methods of Theorem 8.4 would then give Theorem 8.3 with $K = K_1t$. However, when $c > 1$ this $K$ would be unbounded in $t$. Our concern is then with large $t$ though, technically, the proof below works for all $t$.

**Proof.** Let $t$ be the number of nodes of $T_0$ at depth $s_0$. Let $TV(\bar{x}, \bar{y}) = \epsilon$. We again couple $\bar{x}, \bar{y}$. Let $Y$ be the number of nodes in $Ext(T_0)$ at level $s$. Given $Y = y$, let us name these vertices $u_1, \ldots, u_y$. Again we create two pictures. In picture 1, we assign, mutually independently, labels $X_i \in \Sigma$ to $u_i$, with $X_i \sim \bar{x}$, and in picture 2, label $Z_i \sim \bar{y}$. $(X_i, Z_i), 1 \leq i \leq y$ mutually independent, but $X_i, Z_i$ are coupled so that
\[
P[X_i \neq Z_i] = TV(\bar{x}, \bar{y}) = \epsilon.
\] (8.11)
The probability of the event that for at least one $i$, $X_i \neq Z_i$ is then bounded above by $y \cdot \epsilon$.

Suppose $X \in \Sigma$ is the label of the root of $Ext(T_0)$ in picture 1, and $Z$ that in picture 2, determined by using the recursion function $\Gamma$ repeatedly upwards starting at level $s$. Then $X \sim \Psi_c^e(T_0, \bar{x})$, $Z \sim \Psi_c^e(T_0, \bar{y})$.

Recall, from Definition 8.1, that $BAD[T_0]$ is the event that $Ext(T_0)$ is not $s_0$-universal. If $GOOD[T_0] = BAD[T_0]$, then under $GOOD[T_0]$, the Ehrenfeucht value of $Ext(T_0)$ is completely determined by $Ext(T_0)|s_0$, which is $T_0$ itself. And as $T_0$ is fixed, this means that $EV[Ext(T_0)]$ is then independent of $\bar{x}, \bar{y}$. Thus
\[
TV(\Psi_c^e(T_0, \bar{x}), \Psi_c^e(T_0, \bar{y})) \leq P[X \neq Z]
\]
\[
\leq \sum_{y=0}^{\infty} P[Y = y] \cdot y \cdot \epsilon \cdot 1_{BAD[T_0]}
\]
\[
= EV[1_{BAD[T_0]]}] \epsilon.
\]
From Theorem 8.2 and Lemma 8.6
\[
TV(\Psi_c^e(T_0, \bar{x}), \Psi_c^e(T_0, \bar{y})) \leq A(t) \epsilon \text{ with } A(t) = kte^{-\kappa}.
\] (8.12)
Here $A = A(t)$ approaches zero as $t \to \infty$ and so there exists $K_0$ such that $A \leq K_0$ for any choice of $t$. \hfill $\square$
8.5. **Proving Contraction.** We first show Theorem 4.2 in terms of the TV metric. Pick $s_0$ sufficiently large so that, say, the probability that $T_{|s_0}$ is not $s_0$-universal is at most $(2K_0)^{-1}$, $K_0$ given by Theorem 8.8.

This can be done because of Theorem 6.7. Let $\bar{x}, \bar{y} \in D$ with $\varepsilon = TV(\bar{x}, \bar{y})$. We bound $TV(\Psi^s(\bar{x}), \Psi^s(\bar{y}))$ by Theorem 8.4. Consider $TV(\Psi^s(T_0, \bar{x}), \Psi^s(T_0, \bar{y}))$. When $T_0$ is $s_0$-universal this has value zero. Otherwise its value is bounded by $K_0 \varepsilon$ by Theorem 8.8. Theorem 8.4 then gives

$$TV(\Psi^s(\bar{x}), \Psi^s(\bar{y})) \leq \frac{1}{2}K_0 \varepsilon \leq \frac{\varepsilon}{2}.$$  

(8.13)

Finally, we switch to the $L^2$ metric $\rho$. For $B$ a sufficiently large constant the inequaliteis yield, say,

$$\rho(\Psi^s_B(\bar{x}), \Psi^s_B(\bar{y})) \leq \frac{1}{2}\rho(\bar{x}, \bar{y}).$$  

(8.14)

Then Theorem 4.2 is satisfied with $s$ replaced by $sB$ and $\alpha = \frac{1}{2}$.

9. **Implicit Function**

Here we deduce Theorem 3.3 and hence Theorem 2.2 that $\Pr[A]$ is always a $C^\infty$ function of $c$. This follows from three results:

(i) The function $\Delta(c, \bar{x}) = \Psi^c(\bar{x})$ has all derivatives of all orders.

(ii) For each $c > 0$ the function

$$F(\bar{x}) = \Psi^c(\bar{x}) - \bar{x}$$  

(9.1)

has a unique zero $\bar{x} = \bar{x}(c)$.

(iii) The function $\Psi^c : D \rightarrow D$ is contracting in the sense of Theorem 4.2.

Let $A$ be the Jacobian of $\Psi^c$ at $\bar{x}(c)$. From Property (iii) all of the eigenvalues of $A$ lie inside the complex unit circle. Then $A = I$ is the Jacobian of $F$ from Property (ii). Then $(A - I)^{-1} = -\sum_{n=0}^\infty A^n$ is a convergent sequence, and so $A - I$ is invertible. As by Property (ii) the function $\Delta$ is smooth, the Implicit Function Theorem gives that the fixed point function $\bar{x}(c)$ of $F$ is $C^\infty$.

**References**

[1] M. Podder and J. Spencer, First Order Probabilities For Galton-Watson Trees, submitted for publication; Preprint [http://arxiv.org/abs/1510.08832](http://arxiv.org/abs/1510.08832).

[2] J. Spencer, The Strange Logic of Random Graphs, Springer Publishing Company, Inc. 2010; Series : Algorithms and Combinatorics, Vol. 22; ISBN:3642074995 9783642074998.

[3] D. A. Levin, Y. Peres and E. L. Wilmer, Markov Chains and Mixing Times, American Mathematical Society, 2006; ISBN: 978-0-8218-4739-8.