Self-Correction of Transmission on Regular Trees

March 1, 2007

[1]

Alberto Gandolfi* and Roberto Guenzani**

Abstract

We consider noisy binary channels on regular trees and introduce periodic enhancements consisting of locally self-correcting the signal in blocks without break of the symmetry of the model. We focus on the realistic class of within-descent self-correction realized by identifying all descendants k generations down a vertex with their majority. We show that this also allows reconstruction strictly beyond the critical distortion. We further identify the limit at which the critical distortions of within-descent k self-corrected transmission converge, which turns out to be the critical point for ferromagnetic Ising model on that tree. We finally discuss how similar phenomena take place with the biologically more plausible mechanism of eliminating signals which are locally not coherent with the majority.

1 Introduction

We consider a binary channel on a regular tree, as in [?], with a distortion rate $\varepsilon > 0$ at every transmission and are interested in the reconstruction of the starting bit σ_0 from the signals σ_{W_n} at the n-th generation of the tree. We focus on the majority rule, by which σ_0 is reconstructed as the symbol having majority in σ_{W_n}.

In [?] it is shown that for regular trees the majority rule is asymptotically equivalent to the optimal maximum-likelihood rule, and that there is a critical distortion $\bar{\varepsilon}_c = \frac{\sqrt{r}-1}{2\sqrt{r}}$ such that for $\varepsilon > \bar{\varepsilon}_c$ no asymptotic reconstruction takes

[1]**AMS 2000 subject classification:** Primary 60K35; Secondary 90B15, 92C15

Key words and phrases: tree, transmission, Ising model, extremality, correction, cells patterns.

Abbreviated title: Self-Correction of transmission

———————————————————————————

* Research partially supported by italian MIUR PRIN Grant # 2004015228

** Research partially supported by italian MIUR PRIN Grant # 2004015228

place and for $\varepsilon < \bar{\varepsilon}_c$ there is asymptotic reconstruction; see also [?] for a review and [?] for a dynamical version of these results.

The aim of this paper is to investigate how a non-symmetry breaking mechanism of correction performed while transmitting the signal can improve reconstruction by either majority or maximum likelihood. To this purpose, we propose a local self-correction method by which the signal is periodically enhanced in blocks formed within the generations. The enhancement uses majority rule and consists of taking all signals in a block and changing them to all agree with their majority value (with random choice to break tie). The self-correction is based on the information available at the level of interest, and thus can in principle be performed while the signal is transmitted. From every vertex the transmission is then continued as it used to be in the original mechanism and the symmetry of the model is not broken.

It is easy to see that with non-local enhancement one can reconstruct beyond the critical distortion: in fact, by forcing all vertices of each generation to agree with their majority, one can reconstruct for every $\varepsilon \in [0, \frac{1}{2})$. However, such correction involves taking majority on larger and larger blocks, which is not an implementable strategy.

A slightly less expensive self-correction strategy consists of using blocks of fixed size M (as soon as the generation is large enough) and then performing self-correction at every generation. In section 2 we show that for any noise level $\varepsilon < \frac{1}{2}$ it is possible to achieve reconstruction in this way with sufficiently large block size M. This procedure has the advantage of involving only a bounded number of within generation information exchange in self-correcting a block, and thus could in principle be implemented by a real machine. However, it still involves a very large number of within generation operations, performed at each generation: if the cost of each such operation is not zero (as in basically all reasonable situations) then the total cost might become too high.

We, therefore, restrict our attention, in the sequel, to a self-correction mechanism which contains costs by performing self-correction less often, and which has the additional advantage of being performed within the descent of some signal involved in the previous correction. This within descent self-correction reduces implementation costs, and allows signals to be dispersed and loose contact after their involvement in the enhancement, a feature which could be meaningful in a realistic setting. The within-descent self-correction at level k is performed by taking each vertex at some lk-th generation, $l \in \mathbb{N}$, considering its r^k descendants k generations down, and then changing them to agree with their majority (randomly breaking ties).

At first sight, it is not even obvious that such reconstruction improves upon the non self-corrected transmission, but in section 3 we show that, except for $k = 1$ and $r = 2$, the within-descent self-correction at level k strictly increases the critical distortions, and thus is an effective enhancement. The proof is based on the comparison between the self-correction based on the majority transformation with one correction based on random transformation which leaves the critical points unchanged.

The rest of the work is devoted to identifying the limit of the critical distortions of the within-descent self-correction of level k as k diverges. Although it might seem that such mechanism is almost useless for large k, it turns out that instead it improves the transmission further.

To identify the large k limit, in section 4 we exploit the correspondence with the Ising model. In fact, it is easy to see that, for regular trees, the reconstruction problem is equivalent to the free boundary conditions phase transition of the ferromagnetic Ising model on the tree with inverse temperature β such that $1 - 2\varepsilon = \tanh(\beta)$. Such transition occurs at the critical inverse temperature $\bar{\beta}_c$ such that for $\beta > \bar{\beta}_c$ the free boundary Ising model is convex combination of the extremal states (see [?] for a detailed description). On the other hand, the Ising model undergoes its regular phase transition (with boundary conditions) at a lower inverse temperature $\beta_c < \bar{\beta}_c$. In terms of $p = \tanh(\beta)$ and on a regular tree with forward branching rate r, we have $p_c = \tanh(\beta_c) = \frac{1}{r}$ (as shown originally in [?]) and $\bar{p}_c = \tanh(\bar{\beta}_c) = \frac{1}{\sqrt{r}}$ (as shown in [?, ?, ?]).

Our self-correction at level k introduces thus new critical values $1-2\varepsilon_c(k) = p_c(k) = \tanh(\beta_c(k)) < \bar{p}_c$ and our main result is a bound on $p_c(k)$ showing that $\lim_{k\to\infty} p_c(k) = p_c$, the regular Ising model phase transition point. Such estimate is derived by introducing the FK representation of the Ising model and then comparing the information carried by the FK tree of the origin against the external "noise" produced by all other freely fluctuating clusters of vertices. We think that this comparison, which is based on Gaussian approximation and large deviation techniques, has an interest in itself as it gives a very natural way of evaluating the information available on the tree.

In section 5 we remark that the majority self-correction is not biologically feasible, and introduce, instead, a minority removal self-correction which consists of self-correcting a generation by removing the elements not belonging to the majority. Since this leaves at least $r^k/2$ descendants, nothing really changes, and such correction also improves upon normal reconstruction up to the Ising model critical point. As we discuss, this, however, seems to indicate a peculiar phenomenon: it looks like that accepting the risk of creating uniform incorrect regions ("tumors") increases the resistance of inheritance to distortion. Whether this is a biologically meaningful statement should be further investigated with many bits models and realistic parameters.

There remain several open issues. First of all, our bounds on $p_c(k)$ in section 4 are not sharp. Also, our analysis has been performed either for correction each $k = 1$ steps using large block size M or for correction every k steps with $M = r^k$: we do not deal with the generic case of correcting blocks of size M each k generations. Solving the two issues above would then allow to treat the main open problem left by the present work: if one is to reconstruct the signal at a fixed generation n and if within generation transmission has some given cost, it would be natural to introduce a correspondence between within generation transmission costs and gain in reconstruction probability, and then look for the self-correction algorithm with optimal k and M.

2 Large Block Reconstruction

We consider regular trees $T^{(r)}$with forward branching rate $r > 0$. The n-th level of the tree is indicated by $T^{(r)}_n$ and $T^{(r)}_{\to n}$ represents the tree up to and including the n-th level. Vertices v of $T^{(r)}$ are then identified by coordinates $v = (n, s)$ where n is the level and $s = 1, ..., r^n$ numbers the vertices at the same level. Signals or configurations are variables $\{\sigma_v\}_{v\in T^{(r)}}$, $\sigma_v \in \{-1, 1\}$, and their distribution is specified by taking $\varepsilon > 0$, $P_\varepsilon(\sigma_0 = 1) = 1/2$ and for each vertex v and predecessor ${}^{\leftarrow}v$, $P_\varepsilon(\sigma_v = \sigma_{\leftarrow v}) = 1 - \varepsilon$ independently of all other pairs. Reconstruction under majority rule on $(T^{(r)}, P_\varepsilon)$ takes place if

$$\begin{aligned} 0 < \liminf_n \Delta_n(P_\varepsilon) \quad &=: \quad \liminf_n \left(P_\varepsilon(S_n > 0|\sigma_0 = 1) - P_\varepsilon(S_n < 0|\sigma_0 = 1)\right) \\ &= \quad \liminf_n E_\varepsilon|P_\varepsilon(\sigma_0 = 1|S_n) - P_\varepsilon(\sigma_0 = -1|S_n)| \end{aligned} \tag{1}$$

where $S_n = \sum_{v\in T^{(r)}_n} \sigma_v$.

We first consider self-correction performed at each step using large blocks. We fix an integer $M > 0$ and let $\tilde{n} = \max\{k : r^k \leq M\}$. We then consider the $\tilde{n}$-th generation as block 0, and partition each of the following generations into blocks of size M as follows: vertices $v = (n, s) \in T^{(r)}_n$ are partitioned into $\lfloor \frac{r^n}{M} \rfloor$ blocks of vertices with consecutive coordinates s, and possibly one block of $r^n - \lfloor \frac{r^n}{M} \rfloor$M vertices, which is from now on discarded without affecting the argument which follows. Each block B is then connected to all blocks B' such that there are two vertices $v \in B$ and $v' \in B'$ which are connected on $T^{(r)}$. One can easily see that considering blocks as renormalized vertices and connections between them as renormalized bonds we have a new tree $\bar{T}^{(r)}$ with forward branching r at all vertices $\bar{v} \in \bar{T}^{(r)}_n$, $n \geq 1$, and branching rate $r_0 \leq r$ at the starting vertex $\bar{v}_0$. The branching rate of $\bar{T}^{(r)}$ is thus again r.

Next, we consider self-corrected variables, which are required to be constant on blocks:

$$\Sigma_M = \{\sigma \in \{-1, 1\}^{T^{(r)}} \text{ such that } \sigma_v \text{ is constant on each block}\}, \tag{2}$$

and the self-correction map $\Phi_M : \{-1, 1\}^{T^{(r)}} \to \Sigma_M$ defined by

$$(\Phi_M\sigma)_v = \begin{cases} sign(\sum_{v\in B}\sigma_v) & \text{if } v\in B\subseteq T^{(r)}\setminus T^{(r)}_{\to(\tilde n-1)} \\ & \text{and } \sum_{v\in B}\sigma_v \neq 0 \\ Z & \text{if } v\in B\subseteq T^{(r)}\setminus T^{(r)}_{\to(\tilde n-1)} \\ & \text{and } \sum_{v\in B}\sigma_v = 0 \\ \sigma_v & \text{if } v\in T^{(r)}_{\to(\tilde n-1)}, \end{cases} \tag{3}$$

where $Z\in\{-1,1\}$ is a symmetric random variable.

The transmission is then self-corrected by the map Φ_M at every step: $\sigma_{T^{(r)}_{\to(n-1)}} \in \Sigma_M$ generates $\sigma_{T^{(r)}_{\to n}} \in \{-1,1\}^{T^{(r)}}$ as usual, and then we take $\Phi_M\left(\sigma_{T^{(r)}_{\to n}}\right)\in\Sigma_M$. The distribution $P_{\varepsilon,M}$ of the self-corrected configuration is then recursively defined by $P_{\varepsilon,M}\left(\sigma_{T^{(r)}_{\to n}}\middle|\sigma_{T^{(r)}_{\to(n-1)}}\right) = P_\varepsilon\left(\Phi_M^{-1}\sigma_{T^{(r)}_n}\middle|\sigma_{T^{(r)}_{\to(n-1)}}\right)$.

We then take configurations on the renormalized tree $\bar T^{(r)}$ to be $\bar\sigma_{\bar v}$ if $\bar v$ represents the block B and $(\Phi_M\sigma)_v = \bar\sigma_{\bar v}$ for all $v\in B$, and indicate by $\Psi_M : \Sigma_M \to \bar\Sigma_M$, with $\bar\Sigma_M = \{\bar\sigma_{\bar v}, \bar v\in \bar T^{(r)}\} = \{-1,1\}^{\bar T^{(r)}}$, the renormalizing transformation. Renormalized configurations are described by $\bar P_{\varepsilon,M} = \Psi_M\circ P_{\varepsilon,M}$ on (the Borel σ-algebra of) $\bar\Sigma_M$.

Our first result is that, no matter how large the noise level $\varepsilon\in[0,\frac12)$ is, with large enough block size M it is possible to reconstruct the starting signal σ_0 after performing the M-block self-correction at each step.

Theorem 2.1 $\forall\varepsilon\in[0,\frac12)$ $\exists\bar M : \forall M>\bar M$

$$\liminf_n \Delta_n(\bar P_{\varepsilon,M}) > 0.$$

Proof. We first calculate the error rate $\bar\varepsilon_M$ on the renormalized tree $\bar T^{(r)}$: let B be any block of size M of direct descendant of some site $v'\in B'$, where B is a descendant of B' in $\bar T^{(r)}$; then

$$\bar\varepsilon_M = P_\varepsilon(\sum_{v\in B}\sigma_v < 0|\sigma_{v'}=1) + \frac12 P_\varepsilon(\sum_{v\in B}\sigma_v = 0|\sigma_{v'}=1). \tag{4}$$

Given $\sigma_{v'}$, the σ_v's are $\{-1,1\}$-i.i.d. random variables with $P_\varepsilon(\sigma_v=1|\sigma_{v'}=1) = 1-\varepsilon>\frac12$, so that by large deviations theory there exists $c_\varepsilon>0$ such that $\bar\varepsilon_M \le e^{-c_\varepsilon M}$ for all $M>0$. Therefore, for M large enough,

$$(1-2\bar\varepsilon_M)^2 r \ge (1-2e^{-c_\varepsilon M})^2 r > 1. \tag{5}$$

This implies that $\bar{\varepsilon}_M < \varepsilon_c$ and there is reconstruction on the renormalized tree $\bar{T}^{(r)}$. By [**?**] this implies that for such M's:

$$\liminf_n \left(\bar{P}_{\varepsilon,M}(\bar{\sigma}_0 = 1 | \sum_{\bar{v}\in\bar{T}_n^{(r)}} \bar{\sigma}_{\bar{v}} > 0) - \bar{P}_{\varepsilon,M}(\bar{\sigma}_0 = -1 | \sum_{\bar{v}\in\bar{T}_n^{(r)}} \bar{\sigma}_{\bar{v}} > 0) \right) > 0. \tag{6}$$

Now, $\bar{\sigma}_0 = 1$ if $\sum_{v\in T_{\tilde{n}}^{(r)}} \sigma_v > 0$ or, with probability $\frac{1}{2}$, if $\sum_{v\in T_{\tilde{n}}^{(r)}} \sigma_v = 0$. Therefore,

$$\liminf_n \left(P_\varepsilon(\sum_{v\in T_{\tilde{n}}^{(r)}} \sigma_v > 0 | \sum_{\bar{v}\in\bar{T}_n^{(r)}} \bar{\sigma}_{\bar{v}} > 0) - P_\varepsilon(\sum_{v\in T_{\tilde{n}}^{(r)}} \sigma_v < 0 | \sum_{\bar{v}\in\bar{T}_n^{(r)}} \bar{\sigma}_{\bar{v}} > 0) \right) > 0. \tag{7}$$

We now show that by reading the block variables $\bar{\sigma}_{\bar{v}}$ for $\bar{v} \in \bar{T}_n^{(r)}$ one can reconstruct σ_0. To this purpose let

$$\begin{aligned} A &= \{\sigma_0 = +1\}, \\ B &= \{ \sum_{v\in T_{\tilde{n}}^{(r)}} \sigma_v > 0\} \end{aligned}$$

and

$$C = \{ \sum_{\bar{v}\in\bar{T}_n^{(r)}} \bar{\sigma}_{\bar{v}} > 0\}. \tag{8}$$

We then have, by total probabilities theorem, the Markov property and the fact that $P(A|B^c) = P(A^c|B)$ (with the same equality when A and A^c are exchanged),

$$\begin{gathered} P_\varepsilon(A|C) - P_\varepsilon(A^c|C) \\ = P_\varepsilon(A|C\cap B)P_\varepsilon(B|C) + P_\varepsilon(A|C\cap B^c)P_\varepsilon(B^c|C) \\ -(P_\varepsilon(A^c|C\cap B)P_\varepsilon(B|C) + P_\varepsilon(A^c|C\cap B^c)P_\varepsilon(B^c|C)) \\ = P_\varepsilon(A|B)P_\varepsilon(B|C) + P_\varepsilon(A|B^c)P_\varepsilon(B^c|C) \\ -(P_\varepsilon(A^c|B)P_\varepsilon(B|C) + P_\varepsilon(A^c|B^c)P_\varepsilon(B^c|C)) \\ = (P_\varepsilon(A|B) - P_\varepsilon(A^c|B))(P_\varepsilon(B|C) - P_\varepsilon(B^c|C)) > 0; \end{gathered} \tag{9}$$

the last inequality holds since it follows from (**??**) that if M is large enough, $\liminf_n (P_\varepsilon(B|C) - P_\varepsilon(B^c|C)) > 0$, and it follows from the next Lemma that $P_\varepsilon(A|B) - P_\varepsilon(A^c|B) > 0$ for every $\tilde{n}$. ■

Lemma 2.2 *Consider any tree $T^{(r)}$ and a transmission problem described by the distribution P_ε, let $S_n(\sigma) = S_n = \sum_{v\in T_n^{(r)}} \sigma_v$. Then*

i)

$$P_\varepsilon(S_{n-1} > 0|S_n > 0) - P_\varepsilon(S_{n-1} < 0|S_n > 0) > 0$$

ii)

$$P_\varepsilon(S_n > 0|D) - P_\varepsilon(S_n < 0|D) > 0$$

for every $D \subseteq \{-1,1\}^{T_{n-1}^{(r)}}$ such that $\forall\sigma \in D$, $S_{n-1}(\sigma) > 0$.

iii)

$$P_\varepsilon(S_n > 0|\hat\sigma_{T_{n-1}^{(r)}}) - P_\varepsilon(S_n < 0|\hat\sigma_{T_{n-1}^{(r)}}) > 0$$

for every configuration $\hat\sigma_{T_{n-1}^{(r)}} \in \{-1,1\}^{T_{n-1}^{(r)}}$ such that $\sum_{v\in T_{n-1}^{(r)}} \hat\sigma_v = l > 0$.

iv)

$$P_\varepsilon(S_{n-k} > 0|S_n > 0) - P_\varepsilon(S_{n-k} < 0|S_n > 0) > 0$$

for every $k = 1, ..., n$.

Proof. Clearly ii) implies i) taking $D = \{S_{n-1} > 0\}$, and iii) implies ii) since

$$\begin{aligned} &P_\varepsilon(S_n > 0|S_{n-1} > 0) \\ &\qquad = \sum_{\hat\sigma_{T_{n-1}^{(r)}} : \sum_{v\in T_{n-1}^{(r)}} \hat\sigma_v > 0} P_\varepsilon(S_n > 0|\hat\sigma_{T_{n-1}^{(r)}}) P_\varepsilon(\hat\sigma_{T_{n-1}^{(r)}}|S_{n-1} > 0) \end{aligned} \tag{10}$$

To show iii) assume $\sum_{v\in T_{n-1}^{(r)}} \hat\sigma_v = l > 0$. Then $S_n = \sum_{i=1}^{\frac{r^{n-1}-l}{2}} X_i + \sum_{i=\frac{r^{n-1}-l}{2}}^{r^{n-1}-l} Y_i + \sum_{i=r^{n-1}-l+1}^{r^{n-1}} X_i$ with X_i i.i.d, Y_i i.i.d, $X_i, Y_i \in \{-r, r\}$ and $X_i = \sum_{j=1}^r \tilde X_j$, $\tilde X_j$ i.i.d, $\tilde X_j \in \{-1,1\}$, $P(\tilde X_j = 1) = 1-\varepsilon$ and $Y_i = \sum_{j=1}^r \tilde Y_j$, $\tilde Y_j$ i.i.d, $Y_j \in \{-1,1\}$, $P(\tilde Y_j = 1) = \varepsilon$, all these variables being independent. So X_i is distributed like S_1 conditioned to $\sigma_0 = 1$ and, by symmetry of the distribution of S_1, $X_i =^d -Y_i$, so that

$$\bar S_n = \sum_{i=1}^{\frac{r^{n-1}-l}{2}} X_i + \sum_{i=\frac{r^{n-1}-l}{2}+1}^{r^{n-1}-l} Y_i \tag{11}$$

is a symmetric random variable. Therefore,

$$
\begin{aligned}
P_\varepsilon(S_n>0|\hat{\sigma}_{T^{(r)}_{n-1}}) &= P_\varepsilon(\bar{S}_n+\sum_{i=r^{n-1}-l+1}^{r^{n-1}} X_i>0)\\
&= \sum_{l_1=l-r^{n-1}}^{r^{n-1}-l} P_\varepsilon(\bar{S}_n+\sum_{i=r^{n-1}-l+1}^{r^{n-1}} X_i>0|\bar{S}_n=l_1)P_\varepsilon(\bar{S}_n=l_1)\\
&= \sum_{l_1>0}^{r^{n-1}-l}\Bigg[P_\varepsilon(\sum_{i=r^{n-1}-l+1}^{r^{n-1}} X_i>-l_1|\bar{S}_n=l_1)\\
&\quad +P_\varepsilon(\sum_{i=r^{n-1}-l+1}^{r^{n-1}} X_i>l_1|\bar{S}_n=-l_1)\Bigg]P_\varepsilon(\bar{S}_n=l_1)\\
&\quad + P_\varepsilon(\sum_{i=r^{n-1}-l+1}^{r^{n-1}} X_i>0|\bar{S}_n=0)P_\varepsilon(\bar{S}_n=0)\\
&= \sum_{l_1>0}^{r^{n-1}-l}\left[P_\varepsilon(\sum_{i=1}^{l} X_i>-l_1)+P_\varepsilon(\sum_{i=1}^{l} X_i>l_1)\right]P_\varepsilon(\bar{S}_n=l_1)\\
&\quad + P_\varepsilon(\sum_{i=1}^{l} X_i>0)P_\varepsilon(\bar{S}_n=0)
\end{aligned}
\tag{12}
$$

By the analogous expression for $S_n<0$ we then need

$$
\begin{aligned}
&P_\varepsilon(\sum_{i=1}^{l} X_i>-l_1)+P_\varepsilon(\sum_{i=1}^{l} X_i>l_1)\\
&\qquad > P_\varepsilon(\sum_{i=1}^{l} X_i<-l_1)+P_\varepsilon(\sum_{i=1}^{l} X_i<l_1)
\end{aligned}
\tag{13}
$$

For every $l\geq 1$ and $l_1\geq 0$, we have $\sum_{i=1}^{l} X_i=\sum_{j=1}^{rl}\tilde{X}_j$ and

$$
P_\varepsilon(\sum_{j=1}^{rl}\tilde{X}_j>l_1)=\sum_{h=\frac{rl+l_1}{2}}^{rl}\left(\begin{array}{c} rl\\ h\end{array}\right)(1-\varepsilon)^h\varepsilon^{rl-h}
\tag{14}
$$

Also, by the change of variable $rl-h'=h$,

$$
\begin{aligned}
P_\varepsilon(\sum_{j'=1}^{rl} \tilde{X}_{j'} < -l_1) &= \sum_{h'=0}^{\frac{rl-l_1}{2}} \left(\begin{array}{c} rl \\ h' \end{array} \right) (1-\varepsilon)^{h'} \varepsilon^{rl-h'} \\
&= \sum_{h=\frac{rl+l_1}{2}}^{rl} \left(\begin{array}{c} rl \\ h \end{array} \right) (1-\varepsilon)^{rl-h} \varepsilon^{h}
\end{aligned}
\tag{15}
$$

So that, for $\varepsilon < \frac{1}{2}$,

$$
\begin{aligned}
&P_\varepsilon(\sum_{i=1}^{l} X_i > l_1) - P_\varepsilon(\sum_{i=1}^{l} X_i < -l_1) \\
&\qquad \sum_{h=\frac{rl+l_1}{2}}^{rl} \left(\begin{array}{c} rl \\ h \end{array} \right) (1-\varepsilon)^{rl-h} \varepsilon^{rl-h} ((1-\varepsilon)^{2h-rl} - \varepsilon^{2h-rl}) > 0.
\end{aligned}
\tag{16}
$$

This shows (**??**) since we have seen one strict inequality between two terms, and the other two terms satisfy

$$
\begin{aligned}
&P_\varepsilon(\sum_{i=1}^{l} X_i > -l_1') - P_\varepsilon(\sum_{i=1}^{l} X_i < l_1') \\
&\qquad = P_\varepsilon(\sum_{i=1}^{l} X_i \geq l_1') - P_\varepsilon(\sum_{i=1}^{l} X_i \leq -l_1') > 0
\end{aligned}
\tag{17}
$$

for the same inequality (**??**) applied to $l_1 = l_1' - 1 \geq 0$.

Finally, (iv) is shown using iteratively (**??**) for k larger than one with

$$
\begin{aligned}
A &= \{S_{n-k} > 0\} \\
B &= \{S_{n-k+1} > 0\}
\end{aligned}
$$

and

$$
C = \{S_n \geq 0\}.
$$

■

3 Within-descent self-correction: strict inequality of critical points

Our aim is to consider within-descent self-correction at some level k. To this purpose we take a vertex v in some generation mk, $m \in \mathbb{N}$, and look at its r^k descendants k generation down (thus in $T^{(r)}_{(m+1)k}$) as generated by the transmission; we then force all such descendants to agree to their majority (with random choice if there is no majority). Transmission is then resumed as usual from the modified status. This amounts to define a map $\Phi_k : \{-1,1\}^{T^{(r)}} \to \Sigma_k$ given by

$$\Phi_k(\sigma)_v = \begin{cases} 1 \quad \text{with probability } 1 \quad \text{if} \quad \sum_{s_2'=1}^{r^k} \sigma_{mk, s_1 r^k + s_2'} > 0 \\ \\ -1 \quad \text{with probability } 1 \quad \text{if} \quad \sum_{s_2'=1}^{r^k} \sigma_{mk, s_1 r^k + s_2'} < 0 \\ \\ \begin{cases} 1 & \text{with probability } 1/2 \\ -1 & \text{with probability } 1/2 \end{cases} \quad \text{if} \quad \sum_{s_2'=1}^{r^k} \sigma_{mk, s_1 r^k + s_2'} = 0 \end{cases} \tag{18}$$

if $v \in T^{(r)}_{mk}$, with $v = (mk, s_1 r^k + s_2)$, $s_1 = 0, ..., r^{k(m-1)} - 1$, $s_2 = 1, ..., r^k$; otherwise

$$\Phi_k(\sigma)_v = \sigma_v. \tag{19}$$

As before, the transmission is self-corrected by the map Φ_k every k steps: $\sigma_{T^{(r)}_{\to mk}} \in \Sigma_k$ generates $\sigma_{T^{(r)}_{\to (m+1)k}} \in \{-1,1\}^{T^{(r)}}$ as usual, and then we take $\Phi_k(\sigma_{T^{(r)}_{\to (m+1)k}}) \in \Sigma_k$. The distribution $P^{(k)}_\varepsilon$ of the self-corrected configuration is then recursively defined by

$$P^{(k)}_\varepsilon(\sigma_{T^{(r)}_{\to (m+1)k}} | \sigma_{T^{(r)}_{\to mk}}) = P_\varepsilon(\Phi_k^{-1} \sigma_{T^{(r)}_{\to (m+1)k} \setminus T^{(r)}_{\to mr}} | \sigma_{T^{(r)}_{\to mk}}). \tag{20}$$

Notice that $P^{(k)}_\varepsilon$ is no longer a Markov chain but the conditional probabilities satisfy

$$\begin{aligned} P^{(k)}_\varepsilon(\sigma_{T^{(r)}_n} | \sigma_{T^{(r)}_{\to (n-1)}}) &= P_\varepsilon(\sigma_{T^{(r)}_n} | \sigma_{T^{(r)}_{\to (n-1)}}) \\ &= P_\varepsilon(\sigma_{T^{(r)}_n} | \sigma_{T^{(r)}_{(n-1)}}) \end{aligned} \tag{21}$$

for all n not of the form $n = mk$.

Next, for $\sigma \in \Sigma_k$, let $\Psi_k(\sigma) \in T^{(r^k)}$ be defined by

$$\Psi_k(\sigma)_v = \sigma_{(mk, s_1 r^k + 1)} \tag{22}$$

if $v \in T_m^{(r^k)}$, $v = (m, s_1 r^k + s_2)$, $s_1 = 0, ..., r^{k(m-1)} - 1$, $s_2 = 1, ..., r^k$. Note that $\Psi_k(\sigma)$ is a configuration of an almost regular tree $T^{(r^k)}$: $T^{(r^k)}$ has branching rate 1 at the starting vertex and then r^k at all other vertices. As we will see, the initial segment makes no difference in our arguments, and, therefore, we adopt the slight abuse of notation $T^{(r^k)}$ (which in our definitions indicates a regular tree).

Using $P_\varepsilon^{(k)}$ we define the self-corrected critical distortions

$$\varepsilon_{c,r}(k) = \sup\{\varepsilon : \liminf_n \Delta_n(P_\varepsilon^{(k)}) > 0\}. \tag{23}$$

Note that on $\Psi_k(\Phi_k(\{-1,1\}^{T^{(r)}})) = T^{(r^k)}$ the distribution $\Psi_k(P_\varepsilon^{(k)}) = P_\varepsilon^{(k)} \cdot \Psi_k^{-1}$ is a Markov chain, by the definition of $P_\varepsilon^{(k)}$, and thus it is again a transmission model with error rate $\varepsilon(k)$. In other words, $\Psi_k(P_\varepsilon^{(k)}) = P_{\varepsilon(k)}$ on $T^{(r^k)}$.

We first show that reconstruction under $\Psi_k(P_\varepsilon^{(k)})$ on $T^{(r^k)}$ is equivalent to reconstruction under the k-self corrected distribution $P_\varepsilon^{(k)}$.

Lemma 3.1 $\liminf_n \Delta_n(P_\varepsilon^{(k)}) > 0$ *if and only if* $\liminf_n \Delta_n(\Psi_k(P_\varepsilon^{(k)})) > 0$

Proof. First, observe that $\liminf_n \Delta_n(\Psi_k(P_\varepsilon^{(k)})) > 0$ on $\Psi_k(\Sigma_k)$ if and only if $\liminf_n \Delta_n(\Psi_k(P_\varepsilon^{(k)})) > 0$ on $T^{(r^k)}$. In fact, on $\Psi_k(\Sigma_k)$ we obtain

$$\begin{aligned} & P_{\varepsilon(k)}(S_n > 0 | \sigma_0 > 0) \\ &= \liminf_n \big[(1-\varepsilon(k))P_{\varepsilon(k)}(S_n > 0|\sigma_{(1,1)} > 0) \qquad (24) \\ &\quad + \varepsilon(k)P_{\varepsilon(k)}(S_n > 0|\sigma_{(1,1)} < 0)\big] \\ &= (1-2\varepsilon(k))\liminf_n P_{\varepsilon(k)}(S_n > 0|\sigma_{(1,1)} > 0) + \varepsilon(k) \end{aligned} \tag{25}$$

so that

$$\begin{aligned} & P_{\varepsilon(k)}(S_n > 0|\sigma_0 > 0) - P_{\varepsilon(k)}(S_n < 0|\sigma_0 > 0) \\ &= (1-2\varepsilon(k))\Delta_n(\Psi_k(P_\varepsilon^{(k)})); \end{aligned} \tag{26}$$

the $\liminf_n$ of the last expression is positive if and only if $\liminf_n \Delta_n(\Psi_k(P_\varepsilon^{(k)})) > 0$ on $T^{(r^k)}$ as $\varepsilon(k) < 1/2$. Now, observe that $\liminf_n \Delta_n(P_\varepsilon^{(k)}) > 0$ implies $\liminf_{mk} \Delta_{mk}(P_\varepsilon^{(k)}) > 0$, that is $\liminf_n \Delta_n(\Psi_k(P_\varepsilon^{(k)})) > 0$ on $\Psi_k(\{-1,1\}^{T(r)})$.

To show the reverse implication, notice that for every level n of $T^{(r^k)}$ not of the form $n = mk$ we have

$$\Delta_n(P_\varepsilon^{(k)}) = \Delta_{n-1}(P_\varepsilon^{(k)})(P_\varepsilon^{(k)}(S_n > 0|S_{n-1} > 0) - P_\varepsilon^{(k)}(S_n < 0|S_{n-1} > 0)) \tag{27}$$

where if $n-1 = mk$ then $S_{n-1} = \sum_{v \in T^{(r)}_{n-1}} (\Phi_k(\sigma))_v$. In all cases, the event $D = \{S_{n-1} > 0\}$ is such that $\hat{\sigma} \in D$ satisfies $\sum_{v \in T^{(r)}_{n-1}} \hat{\sigma}_v > 0$; this implies $P_\varepsilon^{(k)}(S_n > 0|S_{n-1} > 0) > P_\varepsilon^{(k)}(S_n < 0|S_{n-1} > 0)$ by part ii) of Lemma 2.2 applied to $P_\varepsilon^{(k)}$, since, by (**??**), the conditional probabilities coincide with those of P_ε.

Therefore, computing $\Delta_n(P_\varepsilon^{(k)})$ by finite iteration from the maximum level $mk < n$, $\liminf_{mk} \Delta_{mk}(P_\varepsilon^{(k)}) > 0$ implies $\liminf_n \Delta_n(P_\varepsilon^{(k)}) > 0$. ■

Our next aim is to show that $\varepsilon_{c,r}(k) > \bar{\varepsilon}_{c,r}$, which is to say $p_{c,r}(k) < \bar{p}_{c,r}$, where $\bar{\varepsilon}_{c,r}$ is the critical distortion rate for majority or maximum likelihood reconstruction on $T^{(r)}$.

In order to do this we introduce another random transformation, the fraction identification transform $\tilde{\Phi}_k : \{-1,1\}^{T^{(r)}} \to \{-1,1\}^{T^{(r)}}$ given by

$$\tilde{\Phi}_k(\sigma)_v = \sigma_{\bar{v}} \tag{28}$$

if $v \in T^{(r)}_{mk}$, with $v = (mk, s_1 r^k + s_2)$, $s_1 = 0, ..., r^{k(m-1)} - 1$, $s_2 = 1, ..., r^k$, and $\bar{v} = (mk, s_1 r^k + \bar{s}_2)$, $\bar{s}_2 = 1, ..., r^k$ uniformly chosen at random. Otherwise

$$\tilde{\Phi}_k(\sigma)_v = \sigma_v. \tag{29}$$

As before, for $\sigma \in \tilde{\Phi}_k(\{-1,1\}^{T^{(r)}})$, let $\tilde{\Psi}_k(\sigma) \in T^{(r^k)}$ be defined by

$$\tilde{\Psi}_k(\sigma)_v = \sigma_{(m, s_1 r^k)}. \tag{30}$$

Now, the strict inequality between the self-corrected critical distortion and the original one can be proven. The strict inequality holds for all values of k and r except for the one step correction on binary trees.

Theorem 3.2 *If* $k > 1$ *or* $k = 1$, $r > 2$

$$\varepsilon_{c,r}(k) \quad > \quad \bar{\varepsilon}_{c,r}; \tag{31}$$

$$\varepsilon_{c,2}(1) \quad = \quad \bar{\varepsilon}_{c,2} \tag{32}$$

To prove this fact, we explicitly compute the noise change under the fraction identification. On $\tilde{\Phi}_k(\{-1,1\}^{T(r)})$ the probability distribution $\tilde{P}_\varepsilon(k)$ which implements the fraction transform is defined as $P_\varepsilon^{(k)}$ with $\varepsilon(k)$ replaced by $\tilde{\varepsilon}(k) = 1 - \frac{1}{r^k}\sum_{s_2'=1}^{r^k}(2\sigma_{mk,s_1 r^k + s_2'} - 1)$. Note that $\Psi_k(\tilde{P}_\varepsilon^{(k)}) = P_{\tilde{\varepsilon}(k)}$ on $T^{(r^k)}$. We then have

Lemma 3.3 $\forall \varepsilon$, $\forall k$

$$1 - 2\tilde{\varepsilon}(k) = (1-2\varepsilon)^k \tag{33}$$

therefore the critical distortion $\tilde{\varepsilon}_{c,r}(k) = \sup\{\varepsilon : \liminf_n \Delta_n(\tilde{P}_\varepsilon^{(k)}) > 0\}$ *equals* ε_{c,r^k}.

Proof. Denote by X_k the number of 1's at level k. By definition and linearity of expected values,

$$\tilde{\varepsilon}_k(k) = 1 - \frac{1}{r^k} E_\varepsilon(X_k|\sigma_0 = 1) = 1 - P_\varepsilon(\sigma_{\bar{v}} = 1|\sigma_0 = 1) \tag{34}$$

for every $\bar{v} \in T_k^{(r)}$. The last probability refers to a one-dimensional Markov chain of length k with distortion probability ε, and can be easily computed. Alternatively, (**??**) can be verified by induction, since by the last equality, $1 - 2\tilde{\varepsilon}(1) = 1 - 2\varepsilon$ and

$$\tilde{\varepsilon}(k) = \varepsilon(1 - \tilde{\varepsilon}(k-1)) + (1-\varepsilon)\tilde{\varepsilon}(k-1), \tag{35}$$

so that

$$1 - 2\tilde{\varepsilon}(k) = (1-2\varepsilon)(1 - 2\tilde{\varepsilon}(k-1)) = (1-2\varepsilon)^k. \tag{36}$$

From [**?**], $(1-2\varepsilon_{c,r})^2 r = 1$ and since $\Psi_k(\tilde{P}_\varepsilon(k))$ is on $T^{(r^k)}$, on this second tree criticality is identified by $(1-2\varepsilon_{c,r^k})^2 r^k = 1$ and (**??**) implies $(1-2\tilde{\varepsilon}_{c,r}(k))^2 r^k = ((1-2\varepsilon_{c,r})^k)^2 r^k = ((1-2\varepsilon_{c,r})^2 r)^k = 1$. So $\tilde{\varepsilon}_{c,r}(k) = \varepsilon_{c,r^k}$. ∎

Proof of Theorem 3.2 Introduce

$$T_{k,r}(\varepsilon) = \frac{1}{r^k} \sum_{l=0}^{\frac{r^k-1}{2}} l\left(P_\varepsilon(X_k = l|\sigma_0 = 0) - P_\varepsilon(X_k = l|\sigma_0 = 1)\right) \tag{37}$$

when r is odd, and

$$T_{k,r}(\varepsilon) = \frac{1}{r^k} \sum_{l=0}^{\frac{r^k}{2}-1} l\left(P_\varepsilon(X_k = l|\sigma_0 = 0) - P_\varepsilon(X_k = l|\sigma_0 = 1)\right) \tag{38}$$

when r is even. For r odd, we have

$$
\begin{aligned}
T_{k,r}(\varepsilon) &= \frac{1}{r^k}\sum_{l=\frac{r^k+1}{2}}^{r^k}(r^k-l)P_\varepsilon(X_k=r^k-l|\sigma_0=0)-\frac{1}{r^k}\sum_{l=0}^{\frac{r^k-1}{2}} lP_\varepsilon(X_k=l|\sigma_0=1)\\
&= \tilde{\varepsilon}(k)-\varepsilon(k)
\end{aligned}
\tag{39}
$$

and, for r even

$$
\begin{aligned}
T_{k,r}(\varepsilon) &= \frac{1}{r^k}\sum_{l=\frac{r^k}{2}+1}^{r^k}(r^k-l)P_\varepsilon(X_k=r^k-l|\sigma_0=0)\\
&\quad -\frac{1}{r^k}\sum_{l=0}^{\frac{r^k-1}{2}} lP_\varepsilon(X_k=l|\sigma_0=1)+\frac{1}{2}P_\varepsilon(X_k=\frac{r^k}{2})-\frac{1}{2}P_\varepsilon(X_k=\frac{r^k}{2})\\
&= \tilde{\varepsilon}(k)-\varepsilon(k)
\end{aligned}
\tag{40}
$$

By Lemma 3.3 $\tilde{\varepsilon}_{c,r}(k)=\bar{\varepsilon}_{c,r^k}$ and $T_{1,2}(\bar{\varepsilon}_{c,2})=0$, so it is sufficient to show that $T_{k,r}(\bar{\varepsilon}_{c,r})>0$ for the non trivial cases of k and r. Theorem 1.4 in [**?**] shows that $P_\varepsilon(X_k=l|\sigma_0=0)\geq P_\varepsilon(X_k=l|\sigma_0=1)$ if $r^k-l>l$. To have strict inequality it is sufficient to show that $P_\varepsilon(X_k=1|\sigma_0=0)>P_\varepsilon(X_k=1|\sigma_0=1)$. This will be done by induction in k. We focus on the number i of distortions of σ_0 at the first step. The index i runs from 0 to r, but it is convenient to group together the i-th and the $(r-i)$-th terms. Note that $P_\varepsilon(X_1=i|\sigma_0=0)=\begin{pmatrix} r\\ i\end{pmatrix}\varepsilon^i(1-\varepsilon)^{r-i}$. Assuming $\bar{i}=\frac{r+1}{2}$ for r odd and $\bar{i}=\frac{r}{2}+1$ if r is even and $i\geq\bar{i}$, the terms in $T_{k,r}$ can be collected like this

$$
T_{k,r}(\varepsilon) = \sum_{i=\bar{i}}^{r}\begin{pmatrix} r\\ i\end{pmatrix}T_{k,r,i}(\varepsilon) \tag{41}
$$

with

$$
\begin{aligned}
T_{k,r,i}(\varepsilon) &= \left[\varepsilon^i(1-\varepsilon)^{r-i}-(1-\varepsilon)^i\varepsilon^{r-i}\right]\\
&\quad\cdot\big[iP_\varepsilon(X_{k-1}=1|\sigma_0=1)(P_\varepsilon(X_{k-1}=0|\sigma_0=1))^{i-1}\\
&\quad\cdot(P_\varepsilon(X_{k-1}=0|\sigma_0=0))^{r-i}+(r-i)(P_\varepsilon(X_{k-1}=0|\sigma_0=1))^{i}\\
&\quad\cdot P_\varepsilon(X_{k-1}=1|\sigma_0=0)(P_\varepsilon(X_{k-1}=0|\sigma_0=0))^{r-i-1}\\
&\quad -iP_\varepsilon(X_{k-1}=1|\sigma_0=0)(P_\varepsilon(X_{k-1}=0|\sigma_0=0))^{i-1}\\
&\quad\cdot(P_\varepsilon(X_{k-1}=0|\sigma_0=1))^{r-i}-(r-i)(P_\varepsilon(X_{k-1}=0|\sigma_0=0))^{i}\\
&\quad\cdot P_\varepsilon(X_{k-1}=1|\sigma_0=1)(P_\varepsilon(X_{k-1}=0|\sigma_0=1))^{r-i-1}\big]
\end{aligned}
$$

$$(42)$$

Now, the first factor is negative if $\varepsilon \in (0, 1/2)$ in particular if $\varepsilon = \bar{\varepsilon}_{c,r}$. We now show that the second factor is negative as well under the hypothesis that the statement is true for $k-1$.

The $(r-i)$ terms of the second addend are greater than or equal to $(r-i) \leq i$ terms taken from the third addend since

$$P_\varepsilon(X_{k-1} = 0|\sigma_0 = 0) \geq P_\varepsilon(X_{k-1} = 0|\sigma_0 = 1) \tag{43}$$

again by [?]. The remaining $(2i-r)$ terms from the third addend are strictly less than $(2i-r) \leq i$ terms taken from the first since

$$\begin{aligned} &P_\varepsilon(X_{k-1} = 1|\sigma_0 = 0)P_\varepsilon(X_{k-1} = 0|\sigma_0 = 0) \\ &\qquad > P_\varepsilon(X_{k-1} = 1|\sigma_0 = 1)P_\varepsilon(X_{k-1} = 0|\sigma_0 = 1); \end{aligned} \tag{44}$$

in fact, $P_\varepsilon(X_{k-1} = 0|\sigma_0 = 0) \geq P_\varepsilon(X_{k-1} = 0|\sigma_0 = 1)$ follows from [?], and $P_\varepsilon(X_{k-1} = 1|\sigma_0 = 0) > P_\varepsilon(X_{k-1} = 1|\sigma_0 = 1)$ follows by the induction hypothesis.

Finally, the remaining $(r-i)$ terms in the first addend are greater than or equal to the $(r-i)$ terms in the fourth addend again by [?].

For $r = 2$ and $k = 2$ the statement is true, as, by direct computation, we have, for some $f(\varepsilon)$,

$$\begin{aligned} &P_\varepsilon(X_2 = 1|\sigma_0 = 0) - P_\varepsilon(X_2 = 1|\sigma_0 = 1) \\ &\quad = 4(1-\varepsilon)^5\varepsilon + 2\varepsilon(1-\varepsilon)f(\varepsilon) + 4(1-\varepsilon)\varepsilon^5 - 8(1-\varepsilon)^3 - 2\varepsilon(1-\varepsilon)f(\varepsilon) \\ &\qquad\qquad = 4\varepsilon(1-\varepsilon)((1-\varepsilon)^2 + \varepsilon^2)^2 \end{aligned} \tag{45}$$

which is positive for $\varepsilon \in (0, 1/2)$. For $r > 2$ and $k = 1$ the statement is true as well in the same domain as $P_\varepsilon(X_1 = 1|\sigma_0 = 0) - P_\varepsilon(X_1 = 1|\sigma_0 = 1) = \varepsilon(1-\varepsilon)^{r-1} > (1-\varepsilon)\varepsilon^{r-1}$. ■

4 Limit of within-descent self-corrected critical distortions

The transmission model we are considering can equivalently be rewritten (see [?]) as an Ising model μ_β with inverse temperature β such that

$$\tanh(\beta) = 1 - 2\varepsilon \tag{46}$$

and

$$\mu_{\beta,\eta}(\sigma_{T^{(r)}_{\to n}}) = \frac{1}{Z} e^{-\beta \sum_{(\leftarrow v, v)} \sigma_{\leftarrow v} \sigma_v} \tag{47}$$

where μ_β is any weak limit of $\mu_{\beta,\eta}$. In turn, this can be represented as an FK model, see [?]. The usual FK parameter $p' = 1 - e^{2\beta}$ can then be modified on the tree, to account also for the number of clusters, to $p = \frac{p'}{2-p'} = \tanh(\beta) = 1 - 2\varepsilon$. With $H = \{0,1\}^{\mathcal{E}(T^{(r)})}$, where $\mathcal{E}(T^{(r)})$ are the length 1 edges of $T^{(r)}$ and $\eta \in H$, denoting by $\mathcal{E}(T^{(r)}_{\to n})$ the edges of $T^{(r)}_{\to n}$, we have

$$\nu_p(\eta_{\mathcal{E}(T^{(r)}_{\to n})}) = \prod_{\leftarrow v, v \in T^{(r)}_{\to n}} p^{\eta(\leftarrow v, v)} (1-p)^{1-\eta(\leftarrow v, v)}. \tag{48}$$

Therefore, the FK model is simply an independent Galton-Watson branching process with each descendant generated independently with probability p. The relation between ν_p and μ_β is the usual (see [?])

$$\mu_\beta(\sigma_{T^{(r)}_{\to n}}) = \sum_{\eta_{\mathcal{E}(T^{(r)}_{\to n})} \sim \sigma_{T^{(r)}_{\to n}}} \nu(\eta_{\mathcal{E}(T^{(r)}_{\to n})}) \frac{1}{Cl(\eta_{\mathcal{E}(T^{(r)}_{\to n})})} \tag{49}$$

where $\sim$ means that σ is compatible with η, i.e., $\sigma_{\leftarrow v} \sigma_v \eta_{(\leftarrow v, v)} \geq 0$, and $Cl(\eta_{\mathcal{E}(T^{(r)}_{\to n})})$ equals the number of σ's compatible with the given η, i.e. the number of site clusters determined by 1-edges in η.

In this section we want to show that $\varepsilon_{c,r}(k) \to \varepsilon_{c,r}$, i.e. $p_{c,r}(k) \to p_{c,r}$ and the main results will be

Theorem 4.1 There exist $c_1, c_2 > 0$ and a function $\alpha_k > 0$, $\lim_{k\to\infty} \alpha_k = 0$ such that

$$\frac{1}{r} \vee \frac{1}{c_1^{\frac{1}{2k}} r} \leq p_c(k) \leq \frac{1+\alpha_k}{c_2^{\frac{1}{2k}} r} \tag{50}$$

so that it easily follows

Corollary 4.2

$$\lim_{k\to\infty} p_c(k) = \frac{1}{r}.$$

The FK representation is thus a process in which each edge $e \in \mathcal{E}(T^{(r)})$ is open, i.e. $\eta_e = 1$, independently of all other edges, with probability p. The open edges are then just the (randomly selected) error fre edges in the transmission, in the sense that, given the configuration of the edges, the signal is generated by:

i) fixing the signal σ_0 at the origin;

ii) having the signal transmitted error free through the open edges;

iii) having the signal chosen at random with equal probability through the closed edges.

Seen globally, the set of vertices of $T^{(r)}$ falls apart into maximal connected components connected by open edges, and such components are called clusters. The cluster containing a vertex v is indicated by $C(v)$. Notice that $C(0)$ describes the descendants of a Galton-Watson process with offspring distribution Bernoulli of parameters r and p. The configuration of FK edges can also be described by some $\eta \in \{0,1\}^{\mathcal{E}(T^{(r)})}$.

As before, let $T_n^{(r)}$ be the vertices in the n-th generation of the tree. The vertices of $T_n^{(r)}$ also fall apart into "clusters" connected, via the entire tree, by open edges (these "clusters" are just the intersection of the clusters of $T^{(r)}$ with $T_n^{(r)}$). Given a configuration $\eta \in \{0,1\}^{\mathcal{E}(T^{(r)})}$ of open, i.e. value 1, FK edges, let $Z_i = Z_i(\eta)$, $i = 1, ..., m_n = m_n(\eta)$, be the clusters of $T_n^{(r)}$ in η, $1 \leq m_n \leq r^n$, and let $z_i = |Z_i|$.

Notice that $\Psi_k(\Phi_k(\sigma))$ is a configuration of $T^{(r^k)}$ and that on such tree there is reconstruction if the FK density $p(k) = p_{r^k}$ is such that $p_{r^k}^2 r^k > 1$ (see [**?**]).

On the other hand, by our construction, $p_{r^k} = 1 - 2P(S_k > 0|\sigma_0 = 1)$, so we need a lower bound for this expression. Such lower bound is achieved by estimating the size of $C(0) \cap T_k^{(r)}$, which is the set carrying information, and the value of $\sum_{i=1}^{m_k} Z_i'$, where Z_i' are independent symmetric random variables taking values in $\{-z_i, z_i\}$, i.e. distributed as the Z_i's. This last sum can be estimated via the normal approximation using Berry-Essen estimates of the error. This, however, involves second and third moments of Z_i, and we need to develop a somewhat elaborate bound on these moments since simple ones based on the maximum size of Z_i are not sufficient.

Such bounds on the sums of moments of Z_i's are determined in Theorems 4.2 and 4.3 below, as follows. First, notice that in creating the k-th generation roughly $(1-p)r^{k-1}$ vertices are isolated, thus giving rise to the same number of Z_i's taking values in $\{-1, 1\}$. Therefore, $\sum_{i=1}^{m_k} z_i^2 \geq cr^k$ for some $c > 0$ and our first two estimates show that this bound is nearly optimal. On the other hand, the largest cluster is of size roughly $(pr)^k$, so that $z_i^3 \simeq (pr)^{3k} = (p^2 r)^k (pr^2)^k \leq (1-c)^k (pr^2)^k$ if $p^2 r < 1$. Our last estimate shows that also this bound is nearly optimal. Note that this estimate cannot hold if $p^2 r \geq 1$, so that it provides no information about the reconstruction regime of the original tree.

We first need a large deviation result for the size of the set of vertices $R_n = C(0) \cap T_n^{(r)}$, i.e. for the survival set of the Galton-Watson process in the n- th generation. Let $P_p = P_\varepsilon$ for $p = 1 - 2\varepsilon$.

Lemma 4.3 *Let $\gamma = logr/log(pr) > 1$ and γ^* such that $1/\gamma + 1/\gamma^* = 1$ and let $W = \lim_{n\to\infty} \frac{|R_n|}{(pr)^n}$ (see [**?**]). Indicating by P the distribution of W and by E the expected value with respect to P, if $pr > 1$ then there exist $M, c_1, c_2, c_3 > 0$ such that if $\varepsilon > 0$ is such that $(1+\varepsilon)^{\gamma^*} < (pr)^{1/3}$ and $l \in \mathbb{N}$ is such that*

$\frac{((1+\varepsilon)/2)^\gamma}{\gamma^*(\gamma\tau)^{1/(\gamma-1)}} \le c_1(pr)^{1/3}$ *and* $(1+\varepsilon)^l/2 > M \vee 1$ *with* $\tau = \max_{x<pr} H(x) < \infty$ *and* $H(x) = x^{-\gamma}\log(B_r \cdot \Phi(x))$, $\Phi(s) = E(e^{sW})$ *and* B_r *the Bottcher's function (see [?]), then*

$$P_p(|R_l| \ge (1+\varepsilon)^l p^l r^l) \le c_2 e^{c_3(1+\varepsilon)^{\gamma^* l}} \tag{51}$$

for all $l \in \mathbb{N}$.

Proof. By large deviation properties of W, there exists $M > 0$ such that for all $x > M$

$$P(W \ge x) \le \exp\left(\frac{x^{\gamma^*}}{\gamma^*(\gamma\tau)^{1/(\gamma-1)}}\right) \tag{52}$$

for all x. Also, there exist $c_4, c_5 > 0$ such that

$$P\left(\left|\frac{|R_n|}{(pr)^n} - W\right| \ge 1\right) \le c_4 e^{c_5(pr)^{n/3}}, \tag{53}$$

for all n, see [?], Theorem 5; the conditions of that result are easily met by considering a process with the offspring of R_n plus one additional offspring in each vertex. Therefore, under the current assumptions, for some $c_2 \ge c_4 + 1$ and all $l \in \mathbb{N}$

$$\begin{aligned} P(|R_n| \ge (1+\varepsilon)^l p^l r^l) &\le P\left(\left|\frac{|R_n|}{(pr)^n} - W\right| \ge 1\right) + P(W \ge (1+\varepsilon)^l/2) \\ &\le c_4 e^{c_5(pr)^{l/3}} + \exp\left(\frac{((1+\varepsilon)^l/2)^{\gamma^*}}{\gamma^*(\gamma\tau)^{1/(\gamma-1)}}\right) \\ &\le c_2 e^{c_3(1+\varepsilon)^{\gamma^* l}} \end{aligned} \tag{54}$$

if $c_3 = \frac{1}{\gamma^*(2\gamma^*\gamma\tau)^{1/(\gamma-1)}}$. ■

Theorem 4.4 *$\forall p$ and r with P_p-probability one there exists a constant $c_7 = c_7(\eta) > 0$ such that*

$$\sum_{i=1}^{m_k} z_i^2 \ge c_7 r^k \tag{55}$$

for all k larger than some $\bar{k}_7(\eta)$.

Proof. $\sum_{i=1}^{m_k} z_i^2 \ge \sum_{C: C\cap T_k^{(r)} \ne \emptyset, |C|=1} |C|^2 = |\{C \subseteq T_k^{(r)} : |C| = 1\}| =: I_k$. For every $b = (^{\leftarrow}v, v)$, $^{\leftarrow}v \in T_{k-1}^{(r)}$, η_b is independently chosen to be 0 with probability $1-p$, and in such a case $C(v) = \{v\}$. So, by large deviations estimates for r^k i.i.d. binary random variables, if $c_7 = \frac{1-p}{2}$, $P(I_k \leq c_7 r^k) \leq e^{-c_3 \frac{(1-p)}{2} r^k}$ for some $c_3 > 0$ (see, for instance [?])

Therefore, $\sum_{k=1}^{\infty} P_p(\eta : I_k \leq c_7 r^k) \leq \sum_{k=1}^{\infty} e^{-c_3 \frac{(1-p)}{2} r^k} < \infty$ and by Borel-Cantelli the statement holds with P_p-probability 1 for large k with $c_7 = \frac{1-p}{2}$. ■

Theorem 4.5 *Suppose $p^2 r < 1$ and $pr > 1$. For every $\alpha > 0$ there exist $c_8 = c_8(\alpha) > 0$ and, with P_p-probability one, a finite $\bar{k}_8(\eta) > 0$ such that*

$$\sum_{i=1}^{m_k(\eta)} z_i^2(\eta) \leq c_8 (1+\alpha)^k r^k \tag{56}$$

for all $k \geq \bar{k}_8(\eta)$.

Proof. Let $\gamma = \frac{logr}{log(pr)} > 1$ and γ^* such that $\frac{1}{\gamma} + \frac{1}{\gamma^*} = 1$ and take ε_1 such that $(1+\varepsilon_1)^{\gamma^*} \leq (pr)^{1/3}$ and $(1+\varepsilon_1)^4 p^2 r < 1$. By Lemma 4.1, if $n \in \mathbb{N}$ and $V = V(n) \subseteq T^{(r)}$ is some set of vertices, then, since $(1+\varepsilon_1)^{\gamma^*} \leq (pr)^{1/3}$ we have

$$\begin{aligned} P_p(A_V(n)) &= P_p(\exists v \in V(n) : |C(v) \cap T_n^{(r)}| \geq (1+\varepsilon_1)^{n-|v|} (pr)^{n-|v|}) \\ &\leq \sum_{v \in V(n)} c_5 e^{-c_4 (1+\varepsilon_1)^{\gamma^*(n-|v|)}} \end{aligned} \tag{57}$$

Recursively define V_j and d_j as follows:

$$\begin{aligned} V_1 = V_1(n) &= \left\{ v \in T^{(r)} : |v| \leq d_1 n = n \frac{\log\left((1+\varepsilon_1)^4 p^2 r\right)^{-1}}{\log r} \right\}, \\ V_j = V_j(n) &= \left\{ v \in T^{(r)}, v \notin \bigcup_{j'=1}^{j-1} V_{j'} \right. \\ &\qquad \left. : |v| \leq d_j n = n \frac{\log\left((1+\varepsilon_1)^{4(1-d_{j-1})} p^{2(1-d_{j-1})} r^{1-2d_{j-1}}\right)^{-1}}{\log r} \right\} \end{aligned} \tag{58}$$

we then have

$$\begin{aligned} r^{d_1 n} &= \frac{1}{\left((1+\varepsilon_1)^4 p^2 r\right)^n}, \\ r^{d_j n} &= \frac{1}{\left((1+\varepsilon_1)^{4(1-d_{j-1})} p^{2(1-d_{j-1})} r^{1-2d_{j-1}}\right)^n}, \end{aligned}$$

$$(59)$$

$$
\begin{aligned}
P_p(A_{V_1}(n)) &\le \big((1+\varepsilon_1)^4 p^2 r\big)^{-n} c_5 e^{-\left(\frac{1}{2}\right)^{\gamma^*}(1+\varepsilon_1)^{\gamma^* n(1-d_1)}}, \\
P_p(A_{V_j}(n)) &\le \Big((1+\varepsilon_1)^{4(1-d_{j-1})} p^{2(1-d_{j-1})} r^{(1-2d_{j-1})}\Big)^{-n} \\
&\quad \cdot c_5 e^{-\left(\frac{1}{2}\right)^{\gamma^*}(1+\varepsilon_1)^{\gamma^* n(1-d_j)}}
\end{aligned}
\tag{60}
$$

On $A_{V_j}(n)^c$ we have

$$
\begin{aligned}
\sum_{v\in V_j} |C(v)\cap T_n^{(r)}|^2 &\le r^{d_j n}\left((1+\varepsilon_1)pr\right)^{2n(1-d_{j-1})} \\
&\le (1+\varepsilon_1)^{-2n(1-d_{j-1})} r^n.
\end{aligned}
\tag{61}
$$

Note that for $j = 2, 3, \ldots$

$$d_j = (1-d_{j-1})\frac{\log(1+\varepsilon_1)^4 p^2 r}{\log r} + d_{j-1} = (1-d_{j-1})d_1 + d_{j-1} \tag{62}$$

and that $d_1 \in (0,1)$ since $(1+\varepsilon_1)^4 p^2 r < 1$, so that $\lim_{j\to\infty} d_j = 1$.

On the other hand, for the given $\alpha > 0$ let ρ_1 be such that $r^{\rho_1} < 1+\alpha$; then, if for any cluster C we let $Base(C) = \min\{k : C \cap T_k^{(r)} \neq \emptyset\}$, we have

$$
\begin{aligned}
\sum_{C:Base(C)\ge(1-\rho_1)n} |C\cap T_n^{(r)}|^2 &\le \sum_C |C\cap T_n^{(r)}| \max_{C:Base(C)\ge(1-\rho_1)n} |C\cap T_n^{(r)}| \\
&\le |T_n^{(r)}| r^{\rho_1 n} \\
&\le (1+\alpha)^n r^n.
\end{aligned}
\tag{63}
$$

Next, take $J_1 \in \mathbb{N}$ such that $d_{J_1} \ge (1-\rho_1)$. Then

$$
\begin{aligned}
\sum_{n=1}^{\infty}\sum_{j=1}^{J_1} P_p(A_{V_j}(n)) &\le \sum_{j=1}^{J_1}\sum_{n=1}^{\infty} \Big((1+\varepsilon_1)^{4(1-d_{j-1})} p^{2(1-d_{j-1})} r^{(1-2d_{j-1})}\Big)^{-n} \\
&\quad \cdot c_5 e^{-c_6(1+\varepsilon_1)^{\gamma^* n(1-d_j)}} < +\infty
\end{aligned}
\tag{64}
$$

since for each j the series is of the form $A^n e^{-B^n}$, with $A > 1$ and $B > 0$, thus convergent. This implies that, by Borel-Cantelli, $A_{V_1}(n)\cup A_{V_2}(n)\cup\ldots\cup A_{V_{J_1},}(n)$

occurs only for a finite number of n's with probability one. Thus, for almost all η there exists $\bar{k}_8(\eta)$ such that for all $k > \bar{k}_8(\eta)$, $\bigcap_{j=1}^{J_1} A_{V_j}(k)^c$ occurs and this implies

$$\begin{aligned} \sum_{i=1}^{m_k(\eta)} z_i^2 &= \sum_C |C \cap T_k^{(r)}|^2 \\ &\le \sum_{C:Base(C)\ge(1-\rho_1)k} |C \cap T_k^{(r)}|^2 + \sum_{j=1}^{J_1} \sum_{C:Base(C)\in V_j} |C(v) \cap T_k^{(r)}|^2 \\ &\le (1+\alpha)^k r^k + (1+\varepsilon_1)^{-2k(1-d_{J_1})} r^k J_1 \\ &\le c_8(1+\alpha)^k r^k \end{aligned} \tag{65}$$

for a suitable $c_8 = c_8(J_1)$. ■

Theorem 4.6 *If $p^2 r < 1$ and $pr > 1$, then there exist $\bar{\alpha}' > 0$, $c_9 > 0$ and, with P_p-probability one, a finite $\bar{k}_9(\eta) > 0$ such that for every $\alpha' < \bar{\alpha}'$*

$$\sum_{i=1}^{m_k(\eta)} z_i^3 \le c_9(1-\alpha')^k (pr^2)^k \tag{66}$$

for all $k \ge \bar{k}_{39}(\eta)$.

Proof. We proceed as in the proof of Theorem 4.2 by taking ε_1, V_j, $A_{V_j}(n)$. On $A_{V_j}(n)^c$ we now have

$$\begin{aligned} \sum_{v\in V_j} |C(v) \cap T_n^{(r)}|^3 &\le r^{d_j n}((1+\varepsilon_1)pr)^{3n(1-d_{j-1})} \\ &\le (1+\varepsilon_1)^{-n(1-d_{j.-1})} p^{n(1-d_{j-1})} r^{(2-d_{j-1})n} \\ &\le \frac{1}{((1+\varepsilon_1)^{1-d_{j-1}}(pr)^{d_{j-1}})^n}(pr^2)^n \end{aligned} \tag{67}$$

with d_j's defined as above.

Now, take $\rho_2 > 0$ such that $\rho_2 < \frac{log(pr)}{4logr}$. Then

$$\begin{aligned} \sum_{C:Base(C)\ge(1-\rho_2)n} |C \cap T_n^{(r)}|^3 &\le \sum_C |C \cap T_n^{(r)}| \max_{C:Base(C)\ge(1-\rho_2)n} |C \cap T_n^{(r)}|^2 \\ &\le r^n r^{2\rho_2 n} \\ &\le r^n (pr)^n (pr)^{-n/2} \\ &\le (1-\alpha')^n (pr^2)^n \end{aligned}$$

$$(68)$$

provided that $1-\alpha' \geq \frac{1}{\sqrt{pr}}$.

Next, take $J_2 \in \mathbb{N}$ such that $d_{J_2} \geq 1-\rho_2$ and note that the Borel-Cantelli Lemma applies as above. Take α' also satisfying $1-\alpha' \geq (1+\varepsilon_1)^{-(1-d_{J_2})}$. Then, for $k \geq \bar{k}_9(\eta)$,

$$\begin{aligned}
\sum_{i=1}^{m_k(\eta)} z_i^3 &= \sum_C |C \cap T_k^{(r)}|^3 \\
&\leq \sum_{C:Base(C)\geq(1-\rho_2)k} |C \cap T_k^{(r)}|^3 + \sum_{j=1}^{J_2} \sum_{C:Base(C)\in V_j} |C(v) \cap T_k^{(r)}|^3 \\
&\leq (1-\alpha')^k (pr^2)^k + \frac{1}{(1+\varepsilon_1)^{(1-d_{J_2})k}} (pr^2)^k \\
&\leq c_9 (1-\alpha')^k (pr^2)^k.
\end{aligned} \qquad (69)$$

■

The next result gives the inequality for critical points $p_c(k)$.

For the lower bound we need

Lemma 4.8 *If Z_i's, $i = 1, ..., m$, are independent random variables each taking value in some $\{-l, l\}$, $l \in \mathbb{N}$ such that $Z_i \in \{-1, 1\}$ for all $i = 1, ..., I$ then for every $\alpha > 0$ and $m \geq I > 0$ we have*

$$P\left(\sum_{i=1}^{m} Z_i \in [-\alpha, \alpha]\right) \leq P\left(\sum_{i=1}^{I} Z_i \in [-\alpha, \alpha]\right) \qquad (70)$$

Proof. Since $p_k = P(\sum_{i=1}^{I} Z_i = k) = \binom{I}{(I+k)/2} 2^{-I}$, p_k increases up to $I/2$ and decreases afterwards; then, letting $S_k = \sum_{i=1}^{k} Z_i$, we have

$$
\begin{aligned}
P(S_m \in [-\alpha,\alpha]) &= P(S_I \in [-\alpha,\alpha], S_m \in [-\alpha,\alpha]) \\
&\quad + P(S_I \notin [-\alpha,\alpha], S_m \in [-\alpha,\alpha]) \\
&= P(S_I \in [-\alpha,\alpha], S_m \in [-\alpha,\alpha]) \\
&\quad + \sum_{t\notin[-\alpha,\alpha]} \sum_{l\in[-\alpha-t,\alpha-t]} P(S_I = t, S_{m-I} = l) \\
&\leq P(S_I \in [-\alpha,\alpha], S_m \in [-\alpha,\alpha]) \qquad (71)\\
&\quad + \sum_{t\notin[-\alpha,\alpha]} \sum_{l\in[-\alpha-t,\alpha-t]} P(S_I = t+l, S_{m-I} = -l) \\
&= P(S_I \in [-\alpha,\alpha], S_m \in [-\alpha,\alpha]) \\
&\quad + P(S_I \in [-\alpha,\alpha], S_m \notin [-\alpha,\alpha]) \\
&= P(S_I \in [-\alpha,\alpha])
\end{aligned}
$$

■

For the upper bound we need an estimate for the error rate $\varepsilon(k)$ at distance k, i.e. the value defined by

$$
1-\varepsilon(k) = P_p\left(\sum_{v\in T_k^{(r)}} \sigma_v > 0 \,|\sigma_0 = 1\right) + \frac{1}{2}P_p\left(\sum_{v\in T_k^{(r)}} \sigma_v = 0 \,|\sigma_0 = 1\right) \tag{72}
$$

Lemma 4.7 *If $p^2r < 1$ and $pr > 1$ then there exists $c_{10} > 0$ such that for every $\alpha > 0$ with probability one there exists $\bar{k}_{11}$ finite such that for all $k > \bar{k}_{10}$*

$$
1-\varepsilon(k) \geq \frac{1}{2} + \frac{1}{2}c_{10}\frac{(p\sqrt{r})^k}{(1+\alpha)^{k/2}} \tag{73}
$$

Proof. We have

$$
P_p\left(\sum_{v\in T_k^{(r)}} \sigma_v > 0 \,\middle|\, \sigma_0 = 1\right) + \frac{1}{2}P_p\left(\sum_{v\in T_k^{(r)}} \sigma_v = 0 \,\middle|\, \sigma_0 = 1\right)
$$

$$= \left[P_p \left(\sum_{v \in T_k^{(r)}} \sigma_v > 0 \middle| \sigma_0 = 1, |R_k| < \frac{(pr)^k}{2} \right) \right. \tag{74}$$

$$\left. + \frac{1}{2} P_p \left(\sum_{v \in T_k^{(r)}} \sigma_v = 0 \middle| \sigma_0 = 1, |R_k| < \frac{(pr)^k}{2} \right) \right] \cdot P_p \left(|R_k| < \frac{(pr)^k}{2} \middle| \sigma_0 = 1 \right)$$

$$+ \left[P_p \left(\sum_{v \in T_k^{(r)}} \sigma_v > 0 \middle| \sigma_0 = 1, |R_k| \geq \frac{(pr)^k}{2} \right) \right.$$

$$\left. + \frac{1}{2} P_p \left(\sum_{v \in T_k^{(r)}} \sigma_v = 0 \middle| \sigma_0 = 1, |R_k| \geq \frac{(pr)^k}{2} \right) \right] \cdot P_p \left(|R_k| \geq \frac{(pr)^k}{2} \middle| \sigma_0 = 1 \right)$$

Notice that for each $\eta \in \{-1,1\}^{\mathcal{E}(T^{(r)})}$, $\sum_{v \in T_k^{(r)}} \sigma_v = \sum_{i=1}^{m_k(\eta)} Z_i + |R_k|$, with Z_i symmetric random variables. Therefore,

$$P_p(\sum_{v \in T_k^{(r)}} \sigma_v > 0 | \sigma_0 = 1, |R_k| < \frac{(pr)^k}{2}) + \frac{1}{2} P_p(\sum_{v \in T_k^{(r)}} \sigma_v = 0 | \sigma_0 = 1, |R_k| < \frac{(pr)^k}{2})$$

$$\geq \sum_{\substack{\eta \in \{-1,1\}^{\mathcal{E}(T^{(r)})} \\ |R_k| \leq \frac{(pr)^k}{2}}} \left[P_p \left(\sum_{i=1}^{m_k(\eta)} Z_i > 0 \,|\eta \right) + \frac{1}{2} P_p \left(\sum_{i=1}^{m_k(\eta)} Z_i = 0 \,|\eta \right) \right] P_p(\eta) \geq \frac{1}{2}. \tag{75}$$

For the second part of (**??**) we use that

$$P_p(\sum_{v \in T_k^{(r)}} \sigma_v > 0 | \sigma_0 = 1, |R_k| \geq \frac{(pr)^k}{2}) + \frac{1}{2} P_p(\sum_{v \in T_k^{(r)}} \sigma_v = 0 | \sigma_0 = 1, |R_k| \geq \frac{(pr)^k}{2})$$

$$\geq \sum_{\eta : |R_k(\eta)| \geq \frac{(pr)^k}{2}} \left[P_p \left(\sum_{i=1}^{m_k(\eta)} Z_i > 0 \,|\eta \right) + \frac{1}{2} P_P \left(\sum_{i=1}^{m_k(\eta)} Z_i = 0 \,|\eta \right) \right.$$

$$\left. + P_p \left(\sum_{i=1}^{m_k(\eta)} Z_i \in \left(-\frac{(pr)^k}{2}, 0 \right] |\eta \right) \right] \frac{P_p(\eta)}{P_p(|R_k| \geq \frac{(pr)^k}{2})}$$

$$\geq \frac{1}{2} + \frac{1}{2} \sum_{\eta : |R_k(\eta)| \geq \frac{(pr)^k}{2}} P_p \left(\left| \sum_{i=1}^{m_k(\eta)} Z_i \right| < \frac{(pr)^k}{2} \,|\eta \right) \frac{P_p(\eta)}{P_p(|R_k| \geq \frac{(pr)^k}{2})} \tag{76}$$

Then

$$
\begin{aligned}
&P_p(\sum_{v\in T_k^{(r)}} \sigma_v > 0|\sigma_0 = 1) + \frac{1}{2}P_p(\sum_{v\in T_k^{(r)}} \sigma_v = 0|\sigma_0 = 1) \\
&\geq \frac{1}{2} + \frac{1}{2} \sum_{\eta:|R_k(\eta)|\geq \frac{(pr)^k}{2}} P_p\left(\left|\sum_{i=1}^{m_k(\eta)} Z_i\right| \leq \frac{(pr)^k}{2} |\eta\right) P_p(\eta) \qquad (77)
\end{aligned}
$$

Since the random variable W defined in Lemma 4.1 is absolutely continuous and $E(W) = 1$ (see [?]), then $P(W \geq \frac{1}{2}) > 0$. Moreover, $\frac{|R_k|}{(pr)^k}$ converges in distribution to W, so there exists a non random $\bar{k}_1$ such that for all $k \geq \bar{k}_1$

$$
P_p\left(\frac{|R_k|}{(pr)^k} \geq \frac{1}{2}\right) \geq \frac{1}{2}P(W \geq \frac{1}{2}) > 0. \qquad (78)
$$

We then want to estimate $P_p\left(\left|\sum_{i=1}^{m_k(\eta)} Z_i\right| \leq \frac{(pr)^k}{2} |\eta\right)$ via the Gaussian approximation using the Berry-Essen estimates of the error. To this extent, we will use the results in Theorems 4.2, 4.3 and 4.4 with α of Theorem 4.3 such that $(1+\alpha)^{-1/2} > 1-\alpha'$, with $\alpha' < \bar{\alpha}'$ and $\bar{\alpha}'$ determined as in Theorem 4.4. Such results hold with P_p-probability one for almost all η's, and thus it is possible to find a non random $\bar{k}_2$ such that $P_p(\eta : \bar{k}_2 \geq max(\bar{k}_7(\eta), \bar{k}_8(\eta), \bar{k}_9(\eta)) > 1 - \frac{1}{4}P(W \geq \frac{1}{2})$. Let $\bar{k}_3$ such that $\left(\frac{p^2r}{1+\alpha}\right)^k \frac{1}{4c_8(\alpha)} < -\log\frac{1}{2}$ and $\frac{1}{\sqrt{c_8}(1+\alpha)^k} \geq 2\frac{c_9}{c_7^{3/2}}(1-\alpha')^k$, for $k > \bar{k}_3$.

If we define the non random constant

$$
\bar{k}_{10} = max(\bar{k}_1, \bar{k}_2, \bar{k}_3) \qquad (79)
$$

and

$$
\begin{aligned}
M_k = \Bigg\{\eta \in \{-1,1\}^{\mathcal{E}(T^{(r)})} \Bigg| &\frac{|R_k(\eta)|}{(pr)^k} \geq \frac{1}{2}, c_7r^k \leq \sum_{i=1}^{m_k(\eta)} Z_i^2(\eta) \leq c_8(1+\alpha)^k r^k, \\
&\sum_{i=1}^{m_k(\eta)} |Z_i^3(\eta)| \leq c_9(1-\alpha')^k(pr^2)^k \Bigg\} \qquad (80)
\end{aligned}
$$

then, for $k \geq \bar{k}_{11}$

$$
P_p(M_k) \geq \frac{1}{4}P(W \geq \frac{1}{2}) > 0. \qquad (81)
$$

From (**??**) we then get

$$P_p(\sum_{v\in T_k^{(r)}} \sigma_v > 0|\sigma_0 = 1) + \frac{1}{2} P_p(\sum_{v\in T_k^{(r)}} \sigma_v = 0|\sigma_0 = 1)$$
$$\geq \frac{1}{2} + \frac{1}{2} \sum_{\eta\in M_k} P_p\left(\left|\sum_{i=1}^{m_k(\eta)} Z_i\right| \leq \frac{(pr)^k}{2} |\eta\right) P_p(\eta), \tag{82}$$

which we now estimate using the Gaussian approximation. Given η, the Z_i''s are independent random variables, so we can substitute them with the equally distributed Z_i''s. The Berry-Essen Theorem gives

$$\begin{aligned} P\left(\sum_{i=1}^{m_k(\eta)} Z_i' \in \left[-\frac{(pr)^k}{2}, \frac{(pr)^k}{2}\right]\right) &= P\left(\sum_{i=1}^{m_k(\eta)} \frac{Z_i}{\sqrt{V_k}} \in \left[\frac{-\frac{(pr)^k}{2}}{\sqrt{V_k}}, \frac{\frac{(pr)^k}{2}}{\sqrt{V_k}}\right]\right) \\ &= \int_{-\frac{\frac{(pr)^k}{2}}{\sqrt{V_k}}}^{\frac{\frac{(pr)^k}{2}}{\sqrt{V_k}}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx + E_k \end{aligned} \tag{83}$$

with $|E_k| \leq \frac{s_k}{V_k^{3/2}}$, where $V_k = \sum_{i=1}^{m_k(\eta)} Var(Z_i) = \sum_{i=1}^{m_k(\eta)} z_i^2$ and $s_k = \sum_{i=1}^{m_k(\eta)} E(|z_i|^3) = \sum_{i=1}^{m_k} z_i^3$.

If $\eta \in M_k$, $V_k \leq c_8(1+\alpha)^k r^k$ and

$$E_k \leq \frac{c_9(1-\alpha')^k (pr^2)^k}{(c_1 r^k)^{3/2}} = \frac{c_9}{c_7^{3/2}} (1-\alpha')^k p^k r^{k/2} \tag{84}$$

so that

$$\begin{aligned} P\left(\left|\sum_{i=1}^{m_k(\eta)} Z_i\right| \leq \frac{(pr)^k}{2}\right) &\geq \int_{-\frac{\frac{1}{2}(pr)^k}{\sqrt{c_8(1+\alpha)^k r^k}}}^{\frac{\frac{1}{2}(pr)^k}{\sqrt{c_8(1+\alpha)^k r^k}}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx - \frac{c_9}{c_7^{3/2}} (1-\alpha')^k p^k r^{k/2} \\ &\geq \frac{p^k r^{k/2}}{\sqrt{c_8}(1+\alpha)^{k/2}} e^{-\frac{p^{2k} r^k}{4c_7(1+\alpha)^k}} - \frac{c_9}{c_7^{3/2}} (1-\alpha')^k p^k r^{k/2} \\ &\geq \frac{1}{2} \frac{p^k r^{k/2}}{\sqrt{c_8}(1+\alpha)^{k/2}} \end{aligned} \tag{85}$$

for $k \geq \bar{k}_{10} \geq \bar{k}_3$. Together with (**??**), (**??**) this implies

$$P(\sum_{v\in T_k^{(r)}} \sigma_v > 0|\sigma_0 = 1) + \frac{1}{2} P(\sum_{v\in T_k^{(r)}} \sigma_v = 0|\sigma_0 = 1) \geq \frac{1}{2} + \frac{1}{2} c_{10} \frac{p^k r^{k/2}}{(1+\alpha)^{k/2}} \tag{86}$$

with $c_{10} = \frac{1}{\sqrt{c_8}} P_p(M_k) > 0$, for all $k \geq \bar{k}_{10}$. ■

Proof of Theorem 4.1 From Lemma 4.7, the probability of error free transmission $p(k) = 1 - 2\varepsilon(k)$ satisfies

$$p(k) \geq c_{11} \frac{(p\sqrt{r})^k}{(1+\alpha)^{k/2}} \tag{87}$$

for the binary transmission problem on $T^{(r^k)}$for k large enough. Therefore, there is reconstruction if

$$1 < p(k) r^{k/2} = c_{11} \left(\frac{pr}{\sqrt{1+\alpha}} \right)^k, \tag{88}$$

which is to say

$$p_c(k) \leq \frac{1+\alpha}{c_{11}^{1/k} r} \tag{89}$$

for k large enough. Let α_k be the smallest α s.t. (**??**) holds. Then $\lim_{k\to\infty} \alpha_k = 0$ as required to prove the upper bound of Theorem 4.1.

Similarly to (**??**) we estimate, for $\beta > 0$,

$$
\begin{aligned}
1-\varepsilon(k) \quad = \quad & \left[P_p\left(\sum_{v\in T_k^{(r)}}\sigma_v>0\,\middle|\,\sigma_0=1,|R_k|>(1+\beta)^k(pr)^k\right)\right.\\
&\left.+\frac{1}{2}P_p\left(\sum_{v\in T_k^{(r)}}\sigma_v=0\,\middle|\,\sigma_0=1,|R_k|>(1+\beta)^k(pr)^k\right)\right]\\
&\cdot P_p(|R_k|>(1+\beta)^k(pr)^k|\sigma_0=1)\\
&+\left[P_p\left(\sum_{v\in T_k^{(r)}}\sigma_v>0\,\middle|\,\sigma_0=1,|R_k|\le(1+\beta)^k(pr)^k,\left|\sum_{i=1}^{m_k(\eta)}Z_i\right|\le(1+\beta)^k(pr)^k\right)\right.\\
&\left.+\frac{1}{2}P_p\left(\sum_{v\in T_k^{(r)}}\sigma_v=0\,\middle|\,\sigma_0=1,|R_k|\le(1+\beta)^k(pr)^k,\left|\sum_{i=1}^{m_k(\eta)}Z_i\right|\le(1+\beta)^k(pr)^k\right)\right]\\
&\cdot P_p(|R_k|\le(1+\beta)^k(pr)^k,\left|\sum_{i=1}^{m_k(\eta)}Z_i\right|\le(1+\beta)^k(pr)^k|\sigma_0=1)\\
&+\left[P_p\left(\sum_{v\in T_k^{(r)}}\sigma_v>0\,\middle|\,\sigma_0=1,|R_k|\le(1+\beta)^k(pr)^k,\left|\sum_{i=1}^{m_k(\eta)}Z_i\right|>(1+\beta)^k(pr)^k\right)\right.\\
&\left.+\frac{1}{2}P_p\left(\sum_{v\in T_k^{(r)}}\sigma_v=0\,\middle|\,\sigma_0=1,|R_k|\le(1+\beta)^k(pr)^k,\left|\sum_{i=1}^{m_k(\eta)}Z_i\right|>(1+\beta)^k(pr)^k\right)\right]\\
&\cdot P_p(|R_k|\le(1+\beta)^k(pr)^k,\left|\sum_{i=1}^{m_k(\eta)}Z_i\right|>(1+\beta)^k(pr)^k|\sigma_0=1).
\end{aligned}
\tag{90}
$$

From Lemma 4.3

$$P_p(|R_k|>(1+\beta)^k(pr)^k|\sigma_0=1)\le c_5e^{c_6(1+\beta)^{\gamma^* k}}.$$

In the third term, the expression between square brackets is exactly $\frac{1}{2}$ since Z_i's are independent and symmetric.

Next we consider the second term. Assume first $pr\ge 1$. Let I be the set of vertices in $T_k^{(r)}$ which are isolated FK clusters. Then, by large deviations for i.i.d. random variables, $P_p(|I|<\frac{1}{2}\left(\frac{1-p}{r}\right)r^k\le e^{-cr^k}$. Moreover, from Lemma 4.8, the expression between square brackets in the second term of (**??**) is bounded above by $P_p\left(\left|\sum_{i=1}^{I}Z_i'\right|\le((1+\beta)pr)^k\right)$, with Z_i' i.i.d. symmetric random variables with values in $\{-1,1\}$. In turn, if $|I| \geq (1-p-\varepsilon)r^k \geq \frac{r^k}{2}$ the normal approximation implies that for some $c_{12}>0$, $c_{13}>0$,

$$\begin{aligned} P_p\left(\left|\sum_{i=1}^{I} Z_i'\right| \leq ((1+\beta)pr)^k\right) &\leq c_{12}((1+\beta)p\sqrt{r})^k + c_{13}\frac{1}{\sqrt{I}} \\ &\leq c_1 p^k r^{k/2}(1+\beta)^k \end{aligned} \tag{91}$$

for a suitable c_1 large enough, where the last term comes from the Berry-Essen error estimate for the random variables Z_i', with $|I| \geq \frac{r^k}{2}$ and $\frac{1}{r^{k/2}} \leq p^k r^{k/2}$.

Collecting the above estimates we have

$$\begin{aligned} 1-\varepsilon(k) &\leq e^{-cr^k} + c_{12}((1+\beta)p\sqrt{r})^k + c_{13}p^k r^{k/2}(1-\alpha')^k + \frac{1}{2} \\ &\leq \frac{1}{2} + \frac{1}{2}c_1 p^k r^{k/2}(1+\beta)^k. \end{aligned} \tag{92}$$

Therefore,

$$p(k) \leq c_1 p^k r^{k/2}(1+\beta)^k \tag{93}$$

and the condition for non-reconstruction on the rescaled tree $T^{(r^k)}$ becomes

$$c_1 p^k r^k (1+\beta)^k < 1. \tag{94}$$

This implies

$$p_c(k) \geq \frac{1}{(1+\beta)c_1^{1/k} r} \geq \frac{1}{c_1^{1/k} r}. \tag{95}$$

If, on the other hand, $pr<1$, then for small enough β, $(1+\beta)pr<1$ and the second term in square brackets of (**??**) reduces to $\frac{1}{2}P_p\left(\sum_{v\in T_k^{(r)}} \sigma_v = 0 \middle| \sigma_0 = 1\right)$, but clearly in this case the symmetry is not broken and no reconstruction can take place. ■

From Theorem 4.1 it is obvious that the critical points $p_c(k)$ converge to the Ising model critical point.

5 Minority removal

The self-correction mechanism discussed above is not suitable for biological transmission, in which offsprings, once generated, cannot be changed. However, there is a similar mechanism, which consists of self-correcting a generation by removing the elements not belonging to the majority, which could be implemented in a biological setting. If $r \geq 4$ and such minority removal is carried out every step in blocks of size M, then in the renormalized tree each (macroscopic) vertex has a random number of children larger then or equal to 2, while the error rate is estimated as in (**??**) but on a random number of vertices, between $\frac{M}{2}$ and M; by taking inequalities as done below, one can see that (**??**) still holds with minor changes and thus reconstruction is also possible at every $\varepsilon < \frac{1}{2}$ with a sufficiently large M. It is also the case that if a within-descent minority removal is carried out every k generations, only minor changes in the constants are needed in Theorem 4.1 and the limit of the critical points is still the Ising critical point as in Corollary 4.2.

This highlights a possibly real but rather particular phenomenon. It looks like a bit of information in the parent biological unit is better transmitted, i.e. it is more resistant to random transmission errors, if enhanced by regularly destroying descendants not belonging to the local majority. From the biological point of view this is also likely to improve the functionality of local segments (cells or individuals, for instance). However, the minority removal sometimes preserves the wrong information, thus creating blocks of mutated descendants, a phenomenon similar to tumor formation. In this respect, our findings seem to suggest that tumor generation might be intrinsically connected to improvement in character transmission. Of course, any such claim must be warranted by the study of many bits transmission.

Back to our single bit model, the minority-removal carried out every step by blocks of size M corresponds to first generating a random tree T'_M by means of a transformation Φ'_M analogous to Φ_M and then identifying each block (of random size between $\frac{M}{2}$ and M) by means of a transformation Ψ'_M, analogous to Ψ_M. Let $\bar{P}'_{\varepsilon,M} = \Psi'_M(\Phi'_M(P_\varepsilon))$ be the distribution on the resulting random tree T'_M.

Similarly, the within-descent minority removal carried out every k-steps corresponds to generating a random tree T'_k by means of a transformation Φ'_k, analogous to Φ_k, and then identifying each block (of random size between $\frac{r^k}{2}$ and r^k) by means of a transformation Ψ'_k, analogous to Ψ_k. Let $P'^{(k)}_\varepsilon = \Psi'_k(\Phi'_k(P_\varepsilon))$ be the distribution on the resulting random tree T'_k.

Note that T'_M and T'_k are Galton-Watson trees, since they are random trees with an i.i.d. number of offsprings in each vertex. In generating T'_M at least $M/2$ vertices are preserved in each block of size M; these have at least $rM/2 \geq 2M$ descendants which can be divided into at least 2 blocks of size M (and possibly one remaining smaller block). Thus the number of descendants is at least 2. In generating T'_r on the other hand, at least $r^k/2$ vertices are preserved in each block of size r^k and each such vertex gives rise to one descendant block, so each block (which is a renormalized vertex) has at least $r^k/2$ (and at most r^k) descendants.

The branching numbers, which on the Galton-Watson trees equal the mean offspring number (see [?]), satisfy then $br(T'_M) \geq 2$ and $r^k/2 \leq br(T'_r) \leq r^k$.

We begin with a Lemma stating that if on a subtree $T' \subset T$ maximum likelihood reconstruction takes place, then it does also on T.

Lemma 5.1 *Given trees $T' \subseteq T$, if maximum likelihood reconstruction takes place on T' then it does also on T, i.e. if* $\liminf_n \Delta_n(P_{T'}) > 0$ *then* $\liminf_n \Delta_n(P_T) > 0$.

Proof. Let $A_n = \{\sigma_n \in T_n : P(\sigma_n|\sigma_0 = +1) > P(\sigma_n|\sigma_0 = -1)\}$, let A'_n be the same with T_n replaced by T'_n and let $B' = \{\sigma'_n \in T'_n : P(\sigma'_n|\sigma_0 = +1) = P(\sigma'_n|\sigma_0 = -1)\}$. We know $P(A'_n|\sigma_0 = +1) - P(A'_n|\sigma_0 = -1) \geq \delta > 0$ for some δ for large n, and we want to show the same for A_n. However, denoting by $P^{\pm}(\cdot) = P(\cdot|\pm 1)$ we have $P^{\pm}(A_n \cap (A'_n)^c) = P^{\mp}(A_n^c \cap A'_n)$ by symmetry, and for any event C, by definition of A_n,

$$\begin{aligned} P^+(A_n \cap C) &\geq P^-(A_n \cap C) \\ P^+(A_n^c \cap C) &\leq P^-(A_n^c \cap C). \end{aligned} \tag{96}$$

Then,

$$\begin{gathered} P^+(A_n) - P^-(A_n) \\ = P^+(A_n \cap A'_n) + P^+(A_n \cap (A'_n)^c) + P^+(A_n \cap B') \\ -P^-(A_n \cap A'_n) - P^-(A_n \cap (A'_n)^c) - P^-(A_n \cap B') \\ = P^+(A_n \cap A'_n) + P^-(A_n^c \cap A'_n) + P^+(A_n \cap B') \\ -P^-(A_n \cap A'_n) - P^+(A_n^c \cap A'_n) - P^-(A_n \cap B') \\ \geq P^+(A_n \cap A'_n) + P^+(A_n^c \cap A'_n) \\ -P^-(A_n \cap A'_n) - P^-(A_n^c \cap A'_n) \\ = P^+(A'_n) - P^-(A'_n) \end{gathered} \tag{97}$$

from which the result follows. ■

The results for minority removal can be summarized as follows. Notice that in the proof we use maximum likelihood reconstruction to use Lemma 5.1 and get a bound on the critical point; on the other hand, it is shown in [?] that for binary tree the critical points for majority or maximum likelihood reconstruction coincide.

Theorem 5.2

i) *If $r \geq 4$, in the minority removal carried out every step with blocks of size M, for every $\varepsilon \in [0, 1/2)$ $\exists \bar{\bar{M}} : \forall M > \bar{\bar{M}}$*

$$\liminf_n \Delta_n(\bar{P}'_{\varepsilon,M}) > 0.$$

ii) *In the within-descent minority removal carried out every k steps if $p'_c(k)$ is the critical point then with $c > 0$ as in Theorem 4.1 we have*

$$\frac{1}{2^{\frac{1}{2k}} r} \leq p'_c(k) \leq \frac{4^{\frac{1}{2k}}}{c^{\frac{1}{2k}} r}$$

so that

$$\lim_{k \to \infty} p'_c(k) = \frac{1}{r}.$$

Proof. i) In generating T'_M at least $\frac{M}{2}$ vertices were preserved in each block of size M; these vertices have $r\frac{M}{2} \geq 4\frac{M}{2} = 2M$ descendants which can be divided into at least two blocks of size M (and some remaining others, possibly smaller). Thus, the number of descendants in the renormalized tree is at least 2.

On the other hand, the error rate $\bar{\varepsilon}'_M$ satisfies (**??**) with M replaced by $\frac{M}{2}$. By Lemma 5.1, maximum likelihood reconstruction on T'_M follows from that on $T^{(M/2)}$ which is ensured by

$$2(1 - 2\bar{\varepsilon}'_M)^2 \geq 2(1 - 2e^{-c_\varepsilon M/2})^2 > 1 \tag{98}$$

which is satisfied for large M.

ii) In generating T'_r at least $\frac{r^k}{2}$ vertices are preserved in each block of size r^k; each such vertex gives rise to one descendant block, so the branching number of the renormalized tree is at least $\frac{r^k}{2}$.

Also, it is possible to show bounds on the renormalized error free transmission $p'(k)$ similar to those used to prove Theorem 4.1. By carefully going through that proof, one can see that if $pr \geq 1$

$$p'(k)^2 \leq 2c_1^2 p^{2k} r^k (1 + \beta)^k \tag{99}$$

as in (**??**) if $pr < 1$ again $p(k)$ is exponentially small in k and thus there is no reconstruction; and, finally

$$p'(k) \geq c_2 \frac{p^k \sqrt{\frac{r^k}{2}}}{(1 + \alpha)^{k/2}}. \tag{100}$$

as in (**??**).

Again by Lemma 5.1 this implies

$$\frac{1}{r} \vee \frac{1}{(2c_1)^{\frac{1}{2k}} r} \leq p'_c(k) \leq \frac{4^{\frac{1}{2k}}}{c_2^{\frac{1}{2k}} r}$$

and

$$\lim_{k \to \infty} p_c'(k) = \frac{1}{r}.$$

■

Acknowledgments: We thank G. Giacomin for useful discussions and comments.

References

[1] W. Evans, C. Kenyon, Y. Peres and L. J. Schulman (2000). Broadcasting on Trees and the Ising Model. *Ann. Appl. Probab.* **10**, *410-433*

[2] P. M. Bleher (1990). Extremity of the disordered phase in the Ising model on the Bethe lattice. *Commun. Math. Phys.* **128**, *411-419*

[3] P. M. Bleher, J. Ruiz and V.A. Zagrebnov (1995). On the purity of the limiting Gibbs state for the Ising model on the Bethe lattice. *J. Stat. Phys.* **79**, *473-482*

[4] D. Ioffe (1996). On the Extremality of the Disordered State for the Ising Model on the Bethe Lattice. *Letters in Mthematical Physics* **37**, *137-143*

[5] C.M. Fortuin and P.W. Kasteleyn and J. Ginibre (1981). *Commun. Math. Phys.* **22**, *89*

[6] K. B. Athreya and P. E. Ney (1972). Branching Processes. *Springer-Verlag*

[7] K. B. Athreya (1994). Large Deviation Rates for Branching Processes–I. Single Type Case. *Ann. Appl. Probab.* **Vol. 4, No. 3**, *779-790*

[8] C. J. Preston (1974). Gibbs States on Countable Sets. *Cambridge University Press*

[9] A. Dembo and O. Zeitouni (1998). Large deviations techniques and applications. *Springer-Verlag*

[10] R. Lyons, Y. Peres (2004). (Book in Progress). Probability on Trees and Networks.

[11] E. Mossel (2004). Survey: Information Flow on Trees. *Dimacs Series in Discrete Mathematics an Theoretical Computer Science.*

[12] F. Martinelli, A. Sinclair, D. Witz (2006). Fast Mixing for Independent Sets, Colorings and Other Models on Trees. *Preprint.*

[13] Quansheng Liu (1996) The Growth of an Entire Characteristic Function and the Tail Probabilities of the Limit of a Tree Martingale. *Progr. in Prob.* **Vol. 40**

Alberto Gandolfi, Dipartimento di Matematica “Ulisse Dini”, Viale Morgagni 67a, 50134 FIRENZE, ITALY, gandolfi@math.unifi.it

Roberto Guenzani, Dipartimento di Matematica “Federico Enriques”, Via Cesare Saldini, 50 MILANO ITALY, guenzani@mat.unimi.it