# Information Theory and Phylogenetic trees

MATH 833 - Fall 2012 *Presenter: Daniel Pimentel*

## 1 Introduction

Consider transmitting symbols from an alphabet $\mathcal{A} := \{1, ..., k\}$ over a noisy $d$-ary tree $T = \{V, E\}$ from its root $\rho$ to its nodes $v \in V$ over the i.i.d. channels (edges) $e \in E$ with transition probabilities given by the $k \times k$ matrix $\mathbf{M}$. That is, for any $v \in V$ parent of $w \in V$.

$$\mathbf{M}_{i,j} := \mathbb{P}(\sigma(w) = j | \sigma(v) = i),$$

where $\sigma(v)$ denotes the symbol observed at node $v$.

Let $L_\ell$ denote the set of all the nodes in $V$ at level $\ell$. We are interested on determining whether or not we can estimate $\sigma(\rho)$ from $\sigma(L_\ell)$ for some $\ell$. Specifically, we are interested in knowing whether or not the reconstruction problem is *solvable*, i.e. if there exist $i, j \in \mathcal{A}$ s.t.

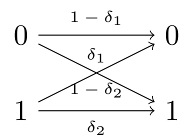
$$\lim_{n\to\infty} |\mathbb{P}(\sigma(L_n)|\sigma(\rho) = i) - \mathbb{P}(\sigma(L_n)|\sigma(\rho) = j)| > 0$$

Intuitively, this condition would imply that in the limit, we would be able to make a sensible decision about which symbol was most likely sent at the root.

## 2 Mossel's Approach

In [1], Mossel is particularly interested in determining solvability by analyzing the second largest eigenvalue of $\mathbf{M}$, $\lambda_2$. It is known that reconstruction is possible if $d\lambda_2^2 > 1$; he now shows that under certain conditions the reconstruction is also possible even if $d\lambda_2^2 < 1$, in particular for the cases of the binary asymmetric channel (Figure 1) and the symmetric channel on $q$ symbols.This can be summarized in the following theorem.

**Main Result.** *Consider the asymmetric binary channel in Figure 1. Suppose that $0 \leq \lambda \leq 1$ and $d\lambda \geq 1$; then there exists a $\delta \geq 0$ s.t. if $\lambda$ is the second largest eigenvalue of **M** and $\delta_1 < \delta$, then the reconstruction problem is solvable for the $d$-ary tree and the channel determined by **M**.*




$$\mathbf{M} = \begin{bmatrix} 1-\delta_1 & \delta_1 \\ 1-\delta_2 & \delta_2 \end{bmatrix}.$$

Figure 1: Binay Asymmetric Channel

### Sketch of the proof

Several lemmas about *$\ell$-diluted $b$-regular* trees, *$\lambda$-percolations*, and *components*[1] (Figure 2) are used, but fundamentally all are based on one:

**Lemma 7.** *Let $T$ be a $d$-ary tree, with percolation parameter*[2] $0 \leq \lambda \leq 1$ *such that* $d\lambda > 1$*. There exists a number* $\epsilon > 0$ *such that, for all* $b$*, there exists a number* $\ell$ *s.t.*

$$\mathbb{P}(|\mathscr{C}(\rho) \cap L_\ell| \geq b) \geq \epsilon.$$

SKETCH OF THE PROOF. The condition that $d\lambda > 1$ is equivalent to $\lambda > 1/d$. Since $\lambda$ is the percolation parameter of $T$, we get that on average, more than $1/d$ of all the edges will be open. Since at level $\ell$ we have $d^\ell$ edges, we only need to find the right $\ell$ for which $\frac{1}{d}d^\ell \geq b$. □

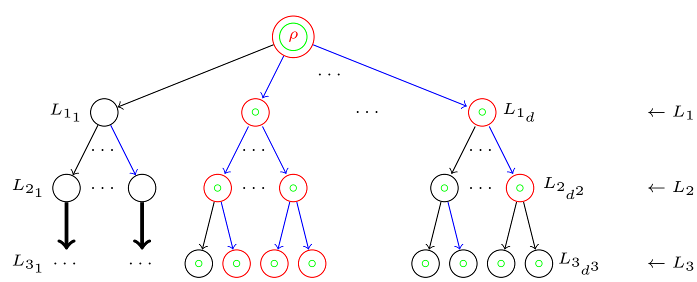



Figure 2: A $d$-ary tree $T$, where $\rho$ is its root, each edge is an i.i.d. channel characterized by $\mathbf{M}$, and blue edges are open. Red vertices represent $\mathscr{C}(\rho)$, and the subtree $T'$ formed by the vertices marked with ○ is an $\ell$-diluted $b$-regular open tree with $\ell = 1$ and $b = 2$

# References

[1] Mossel, E. (2001). Reconstruction on trees: beating the second eigenvalue. *The Annals of Applied Probability*, Vol. 11, No. 1, 285-300.

[1] $\mathscr{C}(v)$:=The set of all the vertices in $V$ connected to $v$ by a path of open edges.

[2] A percolation parameter on a tree is the i.i.d. probability that an edge is open. An edge connecting $v$ and $w$ is open if $\sigma(v) = \sigma(w)$