COMBINATORIAL STATISTICS AND THE SCIENCES

ELCHANAN MOSSEL

ABSTRACT

Combinatorial statistics studies inference in discrete stochastic models. Inference of such models plays an important role in the sciences. We survey research in combinatorial statistics involving the tree broadcast process. We review the mathematical questions that arise in the analysis of this process and its inference via "belief propagation." We discuss the mathematical connections to statistical physics, the social sciences, biological sciences, and theoretical computer science.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 60-02; Secondary 6365, 62-01, 68-02

KEYWORDS

Combinatorial statistics, broadcasting on trees, the reconstruction problem on trees

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 6, pp. 4170–4188
DOI 10.4171/ICM2022/179

Published by EMS Press

1. INTRODUCTION

Discrete probability models are used in many of the hard and soft sciences. Often the scientific challenges lead to novel mathematical questions in combinatorial statistics. The mathematical questions involve inference in such models. The goal of this survey paper is to discuss some of these processes, their inference, and their connections to the sciences. We focus on the tree broadcast process, the mathematical questions that arise in the analysis of this process and its inference via "belief propagation." We review some of the mathematical connections to statistical physics, the social sciences, biological sciences, and theoretical computer science.

1.1. A simple model on trees

In its simplest form, the model in question will be parametrized by four parameters: d, h, q, and θ. The model is defined on the d-ary tree of $h+1$ levels. Level 0 consists of the root, which has d children. Level 1 of the tree consists of the d children of the root. Each of the nodes at level 1 has d children. The collection of d^2 children of the nodes at level 1 makes level 2 of the tree, etc. We will denote the tree by $T=(V,E)$ and the $h+1$ levels by $L_0,\dots,L_h$. We will denote the level of node v by $|v|$. We will denote the root by 0.

We now define a discrete stochastic process indexed by the vertices V of T. We will give two equivalent definitions of this process. First, a recursive definition: the random variable X_0 is chosen uniformly at random from the set $[q] := \{1,\dots,q\}$. Now, for each child v of the root 0, independently, we toss a coin that lands Heads with probability θ. If it lands Heads, we let $X_v = X_0$. If it lands Tails, we sample X_v independently and uniformly at random from $[q]$. We then apply the same procedure recursively to each node at levels $2, 3$, etc.

A moment's thought reveals that the vector $X=(X_v : v \in V)$ has the following probability distribution:

$$\mathbb{P}\big[X=(x_v : v \in V)\big] = \frac{1}{q}\prod_{(u,v)\in E}\left(1(x_u = x_v)\theta + \frac{1-\theta}{q}\right), \tag{1.1}$$

where here and below all edges (u,v) are directed away from the root.

Note that the measure above is well defined for $1 \ge \theta \ge -1/(q-1)$, which will always be assumed. The extreme case $\theta = -1/(q-1)$ corresponds to the uniform measure on q-colorings of the tree. Below we will always exclude the frozen measures, where the root color determines all colors, by assuming $\theta < 1$, and $\theta > -1$, if $q = 2$.

There are many ways in which this process was generalized. Of particular interest are the following two. First, we may consider the process on general rooted trees, random, or deterministic. Second, we may consider more general broadcast processes from parent to child. In particular, assuming the state space is $[q]$, there is no reason that different edges (u,v) will have the same conditional law of X_u given X_v. Moreover, we can consider more general conditional laws of X_u given X_v. Thus for a general, possibly random, finite tree

$V = (T, E)$ and a collection M^e of Markov chains on the state space $[q]$, we may define

$$\mathbb{P}\big[X = (x_v : v \in V)\big] = \pi(x_0) \prod_{(u,v)\in E} M^{(u,v)}(x_u, x_v), \tag{1.2}$$

where π is a given probability distribution on $[q]$. We will always assume all chains $M^{(u,v)}$ are ergodic and that π is not a delta measure.

A lot of what we know about model (1.1) carries over to the more general setting of (1.2). For simplicity, we will mostly discuss (1.1), and sometimes comment on how things generalize.

1.2. Belief propagation and the reconstruction problem

Note that we may define the process $X = (X_v : v \in V)$ recursively also for a d-ary tree $T = (V, E)$ of infinitely many levels. Moreover, if we restrict to $(X_v : |v| \le h)$, it will be distributed according to (1.1).

We are interested in studying if the nodes at level h of the tree are asymptotically independent of X_0 as $h \to \infty$. We first note that by ergodicity of Markov chains we know that X_0 and X_v are asymptotically independent as $|v| \to \infty$, where $|v|$ denotes the level of v (for model (1.2), we need to require a bit more; we will not get into the details). In the language of statistical physics, this means that two point correlations decay exponentially as they do for finite ergodic Markov chains. Instead, we look at point-to-set correlations. More formally, let us denote by X_h the vectors of X_v for v at level h of the tree,

$$X_h = (X_v : |v| = h).$$

Then we are interested in the asymptotic independence of X_0 and X_h as $h \to \infty$. To formalize this question, let

$$Y_0 = \sum_{i=1}^{q} e_i 1(X_0 = i) \in \mathbb{R}^q,$$

where e_i is the ith unit vector.

Definition 1.1. We say that the reconstruction problem is solvable if

$$\lim_{h\to\infty} \mathbb{E}\|\mathbb{E}[Y_0|X_h] - \mathbb{E}[Y_0]\|_2^2 \neq 0. \tag{1.3}$$

In other words, the reconstruction problem is solvable if X_h provides some nonvanishing information on the value of X_0. There are many other equivalent definitions of reconstruction including some involving the limiting mutual information $\lim_{h\to\infty} I(X_0, X_h)$ or softer ones in terms of the tail-triviality of the sequence $X_0, X_1, \ldots$, see, e.g., [34, 43, 92] and the survey [68].

Interestingly, the quantity $f(X_h) = \mathbb{E}[Y_0|X_h]$ can be computed recursively and efficiently as a function of X_h via the belief propagation algorithm. This algorithm is used also for nontree graphical models [80] where it provides an approximation. The accuracy of belief propagation on trees was observed earlier, in specific contexts such as ancestral inference in phylogenetic trees [37, 44] and the study of the Ising model on trees [83]. Note that if

$\mathbb{E}[Y_0|X_h] \to \mathbb{E}[Y_0]$ (say in probability or a.s.) then for large values of h, there is little point in computing $\mathbb{E}[Y_0|X_h]$, as it is most likely trivial. Below we will often write BP instead of belief propagation.

2. LINEAR THEORY AND THE KESTEN–STIGUM BOUND

While there is an easy recursive computation of the function $f(x) = \mathbb{E}[Y_0|X_h = x]$, computing the limiting distribution or the limiting variance of $f(X_h)$ in (1.3) is in general difficult, as f is highly nonlinear and the coordinates of X_h are dependent. To remedy the first difficulty, it is natural to ask if there is a way to linearize the problem so that it is more amenable to analysis.

Interestingly, there are two approaches that lead to studying the same question:

(1) We can introduce an additional noise parameter $\eta > 0$ that will be applied only for the nodes at level h. For a deterministic value x_h of the nodes at level h, define the random vector

$$\tilde{x}_h := (\tilde{x}_v : |v| = h), \tag{2.1}$$

where for nodes v with $|v| = h$, we let $\tilde{x}_v = x_v$ with probability η, and it is independently and uniformly sampled from $[q]$ otherwise. We can then define a new function $\tilde{f}$ of the colors at level h by letting

$$\tilde{f}(x_h) = \frac{d}{d\eta}|_{\eta=0}\mathbb{E}\big[f(\tilde{x_h})\big].$$

Using the chain rule, it is easy to see that this is a linear function of the variables in x_h. More formally, it is a linear function of the $d^h q$ indicator variables $(1(x_v = i) : |v| = h, i \in [q])$. We can now study the correlation between $\tilde{f}(X_h)$ and X_0 instead of the variance in (1.3).

(2) Perhaps the most natural function of the X_h that one may study is $\sum Y_h := \sum_{v \in L_h} Y_v$, where

$$Y_v := \sum_{i=1}^{q} e_i 1(X_v = i) \in \mathbb{R}^q.$$

Of course, $\sum Y_h$ is just the count of how many of each of the q symbols appear at level h.

It is not hard to see that both approaches lead to studying the correlation between $\sum Y_h$ and X_0, see, e.g., [60]. In the work of Kesten and Stigum on multitype branching processes in the 1960s, they proved a law of large numbers for $\sum Y_h$ in [52] and then more refined limit theorems [51] which in particular imply:

Theorem 2.1. *For model* (1.1),

I. $d\theta^2 \leq 1 \implies$ *normalized* $\sum Y_h \xrightarrow[h\to\infty]{(d)}$ a normal law *independent of* X_0.

II. $d\theta^2 > 1 \implies$ *normalized* $\sum Y_h \xrightarrow[n\to\infty]{(d)}$ *a* nonnormal law dependent *on* X_0.

III. *In particular, if* $d\theta^2 > 1$ *then the reconstruction problem is solvable.*

The laws in parts I and II are nondegenerate. The results of [51] are in fact general enough to cover the more general model (1.2) on the d-ary tree if all the M matrices are identical and ergodic. In this case we let $\theta := \max(|\lambda_i| : \lambda_i \neq 1)$, where the λ_is denote the eigenvalues of M. The results further carry to random branching process trees with well-behaved degree distributions, where now d denotes the average number of offsprings.

The original proof of Theorem 2.1 uses the Fourier transform approach though martingale approaches can also be used to prove it as is hinted in [73].

Given Theorem 2.1, it is natural to ask if part I of the theorem implies nonreconstruction when $d\theta^2 \leq 1$. One way for this to work out would be for the higher-order terms in the expansion of BP, $\mathbb{E}[Y_0|X_h]$ to have a bounded contribution in probability. The recursive nature of BP allows proving it in the case of $q = 2$:

Theorem 2.2. *If* $q = 2$ *and* $d\theta^2 \leq 1$, *then the reconstruction problem is not solvable.*

This theorem was first proved by Bleher, Ruiz, and Zagrebnov [10]. Since then many other alternative proofs were presented. In particular, Theorem 2.2 was extended to general infinite trees in [34], where the general definition of d is now in terms of the *branching number* of the tree [56]. See also [81] for the analysis of the critical case for general trees. Proofs by Ioffe [46, 47] are formulated in terms of the FK representation from percolation theory, see, e.g., [41]. There are some recent short proofs based on information inequalities, see, e.g., [1, 82].

Beyond the case $q = 2$, Sly [88] proved nonreconstruction if $d\theta^2 \leq 1$ for $q = 3$ if $d \geq d_{\min}$, where $d_{\min}$ is some constant, and [13] proved it in the case where all the M are identical and given by 2×2 matrices that are almost symmetric.

In terms of the correlation between $\sum Y_h$ and X_0, in the paper [73] it is proven that for all q the distribution of $\sum Y_h$ is asymptotically independent of the root when $d\theta^2 \leq 1$. The paper [49] showed that in the noisy model (2.1), for all q, if $d\theta^2 < 1$, then there exists a constant amount of noise $\eta > 0$ such that $\tilde{X}_h$ is asymptotically independent of X_0 so

$$\lim_{h\to\infty} \mathbb{E}\|\mathbb{E}[Y_0|\tilde{X}_h] - \mathbb{E}[Y_0]\|_2^2 = 0.$$

The last two results say that if reconstruction is possible when $d\theta^2 < 1$ then (1) the information retained about X_0 is not in the count Y_h and (2) the information retained about X_0 is not robust against a fixed amount of noise.

3. NONLINEAR THEORY

Interestingly, for large values of q, reconstruction is possible even for some values of θ such that $d\theta^2 < 1$. First, as $q \to \infty$, the reconstruction threshold θ_q converges to $1/d$ as proven in [67].

Theorem 3.1. *Fix d and let*

$$\theta_q := \inf\bigl(\theta' > 0 : \textit{such that reconstruction is possible for parameters } (d, q, \theta), \forall \theta > \theta'\bigr).$$

Then $\lim_{q\to\infty} \theta_q = 1/d$.

This theorem is proven using branching process techniques. An easy and well-known argument states that reconstruction is impossible when $\theta \le 1/d$. Indeed, if we consider the branching process, where each node has $\mathrm{Bin}(d, \theta)$ children, then Z_h, the population at level h, counts the number of nodes whose colors have been copied from the root. Moreover, conditioned on Z_h, all other colors are independent of the root. Therefore if $Z_h \to 0$, X_0 and X_h are asymptotically independent. Since the branching process is subcritical ($Z_h \to 0$) if and only if $d\theta \le 1$ (see, e.g., [7]), it follows that $\theta_q \ge 1/d$ for all q.

When $d\theta > 1$, $Z_h \to \infty$ with positive probability. However, since we do not know from X_h the location of the colors that were copied from the root, it is still possible that X_h and X_0 are asymptotically independent, as is the case when $q = 2$ and $\theta \in (d^{-1}, d^{-1/2})$. The proof of the harder direction of Theorem 3.1 uses the fact that for large q if two recent descendants of the same node have the same color, it is very likely that node has the same color. Thus the proof uses a function that estimates the root to have a specific value i if a certain fractal-like subtree containing the value i at all of its leaves appears in X_h.

On the other hand, taking the asymptotics as $d \to \infty$, Sly [88] proves:

Theorem 3.2. *Fix $q \ge 5$ and let*

$$\overline{\theta}_d := \inf\bigl(\theta' > 0 : \textit{such that reconstruction is possible for parameters } (d, q, \theta), \forall \theta > \theta'\bigr).$$

Then $\lim_{d\to\infty} d\overline{\theta}_d^2 = C_q < 1$.

Similar results are obtained for $\theta < 0$.

Theorem 3.2 is proven by using a central limit theorem to analyze the basic belief propagation recursion and noting that the nonlinear terms shift the threshold.

A special case that attracted a lot of attention is the case of random coloring where $\theta = -1/(q-1)$. Interestingly, again, for large q the relationship between the critical d and θ is almost linear, see [87, 89].

4. CONNECTIONS TO STATISTICAL PHYSICS

The reconstruction problem on trees was first studied in statistical physics. The case $q = 2$ corresponds to studying the extremality or tail triviality of the Ising model on the tree [92]. Reconstruction solvability for $q = 2$ when $d\theta^2 > 1$ was proven in [43], the author of which was unaware that a more general result is implied by the results of Kesten and Stigum [51].

Interestingly, nonreconstruction for $q = 2$ when $d\theta^2 \le 1$ was first proven in a *spin glass* variant [14, 15, 18]. In this context reconstruction means

$$\lim_{h\to\infty} \mathbb{E}\|\mathbb{E}[Y_0|X_h = B_h] - \mathbb{E}[Y_0]\|_2^2 \ne 0,$$

where B_h are i.i.d. Bernoulli taking each of the two colors with probability $1/2$. The proof in this case is a little easier since in the analysis of the recursion for spin glasses the contributions coming from different subtrees are independent and identically distributed.

The interest in the reconstruction problem in statistical physics saw an explosion as the cavity and replica method played a crucial role in analyzing problems on sparse random graphs, see, e.g., [61–63,78,79]. At a very high level, for many combinatorial problems on sparse random graphs, statistical physics predictions are based on analysis (often nonrigorous) of the reconstruction problem or its variants on a corresponding tree. The connection to the reconstruction problem as defined here was formally made in [61]. In particular, it was conjectured in [61] that the Kesten–Stigum bound predicts the reconstruction threshold for $q = 2, 3$, and does not predict it for $q \geq 5$. As mentioned earlier, this was partially proven [88]. We will not try to summarize the connections between belief propagation and its variants, variants of the reconstruction problem and random constraint satisfaction problems. Some key papers in this area are [20,40,54,65]. This connection is also important in the work leading to the proof of the SAT threshold [21,22,27–29,58].

We will now give more details of one example, the example of *detection* in the *block model*. Here again we will see differences between the linear theory as reflected in the case $q = 2$ vs. nonlinear theory when q is large.

4.1. Detection in the block model

The block model is a random graph model generalizing the famous Erdős–Rényi random graph [33]. The block model is a special case of inhomogeneous random graphs, see, e.g., [11]. The sparse block model may be defined as follows:

Definition 4.1 (The sparse block model). Let $G(n, d, \theta, q)$ denote the model of random, $[q]$-labeled graphs in which each vertex u is assigned (independently and uniformly at random) a label $\sigma_u \in [q]$, and then each possible edge (u, v) is included with probability $(d/n)(1-\theta)$ if $\sigma_u \neq \sigma_v$ and with probability $(d/n)((1-\theta) + q\theta)$ if $\sigma_u = \sigma_v$.

We chose this parametrization so that for a fixed node, the distribution of the number of neighbors of each type will asymptotically agree with the distribution of the number of children of each type in model (1.2) with parameters q and θ on a random tree where each node has a Poisson with parameter d number of children.

The block model was studied extensively in statistics as a model of communities [45], see, e.g., [9,85,90], and in computer science as a model to study the average case behavior of clustering algorithms, see, e.g., [19,23,30,50,59] (interestingly there are very few citations between the two communities of papers even in cases where very similar results are proven). The papers above mostly concentrate on cases where the average degree is at least of order $\log n$, where n is the number of nodes in the graph.

The sparse case in Definition 4.1 became a major object of research due to a landmark paper in statistical physics [26] where the authors predicted that

Conjecture 4.2. *For the block model,*

I. *For all q, belief propagation on the graph G predicts the communities better than random if $d\theta^2 > 1$.*

II. *For $q = 2, 3$, it is information-theoretically impossible to predict better than at random if $d\theta^2 < 1$.*

III. *For $q \geq 5$, it is information-theoretically possible to predict better than at random for some θ with $d\theta^2 < 1$, but not in a computationally efficient way.*

These predictions were based on a linearization of belief propagation for the tree model.

4.2. $q = 2$—linear theory

A major challenge in establishing the algorithmic efficiency of belief propagation for block models stems from a fundamental difference between the application of belief propagation to trees and block models. When applied to trees, the input to belief propagation is the actual colors of the leaves. However, in the block model application, the colors are unknown. So here belief propagation is applied to random colors at all nodes that are independent of the actual colors.

In [55] it was conjectured that the global nonlinear operator that described one iteration of belief propagation on the graph should be linearized around its trivial fixed point to lead to a linear algebra based method to detect the communities. The resulting operator is not normal and its spectrum is complex. It is closely related to the operators used to analyze nonbacktracking walks [6,39,42,91].

This suggestion was followed up by an extensive body of work, including [2,5,12,72], that led proofs that linearized versions of BP detect communities better than at random when $d\theta^2 > 1$, which is in the spirit of part I of Conjecture 4.2.

The original statement of part I of Conjecture 4.2 states, furthermore, that belief propagation is optimal for the problem in the stronger sense that it minimizes the fraction of misclassified nodes. A combination of linearized belief propagation and belief propagation is used in [71] to obtain an efficient algorithm that minimizes the misclassification error when $q = 2$ and $d\theta^2 > C$ for some big constant C [71]. The main ingredient is proving that the estimator in the noisy model (2.1) asymptotically agrees with the original model (1.1):

$$\lim_{h\to\infty} |\mathbb{E}[Y_0|\tilde{X_h}] - \mathbb{E}[Y_0|X_h]| = 0. \tag{4.1}$$

Even earlier, part II of Conjecture 4.2 was partially established as it was shown in [70] that for $q = 2$ it is information-theoretically impossible to detect better than at random if $d\theta^2 < 1$ based on coupling of the graph and tree processes. The case $q = 3$ is still open.

4.3. Nonlinear theory

Parts of the nonlinear predictions in part III of Conjecture 4.2 were confirmed in [8] and [4] which provided exponential-time algorithms to detect for some parameters when

$d\theta^2 < 1$ when $q > 5$ (also when $q = 4$ and $\theta < 0, d\theta^2 < 1$). Of course, we rarely know how to prove that computational problems cannot be solved efficiently, so the support we have for the predicted computational-statistical gap is quite limited, see [3] for a more detailed discussion.

5. CONNECTIONS TO MOLECULAR BIOLOGY

The broadcast process on the tree was independently introduced in mathematical biology as a model of evolution of genetic information such as DNA sequences [16,36,77].

Naturally, the reconstruction problem is interesting in this context. Given the detailed evolutionary tree of some species, we want to infer as much as possible about the genetics of extinct species from the genetics of extant species.

An even more interesting question from a biological perspective is *recovering the species tree from genetic data*. Note that the details of this tree are required to study the reconstruction problem and infer ancestral genetic data.

Since Darwin's Origin of Species [24], a major goal of evolutionary biology is recovering the relationship between different species. Since the 1970s, this is most often done using genetic information collected from extant species. The models introduced in [16,36,77] assume that the genetic distribution of traits $(X_v : v \in V)$ is determined by a binary (rooted) tree $T = (V, E)$ and a collection $\theta_E = (\theta_e : e \in E)$ via the following variant of (1.1):

$$\mathbb{P}\big[X = (x_v : v \in V)\big] = \frac{1}{q} \prod_{e=(u,v)\in E} \left(1(x_u = x_v)\theta_e + \frac{1-\theta_e}{q} \right). \tag{5.1}$$

Note that T and θ_E determine a distribution of traits X and therefore the distribution $D(T, \theta_E)$ of $X_L = (X_v : v \text{ is a leaf})$. A major goal of the *phylogenetic reconstruction problem* is to estimate T and θ_E from independent samples from the distribution $D(T, \theta_E)$. In particular, we are interested in knowing how many samples are needed to recover T with good probability, as this translates to the data requirements needed for accurate estimation.

This ideal model that was introduced in the 1970s has since been generalized to account for many additional biological factors and mechanisms. Key theoretical results in this area include the identifiability of phylogenetic models [17] and efficient polynomial time algorithms to reconstruct phylogenetic trees [31,32,74]. See, e.g., [38,86,93] for general references on the phylogenetic problem.

The connection to the reconstruction problem was predicted by Steel [94] who conjectured that the amount of data needed to reconstruct phylogenetic trees crucially depends on the reconstruction problem.

The easiest setting to understand the connection between phylogenetic inference and the reconstruction problem is when $q = 2$ and the trees are very symmetric. We call a tree an h-level full binary tree if all the leaves are at level h. In the symmetric phylogenetic problem, we assume that the tree is an h-level full binary tree and that $\theta_e = \theta$ for all $e \in E$.

To understand what is inferred in this setup, let us fix $h = 2$. In this case the data given is

$$\left(X_v^i : v \in \{a, b, c, d\}, 1 \le i \le n\right). \tag{5.2}$$

This data can be thought of as four genetic sequences of length n, i.e., the genetic content of species a, b, c, d, where a, b, c, d are the leaves of the tree. Alternatively, the two-dimensional array (5.2) can be viewed as n i.i.d. samples from the process at the four leaves a, b, c, d. The main goal of inference in this simple case is to determine which species are siblings and which are cousins. The three possible sibling relationships are

$$\{\{a, b\}, \{c, d\}, \quad \{\{a, c\}, \{b, d\}, \quad \{\{a, d\}, \{b, c\}.$$

Of course, when h is bigger, we want to determine not just the sibling relation but also higher-order cousin relations.

Theorem 5.1. *Consider the symmetric phylogenetic problem with* $q = 2$ *and* n *independent samples from the distribution* $D(T, \theta_E)$, *where* T *is an* h*-level full binary tree and* $\theta_e = \theta$ *for all* $e \in E$.

I. *If the reconstruction problem is solvable for binary trees at the parameter* θ *(i.e., when* $2\theta^2 > 1$*) then there is an efficient algorithm that, given* $n = O(h)$ *samples, returns the correct tree with probability* $1 - \exp(-\Omega(h))$.

II. *If* θ *is strictly below the reconstruction threshold,* $2\theta^2 < 1$, *then it is information-theoretically impossible to infer the correct tree with probability* $\ge 1/2$ *unless* $n \ge \exp(\Omega(h))$.

The theorem above was first proven in [68] in a more general (and biologically relevant) setting.

For the proof of part I, the basic idea is that we may use correlation between different coordinates in samples from $D(T, \theta_E)$ to identify siblings, cousins, etc., in the tree. We may then estimate the state of their ancestor somewhat accurately since the reconstruction problem is solvable. We then use these estimates to find close relationships between the newly identified nodes and continue recursively.

For part II, one proves that for a node v at distance εh from the root, X_v has an exponentially small in h correlation with X_h. By taking ε sufficiently small, this implies that the same is true for the correlation between $X_{\varepsilon h}$ and X_h. Finally, this allows showing that, unless $n \ge \exp(\Omega(h))$, it is impossible to distinguish between the true tree and modifications of it permuting the nodes at level εh and the trees below them.

More realistic phylogenetic problems are not symmetric and much of the work in [25, 68] and follow up work was devoted to extending part I of the theorem to asymmetric cases.

Note, moreover, that the proof sketch of part I extends to all (q, θ) such that the reconstruction problem is solvable. However, the proofs in [25, 68] that do not have such a strong symmetry assumption do not extend to all such (q, θ) as they require robustness in various steps (the results trivially extend to even q when $2\theta^2 > 1$). Interestingly, the results of part I were extended in [75] to large q for some values of θ where $2\theta^2 < 1$ based on the root estimator in [67]. The results of [75] require the tree to be symmetric but not that $\theta_e = \theta$ for all e. The paper [75] also provided an extension of part II of Theorem 5.1 for (q, θ) when $\theta < \theta'$ and there is no reconstruction for parameter (q, θ').

It is natural to ask if there is a computational or information-theoretical barrier to extending the more realistic phylogenetic results of [25, 68, 84] to all θ's above the reconstruction threshold when $q \geq 5$. An analog of Theorem 5.1 for the limiting case $q = \infty$ is established in [76] for general (asymmetric) trees where the critical value of θ is $1/2$.

6. CONNECTIONS TO THEORETICAL COMPUTER SCIENCE

We have already seen many connections of the reconstruction problem to theoretical computer science. The connections included the role it played in algorithms and determining the satisfiability thresholds of random clustering, random graph, and random constraint satisfaction algorithms in Section 4, and the role it played in the information theoretic and algorithmic analysis of phylogenetic reconstruction in Section 5. Moreover, as belief propagation is a widely used algorithm, the analysis of the reconstruction problem and the robustness of this algorithm provide average case understanding of this important algorithm.

In this section we briefly discuss the computational complexity of the problem of estimating X_0 from X_h or approximately computing $\mathbb{E}[Y_0 | X_h]$. Furthermore, we review the connections between this problem and the classical theory of noisy computation and its connection to deep inference.

This question might seem strange as the belief propagation algorithm computes $\mathbb{E}[Y_0 | X_h]$ exactly in linear time. Note, however, that despite the linear running time it has two complex features:

(1) It uses real numbers. Indeed, the complexity is measured in terms of real arithmetic, but the model we are interested in is discrete.

(2) It is recursive. In other words, the circuit that computed BP has some depth. Is the depth necessary?

6.1. Recursive bounded memory algorithms

Here we only consider the simple model (1.1) with $q = 2$. In this case we know that the reconstruction threshold is given by $d\theta^2 = 1$ and that $\sum Y_h$ provides a good estimator of X_0 when $d\theta^2 > 1$.

Since the definition of the distribution of X_h is recursive, it is natural to ask if there is a simple recursive algorithm that estimates X_0 in a bottom up fashion, i.e., by a recursion of the form $\hat{X}_v = f(\hat{X}_w : w \in L_t(v))$, where $L_t(v)$ is the set of d^t descendants of v exactly t levels below v. The algorithm begins by initializing $\hat{X}_v = X_v$ for nodes v at the bottom level L_h and terminates by estimating X_0 by rounding $\hat{X}_0$ in some fashion.

As mentioned earlier, belief propagation can be written in this way for some real valued function f. The majority estimator $\mathrm{sgn}(\sum Y_h)$ can also be written in this way by computing the sum recursively. However, both of these require the domain of f to be unbounded. Is it possible to estimate X_0 in such recursive fashion using a function f that takes at most a constant B values and a bounded t?

The case $B = 2$ was studied in [66] assuming f is antisymmetric. In this case we can find the optimal function: f is the majority function. And the overall estimator of X_0 is a recursive majority function applied to X_h. This in turn allows computing for each t a critical threshold θ_t such that $\hat{X}_0$ is correlated with X_0 asymptotically if $\theta > \theta_t$ and is uncorrelated if $\theta < \theta_t$. The computation in [66] shows that for all t, $\theta_t < d^{-1/2}$ and $\lim_{t\to\infty} \theta_t = d^{-1/2}$.

There is an interesting connection between the derivation of the thresholds $\theta_1(d)$ and a derivation of von Neumann in the context of noisy computation [35, 95]. In his work on noisy computation, von Neumann considered circuits with noisy gates with the goal of designing circuits that, by duplicating inputs and applying majority gates to correct intermediate computations, are robust to some amount of noise. The derivation of the amount of noise that can be tolerated reduces to the question if the noisy recursive majority function with the all 1 inputs has limiting expectation bounded away from $1/2$. Interestingly, the broadcast model and the noisy computation model yields the same recursion and therefore we derive the same threshold for $\theta_1(d)$. In the noisy computation setting, the case $d = 3$ was derived by von Neumann [95], and was generalized to all d in [35]. The same recursion also appears in other models of noisy broadcast, see, e.g., [57].

In the context of the reconstruction problem, it was conjectured in [34] that any algorithm with bounded B cannot achieve the reconstruction threshold. This was recently established in [48] where it is shown that with B bits of memory, the critical θ_B satisfies $B^{-C} \le \theta_B - \theta \le B^{-c}$, for some positive constants $C > c > 0$.

6.2. The complexity of $\mathbb{P}[X_0|X_h]$

Recent efforts are devoted to studying the complexity of inference of X_0 from X_h in the linear regime when $d\theta^2 > 1$ vs. the nonlinear regime where $d\theta^2 < 1$, q is large, and the reconstruction problem is solvable.

Part of the motivation for studying this problem is to identify natural data-generating processes, where inference is possible but requires some nontrivial complexity.

The polynomial degree is one such measure of complexity. Thus we can ask if there is a low degree polynomial of the $d^h q$ indicator variables $(1(x_v = i) : |v| = h, i \in [q])$ that has nonvanishing correlation with X_0 as $h \to \infty$. We can say that:

(1) In the linear regime where $d\theta^2 > 1$, there is a linear function of the variables in X_h that is correlated with X_0 by Theorem 2.1.

(2) In [67] it is shown that for the model (1.2) on the d-ary tree, where $M^e = M$ for all e, there are chains M with $\theta = 0$ for which the reconstruction problem is solvable. The paper [53] shows that for such chains any polynomial of X_h of degree $\le (2^{ch})$ are uncorrelated with X_0, for some positive constant c.

(3) The authors of [53] ask if a similar phenomenon holds through the nonlinear regime. For example, is it true that polynomials of bounded degree have vanishing correlation with X_0 in the regime where $d\theta^2 < 1$?

In an earlier work [64], circuit complexity measures were used to study the inference of X_0 from X_h. The conjectured gap between the linear and nonlinear cases is reflected in the circuit class $\mathbf{TC}^0$ vs. $\mathbf{NC}^1$:

(1) Since the class $\mathbf{TC}^0$ of bounded depth circuits contains majority gates, it can trivially estimate X_0 better than at random when $d\theta^2 > 1$. Moreover, whenever (4.1) holds, $\mathbf{TC}^0$ can estimate X_0 with minimal error.

(2) It is not too hard to show that the computation of BP can always be carried out in $\mathbf{NC}^1$, the class of circuits of logarithmic depth. The paper [64] constructed a chain M with $\theta = 0$ for which estimating X_0 from X_h better than at random is $\mathbf{NC}^1$-complete.

(3) It is conjectured in [64] that estimating better than at random is $\mathbf{NC}^1$-complete when $d\theta^2 < 1$ and the reconstruction problem is solvable.

It is important to note that it is a major open problem to determine if $\mathbf{NC}^1 = \mathbf{TC}^0$.

In an even earlier work, the paper [69] considered a semisupervised version of the phylogenetic problem in the regime $d\theta^2 < 1$ and proved that in this regime it is information-theoretically impossible to classify the unlabeled data for algorithms that ignore correlations between features in the labeled data, while algorithms that do use high-order correlation can classify the data accurately. Moreover, in the regime $d\theta^2 > 1$, high-order correlations are not needed.

ACKNOWLEDGMENTS

Yuval Peres introduced me to the tree broadcast process at the beginning of my graduate studies. I am indebted to Yuval for fascinating discussions and central insights. Much of my work in this area has greatly benefited from collaborations with Mike Steel, Sebastien Roch, Allan Sly, Joe Neeman, Frederic Koehler, and Colin Sandon. Thanks to Sebastien Roch, Colin Sandon, Allan Sly, Nike Sun, Joe Neeman, and André Dixon for feedback on a draft of this survey.

FUNDING

This work was partially supported by NSF grant DMS-2031883, Vannevar Bush Faculty Fellowship ONR-N00014-20-1-2826 and Simons Investigator award (622132).

REFERENCES

[1] E. Abbe and E. Boix-Adserà, An information-percolation bound for spin synchronization on general graphs. *Ann. Appl. Probab.* **30** (2020), no. 3, 1066–1090.

[2] E. Abbe and C. Sandon, Community detection in general stochastic block models: fundamental limits and efficient algorithms for recovery. In *2015 IEEE 56th annual symposium on foundations of computer science*, pp. 670–688, IEEE, 2015.

[3] E. Abbe and C. Sandon, Detection in the stochastic block model with multiple clusters: proof of the achievability conjectures, acyclic bp, and the information-computation gap. 2015, arXiv:1512.09080.

[4] E. Abbe and C. Sandon, Crossing the KS threshold in the stochastic block model with information theory. In *2016 IEEE international symposium on information theory (ISIT)*, pp. 840–844, IEEE, 2016.

[5] E. Abbe and C. Sandon, Proof of the achievability conjectures for the general stochastic block model. *Comm. Pure Appl. Math.* **71** (2018), no. 7, 1334–1406.

[6] N. Alon, I. Benjamini, E. Lubetzky, and S. Sodin, Non-backtracking random walks mix faster. *Commun. Contemp. Math.* **9** (2007), no. 04, 585–603.

[7] K. B. Athreya and P. E. Ney, *Branching processes*. Springer, New York, 1972.

[8] J. Banks, C. Moore, J. Neeman, and P. Netrapalli, Information-theoretic thresholds for community detection in sparse networks. In *Conference on learning theory*, pp. 383–416, PMLR, 2016.

[9] P. Bickel and A. Chen, A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Natl. Acad. Sci.* **106** (2009), no. 50, 21068–21073.

[10] P. M. Bleher, J. Ruiz, and V. A. Zagrebnov, On the purity of the limiting Gibbs state for the Ising model on the Bethe lattice. *J. Stat. Phys.* **79** (1995), no. 1–2, 473–482.

[11] B. Bollobás, S. Janson, and O. Riordan, The phase transition in inhomogeneous random graphs. *Random Structures Algorithms* **31** (2007), no. 1, 3–122.

[12] C. Bordenave, M. Lelarge, and L. Massoulié, Non-backtracking spectrum of random graphs: community detection and non-regular Ramanujan graphs. In *Foundations of computer science (FOCS), 2015 IEEE 56th annual symposium on*, pp. 1347–1357, IEEE, 2015.

[13] C. Borgs, J. Chayes, E. Mossel, and S. Roch, The Kesten–Stigum reconstruction bound is tight for roughly symmetric binary channels. In *Proceedings of IEEE FOCS 2006*, pp. 518–530, IEEE, 2006.

[14] J. M. Carlson, J. T. Chayes, L. Chayes, J. P. Sethna, and D. J. Thouless, Bethe lattice spin glass: the effects of a ferromagnetic bias and external fields. I. Bifurcation analysis. *J. Stat. Phys.* **61** (1990), no. 5–6, 987–1067.

[15] J. M. Carlson, J. T. Chayes, J. P. Sethna, and D. J. Thouless, Bethe lattice spin glass: the effects of a ferromagnetic bias and external fields. II. Magnetized spin-glass phase and the de Almeida–Thouless line. *J. Stat. Phys.* **61** (1990), no. 5–6, 1069–1084.

[16] J. A. Cavender, Taxonomy with confidence. *Math. Biosci.* **40** (1978), no. 3–4, 271–280.

[17] J. Chang, Full reconstruction of Markov models on evolutionary trees: identifiability and consistency. *Math. Biosci.* **137** (1996), 51–73

[18] J. T. Chayes, L. Chayes, J. P. Sethna, and D. J. Thouless, A mean field spin glass with short-range interactions. *Comm. Math. Phys.* **106** (1986), no. 1, 41–89.

[19] A. Coja-Oghlan, Graph partitioning via adaptive spectral techniques. *Combin. Probab. Comput.* **19** (2010), no. 2, 227–284.

[20] A. Coja-Oghlan, F. Krzakala, W. Perkins, and L. Zdeborová, Information-theoretic thresholds from the cavity method. *Adv. Math.* **333** (2018), 694–795.

[21] A. Coja-Oghlan and K. Panagiotou, The asymptotic k-SAT threshold. *Adv. Math.* **288** (2016), 985–1068.

[22] A. Coja-Oglan and K. Panagiotou, Catching the k-NAESAT threshold. In *Proceedings of the forty-fourth annual ACM symposium on theory of computing*, pp. 899–908, ACM, 2012.

[23] A. Condon and R. Karp, Algorithms for graph partitioning on the planted partition model. *Random Structures Algorithms* **18** (2001), no. 2, 116–140.

[24] C. Darwin, On the origin of species (1859).

[25] C. Daskalakis, E. Mossel, and S. Roch, Evolutionary trees and the Ising model on the Bethe lattice: a proof of Steel's conjecture. *PTRF* **149** (2011), no. 1–2, 149–189.

[26] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E* **84** (2011), 066106.

[27] J. Ding, A. Sly, and N. Sun, Proof of the satisfiability conjecture for large k. In *Proceedings of the forty-seventh annual ACM symposium on theory of computing*, pp. 59–68, ACM, 2015.

[28] J. Ding, A. Sly, and N. Sun, Maximum independent sets on random regular graphs. *Acta Math.* **217** (2016), no. 2, 263–340.

[29] J. Ding, A. Sly, and N. Sun, Satisfiability threshold for random regular NAE-SAT. *Comm. Math. Phys.* **341** (2016), no. 2, 435–489.

[30] M. Dyer and A. Frieze, The solution of some random NP-hard problems in polynomial expected time. *J. Algorithms* **10** (1989), no. 4, 451–489.

[31] P. L. Erdős, M. A. Steel, L. A. Székely, and T. A. Warnow, A few logs suffice to build (almost) all trees (part 1). *Random Structures Algorithms* **14** (1999), no. 2, 153–184.

[32] P. L. Erdős, M. A. Steel, L. A. Székely, and T. A. Warnow, A few logs suffice to build (almost) all trees (part 2). *Theoret. Comput. Sci.* **221** (1999), 77–118.

[33] P. Erdős and A. Rényi, On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* **5** (1960), no. 1, 17–60.

[34] W. S. Evans, C. Kenyon, Y. Y. Peres, and L. J. Schulman, Broadcasting on trees and the Ising model. *Ann. Appl. Probab.* **10** (2000), no. 2, 410–433.

[35] W. S. Evans and L. J. Schulman, Signal propagation and noisy circuits. *IEEE Trans. Inf. Theory* **45** (1999), no. 7, 2367–2373.

[36] J. S. Farris, A probability model for inferring evolutionary trees. *Syst. Zool.* **22** (1973), no. 4, 250–256.

[37] J. Felsenstein, Maximum-likelihood estimation of evolutionary trees from continuous characters. *Am. J. Hum. Genet.* **25** (1973), no. 5, 471.

[38] J. Felsenstein, *Inferring phylogenies*. Sinauer, New York, NY, 2004.

[39] J. Friedman, *A proof of Alon's second eigenvalue conjecture and related problems*. Amer. Math. Soc., 2008.

[40] A. Gershchenfeld and A. Montanari, Reconstruction for models on random graphs. In *Foundations of computer science, annual IEEE symposium on*, pp. 194–204, IEEE Computer Society, 2007.

[41] G. Grimmett, The random-cluster model. In *Probability on discrete structures*, pp. 73–123, Encyclopaedia Math. Sci. 110, Springer, Berlin, 2004.

[42] K-i. Hashimoto, Zeta functions of finite graphs and representations of p-adic groups. In *Automorphic forms and geometry of arithmetic varieties*, pp. 211–280, Elsevier, 1989.

[43] Y. Higuchi, Remarks on the limiting Gibbs states on a $(d + 1)$-tree. *Publ. Res. Inst. Math. Sci.* **13** (1977), no. 2, 335–348.

[44] J. Hilden, GEN EX—An algebraic approach to pedigree probability calculus. *Clinical Genetics* **1** (1970), no. 5–6, 319–348.

[45] P. Holland, K. Laskey, and S. Leinhardt, Stochastic blockmodels: first steps. *Soc. Netw.* **5** (1983), no. 2, 109–137.

[46] D. Ioffe, Extremality of the disordered state for the Ising model on general trees. In *Trees (Versailles, 1995)*, pp. 3–14, Progr. Probab. 40, Birkhäuser, Basel, 1996.

[47] D. Ioffe, On the extremality of the disordered state for the Ising model on the Bethe lattice. *Lett. Math. Phys.* **37** (1996), no. 2, 137–143.

[48] V. Jain, F. Koehler, J. Liu, and E. Mossel, Accuracy-memory tradeoffs and phase transitions in belief propagation. In *Proceedings of the thirty-second conference on learning theory*, edited by A. Beygelzimer and D. Hsu, pp. 1756–1771, Proc. Mach. Learn. Res. 99, PMLR, Phoenix, USA, 2019.

[49] S. Janson and E. Mossel, Robust reconstruction on trees is determined by the second eigenvalue. *Ann. Probab.* **32** (2004), 2630–2649.

[50] M. Jerrum and G. Sorkin, The Metropolis algorithm for graph bisection. *Discrete Appl. Math.* **82** (1998), no. 1–3, 155–175.

[51] H. Kesten and B. P. Stigum, Additional limit theorems for indecomposable multidimensional Galton–Watson processes. *Ann. Math. Stat.* **37** (1966), 1463–1481.

[52] H. Kesten and B. P. Stigum, Limit theorems for decomposable multi-dimensional Galton–Watson processes. *J. Math. Anal. Appl.* **17** (1967), 309–338.

[53] F. Koehler and E. Mossel, Reconstruction on trees and low-degree polynomials, 2021. arXiv:2109.06915.

[54] F. Krząkała, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborová, Gibbs states and the set of solutions of random constraint satisfaction problems. *Proc. Natl. Acad. Sci.* **104** (2007), no. 25, 10318–10323.

[55] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly Z. L, and P. Zhang, Spectral redemption: clustering sparse networks. *Proc. Natl. Acad. Sci.* **100** (2013), no. 52, 20935–20940.

[56] R. Lyons, The Ising model and percolation on trees and tree-like graphs. *Comm. Math. Phys.* **125** (1989), no. 2, 337–353.

[57] A. Makur, E. Mossel, and Y. Polyanskiy, Broadcasting on random directed acyclic graphs. *IEEE Inf. Theory* **66** (2020), no. 2, 780–812.

[58] E. Maneva, E. Mossel, and M. J. Wainwright, A new look at survey propagation and its generalizations. *J. ACM* **54** (2007), 41.

[59] F. McSherry, Spectral partitioning of random graphs. In *Foundations of computer science, 2001. Proceedings 42nd IEEE symposium on*, pp. 529–537, IEEE, 2001.

[60] M. Mézard and A. Montanari, Reconstruction on trees and the spin glass transition. *J. Stat. Phys.* **124** (2006), 1317–1350.

[61] M. Mézard and A. Montanari, *Information, physics, and computation*. Oxford University Press, USA, 2009.

[62] M. Mézard, G. Parisi, and R. Zecchina, Analytic and algorithmic solution of random satisfiability problems. *Science* **297** (2002), no. 5582, 812–815.

[63] M. Mezard and R. Zecchina, Random k-satisfiability: from an analytic solution to an efficient algorithm. *Phys. Rev. E* **66** (2002).

[64] A. Moitra, E. Mossel, and C. Sandon, Parallels between phase transitions and circuit complexity? In *Conference on learning theory*, pp. 2910–2946, PMLR, 2020.

[65] A. Montanari, R. Restrepo, and P. Tetali, Reconstruction and clustering in random constraint satisfaction problems. *SIAM J. Discrete Math.* **25** (2011), no. 2, 771–808.

[66] E. Mossel, Recursive reconstruction on periodic trees. *Random Structures Algorithms* **13** (1998), no. 1, 81–97.

[67] E. Mossel, Reconstruction on trees: beating the second eigenvalue. *Ann. Appl. Probab.* **11** (2001), no. 1, 285–300.

[68] E. Mossel, Survey: Information flow on trees. In *Graphs, morphisms and statistical physics. DIMACS series in discrete mathematics and theoretical computer science*, edited by J. Nestril and P. Winkler, pp. 155–170, 2004.

[69] E. Mossel, Deep learning and hierarchal generative models. 2019, arXiv:1612.09057.

[70] E. Mossel, J. Neeman, and A. Sly, Reconstruction and estimation in the planted partition model. *Probab. Theory Related Fields* **3–4** (2015), 431–461.

[71] E. Mossel, J. Neeman, and A. Sly, Belief propagation, robust reconstruction, and optimal recovery of block models. *Ann. Appl. Probab.* **26** (2016), no. 4, 2211–2256.

[72] E. Mossel, J. Neeman, and A. Sly, A proof of the block model threshold conjecture. *Combinatorica* **38** (2018), no. 3, 665–708.

[73] E. Mossel and Y. Peres, Information flow on trees. *Ann. Appl. Probab.* **13** (2003), no. 3, 817–844.

[74] E. Mossel and S. Roch, Learning nonsingular phylogenies and hidden Markov models. In *Proceedings of the thirty-seventh annual ACM symposium on theory of computing, Baltimore (STOC'05), MD, USA*, pp. 366–376, ACM, 2005.

[75] E. Mossel, S. Roch, and A. Sly, On the inference of large phylogenies with long branches: How long is too long? *Bull. Math. Biol.* **73** (2011), no. 7, 1627–1644.

[76] E. Mossel and M. Steel, A phase transition for a random cluster model on phylogenetic trees. *Math. Biosci.* **187** (2004), no. 2, 189–203.

[77] J. Neyman, Molecular studies of evolution: a source of novel statistical problems. In *Statistical decision theory and related topics*, edited by S. S. Gupta and J. Yackel, pp. 1–27, Elsevier, 1971.

[78] M. M. G. Parisi, A replica analysis of the travelling salesman problem. *J. Phys.* **47** (1986), 1285–1296.

[79] M. M. G. Parisi, On the solution of the random link matching problem. *J. Phys.* **48** (1987), 1451–1459.

[80] J. Pearl, *Reverend Bayes on inference engines: A distributed hierarchical approach*. Cognitive Systems Laboratory, School of Engineering and Applied Science, 1982.

[81] R. Pemantle and Y. Peres, The critical Ising model on trees, concave recursions and nonlinear capacity. *Ann. Probab.* **38** (2010), no. 1, 184–206.

[82] Y. Polyanskiy and Y. Wu, Application of the information-percolation method to reconstruction problems on graphs. *Math. Stat. Learn.* **2** (2020), no. 1, 1–24.

[83] C. J. Preston, *Gibbs states on countable sets: Gibbs states and Markov random fields*. Cambridge University Press, 1974.

[84] S. Roch and A. Sly, Phase transition in the sample complexity of likelihood-based phylogeny inference. *Probab. Theory Related Fields* **169** (2017), no. 1, 3–62.

[85] K. Rohe, S. Chatterjee, and B. Yu, Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** (2011), no. 4, 1878–1915.

[86] C. Semple and M. Steel, *Phylogenetics*. Math. Appl. Ser. 22, Oxford University Press, 2003.

[87] A. Sly, Reconstruction of random colourings. *Comm. Math. Phys.* **288** (2009).

[88] A. Sly, Reconstruction for the Potts model. *Ann. Probab.* **39** (2011), no. 4, 1365–1406.

[89] A. Sly and Y. Zhang, Reconstruction of colourings without freezing. 2016, arXiv:1610.02770.

[90] T. Snijders and K. Nowicki, Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *J. Classification* **14** (1997), no. 1, 75–100.

[91] S. Sodin, Random matrices, nonbacktracking walks, and orthogonal polynomials. *J. Math. Phys.* **48** (2007), 123503.

[92] F. Spitzer, Markov random fields on an infinite tree. *Ann. Probab.* **3** (1975), no. 3, 387–398.

[93] M. Steel, *Phylogeny: Discrete and random processes in evolution*. SIAM, 2016.

[94] M. Steel, My Favourite Conjecture, 2001, http://www.math.canterbury.ac.nz/~mathmas/conjecture.pdf.

[95] J. von Neumann, Probabilistic logics and the synthesis of reliable organisms from unreliable components. In *Automata studies*, pp. 43–98, Ann. of Math. Stud. 34, Princeton University Press, Princeton, NJ, 1956.

ELCHANAN MOSSEL

77 Massachusetts Avenue, Cambridge, MA 02139-4307, USA, elmos@mit.edu