Eötvös Loránd University

Faculty of Science

Dániel Garamvölgyi

Mathematics BSc

Factor of i.i.d. processes and regular graphs

BSc Thesis

Advisor:

Ágnes Backhausz

Department of Probability Theory and Statistics

Budapest, 2016

Acknowledgements

I would like to express my gratitude to my advisor, Ágnes Backhausz, for introducing me to this very current topic of research, for her patience and knowledge in answering my questions, and for the helpful and wide-ranging suggestions she made regarding the thesis.

Contents

Introduction 4

1 Definitions, preliminary results **6**

1.1 Definitions, terminology 6

1.2 Factor of i.i.d. processes 7

2 Independent sets **11**

2.1 A survey of the problem 11

2.2 The Gaussian wave function and its properties 13

2.3 Constructing independent sets 17

3 Limitations of factor of i.i.d. algorithms **21**

3.1 Non-optimality for independent sets 22

3.2 Correlation decay 26

Introduction

There are multiple directions from which one can approach factor of i.i.d. processes. One possible starting point is this: given a graph, we put independent, identically distributed random variables – intuitively, a random label – on each vertex. Then we construct some structure on the graph (e.g. a colouring), with the restriction that for each vertex, our decision regarding that vertex only depends on the labels in some finite neighbourhood of it. The question is, then, what sort of structures can these algorithms produce? Can any random colouring of the graph be constructed in this way? Can we give a procedure of this kind that, with high probability, gives an independent set of the graph that is close to maximal in size? Or, more generally, can we characterize the random processes on the graph that arise in this way?

Factor of i.i.d. processes can be seen as a formalisation (and generalisation) of this concept. The goal of this thesis is to provide a glimpse into the various facets of the current research in this area, especially in the context of regular graphs. We will see concrete examples of graph algorithms based on these notions, as well as some of their limitations.

The thesis is structured as follows: first, in Chapter 1 we recall relevant concepts and notation from graph theory and probability theory. Then, in the second section of the chapter we give a rigorous definition of factor of i.i.d. processes and highlight some of their basic properties.

Chapter 2 is an account of [1], a paper giving an example of algorithms based on factor of i.i.d. processes which construct various graph-theoretic structures, in this case large independent sets in regular graphs. Beside possible practical applications, these algorithms have theoretic relevance in that they can be used to prove lower bounds on the size of the maximal independent set in certain graphs; in fact, the bound provided for finite 3-regular graphs was optimal at the time of publication. We will also prove the existence of the so-called Gaussian wave function, a random process on the d-regular tree which is of separate mathematical interest.

In light of positive results such as the one described in the previous chapter, the question naturally arises whether factor of i.i.d. algorithms can provide optimal, or at least near-optimal answers to various graph optimization problems (such as constructing colourings and matchings, or the aforementioned problem of finding large independent sets). A related and actively researched problem is the characterization of factor of i.i.d. processes, that is, finding necessary and sufficient conditions for a random process to be a factor of an i.i.d. process. We examine both of these problems in Chapter 3. We first summarize a paper of Gamarnik and Sudan [2], in which they prove that for finding independent sets in random regular graphs, no factor of i.i.d.

algorithm can be optimal, provided that the degree d of the graphs in question is large enough. Finally, in the second section of the chapter we prove a theorem from [3], which essentially asserts that, given a factor of i.i.d. process on the d-regular tree (for $d \geq 3$), the correlation of marginal distributions on vertices distant from each other must be small. The result gives a quantitative measure of this "smallness", giving an upper bound on the absolute value of the correlation which decays exponentially as the distance increases. This is among the most effective tools currently known to determine whether a random process is a factor of i.i.d. process.

Chapter 1

Definitions, preliminary results

The purpose of this chapter is to give the basic definitions of the objects that we will focus on throughout the thesis. In the first section we recall the notions from graph theory and probability theory that we will use in the subsequent chapters. Then in the second section we define *factor of i.i.d.* processes and various related concepts, which will be the main object of our investigations.

1.1 Definitions, terminology

Graph terminology. We will use the word *graph* to mean a simple, undirected graph. The vertex and edge set of a graph G will be denoted by $V(G)$ and $E(G)$, respectively. The *girth* of a graph is the length of its shortest cycle. A function $\phi : V(G) \to V(G')$ is an *isomorphism* between the graphs G and G' if it is bijective and for any vertices $u, v \in V(G)$ the pair (u, v) is an edge in G if and only if $(\phi(u), \phi(v))$ is an edge in G'; in the case when $G = G'$ we say that ϕ is an *automorphism* of G. The automorphisms of a graph form a group under composition, which will be denoted by $\mathrm{Aut}(G)$. G is said to be vertex transitive if the action of $\mathrm{Aut}(G)$ on the vertex set is transitive, that is, for every pair of vertices u, v there exists an automorphism ϕ which maps u to v.

Given a positive integer d, we say that a graph is *d-regular* if every vertex of the graph has degree d. There exists a d-regular graph on n vertices if and only if n is greater than d and nd is even: the "only if" direction is trivial by the fact that the sum of the degrees must be even, while sufficiency can be shown with a direct construction, e.g. let the vertex set be $\{0, \dots, n-1\}$ and draw an edge between any vertex k and $k + j \pmod n$ for $j = 1, \dots, \frac{d}{2}$ in the case that d is even and $j = 1, \dots, \frac{(d-1)}{2}, \frac{n}{2}$ otherwise. For $d \geq 2$ we define the *(infinite) d-regular tree*, denoted by T_d, as the unique (up to isomorphism) connected d-regular graph containing no

cycles. Sometimes we distinguish a vertex o of T_d which we call the *root*. A *random d-regular graph* on n vertices is a random variable that is uniformly distributed on the space of d-regular graphs of size n.

Gaussian random variables. We recall some properties of the Gaussian distribution which we will use in Chapter 2. As usual, a real-valued random variable is said to have *standard normal* distribution if its density function is of the form $f(x) = (2\pi)^{-\frac{1}{2}} \exp(-\frac{x^2}{2})$. The *k-dimensional standard normal distribution* is the joint distribution of $k \geq 1$ independent standard normal variables. A (real) vector-valued random variable is said to have a *multivariate Gaussian* or *jointly normal* distribution if it is of the form $AX+b$, where X is a random variable with k-dimensional standard normal distribution, A is an $n \times k$ real matrix and b is a vector of $\mathbb{R}^n$, for some positive n. A multivariate Gaussian is *degenerate* if its covariance matrix is singular; otherwise it is *non-degenerate* (i.e. positive definite, since covariance matrices are always positive semidefinite). It is a standard fact that there exists a jointly normal distribution with prescribed mean μ and covariance matrix Σ precisely if Σ is positive semidefinite, and whenever it exists, it is unique.

We close this section by presenting a proposition concerning trivariate Gaussians which we will need in Chapter 2. The statement and a sketch of the proof can be found in [1].

Proposition 1.1. *For a trivariate, non-degenerate Gaussian (Z_1, Z_2, Z_3) with zero mean we have*

$$\mathbb{P}(Z_1 > 0, Z_2 > 0, Z_3 > 0) = \frac{1}{2} - \frac{1}{4\pi} \sum_{1 \leq i < j \leq 3} \arccos(\mathrm{corr}(Z_i, Z_j)).$$

1.2 Factor of i.i.d. processes

In this section we formalize the concept of a factor of i.i.d. process outlined in the introduction. A *random process* on G is a collection of random variables $X = ((X_v)_{v \in V(G)}), X_v : \Omega \to \mathbb{R}$. The random variable X_v is called a *marginal* of the process. If the marginals are identically distributed and they are independent, we say that they form an *i.i.d. process*.

We first formalize the notion that an algorithm on a graph does not use global information in a rather general sense. In the next definition, F can be seen as an algorithm which, upon receiving information (a real number) from each vertex of the graph, constructs an output on each vertex. Our requirement is that the output on a given vertex should not use global knowledge about "where" the vertex is within the graph: if we permute the nodes (and the information they carry accordingly) with

an automorphism of the graph, the output should be permuted in the same way, but not changed, apart from that. This gives rise to the definition of a factor.

Definition 1.2. Let $(Y, \mathcal{A})$ be a measurable space. We say that a measurable $F : \mathbb{R}^{V(G)} \to Y^{V(G)}$ function is an *Aut(G)-factor* (or simply *factor*), if for any vector $x \in \mathbb{R}^{V(G)}$, vertex $v \in V(G)$ and automorphism $\phi \in \mathrm{Aut}(G)$ we have $F(\phi x)_v = F(x)_{\phi^{-1}(v)}$, where $\phi(x_u)_{u \in V(G)} = (x_{\phi(u)})_{u \in V(G)}$. In other words, F commutes with the action of $\mathrm{Aut}(G)$.

Measurability in the above definition is meant with respect to the natural product σ-algebras obtained from the Borel σ-algebra on $\mathbb{R}$ and from $\mathcal{A}$ on Y; that is to say, the ones generated by the cylinder sets. In most cases Y is either $\mathbb{R}$ or a discrete set.

Definition 1.3. We say that the random process X is a *factor of i.i.d. process* if it is equal in law to $F(Z)$, where Z is an i.i.d. process and $F : \mathbb{R}^{V(G)} \to \mathbb{R}^{V(G)}$ is a factor. Similarly, a *factor of i.i.d. set* is a random set that is acquired as a factor of an i.i.d. process; formally it is a set of the form $\{v \in V(G) : F(Z)_v = 1\}$, where Z is an i.i.d. process and $F : \mathbb{R}^{V(G)} \to \{0,1\}^{V(G)}$ is a factor.

Some authors (e.g. [2], [3], [4]) only allow the random labelling to have marginals with uniform distribution on $[0,1]$. This is not a real restriction, since any distribution on $\mathbb{R}$ can be realized as a (measurable) function of the this one. When presenting a paper, we will always follow the convention that the authors use.

Our main focus will be on factor of i.i.d. processes on d-regular trees. In this context (or generally, in the case when G is vertex-transitive) one can give a different formalization, which is sometimes more convenient to use.

Definition 1.4. Let T_d be the rooted d-regular tree and $(Y, \mathcal{A})$ a measurable space. We say that a function $f : \mathbb{R}^{V(T_d)} \to Y$ is a *rule function* if it is measurable and invariant under the action of automorphisms that fix the root. In other words, f is spherically symmetric around the root. Given a random i.i.d. labelling of $Z = (Z_v)_{v \in V(T_d)}$, the factor of i.i.d. process generated by the rule function is $(X_v)_{v \in V(T_d)}$, with $X_v = f(\phi_v Z)$, where ϕ_v is an automorphism mapping the root to v, acting on the labels as before.

The factor of i.i.d. process in the above process is well defined because of the assumption that f is spherically symmetric around the root. It is easy to see that the two definitions given are equivalent in the sense that (after we fix a root on T_d) a factor uniquely defines a rule-function and vice versa.

A simple example of a factor of i.i.d. process is one that, on each node, outputs the sum of the labels of the node's neighbours. This can be naturally generalized into following class of factor of i.i.d. processes.

Definition 1.5. We say that a process X is a *linear factor* of the i.i.d. process Z if there exists a sequence of real numbers α_0, α_1, ..., such that for each $v \in V(G)$

$$X_v = \sum_{u \in V(G)} \alpha_{d(v,u)} Z_u = \sum_{k=0}^{\infty} \sum_{\substack{u \geq 0 \\ d(v,u)=k}} \alpha_k Z_u,$$

where $d(v, u)$ denotes the distance of v and u in G.

So far we only required locality in the sense that the algorithm couldn't use any information about the global structure of the graph. For both practical and theoretical reasons, it is often desirable to assume that the output of the algorithm on a vertex only depends on a finite neighbourhood of said vertex.

Definition 1.6. We say that a factor of i.i.d. process $X = F(Z)$ on the graph G is a *block* (or *r-local*) *factor of i.i.d. process* if there is a positive integer r such that for any vertex $v \in V(G)$ and labellings $x, y \in \mathbb{R}^{V(T_d)}$, whenever x and y are equal on the r-neighbourhood of v, $F(x)_v = F(y)_v$.

Locality can be similarly defined in terms of rule functions. We will see multiple examples of how block factor of i.i.d. processes allow us to make a connection between results on infinite and finite graphs (in our case between T_d and various finite d-regular graphs). We end this chapter by showing that block factors can, in a sense, approximate any factor of i.i.d. on T_d that has marginals with finite expected value.

Proposition 1.7. *Let $X = (X_v)_{v \in V(T_d)}$ be a factor of i.i.d. process on the rooted d-regular tree, generated by the rule function f and some random i.i.d. labeling Z. Assume that the marginals of X have finite expected value. Then there exists a sequence of rule functions $f_1, f_2, \ldots$, such that the factor of i.i.d. process X^i generated by f_i is a block factor and for any vertex v the marginals X^i_v converge to X_v, both almost surely and in L^1, as $i \to \infty$.*

Proof. The proof is a simple application of well-known results from martingale theory. The basic idea is to consider the rule function f as a random variable and let the approximating functions f_i be the conditional expectation of this random variable with respect to ever larger neighbourhoods of the root o. This can be made rigorous as follows. Let λ denote the probability distribution of the labeling Z on ($\mathbb{R}^{V(T_d)}$, $\mathcal{B}$), where $\mathcal{B}$ is the product of the Borel σ-algebras. Then f, considered as a random variable on the probability space ($\mathbb{R}^{V(T_d)}$, $\mathcal{B}, \lambda$) has the same distribution as X_o.

Let p_v be the projection from $\mathbb{R}^{V(T_d)}$ onto the coordinate corresponding to the vertex v. Consider, for $i \geq 1$, the σ-algebra $\mathcal{F}_i = \sigma(\{p_v : d(v, o) \leq i\})$, i.e. the σ-algebra generated by the projections in the i-neighbourhood of the root. Note that

$\sigma(\cup_{i=1}^{\infty}\mathcal{F}_i) = \mathcal{B}$. We claim that the sequence of functions $f_i = \mathbb{E}(f|\mathcal{F}_i)$ has the desired properties. First, since $\mathbb{E}f < \infty$, the sequence $f_1, f_2, \ldots$ converges to $\mathbb{E}(f|\mathcal{B}) = f$ both almost surely and in L^1; this is essentially the statement of Lévy's zero-one law, which is a corollary of the martingale convergence theorem. Furthermore, since f_i is $\mathcal{F}_i$-measurable, it only depends on the i-neighbourhood of the root. The final step is to show that for positive i, f_i is indeed a rule function, that is, it is spherically symmetric around the root. Let ϕ be an automorphism of T_d that leaves the root fixed and consider a set $A \in \mathcal{F}_i$. Note that the action $\phi : \mathbb{R}^{V(T_d)} \to \mathbb{R}^{V(T_d)}$ is measure-preserving, and so the measures $\lambda \circ \phi^{-1}$ and λ are equal. This means that, using the change of variables formula for Lebesgue integration, we have

$$\int_A f_i(\phi x)\, d\lambda(x) = \int_{\phi A} f_i\, d(\lambda \circ \phi^{-1}) = \int_{\phi A} f_i\, d\lambda = \int_{\phi A} f\, d\lambda = \int_A f\, d\lambda,$$

where the third equality is justified since $\phi A = \{\phi x : x \in A\}$ is easily seen to be in $\mathcal{F}_i$ and f_i is the conditional expectation. This, then, by the uniqueness of the conditional expectation shows that, apart from a set of measure zero, f_i is spherically symmetric. For the sake of completeness we note that we can choose f_i to be spherically symmetric around the root everywhere by choosing the conditional expectation in an arbitrary way and then averaging it over all the (finitely many) root-fixing automorphisms of the i-neighbourhood of the root.

Thus $f_i, i \geq 1$ is a sequence of rule functions generating block factor of i.i.d. processes. An immediate consequence of the construction of the probability space $(\mathbb{R}^{V(T_d)}, \mathcal{B}, \lambda)$ is that the convergence of f_i to f implies the convergence of the marginals X_o^i to X_o, both almost surely and in L^1. The convergence of the marginals on any vertex v can be similarly seen, using the observation that any root-fixing automorphism ϕ is measure preserving as a function on this probability space, as was noted before.

□

Chapter 2

Independent sets

In this chapter we are going to examine a method of finding independent sets (sets of vertices with no edges connecting any two of them) in graphs with certain properties. Besides possible practical applications, this approach will give a lower bound on the ratio of the size of the largest independent set compared to the size of the graph, a number referred to as the *independence ratio* of the graph. Throughout this chapter we will follow [1].

2.1 A survey of the problem

Our focus will be on finite d-regular graphs (for $d \geq 3$, and especially $d = 3$) with large girth, in other words, graphs that locally look like the d-regular tree.

There have been numerous results regarding the independence ratio of d-regular graphs, and its asymptotic behavior as either d or the girth goes to infinity. Since in a tree every other node can be chosen to form an independent set, one could think that the independence ratio would approach $1/2$ as the girth goes to infinity. However, Bollobás [5] showed that this is not the case for uniform random d-regular graphs: while with high probability the number of small cycles in these graphs are small, their independence ratio is bounded away from $1/2$. A detailed survey of previous upper and lower bounds can be found in [1]; examples of more recent development in this area are [6] and [7].

The main result of this chapter is the following theorem (Theorem 1 in [1]):

Theorem 2.1. *Every* 3*-regular graph with sufficiently large girth has independence ratio at least* 0.4361*.*

This was an improvement on the previous lower bound which had been 0.4352, due to Kardoš, Král and Volec [8]. Note that this bound is computer-assisted in that

the authors used a computer to calculate certain numerical integrals; however, one can use simpler estimates to get a bound of 0.43. It is this latter bound that we will prove, and while we will describe the method of proving Theorem 2.1, we will not replicate the computations, which can be found in the appendix of [1]. This lower bound has been improved as well since the publication of said paper, the current best result (at the time of writing) being 0.44533 (see [7]).

We will examine two related methods. Both of them consists of randomly assigning numbers to each node of the graph and then choosing vertices based on a rule in such way that the chosen nodes will form an independent set. If we can compute the expected size of the random set acquired in this way, we get a lower bound on the size of the maximal independent set in the graph (and consequently on the independence ratio).

The first approach is to choose those vertices which have a number assigned to them that is larger than the numbers assigned to their neighbours. The set acquired thus will of course be independent; moreover, if we assigned the numbers in a way that is independent and identically distributed for each vertex, then the probability of being chosen is the same for all nodes, so the expected size of the set is just the size of the graph multiplied by this probability. To illustrate this approach, we prove the following simple proposition using it:

Proposition. *Every finite d-regular graph has independence ratio at least $1/(d+1)$.*

Proof. Given a d-regular graph, let the random variables $X_v, v \in V(G)$ be independent and uniformly distributed on $[0, 1]$. We will choose vertices according to the rule described in the preceding paragraph. For any vertex v_0 with neighbours $v_1, \ldots, v_d$, the probability of the event that v_i has the largest label among these vertices is the same for all $0 \leq i \leq d$, and these are disjoint events, exactly one of which must hold. It follows that the probability of choosing v_0 is $1/(d+1)$. Since this probability is the same for all nodes, the expected size of the resulting independent set is $|V(G)|/(d+1)$, implying that there is an independent set at least this size, which concludes the proof. □

The second approach is more complex, but yields better results. This time we fix a threshold $\tau \in \mathbb{R}$, and keep only those vertices which have a number smaller than τ assigned to them. With the right threshold, the remaining components of the graph will be small with high probability, and so under certain circumstances (such as the graph not containing short cycles) they will often be trees, from which we can choose large independent sets.

It is easily seen that both of these are examples of factor of i.i.d. independent sets (assuming that the random labeling is an i.i.d. process). Furthermore, in both cases

we can easily modify the algorithm in a way that produces another independent set that is disjoint from the previous one (taking the vertices with smaller numbers than their neighbours', and keeping only the vertices having a larger number than the threshold, respectively). The two subsets taken together form a bipartite graph, and so we acquire lower bounds for the relative size of the maximal induced bipartite graph as well. Thus, a byproduct of our proof will be the following theorem (Theorem 2 in [1]):

Theorem 2.2. *Every* 3*-regular graph with sufficiently large girth has an induced subgraph that is bipartite and which contains at least a* 0.86 *fraction of the vertices.*

Before being able to prove Theorem 2.1 and Theorem 2.2, we are going to need some preliminary results concerning the so-called Gaussian wave function and its approximation with factor of i.i.d. processes.

2.2 The Gaussian wave function and its properties

Existence of the Gaussian wave function. As alluded to before, we will first work on the infinite d-regular tree. Our aim is to choose an invariant process in a way that makes it feasible to calculate certain non-trivial probabilites. It turns out that a particular *Gaussian process* will be suitable.

Definition 2.3. A process X on T_d is a *Gaussian process* if for every $v \in V(T_d)$ the random variable X_v has zero mean and for any finite subset $v_1, \dots, v_n$ of $V(T_d)$ the joint distribution of $(X_{v_1}, .., X_{v_n})$ is a multivariate Gaussian. Furthermore, we say that a Gaussian process X is *Aut(T_d)-invariant* or simply *invariant* if for any $\phi \in \mathrm{Aut}(T_d)$ the process $(X_{\phi(v)})_{v \in V(T_d)}$ is equal in law to X.

Since the joint distribution determines the marginal distributions (and since T_d is a transitive graph), in an invariant Gaussian process each X_v must be identically distributed. On the other hand, they need not necessarily be independent.

The joint distribution of a Gaussian process is determined by the covariances $\mathrm{cov}(X_u, X_v)$, $u, v \in V(T_d)$. This, by the Kolmogorov extension theorem, is a consequence of the fact that a multivariate Gaussian is determined (in distribution) by its expected value and covariance matrix.

Note that for an invariant process these covariances only depend on the distance of u and v, which lets us define the *covariance sequence* corresponding to the process.

Definition 2.4. The covariance sequence of an invariant process X on T_d is the sequence $\sigma_0, \sigma_1, \ldots$, where $\sigma_k = \mathrm{cov}(X_v, X_u)$ for some vertices u, v with distance k in T_d.

By the preceding paragraph, whenever there exists an invariant Gaussian process with a given covariance sequence, it is unique.

Proposition 2.5. *For any real λ with $|\lambda| \leq d$ there exists a non-trivial invariant Gaussian process X on T_d such that with probability 1 the equation*

$$\sum_{u \in N(v)} X_u = \lambda X_v$$

holds for every vertex v, where $N(v) \subseteq V(T_d)$ is the neighbourhood set of v. If we also require the variance of each X_v to be 1, then the joint distribution of such a process is unique. We will refer to this process as the Gaussian wave function with eigenvalue λ.

Proof. Suppose for the moment that the process described in the proposition exists; then, multiplying by a constant, we get another such wave function, and so by choosing a suitable constant we may assume that each X_v has variance 1. This process, by the preceding discussion, is determined by the covariances $\sigma_0 = 1, \sigma_1, \ldots$, and, given any vertex v_0 and its neighbours $v_1, \ldots, v_d$ the equation $0 = X_{v_1} + \ldots + X_{v_d} - \lambda X_{v_0}$ holds with probability 1. Thus we get

$$0 = \mathrm{cov}(X_{v_0}, 0) = \mathrm{cov}(X_{v_0}, X_{v_1} + \ldots + X_{v_d} - \lambda X_{v_0}) = d\sigma_1 - \lambda\sigma_0. \tag{2.1}$$

Similarly, if the node u has distance k from v_0, then it has distance $k-1$ from one neighbour of v_0 and distance $k+1$ from the others, and so

$$0 = \mathrm{cov}(X_u, 0) = \mathrm{cov}(X_u, X_{v_1} + \ldots + X_{v_d} - \lambda X_{v_0}) = (d-1)\sigma_{k+1} + \sigma_{k-1} - \lambda\sigma_k \tag{2.2}$$

holds as well.

These equations, taken together with the fact that σ_0 equals 1, give a linear recurrence relation which the covariance sequence must satisfy. Let $\sigma_k, k \geq 0$ be the unique solution to this recurrence. In particular, $\sigma_1 = \lambda/d$ and $\sigma_2 = (\lambda^2 - d)/d(d-1)$. Then it suffices to construct an invariant Gaussian process on T_d with covariance sequence σ_k. We will define the process recursively for every finite subset of $V(T_d)$; the statement will then follow from the Kolmogorov extension theorem.

First we start with two neighbouring vertices u and v. Since the matrix

$$\Sigma = \begin{pmatrix} 1 & \sigma_1 \\ \sigma_1 & 1 \end{pmatrix}$$

is positive semidefinite for $|\lambda| \leq d$, there exists a bivariate Gaussian (X_u, X_v) with covariance matrix Σ and zero mean.

Now suppose that we have already defined $X_v, v \in S$ where S is the vertex set of a finite subtree $T_d[S]$ of T_d, such that for each $u, v \in S$, $\mathrm{cov}(X_u, X_v) = \sigma_{d(u,v)}$. Let v_0 be a leaf (a vertex with degree 1), with its unique neighbour in $T_d[S]$ denoted by v_d, and $d-1$ other neighbours $v_1, \ldots, v_{d-1}$ in T_d. We define X_{v_i} for $i = 1, \ldots, d-1$ as a linear combination of X_{v_0}, X_{v_d} and Y_i, where $(Y_1, \ldots, Y_{d-1})$ is a multivariate Gaussian independent from any of the already constructed X_v, with a covariance matrix to be specified later on.

That is, we set

$$X_{v_i} = \frac{\lambda}{d-1} X_{v_0} - \frac{1}{d-1} X_{v_d} + Y_i,\ i = 1, \ldots, d-1$$

A straightforward calculation shows that $\mathrm{cov}(X_{v_i}, X_u) = \sigma_{d(v_i,u)}$ for any $u \in S$. We also need to ensure that $\mathrm{var}(X_i) = \sigma_0$ and $\mathrm{cov}(X_i, X_j) = \sigma_2$ for $1 \leq i, j \leq d-1, i \neq j$. These conditions can be satisfied by suitably choosing $a = \mathrm{var}(Y_i)$ and $b = \mathrm{cov}(Y_i, Y_j)$: by solving the equations

$$\sigma_0 = \mathrm{var}(X_i) = \left(\frac{\lambda}{d-1}\right)^2 + \left(\frac{1}{d-1}\right)^2 - \frac{2\lambda}{(d-1)^2}\sigma_1 + a, \text{ and}$$

$$\sigma_2 = \mathrm{cov}(X_i, X_j) = \left(\frac{\lambda}{d-1}\right)^2 + \left(\frac{1}{d-1}\right)^2 - \frac{2\lambda}{(d-1)^2}\sigma_1 + b,$$

we get

$$a = \frac{(d-2)(d^2-\lambda^2)}{d(d-1)^2} \text{ and } b = \frac{\lambda^2 - d^2}{d(d-1)^2}.$$

All that remains to show is that there exists a multivariate centered Gaussian with these covariances; to this end, we have to establish that the covariance matrix of $(Y_1, \ldots, Y_{d-1})$ is positive semidefinite, or equivalently, that its eigenvalues are non-negative.

The covariance matrix can be written in the form $(a-b)I + bE$, where every entry of $E_{i,j}$ is 1, and I is the identity matrix. Then for any eigenvalue μ and corresponding eigenvector v,

$$\mu v = ((a-b)I + bE)v = (a-b)v + bv' \iff \frac{b-a+\mu}{b} v = v'$$

where $v'_i = \sum_{j=1}^{d-1} v_j$ for $1 \leq i \leq d-1$. This shows that either $\mu = a - b$ or otherwise v is the same in all coordinates, in which case $\mu = a + (d-2)b$. So the matrix is positive semidefinite precisely if $a \geq b$ and $a \geq -(d-2)b$. Both of these inequalities are satisfied whenever $|\lambda| \leq d$, which shows the existence of $(Y_1, \ldots, Y_{d-1})$.

We succeeded in defining the random variables X_v on a larger connected subset $S' = S \cup \{v_1, \ldots, v_{d-1}\}$. If we choose the leaves in the right order (e.g. first in, first

out) we can define X_v on any finite connected subset by iterating this procedure; so the statement follows from Kolmogorov's extension theorem.

□

In the following, we will state two propositions without proof, although with some explanation. The proofs can be found in (respectively) Section 3 and Section 2.1 of [1].

Approximation with factor of i.i.d. processes. The next proposition provides us with a critical tool; it allows us to approximate the Gaussian wave function in a way that can be "projected" onto finite regular graphs as well (provided that their girth is large enough). This is the reason why we can work on T_d, which, in a sense, has a simpler structure than arbitrary finite regular graphs.

Proposition 2.6. *For any real number λ with $|\lambda| \leq 2\sqrt{d-1}$ there exists a sequence of block factor of i.i.d. processes that converge in distribution to the Gaussian wave function on T_d with eigenvalue λ.*

Here, the factor of i.i.d. processes are meant to be factors of the i.i.d. labelling where each marginal is a standard normal variable. We note that in Chapter 3 we will prove Theorem 3.8, which implies as a corollary that this bound on $|\lambda|$ is sharp, and in fact the Gaussian wave function with eigenvalue λ cannot be approximated by (not necessarily block) factor of i.i.d. processes whenever $|\lambda| > 2\sqrt{d-1}$.

Percolation of the Gaussian wave function. One of our approaches to construct independent factor of i.i.d. sets is to choose a threshold τ and only leave the vertices with labels less than τ. This procedure (and also the resulting set) is called a *percolation* of T_d. As we increase or decrease the threshold, the probability that a given vertex will be part of the percolation changes accordingly. Our goal is to choose τ to be as large as possible, while making sure that the connected components in the resulting set are small with high probability. Intuitively, we want to find a balance between "throwing out" too many vertices and leaving the graph "too connected" (which makes it difficult to choose independent sets). The following proposition provides a suitable τ in the case of $d = 3$.

Proposition 2.7. *Let X be the Gaussian wave function on T_3 with eigenvalue $\lambda = -2\sqrt{2}$ and consider the set $S_\tau = \{v \in V(T_3) : X_v \leq \tau\}$. If $\tau \leq 0.086$, then each connected component of S_τ is finite almost surely.*

In fact, $\tau = 0.086$ is the value used to achieve the result stated in Theorem 2.1. The choice of a negative value for λ is sensible because it ensures that the labels on

neighbouring vertices have negative correlation (see Equation (2.1)), which decreases the probability of both of them being smaller than the threshold, thus making it less likely that the percolation will leave large connected components.

2.3 Constructing independent sets

In this section we will show how to find independent sets in 3-regular graphs with large girth using the tools of the previous sections. The procedures described in the following could be carried out on d-regular graphs for any $d \geq 3$, but the computations required for rigorously proving lower bounds quickly grow in difficulty as d increases. Even for $d = 4$ the method, in its current form, seems to be inefficient for finding theoretical bounds.

We will examine the two approaches outlined in the beginning of the chapter. In both cases we will give a procedure that, given a labelling of the vertices of the graph, constructs an independent set in a way that is invariant (with respect to the automorphisms of the graph) and measurable. Moreover, whether a given node is to be included in the set will only depend on the labels of the vertices in an N'-neighbourhood of the node, for a fixed integer N'. If we choose the labelling in a random manner, we obtain a random independent set in this way, and then the expected size of this set gives a lower bound on the size of the largest independent set.

The value of the results in the preceding section is that they allow us to work on T_3 and choose the Gaussian wave function corresponding to the eigenvalue $-2\sqrt{2}$, from now on denoted by X, as our random labelling. This works as follows: since both X and our procedure is invariant, and T_3 is transitive, the probability p that a given vertex will be an element of the resulting set is the same for all the vertices. Then, by Proposition 2.6, for any $\varepsilon > 0$ we can find a linear block factor of i.i.d. process Y such that if we replace X by Y and carry out the same procedure, the new probability of a node being chosen will be at least $p - \varepsilon$. Since Y is a block factor, for any $v \in V(T_3)$ the random variable Y_v only depends on an N-neighbourhood of v for some fixed N. Thus, whether a vertex is chosen is determined by its $(N + N')$-neighbourhood. As a consequence, we can carry out this procedure on any graph with girth at least $N + N'$ without changing the probability of a vertex being chosen.

To summarize, for any number $p' < p$ there exists a number r such that in any 3-regular graph with girth at least r the independence ratio is at least p'. As a result, our goal in the following is to maximize the probability p.

We recall that, by Equations (2.1) and (2.2), the first three terms of the covariance sequence for X are

$$\sigma_0 = 1;\ \sigma_1 = \frac{-2\sqrt{2}}{3};\ \sigma_2 = \frac{5}{6}.$$

First approach. We choose those vertices which have greater labels assigned to them than their neighbours. To calculate this probability, we will use the following identity from Proposition 1.1 that holds for any trivariate non-degenerate Gaussian (Z_1, Z_2, Z_3):

$$\mathbb{P}(Z_1 > 0, Z_2 > 0, Z_3 > 0) = \frac{1}{2} - \frac{1}{4\pi} \sum_{1 \le i < j \le 3} \arccos(\mathrm{corr}(Z_i, Z_j)). \tag{2.3}$$

Then if $v_0 \in T_3$ is an arbitrary vertex with neighbours v_1, v_2, v_3, and $Z_i = X_{v_0} - X_{v_i}, i = 1, 2, 3$, the correlations in question are

$$\mathrm{corr}(Z_i, Z_j) = \frac{\mathrm{cov}(X_{v_i} - X_{v_0}, X_{v_j} - X_{v_0})}{\sqrt{\mathrm{var}(X_{v_i} - X_{v_0})\mathrm{var}(X_{v_j} - X_{v_0})}} = \frac{\sigma_2 - 2\sigma_1 + \sigma_0}{2\sigma_0 - 2\sigma_1} = \frac{1 + 2\sqrt{2}}{4}.$$

Therefore

$$p = \mathbb{P}(v_0 \text{ is chosen}) = \frac{1}{2} - \frac{3}{4\pi} \arccos(\frac{1 + 2\sqrt{2}}{4}) = 0.42982...$$

So any 3-regular graph with large enough girth has an independent set that contains a 0.4298 fraction of the vertices.

Second approach. We fix a $\tau \in \mathbb{R}$ threshold and consider the connected components of the subgraph with vertex set $S = \{v : X_v \le \tau\} \subseteq V(T_3)$. If the size of a connected component is small (smaller than a number N'), then we choose an independent set from it containing at least half of its vertices in an invariant and measurable way; otherwise, we simply leave the component out. A suitable way of choosing is the following: since the subgraph is a tree, we can partition it into two independent sets in a unique way. If the size of the component is $2k - 1$ for some positive integer k, then we choose the larger one, that is, the one with size k. If the component contains an even number of vertices, we choose the set which contains the vertex with the largest label. This procedure is invariant and measurable, and the choice whether a node is to be included in the resulting set only depends on the labels in its N'-neighbourhood.

Let p_k be the probability that the size of the connected component of a given vertex in the subgraph $T_3[S]$ is k, with $k = 0$ meaning that the vertex is not in S (these probabilities are the same for all nodes). We can ensure that each component is finite almost surely by choosing τ to be under the critical threshold (see Proposition 2.7),

which means that $\sum_{k=0}^{\infty} p_k = 1$. Then the probability that a vertex is chosen by the procedure described in the preceding paragraph is

$$\sum_{k=1}^{N'} \frac{k}{2k-1} p_{2k-1} + \frac{1}{2} \sum_{k=1}^{N'} p_{2k}.$$

If $N' \longrightarrow \infty$, this converges to

$$p = \sum_{k=1}^{\infty} \frac{k}{2k-1} p_{2k-1} + \frac{1}{2}(1 - p_0 - \sum_{k=1}^{\infty} p_{2k-1}) = \frac{1}{2}(1 - p_0) + \sum_{k=1}^{\infty} \frac{1}{2(2k-1)} p_{2k-1}.$$

Our goal then is to calculate (or estimate) the probabilities p_{2k-1}. For the optimal result this is done by choosing τ to be as large as possible and then numerically computing the needed probabilities to high precision. However, for the threshold $\tau = 0$ one can obtain estimates without the use of a computer.

$\tau = 0$: In this case $p_0 = \mathbb{P}(X_v > 0) = 1/2$. Given a vertex v with neighbours (in T_3) v_1, v_2, v_3, the probability p_1 can be expressed as $\mathbb{P}(X_v < 0, X_{v_1} > 0, X_{v_2} > 0, X_{v_3} > 0)$. However, since X is a Gaussian wave function with a negative eigenvalue, whenever $X_{v_1}, X_{v_2}, X_{v_3} > 0$, $X_v < 0$ must also hold (with probability 1). So, using Proposition 1.1, we can write

$$p_1 = \mathbb{P}(X_{v_1} > 0, X_{v_2} > 0, X_{v_3} > 0) = \frac{1}{2} - \frac{3}{4\pi} \arccos(\mathrm{corr}(X_{v_1}, X_{v_2})),$$

where $\mathrm{corr}(X_{v_1}, X_{v_2}) = \sigma_2 = 5/6$. We can use the trivial estimate $p_{2k-1} > 0$ for $k \geq 2$ to arrive at the lower bound

$$\frac{1}{2} - \frac{3}{8\pi} \arccos(\frac{5}{6}) = 0.4300889...$$

which, at the time of the publication of [1] was the best lower bound proved without the use of a computer.

We can also carry out the same procedure for the vertices with labels larger than the threshold, which gives another independent set that is of the same size as the previous one but disjoint from it, thus proving Theorem 2.2.

$\tau = 0.086$: This is the largest value of τ that has been shown to produce only finite components almost surely (see Proposition 2.7). Note that this is not necessarily the optimal threshold value: the authors of [1] claim that by using $\tau = 0.12$ and $N' = 200$, computer simulations suggested that the probability of a vertex being chosen was above 0.438. Nonetheless, a rigorous lower bound required the finiteness of connected components.

Even this case relies heavily on computing certain integrals numerically, the details of which we will leave to [1]. We will only note that the lower bound of 0.4361, proving Theorem 2.1, was achieved by introducing the numbers p'_k, which denote the

probability that any given path of length k in T_3 will actually be a component after the percolation. Of course $p_1 = p'_1$; one can also see that $p_3 = 9p'_3$, since a component of three vertices must be a path, and for any given vertex there are 9 paths of length 3 containing it. For larger values of k, $c_k p'_k \leq p_k$, where c_k is the number of paths of length k containing a given vertex. It is the numbers p'_1, p'_3, p'_5 that the authors numerically compute (and then bound the error of the computation) to arrive at Theorem 2.1.

Chapter 3

Limitations of factor of i.i.d. algorithms

In the preceding chapter we have seen a method of finding independent sets in d-regular graphs with large girth using factor of i.i.d. processes. We noted that while this method works for all $d \geq 3$, computational difficulties make it infeasible to use it for proving theoretical lower bounds on the independence ratio of such graphs even for $d = 4$.

The case of large d has been studied as well in the context of factor of i.i.d. algorithms, which led to a remarkable result, though of a markedly different kind from that of the previous chapter: in 2013, Gamarnik and Sudan [2] proved a theorem pertaining to random d-regular graphs, which, in effect, shows that in that setting, for large enough d, the sort of tools that we have seen in the previous chapter will almost never produce independent sets close to the largest possible.

This shows that, while algorithms based on factor of i.i.d. processes have been used with success for finding various graph-theoretic structures, such as matchings [9], [10] and colourings [11], there are cases when the optimal result is not achievable using them. Thus the result settled, with a negative answer, a conjecture of Hatami, Lovász and Szegedy [12], predicting that these algorithms can provide optimal solutions for various common optimization problems on graphs (see Conjecture 2.4 in [2]).

Though it is not within the scope of this thesis to fully describe the (at times quite computation-heavy) proof of this result, in the first section of this chapter we provide a brief overview of the lines of thought in [2], and of the subsequent work of Rahman and Virág [4], who strengthened the result. The authors of [4] also introduce the concept of *entropy* in the context of random processes on graphs, and prove a general criterion that every factor of i.i.d. process must obey. We will expressly state this condition as it applies to processes T_d.

This leads us to the question of characterizing factor of i.i.d. processes on the infinite d-regular tree. While so far there is no known necessary and sufficient condition for deciding whether a process on T_d is a factor of an i.i.d. process, there have been various results in this area (see, e.g., [13], [14] and the references therein). In the second section of the chapter we prove a theorem from [3], giving us a necessary condition on pairwise correlations which the marginal distributions of such a process must obey. We will introduce the notion of *graphings*, and in particular the so-called Bernoulli graphing, which will make it possible for us to employ tools of functional analysis in our examination of factor of i.i.d. processes.

The aforementioned condition on entropy and the condition on correlations are both tools to decide whether a process can possibly be a factor of i.i.d. process. We note that they are not directly comparable in strength: there are examples of processes that obey one condition but fail the other, in both cases.

3.1 Non-optimality for independent sets

In this section we give the exact formulation of the statement proved by Gamarnik and Sudan, and then outline its proof. Before we can even state the main theorem, we need to introduce some notation and preliminary results.

Asymptotics of the independence ratio. In order to establish the non-optimality of factor of i.i.d. algorithms we have to first formalize the notion of optimality in this context. This has been made possible by results of Bayati, Gamarnik and Tetali [15], and Frieze and Łuczak [16]. Let $\alpha_{d,n}$ denote the independence ratio of a random d-regular graph on n vertices (for the pairs (n, d) when such a graph exists). The first of the aforementioned papers has shown that for any positive d, the sequence of random variables $\{\alpha_{d,n} : n > d, nd \text{ is even}\}$ converges in probability to some (non-random) number α_d. The second result, in turn, is that the sequence α_d for $d = 1, 2, \ldots$, is asymptotically $2\log d/d$. In effect, if d and n are large enough, then with high probability the independence ratio of $G_{d,n}$ is close to $2\log d/d$.

Factor of i.i.d. algorithms. We also need to clarify what we mean by a factor of i.i.d. algorithm. As in the previous chapter, we will consider factor of i.i.d. processes on the d-regular tree which construct independent sets, and then use a method of projecting these sets onto finite graphs. The authors of [2] use the rule function formulation of factor of i.i.d. processes. In the present case this means that we distinguish a vertex o of T_d, called the root, and we construct a random set on T_d using a rule function $f : [0, 1]^{V(T_d)} \to \{0, 1\}$ which is spherically symmetric around

the root (see Definition 1.4). In our case we also require f to have the property that for any neighbouring vertices u,v and any labelling $x \in [0,1]^{V(T_d)}$, at most one of $f(\phi_u x)$ and $f(\phi_v x)$ is equal to one, where ϕ_u and ϕ_v are automorphisms that map u and v respectively to the root, acting on $[0,1]^{V(T_d)}$ in the usual manner. We will call (after [2]) a decision function with this property an *independence function*. A factor of i.i.d. algorithm for constructing independent sets is, then, the procedure of putting i.i.d. labels, each with uniform distribution on $[0,1]$, on the vertices of T_d, and then using an independence function to generate an independent set.

We will measure the optimality of such an algorithm by the probability $\mathbb{P}(f(X) = 1) = \mathbb{E}(f(X))$, which we will denote by $\alpha(f)$; this is also referred to as the *density* of f. Actually, to bridge the gap between T_d and finite d-regular graphs we will consider independence functions which only depend on some finite neighbourhood of the root. However, this is not a real restriction in the sense that, by Proposition 1.7, any independence function can be approximated by a local independence function in a way that the density of the latter is arbitrarily close to the density of the former.

Given an r-local independence function on T_d we can construct an independent set in $G_{d,n}$, in the following way: we decorate the vertices of $G_{d,n}$ with i.i.d. labels, each having uniform distribution on $[0,1]$. Then if a vertex v has an $(r+1)$-neighbourhood isomorphic to a tree, we can apply f (taking v to be the root) to decide whether to include it in the set; we simply leave out the nodes whose $(r+1)$-neighbourhood is not a tree. It is easy to see that the resulting set $I_{G_{d,n}}$ is independent.

This procedure is very similar to the one used in the previous chapter, with the difference that we did not constrain ourselves to graphs with large girth. An immediate question is whether the density of an independence function is adequate in describing the size of the independent set it generates in $G_{d,n}$. It is clear that this method of projecting is only efficient if "most" of the vertices in the graph have tree-like neighbourhoods. This indeed holds, according to Proposition 2.2 of [2].

Proposition 3.1. *Let $B_r(G_{d,n})$ denote the number of vertices in $G_{d,n}$ which have r-neighbourhoods not isomorphic to a tree. Then for any d and r, the sequence of random variables $\frac{1}{n}B_r(G_{d,n})$ converges in probability to zero, as n goes to infinity.*

This immediately implies that $\mathbb{E}(B_r(G_{d,n})) = o(n)$: indeed, for any ε, if n is sufficiently large, then $\mathbb{P}(B_r(G_{d,n}) \geq n\varepsilon) \leq \varepsilon$, and so by separately estimating the expected value on $\{B_r(G_{d,n}) \leq \varepsilon\}$ and $\{B_r(G_{d,n}) > \varepsilon\}$ and using the fact that $B_r(G_{d,n}) \leq n$, we obtain

$$\frac{1}{n}\mathbb{E}(B_r(G_{d,n})) \leq \varepsilon + \varepsilon = 2\varepsilon.$$

Hence the expected relative size of the independent set generated by f is

$$\mathbb{E}\left(\frac{|I_{G_{d,n}}|}{n}\right) = \frac{1}{n}\mathbb{E}\left(\sum_{v\in V(G_{d,n})} I(v \text{ is chosen})\right) = \frac{1}{n}\alpha(f)(n-\mathbb{E}(B_{r+1}(G_{d,n}))) = \alpha(f)(1-o(1)),$$

where the second equality is obtained by interpreting the sum of indicator variables as a random sum of $B_r(G_{d,n})$ i.i.d. variables, and then using the fact that $\mathbb{E}(\sum_{i=1}^{N} X_i) = \mathbb{E}(N)\mathbb{E}(X_i)$, for i.i.d. variables X_i which are independent from N as well.

This shows that the expected size of the "projected" set is close to the density, for sufficiently large n. Thus we can define the number $\hat{\alpha}_d = \sup_r \sup\{\alpha(f_r) : f_r \text{ is an } r\text{-local independence function}\}$ to describe, in some sense, the largest size of independent sets that can be constructed using the factor of i.i.d. algorithms in question. Note that, as mentioned before, Proposition 1.7 implies that $\hat{\alpha}_d = \sup\{\alpha(f) : f \text{ is an independence function}\}$.

The results of Gamarnik and Sudan. Now we are in the position to state the main result of [2] (Theorem 2.5 in the paper):

Theorem 3.2. *For every $\varepsilon > 0$ and for all sufficiently (depending on ε) large d,*

$$\frac{\hat{\alpha}_d}{\alpha_d} \leq \frac{1}{2} + \frac{1}{2\sqrt{2}} + \varepsilon.$$

That is, if d is large enough, then for any factor of i.i.d. algorithm, with probability tending to 1 as n goes to infinity, the expected size of the independent set in a random d-regular graph $G_{d,n}$ given by the algorithm is a multiplicative factor smaller than the largest independent set in the graph.

The proof in [2] has two main building blocks, both of which consists of examining overlaps of independent sets: the first (Theorem 2.6 in the paper) shows that independent sets in random regular graphs exhibit a "clustering" phenomenon, that is, for any pair of large enough independent sets, they either intersect almost nowhere or almost everywhere. To put it in a more quantitative manner, let β and z be chosen such that $1/\sqrt{2} < \beta < 1$ and $0 < z < \sqrt{2\beta^2 - 1}$, and denote by $K(z)$ the integers between $\frac{(1-z)n\log d}{d}$ and $\frac{(1+z)n\log d}{d}$. Then the probability that there exist two independent sets in a random graph $G_{d,n}$ with size at least $n\frac{(1+\beta)\log d}{d}$ such that the size of their intersection is in $K(z)$ goes to zero as n tends to infinity.

As $\beta \to 1$, z can be chosen arbitrarily close to 1 as well, and so we can conclude that in most large enough regular graphs, the intersection of two independent sets that are close to optimal in size is either negligible or very large.

On the other hand, Theorem 2.7 of the paper asserts that this is not the case for independent sets generated by factor of i.i.d. algorithms. They use a technique of

coupling random variables so that their correlation can be controlled by a parameter $p \in [0, 1]$. The idea of this is simple: given two labellings X and Y of the graph G, both with uniform distribution on $[0, 1]^n$ and independent from each other, construct the random variable Z such that $Z_u = X_u$ with probability p, and $Z_u = Y_u$ otherwise, independently for every vertex $u \in V(G)$.

Now for any independence function f with density $\alpha(f)$ we can examine the independent sets I_X and I_Z, generated by f with labels X and Z respectively. For $p = 1$, X is equal to Z, and so the expected size of the intersection of I_X and I_Z is the expected size of the sets themselves, which is $\alpha(f)n + o(n)$. On the other hand, for $p = 0$, X and Z are independent, in which case it is not hard to see that the expected size of the intersection is $\alpha(f)^2 n + o(n)$ (essentially, if a large enough neighbourhood around a vertex v is a tree, then the choice to include v in I_X is independent from the choice to include it in I_Z, and the probability of both is $\alpha(f)$).

Theorem 2.7 of the paper then states that any intermediate intersection size can be produced with the right choice of p. In fact, it is shown that the probability that a vertex is chosen under both labellings depends continuously on the parameter p, so any value between $\alpha(f)^2$ and $\alpha(f)$ can be achieved.

Intuitively, these two propositions imply that factor of i.i.d. independent sets can not be close to optimal in size, and indeed, the proof of Theorem 3.2 depends mainly on their appropriate use.

The result of Rahman and Virág. In their paper, Gamarnik and Sudan conjectured that the multiplicative factor in Theorem 3.2 can be decreased to $1/2$, claiming that the extra factor of $1/2\sqrt{2}$ is just an artifact of analysis. In 2015, Rahman and Virág published [4], which affirms this conjecture. We only note here that their method is to consider a sequence of independent and identically distributed labellings $X_i, i \geq 0$, and the sets I_i generated by a factor f with the labels being the p-coupling of X_0 and X_i, as described earlier. They then derive inequalities for the intersection densities of any k of them, in contrast with [2], where only the case $k = 2$ is analyzed. The main result is then obtained from the careful analysis of these inequalities.

The inequalities, in turn, depend crucially on a condition that the so called *entropy* of the local algorithm in question must satisfy. We conclude this section by stating this condition in the context of processes on the infinite d-regular tree, as formulated in another paper by Rahman [17].

Definition 3.3. The entropy of a discrete probability distribution μ is

$$H(\mu) = \sum_{x \in \text{support}(\mu)} -\mu(x) \log \mu(x).$$

Theorem 3.4. *Let X be a factor of i.i.d. process on $T_d, d \geq 3$, taking values from $\{0,1\}^{V(T_d)}$. Moreover, let $u, v \in V(T_d)$ be two neighbouring vertices, and denote by P_X and π_X the distributions of (X_u, X_v) and X_u, respectively. Then the following holds:*

$$\frac{d}{2}H(P_X) - (d-1)H(\pi_X) \geq 0.$$

3.2 Correlation decay

The results in this section can be seen as an example of exploiting the highly symmetric nature of the d-regular tree to gather information about random processes on it. To elaborate, we will use the symmetry of T_d to construct a special graph named the Bernoulli graphing (denoted by B_d) and a corresponding operator on the Hilbert space $L^2(\Omega_d)$, where $\Omega_d = V(B_d)$. We will also describe a way to transform a rule function f of a factor of i.i.d. process on the tree into a function $\tilde{f}$ on Ω_d in an integral-preserving way. If f is square-integrable (that is, the factor of i.i.d. process has marginals with finite variance), then the transformed function $\tilde{f}$ will be in $L^2(\Omega_d)$, and this will allow us to use tools of functional analysis to prove a theorem about the correlations between the marginals of factor of i.i.d. processes.

First, we give the general definition of the term *graphing*, a concept which arises in the theory of sparse graph limits (as in [12]).

Definition 3.5. Let X be a Polish (i.e. separable, completely metrizable) topological space with a measure ν on its Borel sets. A graphing is a graph $\mathcal{G}$ on $V(\mathcal{G}) = X$ with bounded maximal degree and Borel measurable edge set $E(G) \subseteq X \times X$ such that for all measurable sets $A, B \subseteq X$ we have

$$\int_A e(x, B) d\nu(x) = \int_B e(x, A) d\nu(x),$$

where $e(x, S)$ is the number of edges from $x \in X$ to $S \subseteq X$.

Given a graphing $\mathcal{G}$ we can view it as an operator on the space of functionals from its vertex set to the complex numbers by the following definition: for $f : X \to \mathbb{C}$ let $\mathcal{G}f$ be the function

$$\mathcal{G}f(x) = \sum_{(x,v) \in E(G)} f(v), \quad \text{for } x \in X.$$

Thus we can view $\mathcal{G}$ as linear operator on $L^2(X)$. It turns out that the integral condition in the definition of graphings is equivalent to the self-adjointness of this operator:

Proposition 3.6. *Let X and ν be as before, and let $\mathcal{G}$ be a graph on $V(\mathcal{G}) = X$ with bounded maximal degree and Borel measurable edge set $E(G) \subseteq X \times X$. Then $\mathcal{G}$ is a graphing if and only if the corresponding operator on $L^2(X)$ is self-adjoint.*

Proof. Let A, B be measurable subsets of X with corresponding indicator functions I_A, I_B. The crucial observation is that $\mathcal{G}I_B(x) = e(x, B)$ for all $x \in X$, which is clear from the relevant definitions. Then we have

$$\langle I_A, \mathcal{G}I_B \rangle = \int_X I_A \cdot \mathcal{G}I_B \, d\nu = \int_A \mathcal{G}I_B \, d\nu = \int_A e(x, B) \, d\nu(x).$$

Similarly,

$$\langle \mathcal{G}I_A, I_B \rangle = \int_X \mathcal{G}I_A \cdot I_B \, d\nu = \int_B \mathcal{G}I_A \, d\nu = \int_B e(x, A) \, d\nu(x).$$

This shows that self-adjointness implies that $\mathcal{G}$ is a graphing, and the reverse implication is true for indicator functions. But self-adjointness on the indicator functions implies self adjointness on any function from $L^2(X)$, since by linearity it extends to simple functions, and then the fact that any function in $L^2(X)$ is a limit of such functions, together with the continuity of the inner product, finishes the proof. □

If the graphing $\mathcal{G}$ is d-regular, then the constant functions are eigenfunctions of $\mathcal{G}$ (with eigenvalue d). It is easy to see that orthogonal complement $L^2_0(X)$ of functions with integral zero is an invariant subspace with respect to $\mathcal{G}$; we will, following [3], denote the norm of the graphing on this subspace by $\varrho(\mathcal{G})$.

We now construct the (d-regular) *Bernoulli graphing*, denoted by B_d. Consider the probability space $([0,1]^{V(T_d)}, \mathcal{B}, \lambda)$, that is, the space of labellings of the rooted d-regular tree with the product σ-algebra $\mathcal{B}$ and the product measure λ (taking the Lebesgue measure on $[0,1]$ for each vertex v). We construct an equivalence relation on this space by saying $x \sim^* y$ if and only if $x = \phi y$, where ϕ is a root-preserving automorpishm of T_d. We define $\Omega_d = [0,1]^{V(T_d)}/\sim^*$ to be the vertex set of B_d. We can define a sigma-algebra on Ω_d with the canonical quotient map $q : [0,1]^{V(T_d)} \to \Omega_d$ by $\mathcal{A} = \{A \subseteq Y : q^{-1}(A) \in \mathcal{B}\}$. Then $(\Omega_d, \mathcal{A}, \nu)$ is a probability space, with $\nu = \lambda \circ q^{-1}$ being the pushforward measure. We connect two elements $[x], [y]$ of Ω_d if for the corresponding labellings x, y there is an automorphism ϕ such that ϕ maps the root to one of its neighbours and $x = \phi y$. This is easily seen to be well defined and symmetric. We also note that this implies that two vertices are of distance k in B_d if and only if there is an automorphism that maps one of the corresponding labellings into the other and the root into a vertex that is of distance k in T_d. It is not difficult to see that almost all vertices of B_d have a connected component isomorphic to T_d (in the sense that the set of such vertices has measure 1).

It can be shown that the d-regular graph B_d constructed this way is a graphing (see [12]). We will use Theorem 2 of [3], stating that $\varrho(\mathcal{B}_d) \leq 2\sqrt{d-1}$ for $d \geq 3$. The authors of said paper call a d-regular graphing *Ramanujan* whenever it obeys this inequality.

The following discussion will establish the connection between factor of i.i.d. processes and the Bernoulli graphing. Let f be the rule function of some factor of i.i.d. process on T_d. The fact that f is spherically symmetric around the root allows us to define the function $\tilde{f}$ by setting $\tilde{f}([x]) = f(x)$, for $x \in [0,1]^{V(T_d)}$. In other words, $f = \tilde{f} \circ q$. We also have that for any measurable set $B \in \mathcal{B}$, if the integral of f on B is finite, then the following holds:

$$\int_B f \, d\lambda = \int_B \tilde{f} \circ q \, d\lambda = \int_{q(B)} \tilde{f} d(\lambda \circ q^{-1}) = \int_{q(B)} \tilde{f} \, d\nu,$$

the second equality being a general identity for Lebesgue integrals (for measurable q). In words, the operation $f \to \tilde{f}$, defined for functions that are spherically symmetric around the root, preserves the integral.

The last tool we need is the following result from the domain of functional analysis. A proof can be found in [18] (Theorem 9.9-2).

Proposition 3.7. *Let $T : H \to H$ be a bounded self-adjoint linear operator on a complex Hilbert space H. Then for any polynomial p with real coefficients,*

$$\|p(T)\| \leq \max\{|p(x)| : x \in [-\|T\|, \|T\|]\}.$$

Now we are ready to state and prove the main theorem of this section.

Theorem 3.8. *Let X be a factor of i.i.d. process on the rooted d-regular tree, $d \geq 3$, generated by the rule function $f : [0,1]^{V(T_d)} \to \mathbb{R}$. Let us also assume that X_o has finite variance. Then for any vertices u, v of T_d with distance k, the following holds:*

$$|\mathrm{corr}(X_u, X_v)| \leq \left(k + 1 - \frac{2k}{d}\right)\left(\frac{1}{\sqrt{d-1}}\right)^k.$$

Proof. Since X is an invariant process, the correlation between two of its marginals only depends on their distance, so we can assume that one of the vertices in question is the root o. Furthermore, the marginals of X are identically distributed, and

$$\mathrm{corr}(X_o, X_v) = \mathrm{corr}\left(\frac{X_o - \mathbb{E}X_o}{\sqrt{\mathrm{var}(X_o)}}, \frac{X_v - \mathbb{E}X_v}{\sqrt{\mathrm{var}(X_v)}}\right),$$

so we can assume them to have zero mean and unit variance (i.e. we can replace f with $\frac{f - \mathbb{E}X_o}{\sqrt{\mathrm{var}(X_o)}}$).

The proof consists of two steps. First, we show that the correlation of the marginals can be expressed as the inner product $\langle \tilde{f}, \frac{B_d^{(k)}\tilde{f}}{d(d-1)^{k-1}} \rangle$, where $B_d^{(k)}$ is a graphing on Ω_d obtained by connecting two elements whenever their distance is k in B_d. Then we bound this inner product by expressing the operator $B_d^{(k)}$ as a polynomial of B_d and applying Proposition 3.7. We note that we will actually not use the fact that $B_d^{(k)}$ is a graphing, though it will follow from expressing the corresponding operator as a polynomial of B_d, since this shows its self-adjointness.

Step one: In the following we fix a positive integer k, and write $u \sim v$ whenever $u, v \in V(T_d)$ have distance k. Since we assumed zero mean and unit variance, we have that $\mathrm{corr}(X_o, X_v) = \mathrm{cov}(X_o, X_v) = \mathbb{E}X_oX_v$. Since this covariance only depends on the distance of o and v, we can express it as the average of the covariances $\mathrm{cov}(X_o, X_u)$, for $u \sim o$. Also we recall the fact that X_o has the same distribution as f viewed as a random variable on the probability space $([0,1]^{V(T_d)}, \mathcal{B}, \lambda)$. It follows that

$$\mathrm{corr}(X_o, X_v) = \frac{1}{d(d-1)^{k-1}} \sum_{o\sim u} \mathbb{E}X_oX_u = \frac{1}{d(d-1)^{k-1}} \int f(x) \cdot \sum_{o\sim u} f(\phi_u x)\, d\lambda(x),$$

with ϕ_u being an automorphism such that $\phi_u(o) = u$, for every $u \sim x$. The function $g(x) = \sum_{o\sim u} f(\phi_u x)$ is spherically symmetric around the root. This implies that last term is equal to

$$\frac{1}{d(d-1)^{k-1}} \int_{\Omega_d} \tilde{f} \cdot \tilde{g}\, d\nu.$$

Our goal is to show that $\tilde{g} = B_d^{(k)}\tilde{f}$, as this will imply that the above integal is equal to the claimed inner product. If $[x]$ and $[y]$ are neighbouring vertices of $B_d^{(k)}$, then there is a T_d-automorphism ϕ such that $o \sim \phi(o)$ and $\phi x = y$. This means that for $u = \phi(o)$, the ϕ_u in the definition of g can be taken to be this ϕ. In fact, with probability one (whenever no two coordinates of x are equal) there is a one to one correspondance between neighbours of $[x]$ in $B_d^{(k)}$ and vertices u that are of distance k from x. That means we have, for almost all $[x] \in \Omega_d$,

$$\tilde{g}([x]) = g(x) = \sum_{o\sim u} f(\phi_u x) = \sum_{o\sim u} \tilde{f}([\phi_u x]) = \sum_{([x],x')\in E(B_d^{(k)})} \tilde{f}(x') = (B_d^{(k)}\tilde{f})([x]),$$

as desired.

Step two: For any bounded linear operator T on a Hilbert space and an arbitrary element of the space x, we have $|\langle Tx, x\rangle| \leq \|Tx\| \cdot \|x\| \leq \|T\| \cdot \|x\|^2$, using the Cauchy-Schwarz inequality and the definition of the operator norm. In our case, since $\tilde{f} \in L_0^2(\Omega_d)$ and $\|\tilde{f}\| = 1$, this implies that it's sufficient for us to bound $\varrho\left(\frac{B_d^{(k)}}{d(d-1)^{k-1}}\right)$ from above.

Let U_n for $n \geq 0$ denote the Chebyshev polynomials of the second kind, i.e. $U_0(x) = 1, U_1(x) = 2x$ and for $n \geq 2, U_n(x) = 2xU_{n-1}(x) - U_{n-2}(x)$; also, let $U_{-1} \equiv 0$. We define the polynomials

$$q_n(x) = \sqrt{\frac{d-1}{d}} U_n(x) - \frac{1}{\sqrt{d(d-1)}} U_{n-2}(x), \text{ for } n \geq 1.$$

Note that each q_n is a linear combination of Chebyshev polynomials, thus they satisfy the same recurrence relation.

We claim that

$$\frac{B_d^{(k)}}{d(d-1)^{k-1}} = \frac{1}{\sqrt{d(d-1)^{k-1}}} q_k \left(\frac{B_d}{2\sqrt{d-1}} \right).$$

To show this, the key observation is that the operators $B_d^{(k)}$ satisfy the following recurrence relation:

$$\begin{cases} B_d^{(1)} = B_d, B_d^{(2)} = B_d^2 - dI, \\ B_d^{(k+1)} = B_d B_d^{(k)} - (d-1) B_d^{(k-1)}, k \geq 2. \end{cases}$$

The first line is self-explanatory, and the second can be seen by noting that for a function $g \in L_0^2(X)$ and $x \in X$, we have

$$(B_d(B_d^{(k)} g))(x) = \sum_{(x,v) \in E(B_d)} B_d^{(k)} g(v) = B_d^{(k+1)} g(x) + (d-1) B_d^{(k-1)} g(x),$$

since with probability 1 the connected component of a vertex in B_d is isomorphic to T_d, and so a vertex that is of distance k from a neighbour of x is either of distance $k+1$ or $k-1$ from x. A straightforward calculation shows that the claimed identity holds for $k = 1, 2$. The claim then follows by induction, since

$$\begin{aligned} \frac{1}{\sqrt{d(d-1)^k}} q_{k+1} \left(\frac{B_d}{2\sqrt{d-1}} \right) &= \frac{1}{\sqrt{d(d-1)^k}} \left[\frac{B_d}{\sqrt{d-1}} q_k \left(\frac{B_d}{2\sqrt{d-1}} \right) - q_{k-1} \left(\frac{B_d}{2\sqrt{d-1}} \right) \right] \\ &= \frac{B_d}{d-1} \cdot \frac{B^{(k)}}{d(d-1)^{k-1}} - \frac{1}{d-1} \cdot \frac{B_d^{(k-1)}}{d(d-1)^{k-2}} \\ &= \frac{1}{d(d-1)^k} \left(B_d B_d^{(k)} - (d-1) B_d^{(k-1)} \right) \\ &= \frac{B_d^{(k+1)}}{d(d-1)^k}, \end{aligned}$$

using the induction hyphotesis twice in the second line and the recursion for $B_d^{(k+1)}$ in the fourth We note that this part of the proof is adapted from the proof of Theorem 1.1 in [19].

Using this claim, Proposition 3.7 and the fact that $\varrho(B_d) \leq 2\sqrt{d-1}$ (Theorem 2.2 in [3]), we obtain the upper bound

$$\varrho\left(\frac{B_d^{(k)}}{d(d-1)^{k-1}}\right) \leq \max_{x\in[-1,1]} \frac{|q_k(x)|}{\sqrt{d(d-1)^{k-1}}}$$

The last step then is to bound $|q_k(x)|$. This can be done by observing that

$$\sqrt{d(d-1)}q_k(x) = (d-2)U_k(x) + (U_k(x) - U_{k-2}(x)) = (d-2)U_k(x) + 2T_k(x),$$

where $T_k(x)$ is the k-th Chebyshev polynomial of the first kind, defined by the same recurrence relations as U_k with the difference that $T_1(x) = 1, T_2(x) = x$. Both U_k and T_k have their extrema in the given interval in -1 and 1, with $U_k(1) = (-1)^k U_k(-1) = k+1$ and $T_k(1) = (-1)^k T_k(-1) = 1$ (these well known facts can be easily proved by induction). It follows that the maximum of $|q(k)|$ on the interval is in 1 as well, with

$$|q(1)| = \sqrt{\frac{d}{d-1}}\left(k+1-\frac{2k}{d}\right),$$

after rearranging. Summarizing,

$$|\mathrm{cov}(X_o, X_v)| = \left|\left\langle \tilde{f}, \frac{B_d^{(k)}\tilde{f}}{d(d-1)^{k-1}}\right\rangle\right| \leq \varrho\left(\frac{B_d^{(k)}}{d(d-1)^{k-1}}\right) \leq \left(\frac{1}{\sqrt{d-1}}\right)^k \left(k+1-\frac{2k}{d}\right),$$

which concludes the proof. □

Bibliography

[1] E. Csóka, B. Gerencsér, V. Harangi, and B. Virág, "Invariant Gaussian processes and independent sets on regular graphs of large girth", *Random Structures Algorithms*, vol. 47, no. 2, pp. 284–303, 2015.

[2] D. Gamarnik and M. Sudan, "Limits of local algorithms over sparse random graphs [extended abstract]", in *ITCS'14—Proceedings of the 2014 Conference on Innovations in Theoretical Computer Science*, ACM, New York, 2014, pp. 369–375.

[3] Á. Backhausz, B. Szegedy, and B. Virág, "Ramanujan graphings and correlation decay in local algorithms", *Random Structures Algorithms*, vol. 47, no. 3, pp. 424–435, 2015.

[4] M. Rahman and B. Virág. (Feb. 2014). Local algorithms for independent sets are half-optimal. Preprint. arXiv: `1402.0485 [math.PR]`.

[5] B. Bollobás, "The independence ratio of regular graphs", *Proc. Amer. Math. Soc.*, vol. 83, no. 2, pp. 433–436, 1981.

[6] C. Hoppen and N. Wormald. (Aug. 2013). Local algorithms, regular graphs of large girth, and random regular graphs. Preprint. arXiv: `1308.0266 [math.CO]`.

[7] E. Csóka. (Feb. 2016). Independent sets and cuts in large-girth regular graphs. Preprint. arXiv: `1602.02747 [math.CO]`.

[8] F. Kardoš, D. Král, and J. Volec, "Fractional colorings of cubic graphs with large girth", *Siam J. Discrete Math.*, vol. 25, no. 3, pp. 1454–1476, 2011.

[9] R. Lyons and F. Nazarov, "Perfect matchings as IID factors on non-amenable groups", *European J. Combin.*, vol. 32, no. 7, pp. 1115–1125, 2011.

[10] E. Csóka and G. Lippner. (Nov. 2012). Invariant random matchings in Cayley graphs. Preprint. arXiv: `1211.2374 [math.CO]`.

[11] C. T. Conley, A. S. Marks, and R. Tucker-Drob. (Jan. 2016). Brooks's theorem for measurable colorings. Preprint. arXiv: `1601.03361 [math.LO]`.

[12] H. Hatami, L. Lovász, and B. Szegedy, "Limits of locally-globally convergent graph sequences", *Geom. Funct. Anal.*, vol. 24, no. 1, pp. 269–296, 2014.

[13] R. Lyons. (Jan. 2014). Factors of IID on Trees. Preprint. arXiv: 1401.4197 [math.DS].

[14] Á. Backhausz and B. Szegedy. (Jun. 2014). On large girth regular graphs and random processes on trees. Preprint. arXiv: 1406.4420 [math.PR].

[15] M. Bayati, D. Gamarnik, and P. Tetali, "Combinatorial approach to the interpolation method and scaling limits in sparse random graphs", *Ann. Probab.*, vol. 41, no. 6, pp. 4080–4115, 2013.

[16] A. M. Frieze and T. Łuczak, "On the independence and chromatic numbers of random regular graphs", *J. Combin. Theory Ser. B*, vol. 54, no. 1, pp. 123–132, 1992.

[17] M. Rahman. (Oct. 2014). Factor of iid percolation on trees. Preprint. arXiv: 1410.3745 [math.PR].

[18] E. Kreyszig, *Introductory Functional Analysis with Applications*. Wiley, 1989.

[19] N. Alon, I. Benjamini, E. Lubetzky, and S. Sodin, "Non-backtracking random walks mix faster", *Commun. Contemp. Math.*, vol. 9, no. 4, pp. 585–603, 2007.