Statistical inference on random graphs and networks

Andressa Cerqueira

TESE APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTEÇÃO DO TÍTULO
DE
DOUTOR EM CIÊNCIAS

Programa: Estatística
Orientadora: Prof.ª Dr.ª Florencia Leonardi

Durante o desenvolvimento desta tese a autora recebeu apoio financeiro da CAPES e FAPESP, processo 2014/23526-0, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). Esta tese foi produzida como parte das atividades do Centro de pesquisa, inovação e difusão em Neuromatemática, processo 2013/07699-0, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

São Paulo, 24 de Abril de 2018
Statistical inference on random graphs and networks

Esta versão da tese contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 28/02/2018. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof.ª, Dr.ª. Florencia Graciela Leonardi (orientadora) - IME-USP
- Prof.ª. Dr.ª. Nancy Lopes Garcia - UNICAMP
- Prof. Dr. Cristian Favio Coletti - UFABC
- Prof. Dr. Roberto Imbuzeiro Moraes Felinto de Oliveira - IMPA
- Prof. Dr. Miguel Natalio Abadi - IME-USP
As opiniões, hipóteses e conclusões ou recomendações expressas neste material são de responsabilidade da autora e não necessariamente refletem a visão da FAPESP e da CAPES.
"Um dia de chuva é tão belo como um dia de sol.
Ambos existem; cada um como é."

Fernando Pessoa
Agradecimentos

Em primeiro lugar, eu gostaria de agradecer aos meus pais Rogério e Elizabette pelo apoio dado à mim durante todos os meus anos dedicados a pós-graduação. Também agradeço meu irmão Carlos pelos conselhos e ensinamentos dignos de um irmão mais velho.

Gostaria de agradecer à minha orientadora Florencia Leonardi por todas as reuniões construtivas que tivemos durante meu doutorado e pelo apoio dado para a realização do meu estágio de pesquisa no exterior. Você é em quem me inspirei como pesquisadora.

Quero agradecer Aurélien Garivier que foi meu supervisor durante o período de doutorado sanduíche na Universidade de Toulouse pelas discussões que contribuíram para o meu aprendizado e amadurecimento e que deram origem ao conteúdo da segunda parte dessa tese. Referente ao mesmo período eu gostaria de agradecer aos meus colegas de escritório Hugo, Laure e Guillaume e todos aqueles com quem convivi durante minha estadia na França.

Agradeço a todos os meus amigos do IME e aos amigos com quem compartilho o local de trabalho no NUMEC. Em especial, quero agradecer aos meus amigos Ana Paula, William, Aline, Guilherme Ost pelas experiências vividas dentro e fora do IME.

Quero deixar registrado os meus agradecimentos ao Guilherme pelo companheirismo, apoio e amor durante todos os anos de doutorado. Obrigada por estar presente mesmo estando distante fisicamente.

Por fim, quero agradecer ao Antonio Galves pelas oportunidades concedidas a mim durante todo o meu período trabalhando no NUMEC e por sua confiança em meu trabalho.

Eu quero agradecer ao apoio financeiro da CAPES e FAPESP recebido para o desenvolvimento dessa tese, referente a bolsa que foi concedida no âmbito do convênio FAPESP/CAPES com processo número 2014/23526-0, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).
vi
Resumo

CERQUEIRA, A. Inferência Estatística para grafos aleatórios e redes . 2018. 83 f. Tese (Doutorado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2018.

Nessa tese estudamos dois modelos probabilísticos definidos em grafos: o modelo estocástico por blocos e o modelo de grafos exponenciais. Dessa forma, essa tese está dividida em duas partes. Na primeira parte nós propomos um estimador penalizado baseado na mistura de Krichevsky-Trofimov para o número de comunidades do modelo estocástico por blocos e provamos sua convergência quase certa sem considerar um limitante conhecido para o número de comunidades. Na segunda parte dessa tese nós abordamos o problema de simulação perfeita para o modelo de grafos aleatórios Exponenciais. Nós propomos um algoritmo de simulação perfeita baseado no algoritmo Coupling From the Past usando a dinâmica de Glauber. Esse algoritmo é eficiente apenas no caso em que o modelo é monotóno e nós provamos que esse é o caso para um subconjunto do espaço paramétrico. Nós também propomos um algoritmo de simulação perfeita baseado no algoritmo Backward and Forward que pode ser aplicado a modelos monótonos e não monótonos. Nós provamos a existência de um limitante superior para o número esperado de passos de ambos os algoritmos.

Palavras-chave: modelo estocástico por blocos, estimação, Krichevsky-Trofimov, grafos aleatórios exponenciais, simulação perfeita, coupling from the past, backward and forward algorithm.
Abstract

In this thesis we study two probabilistic models defined on graphs: the Stochastic Block model and the Exponential Random Graph. Therefore, this thesis is divided in two parts. In the first part, we introduce the Krchevsky-Trofimov estimator for the number of communities in the Stochastic Block Model and prove its eventual almost sure convergence to the underlying number of communities, without assuming a known upper bound on that quantity. In the second part of this thesis we address the perfect simulation problem for the Exponential random graph model. We propose an algorithm based on the Coupling From The Past algorithm using a Glauber dynamics. This algorithm is efficient in the case of monotone models. We prove that this is the case for a subset of the parametric space. We also propose an algorithm based on the Backward and Forward algorithm that can be applied for monotone and non monotone models. We prove the existence of an upper bound for the expected running time of both algorithms.

Keywords: Stochastic Block Model, estimation, Krchevsky-Trofimov, Exponential Random Graph, perfect simulation, coupling from the past, backward and forward algorithm.
Contents

1 Introduction 1
  1.1 Objectives .............................................. 3
  1.2 Contributions ........................................... 4
  1.3 Organization of the work ............................... 5

I Stochastic Block Model 7

2 Stochastic Block Model 9
  2.1 Random Graph Definitions ............................. 9
  2.2 Stochastic Block Model ................................. 10
  2.3 Inferential Problems .................................... 13
    2.3.1 Estimation of the number of communities ....... 14
    2.3.2 The KT mixture distribution ..................... 15
  2.4 Computation of KT ..................................... 21

3 The KT order estimator 23
  3.1 Order estimator ........................................ 23
  3.2 Proof of the consistency theorem .................... 24
    3.2.1 Non-overestimation ............................... 24
    3.2.2 Non-underestimation .............................. 28

II Exponential Random Graphs 37

4 Exponential Random Graphs 39
  4.1 Exponential Random Graph ............................. 39
    4.1.1 Definitions and Notations ....................... 39
  4.2 Glauber dynamics ..................................... 41

5 Coupling from the past 45
  5.1 The algorithm .......................................... 45
  5.2 Convergence Speed .................................... 48
6 Backward and Forward algorithm
  6.1 The algorithm ...................................................... 53
  6.2 Kalikow-type decomposition ..................................... 53
  6.3 Backward and Forward algorithms ............................... 62
  6.4 Construction of the process ..................................... 65

7 Simulations
  7.1 The algorithm ..................................................... 71

8 Future directions
  8.1 Final Considerations and related problems .................... 75
     8.1.1 Stochastic Block Model .................................... 75
     8.1.2 Exponential Random Graphs ................................. 76

A Proofs ................................................................. 79

Bibliography ............................................................. 81
Chapter 1

Introduction

In recent years there has been an increasing interest in the study of probabilistic models defined on graphs. One of the main reasons for this interest is that these models are very flexible, making them suitable for the description of real datasets, for instance in social networks, computational networks, biological networks and neural networks. The interactions in a real network can be represented through a graph, where the objects of study are represented by the vertices of the graph and the interactions between these components are identified as the edges of the graph.

In order to describe the random interactions in a system many random graphs models have been vastly studied; however two among all random graphs models will be the focus of this thesis: the Stochastic Block model and the Exponential random graph model.

In some real networks one can observe the presence of communities, where the nodes of the network are split into groups with similar connection patterns. The Stochastic Block Model (SBM) was introduced by Holland et al. (1983) and it is vastly studied in the literature for community detection on random networks. In this model, each node in the network has associated a latent discrete random variable describing its community label, and given two nodes, the presence of a connection between them depends only on the values of the nodes’ latent variables.

From a statistical point of view, some methods have been proposed to address the problem of parameter estimation or label recovering for the SBM. Some examples include maximum likelihood estimation (Amini et al., 2013; Bickel e Chen, 2009), variational methods (Daudin et al., 2008; Latouche et al., 2012), spectral clustering (Rohe et al., 2011) and Bayesian inference (van der Pas et al., 2017). The asymptotic properties of these estimators have also been considered in subsequent works such as Bickel et al. (2013) or Su et al. (2017).

In both problems described above the number of communities (or blocks) of the model is assumed to be known. To our knowledge, the estimation of the number of communities in the SBM was not addressed in the literature until the recent work by Wang et al. (2017). The authors propose a penalized likelihood criterion and show its weak consistency, that is
only valid when the number of communities is upper bounded by a known constant and the network average degree grows as a polylog function on the number of nodes.

The method introduced in Wang et al. (2017) has been subsequently studied in Hu et al. (2016), where the authors propose a modification of the penalty term. However, in practice, the computation of the suggested estimator still remains a demanding task since it depends on the profile maximum likelihood.

The second model studied in this work is the Exponential Random Graph (ERG). Unlike the SBM, it does not take into account the structure of communities in the network. This model is defined using the sufficient statistics of the graph, for example, the number of edges, the number of triangles, the number of stars, and so on. For this reason this model has received a lot of attention from the Social Sciences community. Social networks are used as a way to describe and analyze the interactions between a set of actors where the reciprocity property is desired. Reciprocity assures that if two actors of the model interact with a third one, then it is more likely that they interact between them as well. In this context, the ERG has been vastly explored to model social networks since the sufficient statistics may represent desired properties of the graph; for example, the number of triangles in the graph can be used to represent the reciprocity property.

In terms of parameter estimation for this model, classical inferential methods such as Monte Carlo maximum likelihood of Geyer e Thompson (1992) and the pseudolikelihood estimation introduced by Besag (1974) are widely used in practice. In both these methods, it is necessary to use samples of the ERG distribution; however sampling from this distribution is usually not a simple task due to the fact that the normalizing constant of this distribution depends on a sum over the support of the distribution, that is, the set of all graphs. For a large number of vertices of the graph the distribution’s support is a huge set; for instance, to compute the normalizing constant for a graph with 20 vertices it is necessary to calculate a sum over the support set containing $2^{190}$ graphs.

In general, the algorithms to sample from the ERG distribution are based on a well known and vastly studied method in the literature called Markov chain Monte Carlo (MCMC). The Metropolis Hastings algorithm, introduced in the works of Metropolis et al. (1953) and Hastings (1970), and the Gibbs sampling algorithm, described in the work of Geman e Geman (1984), are the most used MCMC algorithms. None of these algorithms depend on the normalizing constant of the distribution to sample approximately a graph from this model.

The main idea of MCMC algorithms is the construction of an appropriate Markov chain having as limiting distribution the desired distribution. Using this method, the value obtained by running the Markov chain, from an initial state and after a number of steps, is an approximate sample of the target distribution. Notice that after running the chain a finite number of steps the sample obtained is an approximation of the limiting distribution of the chain and, depending on the number of steps, this value may depend on the initial state.

Snijders (2002) has studied and applied the MCMC algorithm for the purpose of simulating from the ERG model. However, the principal drawback of the MCMC methods is
the choice of the number of steps to run the algorithm in order to obtain a representative sample of the target distribution.

Bhamidi and co-authors presented a study of the mixing time of the Markov chain constructed using the Glauber dynamics (also known as Gibbs sampling) for ERG models with positive parameters. They describe the model introducing the concept of high and low temperature regimes and they show that for models for which the parameters belong to the high temperature phase the mixing time of the chain converges quickly to the stationary distribution. Furthermore, for models under the low temperature regime, the convergence takes an exponential time. They also show that for models under the high temperature regime any finite collection of edges in the graph are asymptotically independent, that is, when the number of vertices grows to infinity.

To overcome the issue of the convergence of the MCMC methods, Butts (2015) proposed a novel algorithm called bound sampler, for which the simulation time is fixed and depends only on the number of vertices of the graph. This algorithm and MCMC samplers are approximated methods for sampling from the ERG model. However, the bound sampler is recommended instead of MCMC algorithms when the mixing time of the chain is quite long.

1.1 Objectives

In this thesis we first address the model selection problem for the number of communities in the Stochastic Block Model (SBM). Given a graph obtained from the SBM, our goal is the construction of an estimator for the number of communities of the model.

In this way, we take an information-theoretic perspective and introduce the Krichevsky-Trofimov (KT) mixture estimator (Krichevsky e Trofimov, 1981) in order to determine the number of communities of a SBM based on a finite sample of the network. We propose an estimator without assuming a known upper bound on the number of communities.

We prove the strong consistency of this estimator, in the sense that the empirical value is equal to the correct number of communities in the model with probability one, as long as the number of nodes in the network is sufficiently large. To our knowledge this is the first strong consistency result for an estimator of the number of communities, even in the bounded case.

On the other hand, with respect to the inferential problems for the ERG model we address the following question: can we construct a perfect sampling algorithm to simulate from the distribution of the ERG?

The perfect (also known as exact) simulation algorithms guarantee that after a well defined finite number of steps the value generated by the algorithm has the target distribution.

Propp e Wilson (1996) were the first to introduce the perfect simulation through a Markov chain algorithm called Coupling from the past (CFTP) that generates an exact
sample from the stationary distribution of the chain and determines automatically how long it needs to run.

In our approach, we adapt a version of the CFTP algorithm, which is developed for the spin systems in Propp e Wilson (1996), to construct a perfect simulation algorithm for the ERG model using the Glauber dynamics. However, in this case, this algorithm can be implemented efficiently only for monotone models.

In order to cover all models, we propose another perfect simulation algorithm based in Galves et al. (2010), where the authors applied a perfect simulation algorithm for a Gibbs measure with infinite range interactions. In this work, the algorithm is formed by two stages: in the first stage they determined the set of sites whose spins influence each other and, in the second stage, the spins are assigned using the influence information of the first stage.

In Galves et al. (2013) the authors present a perfect simulation algorithm to simulate from the stationary process of a multicolor system on $\mathbb{Z}^d$ with interactions of infinite range using a Kalikow-type decomposition of the infinite range rates. This decomposition is a convex combination of rates with finite range and it allows us to deal with the infinite range dependence. This decomposition is also applied in the case of processes with long memory as described in the works of Comets et al. (2002) and Garivier (2015).

In spite of the fact that the vertex set of the graph considered in this thesis is finite, we propose a decomposition of the transition probabilities of the Markov chain, constructed using the Glauber dynamics, that considers the local dependence of an edge in the graph through a subset of pair of vertices.

1.2 Contributions

The main contributions of this thesis are the following:

- The construction of a penalized estimator based on the Krichevsky-Trofimov (KT) mixture distribution for the number of communities of the SBM.
  - We construct an estimator without assuming a known upper bound on the number of communities;
  - We prove that the estimator is strongly consistent.

- The construction of alternative algorithms to the MCMC approach to be applied in the estimation methods for ERG models.
  - We construct a perfect sampling algorithm based on the coupling from the past algorithm for monotone ERG models;
  - We construct an alternative perfect sampling algorithm that can also be applied to non-monotone models.
1.3 Organization of the work

This thesis is divided into two parts for the reader’s convenience. Part I covers the model selection problem for the SBM. In Chapter 2 we present the definitions and notations of random graphs used in this thesis. This chapter also contains an overview of the inferential problems for this class of models and the definition of the KT mixture distribution. In the last chapter of this part, Chapter 3, we present the penalized estimator without assuming a known upper bound on the number of communities and we prove its strong consistency. In the second part of this thesis we discuss the perfect simulation problem for the class of ERG models. In Chapter 4 we present the definitions of the ERG models and a brief description of the main concepts used in the second part of this work. The construction of the coupling from the past algorithm is described in Chapter 5. In Chapter 6 we focus our attention on the algorithm called Backward and Forward that can be applied to non-monotone ERG models. Finally, a simulation study of both algorithms is presented in Chapter 7.

In Chapter 8 we present the final considerations about the results obtained in this thesis as well as the future potential research problems in this field.

In Appendix A the reader can find the additional results and proofs omitted in the previous chapters of this thesis.
Part I

Stochastic Block Model
Chapter 2

Stochastic Block Model

In this chapter we present the general definitions and notations about random graphs frequently used in this thesis. The specific definitions and notations required to the development of this thesis are given gradually in the sections. We also introduce the Stochastic Block Model and the Krichevsky - Trofimov mixture distribution.

2.1 Random Graph Definitions

A graph is a pair \((V, E)\), where \(V\) is a finite set of vertices and \(E\) is a set of edges that connect pairs of vertices. In this work we consider undirected graphs, i.e., graphs for which the set \(E\) satisfies the following properties:

1. For all \(i \in V\), \((i, i) \notin E\)
2. If \((i, j) \in E\), then \((j, i) \in E\)

Let \(\mathcal{G}_{|V|}\) be the collection of all undirected graphs with set of vertices \(V\). A graph \((V, E) \in \mathcal{G}_{|V|}\) can be identified with a symmetric binary matrix \(x \in \mathcal{M}_{|V|}(\{0, 1\})\) such that \(x(i, j) = 1\) if and only if \((i, j) \in E\). The matrix \(x\) is called the adjacency matrix of the graph.

We will thus denote by \(x\) a graph, which is completely determined by the values \(x(i, j)\), for \(1 \leq i < j \leq |V|\).

For an undirected random graph its adjacency matrix \(X\) is symmetric, its diagonal entries are zero and its upper diagonal entries are given by a collection of random variables \(\{X(i, j) : 1 \leq i < j \leq |V|\}\).

In the model selection problem for the SBM we are interesting in the case where the number of vertices of the graph grows to infinity. In that case, we assume that \(|V| = n\) and we denote the adjacency matrix of the graph by \(x_{n \times n}\) to emphasize that it depends on \(n\). In the second part of the thesis we consider that the number of vertices of the graph is fixed and finite. Under this assumption, it is convenient to introduce minor notational changes, and we advise the reader to check Section 4.1.1 for further details.
A well known and studied model in the literature is the Erdős-Rényi introduced in \cite{erdos1959random}. In this model the edges are chosen independently of the other edges of the graph and with the same probability, that means, \{X(i, j) : 1 \leq i < j \leq |V|\} is a collection of independent and identically distributed Bernoulli random variables.

### 2.2 Stochastic Block Model

We consider a network with \( n \) nodes with adjacency matrix \( X_{n \times n} = (X(i, j)) \) is given by a symmetric matrix with diagonal entries equal to zero. The classification of the nodes in \( k_0 \) communities are given by a latent random vector \( Z_n = (Z_1, Z_2, \ldots, Z_n) \) of independent random variables taking values in \([k_0] := \{1, \ldots, k_0\}\) for which \( \mathbb{P}(Z_i = k) = \pi^0_k \), \( \pi = (\pi^0_1, \ldots, \pi^0_{k_0}) \) and \( \sum_{i=1}^{k_0} \pi^0_i = 1 \). The event \( \{Z_i = k\} \) means that the node \( i \) belongs to community \( k \).

Given the nodes’ labels, each entry of the adjacency matrix \( X_{n \times n} \) is conditionally independent given the labels of the vertices and it is generated according to the conditional distribution

\[
X(i, j) | (Z_i = a, Z_j = b) \sim \text{Bernoulli}(P^0_{a,b}), \quad 1 \leq i < j \leq n
\]

where \( P = (P^0_{a,b}) \) is a symmetric \( k_0 \times k_0 \) matrix.

The Erdős-Rényi model is a particular case of the SBM when we have \( k_0 = 1 \).

In order to guarantee that the model is identifiable, in the sense that it cannot be reduced to a smaller model (i.e., a model with a smaller number of communities), we assume that each community has positive probability to have at least one member. Furthermore, we assume that two communities cannot have the same probabilities of being connected to other communities. Thus, we assume that the following holds:

**Assumption 2.2.1.** \( P \) does not have two identical columns and \( \pi \) has all positive entries.

We write the vector of model’s parameters as \( \theta = (\pi, P) \) and the parametric space

\[
\Theta^k = \{\theta = (\pi, P) | \pi \in (0, 1)^k, \sum_{a=1}^{k} \pi_a = 1, P \in [0, 1]^{k \times k}, P \text{ symmetric}\}
\]
For $\theta \in \Theta^k$ we write the joint distribution of $X_{n \times n}$ and $Z_n$ as

$$
P_{\theta}(X_{n \times n} = x_{n \times n}, Z_n = z_n) = P_{\theta}(X_{n \times n} = x_{n \times n} | Z_n = z_n) P_{\theta}(Z_n = z_n)$$

$$= \left[ \prod_{a=1}^{k} \pi_{a}^{n_{a}} \right] \prod_{i=1}^{n} \prod_{j=i+1}^{n} P_{\theta}(X(i, j) = x(i, j) | Z_n = z_n) =$$

$$= \left[ \prod_{a=1}^{k} \pi_{a}^{n_{a}} \right] \prod_{i=1}^{n} \prod_{j=i+1}^{n} \prod_{a=1}^{k} \prod_{b=1}^{k} P_{\theta}(X(i, j) = x(i, j) | Z_i = a, Z_j = b)$$

$$= \left[ \prod_{a=1}^{k} \pi_{a}^{n_{a}} \right] \prod_{i=1}^{n} \prod_{j=i+1}^{n} \prod_{a=1}^{k} \prod_{b=1}^{k} P_{a,b} 1\{z_i = a, z_j = b\} x(i, j) (1 - P_{a,b}) 1\{z_i = a, z_j = b\} (1 - x(i, j))$$

$$= \left[ \prod_{a=1}^{k} \pi_{a}^{n_{a}} \right] \prod_{i=1}^{n} \prod_{j=i+1}^{n} \prod_{a=1}^{k} \prod_{b=1}^{k} P_{a,b} O_{a,b} (1 - P_{a,b})^{n_{a,b} - O_{a,b}} \right]^{\frac{1}{2}}$$

(2.2.1)

where the counters $n_a = n_a(z_n), n_{a,b} = n_{a,b}(z_n)$ and $O_{a,b} = O_{a,b}(z_n, x_{n \times n})$ are given by

$$n_a(z_n) = \sum_{i=1}^{n} 1\{z_i = a\}, \quad 1 \leq a \leq k$$

$$n_{a,b}(z_n) = \begin{cases} \sum_{i=1}^{n} \sum_{j \neq i} 1\{z_i = a, z_j = b\} = n_a(z_n) n_b(z_n), & 1 \leq a, b \leq k; a \neq b \\ \sum_{i=1}^{n} \sum_{j \neq i} 1\{z_i = a, z_j = a\} = n_a(z_n) (n_a(z_n) - 1) & 1 \leq a, b \leq k; a = b \end{cases} \quad (2.2.2)$$

and

$$O_{a,b}(z_n, x_{n \times n}) = \sum_{i,j=1}^{n} 1\{z_i = a, z_j = b\} x(i, j), \quad 1 \leq a, b \leq k.$$

As it is usual in the definition of likelihood functions, by convention we define $0^0 = 1$ in (2.2.1) when some of the parameters are 0.

For $\theta \in \Theta^k$ the marginal distribution of $X_{n \times n}$ is given by

$$P_{\theta}(X_{n \times n} = x_{n \times n}) = \sum_{z_n \in [k]^n} P_{\theta}(X_{n \times n} = x_{n \times n}, Z_n = z_n). \quad (2.2.3)$$

The definition of identifiability for this model, as described in Assumption 2.2.1, is vastly used in the literature; however, in this thesis we use an equivalent definition. We define the order of the SBM as the smallest $k$ for which the equality (2.2.1) holds for a pair of parameters $(\pi_0, P_0) \in \Theta^k$ and it is denoted by $k_0$. If a SBM has order $k_0$ then it cannot be
reduced to a model with less communities than $k_0$; this specifically means that $P_0$ does not have two identical columns.

**Example 2.2.2.** Consider the SBM with two blocks where the block assignments and connections are represented by Figure (2.1).

![Figure 2.1: Example of a graph originated from the SBM.](image)

In this case, we have

$$
\begin{align*}
n_a &= 4, \quad n_b = 2 \\
n_{a,a} &= 12, \quad O_{a,a} = 6 \\
n_{a,b} &= 8, \quad O_{a,b} = 2 \\
n_{b,a} &= 8, \quad O_{b,a} = 2 \\
n_{b,b} &= 2, \quad O_{b,b} = 2
\end{align*}
$$

Using the fact that $P$ is symmetric, the joint distribution described in (2.2.1) is expressed as

$$
\mathbb{P}_\theta(X_{n \times n} = x_{n \times n}, Z_n = z_n) = \left[ \prod_{a=1}^{k} \pi_a^{n_a} \right] \left[ \prod_{a=1}^{k} \prod_{b=1}^{k} P_{a,b}^{O_{a,b}} (1 - P_{a,b})^{n_{a,b} - O_{a,b}} \right]^\frac{1}{2}
$$

$$
= \left[ \pi_a^{4 \pi_b^{2}} \right] \left[ P_{a,a}^6 (1 - P_{a,a})^6 P_{a,b}^4 (1 - P_{a,b})^{12} P_{b,b}^2 (1 - P_{b,b})^0 \right]^\frac{1}{2}
$$

$$
= \left[ \pi_a^{4 \pi_b^{2}} \right] \left[ P_{a,a}^3 (1 - P_{a,a})^3 P_{a,b}^2 (1 - P_{a,b})^6 P_{b,b}^1 (1 - P_{b,b})^0 \right]
$$

The next example illustrates the fact that Assumption 2.2.1 is necessary to guarantee identifiability of the model in the sense described above.
Example 2.2.3. Consider the SBM with 3 blocks with parameter $\theta = (\pi, P)$ such that

$$P = \begin{bmatrix} 0.2 & 0.4 & 0.4 \\ 0.4 & 0.5 & 0.5 \\ 0.4 & 0.5 & 0.5 \end{bmatrix} \quad \text{and} \quad \pi = (0.4 \ 0.3 \ 0.3).$$

Observe that this model does not satisfy Assumption 2.2.1. In this way, it can be represented by a smaller order model with parameter $\tilde{\theta} = (\tilde{\pi}, \tilde{P})$ given by

$$\tilde{P} = \begin{bmatrix} 0.2 & 0.4 \\ 0.4 & 0.5 \end{bmatrix} \quad \text{and} \quad \tilde{\pi} = (0.4 \ 0.6).$$

The construction of the smaller order model is detailed in Section 3.2.2 for general models. The main idea behind this construction is to replace label 3 by 2. For example, in Figure (2.2a), consider the graph sampled from the model of order 3 with parameter $\theta$, then we relabeled the nodes 3 by 2 to get the graph in (2.2b).

![Graphs originated by the SBM with (a) 3 blocks and parameter $\theta$ (b) the corresponded relabeled graph.](image)

In this case, for the graph $x_{7 \times 7}$ represented in Figure (2.2a) and the labels $z_7 = (1 \ 2 \ 1 \ 3 \ 3 \ 2)$ we have that $\log P_0(x_{7 \times 7}, z_7) = -8.369482$. For the relabeled model, Figure (2.2b), with labels $\tilde{z}_7 = (1 \ 2 \ 1 \ 2 \ 2 \ 2)$ we have that $\log \tilde{P}_0(\tilde{x}_{7 \times 7}, \tilde{z}_7) = -8.369482$.

2.3 Inferential Problems

Let $(x_{n \times n}, z_0^0)$ be a sample of the SBM of order $k_0$ with parameter $\theta_0 = (\pi_0, P_0)$. We only observe the structure of the network $x_{n \times n}$ and the nodes’ labels $z_0^0$ are hidden. From an inferential point of view we can address three problems:
1. Estimation of the parameter \( \theta_0 = (\pi_0, P_0) \);

2. Estimation of the latent nodes’ labels \( z_n^0 \);

3. Estimation of the number of communities \( k_0 \).

There is an increasing number of works that have addressed the problem of estimating the latent block configuration and the model’s parameters. However, in both problems the number of communities of the model is assumed to be known. In this scenario we can list some relevant works such as the likelihood maximization (Amini et al. (2013) and Bickel e Chen (2009)), variational methods (Daudin et al., 2008; Latouche et al., 2012), spectral clustering (Rohe et al., 2011). There are also some theoretical works about the asymptotic properties of theses methods, as Bickel e Chen (2009) Rohe et al. (2011), Bickel et al. (2013) and Su et al. (2017).

In this part of the thesis we focus on the estimation of the number of communities of the model (order estimation). For this, we define the estimation procedure and the types of estimators.

**Definition 2.3.1.** An order estimation procedure is a sequence of estimators \( \hat{k}_1, \ldots, \hat{k}_n, \ldots \) that, given observed graphs with \( 1, \ldots, n, \ldots \) nodes, it outputs estimates \( \hat{k}_n(x_{n \times n}) \) of \( k_0 \).

**Definition 2.3.2 (Weak and Strong Consistency).** A sequence of estimators \( \hat{k}_1, \ldots, \hat{k}_n, \ldots \) is weakly consistent if \( \Pr(\hat{k}_n = k_0) \) converges to 1 as \( n \to \infty \).

A sequence of estimators \( \hat{k}_1, \ldots, \hat{k}_n, \ldots \) is strongly consistent if \( \hat{k}_n = k_0 \) eventually almost surely as \( n \to \infty \).

### 2.3.1 Estimation of the number of communities

Wang et al. (2017) were the first to propose a penalized likelihood criterion to choose the correct number of communities. Their estimator is given by

\[
\hat{k}_{\lambda}(x_{n \times n}) = \arg \max_k \left\{ \sup_{\theta \in \Theta^k} \log P_\theta(x_{n \times n}) - \lambda \frac{k(k+1)}{2} n \log n \right\}
\]  

where \( \lambda \) is a tuning parameter.

They prove the criterion is weakly consistent, that means, the estimator chooses the correct number of blocks with probability tending to 1, when \( n \) goes to infinity. However, in a practical way, the computation of the log likelihood function (2.2.3) and its supremum is usually not a simple task due to the fact that it depends on a sum over an exponential number of terms. To overcome this difficulty, Wang et al. (2017) proposed some alternative ways to approximate the likelihood function in order to preserve the asymptotic consistency of the suggested estimator. They proposed a variational method as described in Bickel et al. (2013) to obtain a variational estimator for \( \theta \) computed using the EM algorithm of Daudin et al.
Another method that preserves the asymptotic properties of (2.3.1) is the computation of the maximum complete likelihood (2.2.1) by plugging the labels’ estimates. For this purpose, to estimate the node’s labels it is suggested to use the profile maximum likelihood in Bickel e Chen (2009) or the pseudo-likelihood algorithm in Amini et al. (2013).

Hu et al. (2016) proposed a corrected Bayesian information criterion (CBIC) where the estimator is given by

\[
\hat{k}_{\text{CBIC}}(x_{n \times n}) = \arg \max_k \left\{ \max_{z_n \in [k]^n} \sup_{\theta \in \Theta^k} \log P_{\theta}(x_{n \times n}, z_n) - \left( \lambda_n \log k + \frac{k(k+1)}{2} \log n \right) \right\}.
\] (2.3.2)

The penalty function in the CBIC estimator improves that used in Wang et al. (2017); however, in practice, the computation of the suggested estimator still remains a demanding task since it depends on the profile maximum likelihood.

### 2.3.2 The KT mixture distribution

In Chapter 3 we propose an estimator based on the Krichevsky - Trofimov distribution (KT) proposed in Krichevsky e Trofimov (1981). The Krichevsky-Trofimov estimator in the context of a SBM is a regularized estimator for the random adjacency matrix \(X_{n \times n}\).

As it is usual for the KT distributions we choose as “prior” for the pair \(\theta = (\pi, P)\) a product measure obtained by a Dirichlet \((1/2, \cdots , 1/2)\) distribution (the prior distribution for \(\pi\)) and a product of \((k^2 + k)/2\) Beta \((1/2, 1/2)\) distributions (the prior for the symmetric matrix \(P\)). In other words, we define the distribution

\[
\nu_k(\theta) = \left[ \frac{\Gamma \left( \frac{k}{2} \right)}{\Gamma \left( \frac{1}{2} \right)^k} \prod_{a=1}^{k} \pi_a^{-1/2} \right] \left[ \prod_{a=1}^{k} \prod_{a \leq b \leq k} \frac{1}{\Gamma \left( \frac{1}{2} \right)^2} P_{a,b}^{-1/2} (1 - P_{a,b})^{-1/2} \right].
\]

For a given sample \((x_{n \times n}, z_n)\) of the SBM of order \(k\) the Krichevsky-Trofimov mixture distribution on \(G_n\) only depends on the observed network and it is defined by

\[
\text{KT}_k(x_{n \times n}) = \mathbb{E}_{\nu_k} [P_{\theta}(x_{n \times n})] = \int_{\Theta^k} P_{\theta}(x_{n \times n}) \nu_k(\theta) d\theta.
\] (2.3.3)

It is important to emphasize that even though we use prior distributions for the parameters \(\pi\) and \(P\), the KT distribution does not play the role of a posterior distribution.

We start with a result that have an important role in the proof of the consistency of the estimator proposed in Chapter 3. This result gives an upper bound for the maximum log ratio between the maximum likelihood and the Krichevsky-Trofimov mixture distribution.
Proposition 2.3.3. For all \( k \) and all \( n \geq \max(4, k) \) we have

\[
\log C(k, n) \leq \max \left\{ \log \frac{\sup_{\theta \in \Theta^k} P_{\theta}(x_{n \times n})}{KT_k(x_{n \times n})} \right\} \leq \left( \frac{k(k+2)}{2} - \frac{1}{2} \right) \log n + c_{k,n} \tag{2.3.4}
\]

where

\[
c_{k,n} = \frac{k(k+1)}{2} \log \Gamma \left( \frac{1}{2} \right) + \frac{k(k-1)}{4n} + \frac{1}{12n} + \log \frac{\Gamma(\frac{k}{2})}{\Gamma(\frac{k}{2})} + \frac{7k(k+1)}{12} \tag{2.3.5}
\]

and

\[
C(k, n) = \sum_{x_{n \times n}} \sup_{\theta \in \Theta^k} P_{\theta}(x_{n \times n}).
\]

Proof. We start by proving the first inequality in (2.3.4). The proof is based on Csiszár et al. (2004, Theorem 6.2).

\[
C(k, n) = \sum_{y_{n \times n}} \sup_{\theta \in \Theta^k} P_{\theta}(y_{n \times n}) = \sum_{y_{n \times n}} KT_k(y_{n \times n}) \sup_{\theta \in \Theta^k} P_{\theta}(y_{n \times n})
\]

\[
\leq \sum_{y_{n \times n}} KT_k(y_{n \times n}) \max_{x_{n \times n}} \sup_{\theta \in \Theta^k} P_{\theta}(x_{n \times n}) = \max_{x_{n \times n}} \sum_{y_{n \times n}} KT_k(y_{n \times n}) \sup_{\theta \in \Theta^k} P_{\theta}(x_{n \times n})
\]

Thus,

\[
\log C(k, n) \leq \log \left[ \max_{x_{n \times n}} \sup_{\theta \in \Theta^k} P_{\theta}(x_{n \times n}) \right] = \max_{x_{n \times n}} \log \frac{\sup_{\theta \in \Theta^k} P_{\theta}(x_{n \times n})}{KT_k(x_{n \times n})}.
\]

Now we prove the second inequality in (2.3.4). The proof is based on Liu e Narayan (1994, Lemma 3.4). For \( \theta \in \Theta^k \) we have that

\[
P_{\theta}(z_n) = \prod_{a=1}^{k} \pi_a^{n_a} \tag{2.3.6}
\]

and

\[
P_{\theta}(x_{n \times n} | z_n) = \prod_{a=1}^{k} \prod_{a \leq b \leq k} \left[ P_{a,b}^\theta (1 - P_{a,b})^{\tilde{n}_{a,b} - \theta_{a,b}} \right] \frac{1}{2}, \tag{2.3.7}
\]

where

\[
\tilde{n}_{a,b} = \begin{cases} 2n_{a,b}, & 1 \leq a, b \leq k; a \neq b \\ n_{a,b}, & 1 \leq a, b \leq k; a = b \end{cases}
\]
\[ \hat{O}_{a,b} = \begin{cases} 2O_{a,b}, & 1 \leq a, b \leq k; \ a \neq b \\ O_{a,b}, & 1 \leq a, b \leq k; \ a = b. \end{cases} \]

Using that the maximum likelihood estimators of \( \pi_a \) and \( P_{a,b} \) are given by \( \frac{n_a}{n} \) and \( \frac{\hat{O}_{a,b}}{\bar{n}_{a,b}} \) (respectively) we can bound above (2.3.6) and (2.3.7) by

\[ P_{\theta}(z_n) \leq \sup_{\theta \in \Theta^k} P_{\theta}(z_n) = \prod_{a=1}^{k} \left( \frac{n_a}{n} \right)^{n_a}. \quad (2.3.8) \]

Using the fact that the maximum likelihood estimator of \( P_{l,r} \) is given by \( \frac{\hat{O}_{l,r}}{\bar{n}_{l,r}} \) we write (2.3.7) as

\[ P_{\theta}(x_{n \times n} | z_n) \leq \sup_{\theta \in \Theta^k} P_{\theta}(x_{n \times n} | z_n) \]

\[ = \prod_{a=1}^{k} \prod_{a \leq b \leq k} \left[ \left( \frac{\hat{O}_{a,b}}{\bar{n}_{a,b}} \right)^{\hat{O}_{a,b}} \left( 1 - \left( \frac{\hat{O}_{a,b}}{\bar{n}_{a,b}} \right) \right)^{\bar{n}_{a,b} - \hat{O}_{a,b}} \right] \frac{1}{2}. \quad (2.3.9) \]

Observe that the Krichevsky-Trofimov mixture distribution defined in (2.3.3) can be written as

\[ \text{KT}_k(x_{n \times n}) = \sum_{z_n \in [k]^n} \int_{\Theta^k} P_{\theta}(x_{n \times n} | z_n) P_{\theta}(z_n) \nu_k(\theta) d\pi dP \]

\[ = \sum_{z_n \in [k]^n} \int_{\Theta^k} \left( \prod_{a=1}^{k} \pi_a^{n_a} \right) \left( \prod_{a=1}^{k} \prod_{a \leq b \leq k} P_{a,b}^{\hat{O}_{a,b}} (1 - P_{a,b})^{\bar{n}_{a,b} - \hat{O}_{a,b}} \right) \frac{1}{2} \nu_k(\theta) d\pi dP. \quad (2.3.10) \]

Then substituting (2.3.6) and (2.3.7) in (2.3.10)

\[ \text{KT}_k(x_{n \times n}) = \sum_{z_n \in [k]^n} \int_{\Theta^k} \left( \prod_{l=1}^{k} \pi_l^{n_l} \right) \left( \prod_{l=1}^{k} \prod_{l \leq r \leq k} P_{l,r}^{\hat{O}_{l,r}} (1 - P_{l,r})^{\bar{n}_{l,r} - \hat{O}_{l,r}} \right) \frac{1}{2} \nu_k(\theta) d\theta. \]

We can write

\[ \nu_k(\theta) = \nu_k^1(\pi) \nu_k^2(P) \]
where

\[ \nu_k^1(\pi) = \left[ \frac{\Gamma\left(\frac{k}{2}\right)}{\Gamma\left(\frac{1}{2}\right)^k} \prod_{a=1}^{k} \pi_a^{-1/2} \right] \]

\[ \nu_k^2(P) = \prod_{a=1}^{k} \prod_{a \leq b \leq k} \frac{1}{\Gamma\left(\frac{1}{2}\right)^2} P_{a,b}^{-1/2} (1 - P_{a,b})^{-1/2} \]

and define

\[ \Theta_1^k = \{ \pi | \pi \in (0, 1]^k, \sum_{a=1}^{k} \pi_a = 1 \} \]

and

\[ \Theta_2^k = \{ P | P \in [0, 1]^{k \times k}, P \text{ is symmetric} \}. \]

Thus, we rewrite (2.3.10) as

\[ KT_k(x_{n \times n}) = \sum_{z_n \in [k]^n} KT_k(x_{n \times n} | z_n) KT_k(z_n) \]

where

\[ KT_k(x_{n \times n} | z_n) = \int_{\Theta_2^k} \left( \prod_{a=1}^{k} \prod_{a \leq b \leq k} P_{l,r}^{\hat{O}_{a,b}} (1 - P_{a,b})^{\hat{n}_{a,b} - \hat{O}_{a,b}} \right)^{1/2} \nu_k^2(P) \prod_{l=1}^{k} \prod_{a \leq b \leq k} d(P_{a,b}) \]

and

\[ KT_k(z_n) = \int_{\Theta_1^k} \left( \prod_{a=1}^{k} \pi_a^{n_a} \right) \nu_k^1(\pi) \prod_{a=1}^{k} d(\pi_a) \]

We start with the evaluation of \( KT_k(z_n) \).

\[ KT_k(z_n) = \int \left( \prod_{a=1}^{k} \pi_a^{n_a} \right) \left[ \frac{\Gamma\left(\frac{k}{2}\right)}{\Gamma\left(\frac{1}{2}\right)^k} \prod_{a=1}^{k} \pi_a^{-1/2} \right] \prod_{a=1}^{k} d(\pi_a) \]

\[ = \frac{\Gamma\left(\frac{k}{2}\right)}{\Gamma\left(\frac{1}{2}\right)^k} \int \left( \prod_{a=1}^{k} \pi_a^{n_a - 1/2} \right) \prod_{a=1}^{k} d(\pi_a) \]

\[ = \frac{\Gamma\left(\frac{k}{2}\right)}{\Gamma\left(\frac{1}{2}\right)^k} \frac{\prod_{a=1}^{k} \Gamma\left(n_a + \frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)^k \Gamma\left(n + \frac{k}{2}\right)} \quad (2.3.11) \]
Combining (2.3.8) and (2.3.11)

\[
\frac{P(z_n)}{KT_k(z_n)} \leq \frac{k}{\Gamma \left( \frac{k}{2} \right) \prod_{a=1}^{k} \Gamma(n_a + \frac{1}{2})} \frac{\left( \frac{n_a}{n} \right)^{n_a}}{\Gamma \left( \frac{n}{2} \right) \prod_{a=1}^{k} \Gamma(n_a + \frac{1}{2})} \prod_{a=1}^{k} \frac{n_a}{n} \cdot \Gamma \left( \frac{n_a}{2} + 1 \right). \tag{2.3.12}
\]

Using the fact that \(n_1 + \cdots + n_k = n\) and Lemma (A.0.1) we have that

\[
\prod_{a=1}^{k} \frac{n_a}{\Gamma \left( \frac{n_a}{2} + \frac{1}{2} \right)} \leq \frac{1}{\Gamma \left( \frac{k}{2} \right) \prod_{a=1}^{k} \Gamma(n_a + \frac{1}{2})}. \tag{2.3.13}
\]

Using (2.3.13) in equation (2.3.12)

\[
\frac{P(z_n)}{KT_k(z_n)} \leq \frac{1}{\Gamma \left( \frac{k}{2} \right) \prod_{a=1}^{k} \Gamma(n_a + \frac{1}{2})}. \tag{2.3.14}
\]

We have that

\[
KT_k(x_{n \times n} | z_n) = \int_{p} \left( \prod_{a=1}^{k} \prod_{a \leq b \leq k} \frac{\hat{n}_{a,b} - \hat{\tau}_{a,b}}{2} \prod_{a \leq b \leq k} \nu_{\hat{\tau}_{a,b}}^{2}(P) \prod_{a=1}^{k} \prod_{a \leq b \leq k} d(P_{a,b}) \right) \Gamma \left( \frac{k}{2} \right)^{k-1} \prod_{a=1}^{k} \Gamma \left( n_a + \frac{1}{2} \right). \tag{2.3.14}
\]

Combining (2.3.9) and (2.3.14)
\[ P(x_{n \times n} | z_n) \leq \prod_{a=1}^{k} \prod_{a \leq b \leq k} \left( \frac{\hat{O}_{a,b}}{\hat{n}_{a,b}} \right)^{\frac{n_{a,b} - \hat{n}_{a,b}}{2}} \left( 1 - \frac{\hat{O}_{a,b}}{\hat{n}_{a,b}} \right)^{\frac{\hat{n}_{a,b} - \hat{O}_{a,b}}{2}} \frac{1}{\Gamma \left( \frac{k}{2} \right)} \prod_{a=1}^{k} \prod_{a \leq b \leq k} \left( \frac{\hat{O}_{a,b}}{\hat{n}_{a,b}} \right)^{\frac{n_{a,b} - \hat{n}_{a,b}}{2}} \left( 1 - \frac{\hat{O}_{a,b}}{\hat{n}_{a,b}} \right)^{\frac{\hat{n}_{a,b} - \hat{O}_{a,b}}{2}} \frac{\Gamma \left( \frac{n_{a,b} - \hat{n}_{a,b} + 1}{2} \right)}{\Gamma \left( \frac{n_{a,b}}{2} + 1 \right)} \frac{\Gamma \left( \frac{n_{a,b}}{2} + \frac{1}{2} \right)}{\Gamma \left( \frac{k}{2} \right)} \frac{\Gamma \left( \frac{n_{a,b} - \hat{n}_{a,b}}{2} + \frac{1}{2} \right)}{\Gamma \left( \frac{n_{a,b}}{2} + \frac{1}{2} \right)} \right], \]  

For each pair \( a \) and \( b \), \( 1 \leq a \leq k \) and \( 1 \leq b \leq k \), we use the fact that \( \hat{O}_{a,b} + \hat{n}_{a,b} - \hat{O}_{a,b} = \hat{n}_{a,b} \) and Lemma A.0.1 to get

\[ \frac{\Gamma \left( \frac{n_{a,b}}{2} + \frac{1}{2} \right)}{\Gamma \left( \frac{k}{2} \right)} \frac{\Gamma \left( \frac{n_{a,b} - \hat{n}_{a,b}}{2} + \frac{1}{2} \right)}{\Gamma \left( \frac{n_{a,b}}{2} + \frac{1}{2} \right)} \leq \frac{1}{\Gamma \left( \frac{n_{a,b}}{2} + \frac{1}{2} \right)} \Gamma \left( \frac{k}{2} \right) \frac{1}{\Gamma \left( \frac{n_{a,b} - \hat{n}_{a,b}}{2} + \frac{1}{2} \right)} \Gamma \left( \frac{k}{2} \right) \frac{\Gamma \left( \frac{n_{a,b}}{2} + \frac{1}{2} \right)}{\Gamma \left( \frac{k}{2} \right)} \frac{\Gamma \left( \frac{n_{a,b} - \hat{n}_{a,b}}{2} + \frac{1}{2} \right)}{\Gamma \left( \frac{n_{a,b}}{2} + \frac{1}{2} \right)} \]  

Using (2.3.16) in (2.3.15)

\[ P(x_{n \times n} | z_{1:n}) \leq \Gamma \left( \frac{1}{2} \right) \prod_{a=1}^{k} \prod_{a \leq b \leq k} \left[ \frac{\Gamma \left( \frac{n_{a,b}}{2} + 1 \right)}{\Gamma \left( \frac{n_{a,b}}{2} + \frac{1}{2} \right)} \right] \]  

Thus,

\[ \log \frac{P(x_{n \times n})}{\Gamma_k(x_{n \times n})} \leq \log \left( \frac{\Gamma \left( \frac{1}{2} \right) \Gamma \left( \frac{n + k}{2} \right)}{\Gamma \left( \frac{1}{2} \right) \Gamma \left( \frac{n + 1}{2} \right)} \right) + \left( \frac{k(k-1)}{2} \right) \log \Gamma \left( \frac{1}{2} \right) \]  

\[ + \sum_{a=1}^{k} \sum_{a \leq b \leq k} \log \left( \frac{\Gamma \left( \frac{n_{a,b}}{2} + 1 \right)}{\Gamma \left( \frac{n_{a,b}}{2} + \frac{1}{2} \right)} \right). \]  

(2.3.17)
Using Lemma A.0.2 we have that

\[
\log \left( \frac{\Gamma \left( \frac{1}{2} \right) \Gamma \left( n + \frac{k}{2} \right)}{\Gamma \left( \frac{k}{2} \right) \Gamma \left( n + \frac{1}{2} \right)} \right) \leq \left( \frac{k - 1}{2} \right) \log n + \frac{k(k - 1)}{4n} + \frac{1}{12n} + \log \frac{\Gamma \left( \frac{1}{2} \right)}{\Gamma \left( \frac{k}{2} \right)}.
\]

In the same way, for \(1 \leq a \leq k, \ a \leq b \leq k\), we have that

\[
\log \left( \frac{\Gamma \left( \frac{\tilde{n}_{a,b}}{2} + 1 \right)}{\Gamma \left( \frac{\tilde{n}_{a,b}}{2} + \frac{1}{2} \right)} \right) \leq \frac{1}{2} \log \left( \frac{\tilde{n}_{a,b}}{2} \right) + \frac{1}{\tilde{n}_{a,b}} + \frac{1}{6\tilde{n}_{a,b}} \leq \frac{1}{2} \log \left( \frac{\tilde{n}_{a,b}}{2} \right) + \frac{7}{6}.
\]

where the last inequality follows from the fact that \(\tilde{n}_{a,b} \geq 1\).

Using that \(\tilde{n}_{a,b} = 2n_a n_b \leq 2n^2\), for \(a \neq b\) and \(\tilde{n}_{a,a} = n_a (n_a - 1) \leq 2n^2\) we obtain

\[
\log \left( \frac{\Gamma \left( \frac{\tilde{n}_{a,b}}{2} + 1 \right)}{\Gamma \left( \frac{\tilde{n}_{a,b}}{2} + \frac{1}{2} \right)} \right) \leq \log n + \frac{7}{6}.
\]  

(2.3.18)

Setting \(c_{k,n} = \frac{k(k+1)}{2} \log \Gamma \left( \frac{1}{2} \right) + \frac{k(k-1)}{4n} + \frac{1}{12n} + \log \frac{\Gamma \left( \frac{1}{2} \right)}{\Gamma \left( \frac{k}{2} \right)} + \frac{7k(k+1)}{12}\) and combining (2.3.17) and (2.3.18)

\[
\log \frac{\mathbb{P}(x_{n \times n})}{\mathbb{K}_k(x_{n \times n})} \leq \left( \frac{k - 1}{2} \right) \log n + \frac{k(k + 1)}{2} \log n + c_{k,n} = \left( \frac{k(k + 2)}{2} - \frac{1}{2} \right) \log n + c_{k,n}.
\]

Thus, the result follows. \(\square\)

### 2.4 Computation of KT

As discussed before, in practice the computation of the estimator of the number of blocks proposed in the literature can be a challenging task because of the form of the likelihood function. The proof of Proposition 2.3.3 gives us the tools to compute the KT distribution. For notational convenience we rewrite the counters defined in (2.2.2) as

\[
\tilde{n}_{a,b} = \begin{cases} 2n_{a,b} & 1 \leq a, b \leq k; \ a \neq b \\ n_{a,b} & 1 \leq a, b \leq k; \ a = b \end{cases}
\]
and

\[ \hat{\Omega}_{a,b} = \begin{cases} 
2\Omega_{a,b}, & 1 \leq a, b \leq k; \ a \neq b \\
\Omega_{a,b} & 1 \leq a, b \leq k; \ a = b.
\end{cases} \]

By equations (2.3.11) and (2.3.14), we write the KT distribution as

\[ KT_k(x_{n \times n}) = \sum_{z_n \in [k]^n} KT_k(x_{n \times n}|z_n)KT_k(z_n) \]

with

\[ KT_k(z_n) = \frac{\Gamma \left( \frac{k}{2} \right) \prod_{a=1}^{k} \Gamma \left( n_a + \frac{1}{2} \right)}{\Gamma \left( \frac{1}{2} \right)^k \Gamma \left( n + \frac{k}{2} \right)} \]

and

\[ KT_k(x_{n \times n}|z_n) = \frac{1}{\Gamma \left( \frac{1}{2} \right)^{k(k-1)+2k}} \prod_{a=1}^{k} \prod_{a \leq b \leq k} \beta \left( \frac{\hat{\Omega}_{a,b}}{2} + \frac{1}{2}, \frac{n_{a,b} - \hat{\Omega}_{a,b}}{2} + \frac{1}{2} \right) \]

where \( \beta(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \).

Observe that \( KT_k(z_n) \) gives a marginal distribution for the labels and it is a conjugate distribution between the multinomial and Dirichlet distribution. In the same way, \( KT_k(x_{n \times n}|z_n) \) is the marginal conditional density of \( x_{n \times n} \) given \( z_n \) and it is the conjugate distribution of the Bernoulli and Beta distributions.

Using the fact that \( KT_k(z_n) \) is a distribution we can compute an approximation \( \hat{KT}_k(x_{n \times n}) \) of \( KT_k(x_{n \times n}) \) in the following way:

1. Sample \( z_1^n, \ldots, z_S^n \) from a distribution \( KT_k(\cdot) \) with support on \([k]^n\)

2. Compute \( \hat{KT}_k(x_{n \times n}) = \frac{1}{S} \sum_{i=1}^{S} KT_k(x_{n \times n}|z_i^n) \).

This computation avoid the maximization over \( \Theta^k \) in (2.3.1) and (2.3.2).
Chapter 3

The KT order estimator

In this chapter we present the estimator for the number of communities of the SBM based on the Krichevsky - Trofimov mixture distribution without assuming a known upper bound on the number of communities of the model. We also present the proof of strong consistency of the proposed estimator.

3.1 Order estimator

The KT order estimator in the context of a SBM is a regularized estimator of a mixture distribution for the random adjacency matrix $X_{n \times n}$. Given a sample $(x_{n \times n}, z_n)$ of the SBM the estimator only depends on the observed network $x_{n \times n}$. Thus, the KT order estimator is given by

$$
\hat{k}_{KT}(x_{n \times n}) = \arg \max_k \{ \log KT_k(x_{n \times n}) - \text{pen}(k, n) \} \tag{3.1.1}
$$

where $KT_k(x_{n \times n})$ is the mixture distribution for a SBM of order $k$ and $\text{pen}(k, n)$ is a penalizing function.

In order to derive the strong consistency result for the KT order estimator, we need a penalty function in (3.1.1) with a given rate of convergence when $n$ grows to infinity. Although there are a range of possibilities for this penalty function, the specific form we use in this thesis is

$$
\text{pen}(k, n) = \sum_{i=1}^{k-1} \frac{(i(i + 2) + 3 + \epsilon)}{2} \log n
\quad = \left[ \frac{k(k - 1)(2k - 1)}{12} + \frac{k(k - 1)}{2} + \frac{(3 + \epsilon)(k - 1)}{2} \right] \log n \tag{3.1.2}
$$

for some $\epsilon > 0$. The convenience of the expression above will become clear in the proof of the consistency result.

Now we can state the consistency result.
Theorem 3.1.1 (Consistency Theorem). Suppose the SBM has order $k_0$, with parameters $(\pi_0, P_0)$. Then, for a penalty function of the form (3.1.2) we have that
\[
\hat{k}_{\text{KT}} = k_0
\]
eventually almost surely as $n \to \infty$.

The proof of Theorem 3.1.1 is given in the next section.

3.2 Proof of the consistency theorem

The proof of Theorem 3.1.1 is divided in two main parts. The first one, presented in Subsection 3.2.1, proves that $\hat{k}_{\text{KT}}$ does not overestimate the true order $k_0$, eventually almost surely when $n \to \infty$, even without assuming a known upper bound on $k_0$. The second part of the proof, presented in Subsection 3.2.2, shows that $\hat{k}_{\text{KT}}$ does not underestimate $k_0$, eventually almost surely when $n \to \infty$. By combining these two results we prove that $\hat{k}_{\text{KT}} = k_0$ eventually almost surely as $n \to \infty$.

3.2.1 Non-overestimation

The main result in this subsection is given by the following proposition.

Proposition 3.2.1. Let $(x_{n \times n}, z_n^0)$ be a sample of size $n$ from a SBM of order $k_0$, with parameters $\pi_0$ and $P_0$. Then, the $\hat{k}_{\text{KT}}$ order estimator defined in (3.1.1) with penalty function given by (3.1.2) does not overestimate $k_0$, eventually almost surely when $n \to \infty$.

The proof of Proposition 3.2.1 follows straightforward from Lemmas 3.2.3, 3.2.4 and 3.2.5 presented below. The proofs of these Lemmas are inspired in the arguments used in the proof of the consistency of the BIC Markov order estimator in Csiszár et al. (2000) that are also applied to prove the consistency of the code-based estimators of the number of hidden states of a Hidden Markov Model in Gassiat e Boucheron (2003). However, first we state and prove a lemma that is useful to bound the probability of overestimation.

Lemma 3.2.2. For $k > k_0$ we have
\[
\mathbb{P}_{\theta_0}(\hat{k}_{\text{KT}} = k) \leq \exp \left\{ \frac{(k_0(k_0 + 2) - 1)}{2} \log n + c_{k_0,n} + \text{pen}(k_0,n) - \text{pen}(k,n) \right\},
\]
where $c_{k_0,n}$ is given in (2.3.5).
3.2 PROOF OF THE CONSISTENCY THEOREM

Proof. Observe that for \( k > k_0 \) we have

\[
\Pr_{\theta_0}(\hat{k}_{K_T} = k) = \sum_{x_{n \times n}} \Pr_{\theta_0}(x_{n \times n}) 1 \left\{ \arg \max_{k'} \{ \log KT_{k'}(x_{n \times n}) - \text{pen}(k', n) \} = k \right\}
\]

\[
\leq \sum_{x_{n \times n}} \Pr_{\theta_0}(x_{n \times n}) 1 \{ \log KT_k(x_{n \times n}) - \text{pen}(k, n) \geq \log KT_{k_0}(x_{n \times n}) - \text{pen}(k_0, n) \}
\]

\[
= \sum_{x_{n \times n}} \Pr_{\theta_0}(x_{n \times n}) 1 \{ KT_{k_0}(x_{n \times n}) \leq KT_k(x_{n \times n}) \exp [\text{pen}(k_0, n) - \text{pen}(k, n)] \}.
\]

(3.2.1)

By Proposition 2.3.3

\[
\log \Pr_{\theta_0}(x_{n \times n}) \leq \log \sup_{\theta \in \Theta} \Pr_{\theta}(x_{n \times n})
\]

\[
\leq \log KT_{k_0}(x_{n \times n}) + \left( \frac{k_0(k_0 + 2) - 1}{2} \right) \log n + c_{k_0, n}
\]

and therefore

\[
\Pr_{\theta_0}(x_{n \times n}) \leq KT_{k_0}(x_{n \times n}) n^{\left( \frac{k_0(k_0 + 2) - 1}{2} \right)} e^{c_{k_0, n}}.
\]

(3.2.2)

Applying (3.2.2) in (3.2.1) we obtain

\[
\Pr_{\theta_0}(\hat{k}_{K_T} = k)
\]

\[
\leq \sum_{x_{n \times n}} KT_{k_0}(x_{n \times n}) n^{\left( \frac{k_0(k_0 + 2) - 1}{2} \right)} e^{c_{k_0, n}} 1 \left\{ KT_{k_0}(x_{n \times n}) \leq KT_k(x_{n \times n}) e^{\text{pen}(k_0, n) - \text{pen}(k, n)} \right\}
\]

\[
\leq \sum_{x_{n \times n}} KT_k(x_{n \times n}) e^{\text{pen}(k_0, n) - \text{pen}(k, n)} n^{\left( \frac{k_0(k_0 + 2) - 1}{2} \right)} e^{c_{k_0, n}}
\]

\[
= \exp \left\{ \left( \frac{k_0(k_0 + 2) - 1}{2} \right) \log n + c_{k_0, n} + \text{pen}(k_0, n) - \text{pen}(k, n) \right\}
\]

where the last equality follows from the fact that KT_k(\cdot) is a distribution. \( \square \)

Lemma 3.2.3. Under the hypotheses of Proposition 3.2.1 we have that

\[
\hat{k}_{K_T} \notin (k_0, \log n]
\]

eventually almost surely when \( n \to \infty \).
Proof. Using Lemma 3.2.2 we can write

\[ P_{\theta_0}(\hat{k}_{KT} \in (k_0, \log n]) = \sum_{k=k_0+1}^{\log n} P_{\theta_0}(\hat{k}_{KT} = k) \]

\[ \leq \sum_{k=k_0+1}^{\log n} \exp \left\{ \frac{(k_0(k_0 + 2) - 1)}{2} \log n + c_{k_0,n} + \text{pen}(k_0, n) - \text{pen}(k, n) \right\} \]

\[ \leq e^{c_{k_0,n}} \log n \exp \left\{ \frac{(k_0(k_0 + 2) - 1)}{2} \log n + \text{pen}(k_0, n) - \text{pen}(k_0 + 1, n) \right\} \]

where the last inequality follows from the fact that \( \text{pen}(k, n) \) is increasing in \( k \).

Using the penalty given by (3.1.2) we obtain

\[ \frac{(k_0(k_0 + 2) - 1)}{2} \log n + \text{pen}(k_0, n) - \text{pen}(k_0 + 1, n) \]

\[ = \frac{(k_0(k_0 + 2) - 1)}{2} \log n + \sum_{i=1}^{k_0-1} \frac{(i(i + 2) + 3 + \epsilon)}{\log n} \log n - \sum_{i=1}^{k_0} \frac{(i(i + 2) + 3 + \epsilon)}{\log n} \log n \]

\[ = - \frac{1}{2} (2 + \epsilon/2) \log n \]

Let \( C_{k_0} \) be an upper-bound on \( \exp(c_{k_0,n}) \). Thus,

\[ \sum_{n} P_{\theta_0}(\hat{k}_{KT} \in (k_0, \log n]) \leq C_{k_0} \sum_{n} \frac{\log n}{n^{2+\epsilon/2}} < \infty \]

and the result follows by the first Borel Cantelli lemma.

\[ \square \]

Lemma 3.2.4. Under the hypotheses of Proposition 3.2.1 we have that

\( \hat{k}_{KT} \notin [\log n, n] \)

eventually almost surely when \( n \to \infty \).
3.2 PROOF OF THE CONSISTENCY THEOREM

Proof. We use again Lemma 3.2.2 to get

$$\mathbb{P}_{\theta_0}(\hat{k}_{KT} \in (\log n, n]) = \sum_{k=\log n}^{n} \mathbb{P}_{\theta_0}(\hat{k}_{KT} = k)$$

$$\leq \sum_{k=\log n}^{n} \exp\left\{ \frac{(k_0(k_0 + 2) - 1)}{2} \log n + c_{k_0,n} + \text{pen}(k_0, n) - \text{pen}(k, n) \right\}$$

$$\leq e^{c_{k_0,n}} n \exp\left\{ \frac{(k_0(k_0 + 2) - 1)}{2} \log n + \text{pen}(k_0, n) - \text{pen}(\log n, n) \right\}$$

$$= e^{c_{k_0,n}} n \exp\left\{ - \log n \left[ \frac{(k_0(k_0 + 2) - 1)}{2} - \frac{\text{pen}(k_0, n)}{\log n} + \frac{\text{pen}(\log n, n)}{\log n} \right] \right\}.$$ 

Since \( \text{pen}(k, n)/\log(n) \) only depends cubically in \( k \) we have that

$$\liminf_{n \to \infty} \frac{\text{pen}(\log n, n)}{\log n} - \frac{(k_0(k_0 + 2) - 1)}{2} - \frac{\text{pen}(k_0, n)}{\log n} > 3.$$

Thus,

$$\sum_{n} n \exp\left\{ - \log n \left[ \frac{(k_0(k_0 + 2) - 1)}{2} - \frac{\text{pen}(k_0, n)}{\log n} + \frac{\text{pen}(\log n, n)}{\log n} \right] \right\} < \infty.$$

Using the fact that \( \exp(c_{k_0,n}) \) is decreasing on \( n \), the result follows from the first Borel Cantelli lemma. 

Lemma 3.2.5. Under the hypotheses of Proposition 3.2.1 we have that

$$\hat{k}_{KT} \not\in (n, \infty)$$

eventually almost surely when \( n \to \infty \).

Proof. In order to prove the lemma, it is enough to prove that

$$\log KT_{n+m}(x_{n \times n}) - \text{pen}(n + m, n) \leq \log KT_n(x_{n \times n}) - \text{pen}(n, n)$$

for \( m \geq 1 \).

Using Proposition 2.3.3, we have that

$$- \log KT_n(x_{n \times n}) \leq \log - \sup_{\theta \in \Theta^n} \mathbb{P}_{\theta}(x_{n \times n}) + \left( \frac{n(n + 2)}{2} - \frac{1}{2} \right) \log n + c_{n,n}$$

By definition of KT we have

$$KT_{n+m}(x_{n \times n}) \leq \sup_{\theta \in \Theta^{n+m}} \mathbb{P}_{\theta}(x_{n \times n})$$
Thus,

\[
\log KT_{n+m}(x_{n\times n}) - \log KT_n(x_{n\times n}) \\
\leq \sup_{\theta \in \Theta^{n+m}} \mathbb{P}_\theta(x_{n\times n}) - \sup_{\theta \in \Theta^n} \mathbb{P}_\theta(x_{n\times n}) + \left( \frac{n(n+2)}{2} - \frac{1}{2} \right) \log n + c_{n,n} \\
\leq \frac{(n(n+2)-1)}{2} \log n + n(n+1) \left( \frac{\log \Gamma(\frac{1}{2})}{2} + \frac{7}{12} \right) + \frac{n(n-1)}{4n} + \frac{1}{12n} - \log \Gamma(\frac{n}{2}) \\
\leq \text{pen}(n + m, n) - \text{pen}(n, n)
\]

where the last inequality holds for \(n\) big enough.

### 3.2.2 Non-underestimation

In this subsection we deal with the proof of the non-underestimation of the proposed estimator.

**Proposition 3.2.6.** Let \((x_{n\times n}, z_0^n)\) be a sample of size \(n\) from a SBM of order \(k_0\), with parameters \((\pi_0, P_0)\). Then, the \(\hat{k}_{KT}\) order estimator defined in (3.1.1) with penalty function given by (3.1.2) does not underestimate \(k_0\), eventually almost surely when \(n \rightarrow \infty\).

In order to prove this result we need Lemma 3.2.7 below, that explore limiting properties of the under-fitted model. That is we handle with the problem of fitting, under \(\Theta^{k_0-1}\), a SBM with order \(k_0\).

We start by constructing a \((k_0-1)\)-order SBM \((\tilde{Z}_n, \tilde{X}_{n\times n})\) with parameters \((\tilde{\pi}, \tilde{P})\) from \((\pi_0, P_0)\), by a merging procedure that is defined below.

Let \(Q(\tilde{Z}_n, Z_n)\) be the \((k_0-1) \times k_0\) joint matrix of \(\tilde{Z}_n\) and \(Z_n\) whose the \((\tilde{\alpha}, \alpha)\)-th entry is given by

\[
Q_{\tilde{\alpha},\alpha}(\tilde{Z}_n, Z_n) = P(\tilde{Z}_i = \tilde{\alpha}, Z_i = \alpha) \quad \text{for} \ 1 \leq \alpha \leq k_0, \ 1 \leq \tilde{\alpha} \leq k_0 - 1
\]

The matrix \(Q(\tilde{Z}_n, Z_n)\) can be seen as a coupling between the random labels \(Z_n\) of the model of order \(k_0\) and the random labels \(\tilde{Z}_n\) of the model with \(k_0 - 1\) blocks.

An intuitive construction of a \((k_0 - 1)\)-block model is obtained by merging blocks of a model of order \(k_0\). This merging can be constructed in several ways, but we consider the construction given in Wang et al. (2017); however, instead of using the block proportions as considered in their work, we use the true probabilities \((\pi_0, P_0)\).

Given the probabilities \(\pi_0 = (\pi_0^1, \pi_0^2, \cdots, \pi_0^{k_0})\) we define a merging operation \(M_{a,b}(P_0^0, \pi_0)\) which combines blocks \(a\) and \(b\) in \(P_0^0\) by taking weighted averages with probabilities given
by $\pi^0$. For example, for $P = M_{k_0,k_0-1}(\pi^0, P^0)$

\[
P_{l,k} = P^0_{l,k} \quad \text{for } 1 \leq l, k \leq k_0 - 2
\]

\[
P_{l,0} = \frac{\pi^0_{l,k_0-1} P^0_{l,k_0-1} + \pi^0_k P^0_{l,k}}{\pi^0_{l,k_0-1} + \pi^0_k} \quad \text{for } 1 \leq l \leq k_0 - 2
\]

\[
P_{0,0} = \frac{\pi^0_{k_0-1} P^0_{k_0-1, k_0-1} + 2 \pi^0_{k_0-1} P^0_{k_0-1, k_0} + \pi^0_{k_0} P^0_{k_0, k_0}}{\pi^0_{k_0-1} \pi^0_{k_0-1} + 2 \pi^0_{k_0-1} \pi^0_{k_0} + \pi^0_{k_0} \pi^0_{k_0}}.
\]

In the same way, we consider the merged probabilities $\pi = M_{k_0,k_0-1}(\pi^0)$ as

\[
\pi_k = \pi^0_k \quad \text{for } 1 \leq k \leq k_0 - 2
\]

\[
\pi_{k_0-1} = \pi^0_{k_0-1} + \pi^0_{k_0}.
\]

In the general merging $(\pi, P) = (M_{a,b}(\pi^0), M_{a,b}(P^0, \pi^0))$, $b > a$, the nodes are relabeled using

\[
\tilde{c} = \begin{cases} 
c, & \text{if } c < b \\
-c+1, & \text{otherwise} 
\end{cases}
\]  

(3.2.3)

for $\tilde{c} = 1, 2, \ldots, k_0 - 1$.

Given $(x_{n \times n}, z_n)$ originated from the SBM of order $k_0$ and parameters $(\pi^0, P^0)$, we define the profile likelihood estimator of the label assignement under the $(k_0 - 1)$-block model as

\[
z^*_n = \arg \max_{z_n \in [k_0-1]^{n}} \sup_{\theta \in \Theta^{k_0-1}} \mathbb{P}_\theta(x_{n \times n}, z_n).
\]

The next lemma gives an upper bound for the limit of logarithm of the ratio between the true likelihood under the corrected model and the maximum profile likelihood function under the underfitting model that depends on the merging operation described above.

**Lemma 3.2.7.** Let $(x_{n \times n}, z^*_n)$ be a sample of size $n$ from a SBM of order $k_0$, with parameters $(\pi^0, P^0)$. We have that there exist $r, s \in [k_0]$ such that

\[
\liminf_{n \to \infty} \frac{1}{n^2} \log \frac{\mathbb{P}_{\theta_1}(x_{n \times n}, z^*_n)}{\sup_{\theta \in \Theta^{k_0-1}} \mathbb{P}_\theta(x_{n \times n}, z^*_n)} 
\geq \frac{1}{2} \sum_{a,b=1}^{k_0} \pi^0_a \pi^0_b \gamma(P^0_{ab}) - \frac{1}{2} \sum_{\tilde{a}, \tilde{b}=1}^{k_0-1} [M_{r,s}(\pi^0)]_{\tilde{a}} [M_{r,s}(\pi^0)]_{\tilde{b}} \gamma([M_{r,s}(\pi^0), P^0])_{\tilde{a}, \tilde{b}}
\]

(3.2.4)

\[
> 0.
\]

eventually almost surely when $n \to \infty$. In the formula above $\gamma(x) = x \log x + (1-x) \log(1-x)$.
Proof of Lemma 3.2.7. Define

$$
\hat{\pi}_a(z_n) = \frac{n_a(z_n)}{n}, \quad a = 1, \cdots, k_0
$$

$$
\hat{P}_{a,b}(z_n) = \frac{O_{a,b}(z_n, x_{n \times n})}{n_{a,b}(z_n)}, \quad a = 1, \cdots, k_0, \ b = a, \cdots, k_0.
$$

Using the fact that $x_{n \times n}$ given $z_n^0$ has distribution $P^0$ and $z_n^0$ is originated from $\pi^0$ we have that

$$
\log P_{\theta_0}(x_{n \times n}, z_n^0) = \sum_{a=1}^{k_0} n_a(z_n^0) \log \pi_a^0 \\
+ \frac{1}{2} \sum_{a=1}^{k_0} \sum_{b=a}^{k_0} O_{a,b}(z_n^0) \log P_{a,b}^0 + (n_a(z_n^0) - O_{a,b}(z_n^0)) \log (1 - P_{a,b}^0) \\
= \sum_{a=1}^{k_0} n_a(z_n^0) \log \pi_a^0 - \frac{1}{2} \sum_{a=1}^{k_0} n_a(z_n^0) \left[ \frac{O_{aa}(z_n^0)}{n_{a,a}(z_n^0)} \log P_{a,a}^0 + \left(1 - \frac{O_{aa}(z_n^0)}{n_{a,a}(z_n^0)} \right) \log (1 - P_{a,a}^0) \right] \\
+ \frac{1}{2} \sum_{a=1}^{k_0} \sum_{b=1}^{k_0} n_a(z_n^0) n_b(z_n^0) \left[ \frac{O_{a,b}(z_n^0)}{n_{a,b}(z_n^0)} \log P_{a,b}^0 + \left(1 - \frac{O_{a,b}(z_n^0)}{n_{a,b}(z_n^0)} \right) \log (1 - P_{a,b}^0) \right] \\
= \sum_{a=1}^{k_0} n_a(z_n) \log \pi_a - \frac{1}{2} \sum_{a=1}^{k_0} n_a(z_n) \left[ \hat{P}_{a,a}(z_n^0) \log P_{a,a}^0 + \left(1 - \hat{P}_{a,a}(z_n^0) \right) \log (1 - P_{a,a}^0) \right] \\
+ \frac{1}{2} \sum_{a=1}^{k_0} \sum_{b=1}^{k_0} n_a(z_n) n_b(z_n) \left[ \hat{P}_{a,b}(z_n^0) \log P_{a,b}^0 + \left(1 - \hat{P}_{a,b}(z_n^0) \right) \log (1 - P_{a,b}^0) \right]
$$

(3.2.5)

and, eventually almost surely when $n \to \infty$,

$$
\hat{\pi}_a(z_n^0) \xrightarrow{n \to \infty} \pi_a^0, \quad a = 1, \cdots, k_0
$$

$$
\hat{P}_{a,b}(z_n^0) \xrightarrow{n \to \infty} P_{a,b}^0, \quad a = 1, \cdots, k_0, \ b = a, \cdots, k_0.
$$

The first two terms on the right hand side of (3.2.5) is of smaller order compared to $n^2$, so eventually almost surely when $n \to \infty$,

$$
\lim_{n \to \infty} \frac{1}{n^2} \log P_{\theta_0}(x_{n \times n}, z_n^0) = \frac{1}{2} \sum_{a=1}^{k_0} \sum_{b=1}^{k_0} \pi_a^0 \pi_b^0 \log P_{a,b}^0 + \pi_a^0 \pi_b^0 (1 - P_{a,b}^0) \log (1 - P_{a,b}^0)
$$

(3.2.6)
PROOF OF THE CONSISTENCY THEOREM

3.2 Under $\Theta^{k_0 - 1}$ we have that

$$\log \sup_{\theta \in \Theta^{k_0 - 1}} \mathbb{P}_\theta(x_{n \times n}, z^*_n) = \sum_{a=1}^{k_0 - 1} n \hat{a}_n(z^*_n)\log \hat{a}_n(z^*_n) +$$

$$\sum_{a=1}^{k_0 - 1} \sum_{b=1}^{k_0 - 1} \frac{n^2}{2} \left[ \hat{\pi}_a(z^*_n)\hat{\pi}_b(z^*_n)\hat{P}_{a,b}(z^*_n) + \hat{\pi}_a(z^*_n)\hat{\pi}_b(z^*_n)(1 - \hat{P}_{a,b}(z^*_n)) \log(1 - \hat{P}_{a,b}(z^*_n)) \right]$$

$$- \sum_{a=1}^{k_0 - 1} \frac{n^2}{2} \left[ \hat{\pi}_a(z^*_n)\hat{P}_{a,a}(z^*_n) + \hat{\pi}_a(z^*_n)(1 - \hat{P}_{a,a}(z^*_n)) \log(1 - \hat{P}_{a,a}(z^*_n)) \right]$$

$$= \frac{n}{2} \sum_{a=1}^{k_0 - 1} \left[ 2 \log \hat{\pi}_a(z^*_n) - \gamma \left( \hat{P}_{a,a}(z^*_n) \right) \right] + \sum_{a=1}^{k_0 - 1} \sum_{b=1}^{k_0 - 1} \frac{n^2}{2} \left[ \hat{\pi}_a(z^*_n)\hat{\pi}_b(z^*_n) \gamma \left( \hat{P}_{a,b}(z^*_n) \right) \right].$$

(3.2.7)

First, observe that each term in the first sum of the right hand side of (3.2.7) is limited and it is of order smaller than $n^2$.

Using the fact that each term of the second sum of the right hand side of (3.2.7) is limited we have that $\lim \sup_{n \to \infty} \frac{1}{n^2} \log \sup_{\theta \in \Theta^{k_0 - 1}} \mathbb{P}_\theta(x_{n \times n}, z^*_n)$ exists.

Hence, there exists a subsequence $(z^*_{n_j})_{j \geq 1}$ of $(z^*_n)_{n \geq 1}$ such that

$$\lim_{j \to \infty} \frac{1}{n^2} \log \sup_{\theta \in \Theta^{k_0 - 1}} \mathbb{P}_\theta(x_{n \times n}, z^*_{n_j}) = \lim_{n \to \infty} \frac{1}{n^2} \log \sup_{\theta \in \Theta^{k_0 - 1}} \mathbb{P}_\theta(x_{n \times n}, z^*_n).$$

(3.2.8)

Given any string $z^*_n \in [k_0 - 1]^n$, each of the terms $\hat{\pi}_a(z^*_n)$ and $\hat{P}_{a,b}(z^*_n)$ are bounded in absolute value by 1. In particular, this holds true for $z^*_n = z^*_{n_j}$ being any element in the subsequence $(z^*_{n_j})$ constructed above. This implies that we can extract a subsequence $(z^*_{m_k})$ of $(z^*_{n_j})$ for which

$$\lim_{k \to \infty} \hat{\pi}_a(z^*_{m_k}) = \bar{\pi}_a, \quad \lim_{k \to \infty} P_{a,b}(z^*_{m_k}) = \bar{P}_{a,b},$$

for some numbers $\bar{\pi}_a$, $\bar{P}_{a,b}$, and any $a, b \in \{1, \ldots, k_0 - 1\}$. Combined with (3.2.8), we thus get that

$$\lim \sup_{n \to \infty} \frac{1}{n^2} \log \sup_{\theta \in \Theta^{k_0 - 1}} \mathbb{P}_\theta(x_{n \times n}, z^*_n) = \frac{1}{2} \sum_{a=1}^{k_0 - 1} \sum_{b=1}^{k_0 - 1} \bar{\pi}_a \bar{\pi}_b \gamma(\bar{P}_{a,b}).$$

(3.2.9)
Combining (3.2.6) and (3.2.9) we have that

\[
\lim_{n \to \infty} \frac{1}{n^2} \log \frac{\mathbb{P}_{\theta_0}(x_{n \times n}, z_{n}^0)}{\mathbb{P}(x_{n \times n}, z_{n}^0)} \sup_{(\theta) \in \Theta_{k_0 - 1}} \mathbb{P}(x_{n \times n}, z_{\ast})
\]

\[
= \frac{1}{2} \sum_{a=1}^{k_0} \sum_{b=1}^{k_0} \pi_a^0 \pi_b^0 \gamma(P_{a,b}^0) - \lim_{n \to \infty} \frac{1}{n^2} \log \sup_{(\theta) \in \Theta_{k_0 - 1}} \mathbb{P}(x_{n \times n}, z_{\ast})
\]

\[
= \frac{1}{2} \sum_{a=1}^{k_0} \sum_{b=1}^{k_0} \pi_a^0 \pi_b^0 \gamma(P_{a,b}^0) - \frac{1}{2} \sum_{a=1}^{k_0 - 1} \sum_{b=1}^{k_0 - 1} \pi_a \pi_b \gamma(\tilde{P}_{a,b})
\]

eventually almost surely when \( n \to \infty \). To obtain a lower bound of (3.2.10), we need to obtain \((\tilde{\pi}, \tilde{P})\) that minimizes

\[
\sum_{a=1}^{k_0} \sum_{b=1}^{k_0} \pi_a^0 \pi_b^0 \gamma(P_{a,b}^0) - \sum_{a=1}^{k_0 - 1} \sum_{b=1}^{k_0 - 1} \pi_a \pi_b \gamma(\tilde{P}_{a,b})
\]

This is equivalent to obtain \((\tilde{\pi}, \tilde{P})\) that maximizes

\[
\sum_{a=1}^{k_0 - 1} \sum_{b=1}^{k_0 - 1} \pi_a \pi_b \gamma(\tilde{P}_{a,b}).
\]

By definition

\[
\tilde{P}_{a,b} = \frac{P(X_{i,j} = 1, \tilde{Z}_i = a, \tilde{Z}_j = b)}{P(Z_i = a, Z_j = b)}
\]

The numerator equals

\[
\sum_{a=1}^{k_0} \sum_{b=1}^{k_0} P(\tilde{Z}_i = a, \tilde{Z}_j = b | Z_i = a, Z_j = b)P(X_{i,j} = 1 | Z_i = a, Z_j = b)P(Z_i = a)P(Z_j = b)
\]

\[
= \sum_{a=1}^{k_0} \sum_{b=1}^{k_0} P(\tilde{Z}_i = a | Z_i = a)P(Z_i = a) P_{a,b}^0 P(\tilde{Z}_j = b | Z_j = b)P(Z_j = b)
\]

\[
= \sum_{a=1}^{k_0} \sum_{b=1}^{k_0} P(\tilde{Z}_i = a, Z_i = a) P_{a,b}^0 P(\tilde{Z}_j = b, Z_j = b) = (QP^0 Q^T)_{a,b}
\]
and the denominator equals
\[
\sum_{a=1}^{k_0} \sum_{b=1}^{k_0} P(\tilde{Z}_i = \tilde{a}, \tilde{Z}_j = \tilde{b} \mid Z_i = a, Z_j = b) P(Z_i = a, Z_j = b)
\]
\[
= \sum_{a=1}^{k_0} \sum_{b=1}^{k_0} P(\tilde{Z}_i = \tilde{a} \mid Z_i = a) P(\tilde{Z}_j = \tilde{b} \mid Z_j = b) P(Z_i = a) P(Z_j = b)
\]
\[
= \sum_{a=1}^{k_0} \sum_{b=1}^{k_0} P(\tilde{Z}_i = \tilde{a}, Z_i = a) P(\tilde{Z}_j = \tilde{b}, Z_j = b) = (Q(11^T)Q^T)_{\tilde{a}, \tilde{b}}.
\]

Then rewrite (3.2.11) as
\[
\sum_{a=1}^{k_0-1} \sum_{b=1}^{k_0-1} (Q(11^T)Q^T)_{a,b} \gamma \left[ \frac{(QP^0Q^T)_{a,b}}{(Q(11^T)Q^T)_{a,b}} \right].
\]

Observe that the conditional probabilities \( \tilde{P}_{a,b} \) are completely determined by the coupling \( Q_{a,a}(Z_a, \tilde{Z}_n) \) and the parameters \( P_{a,b}^0 \). Therefore, to find a pair \( (\tilde{\pi}, \tilde{P}) \) maximizing (3.2.11) is equivalent to find an optimal coupling maximizing (3.2.12).

In Wang et al. (2017) it has been shown that there exists \( r \) and \( s \), \( r, s \in [k_0] \), such that (3.2.12) achieves its maximum for \( \tilde{\pi} = M_{r,s}(\pi^0) \) and \( \tilde{P} = M_{r,s}(\pi^0, P^0) \). This concludes the proof of the first inequality in (3.2.4).

In order to prove the second inequality in (3.2.4), we consider, for convenience and without loss of generality, \( r = k_0 \) and \( s = k_0 - 1 \). Define the function
\[
\varphi(x, y) = x \log \left( \frac{x}{y} \right)
\]

We rewrite
\[
\sum_{a=1}^{k_0} \sum_{b=1}^{k_0} P_{a,b} \log P_{a,b} = \sum_{a=1}^{k_0-2} \sum_{b=1}^{k_0-2} \varphi(\pi_a P_{a,b}, \pi_b)
\]
\[
+ \sum_{a=1}^{k_0-2} \left( \varphi(\pi_a P_{a,k_0}, \pi_a \pi_{k_0}) + \varphi(\pi_a P_{a,k_0-1}, \pi_a \pi_{k_0}) \right)
\]
\[
+ \varphi(\pi_{k_0-1} P_{k_0-1,k_0-1}, \pi_{k_0-1}) + \varphi(P_{k_0-1,k_0-1}, \pi_{k_0-1})
\]
\[
+ \varphi(\pi_{k_0-1} P_{k_0-1,k_0}, \pi_{k_0-1})
\]
and, by definition of the merging \( M_{k_0-1,k_0}(\pi^0, P^0) \),
To prove Proposition 3.2.6 it is enough to show that for all \( k_0 - 1 \leq k_0 - 1 \)
\[
\sum_{a=1}^{k_0-2} \sum_{b=1}^{k_0-2} [M_{k_0-1,k_0}(\pi^0)]_a [M_{k_0-1,k_0}(\pi^0)]_b \log([M_{k_0-1,k_0}(\pi^0, P^0)]_a)_b 
\]
\[
= \sum_{a=1}^{k_0-2} \sum_{b=1}^{k_0-2} \varphi(\pi^0_{a,b} P_{a,b}, \pi^0_{a,b}) + \sum_{a=1}^{k_0-2} \varphi(\pi^0_{a,k_0} P_{a,k_0}, \pi^0_{a,k_0} + \pi^0_{a,k_0-1}) 
\]
\[
+ \varphi(\pi^0_{k_0-1} P_{k_0-1,k_0-1} + 2\pi^0_{k_0-1} P^0_{k_0-1,k_0}, (\pi^0_{k_0-1})^2) 
\]
\[
+ \varphi(\pi^0_{k_0-1} P_{k_0-1,k_0-1} + 2\pi^0_{k_0-1} P^0_{k_0-1,k_0}, (\pi^0_{k_0-1})^2) 
\]

Using the log-sum inequality we have that
\[
\sum_{a=1}^{k_0-2} \varphi(\pi^0_{a,k_0} P_{a,k_0}, \pi^0_{a,k_0}) + \varphi(\pi^0_{a,k_0-1} P_{a,k_0-1}, \pi^0_{a,k_0-1}) \geq \sum_{a=1}^{k_0-2} \varphi(\pi^0_{a,k_0} P_{a,k_0} + \pi^0_{a,k_0-1} P_{a,k_0-1}, \pi^0_{a,k_0} + \pi^0_{a,k_0-1}) 
\]

and
\[
\varphi((\pi^0_{k_0})^2 P_{k_0,k_0}, (\pi^0_{k_0})^2) + \varphi((\pi^0_{k_0-1})^2 P_{k_0-1,k_0-1}, (\pi^0_{k_0-1})^2) + \varphi(\pi^0_{k_0-1} P_{k_0-1,k_0-1}, \pi^0_{k_0-1} \pi^0_{k_0}) 
\]
\[
+ \varphi(\pi^0_{k_0-1} P_{k_0-1,k_0-1}, \pi^0_{k_0-1} \pi^0_{k_0}) 
\]
\[
\geq \varphi((\pi^0_{k_0})^2 P^0_{k_0-1,k_0-1} + 2\pi^0_{k_0-1} P^0_{k_0-1,k_0}, (\pi^0_{k_0})^2 P^0_{k_0,k_0}, (\pi^0_{k_0-1})^2 + 2\pi^0_{k_0-1} \pi^0_{k_0} + (\pi^0_{k_0})^2). 
\]

The equality in (3.2.13) and (3.2.14) hold if and only if for \( 1 \leq a \leq b \leq k_0 \)
\[
P_{a,k_0} = P_{a,k_0-1}. 
\]

Then we conclude that the matrix \( P^0 \) has the last two columns equal. However, by assumption, \( k_0 \) is the order of the SBM that originated the graph \( x_{n \times n} \), that means, \( P^0 \) does not have identical columns. So, there exists \( a_1 \in [k_0] \), such that \( P_{a_1,k_0} \neq P_{a_1,k_0-1} \). Hence the result follows.

Proof of Proposition 3.2.6. Let \( (x_{n \times n}, z^0_n) \) be a sample of size \( n \) from a SBM of order \( k_0 \).

To prove Proposition 3.2.6 it is enough to show that for all \( k'< k_0 \)
\[
\log KT_{k_0}(x_{n \times n}) - \text{pen}(k_0, n) \geq \log KT_{k'}(x_{n \times n}) - \text{pen}(k', n) 
\]

eventually almost surely when \( n \to \infty \).

Using the fact that
it suffices to show that, eventually almost surely when \( n \to \infty \),

\[
\lim_{n \to \infty} \frac{\text{pen}(k_0, n) - \text{pen}(k', n)}{n^2} = 0
\]

We write

\[
\log \frac{\text{KT}_{k_0}(x_{n \times n})}{\text{KT}_{k'}(x_{n \times n})} = \log \frac{\text{KT}_{k_0}(x_{n \times n})}{\mathbb{P}_{\theta_0}(x_{n \times n})} + \log \frac{\mathbb{P}_{\theta_0}(x_{n \times n})}{\text{KT}_{k'}(x_{n \times n})}.
\]

We also have that

\[
\log \frac{\mathbb{P}_{\theta_0}(x_{n \times n})}{\text{KT}_{k'}(x_{n \times n})} \geq \log \frac{\mathbb{P}_{\theta_0}(x_{n \times n})}{\sup_{\theta \in \Theta} \mathbb{P}_{\theta}(x_{n \times n})} + \log \frac{\sup_{\theta \in \Theta'} \mathbb{P}_{\theta}(x_{n \times n})}{\text{KT}_{k'}(x_{n \times n})}.
\]

Combining (3.2.16) and (3.2.17) we have

\[
\frac{1}{n^2} \log \frac{\text{KT}_{k_0}(x_{n \times n})}{\text{KT}_{k'}(x_{n \times n})} \geq - \left( \frac{k_0(k_0 + 2)}{2} - \frac{1}{2} \right) \frac{n^2}{n^2} - \frac{c_{k_0,n}}{n}.
\]

Thus, to show (3.2.15) it suffices to show, for \( k' < k_0 \), that

\[
\liminf_{n \to \infty} \frac{1}{n^2} \log \frac{\mathbb{P}_{\theta_0}(x_{n \times n})}{\sup_{\theta \in \Theta'} \mathbb{P}_{\theta}(x_{n \times n})} > 0.
\]

3.2.18
Using $z^*_n = \arg \max_{z_n \in \{k_0 - 1\}^n} \sup_{\theta \in \Theta^{k_0-1}} \mathbb{P}_\theta(x_{n \times n}, z_n)$ we have

$$
\sup_{\theta \in \Theta^{k_0-1}} \mathbb{P}_\theta(x_{n \times n}) = \sup_{\theta \in \Theta^{k_0-1}} \sum_{z_n \in \{k_0 - 1\}^n} \mathbb{P}_\theta(x_{n \times n}, z_n) \leq \sum_{z_n \in \{k_0 - 1\}^n} \sup_{\theta \in \Theta^{k_0-1}} \mathbb{P}_\theta(x_{n \times n}, z_n^*) \leq (k_0 - 1)^n \sup_{\theta \in \Theta^{k_0-1}} \mathbb{P}_\theta(x_{n \times n}, z_n^*).$$

Taking the log

$$
\log \sup_{\theta \in \Theta^{k_0-1}} \mathbb{P}_\theta(x_{n \times n}) \leq n \log(k_0 - 1) + \log \sup_{\theta \in \Theta^{k_0-1}} \mathbb{P}_\theta(x_{n \times n}, z_n^*).
$$

Thus,

$$
\log \mathbb{P}_{\theta_0}(x_{n \times n}) = \log \sum_{z_n \in \{k_0\}^n} \mathbb{P}_{\theta_0}(x_{n \times n}, z_n) \geq \log \mathbb{P}_{\theta_0}(x_{n \times n}, z_n^0)
$$

and

$$
\log \frac{\mathbb{P}_{\theta_0}(x_{n \times n})}{\sup_{\theta \in \Theta^{k_0-1}} \mathbb{P}(x_{n \times n})} \geq \log \frac{\mathbb{P}_{\theta_0}(x_{n \times n}, z_n^0)}{(k_0 - 1)^n \sup_{\theta \in \Theta^{k_0-1}} \mathbb{P}_\theta(x_{n \times n}, z_n^*)} \geq \log \frac{\mathbb{P}_{\theta_0}(x_{n \times n}, z_n^0)}{\sup_{\theta \in \Theta^{k_0-1}} \mathbb{P}_\theta(x_{n \times n}, z_n^*)} - n \log(k_0 - 1).
$$

Thus, to show (3.2.18) it is enough to have

$$
\lim_{n \to \infty} \frac{1}{n^2} \log \frac{\mathbb{P}_{\theta_0}(x_{n \times n}, z_n^0)}{\sup_{\theta \in \Theta^{k_0-1}} \mathbb{P}_\theta(x_{n \times n}, z_n^*)} > 0
$$

eventually almost surely when $n \to \infty$. Using Lemma (3.2.7) the result follows.

To complete the proof, for $k' < k_0 - 1$ we write

$$
\lim_{n \to \infty} \frac{1}{n^2} \log \frac{\mathbb{P}_{\theta_0}(x_{n \times n})}{\sup_{\theta \in \Theta^{k_0-1}} \mathbb{P}(x_{n \times n})} = \lim_{n \to \infty} \frac{1}{n^2} \log \frac{\mathbb{P}_{\theta_0}(x_{n \times n})}{\sup_{\theta \in \Theta^{k_0-1}} \mathbb{P}_\theta(x_{n \times n})} + \lim_{n \to \infty} \frac{1}{n^2} \log \frac{\sup_{\theta \in \Theta^{k_0-1}} \mathbb{P}_\theta(x_{n \times n})}{\sup_{\theta \in \Theta^{k_0'}} \mathbb{P}_\theta(x_{n \times n})} \\
\geq \lim_{n \to \infty} \frac{1}{n^2} \log \frac{\mathbb{P}_{\theta_0}(x_{n \times n})}{\sup_{\theta \in \Theta^{k_0-1}} \mathbb{P}_\theta(x_{n \times n})} > 0
$$

where the first inequality follows from the fact that $\sup_{\theta \in \Theta^{k_0-1}} \mathbb{P}_\theta(x_{n \times n}) \geq \sup_{\theta \in \Theta^{k_0'}} \mathbb{P}_\theta(x_{n \times n})$. □
Part II

Exponential Random Graphs
Chapter 4

Exponential Random Graphs

In this chapter we introduce the concept of Exponential Random Graphs (ERG) and we construct a time-reversible Markov chain for which the invariant measure is the ERG distribution using the Glauber dynamics.

4.1 Exponential Random Graph

4.1.1 Definitions and Notations

In this model, we consider that the number of vertices of the graph is fixed and finite, that is, \(V = \{1, 2, \ldots, N\}\).

Let \(G_N\) be the collection of all undirected graphs with set of vertices \(V\). For convenience, we omit the notation of dependence of the graph on \(N\). So, let \(x\) be an observed graph in \(G_N\). We write \(x_{-ij}\) to represent the set of values \(\{x(l, k) : (i, j) \neq (l, k)\}\). For a given graph \(x\) and a set of vertices \(U \subseteq V\), we denote by \(x(U)\) the subgraph induced by \(U\), that is \(x(U)\) is a graph with set of vertices \(U\) and such that \(x(U)(i, j) = x(i, j)\) for all \(i, j \in U\).

The empty graph \(x^{(0)}\) is such that \(x^{(0)}(i, j) = 0\), for \(1 \leq i < j \leq N\). In the same way, the complete graph \(x^{(1)}\) is such that \(x^{(1)}(i, j) = 1\), for \(1 \leq i < j \leq N\).

Let \(E = \{(i, j) : i \in V, j \in V\text{ and } i < j\}\) be the set of possible edges in a graph with set of vertices \(V\). For any graph \(x \in G_N\), a pair of vertices \((i, j) \in E\) and \(a \in \{0, 1\}\), the modified graph \(x^a_{ij}\) is given by

\[
x^a_{ij}(k, l) = \begin{cases} x(k, l), & \text{if } (k, l) \neq (i, j); \\ a, & \text{if } (k, l) = (i, j). \end{cases}
\]

For the ERG model the probability of selecting a graph in \(G_N\) depends on its structure; for example, the number of edges, the number of triangles, the length of the longest path
and so on. The probability of observing a graph $x$ is given by

$$p_N(x|\theta) = \frac{\exp(\theta^T s(x))}{z_N(\theta)}$$

(4.1.1)

where $\theta$ is a $s \times 1$ vector of parameters, $s(x)$ is a $s \times 1$ vector of graph’s statistics and $z_N(\theta)$ is the normalizing constant of the model given by $z_N(\theta) = \sum_{x \in \mathcal{G}_N} \exp(\theta^T s(x))$.

A specific model of ERG explored in this thesis is called edges-2-stars-triangles model. In this model, the graph’s statistics are given by the number of edges, 2-stars and triangles. In terms of the probability (4.1.1), we write

$$p_N(x|\theta) = \frac{1}{z_N(\theta)} \exp \left( \theta_1 \sum_{i<j} x(i,j) + \theta_2 \sum_{i<j,k \neq i,j} x(i,k)x(j,k) + \theta_3 \sum_{i<j<k} x(i,j)x(i,k)x(j,k) \right).$$

The Erdős-Rényi model is a particular case of the ERG model when we have $\theta_2 = 0$ and $\theta_3 = 0$.

In different works, such as in Chatterjee et al. (2013) and Bhamidi et al. (2011), an alternative definition of the ERG distribution is used for models for which the graph’s statistics of $x$ are only represented by subgraphs of $x$. In this representation, the vector of statistics used in (4.1.1) can be replaced by the number of subgraphs contained in the graph $x$. For example, the statistic that counts the number of triangles in a graph can be computed by counting the number of completed graphs with three vertices that are contained in $x$.

In order to define subgraphs counts in a graph, let $V_m$ be the set of all possible permutations of $m$ distinct elements of $V$. For $v_m \in V_m$, define $x(v_m)$ as the subgraph of $x$ induced by $v_m$, that is $x(v_m)$ is a graph with set of vertices $v_m$ such that if there exist the edge $(i,j)$ in the graph $x$, for two vertices $i$ and $j$ in $v_m$, then there exist the edge $(i,j)$ in $x(v_m)$. For a graph $g$ with $m$ vertices, we say $x(v_m)$ contains $g$ (and we write $x(v_m) \succeq g$) if $g(i,j) = 1$ implies $x(v_m)(i,j) = 1$. Then, for $x \in \mathcal{G}_N$ and $g \in \mathcal{G}_m$, $m \leq N$, we define the number of subgraphs $g$ in $x$ by the counter

$$N_g(x) = \sum_{v_m \in V_m} \mathbf{1}\{x(v_m) \succeq g\}.$$  

(4.1.2)

Let $g_1, \ldots, g_s$ be a sequence of fixed graphs, where $g_i$ has $m_i$ vertices, $m_i \leq N$ and $\beta \in \mathbb{R}^s$ a vector of parameters. By convention we set $g_1$ as the graph with only two vertices and one edge. Following Chatterjee et al. (2013) and Bhamidi et al. (2011) we define the probability of graph $x$ by

$$p_N(x|\beta) = \frac{1}{Z_N(\beta)} \exp \left( \sum_{i=1}^s \beta_i \frac{N_{g_i}(x)}{N_{m_i}-2} \right).$$  

(4.1.3)
It is possible to relate the parameters $\theta$ and $\beta$ of the two equivalent definitions of the probability of the ERG. For instance, consider the model that counts the number of edges and triangles of the graph. Since (4.1.2) counts twice the number of edges of the graph and six times the number of triangles, we have that $\theta_1 = 2\beta_1$ and $\theta_2 = \frac{6\beta_2}{N}$. More generally, if $g_s$ is a fixed graph with $m_s$ vertices and $A(g_s)$ is the number of automorphisms of $g_s$, then $\theta_s = \frac{A(g_s)\beta_s}{N^{m_s-2}}$. A graph automorphism is a permutation of the vertices of the graph that preserves the edges of the graph, as shown in Figure (4.1).

The model (4.1.3) is used in some proofs of this thesis.

\begin{figure}[h]
\centering
\begin{tikzpicture}
  \node (1) at (0,0) {1};
  \node (2) at (1,1) {2};
  \node (3) at (1,-1) {3};
  \node (4) at (2,0) {4};

  \draw (1) -- (2);
  \draw (1) -- (3);
  \draw (2) -- (4);
  \draw (3) -- (4);
\end{tikzpicture}
\caption{Example of a graph automorphism.}
\end{figure}

4.2 Glauber dynamics

In the ERG models, the computation of the normalizing constant $z_N(\theta) = \sum_{x \in G_N} \exp(\theta^T s(x))$ (or $Z_N(\beta)$) might be extremely demanding due to the size of $G_N$. For example, for a graph with 10 vertices, we have that $|G_N| = 2^{45}$.

The algorithms used to simulate from the ERG distribution avoid the computation of this constant applying the MCMC methodology. In this work we develop techniques that consider the main idea of the MCMC algorithms: the construction of a time-reversible Markov chain for which the invariant measure is the ERG distribution.

In this section we show how to construct the time-reversible Markov chain with stationary distribution $p_N(\cdot; \theta)$ using a local update algorithm called Glauber dynamics, also known as Gibbs sampler. This Markov chain has as states the graphs with $N$ vertices, so the state space is the set $G_N$.

Using the Glauber dynamics the probability of a transition from the graph $x$ to the modified graph $x'_{ij}$ is

$$p_{ij}(a|\theta, x_{-ij}) = \frac{p_N(x''_{ij}|\theta)}{p_N(x'_{ij}|\theta) + p_N(x''_{ij}|\theta)},$$
In the case of ERG we can write

\[
p_{ij}(1|\theta, x_{-ij}) = \frac{p_N(x_{ij}^1|\theta)}{p_N(x_{ij}^0|\theta) + p_N(x_{ij}^1|\theta)} = \frac{1}{1 + \frac{p_N(x_{ij}^0|\theta)}{p_N(x_{ij}^1|\theta)}} = \frac{1}{1 + \exp(\theta^T s(x_{ij}^0) - s(x_{ij}^1))}.
\]

In the same way,

\[
p_{ij}(0|\theta, x_{-ij}) = 1 - p_{ij}(1|\theta, x_{-ij}) = \frac{1}{1 + \exp(\theta^T s(x_{ij}^0) - s(x_{ij}^1))}.
\]

For the specific model edges-2-stars-triangles, we have that

\[
p_{ij}(1|\theta, x_{-ij}) = \frac{1}{1 + \exp(-\theta_1 - \theta_2 \sum_{k \neq i,j} [x(i,k) + x(j,k)] - \theta_3 \sum_{k \neq i,j} x(i,k)x(j,k))}
\]

\[
p_{ij}(0|\theta, x_{-ij}) = \frac{1}{1 + \exp(\theta_1 + \theta_2 \sum_{k \neq i,j} [x(i,k) + x(j,k)] + \theta_3 \sum_{k \neq i,j} x(i,k)x(j,k))}.
\]

The construction of a Markov chain with invariant distribution \(p_N(\cdot; \theta)\) can be done with the help of the local update function \(\phi : G_N \times E \times [0,1] \rightarrow G_N\) defined as

\[
\phi(x, (i,j), u) = \begin{cases} 
  x_{ij}^0, & \text{if } u \leq p_{ij}(0|\theta, x_{-ij}) \\
  x_{ij}^1, & \text{otherwise}. 
\end{cases}
\]

A transition in the Markov chain from graph \(x\) to graph \(y\) is obtained by the random function

\[
y = \phi(x, e, u), \quad \text{with } e \sim \text{Uniform}(E), \: u \sim \text{Uniform}(0,1).
\]

In this way, the Markov chain with invariant distribution \(p_N(\cdot; \theta)\) is constructed in the following manner

1. Set \(n = 1\) and \(x_0\) as the initial state.
2. Select a pair of vertices \((i, j)\) uniformly on the set \(E\).
3. Generate the random number \(u_n\) from the Uniform distribution on the interval \([0,1]\).
4. Set \(x_n = \phi(x_{n-1}, (i, j), u_n)\).
5. Update \(n\) to \(n + 1\).
6. Repeat 2 – 6 until $n > T$

7. Return $x_0, \ldots, x_T$

The time-reversible Markov chain $(X_n)_{n \geq 0}$ with the desired invariant measure $p_N(\cdot | \theta)$ has the transition probabilities given by

$$P(x, y) = \begin{cases} \frac{1}{|E|} p_{ij}(a | \theta, x_{-ij}), & \text{if } y = x_a^{ij}, a \in \{0, 1\} \\ 0, & \text{otherwise}. \end{cases}$$ (4.2.3)

For the reader’s convenience we prove in Proposition 4.2.1 that the construction described above generates a Markov chain with the desired invariant measure.

**Proposition 4.2.1.** The Markov chain $(X_n)_{n \geq 0}$ with transition probabilities given by (4.2.3) is reversible with respect to the ERG probability distribution $p_N(\cdot | \theta)$.

**Proof.** We need to verify that $p_N(x | \theta) P(x, y) = p_N(y | \theta) P(y, x)$, for all $x, y \in \mathcal{G}_N$.

If $y$ is such that $x_{-ij} \neq y_{-ij}$ the equality above is verified, since $P(x, y) = P(y, x) = 0$.

Without loss of generality, let $x$ be a graph with $x(i, j) = a$, for some $i, j \in V$.

Consider the case that $y_{-ij} = x_{-ij}$.

If $y(i, j) = a$, then $x = y$ and the equality follows.

If $y(i, j) = 1 - a$ we have

$$p_N(x | \theta) P(x, y) = p_N(x_a^{ij} | \theta) P(x_a^{ij}, x_{ij}^{1-a}) = p_N(x_a^{ij} | \theta) \frac{1}{|E|} p_{ij}(1 - a | \theta, x_{-ij})$$

$$= p_N(x_a^{ij} | \theta) \frac{1}{|E|} \frac{p_N(x_a^{1-a} | \theta)}{p_N(x_a^{0} | \theta) + p_N(x_a^{1} | \theta)} = p_N(x_a^{1-a} | \theta) \frac{1}{|E|} \frac{p_N(x_a^{0} | \theta)}{p_N(x_a^{0} | \theta) + p_N(x_a^{1} | \theta)}$$

$$= p_N(x_{ij}^{1-a} | \theta) \frac{1}{|E|} p_{ij}(a | \theta, x_{-ij}) = p_N(x_{ij}^{1-a} | \theta) P(x, x_a^{ij})$$

$$= p_N(y | \theta) P(y, x).$$

Thus the result follows. \qed
Chapter 5

Coupling from the past

In this chapter we present the construction of the Coupling from the past (CFTP) algorithm for the ERG models.

5.1 The algorithm

In order to get a sample from the stationary distribution of a Markov chain, an alternative approach introduced by Propp e Wilson (1996) is called Coupling From The Past (CFTP). This method yields a value distributed under the exact target distribution, and is thus called perfect (also known as exact) simulation. Like MCMC, it does not require computing explicitly the normalising constant; but it does not require any burn-in phase either, nor any detection of convergence to the target distribution. In our approach, we adapt a version of the CFTP algorithm, which is developed for the spin systems in Propp e Wilson (1996), to construct a perfect simulation algorithm for the ERG model using the Glauber dynamics. The ERG model can be interpreted as a system of particles, where in this case the particles are the edges having spin 0 or 1.

For random vectors $e_{-n}^{-1} = (e_{-n}, e_{-n+1}, \ldots, e_{-1})$ and $u_{-n}^{-1} = (u_{-n}, u_{-n+1}, \ldots, u_{-1})$, with $e_{-k} \in E$ and $u_{-k} \in [0, 1]$, we define the random map $F_{-n}^0$ given by the following induction:

$$F_{-1}^0(x, e_{-1}^{-1}, u_{-1}^{-1}) = \phi(x, e_{-1}, u_{-1}), \text{ and}$$

$$F_{-n}^0(x, e_{-n}^{-1}, u_{-n}^{-1}) = F_{-n+1}^0(F_{-1}^0(x, e_{-n}^{-1}, u_{-n}^{-1}), e_{-n+1}^{-1}, u_{-n+1}^{-1}), \text{ for } n \geq 2.$$

The CFTP protocol of (Propp e Wilson, 1996) relies on the following elementary observation. If, for some time $-n$ and for some uniformly distributed $e_{-n}^{-1} = (e_{-n}, \ldots, e_{-1})$ and $u_{-n}^{-1} = (u_{-n}, \ldots, u_{-1})$, the mapping $F_{-n}^0(\cdot, e_{-n}^{-1}, u_{-n}^{-1})$ is constant, that is if it takes the same value at time 0 on all possible graphs $x$, then this constant value is readily seen to be a sample from the invariant measure of the Markov chain. Moreover, if $F_{-n}$ is constant for some positive $n$, then it is also constant for all larger values of $n$. Thus, it suffices to find a value of $n$ large enough so as to obtain a constant map, and to return the constant value of this map. For example, one may try some arbitrary value $n_1$, check if the map $F_{-n_1}^0(\cdot, e_{-n_1}^{-1}, u_{-n_1}^{-1})$
is constant, try \( n_2 = 2n_1 \) otherwise, and so on...

Since the state space has the huge size \( |G_N| = 2^{\frac{N(N-1)}{2}} \), it would be computationally intractable to compute the value of \( F_n^{0}(x, e^{-1}, u^{-1}) \) for all \( x \in G_N \) in order to determine if they coincide.

This is where the monotonicity property of the distribution helps: it is sufficient to inspect the maximal and minimal elements of \( G_N \). If they both yield the same value, then all other initial states will also coincide with them, and the mapping will be constant.

To apply the monotone CFTP algorithm we define an ordering \( \preceq \) on \( G_N \) in the following way: two graphs satisfy \( x \preceq y \) if, and only if, \( x(i, j) \leq y(i, j) \), for all \( i, j \in V \). In other words, \( x \preceq y \) if and only if \( x(i, j) = 1 \) implies \( y(i, j) = 1 \).

Notice that the ordering defined above is not a total ordering; that is, there are graphs in \( G_N \) that cannot be pairwise compared. On the other hand, there exists a maximal graph \( x^{(1)} \) (the complete graph) such that \( z \preceq x^{(1)} \), for all \( z \in G_N \) and a minimal graph \( x^{(0)} \) (the empty graph) such that \( x^{(0)} \preceq z \), for all \( z \in G_N \).

Considering this partial order, we say \( p_N(\cdot | \beta) \) is monotone if the conditional distribution of \( x(i, j) = 1 \) given \( x_{-ij} \) is a monotone increasing function. That is, \( p_N(\cdot | \beta) \) is monotone if, and only if \( x \preceq y \) implies

\[
p_{ij}(1|\theta, x_{-ij}) \leq p_{ij}(1|\theta, y_{-ij})
\]

(5.1.2)

Observe that

\[
\frac{p_{ij}(1|\theta, x_{-ij})}{p_N(x_{ij}^1|\theta)} \leq \frac{p_N(y_{ij}^1|\theta)}{p_N(x_{ij}^1|\theta) + p_N(y_{ij}^1|\theta)}
\]

\[
1 + \frac{p_N(x_{ij}^0|\theta)}{p_N(x_{ij}^1|\theta)} \leq \frac{1}{1 + \frac{p_N(y_{ij}^1|\theta)}{p_N(y_{ij}^1|\theta)}}
\]

\[
\frac{p_N(x_{ij}^0|\theta)}{p_N(x_{ij}^1|\theta)} \leq \frac{p_N(y_{ij}^0|\theta)}{p_N(y_{ij}^1|\theta)}
\]

Then, Inequality (5.1.2) is equivalent to

\[
\frac{p_N(x_{ij}^1|\theta)}{p_N(x_{ij}^0|\theta)} \leq \frac{p_N(y_{ij}^1|\theta)}{p_N(y_{ij}^0|\theta)}
\]

(5.1.3)

The main difficult to verify that the distribution of the ERG is monotone is the size of
5.1 THE ALGORITHM

\( G_N \), since the equation (5.1.3) should hold for all graphs \( x, y \in G_N \) that satisfy \( x \preceq y \). In such a way, the following proposition gives a sufficient condition on the parameter vector \( \theta \) under which the corresponding ERG distribution is monotone.

**Proposition 5.1.1.** Consider the ERG model with probability distribution given by (4.1.1). If \( \theta_i \geq 0 \), for \( i = 2, \ldots, s \), then the distribution \( p_N(\cdot; \theta) \) is monotone.

**Remark 5.1.2.** If \( \theta_1 \in \mathbb{R} \) and \( \theta_2, \theta_3 \geq 0 \), we have that the model edges-2-stars-triangles is monotone.

**Proof.** Let \( x, z \in G_N \) such that \( x \preceq z \). Without loss of generality, we consider the parametrization of the ERG given by (4.1.3).

For a pair of vertices \( (i, j) \), we define a new count that considers only the subgraphs of \( x \) that contain the vertices \( i \) and \( j \), that is

\[
N_{g_k}(x, (i, j)) = \sum_{v_{m_k} \in V_{m_k}, i, j \in v_{m_k}} 1\{v_{m_k} \preceq g_k\}
\]

Since \( x_{ij}^1 \) is different from \( x_{ij}^0 \) only in the edge \( (i, j) \), we have that \( x_{ij}^0 \preceq x_{ij}^1 \) implying that \( N_{g_k}(x_{ij}^0) \leq N_{g_k}(x_{ij}^1) \). In the same way, we have that \( N_{g_k}(z_{ij}^0) \leq N_{g_k}(z_{ij}^1) \). So,

\[
0 \leq N_{g_k}(x_{ij}^1) - N_{g_k}(x_{ij}^0) = N_{g_k}(x_{ij}^1, (i, j))
\]

\[
0 \leq N_{g_k}(z_{ij}^1) - N_{g_k}(z_{ij}^0) = N_{g_k}(z_{ij}^1, (i, j))
\]

Since \( x \preceq z \) we have that \( N_{g_k}(x_{ij}^1, (i, j)) \leq N_{g_k}(z_{ij}^1, (i, j)) \). For \( \beta_k \geq 0 \), \( k = 2, \ldots, s \), we have that

\[
\sum_{k=2}^{s} \frac{\beta_k}{N_{m_k} - 2} \left( N_{g_k}(x_{ij}^1) - N_{g_k}(x_{ij}^0) - N_{g_k}(z_{ij}^1) + N_{g_k}(z_{ij}^0) \right) =
\]

\[
\sum_{k=2}^{s} \frac{\beta_k}{N_{m_k} - 2} \left( N_{g_k}(x_{ij}^1, (i, j)) - N_{g_k}(z_{ij}^1, (i, j)) \right) \leq 0.
\]

Finally,

\[
\frac{p_N(y_{ij}^1 | \beta)}{p_N(y_{ij}^0 | \beta)} \frac{p_N(z_{ij}^0 | \beta)}{p_N(z_{ij}^1 | \beta)} = \exp \left\{ \sum_{k=2}^{s} \frac{\beta_k}{N_{m_k} - 2} \left( N_{g_k}(x_{ij}^1, (i, j)) - N_{g_k}(z_{ij}^1, (i, j)) \right) \right\}
\]

\[
\leq 1
\]

Since \( \theta_s = \frac{A(g_s) \beta_s}{N_{m_s} - 2} \), the result follows.

\( \square \)
To apply the monotone CFTP algorithm, we use the partial order defined by (5.1.1) and the extremal elements \( x^{(0)} \) and \( x^{(1)} \). The resulting procedure, which simulates the ERGM using the CFTP protocol, is described in Algorithm 1.

**Algorithm 1 CFTP for ERGM**

1: Input: \( E, \phi \)
2: Output: \( x \)
3: \( \text{upper}(i, j) \leftarrow 1 \), for all \((i, j) \in E\)
4: \( \text{lower}(i, j) \leftarrow 0 \), for all \((i, j) \in E\)
5: \( n \leftarrow 0 \)
6: while upper \( \neq \) lower do
7: \( n \leftarrow n + 1 \)
8: \( \text{upper}(i, j) \leftarrow 1 \), for all \((i, j) \in E\)
9: \( \text{lower}(i, j) \leftarrow 0 \), for all \((i, j) \in E\)
10: Choose a pair of vertices \( e_{-n} \) randomly on \( E \)
11: Simulate \( u_{-n} \) with distribution uniform on \([0, 1]\)
12: upper \( \leftarrow F_{0}^{-n}(\text{upper}, e_{-1}^{-n}, u_{-1}^{-n}) \)
13: lower \( \leftarrow F_{0}^{-n}(\text{lower}, e_{-1}^{-n}, u_{-1}^{-n}) \)
14: end while
15: \( x \leftarrow \text{upper} \)
16: Return: \( n, x \)

Let

\[
T_{N}^{\text{stop}} = \min\{n > 0 : F_{0}^{-n}(x^{(0)}, e_{-n}, u_{-n}) = F_{0}^{-n}(x^{(1)}, e_{-n}, u_{-n})\}
\]

be the stopping time of Algorithm 1. Proposition 5.1.3 below guarantees that the law of the output of Algorithm 1 is the target distribution \( p_{N}(\cdot; \theta) \). The proof of Proposition 5.1.3 follows directly from Theorem 2 in Propp e Wilson (1996) and is omitted here.

**Proposition 5.1.3.** Suppose that \( \mathbb{P}(T_{N}^{\text{stop}} < \infty) = 1 \). Then the graph \( x \) returned by Algorithm 1 has law \( p_{N}(\cdot; \theta) \).

### 5.2 Convergence Speed

By the construction of the perfect simulation algorithm it is expected that the stopping time \( T_{N}^{\text{stop}} \) of Algorithm 1 is related with the mixing time of the chain constructed using the Glauber dynamics. The proof of this relation is based in the general proof given by Propp e Wilson (1996) and for convenience of the reader we give the complete proof here.

Given vectors \( e_{n} = (e_{1}, e_{2}, \ldots, e_{n}) \) and \( u_{n} = (u_{1}, u_{2}, \ldots, u_{n}) \) define the forward random map \( F_{0}^{n}(x, e_{n}, u_{n}) \) as

\[
F_{0}^{1}(x, e_{1}, u_{1}) = \phi(x, e_{1}, u_{1}); \quad F_{0}^{n}(x, e_{n}, u_{n}) = F_{0}^{1}(F_{0}^{n-1}(x, e_{1}^{n-1}, u_{1}^{n-1}), e_{n}^{n-1}, u_{n}^{n-1}), \quad \text{for } n \geq 2.
\]
5.2 CONVERGENCE SPEED

For $x \in G_N$, let the Markov chain $\{Y_n^x\}_{n \in \mathbb{N}}$ be given by

\[
Y_0^x = x \\
Y_n^x = F_n^0(x, e_n^n, u_n^n), \quad n \geq 1
\]

and denote by $p_n^x$ the distribution of this Markov chain at time $n$. Now define

\[
\bar{d}(n) = \max_{x, x' \in G_N} ||p_n^x - p_n^{x'}||_{TV},
\]

where $\cdot||_{TV}$ stands for the total variation distance.

Let $T_N^{mix}$ be given by

\[
T_N^{mix} = \min \left\{ n > 0 : \bar{d}(n) \leq \frac{1}{e} \right\}.
\] (5.2.2)

The following theorem shows that the expected value of $T_N^{stop}$ is upper bounded by $T_N^{mix}$, up to a explicit multiplicative constant that depends on the number of vertices of the graph.

**Theorem 5.2.1.** Let $T_N^{mix}$ be the mixing time defined by (5.2.2) for the forward map (5.2.1). Then

\[
\mathbb{E}[T_N^{stop}] \leq 2(\log(N(N-1)/2) + 1) T_N^{mix}.
\]

**Proof.** Define the stopping time of the forward algorithm by

\[
\tilde{T}_N^{stop} = \min\{n > 0 : F_n^0(x^{(0)}, e_1^n, u_1^n) = F_n^0(x^{(1)}, e_1^n, u_1^n)\}
\]

To prove the theorem claim we use the random variable $\tilde{T}_N^{stop}$, since $\tilde{T}_N^{stop}$ and $T_N^{stop}$ have the same probability distribution. In fact,

\[
\mathbb{P}(T_N^{stop} > n) = \mathbb{P}(F_n^0(x^{(0)}, e_{-n}^{-1}, u_{-n}^{-1}) \neq F_n^0(x^{(1)}, e_{-n}^{-1}, u_{-n}^{-1}))
\]

\[
= \mathbb{P}(F_0^0(x^{(0)}, e_1^n, u_1^n) \neq F_0^0(x^{(1)}, e_1^n, u_1^n)) = \mathbb{P}(\tilde{T}_N^{stop} > n).
\]

Let $\{Y_n^1\}_{n \in \mathbb{N}}$ and $\{Y_n^0\}_{n \in \mathbb{N}}$ be the Markov chains obtained by (5.2.1) with initial states given by $x^{(0)}$ and $x^{(1)}$, respectively. Define $l(y)$ as the length of the longest increasing chain such that the top element is $y$. In the case $Y_n^0 = Y_n^1$ we have $l(Y_n^0) = l(Y_n^1)$. Otherwise, if $Y_n^0 \neq Y_n^1$ we have that $Y_n^0$ has at least one different edge of $Y_n^1$, since our algorithm use a local update. Then, $l(Y_n^0) + 1 \leq l(Y_n^1)$. So, we have that

\[
\mathbb{P}(\tilde{T}_N^{stop} > n) = \mathbb{P}(Y_n^0 \neq Y_n^1) = \mathbb{P}(l(Y_n^0) + 1 \leq l(Y_n^1))
\]

\[
\leq \mathbb{E}[|l(Y_n^1) - l(Y_n^0)|] = |\mathbb{E}_{p_0^1}[l(Y)] - \mathbb{E}_{p_0^0}[l(Y)]|
\]

\[
\leq ||p_1^n - p_0^n||_{TV} \max_{x \in G_N} l(x) \leq \bar{d}(n) \max_{x \in G_N} l(x)
\] (5.2.3)

Since the update function of the transitions of the Markov chain given by (4.2.2) updates
only one edge at each step, we have that \( \max_{x \in G_N} l(x) = l(x^{(1)}) \) and \( l(x^{(1)}) \) is the length of the longest increasing chain started on \( x^{(0)} \) with top element is \( x^{(1)} \), that is \( l(x^{(1)}) = \frac{N(N-1)}{2} \).

Thus

\[
\bar{d}(n) \geq \frac{2}{N(N-1)} \mathbb{P}(\tilde{T}_N^{\text{stop}} > n).
\]

Since \( \mathbb{P}(\tilde{T}_N^{\text{stop}} > n) \) is submultiplicative (Theorem 6 in Propp e Wilson (1996)) we have that

\[
\mathbb{E}[\tilde{T}_N^{\text{stop}}] \leq \sum_{i=1}^{\infty} n \mathbb{P}(\tilde{T}_N^{\text{stop}} > i n) \leq \sum_{i=1}^{\infty} n \mathbb{P}(\tilde{T}_N^{\text{stop}} > n)^i = \frac{n}{\mathbb{P}(\tilde{T}_N^{\text{stop}} \leq n)}.
\] (5.2.4)

Then, (5.2.3) and (5.2.4) imply that

\[
\mathbb{E}[\tilde{T}_N^{\text{stop}}] \leq \frac{n}{1 - \bar{d}(n)\frac{N(N-1)}{2}}.
\] (5.2.5)

Set \( n = \left( \log \left( \frac{N(N-1)}{2} \right) + 1 \right) T_N^{\text{mix}} \). Using the submultiplicative property of \( \bar{d}(n) \) (Lemma 4.12 in Levin et al. (2009)) we have that

\[
\bar{d}(n) \leq \left[ \bar{d}(T_N^{\text{mix}}) \right]^{\log \left( \frac{N(N-1)}{2} + 1 \right)} \leq \frac{2}{eN(N-1)}.
\] (5.2.6)

and by (5.2.5) and (5.2.6) we conclude that

\[
\mathbb{E}[\tilde{T}_N^{\text{stop}}] \leq \left( \log \left( \frac{N(N-1)}{2} \right) + 1 \right) T_N^{\text{mix}} \leq 2 \left( \log \left( \frac{N(N-1)}{2} \right) + 1 \right) T_N^{\text{mix}}.
\]

Theorem 5.2.1 guarantees that the CFTP algorithm is successful when it is applied for models for which the mixing time of the chain is small, that is, when the distribution of the chain constructed using the Glauber dynamics converges quickly to the stationary distribution. For this models, the CFTP algorithm can be implemented instead of using the MCMC method to get a exact sample from the ERG models.

The upper bound in Theorem 5.2.1 could be also combined with other results on the mixing time of the chain to obtain an estimate of the expected value of \( T_N^{\text{stop}} \), when they are available. As an example, we cite the results obtained by Bhamidi et al. (2011), where they present a study of the mixing time of the Markov chain constructed using the Glauber dynamics for ERG model under the conditions of Proposition 5.1.1 as we consider here.

They show that for models where \( \beta \) belongs to the high temperature regime (we will not get into details about this definition) the mixing time of the chain is \( \Theta(N^2 \log N) \). On the other hand, for models under the low temperature regime, the mixing time is exponentially slow.

The mixing time of models for which the parameter vector does not satisfy the conditions
of Proposition 5.1.1, that is, $\beta_i \leq 0$, for $i = 2, \cdots, s$, not in either phase are not studied in this work. Without loss of generality, the same result follows considering the model (4.1.1) with the appropriate relation between $\beta$ and $\theta$.

The mixing time of models with parameters not in either phase are not studied in this work, as well as, the models with parameters that do not satisfy the conditions of Proposition 5.1.1. Without loss of generality, the same result follows considering the model (4.1.1) with the appropriate relation between $\beta$ and $\theta$.

In order to construct an algorithm that can be applied for non monotone models a perfect simulation algorithm called Backward and Forward algorithm is described in the next chapter.

We come back to the CFTP algorithm in Chapter 7, where we present a few simulation results about both algorithms studied in this part of the thesis.
Chapter 6

Backward and Forward algorithm

In this chapter we present an alternative algorithm to simulate a graph from the ERG model for which it is not required monotonicity properties. This algorithm is based on the Kalikow-type decomposition of the transition probabilities of the Markov chain, constructed using the Glauber dynamics, having as stationary distribution the ERG distribution.

6.1 The algorithm

In this chapter we propose another perfect simulation algorithm based in Galves et al. (2010) and Galves et al. (2013). Galves et al. (2010) applied a perfect simulation algorithm for a Gibbs measure with infinite range interactions. In Galves et al. (2013) the authors present a perfect simulation algorithm to simulate from the stationary process of a multi-color system on $\mathbb{Z}^d$ with interactions of infinite range. In both works the approach is based on the Kalikow-type decomposition of the infinite range rates. This decomposition is a convex combination of rates with finite range and it allows us to deal with the infinite range dependence. This decomposition is also applied in the case of processes with long memory as described in the works of Comets et al. (2002) and Garivier (2015).

In spite of the fact that the vertex set of the graph considered in this thesis is finite, we propose a decomposition of the transition probabilities of the Markov chain, constructed using the Glauber dynamics, that considers the local dependence of an edge in the graph through a subset of pair of vertices.

6.2 Kalikow-type decomposition

In the model edges-2-stars-triangles the probability of the transition from a graph $x$ to $x^*_{ij}$, defined in (4.2.1), depends on the possible edges of $x$ for which the structure of triangles or 2-stars are obtained using the edge $(i, j)$. In this way, the presence or absence of the edge $(i, j)$ influences the number of triangles and 2-stars of the graph. The set of possible edges that can formed triangles and 2-stars using the edge $(i, j)$ is $E_{ij}^{N-2} = \bigcup_{v \in V \setminus \{i,j\}} \{(i,v), (j,v)\}$. 
Note that the use of the word “possible” edges here is just to emphasize that we need to look to these pairs of vertices to count triangles and 2-stars that depend on the pair \((i, j)\).

The idea is to restrict the set of edges \(E_{ij}^{N-2}\) that we need to look in the graph to compute the transition probability \(p_{ij}(1|\theta, x_{-ij})\). To this end, we rewrite the set of possible edges \(E_{ij}^{N-2}\), by defining the subsets \(E_{ij}^0, E_{ij}^1, \ldots, E_{ij}^{N-2}\) in the following way:

\[
E_{ij}^0 = \emptyset \\
E_{ij}^k = \bigcup_{v_l \in V \setminus \{i, j\}} \{(i, v_l), (j, v_l)\} \\
(6.2.1)
\]

By the construction described above we have that

A1. \(E_{ij}^k \subset E_{ij}^{k+1}\)

A2. \(E_{ij}^{N-2} = \bigcup_{k \in \{0, \ldots, N-2\}} E_{ij}^k\)

A3. \(|E_{ij}^k| = 2k\)

**Example 6.2.1.** Suppose that we have a graph with set of vertices \(V = \{1, 2, 3, 4, 5\}\). We fix the pair of vertices \((1, 3)\) and we construct the subsets defined in (6.2.1) in the following way:

\[
E_{13}^0 = \emptyset \\
E_{13}^1 = \{(1, 2), (3, 2)\} \\
E_{13}^2 = \{(1, 2), (3, 2), (1, 4), (3, 4)\} \\
E_{13}^3 = \{(1, 2), (3, 2), (1, 4), (3, 4), (5, 1), (5, 3)\}
\]

The sets defined above can be interpreted using the Figure (6.1). The set \(E_{13}^1\) is composed by the edges for which the structures of triangles and 2-stars are obtained considering the edge \((i, j)\) and the vertex 2. In the same way, the set \(E_{13}^2\) is obtained by considering the vertices 2 and 4. The last set \(E_{13}^3\) is obtained by considering all the remaining vertices \(V \setminus \{i, j\}\), that is, 2, 4 and 5.

The sets defined in (6.2.1) are interpreted in the following way: the first set \(E_{ij}^0\) means that we do not need to consider the other edges in the graph to add (or remove) the edge \((i, j)\). The set \(E_{ij}^k\) represents the set of pairs of vertices that we need to look in the graph to add (or remove) the edge \((i, j)\) whether I just want to look at \(2k\) pairs of vertices in the graph.

At this point, we need to define the probability of adding or removing the edge \((i, j)\) in the graph given the set \(E_{ij}^k\).
Define, for \( a \in \{0, 1\} \), \((i, j) \in E\) and \(1 \leq k \leq N - 2\)

\[
\begin{align*}
  r_{ij}^0(a|\theta) &= \min \{ p_{ij}(a|\theta, y_{-ij}) : y_{-ij} \in G_N \}, \\
  r_{ij}^k(a|\theta, x(E_{ij}^k)) &= \min \{ p_{ij}(a|\theta, y_{-ij}) : y(E_{ij}^k) = x(E_{ij}^k) \}.
\end{align*}
\]  

(6.2.2)

For \(0 \leq k \leq N - 2\), define

\[
\begin{align*}
  \alpha_{ij}^0(\theta) &= r_{ij}^0(0|\theta) + r_{ij}^0(1|\theta), \\
  \alpha_{ij}^k(\theta) &= \min_{x \in G_N} \{ r_{ij}^k(0|\theta, x(E_{ij}^k)) + r_{ij}^k(1|\theta, x(E_{ij}^k)) \}.
\end{align*}
\]  

(6.2.3)

and

\[
\begin{align*}
  \lambda_{ij}^0(\theta) &= \alpha_{ij}^0(\theta), \\
  \lambda_{ij}^k(\theta) &= \alpha_{ij}^k(\theta) - \alpha_{ij}^{k-1}(\theta).
\end{align*}
\]  

(6.2.4)

The quantities defined by (6.2.3) and (6.2.4) are all well defined in terms of theory. However, in the practical use, they can be difficult to calculate, due to the fact that its expressions involve a minimum over the set \( G_N \). It is also computationally demanding, since
the size of $\mathcal{G}_N$ is $2^{N(N-1)/2}$.

Thereby, the target point before proceeding with the Kalikow-type decomposition is to show that all the quantities involved in this construction can be easily calculated for some models of ERG, even in the case of large $N$.

**Proposition 6.2.2.** Consider the ERG model edges-2-stars-triangles with $\theta_1 \in \mathbb{R}$ and $\theta_2, \theta_3 \geq 0$. For $(i, j) \in \mathcal{E}$ we have that

$$r_{ij}^0(a|\theta) = r_{lm}^0(a|\theta), \ (l, m) \in \mathcal{E} \text{ and } a \in \{0, 1\}$$

and

$$r_{ij}^0(1|\theta) = p_{ij}(1|\theta, x^{(0)})$$
$$r_{ij}^0(0|\theta) = p_{ij}(0|\theta, x^{(1)})$$
$$r_{ij}^k(1|\theta, x(E_{ij}^k)) = p_{ij}(1|\theta, x^{(0)}(E_{ij}^k))$$
$$r_{ij}^k(0|\theta, x(E_{ij}^k)) = p_{ij}(0|\theta, x^{(1)}(E_{ij}^k))$$

where

$$x^{(0)}(E_{ij}^k)(l, m) = \begin{cases} 0, & (l, m) \in \mathcal{E} \setminus E_{ij}^k \\ x(l, m), & (l, m) \in E_{ij}^k \end{cases}$$
$$x^{(1)}(E_{ij}^k)(l, m) = \begin{cases} 1, & (l, m) \in \mathcal{E} \setminus E_{ij}^k \\ x(l, m), & (l, m) \in E_{ij}^k \end{cases}$$

**Proof.** Define for $(i, j) \in \mathcal{E}$ and $1 \leq l \leq N - 2$

$$a_0 = \arg \max_{y_{-ij} \in \mathcal{G}_N} \left\{ \exp \left( -\theta_1 - \theta_2 \sum_{k \neq i, j} [y(i, k) + y(j, k)] - \theta_3 \sum_{k \neq i, j} y(i, k)y(j, k) \right) \right\}$$
$$a_l(x(E_{ij}^l)) = \arg \max_{y_{-ij} \in \mathcal{G}_N: y(E_{ij}^l) = x(E_{ij}^l)} \left\{ \exp \left( -\theta_1 - \theta_2 \sum_{k \neq i, j} [y(i, k) + y(j, k)] - \theta_3 \sum_{k \neq i, j} y(i, k)y(j, k) \right) \right\}$$

and

$$b_0 = \arg \max_{y_{-ij} \in \mathcal{G}_N} \left\{ \exp \left( \theta_1 + \theta_2 \sum_{k \neq i, j} [y(i, k) + y(j, k)] + \theta_3 \sum_{k \neq i, j} y(i, k)y(j, k) \right) \right\}$$
$$b_l(x(E_{ij}^l)) = \arg \max_{y_{-ij} \in \mathcal{G}_N: y(E_{ij}^l) = x(E_{ij}^l)} \left\{ \exp \left( \theta_1 + \theta_2 \sum_{k \neq i, j} [y(i, k) + y(j, k)] + \theta_3 \sum_{k \neq i, j} y(i, k)y(j, k) \right) \right\}$$
For $\theta_1 \in \mathbb{R}$ and $\theta_2, \theta_3 \geq 0$ we have

\[
a_0 = x^{(0)} \\
b_0 = x^{(1)}
\]

and

\[
a_l(x(E^l_{ij})) = \arg \max_{y_{-ij} \in \Theta_N: y(E^l_{ij}) = x(E^l_{ij})} \left\{ \exp \left( -\theta_1 - \sum_{k \in E^l_{ij}} (\theta_2 [x(i, k) + x(j, k)] + \theta_3 x(i, k)x(j, k) ) \right. \\
- \sum_{k \in V \setminus E^l_{ij}} (\theta_2 [y(i, k) + y(j, k)] + \theta_3 y(i, k)y(j, k) ) \left. \right\} = x^{(0)}(E^l_{ij})
\]

\[
b_l(x(E^l_{ij})) = \arg \max_{y_{-ij} \in \Theta_N: y(E^l_{ij}) = x(E^l_{ij})} \left\{ \exp \left( \theta_1 + \sum_{k \in E^l_{ij}} (\theta_2 [x(i, k) + x(j, k)] + \theta_3 x(i, k)x(j, k) ) \right. \\
+ \sum_{k \in V \setminus E^l_{ij}} (\theta_2 [y(i, k) + y(j, k)] + \theta_3 y(i, k)y(j, k) ) \left. \right\} = x^{(1)}(E^l_{ij})
\]

By the definition (6.2.2) the result follows. \(\square\)

**Proposition 6.2.3.** Consider the ERG model edges-2-stars-triangles with $\theta_1 \in \mathbb{R}$ and $\theta_2, \theta_3 \leq 0$. For $(i, j) \in \mathcal{E}$ we have that

\[
r^0_{ij}(a|\theta) = r^0_{lm}(a|\theta), \ (l, m) \in \mathcal{E} \text{ and } a \in \{0, 1\}
\]

and

\[
\begin{align*}
r^0_{ij}(1|\theta) &= p_{ij}(1|\theta, x^{(1)}) \\
r^0_{ij}(0|\theta) &= p_{ij}(0|\theta, x^{(0)}) \\
r^k_{ij}(1|\theta, x(E^k_{ij})) &= p_{ij}(1|\theta, x^{(1)}(E^k_{ij})) \\
r^k_{ij}(0|\theta, x(E^k_{ij})) &= p_{ij}(0|\theta, x^{(0)}(E^k_{ij}))
\end{align*}
\]

where

\[
x^{(0)}(E^k_{ij})(l, m) = \begin{cases} 
0, & (l, m) \in \mathcal{E} \setminus E^k_{ij} \\
x(l, m), & (l, m) \in E^k_{ij}
\end{cases}
\]
\[ x^{(1)}(E^k_{ij})(l,m) = \begin{cases} 1, & (l,m) \in \mathcal{E} \setminus E^k_{ij} \\ x(l,m), & (l,m) \in E^k_{ij} \end{cases} \]

**Proof.** This proof follows the same arguments of the proof of Proposition 6.2.2. 

Now, we describe a recursive way to compute (6.2.3), considering the model edges-stars-triangles, for which it is not necessary to explore the set of all graphs \( \mathcal{G}_N \).

We start by finding the graph \( \mathbf{x} \) that minimizes

\[ r^1_{ij}(0|\theta, x(E^1_{ij})) + r^1_{ij}(1|\theta, x(E^1_{ij})) \]

The idea is calculate the minimum over a set of three graphs instead of \( 2^{N(N-1)/2} \) graphs. Recall that \( r^1_{ij}(0|\theta, x(E^1_{ij})) \) depends on the graph \( x \) through the edges in \( E^1_{ij} \). So, it is a easy task to compute \( y^1 = \arg\min_{x \in \mathcal{G}_N} (r^1_{ij}(0|\theta, x(E^1_{ij})) + r^1_{ij}(1|\theta, x(E^1_{ij}))) \) since we just need to calculate the \( \arg\min \) over the graphs \( \mathbf{x} \) with the possible edges on \( E^1_{ij} \) described in Table 6.1.

| \( x(i,v_1) \) | \( x(j,v_1) \) |
|---|---|
| 0 | 0 |
| 0 | 1 |
| 1 | 1 |

**Table 6.1:** Possible edges in the set \( E^1_{ij} = \{(i,v_1),(j,v_1)\} \).

It is important to make two remarks:

1. the remaining edges of the graph \( y^1 \) can assume any value in \{0,1\};
2. \( y^1 \) does not depends on the choice of the pair \((i,j)\), since we only consider undirected graphs.

For \( E^2_{ij} = \{(i,v_1),(j,v_1),(i,v_2),(j,v_2)\} \), the idea is to add (1) or remove (0) edges in the graph \( y^1 \) for the pair of vertices in \( E^2_{ij} \setminus E^1_{ij} \). Thus, to compute \( y^2 = \arg\min_{x \in \mathcal{G}_N} (r^2_{ij}(0|\theta, x(E^2_{ij}))+r^1_{ij}(1|\theta, x(E^1_{ij}))) \) we just need to compute the \( \arg\min \) over the graphs \( \mathbf{x} \) with \( x(E^1_{ij}) = y^1(E^1_{ij}) \) and the possible edges in \( E^2_{ij} \setminus E^1_{ij} \) given by Table (6.1) (replacing \( v_1 \) by \( v_2 \)).

We use the same framework described above to compute \( y^3, \cdots, y^{N-2} \). By (6.2.6) and the remark (2) above we have that

\[ \alpha^0(\theta) = r^0(0|\theta) + r^0(1|\theta) \]
\[ \alpha^k(\theta) = r^k_{ij}(0|\theta, y^k(E^k_{ij}))+r^1_{ij}(1|\theta, y^k(E^k_{ij})) \]

and

\[ \lambda^0(\theta) = \alpha^0(\theta) \]
\[ \lambda^k(\theta) = \alpha^k(\theta) - \alpha^{k-1}(\theta) \]
where

\[ r^0(0|\theta) \equiv r^0_{ij}(0|\theta), \text{ for some } (i, j) \in \mathcal{E} \]

The Kalikow-type decomposition of the ERG model edges-2-stars-triangles is given in Proposition (6.2.4).

**Proposition 6.2.4.** Consider a ERG model edges-2-stars-triangles. For any \((i, j) \in \mathcal{E}\) and any graph \(x \in \mathcal{G}_N\) the following decomposition holds

\[
p_{ij}(a|\theta, x_{-ij}) = \lambda^0(\theta)q^0(a|\theta) + \sum_{k=1}^{N-2} \lambda^k(\theta)q^k_{ij}(a|\theta, x(E^k_{ij})). \tag{6.2.5}
\]

**Remark 6.2.5.** The probabilities \(q^k_{ij}(|\theta, x(E^k_{ij}))\) are defined by (6.2.10) in the proof of Proposition 6.2.4.

The first component of the mixture described in (6.2.5) is the only term that does not depend on the structure of the graph and it can be interpreted as the mixture term associated with the distribution of the Erdős-Rényi model with parameter \(q^0(1|\theta)\).

**Proof of Proposition 6.2.4.** Notice that the transition probabilities defined in (4.2.1) depend on the graph \(x_{-ij}\), more precisely they depend on \(x_{-ij}(E^{N-2}_{ij})\). So, by definition (6.2.2)

\[
p_{ij}(a|\theta, x_{-ij}) = r^0_{ij}N^{-2}(a|\theta, x(E^{N-2}_{ij})).
\]

Besides, by (A1.) (in pg. 54) we have that

\[
r^{k-1}_{ij}(a|\theta, x(E^{k-1}_{ij})) \leq r^k_{ij}(a|\theta, x(E^k_{ij})). \tag{6.2.6}
\]

Now, define for \(a \in \{0, 1\}\) and \(1 \leq k \leq N - 2\)

\[
\Delta^0(a|\theta) = r^0(a|\theta)
\]

\[
\Delta^k_{ij}(a|\theta, x(E^k_{ij})) = r^k_{ij}(a|\theta, x(E^k_{ij})) - r^{k-1}_{ij}(a|\theta, x(E^{k-1}_{ij})).
\]

Observe that

\[
\lambda^0(\theta) = \Delta^0(0|\theta) + \Delta^0(1|\theta)
\]

and define for \(1 \leq k \leq N - 2\)

\[
\tilde{\lambda}^k_{ij}(\theta, x(E^k_{ij})) = \Delta^k_{ij}(0|\theta, x(E^k_{ij})) + \Delta^k_{ij}(1|\theta, x(E^k_{ij})).
\]
60 BACKWARD AND FORWARD ALGORITHM

By construction,

\[ \lambda^0_{ij}(\theta) + \sum_{k=1}^{N-2} \tilde{\lambda}_ij^k(\theta, x(E^k_{ij})) = \Delta^0_{ij}(0|\theta) + \Delta^0_{ij}(1|\theta) + \sum_{k=1}^{N-2} \sum_{a \in \{0,1\}} \Delta_{ij}^k(a|\theta, x(E^k_{ij})) \]

\[ = \sum_{a \in \{0,1\}} \sum_{k=1}^{N-2} (r^0_{ij}(a|\theta) + r^k_{ij}(a|\theta, x(E^k_{ij}))) - r^{k-1}_{ij}(a|\theta, x(E^{k-1}_{ij}))) \]

\[ = r^{N-2}_{ij}(0|\theta, x(E^{N-2}_{ij})) + r^{N-2}_{ij}(1|\theta, x(E^{N-2}_{ij})) \]

\[ = p_{ij}(0|\beta, x_{-ij}) + p_{ij}(1|\beta, x_{-ij}) = 1. \]

So, we define the probability distribution

\[ q^0(a|\theta) = \frac{\Delta^0(a|\theta)}{\lambda^0(\theta)} = \frac{r^0(a|\theta)}{\lambda^0(\theta)} \]

\[ \tilde{q}_ij^k(a|\theta, x(E^k_{ij})) = \frac{\Delta_{ij}^k(a|\theta, x(E^k_{ij}))}{\tilde{\lambda}_ij^k(\theta, x(E^k_{ij}))}. \]

Finally, we write

\[ p_{ij}(1|\theta, x_{-ij}) = \lambda^0(\theta)q^0(a|\theta) + \sum_{k=1}^{N-2} \tilde{\lambda}_ij^k(\theta, x(E^k_{ij}))\tilde{q}_ij^k(a|\theta, x(E^k_{ij})). \quad (6.2.7) \]

Note that in the above decomposition the values of \( \tilde{\lambda}_ij^k(\theta, x(E^k_{ij})) \) still depends on the graph \( x \) through \( E^k_{ij} \). In order to eliminate this dependence, we define the following quantities

\[ \beta^0(\theta) = \lambda^0(\theta) = \alpha^0(\theta) \]

\[ \beta_{ij}(\theta, x(E^k_{ij})) = \sum_{l \leq k} \tilde{\lambda}_ij^l(\theta, x(E^l_{ij})). \]

Observe that

\[ \beta_{ij}^k(\theta, x(E^k_{ij})) = \sum_{l \leq k} \tilde{\lambda}_ij^l(\theta, x(E^l_{ij})) \]

\[ = \sum_{l \leq k} \sum_{a \in \{0,1\}} \Delta_{ij}^l(a|\theta, x(E^l_{ij})) \]

\[ = \sum_{l \leq k} \sum_{a \in \{0,1\}} (r^l_{ij}(a|\theta, x(E^l_{ij}))) - r^{l-1}_{ij}(a|\theta, x(E^{l-1}_{ij}))) \quad (6.2.8) \]

\[ = \sum_{a \in \{0,1\}} r^k_{ij}(a|\theta, x(E^k_{ij})) \geq \alpha_{ij}^k(\theta). \]
By (6.2.6)

\begin{align*}
0 \leq & \alpha^0(\theta) \leq \alpha^1(\theta) \leq \cdots \leq \alpha^{N-2}(\theta) = 1 \\
0 \leq & \beta^0(\theta) \leq \beta^1(\theta, x(E_{ij}^1)) \leq \cdots \leq \beta^{N-2}(\theta, x(E_{ij}^{N-2})) = 1.
\end{align*}

(6.2.9)

In this way, for every \((i,j) \in E\) we construct the following partition of the interval \([0,1]\) that does not depend on \(x(B_{ij}^k)\)

| \(\lambda^0(\theta)\) | \(\lambda_1(\theta)\) | \(\lambda_2(\theta)\) | \(\cdots\) | \(\lambda_{N-2}(\theta)\) |
|------------------|------------------|------------------|----------|
| \(\alpha^0(\theta)\) | \(\alpha^1(\theta)\) | \(\alpha^2(\theta)\) | \(\cdots\) | \(\alpha^{N-3}(\theta)\) |

and we construct the partition of the interval \([0,1]\) that depends on \(x(B_{ij}^k)\), for every \((i,j) \in E\) and \(x \in G_N\)

| \(\lambda^0(\theta)\) | \(\lambda_1(\theta)\) | \(\lambda_2(\theta)\) | \(\cdots\) | \(\lambda_{N-2}(\theta)\) |
|------------------|------------------|------------------|----------|
| \(\beta^0(\theta)\) | \(\beta^1(\theta, x(E_{ij}^1))\) | \(\beta^2(\theta, x(E_{ij}^2))\) | \(\cdots\) | \(\beta^{N-3}(\theta, x(E_{ij}^{N-3}))\) |

To define the probability measure \(q_i^k(\cdot|\theta, x(E_{ij}^k))\) on \([\alpha_i^{-1}(\theta), \alpha_i^{k}(\theta)]\) we use the probabilities \(\tilde{q}_{ij}(\cdot|\theta, x(E_{ij}^k))\), \(l = 1, \ldots, N - 2\).

Since \(\tilde{q}_{ij}(\cdot|\theta, x(E_{ij}^k))\) is defined on the interval \([\beta_i^{k-1}(\theta, x(E_{ij}^k)), \beta_i^k(\theta, x(E_{ij}^k)))\), it is necessary to consider only \(\tilde{q}_{ij}(\cdot|\theta, x(E_{ij}^k))\), for \(l\) such that \(\alpha_i^{k}(\theta) \leq \beta_i^l(\theta, x(E_{ij}^l)) \leq \alpha_i^{k+1}(\theta)\).

Fix \(k \geq 0\) and define \(l_1^k(x) = \min\{m : \alpha_i^{k-1}(\theta) \leq \beta_i^m(\theta, x(E_{ij}^m))\}\) and \(l_2^k(x) = \max\{m : \alpha_i^{k}(\theta) \geq \beta_i^m(\theta, x(E_{ij}^m))\}\). Thus, by (6.2.9), we represent the partition of the interval \([\alpha_i^{k-1}(\theta), \alpha_i^{k}(\theta)]\) for the graph \(x\) as follows (for easy of notation we consider in this representation \(l_1^k(x) = l_1\) and \(l_2^k(x) = l_2\))

\[
\begin{array}{cccc}
\tilde{q}_{ij}^k(a|\theta, E_{ij}^k) & \tilde{q}_{ij}^{k+1}(a|\theta, E_{ij}^{k+1}) & \tilde{q}_{ij}^{k+1}(a|\theta, E_{ij}^{k+1}) & \tilde{q}_{ij}^{k+1}(a|\theta, E_{ij}^{k+1}) \\
\beta_{ij}^{k-1}(\theta, E_{ij}^k) & \alpha_i^{k-1}(\theta) & \beta_i^k(\theta, E_{ij}^k) & \beta_i^{k+1}(\theta, E_{ij}^k) \\
\end{array}
\]
Formally,

\[
[\alpha_{ij}^{k-1}(\theta), \alpha_{ij}^{k}(\theta)] = (\alpha_{ij}^{k-1}(\theta), \beta_{ij}^{k}(\theta), x(E_{ij}^{k})) \bigcup \left( \bigcup_{m=I_{ij}^{l}+1}^{I_{ij}^{l}} (\beta_{ij}^{m-1}(\theta, x(E_{ij}^{m-1})), \beta_{ij}^{m}(\theta, x(E_{ij}^{m}))) \right) \bigcup (\beta_{ij}^{k}(\theta, x(E_{ij}^{k})), \alpha_{ij}^{k}(\theta)) .
\]

In each sub-interval, we use the corresponding probability \( \tilde{q}_{ij}^{k}(a|\theta, x(E_{ij}^{k})) \) to define

\[
q_{ij}^{k}(a|\theta, x(E_{ij}^{k})) = \frac{\beta_{ij}^{k}(\theta, x(E_{ij}^{k})) - \alpha_{ij}^{k-1}(\theta)}{\lambda_{ij}^{k}(\theta)} q_{ij}^{k}(a|\theta, x(E_{ij}^{k})) + \sum_{m=I_{ij}^{l}+1}^{I_{ij}^{l}} \frac{\tilde{\lambda}_{ij}^{m}(\theta, x(E_{ij}^{m}))}{\lambda_{ij}^{k}(\theta)} \tilde{q}_{ij}^{m}(a|\theta, x(E_{ij}^{m-1})) + \frac{\alpha_{ij}^{k}(\theta) - \beta_{ij}^{k}(\theta, x(E_{ij}^{k})))}{\lambda_{ij}^{k}(\theta)} \tilde{q}_{ij}^{k+1}(a|\theta, x(E_{ij}^{k+1})).
\]

(6.2.10)

Notice that the probability defined above depends only on \( E_{ij}^{k} \), since \( I_{ij}^{l}(x) \leq k \) by (6.2.8) and for \( l \leq k \) we have that \( E_{ij}^{l} \subseteq E_{ij}^{k} \).

Since \( I_{ij}^{l}(x) - 1 = I_{ij}^{k-2}(x) \) and by (6.2.7), we finally get

\[
\lambda^{0}(\theta)q^{0}(a|\theta) + \sum_{k=1}^{N-2} \lambda_{ij}^{k}(\theta)q_{ij}^{k}(a|\theta, x(E_{ij}^{k})) = \lambda^{0}(\theta)q^{0}(a|\theta) + \sum_{k=1}^{N-2} \tilde{\lambda}_{ij}^{k}(\theta, x(E_{ij}^{k}))\tilde{q}_{ij}^{k}(a|\theta, x(E_{ij}^{k}))
\]

\[
= p_{ij}(a|\theta, x_{-ij}).
\]

\[\square\]

### 6.3 Backward and Forward algorithms

In this section we consider the time-reversible Markov chain \((X_{n})_{n \geq 0}\) with respect to the probability distribution \( p_{N}(\cdot|\theta) \) for which the construction is described in Section 4.2. As a consequence of Proposition 6.2.4 we can rewrite the transition probabilities (6.3.1) by

\[
\tilde{P}(x, y) = \begin{cases} 
\frac{1}{|E|} \left( \lambda^{0}(\theta)q^{0}(a|\theta) + \sum_{k=1}^{N-2} \lambda_{ij}^{k}(\theta)q_{ij}^{k}(a|\theta, x(E_{ij}^{k})) \right), & \text{if } y = x_{ij}, a \in \{0, 1\} \\
0, & \text{otherwise.}
\end{cases}
\]

(6.3.1)

In a nutshell, a transition from a graph \( x_{n} \) can be done in the following manner:
1. **(Select a pair of vertices to update)** Select a pair of vertices \((i, j)\) uniformly on \(\mathcal{E}\);

2. **(Select a mixture component)** Select \(k \in \{0, 1, \ldots, N - 2\}\) according to the probability distribution \(P(K = k) = \lambda^k(\theta)\);

3. **(Add or remove an edge)** The existence of an edge between \(i\) and \(j\) depends on the subgraph \(x_n(E^k_{ij})\). The probability of \(x_{n+1}(i, j) = 1\) given \(x_n(E^k_{ij})\) is \(q^k_{ij}(1|\theta, x_n(E^k_{ij}))\).

Step (2) gives us the information about the set of pairs of vertices of \(x_n\) (a subgraph of \(x\)) for which the existence of an edge between \(i\) and \(j\) depends on. Observe that if, at the time \(n + 1\), we chose \(k = 0\), then the existence of an edge between \(i\) and \(j\) does not depend on the graph \(x_n\) and it also does not depend on the predecessor graphs \(x_{n-1}, x_{n-2}, \ldots\).

Using this dynamics based on the decomposition given in Proposition 6.2.4 we construct a perfect simulation algorithm composed by two procedures. To well explain both procedures, we suppose that we want to simulate the existence of an edge between a fixed pair of vertices \(i\) and \(j\).

In the first procedure, we determine the set \(C^{ij}_n\) of pair of vertices of the graph that influence the existence (or absence) of an edge between \(i\) and \(j\), at time \(n\).

In order to do this, we set \(C^{ij}_n = \{(i, j)\}\) and we select \(k\) according to step (2). If \(k = 0\), we do not need to consider the structure of the graph at time \(n - 1\), then we stop the procedure. If \(k \neq 0\), the existence of the edge \((i, j)\) depends on the pairs of vertices \(E^k_{ij}\) in the graph at time \(n - 1\), then we need to add \(E^k_{ij}\) in the set \(C^{ij}_n\). So, we have \(C^{ij}_{n-1} = C^{ij}_n \cup E^k_{ij}\). Now, we choose randomly a pair \((l, m)\) in \(C^{ij}_{n-1}\) to determine the existence of an edge with the information about the structure of the graph at time \(n - 2\). Thus, we follow the same sequence of steps as before. We select \(k\) according to step (2). If \(k = 0\), we do not need to consider the structure of the graph at time \(n - 2\), then we remove \((l, m)\) from \(C^{ij}_{n-1}\). If \(k \neq 0\), the existence of the edge \((l, m)\) depends on the pairs of vertices \(E^k_{lm}\) in the graph at time \(n - 2\), then we need to add \(E^k_{lm}\) in the set \(C^{ij}_{n-1}\). The procedure stops when the set \(C^{ij}_m\), \(m < n\), is empty, that is, we do not need to use more information about the predecessor graphs.

The second procedure of the algorithm consists in adding, or removing, edges in the graph using the probabilities \(\{q^k_{ij}: k = 1, \ldots, N - 2, (i, j) \in \mathcal{E}\}\), from the past to the future. This algorithm starts by adding, or removing edges, between the pairs of vertices that are chosen in the end of the first procedure. Without loss of generality we set \(n = 0\). To sample the graph we use both procedures with \(C^{ij}_0 = \mathcal{E}\).

These two procedures are formally described in Algorithms (2) and (3).
Algorithm 2 Backward sketch procedure

1: Input: $\mathcal{E}$
2: Output: $N_{\text{stop}}, E$
3: $N \leftarrow 0$, $N_{\text{stop}} \leftarrow 0$, $E = \emptyset$, $C \leftarrow \mathcal{E}$
4: while $C \neq \emptyset$ do
5:  \hspace{1em} $N \leftarrow N + 1$
6:  \hspace{1em} Choose a pair of vertices $(I, J) \in C$ uniformly on $C$
7:  \hspace{1em} Choose $K \in \{0, 1, \ldots, N - 2\}$ randomly according to the distribution $P(K = k) = \lambda^k(\theta)$
8:  \hspace{1em} if $K = 0$ then
9:  \hspace{2em} $C \leftarrow C \setminus \{(I, J)\}$
10:  \hspace{1em} else
11:  \hspace{2em} $C \leftarrow C \cup E^K_{IJ}$
12:  \hspace{1em} end if
13:  \hspace{1em} $E(N) \leftarrow (I, J, K)$
14: end while
15: $N_{\text{stop}} \leftarrow N$
16: Return $N_{\text{stop}}, E$

Algorithm 3 Forward edges assignment procedure

1: Input: $N_{\text{stop}}, E$
2: Output: $\{x(i, j) : (i, j) \in \mathcal{E}\}$
3: $N \leftarrow N_{\text{stop}}$
4: $x(i, j) \leftarrow \Delta$, for all $i, j \in V$
5: while $N \geq 1$ do
6:  \hspace{1em} $(I, J, K) \leftarrow E(N)$
7:  \hspace{1em} if $K = 0$ then
8:  \hspace{2em} Choose $A \in \{0, 1\}$ according to the probability distribution $P(A = a) = q^0(a|\theta)$
9:  \hspace{1em} else
10:  \hspace{2em} Choose $A \in \{0, 1\}$ according to the probability distribution $P(A = a) = q^K_{fj}(a|\theta, x(E^K_{fj}))$
11:  \hspace{2em} end if
12:  \hspace{1em} $x(I, J) \leftarrow A$
13:  \hspace{1em} $N \leftarrow N - 1$
14: end while
15: Return $\{x(i, j) : (i, j) \in \mathcal{E}\}$
6.4 Construction of the process

In this section we describe formally the stochastic process behind the backward sketch process procedure described in Algorithm (2). Define $C_{ij}^0 = \{(i, j)\}$ and define $C_{ij}^n$ as the set of pairs of vertices at time $-n$ for which the structure of the graph at time 0 depends on. This process has as state space the set $\mathcal{P}(\mathcal{E})$, that is, the set of subsets of $\mathcal{E}$.

The main idea of this construction is to select, at each time $n$, a pair of vertices $(l, m)$ from $C_{ij}^n$ and to change the set $C_{ij}^n$ according to a transformation $\pi^{lm,k}(C_{ij}^n)$.

For $B \in \mathcal{P}(\mathcal{E})$, we define

$$\pi^{lm,k}(B) = \begin{cases} B \setminus \{(l, m)\}, & \text{if } k = 0 \\ B \cup E_{lm}^k, & \text{if } k \neq 0 \end{cases}$$

In a nutshell, this transformation change the set $B$ in the following way: if $k = 0$, then the edge $(l, m)$ is removed from $B$; otherwise the set of pair of vertices $E_{lm}^k$ is added to the set $B$.

We construct the Markov chain, at time $n$, in the following manner

1. Choose $(l, m)$ uniformly on $C_{ij}^{n-1}$;
2. Choose $k$ according to the probability distribution $P(K = k) = \lambda^k(\theta)$;
3. Let $C_{ij}^n = \pi^{lm,k}(C_{ij}^{n-1})$.

By the construction above, $(C_{ij}^n)_{n \geq 0}$ is a Markov chain with initial state $C_{ij}^0 = \{(i, j)\}$ and transitions given by

$$P(A, B) = \begin{cases} \frac{\lambda_k}{|A|}, & B = \pi^{lm,k}(A), (l, m) \in A, k \in \{0, \cdots, N - 2\} \\ 0, & \text{otherwise} \end{cases} \quad (6.4.1)$$

For the set $\mathcal{E}$ of all possible edges of the graph define the process

$$\mathcal{C}_n^\mathcal{E} = \bigcup_{(i, j) \in \mathcal{E}} C_{ij}^n \quad (6.4.2)$$

The steps 6 – 11 of the Algorithm (2) are justified by the transition probability (6.4.1) of the chain.

Define the stopping times

$$N_{\text{stop}}^{ij} = \inf \{n : C_{ij}^n = \emptyset\}$$

$$N_{\text{stop}} = \inf \{n : \mathcal{C}_n^\mathcal{E} = \emptyset\}$$
Notice that $N_{stop}$ is exactly the number of steps of Algorithm (2). In this way, the number of steps $N_{stop}$ must be finite for the perfect simulation algorithm to be successful. For this reason, we need to study the conditions to guarantee that $N_{stop} < \infty$.

The following theorem assures that the law of output of Algorithm (3) is the invariant measure of the process $(x_n)_{n \geq 0}$ and, as a consequence of the construction of the process, the invariant measure is the ERG distribution. This theorem is proved in Galves et al. (2010) and Galves et al. (2013) and for convenience of the reader we give the complete proof here.

**Theorem 6.4.1.** Assume that $P(N_{stop} < \infty) = 1$. Thus, the law of the output $\{x(i, j) : (i, j) \in \mathcal{E}\}$ printed at the end of Algorithms (2) and (3) is the invariant measure of the process $(x_n)_{n \geq 0}$.

**Proof.** Let $\mu$ be the law of the output $\{x(i, j) : (i, j) \in \mathcal{E}\}$ of Algorithms 2 and 3. To prove that $\mu$ is the invariant measure of the process $(x_t)_{t \geq 0}$ we need to show that the law of the process $(x^y_t)_{t \geq 0}$, starting at time 0 with the graph $y$, converges to $\mu$, as $t \to \infty$.

In order to construct the process $(x^y_t)_{t \geq 0}$ we need to construct a modification of Algorithms 2 and 3.

Let $T_{stop}$ be a random variable taking values in $\{0, 1, \cdots\}$. We constructed a modified version of Algorithm 2 in the following way:

1. Input: $\mathcal{E}$
2. Output: $T_{stop}$, $E$, $C$
3. $B = \emptyset$, $C \leftarrow \mathcal{E}$, $T_{stop} \leftarrow 0$
4. While $T_{stop} < t$ and $C \neq \emptyset$ do
   
   $T_{stop} \leftarrow T_{stop} + 1$

5. Repeat steps 6–13 of Algorithm 2.
16. Return $T_{stop}$, $B$ and $C$.

In the modified version of Algorithm 3 we replace steps 1, 3, 4 and 15 by:

1. Input: $T_{stop}$, $B$, $C$ and $y$
3. $N_{stop} \leftarrow T_{stop}$
4. $x^y(i, j) \leftarrow y(i, j)$, for all $(i, j) \in C$ and $x^y(i, j) \leftarrow \Delta$, for all $(i, j) \in \mathcal{E} \setminus C$
15. Return $\{x^y(i, j) : (i, j) \in \mathcal{E}\}$

The law of the output of the unmodified Algorithm 3 is the law of $(x^y_t)_{t \geq 0}$.

The output $C$ of the modified version of Algorithm 2 is the set $C^E_t$ of pair of vertices of the graph at time $-t$ which the structure of the graph at time 0 depend. Notice that in this version of the Algorithm if $C = \emptyset$ then $T_{stop} < t$ and $T_{stop} = N_{stop}$, where $N_{stop}$ is
the stopping time of Algorithm 2. In this case, the output \( x^y \) of the modified Algorithm 3 is equal to the output \( x \) of the unmodified Algorithm 3.

To prove that the law of \( x^y \) converges to \( \mu \), the law of the output \( \{x(i,j): (i,j) \in E\} \), we use the following result: we say that a sequence of random variables \( Z_n \) converges in law to \( Z \) if \( \mathbb{E}[f(Z_n)] \to \mathbb{E}[f(Z)] \), as \( n \to \infty \), for all continuous and bounded function \( f \).

So, let \( f: G_N \to \mathbb{R} \) be a bounded continuous function. We have that

\[
\mathbb{E}[f(x^y), T_{\text{stop}} \geq t] = \mathbb{E}[f(x^y), T_{\text{stop}} < t] + \mathbb{E}[f(x^y), T_{\text{stop}} \geq t] = \mathbb{E}[f(x), T_{\text{stop}} < t] + \mathbb{E}[f(x^y), T_{\text{stop}} \geq t] \]

If \( T_{\text{stop}} \geq t \) we have that \( C \neq \emptyset \) and \( \{T_{\text{stop}} \geq t\} \subseteq \{N_{\text{stop}} \geq t\} \). Since \( f \) is bounded we have that

\[
\mathbb{E}[f(x^y), T_{\text{stop}} \geq t] \leq \|f\|_{\infty} P(T_{\text{stop}} \geq t) \leq \|f\|_{\infty} P(N_{\text{stop}} \geq t) \]

By assumption \( P(N_{\text{stop}} < \infty) = 1 \). So, for \( t \to \infty \)

\[
\mathbb{E}[f(x^y), T_{\text{stop}} \geq t] \leq \|f\|_{\infty} P(N_{\text{stop}} \geq t) \to 0
\]

We conclude that

\[
\mathbb{E}[f(x^y)] \to \mathbb{E}[f(x)]
\]

Then, the law of the process \( (x^y_t)_{t \geq 0} \), starting at time 0 with the graph \( y \) converges to the law \( \mu \) of the output \( x \) of Algorithm 3, as \( t \to \infty \). So, the invariant measure of the process is equal to the law of the output \( x \) of Algorithm 3. \( \square \)

The following theorem assures, under a certain condition, that the Algorithm (2) stops almost surely. The proof of this result follows the steps of the proof present in Galves et al. (2013).

**Theorem 6.4.2.** Assume that

\[
\delta = \sum_{k=1}^{N-2} \lambda^k(\theta)(2k + 1) < 1. \quad (6.4.3)
\]

We have that Algorithm (2) stops after a finite number of steps almost surely and

\[
\mathbb{E}[N_{\text{stop}}] \leq \frac{N(N-1)}{2(1-\delta)}.
\]
Proof. Let \( L_{ij}^n = |C_{ij}^n| \) be the number of possible edges in the set \( C_{ij}^n \) after \( n \) steps of the algorithm, for \( (i, j) \in \mathcal{E} \).

The main idea of this proof is to construct a multi-type branching process \( Z_{ij}^n \).

To describe the offspring distribution, define \((\tilde{\xi}_{lm}^n)_{n \geq 1, (l,m) \in \mathcal{E}}\) the sequence of independent random variables, taking values in \( \{0, |E_{lm}^1| + 1, \ldots, |E_{lm}^{N-2}| + 1\} \). The individuals of the branching process are represented by the possible edges of the graph and the individual \((l, m)\) has type \((l, m)\). We start with one individual with type \((i, j)\), so \( Z_{ij}^{0} = 1 \).

We construct the offspring of \((l, m)\) “adding” new possible edges, that is \( \tilde{\xi}_{lm}^n \in \{|E_{lm}^1| + 1, \ldots, |E_{lm}^{N-2}| + 1\} \). The children of \((l, m)\) have type \( E_{lm}^k \cup (l, m) \). In this way, we define the offspring distribution by

\[
P(\tilde{\xi}_{lm}^n = |E_{lm}^k| + 1) = \lambda^k(\theta) \\
P(\tilde{\xi}_{lm}^n = 0) = \lambda^0(\theta)
\]

Thus, we define the multi-type branching process by

\[
Z_{ij}^{0} = 1 \\
Z_{ij}^n = \sum_{n=1}^{Z_{ij}^{n-1}} \sum_{(l,m) \in \mathcal{E}} \tilde{\xi}_{lm}^n \mathbf{1}\{\text{type} = (l, m)\}
\]

and, by condition (6.4.3)

\[
E[\tilde{\xi}_1^{lm}] = \sum_{k=1}^{N-2} (|E_{lm}^k| + 1) \lambda^k_{lm}(\theta) < 1
\]

Define the stopping time \( V_{ij} = \inf\{n : Z_{ij}^n = 0\} \). We have that \( P(V_{ij} < \infty) = 1 \), since \( E[\tilde{\xi}_1^{lm}] < 1 \), for all \((l, m) \in \mathcal{E}\).

Now, we describe the multi-type branching process \( Z_{ij}^n \) through a random walk. The offspring of each individual in the generation \( n \) is obtained one by one, in such a way that if \( Z_{ij}^n = k \) it takes \( k \) steps to obtain \( Z_{ij}^{n+1} \). To define the random walk associated with \( Z_{ij}^n \), we define

\[
P(\xi_{lm}^n = |E_{lm}^k|) = \lambda^k(\theta) \\
P(\xi_{lm}^n = -1) = \lambda^0(\theta)
\]

For any \((i, j) \in \mathcal{E}\), let \( I_{ij}^n \) be the edge that has been chosen at the \( n \)-th jump of the
process \((C_{ij}^{n})_{n \geq 1}\). The random variable \(I_{n}^{ij}\) is conditionally independent given \(C_{n-1}^{ij}\) and

\[
P(I_{n}^{ij} = (l, m) | C_{n-1}^{ij}) = \frac{1}{|C_{n-1}^{ij}|}
\]

Now, define the random walk such that \(S_{0}^{ij} = 0\) and

\[
S_{n}^{ij} = \sum_{k=1}^{n} \xi_{k}^{ij}
\]

We have that

\[
E[\xi_{1}^{lm}] = -\lambda^{1}(\theta) + \sum_{k=1}^{N-2} |E_{lm}^{k}| \lambda^{k}(\theta) = -1 + \sum_{k=1}^{N-2} \lambda^{k}(\theta)(|E_{lm}^{k}| + 1) = \delta - 1 < 0
\]

By construction \(S_{n}^{ij} + n(1-\delta)\) is a super-martingale with respect to \(\mathcal{F}_{n}^{ij} = \{C_{1}^{ij}, C_{2}^{ij}, \ldots, C_{n}^{ij}\}\), since

\[
E[S_{n}^{ij} + n(1-\delta) | \mathcal{F}_{n-1}^{ij}] = E[S_{n-1}^{ij} + n(1-\delta) + \xi^{E_{n}^{ij}}_{n} | \mathcal{F}_{n-1}^{ij}] = S_{n-1}^{ij} + n(1-\delta) + E[\xi^{E_{n}^{ij}}_{n} | \mathcal{F}_{n-1}^{ij}]
\]

\[
= S_{n-1}^{ij} + n(1-\delta) + E\left\{ E[\xi^{E_{n}^{ij}}_{n} | I_{n}^{ij}] | \mathcal{F}_{n-1}^{ij}\right\}
\]

\[
= S_{n-1}^{ij} + n(1-\delta) - \sum_{(k,l) \in C_{n-1}^{ij}} \frac{(1-\delta)}{|C_{n-1}^{ij}|} = S_{n-1}^{ij} - (n-1)(1-\delta)
\]

By defining the stopping time \(V_{n}^{ij} = \min\{k : S_{k}^{ij} = 0\}\), we have

\[
I_{n}^{ij} \leq S_{n}^{ij} \quad \text{as long as} \quad n \leq V_{n}^{ij}
\]

Recall the following result of the martingale theory: if \(N\) is a stopping time and \(X_{n}\) is a super martingale, then \(X_{N \wedge n}\) is a super martingale. Furthermore, \(\mathbb{E}[X_{N \wedge n}] \leq \mathbb{E}[X_{0}]\) (Theorem 5.2.6 in Durrett (2010)).

In this way, we have that

\[
\mathbb{E}[S_{n \wedge V_{n}^{ij}}^{ij} + (1-\delta)(n \wedge V_{n}^{ij})] \leq \mathbb{E}[S_{0}^{ij}] = 0 \quad (6.4.4)
\]
and

$$
\mathbb{E}[S_{n\land V_{ij}^{\text{stop}}}^{ij}] = -P(V_{ij}^{\text{stop}} < n) + \mathbb{E}[S_n^{ij}; n < V_{ij}^{\text{stop}}]
$$

Since $S_n^{ij} \geq 0$ on $\{n < V_{ij}^{\text{stop}}\}$, we have

$$
\mathbb{E}[S_{n\land V_{ij}^{\text{stop}}}^{ij}] \geq -P(V_{ij}^{\text{stop}} < n)
$$

Combining (6.4.4) and (6.4.5) we get

$$
\mathbb{E}[n \land V_{ij}^{\text{stop}}] \leq \frac{1}{(1 - \delta)} P(V_{ij}^{\text{stop}} < n)
$$

We have that $S_{T_n}^{ij} \leq Z_n^{ij}$, for $T_0 = 0$ and $T_{n+1} = T_n + S_{T_n}^{ij}$.

Since, $S_{T_{V_{ij}}}^{ij} \leq Z_{V_{ij}}^{ij} = 0$ we have that $V_{ij}^{\text{stop}} \leq T_{V_{ij}}$ and $P(V_{ij}^{\text{stop}} < \infty) = 1$. Letting $n \rightarrow \infty$

$$
\mathbb{E}[V_{ij}^{\text{stop}}] \leq \frac{1}{(1 - \delta)}
$$

By construction $N_{\text{stop}}^{ij} \leq V_{ij}^{\text{stop}}$ and by (6.4.2) we conclude that

$$
\mathbb{E}[N_{\text{stop}}] \leq \sum_{(i,j) \in \mathcal{E}} \mathbb{E}[N_{\text{stop}}^{ij}] \leq \sum_{(i,j) \in \mathcal{E}} \mathbb{E}[V_{ij}^{\text{stop}}] \leq \frac{|\mathcal{E}|}{(1 - \delta)}
$$

The study of the Backward and Forward algorithm are detailed in Chapter 7.
Chapter 7

Simulations

In this section we present the simulation results of the two algorithms described in this thesis. Both algorithms were implemented using the software $R$ (R Core Team, 2015).

7.1 The algorithm

We start the simulations comparing the MCMC algorithm with the CFTP algorithm. As we know, the output of the MCMC method depends on both the initial graph used to start the algorithm and the number of steps used to run the chain. This can be observed in practice in Figure 7.1: the convergence time for both MCMC and CFTP are of the same order of magnitude, as suggested by Theorem 6.4.2. In this figure we compare the number of edges and 2-stars of the graphs obtained by MCMC at different sample sizes of the forward Markov chain and by CFTP at its convergence time. In this case, we consider the model edges and 2-stars with parameters $(\theta_1, \theta_2) = (-2.2, 0.03)$ for a graph with 80 vertices and the mean proportion of edges in the graphs is approximately $p = 0.86$. We observe in Figure 8.1a that MCMC remains a long time strongly dependent on the initial value of the chain when this value is chosen in a region of low probability. On the other hand, when the initial value is chosen appropriately, then MCMC and CFTP have comparable performances, as shown in Figure 8.1b. In this example, we use a sample with 500 graphs for each method.
(a) Initial state for MCMC chosen as Erdős-Rényi graph with parameter 0.1.

(b) Initial state for MCMC chosen as Erdős-Rényi graph with parameter 0.8.

**Figure 7.1:** Boxplots of number of edges (left) and number of 2-stars (right) for the ERGM with $N = 80$ and parameter $(\theta_1, \theta_2) = (-2.2, 0.03)$. The MCMC algorithm was run for $n = 80,000$ time steps (a) (respectively $n = 100,000$ (b) and $n = 130,000$ (c)) and two different initial states. The results for the CFTP algorithm at the time of convergence are shown in (d), with mean 130.500 time steps.
In order to study the convergence of the Backward and Forward algorithm, the reader should observe in Figure 7.2 the relation between the size of the set \( C_n, n = 1, \ldots, N_{\text{stop}} \), with the parameter \( \delta \), defined in Theorem 6.4.2, for the model edges-2-stars-triangles with 10 vertices and different values of the parameters. In Figure 7.2a the size of set \( C_n \) converges quickly to 0 for the model with parameter \( \theta = (-1.8, 0.02, 0.03) \). Observe that \( \delta = 0.86 \) so this example satisfies the condition (6.4.3) of Theorem 6.4.2.

In the second model with \( \theta = (-1.8, 0.05, 0.04) \), Figure 7.2b, the condition (6.4.3) is not verified (\( \delta = 2.1 \)); however, the algorithm BF stops after 25723 steps. This example reinforces the fact that condition (6.4.3) is not a necessary condition for the algorithm stops.

Finally, the model with \( \theta = (-1.8, 0.07, 0.05) \) does not satisfy the condition (6.4.3) (\( \delta = 3.15 \)) and, as we can observe in Figure 7.2c, the values of \( |C_n| \) oscillate between 15 and 45, making the algorithm unfinished (the algorithm ran until step 1000000).

\[ \theta = (-1.8, 0.02, 0.03). \]
\[ \theta = (-1.8, 0.05, 0.04). \]
\[ \theta = (-1.8, 0.07, 0.05). \]

**Figure 7.2:** The values of \( |C_n| \) for the model edges-2-stars-triangles with 10 vertices and parameter \( \theta \).
Despite the application of the Backward-Forward method for non monotone ERG models for which the sufficient statistics are edges, 2-stars and triangles, we have that the condition (6.4.3) of Theorem 6.4.2 restricts the models that can be simulated with mean number of steps of order $N^2$, as $N$ grows. The reader can observe in Figure 7.3 that the number of models that satisfies condition (6.4.3) decreases when the size of the graph grows.

![Graphs](image)

**Figure 7.3:** The values of $\delta$ for the model edges-2-stars-triangles with parameter $\theta = (\theta_1, 0, \theta_3)$.

However, it is important to emphasize that condition (6.4.3) is a sufficient condition for the Backward-Forward algorithm to stop. In Table 7.1 only the first three models satisfy this condition and the algorithm still works for the other models. In the case the condition is satisfied the mean number of steps of the Backward-Forward algorithm is smaller than the mean number of steps of the CFTP algorithm using a sample of size 10000.

**Table 7.1:** Comparison between the CFTP and BF algorithms for the model edges-2-stars-triangles with $N = 10$ and parameters $(\theta_1, \theta_2, \theta_3) = (-1.8, 0.02, \theta_3)$

| $\theta_3$ | $(\text{BF})$ $\mathbb{E}[N_{\text{stop}}]$ | $(\text{CFTP})$ $\mathbb{E}[T_{\text{stop}}^N]$ | $\delta$ | $N(N-1)$ | mean number of steps | mean number of edges | mean number of 2-stars | mean number of triangles |
|------------|---------------------------------|---------------------------------|---------|----------|---------------------|----------------------|------------------------|--------------------------|
| 0.01       | 87.83                           | 204.54                          | 0.58    | 106.91   | 6.69                | 8.04                 | 0.41                   |                          |
| 0.02       | 107.51                          | 207.73                          | 0.72    | 159.51   | 6.65                | 8.08                 | 0.44                   |                          |
| 0.03       | 142.07                          | 208.99                          | 0.86    | 332.04   | 6.64                | 8.00                 | 0.41                   |                          |
| 0.04       | 196.38                          | 208.19                          | 1.02    | -        | 6.68                | 8.10                 | 0.42                   |                          |
| 0.05       | 298.61                          | 209.07                          | 1.18    | -        | 6.65                | 7.98                 | 0.42                   |                          |
| 0.06       | 507.07                          | 208.71                          | 1.35    | -        | 6.76                | 8.34                 | 0.45                   |                          |
| 0.07       | 827.97                          | 213.10                          | 1.53    | -        | 6.81                | 8.47                 | 0.47                   |                          |
Chapter 8

Future directions

In this chapter we present the final considerations about the results obtained in this thesis as well as the future potential research problems related with the problem of estimating the number of communities in the SBM and simulating graphs from the ERG model.

8.1 Final Considerations and related problems

8.1.1 Stochastic Block Model

In the first part of this thesis we introduced the Krichevsky-Trofimov estimator for the number of communities in the Stochastic Block Model and prove its eventual almost sure convergence to the underlying number of communities, without assuming a known upper bound on that quantity. The family of penalty functions of the form (3.1.2) improves the penalty function used in the penalized likelihood estimator (2.3.1) proposed by Wang et al. (2017) in the sense that the penalty function used in this thesis is of smaller order compared with the penalty used in the previous work.

The penalty function used in the corrected Bayesian information criterion proposed by Hu et al. (2016) also improves that used in the penalized likelihood estimator of Wang et al. (2017). In the case that \( k \) is of smaller order than \( \log n \) the penalty function of the KT order estimator improves that used in the corrected Bayesian information criterion; however, in the case that \( k \) is close to \( n \) the penalty function used in the KT estimator is of order \( n^2 \log n \) while the penalty used in the corrected Bayesian information estimator is of order \( n^2 \log n \). Having said that, a possible future research concerning the KT estimator might be an improvement in the penalty function (3.1.2). However, it is important to emphasize that even though the penalty function is not of smaller order compared with the corrected Bayesian information criterion, the KT estimator is the only strongly consistent estimator known at the moment.

Another future line of questioning can be the problem of constructing a strong consistent penalized estimator for the number of communities of the SBM based on the profile maximum likelihood instead of the KT estimator.
In this thesis we considered a SBM for which the matrix $P$ does not depend on $n$, that is, the model has the same expected degree for all nodes in the graph. An interesting extension of the estimator proposed in this thesis is to consider that $P_n = \rho_n \tilde{P}$, where $\tilde{P}$ is a matrix with constant non-negative entries and $\rho_n$ is the degree density. The expected degree of any node of the graph is given by $n \rho_n \sum_{a,b=1}^{k_0} \tilde{P}_{a,b} \pi_a \pi_b$. Wang et al. (2017) and Hu et al. (2016) also proved the weak consistency for the communities’ estimator in the regime where the network average degree grows at a polylog rate, that means, $\rho_n$ goes to zero at a rate $\frac{n \rho_n \log n}{\log n}$.

Another future direction concerning the estimation of the number of communities is to consider the degree corrected stochastic block model (DCSBM) introduced in Karrer e Newman (2011). In this model, the constraint of having the same expected degree for all nodes within a community is removed by defining $\mathbb{P}(X(i,j) = 1|Z_i = 1, Z_j = b) = \theta_i \theta_j P_{a,b}$.

### 8.1.2 Exponential Random Graphs

In the second part of this thesis we proposed a perfect simulation algorithm for the Exponential Random Graph Model, based on the Coupling From The Past method of Propp e Wilson (1996) and on the Backward and Forward algorithm as used in Galves et al. (2010) and Galves et al. (2013).

The justification for the use of the CFTP algorithm is generally restricted to the monotonicity property of the dynamics, given by Proposition 5.2.1, and to prove that $T_{\text{stop}}^N$ is almost-surely finite, but with no clue about the expected time to convergence. Here, in contrast, we prove a much stronger result: not only do we upper-bound the expectation of the waiting time, but we show that CFTP compares favorably with standard MCMC forward simulation: the waiting time is (up to a logarithmic factor) not larger than the time one has to wait before a run of the chain approaches the stationary distribution. Moreover, CFTP has the advantage of providing a draw from the exact stationary distribution, in contrast to MCMC which may keep for a long time a strong dependence on the initial state of the chain, as shown in the simulation example of Figure 7.1. We thus argue that for these models the CFTP algorithm is a better tool than forward MCMC simulation in order to get a sample from the Exponential Random Graph Model.

In order to overcome the monotonicity constraint of the model one can apply the Backward and Forward algorithm; however, as the reader can observe in the simulations presented in Chapter 7, the efficiency of the algorithm seems to decay when the number of vertices of the graph grows. This relation can be explained by the structure of the model: when one vertex is added in the graph, all the other vertices can be linked with it. This happens because the model takes into account the triangles and stars structures of the graph.

For the efficiency of the algorithm it is necessary to have high mixture probability $\lambda_k(\vec{\theta})$ in the Kalikow-type decomposition (6.2.5) for small values of $k$. However, because of the dependence between the edges described above, the mixture probabilities $\lambda_k(\vec{\theta})$ do not decrease when the size of the set of possible edges $E_{ij}^k$ increases, that means, when $k$ increases. In this
way, when $N$ grows to infinity, the algorithm works only for models with $\lambda_0(\theta)$ tending to 1, that means, only for the Erdős-Rényi model (recall that the mixture component related with the mixture distribution $\lambda_0(\theta)$ does not take into account the other edges of the graph to decide if an edge is added in the graph).

Having said that, one possible way to construct a Kalikow-type decomposition in such a way that the mixture probabilities decrease when the size of the set of possible edges grows is to consider a model for which the vertices are located in a metric space and the existence of an edge between a pair of vertices depends on the geometric distance between these vertices.

The graph for which the vertices are located in a space equipped with a metric is called spatial network. A simplest example of spatial network is the geometric graph where the vertices are randomly located in the plane and there exists an edge between a pair of vertices if the distance between the vertices is smaller than $r$. So an interesting extension of the simulation algorithm proposed for the ERG models is to consider a spatial network model where the probability of observing a graph is based on the graph’s statistics as considered in the ERG model; however, in this case, we want to penalize connections between nodes that are very distant.

From the theoretical point of view, the study of such a model seems promising since Bhamidi et al. (2011) have proved that for ERG models under the high temperature regime any finite collection of edges in the graph is asymptotically independent when the number of vertices grows to infinity. As a complement of this work, Chatterjee et al. (2013) shows that for $n$ large the ERG model is essentially the same as the Erdős-Rényi model. In a nutshell, this result shows that the ERG model may not be a good choice to model real networks with large number of vertices when we expect to have strong dependencies in the network.
FUTURE DIRECTIONS
Appendix A

Proofs

Lemma A.0.1. For \( n = n_1 + \cdots + n_J \), we have that

\[
\prod_{j=1}^{J} \left( \frac{n_j}{n} \right)^{n_j} \leq \prod_{j=1}^{J} \Gamma \left( n_j + \frac{1}{2} \right) \Gamma \left( \frac{1}{2} \right)^{J-1}.
\]

Proof. We write as in Davisson et al. (1981) [eq. 56]

\[
\frac{\prod_{j=1}^{J} \Gamma \left( n_j + \frac{1}{2} \right)}{\Gamma \left( n + \frac{1}{2} \right) \Gamma \left( \frac{1}{2} \right)^{J-1}} = \frac{\prod_{j=1}^{J} (n_j + 1)(n_j + 2) \cdots (2n_j)}{(n + 1)(n + 2) \cdots (2n)}.
\]

Define, for \( x \geq 0 \)

\[
g_k(x) = \prod_{i=1}^{k} \left( x + \frac{i}{k} \right).
\]

It was shown in the Appendix of Davisson et al. (1981) that

\[
g_n(x) \leq \prod_{j=1}^{J} g_{n_j}(x).
\]

Using this result for \( x = 1 \) we have that

\[
\prod_{i=1}^{n} \left( n + i \right) \leq \prod_{j=1}^{n} (n_j + 1)(n_j + 2) \cdots (2n_j) .
\]

Rearranged, this is

\[
\prod_{j=1}^{J} \left( \frac{n_j}{n} \right)^{n_j} \leq \frac{\prod_{j=1}^{n} (n_j + 1)(n_j + 2) \cdots (2n_j)}{(n + 1)(n + 2) \cdots (2n)}.
\]
Lemma A.0.2. For \( n \geq 4 \) and for all \( k \) smaller than \( n \)

\[
\log \left( \frac{\Gamma \left( \frac{1}{2} \right) \Gamma \left( n + \frac{k}{2} \right)}{\Gamma \left( \frac{k}{2} \right) \Gamma \left( n + \frac{1}{2} \right)} \right) \leq \left( \frac{k-1}{2} \right) \log n + \frac{k(k-1)}{4n} + \frac{1}{12n} + \log \frac{\Gamma \left( \frac{1}{2} \right)}{\Gamma \left( \frac{k}{2} \right)}.\]

Proof. This proof is given in Appendix I of Gassiat e Boucheron (2003). For \( x > 0 \) we have the following inequality

\[
x^{-\frac{1}{2}}e^{-x\sqrt{2\pi}} \leq \Gamma(x) \leq x^{-\frac{1}{2}}e^{-x\sqrt{2\pi}}e^{\frac{1}{12x}}
\]

So

\[
\log \left( \frac{\Gamma \left( \frac{1}{2} \right) \Gamma \left( n + \frac{k}{2} \right)}{\Gamma \left( \frac{k}{2} \right) \Gamma \left( n + \frac{1}{2} \right)} \right) \leq \left( n + \frac{k-1}{2} \right) \log \left( n + \frac{k}{2} \right) - \left( n + \frac{k}{2} \right) + \frac{1}{12} \left( n + \frac{k}{2} \right) \\
- n \log \left( n + \frac{1}{2} \right) + \left( n + \frac{1}{2} \right) + \log \frac{\Gamma \left( \frac{1}{2} \right)}{\Gamma \left( \frac{k}{2} \right)} \\
\leq \left( n + \frac{k-1}{2} \right) \log \left( n \left( 1 + \frac{k}{2n} \right) \right) + \frac{1}{12n} - n \log \left( n \left( 1 + \frac{1}{2n} \right) \right) \\
- \frac{k-1}{2} + \log \frac{\Gamma \left( \frac{1}{2} \right)}{\Gamma \left( \frac{k}{2} \right)} \\
\leq \left( \frac{k-1}{2} \right) \log n + \left( n + \frac{k-1}{2} \right) \log \left( 1 + \frac{k}{2n} \right) - n \log \left( 1 + \frac{1}{2n} \right) \\
+ \frac{1}{12n} - \frac{k-1}{2} + \log \frac{\Gamma \left( \frac{1}{2} \right)}{\Gamma \left( \frac{k}{2} \right)}.
\]

Using \( 1 - \frac{1}{x} \leq \log x \leq x - 1, \ x > 0, \)

\[
\log \left( \frac{\Gamma \left( \frac{1}{2} \right) \Gamma \left( n + \frac{k}{2} \right)}{\Gamma \left( \frac{k}{2} \right) \Gamma \left( n + \frac{1}{2} \right)} \right) \leq \left( \frac{k-1}{2} \right) \log n + \left( n + \frac{k-1}{2} \right) \left( \frac{k}{2n} \right) + \frac{1}{12n} - \frac{k-1}{2} + \log \frac{\Gamma \left( \frac{1}{2} \right)}{\Gamma \left( \frac{k}{2} \right)}
\]

\[
\leq \left( \frac{k-1}{2} \right) \log n + \frac{k(k-1)}{4n} + \frac{1}{12n} + \log \frac{\Gamma \left( \frac{1}{2} \right)}{\Gamma \left( \frac{k}{2} \right)}.
\]

\( \square \)
Bibliography

Amini et al. (2013) Arash A Amini, Aiyou Chen, Peter J Bickel, Elizaveta Levina et al. Pseudo-likelihood methods for community detection in large sparse networks. *The Annals of Statistics*, 41(4):2097–2122. Citado na pág. 1, 14, 15

Besag (1974) Julian Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, páginas 192–236. Citado na pág. 2

Bhamidi et al. (2011) Shankar Bhamidi, Guy Bresler e Allan Sly. Mixing time of exponential random graphs. *The Annals of Applied Probability*, 21(6):2146–2170. Citado na pág. 40, 50, 77

Bickel et al. (2013) Peter Bickel, David Choi, Xiangyu Chang e Hai Zhang. Asymptotic normality of maximum likelihood and its variational approximation for stochastic block-models. *The Annals of Statistics*, páginas 1922–1943. Citado na pág. 1, 14

Bickel e Chen (2009) Peter J Bickel e Aiyou Chen. A nonparametric view of network models and newman–girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50):21068–21073. Citado na pág. 1, 14, 15

Butts (2015) Carter T Butts. A novel simulation method for binary discrete exponential families, with application to social networks. *The Journal of mathematical sociology*, 39 (3):174–202. Citado na pág. 3

Chatterjee et al. (2013) Sourav Chatterjee, Persi Diaconis et al. Estimating and understanding exponential random graph models. *The Annals of Statistics*, 41(5):2428–2461. Citado na pág. 40, 77

Comets et al. (2002) Francis Comets, Roberto Fernández e Pablo A Ferrari. Processes with long memory: regenerative construction and perfect simulation. *Annals of Applied Probability*, páginas 921–943. Citado na pág. 4, 53

Csiszár et al. (2000) Imre Csiszár, Paul C Shields et al. The consistency of the bic markov order estimator. *The Annals of Statistics*, 28(6):1601–1619. Citado na pág. 24

Csiszár et al. (2004) Imre Csiszár, Paul C Shields et al. Information theory and statistics: A tutorial. *Foundations and Trends® in Communications and Information Theory*, 1(4):417–528. Citado na pág. 16

Daudin et al. (2008) J-J Daudin, Franck Picard e Stéphane Robin. A mixture model for random graphs. *Statistics and computing*, 18(2):173–183. Citado na pág. 1, 14
Davisson et al. (1981) L Davisson, R McEliece, M Pursley e Mark Wallace. Efficient universal noiseless source codes. *IEEE Transactions on Information Theory*, 27(3):269–279. Citado na pág. 79

Durrett (2010) Rick Durrett. Probability: Theory and examples, 2010. Citado na pág. 69

Erdös e Rényi (1959) Paul Erdös e Alfréd Rényi. On random graphs, i. *Publicationes Mathematicae (Debrecen)*, 6:290–297. Citado na pág. 10

Galves et al. (2010) Antonio Galves, Eva Löcherbach e Enza Orlandi. Perfect simulation of infinite range gibbs measures and coupling with their finite range approximations. *Journal of Statistical Physics*, 138(1-3):476–495. Citado na pág. 4, 53, 66, 76

Galves et al. (2013) Antonio Galves, NL Garcia, E Löcherbach, Enza Orlandi et al. Kalikow-type decomposition for multicolor infinite range particle systems. *The Annals of Applied Probability*, 23(4):1629–1659. Citado na pág. 4, 53, 66, 67, 76

Garivier (2015) Aurélien Garivier. Perfect simulation of processes with long memory: A coupling into and from the past algorithm. *Random Structures & Algorithms*, 46(2):300–319. Citado na pág. 4, 53

Gassiat e Boucheron (2003) Elisabeth Gassiat e Stéphane Boucheron. Optimal error exponents in hidden markov models order estimation. *IEEE Transactions on Information Theory*, 49(4):964–980. Citado na pág. 24, 80

Geman e Geman (1984) Stuart Geman e Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6):721–741. Citado na pág. 2

Geyer e Thompson (1992) Charles J Geyer e Elizabeth A Thompson. Constrained monte carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society. Series B (Methodological)*, páginas 657–699. Citado na pág. 2

Hastings (1970) W Keith Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109. Citado na pág. 2

Holland et al. (1983) Paul W Holland, Kathryn Blackmond Laskey e Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137. Citado na pág. 1

Hu et al. (2016) Jianwei Hu, Hong Qin, Ting Yan e Yunpeng Zhao. On consistency of model selection for stochastic block models. *arXiv preprint arXiv:1611.01238*. Citado na pág. 2, 15, 75, 76

Karrer e Newman (2011) Brian Karrer e Mark EJ Newman. Stochastic blockmodels and community structure in networks. *Physical review E*, 83(1):016107. Citado na pág. 76

Krichevsky e Trofimov (1981) Raphail E. Krichevsky e Victor K. Trofimov. The performance of universal encoding. *IEEE Trans. Information Theory*, 27(2):199–206. Citado na pág. 3, 15

Latouche et al. (2012) Pierre Latouche, Etienne Birmele e Christophe Ambroise. Variational bayesian inference and complexity control for stochastic block models. *Statistical Modelling*, 12(1):93–115. Citado na pág. 1, 14
Levin et al. (2009) David Asher Levin, Yuval Peres e Elizabeth Lee Wilmer. *Markov chains and mixing times*. American Mathematical Soc. Citado na pág. 50

Liu e Narayan (1994) Chuang-Chun Liu e Prakash Narayan. Order estimation and sequential universal data compression of a hidden markov source by the method of mixtures. *IEEE Transactions on Information Theory*, 40(4):1167–1180. Citado na pág. 16

Metropolis et al. (1953) Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller e Edward Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092. Citado na pág. 2

Propp e Wilson (1996) James Gary Propp e David Bruce Wilson. Exact sampling with coupled markov chains and applications to statistical mechanics. *Random structures and Algorithms*, 9(1-2):223–252. Citado na pág. 3, 4, 45, 48, 50, 76

R Core Team (2015) R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015. URL http://www.R-project.org/. Citado na pág. 71

Rohe et al. (2011) Karl Rohe, Sourav Chatterjee, Bin Yu et al. Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, 39(4):1878–1915. Citado na pág. 1, 14

Snijders (2002) Tom AB Snijders. Markov chain monte carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2):1–40. Citado na pág. 2

Su et al. (2017) Liangjun Su, Wuyi Wang e Yichong Zhang. Strong consistency of spectral clustering for stochastic block models. *arXiv preprint arXiv:1710.06191*. Citado na pág. 1, 14

van der Pas et al. (2017) SL van der Pas, AW van der Vaart et al. Bayesian community detection. *Bayesian Analysis*. Citado na pág. 1

Wang et al. (2017) YX Rachel Wang, Peter J Bickel et al. Likelihood-based model selection for stochastic block models. *The Annals of Statistics*, 45(2):500–528. Citado na pág. 1, 2, 14, 15, 28, 33, 75, 76