

# **Community Detection and Stochastic Block Models**

**Other titles in Foundations and Trends® in Communications and Information Theory**

*Redundancy of Lossless Data Compression for Known Sources by Analytic Methods*

Michael Drmota and Wojciech Szpankowski

ISBN: 978-1-68083-284-6

*Multiterminal Secrecy by Public Discussion*

Prakash Narayan and Himanshu Tyagi

ISBN: 978-1-68083-186-3

*Channel Coding Methods for Non-Volatile Memories*

Lara Dolecek and Frederic Sala

ISBN: 978-1-68083-042-2

*Multi-way Communications: An Information Theoretic Perspective*

Anas Chaaban and Aydin Sezgin

978-1-60198-788-4

*An Approximation Approach to Network Information Theory*

A. Salman Avestimehr, Suhas N. Diggavi, Chao Tian and

David N. C. Tse

978-1-68083-026-2

# Community Detection and Stochastic Block Models

---

**Emmanuel Abbe**  
Princeton University  
[eabbe@princeton.edu](mailto:eabbe@princeton.edu)

**now**

the essence of knowledge

Boston — Delft

# Foundations and Trends<sup>®</sup> in Communications and Information Theory

*Published, sold and distributed by:*

now Publishers Inc.  
PO Box 1024  
Hanover, MA 02339  
United States  
Tel. +1-781-985-4510  
[www.nowpublishers.com](http://www.nowpublishers.com)  
[sales@nowpublishers.com](mailto:sales@nowpublishers.com)

*Outside North America:*

now Publishers Inc.  
PO Box 179  
2600 AD Delft  
The Netherlands  
Tel. +31-6-51115274

The preferred citation for this publication is

E. Abbe. *Community Detection and Stochastic Block Models*. Foundations and Trends<sup>®</sup> in Communications and Information Theory, vol. 14, no. 1-2, pp. 1–162, 2018.

ISBN: 978-1-68083-477-2

© 2018 E. Abbe

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, mechanical, photocopying, recording or otherwise, without prior written permission of the publishers.

Photocopying. In the USA: This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by now Publishers Inc for users registered with the Copyright Clearance Center (CCC). The 'services' for users can be found on the internet at: [www.copyright.com](http://www.copyright.com)

For those organizations that have been granted a photocopy license, a separate system of payment has been arranged. Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. In the rest of the world: Permission to photocopy must be obtained from the copyright owner. Please apply to now Publishers Inc., PO Box 1024, Hanover, MA 02339, USA; Tel. +1 781 871 0245; [www.nowpublishers.com](http://www.nowpublishers.com); [sales@nowpublishers.com](mailto:sales@nowpublishers.com)

now Publishers Inc. has an exclusive license to publish this material worldwide. Permission to use this content must be obtained from the copyright license holder. Please apply to now Publishers, PO Box 179, 2600 AD Delft, The Netherlands, [www.nowpublishers.com](http://www.nowpublishers.com); e-mail: [sales@nowpublishers.com](mailto:sales@nowpublishers.com)

# Foundations and Trends<sup>®</sup> in Communications and Information Theory

Volume 14, Issue 1-2, 2018

## Editorial Board

### Editor-in-Chief

**Sergio Verdú**

Princeton University  
United States

### Editors

Venkat Anantharam  
*UC Berkeley*

Helmut Bölcske  
*ETH Zurich*

Giuseppe Caire  
*TU Berlin*

Daniel Costello  
*University of Notre Dame*

Anthony Ephremides  
*University of Maryland*

Andrea Goldsmith  
*Stanford University*

Albert Guillen i Fabregas  
*Pompeu Fabra University*

Dongning Guo  
*Northwestern University*

Dave Forney  
*MIT*

Te Sun Han  
*University of Tokyo*

Babak Hassibi  
*Caltech*

Michael Honig  
*Northwestern University*

Tara Javidi  
*UC San Diego*

Ioannis Kontoyiannis  
*Cambridge University*

Gerhard Kramer  
*TU Munich*

Amos Lapidoth  
*ETH Zurich*

Muriel Medard  
*MIT*

Neri Merhav  
*Technion*

David Neuhoff  
*University of Michigan*

Alon Orlitsky  
*UC San Diego*

Yury Polyanskiy  
*MIT*

Vincent Poor  
*Princeton University*

Maxim Raginsky  
*UIUC*

Kannan Ramchandran  
*UC Berkeley*

Igal Sason  
*Technion*

Shlomo Shamai  
*Technion*

Amin Shokrollahi  
*EPF Lausanne*

Yossef Steinberg  
*Technion*

Wojciech Szpankowski  
*Purdue University*

David Tse  
*Stanford University*

Antonia Tulino  
*Bell Labs*

Rüdiger Urbanke  
*EPF Lausanne*

Emanuele Viterbo  
*Monash University*

Tsachy Weissman  
*Stanford University*

Frans Willems  
*TU Eindhoven*

Raymond Yeung  
*CUHK*

Bin Yu  
*UC Berkeley*

## Editorial Scope

### Topics

Foundations and Trends<sup>®</sup> in Communications and Information Theory publishes survey and tutorial articles in the following topics:

- Coded modulation
- Coding theory and practice
- Communication complexity
- Communication system design
- Cryptology and data security
- Data compression
- Data networks
- Demodulation and Equalization
- Denoising
- Detection and estimation
- Information theory and statistics
- Information theory and computer science
- Joint source/channel coding
- Modulation and signal design
- Multiuser detection
- Multiuser information theory
- Optical communication channels
- Pattern recognition and learning
- Quantization
- Quantum information processing
- Rate-distortion theory
- Shannon theory
- Signal processing for communications
- Source coding
- Storage and recording codes
- Speech and Image Compression
- Wireless Communications

### Information for Librarians

Foundations and Trends<sup>®</sup> in Communications and Information Theory, 2018, Volume 14, 4 issues. ISSN paper version 1567-2190. ISSN online version 1567-2328 . Also available as a combined paper and online subscription.

# Contents

---

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Community detection, clustering and block models . . . . .	3
1.2	Fundamental limits: information and computation . . . . .	7
1.3	An example on real data . . . . .	8
1.4	Historical overview of the recent developments . . . . .	10
1.5	Outline . . . . .	13
1.6	Notations . . . . .	13
<b>2</b>	<b>The stochastic block model</b>	<b>15</b>
2.1	The general SBM . . . . .	15
2.2	The symmetric SBM . . . . .	17
2.3	Recovery requirements . . . . .	17
2.4	SBM regimes and topology . . . . .	22
2.5	Learning the model . . . . .	24
<b>3</b>	<b>Tackling the stochastic block model</b>	<b>28</b>
3.1	The block MAP estimator . . . . .	29
3.2	Computing block MAP: spectral and SDP relaxations . . . . .	31
3.3	The bit MAP estimator . . . . .	35
3.4	Computing bit MAP: belief propagation . . . . .	36

<b>4</b>	<b>Exact recovery for two communities</b>	<b>39</b>
4.1	Warm up: genie-aided hypothesis test . . . . .	41
4.2	The information-theoretic threshold . . . . .	42
4.3	Achieving the threshold . . . . .	50
<b>5</b>	<b>Weak recovery for two communities</b>	<b>64</b>
5.1	Warm up: broadcasting on trees . . . . .	66
5.2	The information-theoretic threshold . . . . .	71
5.3	Achieving the threshold . . . . .	80
<b>6</b>	<b>Partial recovery for two communities</b>	<b>99</b>
6.1	Almost Exact Recovery . . . . .	99
6.2	Partial recovery at finite SNR . . . . .	103
6.3	Mutual Information-SNR tradeoff . . . . .	105
6.4	Proof technique and connections to spiked Wigner models	107
6.5	Partial recovery for constant degrees . . . . .	109
<b>7</b>	<b>The general SBM</b>	<b>112</b>
7.1	Exact recovery and CH-divergence . . . . .	112
7.2	Weak recovery and generalized KS threshold . . . . .	122
7.3	Partial recovery . . . . .	124
<b>8</b>	<b>The information-computation gap</b>	<b>125</b>
8.1	Crossing KS: typicality . . . . .	126
8.2	Nature of the gap . . . . .	130
8.3	Proof technique for crossing KS . . . . .	131
<b>9</b>	<b>Other block models</b>	<b>135</b>
<b>10</b>	<b>Concluding remarks and open problems</b>	<b>140</b>
	<b>Acknowledgements</b>	<b>144</b>
	<b>References</b>	<b>145</b>



# Community Detection and Stochastic Block Models

Emmanuel Abbe<sup>1</sup>

<sup>1</sup>*Program in Applied and Computational Mathematics, and Department of Electrical Engineering, Princeton University, Princeton, USA.*

*Email: [eabbe@princeton.edu](mailto:eabbe@princeton.edu), URL: [www.princeton.edu/~eabbe](http://www.princeton.edu/~eabbe).*

*This research was partly supported by the Bell Labs Prize, the NSF CAREER Award CCF-1552131, the NSF Center for the Science of Information CCF-0939370, and the Google Faculty Research Award.*

---

## ABSTRACT

The stochastic block model (SBM) is a random graph model with different group of vertices connecting differently. It is widely employed as a canonical model to study clustering and community detection, and provides a fertile ground to study the information-theoretic and computational tradeoffs that arise in combinatorial statistics and more generally data science.

This monograph surveys the recent developments that establish the fundamental limits for community detection in the SBM, both with respect to information-theoretic and computational tradeoffs, and for various recovery requirements such as exact, partial and weak recovery. The main results discussed are the phase transitions for exact recovery at the Chernoff-Hellinger threshold, the phase transition for weak recovery at the Kesten-Stigum threshold, the optimal SNR-mutual information tradeoff for partial recovery, and the gap between information-theoretic and computational thresholds.

The monograph gives a principled derivation of the main algorithms developed in the quest of achieving the limits, in particular two-round algorithms via graph-splitting, semi-definite programming, (linearized) belief propagation, classical/nonbacktracking spectral methods and graph powering. Extensions to other block models, such as geometric block models, and a few open problems are also discussed.

---

# 1

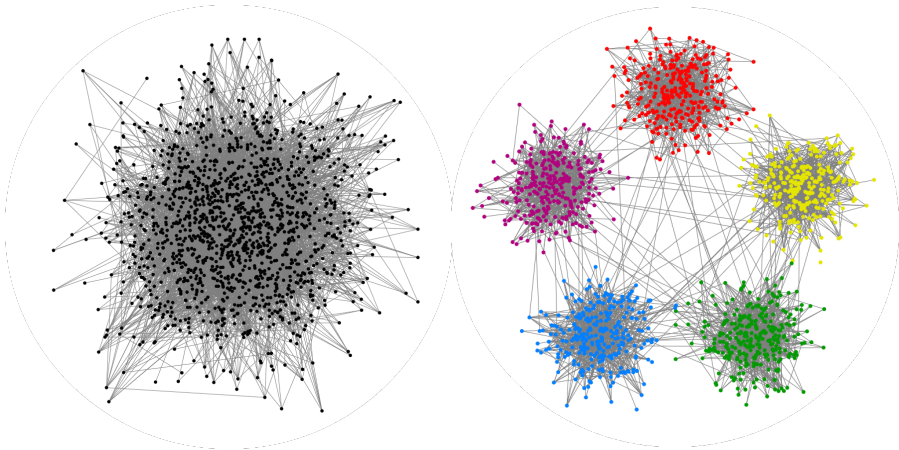
---

## Introduction

---

### 1.1 Community detection, clustering and block models

The most basic task of *community detection*, or *graph clustering*, consists in partitioning the vertices of a graph into clusters that are more densely connected. From a more general point of view, community structures may also refer to groups of vertices that connect similarly to the rest of the graph without having necessarily a higher inner density, such as disassortative communities that have higher external connectivity. Note that the terminology of ‘community’ is sometimes used only for assortative clusters in the literature, but we adopt here the more general definition. Community detection may also be performed on graphs where edges have labels or intensities, and if these labels represent similarities among data points, the problem may be called *data clustering*. In this monograph, we will use the terms communities and clusters exchangeably. Further, one may also have access to interactions that go beyond pairs of vertices, such as in hypergraphs, and communities may not always be well separated due to overlaps. In the most general context, community detection refers to the problem of inferring similarity classes of vertices in a network by having access to measurements of local interactions.



**Figure 1.1:** The above two graphs are the same graph re-organized and drawn from the SBM model with 1000 vertices, 5 balanced communities, within-cluster probability of  $1/50$  and across-cluster probability of  $1/1000$ . The goal of community detection in this case is to obtain the right graph (with five communities) from the left graph (scrambled) up to some level of accuracy. In such a context, community detection may be called graph clustering. In general, communities may not only refer to denser clusters but more generally to groups of vertices that behave similarly.

Community detection and clustering are central problems in machine learning and data science. A large number of data sets can be represented as a network of interacting items, and one of the first features of interest in such networks is to understand which items are “alike,” as an end or as a preliminary step towards other learning tasks. Community detection is used in particular to understand sociological behavior [3, 146, 134], protein to protein interactions [52, 121], gene expressions [57, 62], recommendation systems [115, 148, 158], medical prognosis [151], DNA 3D folding [50], image segmentation [97], natural language processing [29], product-customer segmentation [56], webpage sorting [109], and more.

The field of community detection has been expanding greatly since the 1980’s, with a remarkable diversity of models and algorithms developed in different communities such as machine learning, computer science, network science, social science and statistical physics. These rely

on various benchmarks for finding clusters, in particular, cost functions based on cuts or modularities [82]. We refer to [118, 146, 3, 134] for an overview of these developments.

Nonetheless, various fundamental questions remain unsettled, such as:

- When are there really communities? Algorithms may output community structures, but are these meaningful or artefacts?
- Can we always extract the communities when they are present; fully, partially?
- What is a good benchmark to measure the performance of algorithms, and how good are the current algorithms?

The goal of this monograph is to describe recent developments aimed at answering these questions in the context of block models. Block models are a family of random graphs with planted clusters. The “mother model” is the *stochastic block model (SBM)*, which has been widely employed as a canonical model for community detection. It is arguably the simplest model of a graph with communities (see definitions in the next section). Since the SBM is a generative model, it benefits from a ground truth for the communities, which allows us to consider the previous questions in a formal context. Like any model, it is not necessarily realistic, but it is insightful - judging for example from the powerful algorithms that have emerged from its study.

In a sense, the SBM plays a similar role to the discrete memoryless channel (DMC) in information theory. While the task of modelling external noise may be more amenable to simplifications than real data sets, the SBM captures some of the key bottleneck phenomena for community detection and admits many possible refinements that improve its fit to real data. Our focus here will be on the fundamental understanding of the “canonical SBM,” without diving too much into the refined extensions.

The SBM is defined as follows. For positive integers  $k, n$ , a probability vector  $p$  of dimension  $k$ , and a symmetric matrix  $W$  of dimension  $k \times k$  with entries in  $[0, 1]$ , the model  $\text{SBM}(n, p, W)$  defines an  $n$ -vertex

random graph with vertices split in  $k$  communities, where each vertex is assigned a community label in  $\{1, \dots, k\}$  independently under the community prior  $p$ , and pairs of vertices with labels  $i$  and  $j$  connect independently with probability  $W_{i,j}$ .

Further generalizations allow for labelled edges and continuous vertex labels, connecting to low-rank approximation models and graphons (using the latter terminology as adapted in the statistics literature). For example, a spiked Wigner model with observation  $Y = XX^T + Z$ , where  $X$  is an unknown vector and  $Z$  is Wigner, can be viewed as a labeled graph where edge  $(i, j)$ 's label is given by  $Y_{ij} = X_i X_j + Z_{ij}$ . If the  $X_i$ 's take discrete values, e.g.,  $\{1, -1\}$ , this is closely related to the stochastic block model—see [162] for a precise connection. Continuous labels can also model Euclidean connectivity kernels, an important setting for data clustering. In general, models where a collection of variables  $\{X_i\}$  have to be recovered from noisy observations  $\{Y_{ij}\}$  that are stochastic functions of  $X_i, X_j$ , or more generally that depend on local interactions of some of the  $X_i$ 's, can be viewed as inverse problems on graphs or hypergraphs that bear similarities with the basic community detection problems discussed here. This concerns in particular topic modelling, ranking, synchronization problems and other unsupervised learning problems. We refer to Section 9 for further discussion on these. The specificity of the stochastic block model is that the input variables are discrete.

A first hint at the centrality of the SBM comes from the fact that the model appeared independently in numerous scientific communities. It appeared under the SBM terminology in the context of social networks in the machine learning and statistics literature [93], while the model is typically called the planted partition model in theoretical computer science [49, 119, 41], and the inhomogeneous random graph in the mathematics literature [40]. The model takes also different interpretations, such as a planted spin-glass model [63], a sparse-graph code [13, 68] or a low-rank (spiked) random matrix model [123, 154, 162] among others.

In addition, the SBM has recently turned into more than a model for community detection. It provides a fertile ground for studying various central questions in machine learning, computer science and statistics: It is rich in phase transitions [63, 122, 128, 13, 68], allowing us to study

the interplay between statistical and computational barriers [161, 18, 32, 20], as well as the discrepancies between probabilistic and adversarial models [125], and it serves as a test bed for algorithms, such as SDPs [13, 159, 89, 85, 1, 22, 126, 141], spectral methods [154, 160, 122, 108, 42, 147], and belief propagation [63, 17].

## 1.2 Fundamental limits: information and computation

This monograph focuses on the *fundamental limits* of community detection. The term ‘fundamental limit’ is used to emphasize the fact that we seek conditions for recovering the communities that are *necessary and sufficient*. In the information-theoretic sense, this means finding conditions under which a given task can or cannot be resolved irrespective of complexity or algorithmic considerations, whereas in the computational sense, this further constrains the algorithms to run in polynomial time in the number of vertices. As we shall see in this monograph, such fundamental limits are often expressed through *phase transitions*, which provide sharp transitions in the relevant regimes between phases where the given task can or cannot be resolved.

Fundamental limits have proved to be instrumental in the developments of algorithms. A prominent example is Shannon’s coding theorem [149], which gives a sharp threshold for coding algorithms at the channel capacity, and which has led the development of coding algorithms for more than 60 years (e.g., LDPC, turbo or polar codes) at both the theoretical and practical level [153]. Similarly, the SAT threshold [60] has driven the developments of a variety of satisfiability algorithms such as survey propagation [117].

In the area of clustering and community detection, where establishing rigorous benchmarks is a long standing challenge, the quest of fundamental limits and phase transitions is also impacting the development of algorithms. As discussed in this monograph, this has already lead to developments of algorithms such as sphere-comparisons, linearized belief propagation, nonbacktracking spectral methods. Fundamental limits also shed light on the limitations of the model versus those of the algorithms used; see Section 1.3. However, unlike in the data transmission context of Shannon, information-theoretic limits

may not always be efficiently achievable in community detection, with information-computation gaps that may emerge as discussed in Section 8.

### 1.3 An example on real data

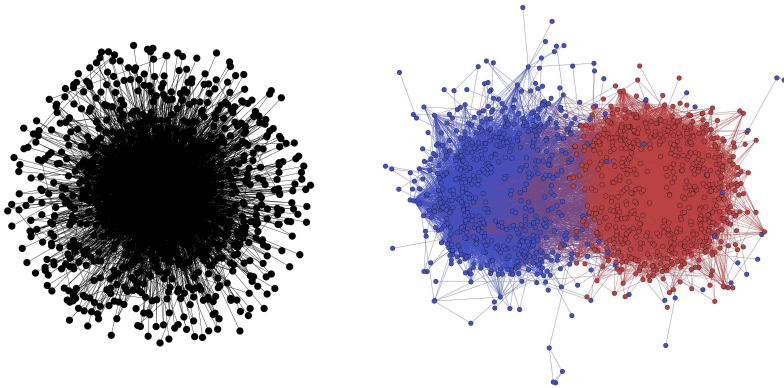
This monograph focuses on the fundamentals of community detection, but we want to give an application example here. We use the blogosphere data set from the 2004 US political elections [110] as an archetype example.

Consider the problem where one is interested in extracting features about a collection of items, in our case  $n = 1,222$  individuals writing about US politics, observing only some of their interactions. In our example, we have access to which blogs refers to which (via hyperlinks), but nothing else about the content of the blogs. The hope is to extract knowledge about the individual features from these simple interactions.

To proceed, build a *graph of interaction* among the  $n$  individuals, connecting two individuals if one refers to the other, ignoring the direction of the hyperlink for simplicity. Assume next that the data set is generated from a stochastic block model; assuming two communities is an educated guess here, but one can also estimate the number of communities (e.g., as in [18]). The type of algorithms developed in Sections 7.2 and 7.1 can then be run on this data set, and two assortative communities are obtained. In the paper [110], Adamic and Glance recorded which blogs are right or left leaning, so that we can check how much agreement the algorithms give with the true partition of the blogs. The results give about 95% agreement on the blogs' political inclinations (which is roughly the state-of-the-art [133, 101, 80]).

Despite the fact that the blog data set is particularly 'well behaved'—there are two dominant clusters that are well balanced and well separated—the above approach can be applied to a broad collection of data sets to extract knowledge about the data from graphs of similarities or interactions. In some applications, the graph is obvious (such as in social networks with friendships), while in others, it is engineered from the data set based on metrics of similarity/interactions that need to be chosen properly (e.g, similarity of pixels in image segmentation). The





**Figure 1.2:** The above graphs represent the real data set of the political blogs from [110]. Each vertex represents a blog and each edge represents the fact that one of the blogs refers to the other. The left graph is plotted with a random arrangement of the vertices, and the right graph is the output of the ABP algorithm described in Section 7.2, which gives 95% accuracy on the reconstruction of the political inclination of the blogs (blue and red colors correspond to left and right leaning blogs).

goal is to apply such approaches to problems where the ground truth is unknown, such as to understand biological functionality of protein complexes; to find genetically related sub-populations; to make accurate recommendations; medical diagnosis; image classification; segmentation; page sorting; and more (see references in the introduction).

In such cases where the ground truth is not available, a key question is to understand how reliable the algorithms' outputs may be. We now discuss how the results presented in this monograph add to this question. Following the definitions from Sections 7.2 and 7.1, the parameters estimated by fitting an SBM on this data set in the constant degree regime are

$$p_1 = 0.48, \quad p_2 = 0.52, \quad Q = \begin{pmatrix} 52.06 & 5.16 \\ 5.16 & 47.43 \end{pmatrix}. \quad (1.1)$$

and in the logarithmic degree regime

$$p_1 = 0.48, \quad p_2 = 0.52, \quad Q = \begin{pmatrix} 7.31 & 0.73 \\ 0.73 & 6.66 \end{pmatrix}. \quad (1.2)$$

Following the definitions of Theorem 7.9 from Section 7.2, we can now compute the SNR for these parameters in the constant-degree regime,

obtaining  $\lambda_2^2/\lambda_1 \approx 18$  which is much greater than 1. Thus, under an SBM model, the data is largely in a regime where communities can be detected, i.e., above the weak recovery threshold. Following the definitions of Theorem 7.1 from Section 7.1, we can also compute the CH-divergence for these parameters in the logarithmic-degree regime, obtaining  $J(p, Q) \approx 2$  which is also greater than 1. Thus, under an SBM and with an asymptotic approximation, the data is in a regime where the graph communities can in fact be recovered entirely, i.e, above the exact recovery threshold. This does not answer whether the SBM is a good or a bad model, but it gives that under this model, the data appears to be in a strong ‘clusterable regime.’

Note also that such a conclusion may not appear using a specific algorithm, e.g., one that is sensitive to the degree variations and that may split the vertices into high vs. low-degree vertices. This prompted for example the development of degree-corrected SBMs in [27], as the algorithm used in [27] for the blog data set with the fitting of an SBM failed for such reasons. However, how do we know whether the failure is due to the model or the algorithm? By establishing the fundamental limits on the SBM, we will find algorithms that are ‘maximally’ robust by succeeding in the most challenging regimes, i.e., down to the fundamental limits, which achieve in particular the positive accuracy for the blog data set described in Figure 1.2. We also refer to Section 5.3.2 for discussions on the robustness of algorithms to degree variations.

## 1.4 Historical overview of the recent developments

This section provides a brief historical overview of the recent developments discussed in this monograph. The resurgent interest in the SBM and its ‘modern study’ have been initiated in part due to the paper of Decelle, Krzakala, Moore and Zdeborová [63], which conjectured<sup>1</sup> phase

---

<sup>1</sup>The conjecture of the Kesten-Stigum threshold in [63] was formulated with what we call in this note the max-detection criteria, asking for an algorithm to output a reconstruction of the communities that strictly improves on the trivial performance achieved by putting all the vertices in the largest community. This conjecture is formally incorrect for general SBMs, see [20] for a counter-example, as the notion of max-detection is too strong in some cases. The conjecture is believed to hold for symmetric SBMs, as re-stated in [130], but it requires a different notion of detection

transition phenomena for the weak recovery (a.k.a. detection) problem at the Kesten-Stigum threshold and the information-computation gap at 4 symmetric communities in the symmetric case. These conjectures are backed in [63] with insights from statistical physics, based on the cavity method (belief propagation), and provide a detailed picture of the weak recovery problem, both for the algorithmic and information-theoretic behavior. With such insights, a new research program started driven by the phase transition phenomena.

One of the first papers that obtained a non-trivial algorithmic result for the weak recovery problem is [2] from 2010, which appeared before the conjecture (and does not achieve the threshold by a logarithmic degree factor). The first paper that made progress on the conjecture is [130] from 2012, which proved the impossibility part of the conjecture for two symmetric communities, introducing various key concepts in the analysis of block models. In 2013, [72] also obtained a result on the partial recovery of the communities, expressing the optimal fraction of mislabelled vertices when the signal-to-noise ratio is large enough in terms of the broadcasting problem on trees [105, 156].

The positive part of the conjecture for efficient algorithm and two communities was first proved in 2014 with [122] and [128], using respectively a spectral method from the matrix of self-avoiding walks and weighted non-backtracking walks between vertices.

In 2014, [10, 13] and [73] found that the exact recovery problem for two symmetric communities has also a phase transition, in the logarithmic rather than constant degree regime, shown to be also efficiently achievable. This relates to a large body of work from the first decades of research on the SBM [49, 119, 41, 150, 59, 123, 38, 55, 154, 161], driven by the exact or almost exact recovery problems without sharp thresholds.

In 2015, the phase transition for exact recovery was obtained for the general SBM [68, 18], and shown to be efficiently achievable irrespective of the number of communities. For the weak recovery problem, [42] showed that the Kesten-Stigum threshold can be achieved with a spectral method based on the nonbacktracking (edge) operator in a fairly general

---

to hold for general SBMs; see definitions from [20] discussed in Section 7.2.

setting (covering SBMs that are not necessarily symmetric), but felt short of settling the conjecture for more than two communities in the symmetric case due to technical reasons. The approach of [42] is based on the ‘spectral redemption’ conjecture made in 2013 in [108], which introduced the use of the nonbacktracking operator as a linearization of belief propagation. This is one of the most elegant approaches to the weak recovery problem, except perhaps for the fact that the matrix is not symmetric (note that the first proof of [122] does provide a solution with a symmetric matrix via the count of self-avoiding walks, albeit less direct to construct). The general conjecture for arbitrary many symmetric or asymmetric communities is settled later in 2015 with [17, 20], relying on a higher-order nonbacktracking matrix and a message passing implementation. It was further shown in [17, 20] that it is possible to cross information-theoretically the Kesten-Stigum threshold in the symmetric case at 4 communities, settling both positive parts of the conjectures from [63]. Crossing at 5 rather than 4 communities is also obtained in [30, 32], which further obtains the scaling of the information-theoretic threshold for a growing number of communities.

In 2016, a tight expression was obtained for partial recovery with two communities in the regime of finite SNR with diverging degrees in [162] and [129] for a different distortion measure. This also gives the threshold for weak recovery in the regime where the SNR is finite while the degrees are diverging.

Other major lines of work on the SBM have been concerned with the performance of SDPs, with a precise picture obtained in [85, 126, 99] for the weak recovery problem and in [13, 159, 1, 5, 22, 141] for the (almost) exact recovery problem; as well as with spectral methods on classical operators [123, 2, 138, 160, 154, 147, 165]. A detailed picture has also been developed for the problem of a single planted community in [4, 91, 90, 51]. Recently, attention has been paid to graphs that have a larger number of short loops [16, 9, 79, 14]. There is a much broader list of works on the SBMs that is not covered in this monograph, especially before the ‘recent developments’ discussed above but also after. It is particularly challenging to track the vast literature on this subject as it is split between different communities of statistics, machine learning, mathematics, computer science, information theory, social sciences

and statistical physics. This monograph mainly covers developments until 2016, with some references from 2017. There are a few additional surveys available; community detection and statistical network models are discussed in [118, 146, 3], and C. Moore has a recent overview paper [127] that focuses on the weak recovery problem with emphasis on the cavity method.

In the table below, we summarize the main thresholds proved for weak and exact recovery, covered in several chapters of this monograph:

	Exact recovery (logarithmic degrees)	Weak recovery (detection) (constant degrees)
2-SSBM	$ \sqrt{a} - \sqrt{b}  > \sqrt{2}$ [10, 73]	$(a - b)^2 > 2(a + b)$ [122, 128]
General SBM	$\min_{i < j} D_+((PQ)_i, (PQ)_j) > 1$ [68]	$\lambda_2^2(PQ) > \lambda_1(PQ)$ [42, 17]

## 1.5 Outline

In the next section, we formally define the SBM and various recovery requirements for community detection, namely exact, weak, and partial recovery. We then start with a quick overview of the key approaches for these recovery requirements in Section 3, introducing the key new concepts obtained in the recent developments. We then treat each of these three recovery requirements separately for the two community SBM in Sections 7.1, 7.2 and 6 respectively, discussing both fundamental limits and efficient algorithms. We give complete (and revised) proofs for exact recovery and partial proofs for weak and partial recovery. We then move to the results for the general SBM in Section 7. In Section 9 we discuss other block models, such as geometric block models, and in Section 10 we give concluding remarks and open problems.

## 1.6 Notations

We use the standard little-o and big-o notations. Recall that  $a_n = \Omega(b_n)$  means that  $b_n = O(a_n)$ , and  $a_n = \omega(b_n)$  means that  $b_n = o(a_n)$ . In particular,  $a_n = o(1)$  means that  $a_n$  is vanishing,  $a_n = \Omega(1)$  means that  $a_n$  is non-vanishing, and  $a_n = \omega(1)$  means that  $a_n$  is diverging. We use  $a_n \lesssim b_n$  when  $a_n = \Omega(b_n)$ ;  $a_n \ll b_n$  when  $a_n = o(b_n)$  (and

$a_n \gg b_n$  when  $b_n = o(a_n)$ ;  $a_n = \Theta(b_n)$ , or equivalently  $a_n \asymp b_n$ , when we simultaneously have  $a_n = \Omega(b_n)$  and  $a_n = O(b_n)$ ;  $a_n \sim b_n$  when  $a_n = b_n(1 + o(1))$ .

We say that an event  $E_n$  takes place with high probability if its probability tends to 1 as  $n$  diverges, i.e.,  $\mathbb{P}\{E_n\} = 1 - o(1)$ . We also use a.a.e. and a.a.s. for asymptotically almost everywhere and asymptotically almost surely (respectively).

We usually use superscripts to specify the dimensions of vectors; in particular,  $1^n$  is the all-one vector of dimension  $n$ ,  $0^n$  the all-zero vector of dimension  $n$ , and  $x^n = (x_1, \dots, x_n)$ .

## References

---

- [1] A. Amini and E. Levina. 2014. “On semidefinite relaxations for the block model”. *arXiv:1406.5647*. June.
- [2] A. Coja-Oghlan. 2010. “Graph partitioning via adaptive spectral techniques”. *Comb. Probab. Comput.* 19(2): 227–284. ISSN: 0963-5483. DOI: [10.1017/S0963548309990514](https://doi.org/10.1017/S0963548309990514). URL: <http://dx.doi.org/10.1017/S0963548309990514>.
- [3] A. Goldenberg, A. X. Zheng, S. E. Fienberg, and E. M. Airoldi. 2010. “A survey of statistical network models”. *Foundations and Trends in Machine Learning*. 2(2): 129–233.
- [4] A. Montanari. 2015. “Finding One Community in a Sparse Graph”. *arXiv:1502.05680*.
- [5] A. S. Bandeira. 2015. “Random Laplacian matrices and convex relaxations”. *arXiv:1504.03987*.
- [6] A. Saade, F. Krzakala, M. Lelarge, and L. Zdeborová. 2015. “Spectral Detection in the Censored Block Model”. Jan. arXiv: [1502.00163](https://arxiv.org/abs/1502.00163).
- [7] Abbe, E. 2016. “Graph compression: The effect of clusters”. In: *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. 1–8. DOI: [10.1109/ALLERTON.2016.7852203](https://doi.org/10.1109/ALLERTON.2016.7852203).
- [8] Abbe, E. 2017. “Community detection and stochastic block models: recent developments”. *ArXiv:1703.10146*. Mar.

- [9] Abbe, E., F. Baccelli, and A. Sankararaman. 2017a. “Community Detection on Euclidean Random Graphs”. *ArXiv:1706.09942*. June.
- [10] Abbe, E., A. S. Bandeira, and G. Hall. 2014a. “Exact Recovery in the Stochastic Block Model”. *ArXiv:1405.3267*. May.
- [11] Abbe, E., A. Bandeira, A. Bracher, and A. Singer. 2014b. “Decoding Binary Node Labels from Censored Edge Measurements: Phase Transition and Efficient Recovery”. *Network Science and Engineering, IEEE Transactions on*. 1(1): 10–22. ISSN: 2327-4697. DOI: [10.1109/TNSE.2014.2368716](https://doi.org/10.1109/TNSE.2014.2368716).
- [12] Abbe, E., A. Bandeira, A. Bracher, and A. Singer. 2014c. “Linear inverse problems on Erdős-Rényi graphs: Information-theoretic limits and efficient recovery”. In: *Information Theory (ISIT), 2014 IEEE International Symposium on*. 1251–1255. DOI: [10.1109/ISIT.2014.6875033](https://doi.org/10.1109/ISIT.2014.6875033).
- [13] Abbe, E., A. Bandeira, and G. Hall. 2016. “Exact Recovery in the Stochastic Block Model”. *Information Theory, IEEE Transactions on*. 62(1): 471–487. ISSN: 0018-9448. DOI: [10.1109/TIT.2015.2490670](https://doi.org/10.1109/TIT.2015.2490670).
- [14] Abbe, E., E. Boix, and C. Sandon. 2017b. “Graph powering and spectral gap extraction”. *Manuscript. Results partly presented at the Simons Institute and partly available in E. Boix PACM Thesis, Princeton University*.
- [15] Abbe, E., J. Fan, K. Wang, and Y. Zhong. 2017c. “Entrywise Eigenvector Analysis of Random Matrices with Low Expected Rank”. *ArXiv:1709.09565*. Sept.
- [16] Abbe, E., L. Massoulié, A. Montanari, A. Sly, and N. Srivastava. 2017d. “Group Synchronization on Grids”. *ArXiv:1706.08561*. June.
- [17] Abbe, E. and C. Sandon. 2015a. “Detection in the stochastic block model with multiple clusters: proof of the achievability conjectures, acyclic BP, and the information-computation gap”. *ArXiv e-prints 1512.09080*. Dec. arXiv: [1512.09080](https://arxiv.org/abs/1512.09080) [[math.PR](https://arxiv.org/abs/1512.09080)].



- [18] Abbe, E. and C. Sandon. 2015b. “Recovering Communities in the General Stochastic Block Model Without Knowing the Parameters”. In: *Advances in Neural Information Processing Systems (NIPS) 28*. Ed. by C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett. Curran Associates, Inc. 676–684.
- [19] Abbe, E. and C. Sandon. 2016. “Achieving the KS threshold in the general stochastic block model with linearized acyclic belief propagation”. In: *Advances in Neural Information Processing Systems 29*. Ed. by D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett. Curran Associates, Inc. 1334–1342.
- [20] Abbe, E. and C. Sandon. 2017. “Proof of the Achievability Conjectures for the General Stochastic Block Model”. *Communications on Pure and Applied Mathematics*. 71(7): 1334–1406. DOI: [10.1002/cpa.21719](https://doi.org/10.1002/cpa.21719). eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpa.21719>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.21719>.
- [21] Achlioptas, D. and A. Naor. 2005. “The Two Possible Values of the Chromatic Number of a Random Graph”. *Annals of Mathematics*. 162(3): 1335–1351. ISSN: 0003486X. URL: <http://www.jstor.org/stable/20159944>.
- [22] Agarwal, N., A. S. Bandeira, K. Koiliaris, and A. Kolla. 2015. “Multisection in the Stochastic Block Model using Semidefinite Programming”. *ArXiv:1507.02323*. July.
- [23] Airoldi, E. M., D. M. Blei, S. E. Fienberg, and E. P. Xing. 2008. “Mixed Membership Stochastic Block Models”. *J. Mach. Learn. Res.* 9(June): 1981–2014. ISSN: 1532-4435. URL: <http://dl.acm.org/citation.cfm?id=1390681.1442798>.
- [24] Aldous, D. J. 1981. “Representations for partially exchangeable arrays of random variables”. *Journal of Multivariate Analysis*. 11(4): 581–598. ISSN: 0047-259X. DOI: [http://dx.doi.org/10.1016/0047-259X\(81\)90099-3](http://dx.doi.org/10.1016/0047-259X(81)90099-3). URL: <http://www.sciencedirect.com/science/article/pii/0047259X81900993>.

- [25] Alon, N. and N. Kahale. 1997. “A Spectral Technique for Coloring Random 3-Colorable Graphs”. *SIAM Journal on Computing*. 26(6): 1733–1748. DOI: [10.1137/S0097539794270248](https://doi.org/10.1137/S0097539794270248). eprint: <http://dx.doi.org/10.1137/S0097539794270248>. URL: <http://dx.doi.org/10.1137/S0097539794270248>.
- [26] Asadi, A. R., E. Abbe, and S. Verdú. 2017. “Compressing data on graphs with clusters”. In: *2017 IEEE International Symposium on Information Theory (ISIT)*. 1583–1587. DOI: [10.1109/ISIT.2017.8006796](https://doi.org/10.1109/ISIT.2017.8006796).
- [27] B. Karrer and M. E. J. Newman. 2011. “Stochastic blockmodels and community structure in networks”. *Phys. Rev. E*. 83(1): 016107. DOI: [10.1103/PhysRevE.83.016107](https://doi.org/10.1103/PhysRevE.83.016107). URL: <http://link.aps.org/doi/10.1103/PhysRevE.83.016107>.
- [28] Baik, J., G. Ben Arous, and S. Péché. 2005. “Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices”. *Ann. Probab.* 33(5): 1643–1697. DOI: [10.1214/009117905000000233](https://doi.org/10.1214/009117905000000233). URL: <http://dx.doi.org/10.1214/009117905000000233>.
- [29] Ball, B., B. Karrer, and M. E. J. Newman. 2011. “An efficient and principled method for detecting communities in networks”. *Phys. Rev. E*. 84(3): 036103. DOI: [10.1103/PhysRevE.84.036103](https://doi.org/10.1103/PhysRevE.84.036103). URL: <http://link.aps.org/doi/10.1103/PhysRevE.84.036103>.
- [30] Banks, J. and C. Moore. 2016. “Information-theoretic thresholds for community detection in sparse networks”. *ArXiv:1601.02658*. Jan.
- [31] Banks, J., C. Moore, N. Verzelen, R. Vershynin, and J. Xu. 2016a. “Information-theoretic bounds and phase transitions in clustering, sparse PCA, and submatrix localization”. July. arXiv: [1607.05222](https://arxiv.org/abs/1607.05222) [[math.ST](https://arxiv.org/abs/1607.05222)].
- [32] Banks, J., C. Moore, J. Neeman, and P. Netrapalli. 2016b. “Information-theoretic thresholds for community detection in sparse networks”. *Proc. of COLT*.
- [33] Bayati, M. and A. Montanari. 2011. “The dynamics of message passing on dense graphs, with applications to compressed sensing”. *Information Theory, IEEE Transactions on*. 57(2): 764–785.

- [34] Bean, D., P. J. Bickel, N. El Karoui, and B. Yu. 2013. “Optimal M-estimation in high-dimensional regression”. *Proceedings of the National Academy of Sciences*. 110(36): 14563–14568.
- [35] Berthet, Q. and P. Rigollet. 2013. “Optimal detection of sparse principal components in high dimension”. *Ann. Statist.* 41(4): 1780–1815. DOI: [10.1214/13-AOS1127](https://doi.org/10.1214/13-AOS1127). URL: <http://dx.doi.org/10.1214/13-AOS1127>.
- [36] Berthet, Q., P. Rigollet, and P. Srivastava. 2016. “Exact recovery in the Ising blockmodel”. *ArXiv:1612.03880*. Dec.
- [37] Bhattacharyya, S. and P. J. Bickel. 2014. “Community Detection in Networks using Graph Distance”. *ArXiv:1401.3915*. Jan.
- [38] Bickel, P. J. and A. Chen. 2009. “A nonparametric view of network models and Newman-Girvan and other modularities”. *Proceedings of the National Academy of Sciences*. 106(50): 21068–21073. DOI: [10.1073/pnas.0907096106](https://doi.org/10.1073/pnas.0907096106). eprint: <http://www.pnas.org/content/106/50/21068.full.pdf>. URL: <http://www.pnas.org/content/106/50/21068.abstract>.
- [39] Bleher, P. M., J. Ruiz, and V. A. Zagrebnoy. 1995. “On the purity of the limiting Gibbs state for the Ising model on the Bethe lattice”. *Journal of Statistical Physics*. 79(1): 473–482. DOI: [10.1007/BF02179399](https://doi.org/10.1007/BF02179399). URL: <http://dx.doi.org/10.1007/BF02179399>.
- [40] Bollobás, B., S. Janson, and O. Riordan. 2007. “The Phase Transition in Inhomogeneous Random Graphs”. *Random Struct. Algorithms*. 31(1): 3–122. ISSN: 1042-9832. DOI: [10.1002/rsa.v31:1](https://doi.org/10.1002/rsa.v31:1). URL: <http://dx.doi.org/10.1002/rsa.v31:1>.
- [41] Boppana, R. 1987. “Eigenvalues and graph bisection: An average-case analysis”. In *28th Annual Symposium on Foundations of Computer Science*: 280–285.
- [42] Bordenave, C., M. Lelarge, and L. Massoulié. 2015. “Non-backtracking Spectrum of Random Graphs: Community Detection and Non-regular Ramanujan Graphs”. In: *Proceedings of the 2015 IEEE 56th Annual Symposium on Foundations of Computer Science (FOCS)*. *FOCS '15*. Washington, DC, USA: IEEE Computer Society. 1347–1357. ISBN: 978-1-4673-8191-8. DOI: [10.1109/FOCS.2015.86](https://doi.org/10.1109/FOCS.2015.86). URL: <http://dx.doi.org/10.1109/FOCS.2015.86>.

- [43] Borgs, C., J. T. Chayes, H. Cohn, and S. Ganguly. 2015a. “Consistent nonparametric estimation for heavy-tailed sparse graphs”. *ArXiv:1508.06675*. Aug.
- [44] Borgs, C., J. T. Chayes, H. Cohn, and Y. Zhao. 2014. “An  $L^p$  theory of sparse graph convergence I: limits, sparse random graph models, and power law distributions”. *ArXiv:1401.2906*. Jan.
- [45] Borgs, C., J. Chayes, C. E. Lee, and D. Shah. 2017. “Iterative Collaborative Filtering for Sparse Matrix Estimation”. *ArXiv:1712.00710*. Dec.
- [46] Borgs, C., J. Chayes, L. Lovasz, V. Sos, and K. Vesztegombi. 2008. “Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing”. *Advances in Mathematics*. 219(6): 1801–1851. ISSN: 0001-8708. DOI: <http://dx.doi.org/10.1016/j.aim.2008.07.008>. URL: <http://www.sciencedirect.com/science/article/pii/S0001870808002053>.
- [47] Borgs, C., J. Chayes, E. Mossel, and S. Roch. 2006. “The Kesten-Stigum Reconstruction Bound Is Tight for Roughly Symmetric Binary Channels”. In: *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science. FOCS '06*. Washington, DC, USA: IEEE Computer Society. 518–530. ISBN: 0-7695-2720-5. DOI: [10.1109/FOCS.2006.76](https://doi.org/10.1109/FOCS.2006.76). URL: <http://dx.doi.org/10.1109/FOCS.2006.76>.
- [48] Borgs, C., J. Chayes, and A. Smith. 2015b. “Private Graphon Estimation for Sparse Graphs”. In: *Advances in Neural Information Processing Systems 28*. Ed. by C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett. Curran Associates, Inc. 1369–1377. URL: <http://papers.nips.cc/paper/5828-private-graphon-estimation-for-sparse-graphs.pdf>.
- [49] Bui, T., S. Chaudhuri, F. Leighton, and M. Sipser. 1987. “Graph bisection algorithms with good average case behavior”. *Combinatorica*. 7(2): 171–191. ISSN: 0209-9683. DOI: [10.1007/BF02579448](https://doi.org/10.1007/BF02579448). URL: <http://dx.doi.org/10.1007/BF02579448>.
- [50] Cabrerós, I., E. Abbe, and A. Tsigros. 2015. “Detecting Community Structures in Hi-C Genomic Data”. *Conference on Information Science and Systems, Princeton University. ArXiv e-prints 1509.05121*. Sept. arXiv: [1509.05121](https://arxiv.org/abs/1509.05121) [q-bio.GN].

- [51] Caltagirone, F., M. Lelarge, and L. Miolane. 2016. “Recovering asymmetric communities in the stochastic block model”. *Allerton*.
- [52] Chen, J. and B. Yuan. 2006. “Detecting functional modules in the yeast protein-protein interaction network”. *Bioinformatics*. 22(18): 2283–2290. DOI: [10.1093/bioinformatics/btl370](https://doi.org/10.1093/bioinformatics/btl370).
- [53] Chen, Y. and A. J. Goldsmith. 2014. “Information Recovery from Pairwise Measurements”. In *Proc. ISIT, Honolulu*.
- [54] Chen, Y., Q.-X. Huang, and L. Guibas. 2014. “Near-Optimal Joint Object Matching via Convex Relaxation”. Available at *arXiv:1402.1473*.
- [55] Choi, D. S., P. J. Wolfe, and E. M. Airoldi. 2012. “Stochastic blockmodels with a growing number of classes”. *Biometrika*: 1–12. DOI: [10.1093/biomet/asr053](https://doi.org/10.1093/biomet/asr053).
- [56] Clauset, A., M. E. J. Newman, and C. Moore. 2004. “Finding community structure in very large networks”. *Phys. Rev. E*. 70(6): 066111. DOI: [10.1103/PhysRevE.70.066111](https://doi.org/10.1103/PhysRevE.70.066111). URL: <http://link.aps.org/doi/10.1103/PhysRevE.70.066111>.
- [57] Cline, M., M. Smoot, E. Cerami, A. Kuchinsky, N. Landys, C. Workman, R. Christmas, I. Avila-Campilo, M. Creech, B. Gross, K. Hanspers, R. Isserlin, R. Kelley, S. Killcoyne, S. Lotia, S. Maere, J. Morris, K. Ono, V. Pavlovic, A. Pico, A. Vailaya, P. Wang, A. Adler, B. Conklin, L. Hood, M. Kuiper, C. Sander, I. Schmulevich, B. Schwikowski, G. J. Warner, T. Ideker, and G. Bader. 2007. “Integration of biological networks and gene expression data using Cytoscape”. *Nature Protocols*. 2(10): 2366–2382. ISSN: 1754-2189. DOI: [10.1038/nprot.2007.324](https://doi.org/10.1038/nprot.2007.324). URL: <http://dx.doi.org/10.1038/nprot.2007.324>.
- [58] Coja-Oghlan, A., F. Krzakala, W. Perkins, and L. Zdeborova. 2016. “Information-theoretic thresholds from the cavity method”. *ArXiv:1611.00814*. Nov.
- [59] Condon, A. and R. M. Karp. 1999. “Algorithms for Graph Partitioning on the Planted Partition Model”. *Lecture Notes in Computer Science*. 1671: 221–232.
- [60] D. Achlioptas, A. Naor, and Y. Peres. 2005. “Rigorous Location of Phase Transitions in Hard Optimization Problems”. *Nature*. 435: 759–764.

- [61] D. Hoover. 1979. *Relations on Probability Spaces and Arrays of Random Variables*. Preprint, Institute for Advanced Study, Princeton.
- [62] D. Jiang, C. Tang, and A. Zhang. 2004. “Cluster analysis for gene expression data: a survey”. *Knowledge and Data Engineering, IEEE Transactions on*. 16(11): 1370–1386. ISSN: 1041-4347. DOI: [10.1109/TKDE.2004.68](https://doi.org/10.1109/TKDE.2004.68).
- [63] Decelle, A., F. Krzakala, C. Moore, and L. Zdeborová. 2011. “Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications”. *Phys. Rev. E*. 84(6): 066106.
- [64] Deshpande, Y. and A. Montanari. 2014. “Information-theoretically optimal sparse PCA”. In: *Information Theory (ISIT), 2014 IEEE International Symposium on*. IEEE. 2197–2201.
- [65] Diaconis, P. and S. Janson. 2007. “Graph limits and exchangeable random graphs”. *ArXiv:0712.2749*. Dec.
- [66] Donoho, D. L., A. Maleki, and A. Montanari. 2009. “Message-passing algorithms for compressed sensing”. *Proceedings of the National Academy of Sciences*. 106(45): 18914–18919. DOI: [10.1073/pnas.0909892106](https://doi.org/10.1073/pnas.0909892106). eprint: <http://www.pnas.org/content/106/45/18914.full.pdf>. URL: <http://www.pnas.org/content/106/45/18914.abstract>.
- [67] E. Abbe and A. Montanari. 2015. “Conditional Random Fields, Planted Constraint Satisfaction, and Entropy Concentration”. *Theory of Computing*. 11(17): 413–443. DOI: [10.4086/toc.2015.v011a017](https://doi.org/10.4086/toc.2015.v011a017). URL: <http://www.theoryofcomputing.org/articles/v011a017>.
- [68] E. Abbe and C. Sandon. 2015a. “Community Detection in General Stochastic Block models: Fundamental Limits and Efficient Algorithms for Recovery”. In: *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS 2015, Berkeley, CA, USA, 17-20 October, 2015*. 670–688. DOI: [10.1109/FOCS.2015.47](https://doi.org/10.1109/FOCS.2015.47). URL: <http://dx.doi.org/10.1109/FOCS.2015.47>.
- [69] E. Abbe and C. Sandon. 2015b. “Community detection in general stochastic block models: fundamental limits and efficient recovery algorithms”. *arXiv:1503.00609*. Mar.

- [70] E. Airoldi, T. Costa, and S. Chan. 2013. “Stochastic blockmodel approximation of a graphon: Theory and consistent estimation”. *arXiv:1311.1731*.
- [71] E. Mossel. 2017. Private communications, 2017.
- [72] E. Mossel, J. Neeman, and A. Sly. 2013. “Belief Propagation, Robust Reconstruction, and Optimal Recovery of Block Models”. *Arxiv:arXiv:1309.1380*.
- [73] E. Mossel, J. Neeman, and A. Sly. 2014. “Consistency Thresholds for Binary Symmetric Block Models”. *Arxiv:arXiv:1407.1591*. In *proc. of STOC15*. July.
- [74] E. Mossel and Y. Peres. 2003. “Information flow on trees”. *Ann. Appl. Probab.* 13: 817–844.
- [75] E. Szemerédi. 1976. “Regular Partitions of Graphs”. *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*.
- [76] Erdős, P. and A. Rényi. 1960. “On the Evolution of Random Graphs”. In: *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*. 17–61.
- [77] Feige, U. and J. Kilian. 2001. “Heuristics for semirandom graph problems”. *Journal of Computer and System Sciences*. 63(4): 639–671.
- [78] Feige, U. and E. Ofek. 2005. “Spectral techniques applied to sparse random graphs”. *Random Structures & Algorithms*. 27(2): 251–275. ISSN: 1098-2418. DOI: [10.1002/rsa.20089](https://doi.org/10.1002/rsa.20089). URL: <http://dx.doi.org/10.1002/rsa.20089>.
- [79] Galhotra, S., A. Mazumdar, S. Pal, and B. Saha. 2017. “The Geometric Block Model”. *ArXiv:1709.05510*. Sept.
- [80] Gao, C., Z. Ma, A. Y. Zhang, and H. H. Zhou. 2015. “Achieving Optimal Misclassification Proportion in Stochastic Block Model”. *ArXiv:1505.03772*. May.
- [81] Ghasemian, A., P. Zhang, A. Clauset, C. Moore, and L. Peel. 2016. “Detectability Thresholds and Optimal Algorithms for Community Structure in Dynamic Networks”. *Physical Review X*. 6(3), 031005: 031005. DOI: [10.1103/PhysRevX.6.031005](https://doi.org/10.1103/PhysRevX.6.031005). arXiv: [1506.06179](https://arxiv.org/abs/1506.06179) [[stat.ML](https://arxiv.org/abs/1506.06179)].

- [82] Girvan, M. and M. E. J. Newman. 2002. “Community structure in social and biological networks”. *Proceedings of the National Academy of Sciences*. 99(12): 7821–7826. DOI: [10.1073/pnas.122653799](https://doi.org/10.1073/pnas.122653799). eprint: <http://www.pnas.org/content/99/12/7821.full.pdf+html>. URL: <http://www.pnas.org/content/99/12/7821.abstract>.
- [83] Goemans, M. X. and D. P. Williamson. 1995. “Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming”. *Journal of the Association for Computing Machinery*. 42: 1115–1145.
- [84] Gopalan, P. K. and D. M. Blei. 2013. “Efficient discovery of overlapping communities in massive networks”. *Proceedings of the National Academy of Sciences*.
- [85] Guédon, O. and R. Vershynin. 2016. “Community detection in sparse networks via Grothendieck’s inequality”. *Probability Theory and Related Fields*. 165(3): 1025–1049. ISSN: 1432-2064. DOI: [10.1007/s00440-015-0659-z](https://doi.org/10.1007/s00440-015-0659-z). URL: <http://dx.doi.org/10.1007/s00440-015-0659-z>.
- [86] Gulikers, L., M. Lelarge, and L. Massoulié. 2015a. “A spectral method for community detection in moderately-sparse degree-corrected stochastic block models”. *ArXiv:1506.08621*. June.
- [87] Gulikers, L., M. Lelarge, and L. Massoulié. 2015b. “An Impossibility Result for Reconstruction in a Degree-Corrected Planted-Partition Model”. *ArXiv:1511.00546*. Nov.
- [88] Guo, D., S. Shamai, and S. Verdú. 2005. “Mutual information and minimum mean-square error in Gaussian channels”. *Information Theory, IEEE Transactions on*. 51(4): 1261–1282.
- [89] Hajek, B., Y. Wu, and J. Xu. 2015a. “Achieving Exact Cluster Recovery Threshold via Semidefinite Programming: Extensions”. *ArXiv:1502.07738*. Feb.
- [90] Hajek, B., Y. Wu, and J. Xu. 2015b. “Information Limits for Recovering a Hidden Community”. *ArXiv:1509.07859*. Sept.
- [91] Hajek, B., Y. Wu, and J. Xu. 2015c. “Recovering a Hidden Community Beyond the Spectral Limit in  $O(|E| \log^* |V|)$  Time”. *ArXiv:1510.02786*.



- [92] Heimlicher, S., M. Lelarge, and L. Massoulié. 2012. “Community Detection in the Labelled Stochastic Block Model”. Sept. arXiv: [1209.2910](https://arxiv.org/abs/1209.2910).
- [93] Holland, P. W., K. Laskey, and S. Leinhardt. 1983. “Stochastic blockmodels: First steps”. *Social Networks*. 5(2): 109–137. URL: [http://d.wanfangdata.com.cn/NSTLQK%5C\\_10.1016-0378-8733\(83\)90021-7.aspx](http://d.wanfangdata.com.cn/NSTLQK%5C_10.1016-0378-8733(83)90021-7.aspx).
- [94] Hopkins, S. B. and D. Steurer. 2017. “Bayesian estimation from few samples: community detection and related problems”. *ArXiv:1710.00264*. Sept.
- [95] I. Csiszár. 1963. “Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten”. *Magyar. Tud. Akad. Mat. Kutató Int. Közl.* 8: 85–108.
- [96] J. Lafferty. 2001. “Conditional random fields: Probabilistic models for segmenting and labeling sequence data”. In: Morgan Kaufmann. 282–289.
- [97] J. Shi and J. Malik. 1997. “Normalized Cuts and Image Segmentation”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 22: 888–905.
- [98] Janson, S. and E. Mossel. 2004. “Robust reconstruction on trees is determined by the second eigenvalue”. *Ann. Probab.* 32(3B): 2630–2649. DOI: [10.1214/009117904000000153](https://doi.org/10.1214/009117904000000153). URL: <http://dx.doi.org/10.1214/009117904000000153>.
- [99] Javanmard, A., A. Montanari, and F. Ricci-Tersenghi. 2016. “Performance of a community detection algorithm based on semidefinite programming”. *ArXiv:1603.09045*. Mar.
- [100] Javanmard, A. and A. Montanari. 2015. “De-biasing the lasso: Optimal sample size for Gaussian designs”. arXiv: [1508.02757](https://arxiv.org/abs/1508.02757).
- [101] Jin, J. 2015. “Fast community detection by SCORE”. *Ann. Statist.* 43(1): 57–89. DOI: [10.1214/14-AOS1265](https://doi.org/10.1214/14-AOS1265). URL: <http://dx.doi.org/10.1214/14-AOS1265>.
- [102] Jog, V. and P.-L. Loh. 2015. “Information-theoretic bounds for exact recovery in weighted stochastic block models using the Renyi divergence”. *ArXiv:1509.06418*. Sept.
- [103] Joseph, A. and B. Yu. 2013. “Impact of regularization on Spectral Clustering”. *ArXiv:1312.1733*. Dec.

- [104] Kawamoto, T. and Y. Kabashima. 2015. “Limitations in the spectral method for graph partitioning: Detectability threshold and localization of eigenvectors”. *Phys. Rev. E*. 91(6): 062803.
- [105] Kesten, H. and B. P. Stigum. 1966. “A Limit Theorem for Multidimensional Galton-Watson Processes”. *Ann. Math. Statist.* 37(5): 1211–1223. DOI: [10.1214/aoms/1177699266](https://doi.org/10.1214/aoms/1177699266). URL: <http://dx.doi.org/10.1214/aoms/1177699266>.
- [106] Kim, C., A. S. Bandeira, and M. X. Goemans. 2017. “Community Detection in Hypergraphs, Spiked Tensor Models, and Sum-of-Squares”. *ArXiv:1705.02973*. May.
- [107] Krauthgamer, R. and U. Feige. 2006. “A Polylogarithmic Approximation of the Minimum Bisection”. *SIAM Review*. 48(1): 99–130. DOI: [10.1137/050640904](https://doi.org/10.1137/050640904). eprint: <http://dx.doi.org/10.1137/050640904>. URL: <http://dx.doi.org/10.1137/050640904>.
- [108] Krzakala, F., C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborova, and P. Zhang. 2013. “Spectral redemption in clustering sparse networks”. *Proceedings of the National Academy of Sciences*. 110(52): 20935–20940. DOI: [10.1073/pnas.1312486110](https://doi.org/10.1073/pnas.1312486110). eprint: <http://www.pnas.org/content/110/52/20935.full.pdf>. URL: <http://www.pnas.org/content/110/52/20935.abstract>.
- [109] Kumar, R., P. Raghavan, S. Rajagopalan, and A. Tomkins. 1999. “Trawling the Web for Emerging Cyber-communities”. *Comput. Netw.* 31(11-16): 1481–1493. ISSN: 1389-1286. DOI: [10.1016/S1389-1286\(99\)00040-7](https://doi.org/10.1016/S1389-1286(99)00040-7). URL: [http://dx.doi.org/10.1016/S1389-1286\(99\)00040-7](http://dx.doi.org/10.1016/S1389-1286(99)00040-7).
- [110] L. Adamic and N. Glance. 2005. “The Political Blogosphere and the 2004 U.S. Election: Divided They Blog”. In: *Proceedings of the 3rd International Workshop on Link Discovery. LinkKDD '05*. Chicago, Illinois. 36–43. ISBN: 1-59593-215-1. DOI: [10.1145/1134271.1134277](https://doi.org/10.1145/1134271.1134277). URL: <http://doi.acm.org/10.1145/1134271.1134277>.
- [111] L. Lovász and B. Szegedy. 2006. “Limits of dense graph sequences”. *Journal of Combinatorial Theory, Series B*. 96(6): 933–957. ISSN: 0095-8956. DOI: [10.1016/j.jctb.2006.05.002](https://doi.org/10.1016/j.jctb.2006.05.002). URL: <http://www.sciencedirect.com/science/article/pii/S0095895606000517>.

- [112] Le, C. M., E. Levina, and R. Vershynin. 2015. “Sparse random graphs: regularization and concentration of the Laplacian”. *ArXiv:1502.03049*. Feb.
- [113] Lelarge, M. and L. Miolane. 2016. “Fundamental limits of symmetric low-rank matrix estimation”. arXiv: [1611.03888](https://arxiv.org/abs/1611.03888).
- [114] Lesieur, T., F. Krzakala, and L. Zdeborová. 2015. “MMSE of probabilistic low-rank matrix estimation: Universality with respect to the output channel”. *ArXiv:1507.03857*. July.
- [115] Linden, G., B. Smith, and J. York. 2003. “Amazon.Com Recommendations: Item-to-Item Collaborative Filtering”. *IEEE Internet Computing*. 7(1): 76–80. ISSN: 1089-7801. DOI: [10.1109/MIC.2003.1167344](https://doi.org/10.1109/MIC.2003.1167344). URL: <http://dx.doi.org/10.1109/MIC.2003.1167344>.
- [116] Lovász, L. 2012. *Large Networks and Graph Limits*. American Mathematical Society colloquium publications. American Mathematical Society. ISBN: 9780821890851. URL: <http://books.google.com/books?id=FsFqHLid8sAC>.
- [117] M. Mézard, G. Parisi, and R. Zecchina. 2003. “Analytic and Algorithmic Solution of Random Satisfiability Problems”. *Science*. 297: 812–815.
- [118] M. Newman. 2010. *Networks: an introduction*. Oxford: Oxford University Press.
- [119] M.E. Dyer and A.M. Frieze. 1989. “The solution of some random NP-hard problems in polynomial expected time”. *Journal of Algorithms*. 10(4): 451–489. ISSN: 0196-6774. DOI: [10.1016/0196-6774\(89\)90001-1](https://doi.org/10.1016/0196-6774(89)90001-1). URL: <http://www.sciencedirect.com/science/article/pii/0196677489900011>.
- [120] Makarychev, K., Y. Makarychev, and A. Vijayaraghavan. 2015. “Learning Communities in the Presence of Errors”. Nov. arXiv: [1511.03229](https://arxiv.org/abs/1511.03229) [cs.DS].
- [121] Marcotte, E., M. Pellegrini, H.-L. Ng, D. Rice, T. Yeates, and D. Eisenberg. 1999. “Detecting Protein Function and Protein-Protein Interactions from Genome Sequences”. *Science*. 285(5428): 751–753. DOI: [10.1126/science.285.5428.751](https://doi.org/10.1126/science.285.5428.751).

- [122] Massoulié, L. 2014. “Community detection thresholds and the weak Ramanujan property”. In: *STOC 2014: 46th Annual Symposium on the Theory of Computing*. New York, United States. 1–10. URL: <https://hal.archives-ouvertes.fr/hal-00969235>.
- [123] McSherry, F. 2001. “Spectral partitioning of random graphs”. In: *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on*. 529–537. DOI: [10.1109/SFCS.2001.959929](https://doi.org/10.1109/SFCS.2001.959929).
- [124] Mézard, M. and A. Montanari. 2006. “Reconstruction on Trees and Spin Glass Transition”. English. *Journal of Statistical Physics*. 124(6): 1317–1350. ISSN: 0022-4715. DOI: [10.1007/s10955-006-9162-3](https://doi.org/10.1007/s10955-006-9162-3). URL: <http://dx.doi.org/10.1007/s10955-006-9162-3>.
- [125] Moitra, A., W. Perry, and A. S. Wein. 2016. “How robust are reconstruction thresholds for community detection?” In: *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing*. ACM. 828–841.
- [126] Montanari, A. and S. Sen. 2016. “Semidefinite Programs on Sparse Random Graphs and Their Application to Community Detection”. In: *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing. STOC 2016*. Cambridge, MA, USA: ACM. 814–827. ISBN: 978-1-4503-4132-5. DOI: [10.1145/2897518.2897548](https://doi.org/10.1145/2897518.2897548). URL: <http://doi.acm.org/10.1145/2897518.2897548>.
- [127] Moore, C. 2017. “The Computer Science and Physics of Community Detection: Landscapes, Phase Transitions, and Hardness”. *ArXiv:1702.00467*. Feb.
- [128] Mossel, E., J. Neeman, and A. Sly. 2014. “A Proof Of The Block Model Threshold Conjecture”. Jan. arXiv: [1311.4115](https://arxiv.org/abs/1311.4115).
- [129] Mossel, E. and J. Xu. 2015. “Density Evolution in the Degree-correlated Stochastic Block Model”. *ArXiv:1509.03281*. Sept.
- [130] Mossel, E., J. Neeman, and A. Sly. 2015. “Reconstruction and estimation in the planted partition model”. *Probability Theory and Related Fields*. 162(3): 431–461. ISSN: 1432-2064. DOI: [10.1007/s00440-014-0576-6](https://doi.org/10.1007/s00440-014-0576-6). URL: <http://dx.doi.org/10.1007/s00440-014-0576-6>.

- [131] Murphy, K. P., Y. Weiss, and M. I. Jordan. 1999. “Loopy Belief Propagation for Approximate Inference: An Empirical Study”. In: *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence. UAI'99*. Stockholm, Sweden: Morgan Kaufmann Publishers Inc. 467–475. ISBN: 1-55860-614-9. URL: <http://dl.acm.org/citation.cfm?id=2073796.2073849>.
- [132] Nadakuditi, R. R. and M. E. J. Newman. 2012. “Graph Spectra and the Detectability of Community Structure in Networks”. *Phys. Rev. Lett.* 108(18): 188701. DOI: [10.1103/PhysRevLett.108.188701](https://doi.org/10.1103/PhysRevLett.108.188701). URL: <http://link.aps.org/doi/10.1103/PhysRevLett.108.188701>.
- [133] Newman, M. E. J. 2011. “Communities, modules and large-scale structure in networks”. *Nature Physics*. 8(1): 25–31. ISSN: 1745-2473. DOI: [10.1038/nphys2162](https://doi.org/10.1038/nphys2162). URL: <http://dx.doi.org/10.1038/nphys2162>.
- [134] Newman, M. E. J., D. J. Watts, and S. H. Strogatz. “Random Graph Models of Social Networks”. *Proc. Natl. Acad. Sci. USA*. 99: 2566–2572.
- [135] Newman, M. E. and T. P. Peixoto. 2015. “Generalized communities in networks”. *Phys. Rev. Lett.* 115(8): 088701.
- [136] O’Rourke, S., V. Vu, and K. Wang. 2013. “Random perturbation of low rank matrices: Improving classical bounds”. Nov. arXiv: [1311.2657 \[math.NA\]](https://arxiv.org/abs/1311.2657).
- [137] Olhede, S. C. and P. J. Wolfe. 2014. “Network histograms and universality of blockmodel approximation”. *Proceedings of the National Academy of Sciences*. 111(41): 14722–14727. DOI: [10.1073/pnas.1400374111](https://doi.org/10.1073/pnas.1400374111). eprint: <http://www.pnas.org/content/111/41/14722.full.pdf>.
- [138] P. Chin, A. Rao, and V. Vu. 2015. “Stochastic Block Model and Community Detection in the Sparse Graphs: A spectral algorithm with optimal rate of recovery”. *arXiv:1501.05021*. Jan.
- [139] Palla, G., I. Derenyi, I. Farkas, and T. Vicsek. 2005. “Uncovering the overlapping community structure of complex networks in nature and society”. *Nature*. 435: 814–818.

- [140] Peixoto, T. P. 2015. “Model selection and hypothesis testing for large-scale network models with overlapping groups”. *Phys. Rev. X*. 5(1): 011033.
- [141] Perry, A. and A. S. Wein. 2015. “A semidefinite program for unbalanced multisection in the stochastic block model”. July. arXiv: [1507.05605](https://arxiv.org/abs/1507.05605) [cs.DS].
- [142] Perry, A., A. S. Wein, and A. S. Bandeira. 2016a. “Statistical limits of spiked tensor models”. *ArXiv:1612.07728*. Dec.
- [143] Perry, A., A. S. Wein, A. S. Bandeira, and A. Moitra. 2016b. “Message-passing algorithms for synchronization problems over compact groups”. *ArXiv:1610.04583*. Oct.
- [144] Perry, A., A. S. Wein, A. S. Bandeira, and A. Moitra. 2016c. “Optimality and Sub-optimality of PCA for Spiked Random Matrices and Synchronization”. *ArXiv:1609.05573*. Sept.
- [145] Reichardt, J. and M. Leone. 2008. “(Un)detectable cluster structure in sparse networks”. *Phys. Rev. Lett.* 101(7): 078701.
- [146] S. Fortunato. 2010. “Community detection in graphs”. *Physics Reports*. 486 (3-5): 75–174.
- [147] S. Yun and A. Proutiere. 2014. “Accurate Community Detection in the Stochastic Block Model via Spectral Algorithms”. *arXiv:1412.7335*. Dec.
- [148] Sahebi, S. and W. Cohen. 2011. “Community-Based Recommendations: a Solution to the Cold Start Problem”. In: *Workshop on Recommender Systems and the Social Web (RSWEB), held in conjunction with ACM RecSys11*. URL: <http://d-scholarship.pitt.edu/13328/>.
- [149] Shannon, C. E. 1948. “A Mathematical Theory of Communication”. *The Bell System Technical Journal*. 27(July): 379–423. URL: <http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf>.
- [150] Snijders, T. A. B. and K. Nowicki. 1997. “Estimation and Prediction for Stochastic Blockmodels for Graphs with Latent Block Structure”. *Journal of Classification*. 14(1): 75–100. ISSN: 0176-4268. DOI: [10.1007/s003579900004](https://doi.org/10.1007/s003579900004). URL: <http://dx.doi.org/10.1007/s003579900004>.

- [151] Sorlie, T., C. Perou, R. Tibshirani, T. Aas, S. Geisler, H. Johnsen, T. Hastie, M. Eisen, M. van de Rijn, S. Jeffrey, T. Thorsen, H. Quist, J. Matese, P. Brown, D. Botstein, P. Lonning, and A. Borresen-Dale. 2001. “Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications”. 98(19): 10869–10874. DOI: [10.1073/pnas.191367098](https://doi.org/10.1073/pnas.191367098).
- [152] Spielman, D. A. and N. Srivastava. 2011. “Graph Sparsification by Effective Resistances”. *SIAM Journal on Computing*. 40(6): 1913–1926. DOI: [10.1137/080734029](https://doi.org/10.1137/080734029). eprint: <https://doi.org/10.1137/080734029>. URL: <https://doi.org/10.1137/080734029>.
- [153] T. Richardson and R. Urbanke. 2001. “An Introduction to the Analysis of Iterative Coding Systems”. In: *Codes, Systems, and Graphical Models. IMA Volume in Mathematics and Its Applications*. Springer. 1–37.
- [154] Vu, V. 2014. “A simple SVD algorithm for finding hidden partitions”. Available at *arXiv:1404.3918*. To appear in *CPC*. Apr.
- [155] Vu, V. H. 2007. “Spectral Norm of Random Matrices”. *Combinatorica*. 27(6): 721–736. ISSN: 0209-9683. DOI: [10.1007/s00493-007-2190-z](https://doi.org/10.1007/s00493-007-2190-z). URL: <http://dx.doi.org/10.1007/s00493-007-2190-z>.
- [156] W. Evans, C. Kenyon, Y. Peres, and L. J. Schulman. 2000. “Broadcasting on trees and the Ising model”. *Ann. Appl. Probab.* 10: 410–433.
- [157] Wolfe, P. J. and S. C. Olhede. 2013. “Nonparametric graphon estimation”. *ArXiv:1309.5936*. Sept.
- [158] Wu, R., J. Xu, R. Srikant, L. Massoulié, M. Lelarge, and B. Hajek. 2015a. “Clustering and Inference From Pairwise Comparisons”. *ArXiv:1502.04631*. Feb.
- [159] Wu, Y., J. Xu, and B. Hajek. 2015b. “Achieving exact cluster recovery threshold via semidefinite programming under the stochastic block model”. In: *2015 49th Asilomar Conference on Signals, Systems and Computers*. 1070–1074. DOI: [10.1109/ACSSC.2015.7421303](https://doi.org/10.1109/ACSSC.2015.7421303).
- [160] Xu, J., M. Lelarge, and L. Massoulié. 2014. “Edge Label Inference in Generalized Stochastic Block Models: from Spectral Theory to Impossibility Results”. *Proceedings of COLT 2014*.

- [161] Y. Chen, J. X. 2014. “Statistical-Computational Tradeoffs in Planted Problems and Submatrix Localization with a Growing Number of Clusters and Submatrices”. *arXiv:1402.1267*. Feb.
- [162] Y. Deshpande, E. Abbe, and A. Montanari. 2015. “Asymptotic Mutual Information for the Two-Groups Stochastic Block Model”. *arXiv:1507.08685*.
- [163] Young, J.-G., P. Desrosiers, L. Hébert-Dufresne, E. Laurence, and L. J. Dubé. 2016. “Finite size analysis of the detectability limit of the stochastic block model”. *arXiv:1701.00062*.
- [164] Yun, S.-Y. and A. Proutiere. 2014. “Community Detection via Random and Adaptive Sampling”. *ArXiv e-prints 1402.3072*. In *proc. COLT14*. Feb.
- [165] Yun, S.-Y. and A. Proutiere. 2015. “Optimal Cluster Recovery in the Labeled Stochastic Block Model”. *ArXiv:1510.05956*. Oct.
- [166] Zhang, P., C. Moore, and M. E. J. Newman. 2016. “Community detection in networks with unequal groups”. *Phys. Rev. E*. 93(1): 012303. DOI: [10.1103/PhysRevE.93.012303](https://doi.org/10.1103/PhysRevE.93.012303). URL: <http://link.aps.org/doi/10.1103/PhysRevE.93.012303>.
- [167] Zhang, P., C. Moore, and L. Zdeborová. 2014. “Phase transitions in semisupervised clustering of sparse networks”. *Phys. Rev. E*. 90(5): 052802. DOI: [10.1103/PhysRevE.90.052802](https://doi.org/10.1103/PhysRevE.90.052802). URL: <http://link.aps.org/doi/10.1103/PhysRevE.90.052802>.
- [168] Zhong, Y. and N. Boumal. 2017. “Near-optimal bounds for phase synchronization”. *arXiv preprint arXiv:1703.06605*.