

Washington University in St. Louis
Washington University Open Scholarship

All Theses and Dissertations (ETDs)

6-25-2012

Statistical Mechanics of the Community Detection Problem: Theory and Application

Dandan Hu

Washington University in St. Louis

Follow this and additional works at: <http://openscholarship.wustl.edu/etd>

Recommended Citation

Hu, Dandan, "Statistical Mechanics of the Community Detection Problem: Theory and Application" (2012). *All Theses and Dissertations (ETDs)*. 959.
<http://openscholarship.wustl.edu/etd/959>

This Dissertation is brought to you for free and open access by Washington University Open Scholarship. It has been accepted for inclusion in All Theses and Dissertations (ETDs) by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Physics

Dissertation Examination Committee:

Zohar Nussinov, Chair

Samuel Achilefu

Anders Carlsson

John Clark

Tao Ju

Ralf Wessel

Statistical Mechanics of the Community Detection Problem: Theory and

Application

by

Dandan Hu

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2012

Saint Louis, Missouri

© copyright by

Dandan Hu

2012

Abstract

We study phase transitions in spin glass type systems and in related computational problems. In the current work, we focus on the “community detection” problem when cast in terms of a general Potts spin glass type problem. We report on phase transitions between solvable and unsolvable regimes. Solvable region may further split into easy and hard phases. Spin glass type phase transitions appear at both low and high temperatures. Low temperature transitions correspond to an order by disorder type effect wherein fluctuations render the system ordered or solvable. Separate transitions appear at higher temperatures into a disordered (or an unsolvable) phases. Different sorts of randomness lead to disparate behaviors. We illustrate the spin glass character of both transitions and report on memory effects. We further relate Potts type spin systems to mechanical analogs and suggest how chaotic-type behavior in general thermodynamic systems can indeed naturally arise in hard-computational problems and spin-glasses. In this work, we also examine large networks (with a power law distribution in cluster size) that have a large number of communities. We infer that large systems at a constant ratio of q to the number of nodes N asymptotically tend toward insolvability in the limit of large N for any positive temperature. We

further employ multivariate Tutte polynomials to show that increasing q emulates increasing T for a general Potts model, leading to a similar stability region at low T .

We further apply the replica inference based Potts model method to unsupervised image segmentation on multiple scales. This approach was inspired by the statistical mechanics problem of “community detection” and its phase diagram. The problem is cast as identifying tightly bound clusters against a background. Within our multiresolution approach, we compute information theory based correlations among multiple solutions of the same graph over a range of resolutions. Significant multiresolution structures are identified by replica correlations as manifest in information overlaps. With the aid of these correlations as well as thermodynamic measures, the phase diagram of the corresponding Potts model is analyzed both at zero and finite temperatures. Optimal parameters corresponding to a sensible unsupervised segmentation correspond to the easy phase of the Potts model. Our algorithm is fast and shown to be at least as accurate as the best algorithms to date and to be especially suited to the detection of camouflage images.

Acknowledgements

First of all I would like to thank my thesis advisor Professor Zohar Nussinov, for suggesting phase transition as a possible dissertation topic, and for the truly commendable patience and willingness to help in the past four years. No single section of this thesis would be possible without his guidance and supervision. I am appreciative of the time and effort of my research committee members Professor John Clark and Ralf Wessel. This thanks extend also to my dissertation examination committee Professor Anders Carlsson, Samuel Achilefu and Tao Ju. I am also grateful to Dr Peter Ronhovde for many useful discussions over the years. I wish to thank the group-mates Saurish Chakrabarty and Patrick Jonhson for many useful conversations. I also wish to thank Sai Iyer and Richard A. Schmaeng for helping me with the computer problems during the years. I really appreciate the time spent together with my officemates Nilushi Dasanayake and Simin Mahmoodifar. My life at WUSTL was made much easier by the work of many secretaries in the WUSTL physics department, including Julia Hamilton, Sarah Hedley, Tammy White-Devine, Christina Saldivar and Trecia Stumbaugh. I would like to thank my Chinese friends, including Yuan Yao, Yan Sang, Wei Wang, Shouting Huang, Zhenyu Zhou, Ruizhe Wang, Jian Wu,

who made my PhD life colorful. Finally I wish to thank my parents Xianqiang Hu and Xiaojuan Jiang for all their love and support.

Contents

Abstract	ii
Acknowledgements	iv
Table of Contents	vi
List of Figures	ix
List of Tables	xi
1 Introduction	1
1.1 Community detection algorithm	1
1.2 Main results of this thesis	6
1.3 Phase transition	6
1.4 Image segmentation	12
1.5 Overview of thesis	18
2 Phase transition in community detection problem	20
2.1 Introduction	20
2.2 The Potts model	23
2.3 Definitions: Trials and Replicas	25
2.4 Heat Bath Algorithm	29
2.5 Information theory and complexity measures	32
2.6 The construction of random graphs and noise test	36
2.7 Theoretical analysis of the community detection problem	39
2.8 Spin glass type transitions	43
2.8.1 Information theory correlation and thermodynamic quantities	43
2.8.2 Phase diagram as ascertained by numerical data	44
2.8.3 Spin glass character of the two transitions	57
2.8.4 General discussion	66
2.9 Dynamical Aspects	72
2.9.1 Relating discrete Hamiltonians to continuous dynamics	73
2.9.2 Numerical results for the continuous dynamical analog	74
2.9.3 Node trajectories in the presence of a biasing field	79
2.9.4 From the classical overdamped system to quantum system	79

2.10 Conclusion	81
3 The stability to instability transition	83
3.1 Introduction	83
3.2 Potts Hamiltonian	87
3.3 Construction of embedded graphs and the noise test	88
3.4 Spin glass type transitions	93
3.4.1 $\chi(T, p_{out})$ at fixed $\alpha = q/N$	94
3.4.2 $\chi(T, p_{out})$ at fixed q	110
3.4.3 Other information theoretic and thermodynamic quantities .	119
3.5 non-interacting cliques	120
3.5.1 Partition function	122
3.5.2 Symmetry Breaking	123
3.5.3 Simulated crossover temperature	126
3.5.4 A discussion of the crossover temperature	128
3.6 An example of a Phase transition in an image segmentation problem .	132
3.7 free energy: Simple results	132
3.7.1 Free energy of non-interacting cliques under a large q expansion	133
3.7.2 Free energy of non-interacting cliques under a high T expansion	134
3.7.3 Free energy of “circle-of-cliques” in the high q/T expansion .	137
3.7.4 Free energy of a random graph in a large q/T expansion . .	138
3.7.5 Free energy of an arbitrary graph G in the large T expansion .	141
3.7.6 Annealed versus quenched averages	143
3.7.7 Free energy for an arbitrary Potts model in a large T expansion	144
3.7.8 Free energy for an arbitrary Potts model in a large q expansion	146
3.8 Tutte polynomials	147
3.8.1 Unweighted graph G	149
3.8.2 Weighted graph G	151
3.9 Derivation of the Tutte polynomial for a circle of cliques	153
3.10 Random knot “transitions”	154
3.11 Conclusions	156
4 Image segmentation	159
4.1 Introduction	159
4.2 Potts Models	162
4.2.1 Representation	163
4.2.2 Potts model Hamiltonian for unweighted graphs	163
4.2.3 Potts Model Hamiltonian for Weighted Graphs	165
4.3 Casting Images as Networks	167
4.3.1 Edge definition for images with uniform intensity	168
4.3.2 Edge definition for images with varying intensities	169
4.3.3 Edge definition for images with “mixed type” intensities . .	176
4.4 The community detection algorithm	176

4.5	Multi-scale networks	178
4.5.1	Apply multiresolution algorithm for a hierarchical network	180
4.6	Replica correlations as weights in a graph	183
4.7	Summary of parameters	185
4.8	The determination of optimal parameters	187
4.9	Results	190
4.9.1	Brain Image	190
4.9.2	A painting by Salvador Dali	198
4.9.3	Benchmarks	198
4.9.4	Images with spatially varying intensities	206
4.9.5	Detection of camouflaged objects	210
4.9.6	Phase Diagram	218
4.10	Conclusions	226
5	Segmenting Intracellular Distribution Images	229
5.1	Introduction	229
5.2	Near-Infrared Fluorescence Lifetime Imaging Microscopy	232
5.3	Segmentation of Imaging Described by 2D Spatial Information	234
5.3.1	Community Detection Method	235
5.3.2	Mixture of Gaussian Distribution Method	236
5.4	Segmentation of Images Described by 3D Spatio-Temporal Information	238
5.4.1	Community Detection Method	238
5.4.2	Mixture of Gaussian Distribution Method	240
5.4.3	The Classification Expectation Maximization Algorithm	240
5.5	Results	242
5.6	Conclusion	244

List of Figures

1.1	Example depiction of a graph with natural communities	3
1.2	Examples of the phase transition in the computational problems	9
1.3	Applications of image segmentation to social life and military	14
1.4	Application of image segmentation to medical image	15
2.1	A caricature of the information theory correlations between replicas .	25
2.2	A schematic plot of the convergence time and trials	26
2.3	Comparing the accuracy with other algorithms	27
2.4	Plots of various measures as a function of the noise	35
2.5	Plots of various measures as the function of temperature and noise .	37
2.6	A caricature of the accessible energy landscape at different T	47
2.7	I_N as a function of noise for system $N = 2048$ at zero temperature .	48
2.8	A comparison of I_N as a function of noise between the optimal γ and 1	49
2.9	The plot of E versus t for the system $N = 1024$, $q = 70$	51
2.10	The hysteresis-like effect for E as a function of T	52
2.11	The hysteresis-like effect for I_N as a function of noise	53
2.12	The hysteresis-like effect in the autocorrelation function versus time .	54
2.13	The autocorrelation as a function of time for system $N = 512, q = 40$	55
2.14	A validation of the spin glass character of the low T hard phase . .	56
2.15	An illustration of the spin glass character of the high T hard phase .	60
2.16	The Cartesian components of the node trajectories as a function of t .	75
2.17	The Cartesian components of the node trajectories under applied fields	76
3.1	The illustration of a partition into communities	86
3.2	3D plot of $\chi(T, p_{out})$ for systems with the fixed ratio of $\alpha = q/N$. .	92
3.3	The boundaries of the hard phase for systems with a fixed $\alpha = q/N$.	95
3.4	The first transition point vs T for systems with a fixed $\alpha = q/N$. .	96
3.5	The convergence time vs noise for systems with a fixed $\alpha = q/N$. .	97
3.6	3D plots of $\chi(T, p_{out})$ for systems with a fixed q	100
3.7	The boundaries of the hard phase for systems with a fixed q	101
3.8	The first transition point vs T for systems with a fixed q	102
3.9	The convergence time vs noise for systems with fixed a q	103
3.10	3D plots of various measures show easy-hard-unsolvable phases . . .	114
3.11	The consistency of transition points in various measures for $N = 2048$	115

3.12	The consistency of transition points in various measures for $N = 1024$	116
3.13	A collapse of the correlation curves for $N = 2048$	117
3.14	A schematic Plot of the clique	121
3.15	The consistency of crossover temperature by simulation and analysis	127
3.16	The segmentation results of the bird image	129
3.17	The 3D phase diagram of $I_N(T, \gamma)$ for the bird image	130
3.18	A schematic plot of the circle of clique	137
3.19	A sample depiction of a random graph with N nodes	139
3.20	The derivation of the Tutte polynomial for cliques	148
3.21	A sample depiction of the trefoil knot	155
4.1	Examples of currently challenging problems in image segmentation	160
4.2	An example of overlapping blocks	170
4.3	Heterogeneous hierarchical system	171
4.4	Plot of information measures vs the resolution parameter γ	172
4.5	The brain image and its unweighted segmentation results	191
4.6	The brain image and its weighted segmentation results (I)	192
4.7	The brain image and its weighted segmentation results (II)	193
4.8	The “Lincoln” image and its segmentation result	197
4.9	The results on Berkeley image segmentation benchmark	199
4.10	The improvement of “F” value by removing high precision features	200
4.11	The results on Microsoft Research benchmark	204
4.12	The segmentation results for the quasicrystal images	205
4.13	Examples of the image segmentation results with Fourier weights	207
4.14	The segmentation results for the “camouflage” images	208
4.15	The multiresolution results of zebra with $\gamma = 1$	209
4.16	The multiresolution result of zebra with $q = 3$ and $\gamma = 1$	212
4.17	The image segmentation of zebra with negative background	213
4.18	Results of our algorithm vs the length scale for the dalmatian dog	215
4.19	The 3D phase diagram of various parameters for the dalmatian dog	219
4.20	The squares of gradients of various measures to the dalmatian dog	220
4.21	The 3D phase diagram of various measures for the bird image	221
4.22	The segmentation results of the bird image	224
5.1	Segmentation results for an liver tissue image with 2D data	246
5.2	Comparison between the CD and MGD method for 2D data	247
5.3	Comparison between the CD and MGD method for 3D data	248

List of Tables

3.1	The crossover temperatures for different clique size	124
4.1	The comparison of F measure by our algorithm with the gPb algorithm	202
4.2	The improvement of F value by removing small noise	202

Chapter 1

Introduction

1.1 Community detection algorithm

“Community detection” [1] describes the problem of finding the closely related subgroups within a general network, as shown in Fig. 1.1. In the study of complex networks, a network is said to have community structure if the nodes of the network can be easily grouped into sets of nodes such that each set of nodes is densely connected internally. In the particular case of non-overlapping community finding, this implies that the network divides naturally into groups of nodes with dense connections internally and sparser connections between groups. A more general definition of community detection is to aim to place pairs of nodes in the same community if they are connected to each other.

In the study of networks, such as computer and information networks [2, 3, 4, 5], social networks [6] or biological networks [7, 8], a number of different characteristics

have been found to occur commonly, including clustering, heavy-tailed degree distributions, and the small-world property. Community structures are quite common in real networks. Social networks often include community groups based on common interests, location, occupation, etc. Metabolic networks [6] have communities based on functional groupings. Citation networks form communities by research topic. Being able to identify these sub-structures within a network can provide insight into how network function and topology affect each other. An introduction to the “physics of networks” can be found in [1]. Reviews of the field are found in Refs.[9, 10], and the most recent thorough review is found in Ref. [11].

Finding communities within an arbitrary network can be a difficult task. The number of communities within the network is typically unknown and the communities are often of unequal size and density. Several methods for community finding have been developed and employed with varying levels of success. The methods include minimum-cut method [12], hierarchical clustering, Girvan-Newman algorithm [6], modularity maximization [13], “surprise” maximization [14], the Louvain method [15], clique based methods [16], etc.

One method of quantitatively assessing community structure is through the use of a “quality function”. A quality function essentially provides an objective measure of how “clustered” or “modular” a network is. Examples include the prominent modularity measure defined by Newman and Girvan [6], a Potts model originally proposed by Reichardt and Bornholdt (RB) [17], our Potts model [18] that eliminates the random partition applied by RB, an application of a Potts model utilizing a mean-field

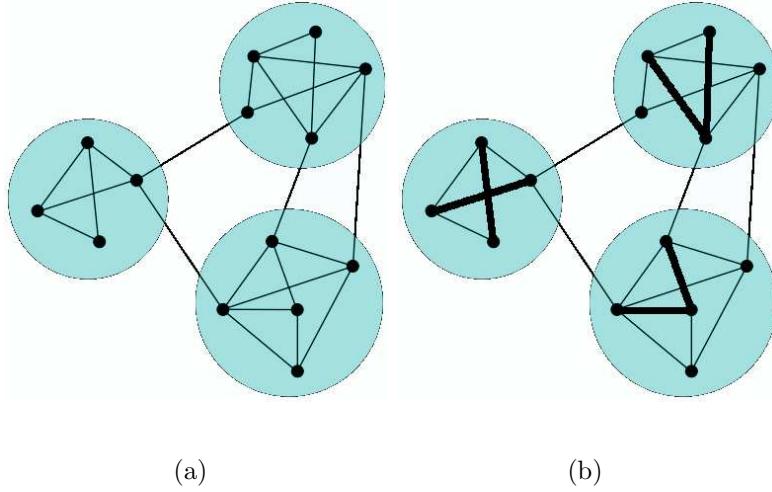


Figure 1.1: The two panels show a small network with 3 natural communities. Panel(a) depicts an unweighted version, and panel(b) shows the weighted network. In general, the edges could also be assigned in a particular direction between nodes. Regardless of the approach used, the goal in community detection is to identify any such strongly related clusters of nodes based on their defined edge relationships. In panel (a), lines depict links corresponding to complimentary or attractive relationships where $a_{ij} > 0$ and $b_{ij} = 0$ in Eq. (??). In panel (b), the thin lines depict missing, adversarial, or repulsive relationships where $a_{ij} = 0$ and $b_{ij} > 0$. The relative link weight is indicated by the respective line thickness. For presentation purpose, missing intercommunity relationships are not depicted.

approximation with “belief-propagation” [19], and other measures such as “fitness” [3]. Other approaches to community detection include clique percolation [20], spectral [21], “label propagation” [22], dynamical [23], and maximum likelihood [24] method. Some efforts enhance or expand applications to more general systems such as weighted networks [25], heterogeneous systems [18], and multiresolution methods [26, 27].

Our approach is a physics-inspired method that casts community detection as a Potts model spin glass. Individual communities correspond to different Potts spin states, and the associated system energy indicates the quality of a candidate partition. Our particular model was originally inspired by a minimal cut method by Djidjev [28] which is equivalent to modularity. The resulting generalized Hamiltonian was previously presented by RB [17]. In their specific implementation, RB generalized null model based approaches to community detection, including modularity as special case, and elaborated on the connection between physics and community detection.

The characterization of which algorithms are better to detect the community structure is still an open question. Commonly used computer-generated benchmarks start with a network of well-defined communities. Then, this structure is degraded by rewiring or removing links and it gets harder and harder for the algorithms to detect the original partition. At the end, the network reaches a point where it is essentially random. This kind of benchmark may be called “open”. The performance on these benchmarks is evaluated by measures such as normalized mutual information or variation of information [29]. These measures compare the solution obtained by an algorithm with the original community structure by evaluating the similarity of

both partitions.

When tested on the commonly used benchmarks and compared with other well-known algorithms, our community detection algorithm proved to be strong, exceptionally accurate, solvable for the community detection problem, and be able to address two particular challenges in the field. First, certain very popular models have been shown to have an implicit limitation in the smallest communities that can be resolved in a large system, the so-called “resolution limit” problem [10]. Second, a more recent focus in the field is the question of how to determine the “best” scale(s) at which to solve a system.

Our approach corrects known resolution limit problems encountered in some models by avoiding a null model comparison [27]. Instead, it penalizes for missing edges directly in the energy sum. In effect, a community is defined by its edge density as opposed to allowing each graph to independently define a community through the use of a relative null model. One consequence of our approach is that it removes the ability of the model to automatically scale the solution based on global properties of graph, but the change results in a robust model with significant improvements to several desirable properties. Further, the multiresolution algorithm used in this thesis quantitatively determines the “best” network scales by evaluating the strength of correlations among independent partitions (“replicas”) of the same graph over a range of resolutions.

1.2 Main results of this thesis

My research focused on two aspects. Most of the new results concern:

- (a) Phase transitions in the community detection problem.

For complex graphs with an arbitrary number of clusters, we found the existence of the easy-hard-unsolvable phases as the system noise increases. This is in analogy to the SAT-hard-unSAT phases in k-SAT problem [Science 297,812 (2002)]. We further discovered that both the low and high temperature hard phases were of the spin-glass type. In addition to random systems, we also investigated for other systems. The transitions became sharper in the thermodynamic limit in all systems.

- (b) Image segmentation and application in medical images.

We applied the community detection algorithm and the corresponding phase transitions to the image segmentation problem. The optimal segmentations were obtained by choosing parameters at the easy phase of the Potts model. The determination of the phase diagram in the analysis of the image segmentation was proved to be efficient.

In below sections, I will briefly elaborate on these.

1.3 Phase transition

One of the highly significant recent applications of statistical mechanics concerns a topic of great broad interest. This arena concerns the developing challenging quest of studying computational problems in various arenas. Bringing tools from physics into the fore has enabled very significant advances in the design of new algorithms and

the identification and understanding of various “phases” of problems in way that has dramatically advanced older computer science related approaches.

Phase transition effects appear in many computational problems, and they constitute one of the most important applications of statistical methods to these problems. In these applications, a phase transition is defined as a situation where small changes in local behavior will significantly change the overall algorithm performance (accuracy or computational cost). Examples include the k-Satisfiability(k-SAT) problem [30, 31, 32] (panel (a) of Fig. 1.2), search problems in artificial intelligence [33] (panel (b) of Fig. 1.2), Steiner trees [34], random vertex-covers [35], Hamiltonian circuits [36], graph coloring [36], image restoration and error-correction [37], graph coloring [38] (panel (c) of Fig. 1.2) and others (examples shown in Fig. 1.2).

As shown in panel (a) of Fig. 1.2, the transition between SAT(satisfiable) to UNSAT(unsolvable) phases of the 3-SAT problem occurs at $\alpha = \alpha_c \sim 4.256$ (wherein $\alpha = M/N$ is defined as the ratio of the number of logical clauses M to the number of variables N). In the k-SAT there are three phases (solvable, hard (typically easy to check in polynomial time proposed solutions but it is very hard to find solutions), and unsolvable phases). As the k-SAT problem is NP complete, the same three phases may appear in all NP problems. Same three phases also appear in our problem (will show in more details in Chapters 2 and 4). In panel (b) of Fig. 1.2, the search cost c to find the first solution, if any, or determine there are none for 3-coloring a randomly generated set of connected graphs with 10 nodes, versus the number of edges e in the graph. A solution exists for those cases with 19 or fewer edges. Thus the peak in the

search cost occurs at the point where the solutions just disappear. Panel (c) shows the number of distinct optimal colorings of each graph in H_1, \dots, H_N for $n = 10$ (H_N is obtained by stripping the graph G_N of its isolated nodes). The number of distinct optimal colorings increases dramatically at each step in the chromatic number, and after that decreases more or less steadily until immediately before the next step. Panel (d) of Fig. 1.2 shows the phase transition of the susceptibility χ_n (defined as $\chi_n = I_N(t = n) - I_N(t = 4)$) versus Z_{out} of our absolute Potts model. Additional trials are unnecessary in the easy region. The benefit of extra trials is largest in the short transition region. Afterward, the benefit diminishes into the hard region, where the accuracy improvement is small even with a large number of attempted optimization trials.

Thus far, the broad study of phase transition has focused on satisfiability problems [30, 31, 32] and related problems- many of which are theoretically important [39] but not always of great practical interest. Insofar as the classification of problems was concerned, the main tools of analysis to date of the phase diagrams of problems were of a static nature and often focused on analogies to glass and spin-glass type physics.

As alluded to above, community detection is a very practical problem of growing interest. This problem has incarnations in physics and many other disciplines (e.g., biology [6], computer science, medicine [10], and economics to name a few). In this thesis [40], we provide strong evidence of a physical phase transition in this problem. We do so by employing two complimentary approaches. These approaches include (1) a spin-glass-type transition by a statistical mechanics analysis of Hamiltonian

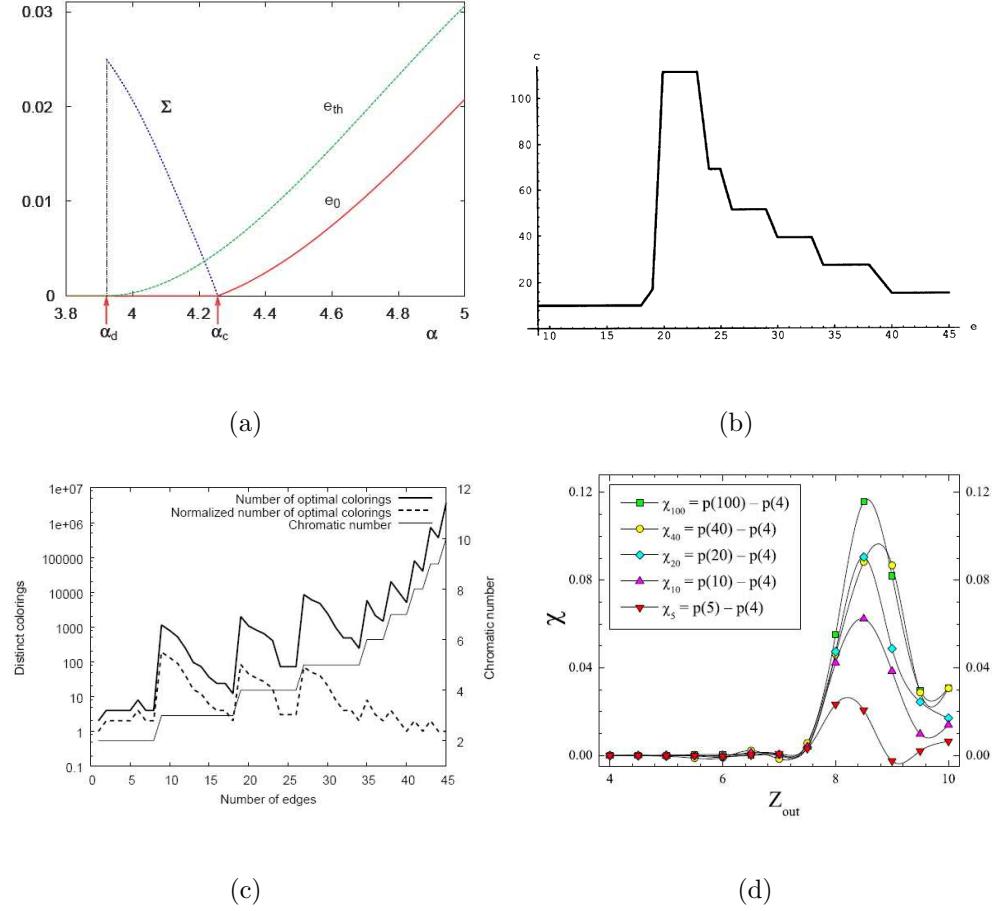


Figure 1.2: Panel(a) [30], the phase diagram of the random 3sat problem. Panel(b) [33], phase transition of the search cost c to find the first solution. Panel(c) [38], the phase transitions of the number of distinct optimal colorings of each graph in H_1, \dots, H_N for $n = 10$. Panel(d) [27], the phase transition of the susceptibility χ_n (defined as $\chi_n = I_N(t = n) - I_N(t = 4)$) versus Z_{out} of our absolute Potts model.

as a function of graph parameters and temperature. (2) a dynamical transition by node trajectories as a function of time. We locate the hard-to-solve region in the above phase transition process and memory effects have been shown in this region. Our results constitute a new perspective for a singular transition from a typical-easy to a rare-hard region in the community detection problem. As far as we are aware no earlier study has focused neither on this important problem’s phase diagram nor on the relation (nor even the physical use) of physical dynamics to the spin-glass transition in computational problems such as this.

The static phase transition is studied via the system energy, “noise” (density of intercommunity edges), time and temperature curves, and also the time-correlation function. For a particular system, the energy as a function of noise p_{out} reaches a peak at some critical value of p_{out} which is a sign of crossing from an easy to a hard solution region much like that which occurs in the k-SAT problem. We determine the critical value of the noise for this transition, and we then study the properties of system in the transition region.

The dynamic transition is analyzed by testing the fluctuations in the node trajectories (community memberships) as a function of time. The node trajectories change from convergent (well-defined memberships) to a more chaotic form (rapidly changing memberships) as more noise is added. The convergent behavior in low noise corresponds to system ergodicity. When we have dissipation (nevertheless on long yet short enough time scales the dynamics may be chaotic as we observe), the chaotic behavior in high noise corresponds to the breakdown of ergodicity. For a fixed sys-

tem, the transition found by dynamic and static facets always correspond to the same transition point which indicates that the phase transition is an inherent property of the community detection problem.

We further systematically examined [41] the phase transitions for the community detection problem via a “noise test” across a range of parameters. The noise test consists of a structured graph with a strongly-defined ground state. We add increasing numbers of extraneous intercommunity edges (noise) and test the performance of a stochastic community detection algorithm in solving for the well-defined ground state. Specifically, we studied two sequences of systems. In the first such sequence, we fixed the ratio $\alpha = q/N$ of the number of communities q to the number of nodes N . We fixed q at different values and varied N in the second sequence of systems. All of these systems showed regions with distinct phase transitions in the large N limit. Deviations occurred most often in smaller systems indicating a definite finite-size effect.

The spin-glass-type phase transitions in our noise test occurred between solvable and unsolvable regions of the community detection problem. A hard, but solvable, region lies at the transition itself where it is difficult, in general, for any community detection algorithm to obtain the correct solution. We analyzed a system of non-interacting cliques and illustrated that in the large q limit, the system experiences a thermal disorder in the thermodynamic limit for any non-zero temperature. When in contact with a heat bath, the asymptotic behavior of the temperatures beyond which the system is permanently disordered varies slowly with the number of communities

q , specifically, $T_x \approx [1/\log q]$. This implies that problems of practical size maintain a definite region of solvability.

We further studied the free energy of arbitrary graphs arriving at the same conclusion. Increasing the number of communities q emulates increasing T in arbitrary graphs for a general Potts model. The effective interaction strength for increasing q scales such that this disorder is circumvented by the often standard use of a simulated annealing algorithm, but the “glassy” (high noise) region remains a challenge for any community detection algorithm.

1.4 Image segmentation

In computer vision, segmentation is the process of partitioning a digital image into multiple segments [42, 43, 44]. The goal of segmentation is to simplify and change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics.

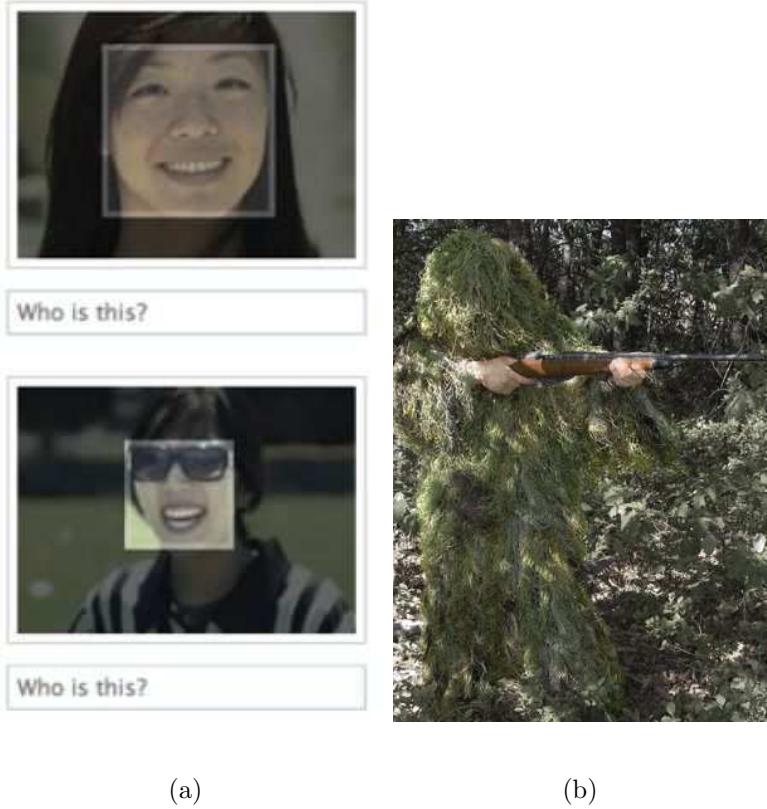
The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image. Each of the pixels in a region are similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect

to the same characteristics. When applied to a stack of images, typical in medical imaging, the resulting contours after image segmentation can be used to create 3D reconstructions with the help of interpolation algorithms like Marching cubes [45].

There are plenty of practical applications of image segmentation. One is medical imaging [46]—includes locate tumors and other pathologies, measure tissue volumes, computer-guided surgery, diagnosis, treatment planning and study of anatomical structure, Other applications include locating objects in satellite images, face recognition [47], also see panel (a) of Fig. 1.3, fingerprint recognition [48], traffic control systems [49], brake light detection, machine vision and agricultural imaging-crop disease detection [50].

Several general purpose algorithm and techniques have been developed for image segmentation. Since there is no general solution to the image segmentation problem, these techniques often have to be combined with domain knowledge in order to effectively solve an image segmentation problem for a problem domain. The methods include thresholding [51], clustering methods [52], compression-based methods [53], histogram-based methods [54], edge detection [55], region-growing methods [56], split-and-merge methods [57], partial differential equation-based methods, graph partitioning methods [58], watershed transformation [59], model based segmentation [60], multi-scale segmentation [61], semi-automatic segmentation [62], trainable segmentation, and etc.

In this thesis, we will apply the above mentioned “community detection” algorithm to image segmentation. This method belongs to the graph partitioning category.



(a)

(b)

Figure 1.3: Two applications of the image segmentation. Left (downloaded from <http://www.insidefacebook.com/2010/12/15/facebook-facial-recognition-suggestions/>): the face detection in Facebook. Facebook has a feature that had some face detection built in—if someone appeared in multiple photos in the same album, Facebook would group them together and ask you who that person was, allowing you to tag them in one fell swoop. Facebook is currently improving on that by using face recognition technology to guess who the person is, using your network of friends as a reference. Right (downloaded from <http://www.amazon.com/Rothco-Lightweight-Purpose-2PC-Ghillie/dp/B002PD43LC>): Camouflage detection in military applications. Camouflage has become an essential part of modern military tactics. Thus it is critical for reconnaissance aircrafts to detect them using the techniques of image segmentation.

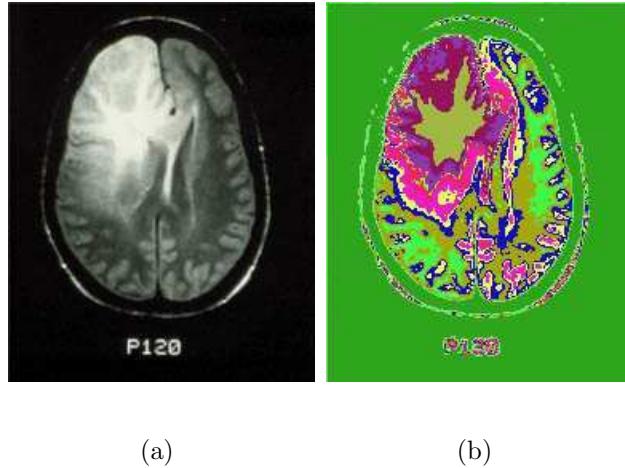


Figure 1.4: Left: the brain image is reproduced with permission from Dr.Edward Aul and the Iowa Neuroradiology Library. Right: the image segmentation result by our community detection algorithm.

Community detection Sec. 1.1 seeks to identify groups of nodes densely connected within their own group (“community”) and more weakly connected to other groups. A solution enables the partition of a large physically interacting system into optimally decoupled communities. The image is then divided into different regions (“communities”) based on a certain criterion, and each resulting region corresponds to an object in the original image.

It is notable that by virtue of its graph theoretical nature, community detection is suited for the study of arbitrary dimensional data. However, unlike general high dimensional graphs, images are two (or three) dimensional. Thus, real images are far simpler than higher dimensional data sets as, e.g., evinced by the four color theorem stating that four colors suffice to mark different neighboring regions in a segmentation

of any two dimensional image. Thus, geometrical (and topological) constraints can be used to further improve the efficiency of the bare graph theoretical method. In [63, 64], in the context of analyzing structures of complex physical systems such as glasses, we used geometry dependent physical potentials to set the graph weights in various two and three dimensional systems. In the case of image segmentation, in the absence of underlying physics, we will invoke geometrical cut-off scales.

In this thesis, we will discuss “unsupervised” image segmentation. By this term, we allude to a general multi-purpose segmentation method based on a general physical intuition. The current method does not take into account initial “training” of the algorithm- i.e., provide the system with known examples in which specific patterns are told to correspond to specific objects. We leave the study of supervised image segmentation and more sophisticated extensions of our inference procedure to a future work. One possible avenue which can be explored is the use of inference beyond that relating to different “replicas” in the simple form discussed in this manuscript that is built on prior knowledge (and prior probabilities in a Bayesian type analysis) of expected patterns in the studied images.

We will heavily employ multi-resolution community detection Sec. 4.5.1. Specifically, we will apply of a particular multi-resolution scheme of the type introduced in [27] to investigate the overall structure at different resolutions in the test images. Similar to [27], we will employ information based measures (e.g., the normalized mutual information and the variation of information) to determine the significant structures at which the “replicas” (independent solutions of the same community detection algo-

rithm) are strongly correlated. With the aid of these information theory correlations, we illustrate how we may discern structures at different pertinent resolutions (or spatial scales). An image may be segmented at different levels of detail and scales by setting the resolution parameters to these pertinent values. We demonstrate in a detailed study of various test cases, how our method works in practice and the resulting high accuracy of our image segmentation method.

We further work on the medical image segmentation Fig. 1.4 in this thesis. Medical imaging is the technique and process used to create images of the human body for clinical purposes or medical science. As a discipline and its widest sense, it is part of biological imaging and incorporates radiology, nuclear medicine, investigative radiological sciences, endoscopy, thermography, medical photography and microscopy.

Image segmentation plays a crucial role in many medical imaging applications by detecting anatomical structures and other regions of interest. Examples of the medical image segmentation methods are listed in [65, 66], include the thresholding approaches, region growing approaches, clustering approaches, Markov random field models, artificial neural networks, deformable models, atlas-guided approaches, clustering, graph partitioning methods and normalized cuts and mixed Gaussian distribution method [67, 68].

We will examine the medical images towards the end of this thesis. These medical images, represented by the fluorescence lifetime of molecular probes injected in liver tissue, will be segmented into different “communities” based on a certain criterion, and each resulting region corresponds to a particular decay trace of photons in the

original image.

We will work on both two and three dimensional data. The two dimensional data refers to the scale value of the fluorescence lifetime of molecule probes injected in liver tissue in individual pixel. The three dimensional data corresponds to the vector form of a decay trace of photons in each pixel. In other words, the two dimensional data represents a static planar image and the three dimensional data corresponds to the evolution of planar images with time. For the two dimensional data, we calculate the intensity (dis)similarity for each pixel pair as the edge weight. For three dimensional data, we apply the Gaussian distribution to fit the vector curve of each pixel, and calculate the mean and variance of the distribution. We will use the combination of the mean and variance value difference for each pixel pair as the edge weight. The replica correlation method will also be employed to the image segmentation of the three dimensional data.

1.5 Overview of thesis

This dissertation contains information related to the following publications or manuscripts in preparation roughly arranged into chapter divisions as indicated:

- Chapters 2: Dandan Hu, Peter Ronhovde and Zohar Nussinov, *Phase transitions in random Potts systems and the community detection problem: spin-glass type and dynamic perspectives*, Philosophical Magazine 92, 4, (2012).
- Chapter 3: Dandan Hu, Peter Ronhovde and Zohar Nussinov, *The stability to in-*

stability transition in the structure of large scale networks, <http://arxiv.org/abs/1204.4167>,
also submitted to Phys. Rev. E (2012).

- Chapter 4: Dandan Hu, Peter Ronhovde and Zohar Nussinov, *Replica inference approach to unsupervised multiscale image segmentation*, Phys. Rev. E 85, 016101, (2012).
- Chapter 5: Dandan Hu, Pinaki Sarder, Peter Ronhovde, Samuel Achilefu and Zohar Nussinov, *Segmenting interacellular distribution images of near-infrared fluorescent dye derived from fluorescence lifetime imaging microscopy*, in preparation (2012).

Chapter 2

Phase transition in community detection problem

2.1 Introduction

One of the highly significant recent applications of statistical mechanics concerns a topic of broad interest— that of community detection[6, 11, 15, 17, 18, 19, 22, 27, 69, 70, 71, 72, 73] in complex networks [11, 72, 74] and related computational problems [34, 35, 37]. In this chapter, we address further development in the challenging quest of studying these difficult computational problems by bringing additional tools from physics into the fore. Our aim is not only to study the community detection problem itself. Rather, we use the community detection problem as a platform for a detailed investigation of phase transitions [30, 31, 32, 33] associated with complex computational problems and, generally, Potts spin glass type systems. Various applications

of physics to computational problems have enabled significant advances in the design of new algorithms and the identification and understanding of various “*phases*” of computational problems in way that has dramatically advanced previous approaches.

In this chapter, we provide direct evidence for earlier indications of phase transitions in the community detection problem and more generally in Potts type spin glass systems. These Potts type spin glass transitions may occur at *both* low and high temperatures (or, similarly, at low and high levels of randomness or noise). These transitions reflect different underlying physics. Earlier reports of such transitions were afforded by information theory measures (as in Appendix E of [18]) and a computational “computational susceptibility” to be defined in later sections of the current work that monitors the onset of a large number of local minima, or large computational complexity (as in Appendix B of [27]). As was earlier shown (e.g., Fig. 11 in [27]), overlap parameters (to be defined herein) such as the normalized mutual information I_N exhibit progressively sharper changes as the system size N increases. This suggests the existence of bona fide thermodynamic transitions. In this chapter, we will investigate “fixed” spin glass type Potts Hamiltonian. By “fixed”, we allude to new spin glass systems with fixed parameters which are not dependent on the problem itself. Thus, this fixed approach contrasts with, e.g., “modularity” [6, 11, 75] or other models that involved comparisons to random case systems- so called “null models” [6, 11, 17] that have been earlier invoked on the community detection problem. When cast in terms of canonical fixed Potts spin Hamiltonian, the system exhibits sharper phase transitions [27]. By applying our model to a general random graph, we can lo-

cate phase transitions between *solvable* and *unsolvable* regions. Solvable regions may further splinter into “*easy*” and “*hard*” phases. We further elaborate on disparate phase transitions (at low and high temperatures) in these rather general Potts spin glass type systems. It is noteworthy that a similar analysis can be done for any other method for detecting communities. Within most of the easy phase, all of the known methods agree on the solutions. The results of our analysis are not relevant to only one specific method.

Insofar as the classification of computational problems is concerned, the main tools of analysis to date were of a static nature and further invoke various forms of “cavity” type approximations [76, 77] and extremely powerful related approaches such as “belief propagation” [19, 78, 79]. Cavity type methods were of immense success early on in studying mean-field type theories in spin-glasses and, in the last decade, have seen a rapid resurgence in enabling new and very potent algorithms and in better enabling an understanding of complex problems.

In this chapter, we directly study the phase transition in computational problems such as community detection from both static (i.e., thermodynamic) and dynamic aspects. We directly numerically investigate, sans any analytical approximations, thermodynamic quantities characterizing the transition augmented by further direct measures of the energy landscape of these systems by use of a “computational susceptibility” that we will introduce later on that monitors the increase number of local minima and convergence with local minima. In the dynamic approach, in order to relate hard computational problems to classical dynamics, we will dualize, via a

Hubbard-Stratonovich transformation, the original (discrete) system to be optimized by a continuous theory for which equations of motion can be written down and the dynamics investigated. By employing these two complimentary approaches ((i) static thermodynamic of information measures of the energy landscape and (ii) classical dynamics), we correspondingly report on the existence of (i) static spin-glass-type transitions as well as (ii) dynamical transitions, i.e., the transition of nodes from stable orbits to “chaos”. The transitions as ascertained by both approaches occur at precisely the same set of parameters describing the problem. As far as we are aware, earlier studies have not investigated the general phase diagram of this important problem. To date, links between dynamical mechanical transitions and spin-glass type transitions in computational problems such as this have, furthermore, not been discovered.

2.2 The Potts model

We will employ a, rather general, spin-glass type Potts model Hamiltonian (denoted, henceforth, as the “Absolute Potts Model” (APM)) [18] for solving the community detection problem. The Hamiltonian reads

$$H(\sigma) = -\frac{1}{2} \sum_{i \neq j} (A_{ij} - \gamma(1 - A_{ij}))\delta(\sigma_i, \sigma_j). \quad (2.1)$$

Here, A_{ij} is an adjacency matrix element which assumes a value of 1 if nodes i and j are connected and a value of 0 otherwise. The spins $\{\sigma_i\}_{i=1}^N$ attain integer values: $1 \leq \sigma_i \leq q$. Their values reflect the community membership. That is, if $\sigma_i = a$ then

node i belongs to community number a . The parameter q denotes the total number of communities. To simplify the analysis, we will, unless stated otherwise, set (the so-called resolution parameter [27]) $\gamma = 1$. (In our other work [80], we reported on similar results for general γ and a weighted version of Eq. (2.1). In particular, in physics related applications for many particle systems, the weights A_{ij} were determined by the two-body interactions [63, 64].)

Although, as we elaborate on in Sec. 2.7, we can achieve analytic solutions for certain cases of graphs (e.g., employing the cavity method [39, 78, 81, 82, 83] when all of the nodes are of a fixed degree of $k = 3$ or Ising systems (i.e., systems with $q = 2$ communities)), most *general graphs* (with arbitrary degree and cluster size distributions) require computer simulation. To this end, we will undertake a direct numerical investigation of the system at hand without the need to invoke analytical approximations or assumptions. Our (“zero-temperature”) community detection algorithm for minimizing Eq. (2.1) was discussed at length in Refs. [18, 27, 63, 64, 80]. In the current work, we investigate the above Hamiltonian of Eq. (2.1) at zero temperature [18] and also at finite temperatures ($T > 0$) with the use of a heat bath algorithm (HBA). In brief, within the HBA, we will sequentially allow each node an opportunity to change the community membership during each time step with probabilities determined by a Boltzmann weight $e^{-\beta\Delta E}$ ($\beta = \frac{1}{T}$) at a specified temperature T and the energy change (ΔE) as the node were moved to each connected cluster (or to a new cluster). Similarly, as elaborated on in more detail, following each step, we further allow the possibility of community merges based on a Boltzmann weight.

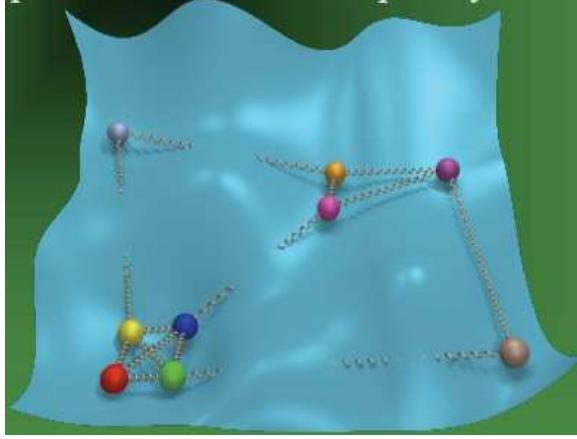


Figure 2.1: A caricature of the information theory correlations (springs) between “replicas” (denoted symbolically by balls) in a high dimensional energy landscape (in our case, graph partitions or Potts spin configurations). “Replicas” are obtained from multiple solutions of the same problem (in this case, the minimization of the Potts model Hamiltonian of Eq. (2.1)). The information theory correlations measure the agreement or overlap between the candidate solutions (“replicas”). In earlier works and in the current work, we use such correlations to ascertain system parameters (e.g., γ of Eq. (2.1)) for which clearly defined solutions appear. Throughout most of the current work, we will not employ inter-replica correlations but rather the average of the correlations between all of the replicas and a known (or “planted”) solution to the community detection problem (a minimum of the Hamiltonian). For detailed definitions of replicas and information theory correlations, see Sec. 2.3 and Sec. 2.5, respectively.

2.3 Definitions: Trials and Replicas

Before turning to the specifics of our results, we need to introduce several basic notions. We start by discussing two concepts which underlie our approach. Both

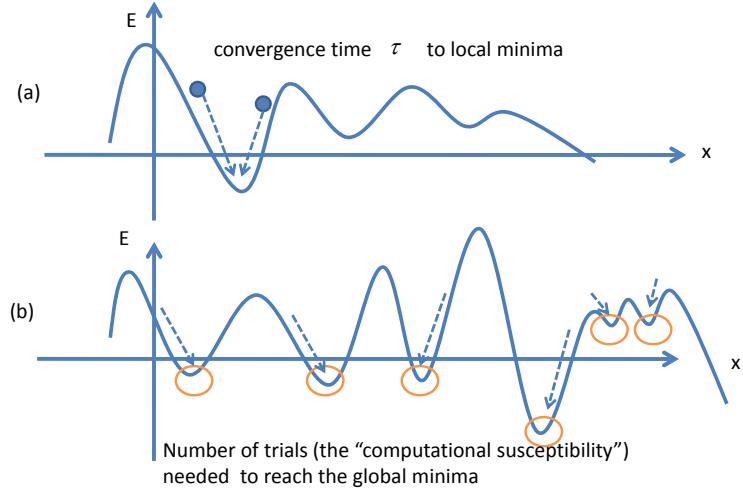


Figure 2.2: A schematic of the physical content of parameters that we employ: (a) The convergence time τ is the number of steps the algorithm needed to reach local minima, (b) When the energy landscape becomes complex, more “trials” are needed in order to veer towards the global minimum (or minima). This requisite number of trials relates to the “computational susceptibility” χ of Eq.(2.8) that as will be explained later records the improvement in the quality of the solutions (as seen by the normalized mutual information I_N) as the number of trials s (different trajectories in panel (b)) is increased.

concepts pertain to the use of multiple identical copies of the same system which differ from one another by a permutation of the site indices. In the definitions of “*trials*” and “*replicas*” given below, we build on the existence of a given algorithm (any algorithm) that may minimize a given energy or cost function. In our particular case, we minimize the Hamiltonian of Eq. (2.1). However, these ideas and concepts are more general.

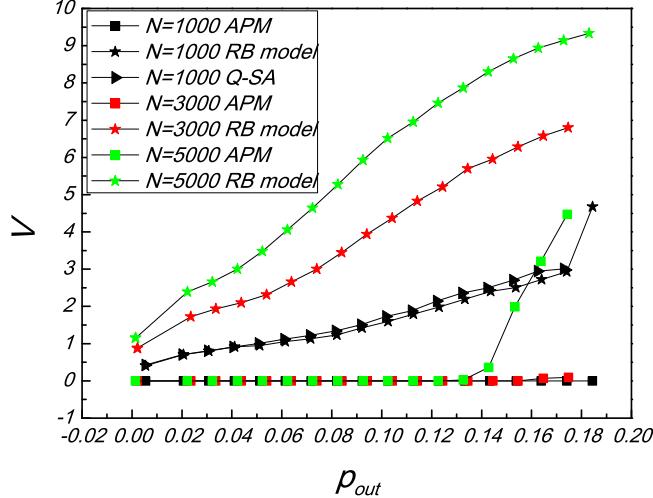


Figure 2.3: The average variation of information V of the “noise” p_{out} (the density of links connecting different communities). V is calculated between the proposed solution and the embedded constructed sample graph whose solution is known. (The graph has a power-law distribution of community sizes with a minimum $n_{min} = 8$, maximal $n_{max} = 40$, and with the exponent determining the community size distribution set equal to -1). We show results obtained by using our absolute Potts model (denoted as “APM” in Eq. (2.1)). For comparison, we also plot the results determined by “RB Potts” [17, 18] model and modularity optimization (“Q-SA”) [6] using simulated annealing. With the “APM”, our algorithm demonstrates extremely high accuracy for the small and large systems shown above.

- Trials. We use trials alone in our bare community detection algorithm [18, 27]. We run the algorithm on the same problem “ s ” independent times. This may generally lead to different contending states that minimize Eq. (2.1). Out of these s trials, we

will pick the lowest energy state and use that state as the solution. In the current work, $4 \leq s \leq 20$. We will canonically employ $s = 4$ trials. We will use $s > 4$ trials in the calculation of the computational susceptibility of Eq.(2.8).

- Replicas. Each sequence of the above described s trials is termed a *replica* (see the schematic plot Fig. 2.1 of *replicas*). When using “replicas” in the current context, we run the aforementioned s trials (and pick the lowest solution) “ r ” independent times. By examining information theory correlations between the “ r ” *replicas* and the known (or “planted”) solution, we can assess the quality of candidate solutions. In this work, we set $r = 100$.

In this chapter, we will briefly remark on the determination of optimal parameters of the system. To this end, we will compute the average inter-replica information theory correlations within the ensemble of r replicas. Specifically, *information theory extrema* as a function of the scale parameters, generally correspond to more pertinent solutions that are locally stable to a continuous change of scale. It is in this way that we will detect the important physical scales and parameters in the system.

In this chapter, we will compute the *average* information measures between the disparate candidate solutions found by different “*replicas*” and the known (or “planted”) solution to the problem which we label below as “ K ”. In general, with A denoting graph partitions in different “*replicas*” and $Q(A, K)$ denoting the information theory overlap between replica A and the known solution K , the average for a general

quantity Q that we will employ are, rather explicitly,

$$\langle Q \rangle = \frac{1}{r} \sum_A Q(A, K). \quad (2.2)$$

In earlier works [18, 63, 64, 80] we employed the average inter-replica information theory overlaps. We will invoke this method once when discussing the optimal value of the resolution parameter γ of Eq. (2.1). Apart from that single case, we will generally not use these average inter-replica measures here but rather their comparison to a known solution K .

In the context of the Potts model Hamiltonian of Eq. (2.1), by “replicas”, we allude [27] to systems that initially constitute identical copies of the system that differ only by a permutation the Potts spin label. Different replicas will, generally, lead to disparate final contending solutions. By the use of an *ensemble* of such replicas, we can attain accurate result and determine information theory correlations between candidate solutions and infer from these a detailed picture of the system.

2.4 Heat Bath Algorithm

We extend the zero-temperature (greedy) algorithm of [18, 27] to finite temperature via a heat bath algorithm. This algorithm allows each node to become a member of one community with probability set by a thermal distribution [6]. The probability is

$$p_{a \rightarrow b} = \frac{\exp(-\Delta E_{a \rightarrow b}/T)}{\sum_d \exp(-\Delta E_{a \rightarrow d}/T)}. \quad (2.3)$$

Here $\Delta E_{a \rightarrow b}$ is the change of energy for moving this node from cluster a to cluster b , and d runs through all connected clusters (neighbors) of this node (including the

case that $d = a$, i.e., this node remains in cluster a ; and the case that d is a newly added cluster, i.e., this node becomes a new sole-node cluster).

The steps of our heat bath algorithm are as follows:

(1) *Initialize the system.* Symmetrically initialize the system by assigning each node to its own community. (i.e., $q_0 = N$). If the number of communities is constrained to some value q , we instead randomly initialize the system into q communities.

(2) *Find the best cluster for node i .* Sequentially “pick up” each node and scan its neighbor list (include its current cluster and the newly added cluster). Calculate the energy change as if it were moved to each connected cluster. Then calculate the probability for an arbitrary node in cluster i to be moved to a connected cluster b using Eq. (2.3). Then we use all the probabilities for different j ’s to determine which cluster to be moved to; i.e., generate a random number between 0 and 1, then determine which probability range the random number is in, and move the node from cluster a to the selected cluster b .

(3) *Repeat step 2 for all nodes in the system.* A node is frozen for the current iteration once it has been considered for a move.

(4) *Merge clusters.* Allow for the merger of two communities together based on the merge probability. Towards this end, we calculate the energy change as if the current community is merged with its neighbors. We then use Eq. (2.3) to calculate merge probabilities.

(5) *Repeat the above two steps.* Repeat step 2 to 4 until the maximum number of

iterations is reached.

(6) *Repeat all the above steps for s trials.* Repeat step 1-5 for s trials and select the lowest energy result as the best solution. Each trial randomly permutes the order of nodes in the symmetric initial state.

The new algorithm is similar to the earlier greedy algorithm [18, 27] except for steps (2) and (4). The nodes are moved based on a random process. Thus, the outcome may be sometimes sensitive to the initial random seed state. As noted within the main text, when the system is within the easy phase, all seeds lead to the same final outcome. However, when the system is within the hard phase changing the random seed may significantly alter the final result. In such a case, different initial conditions enable the system to get stuck in different local minima (each corresponding to a different partition of the system into disparate communities). This is why we repeat the procedures 1 – 5 for s trials (usually s is set to be 4). The additional trials sample different solutions evenly with the symmetric initialization, and it will reduce the dependence on initial conditions. In the unsolvable phase, for any finite number of trials s , the quality of the solutions does not visibly change.

We should note that the new “heat bath algorithm” that we introduced above is *different from the commonly used “simulated annealing algorithm”*. (The latter is a generalization of the “Metropolis Monte Carlo” procedure (MMC) [84]).

Within the conventional MMC procedure, the probability for an arbitrary node to be moved in cluster i to a connected cluster j is given by $\min(1, \exp(-\beta(E_b - E_a)))$. This implies that a node i in community a will (with certainty) be moved to cluster

b if the energy change is negative. Such an algorithm precludes for a lower energy move (if such a later move will be found later on). By contrast, within our “heat bath algorithm”, nodes are not immediately moved to the first tried clusters if the energy change is negative. We compute the probabilities of connected clusters. Obviously, the cluster with the largest energy decrease would have the largest probability to be the “candidate of absorption” for node i . Thus, in contrasting the commonly used MMC procedure and our HBA, it seems be easier to get to the lowest energy state of the studied system within our algorithm. Our procedure allows nodes to explore more energy states in each step and better equilibrate.

The results obtained at low temperature by our HBA are very close to the results obtained by the zero temperature “greedy algorithms” [18, 27].

2.5 Information theory and complexity measures

In this section, we introduce and review information theory measures (see the schematic plot Fig. 2.1 depicting the information theory correlations) (as they pertain to the community detection problem) that we will employ in our analysis.

- *Shannon Entropy.* If there are q communities in a partition A , then the Shannon entropy is

$$H_A = - \sum_{a=1}^q \frac{n_a}{N} \log_2 \frac{n_a}{N}. \quad (2.4)$$

The ratio $\frac{n_a}{N}$ is the probability for a randomly selected node to be in a community

a with n_a the number of nodes in community a and N the total number of nodes.

With the aid of this probability distribution the Shannon entropy of Eq. (2.4) follows.

- *The mutual information.* The mutual information $I(A, B)$ between candidate partitions (A and B) that are found by two replicas is

$$I(A, B) = \sum_{a=1}^{q_A} \sum_{b=1}^{q_B} \frac{n_{ab}}{N} \log_2 \frac{n_{ab}N}{n_a n_b}. \quad (2.5)$$

Here, n_{ab} is the number of nodes of community a of partition A that are shared with community b of partition B , q_A/q_B is the number of communities in partition A (or B), and (as earlier) n_a (or n_b) is the number of nodes in community a (or b).

- *The variation of information.*

The variation of information $0 \leq V(A, B) \leq \log_2 N$ between two partitions A and B is given by

$$V(A, B) = H_A + H_B - 2I(A, B). \quad (2.6)$$

- *The normalized mutual information.* The normalized mutual information $0 \leq I_N(A, B) \leq 1$ is

$$I_N(A, B) = \frac{2I(A, B)}{H_A + H_B}. \quad (2.7)$$

High I_N and low V values generally indicate high agreement between different the partitions (or general Potts spin configurations) A and B .

The physical significance of two of the following concepts is sketched in Fig. 2.2.

- *The convergence time.* The convergence time τ is the number of the algorithm steps needed to find the local minimum following a greedy algorithm. As just noted above, a schematic plot explaining the physical meaning of the convergence time τ is shown in Fig. 2.2.
- *The complexity.* The complexity customarily denoted as $\Sigma(e)$, can be derived from the number of states $\mathcal{N}(E)$ with energy E . Specifically, $\mathcal{N}(E) \sim \exp[N\Sigma(e)]$, [30] where the energy density $e = E/N$. In this chapter, we will numerically determine the onset of the high complexity (which probes the number of local minima) without any prior assumptions or approximations by directly computing the “computational susceptibility” ([27]) that we will briefly define next.
- *The “computational susceptibility”.*

A “computational susceptibility” monitoring the onset of high complexity can be defined as:

$$\chi_n = I_N(s=n) - I_N(s=4). \quad (2.8)$$

That is, χ is the increase in the normalized mutual information I_N as the number of trials (number of initial starting points in the energy landscape) $s = n$ is increased. Physically, we ask how many different initial starting points in the energy landscape (i.e., how many different initial “trials”) are required to achieve a certain desired threshold accuracy as measured by information theory measures.

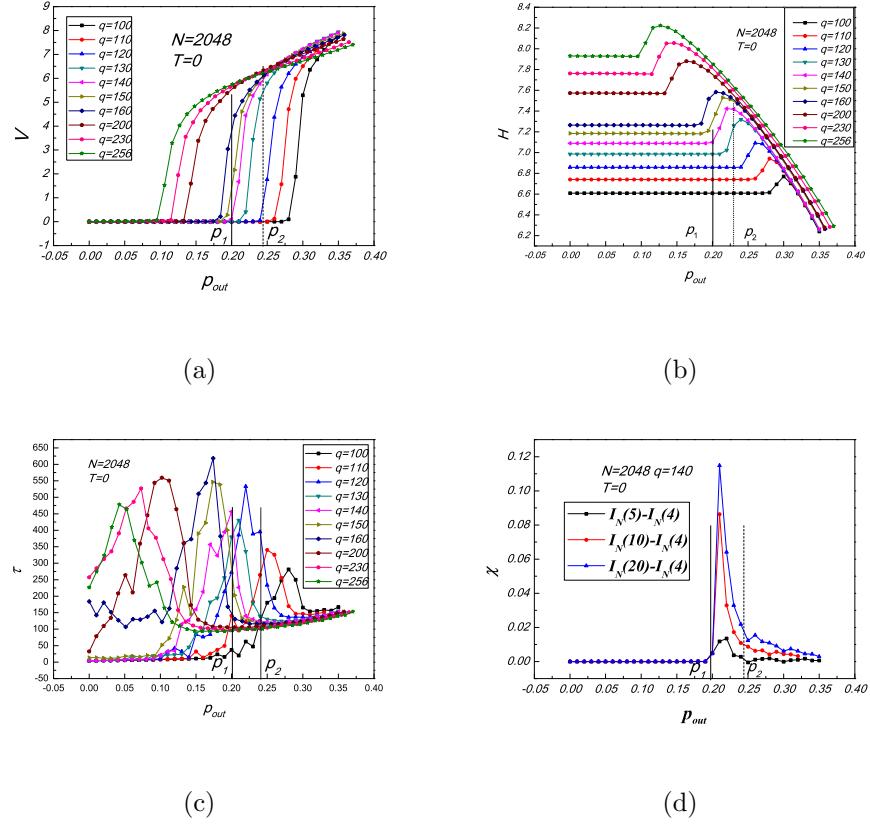


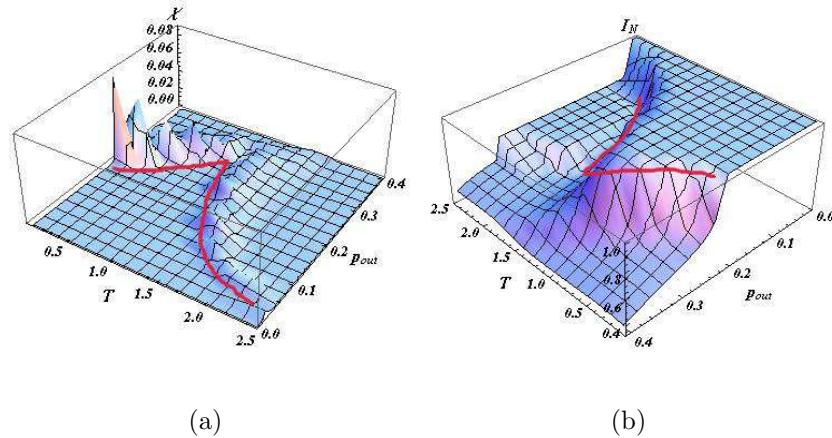
Figure 2.4: Plots of various measures as a function of the noise level p_{out} . V is the variation of information. H is the shannon entropy ([27]). τ is the number of steps needed to reach local low energy state (see also Fig. 2.6). The “computational susceptibility” χ is defined in Eq. (2.8). In the examined system of $N = 2048$ nodes with $q = 140$ communities, all of the plots show three phases as noise varies. (1) Below a noise threshold value of $p_1 = 0.2$, the system can be “easily” solved. (2) When $0.2 < p_{out} < 0.24$, the benefit of extra trials is most significant (shown in (c)) and it is “hard” to solve the system. (3) Above noise levels about $p_2 = 0.24$, the system cannot be perfectly solved. As we will outline, the two transitions at $p_{out} = p_1, p_2$ are both of the spin-glass type.

2.6 The construction of random graphs and noise test

Similar to [85], we will use a “noise test” benchmark as a workhorse to study phase transitions in random graphs [18].

We define the system “noise” in community detection problem as edges that connect a given node to communities other than its original community assignment (“inter-community” edges). In general [18], we cannot initially distinguish between edges contributing to noise and those constituting edges within communities of the best partition(s).

Specifically, for each constructed benchmark graph, we start with N nodes divided into q communities with a power law size distribution (with the exponent determining the community size distribution [85] set equal to (-1) , i.e., the community size n scales as $n^{\bar{\beta}}$ with $\bar{\beta} = -1$). We connect all “intra-community” edges at a high average edge density $p_{in} = 0.95$. Thus, when $p_{out} = 0$ we have decoupled clusters with no



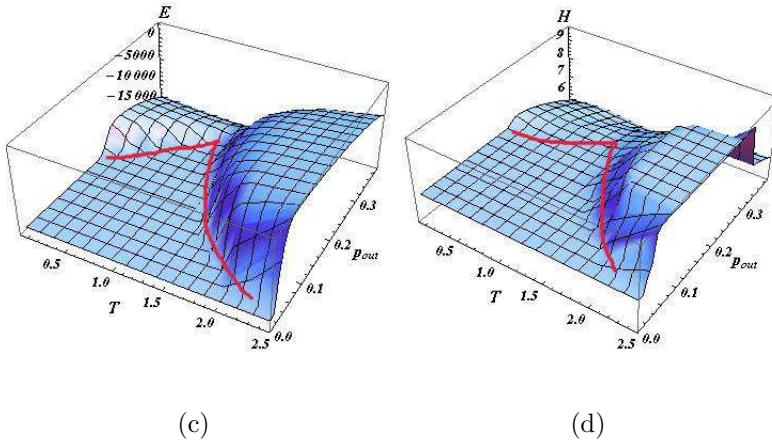


Figure 2.5: The computational susceptibility χ , normalized mutual information I_N , Shannon entropy H and energy E in terms of temperature T and the inter-community noise p_{out} for systems with $N = 2048$ nodes and $q = 140$ communities. All of the plots show three different phases which correspond to the three panels ((a)-(c)) shown in Fig. 2.6, denoted as “hard-easy-hard”. The first “ridge” in the low temperature in panel (a)-(d) (computational susceptibility χ /normalized mutual information I_N /energy E /entropy H) corresponds to the “hard” phase shown in panel (a) in Fig. 2.6. A higher temperature hard phase is also present. A guide to the eye is drawn to emphasize the manifestation of the hard phases in all measured quantities. The middle “flat” region in panels (a)-(d) is the “easy” phase.

inter-community links. We then add *random* “inter-community” edges (“noise”) at a density of $p_{out} < 0.5$. Specifically, p_{in} is defined as the ratio of the existing intra-community edges over the maximal intra-community edges, and p_{out} is defined as the ratio of the existing “inter-community” edges over the maximal inter-community edges. If we denote the *average external degree* for each node by Z_{out} (i.e., the average number of links between a given node to nodes in communities other than its own) and the *average internal degree* by Z_{in} (i.e., the average number of links to nodes in the same community- $Z_{in} + Z_{out} = Z$ with Z the average coordination number), then we define [18]

$$p_{in} = \frac{NZ_{in}}{\sum_{a=1}^q n_a(n_a - 1)}, \quad (2.9)$$

and

$$p_{out} = \frac{NZ_{out}}{\sum_{a=1}^q \sum_{b \neq a} n_a n_b}. \quad (2.10)$$

In the above, as throughout, n_a denotes the number of nodes in community a . When the noise is low (i.e., when p_{out} is small), all the communities are well defined. As more and more external links are progressively added to the system (p_{out} increases), the communities become harder and harder to detect. In some stage, when the external link density is sufficiently high, the system cannot be detected. As alluded to earlier, we investigate the phase transition from the “solvable” to “unsolvable” at

both low and high temperature with the use of the heat bath algorithm (“HBA” in Sec. 2.4) in the following section.

2.7 Theoretical analysis of the community detection problem

In this section, we follow the description of cavity method in [39] and merely generalize it to all graphs (general q and unequal size communities). The uninitiated reader is encouraged to peruse [39] in order to familiarize him/herself with basic the cavity method (and the notations) used that we expand on below. The brief introduction below is not self-contained.

Within the cavity approach, each node passes *a message* along edges. A message from node i to j is a q -dimensional vector of zeros and ones. Node i takes the messages from all the other nodes $k \neq j$ connected to i and sums them. Then the cavity field defined as $\mathbf{h}_{i \rightarrow j} = \sum_{k \neq j} J_{ki} \mathbf{u}_{k \rightarrow i}$ is obtained through the above process. Finally, node i converts this cavity field into a message to j by picking and setting the maximal components in \mathbf{h} to one and the rest to zero. The probability distribution of messages being sent in the system is denoted as $\mathcal{Q}^s(\mathbf{u})$. The superscript s denotes a possible dependence of this distribution on the index of the pre-defined cluster to which the sending node belongs.

To be consistent with the notations in [39], in what follows in this section (and only in this section), we will employ the same definition for p_{in} and p_{out} as that of

[39]. For a fixed cluster A , $p_{in}^A = p(A|A)$ is the conditional probability that a link starting with a node in A also ends in A . Given two (different) clusters A and B , $p_{out}^{B|A} = p(B|A)$ denotes the conditional probability that a link starting with a node in A would end in B . It follows directly from these definitions that,

$$p_{in}^A + \sum_{B \neq A} p_{out}^{B|A} = 1. \quad (2.11)$$

In particular, when $q = 2$, there are only two clusters/states which A, B , and we have $p_{in}^A + p_{out}^{B|A} = 1$.

Following the same calculation process in [39], we also test the phase transition of community detection in a random Bethe lattice with exact degree $k = 3$. But there is an essential difference that our Hamiltonian *does not have the constraint of equal-size clusters*, which means we do not have the symmetric condition for the order parameter $Q^s(\mathbf{u}) = \eta_{c\omega}$, where, $c = 1$ denotes the “correct” component, and $\omega \in \{1 - c, \dots, q - 1\}$ denotes the number label of a “wrong” component. In our case, ω now can not be necessarily written as $\omega = ||\mathbf{u}|| - c$. We denote this, more generally, as $\eta(\mathbf{u})^{state}$.

We first discuss the case of $q = 2$ and then proceed to its generalization.

In systems with two clusters ($q = 2$), there are two (Potts) spin states. We will denote these herein as A and B (once again, we do so to be consistent with the notations in [39], in particular Eqs. (6.60)-(6.63) therein). In this case, there are 6 different “order parameters”. We will denote these as η_{01}^A , η_{10}^A , η_{11}^A , η_{01}^B , η_{10}^B and η_{11}^B .

In the following, we present the expressions for η_{01}^A , η_{10}^A and η_{11}^A . The expressions for η_{01}^B , η_{10}^B and η_{11}^B have an identical form with a permutation of the superscripts

$A \leftrightarrow B$.

$$\begin{aligned} \eta_{11}^A = & (p_{in}^A \eta_{11}^A + p_{out}^{A|B} \eta_{11}^B)^2 \\ & + 2(p_{in}^A \eta_{10}^A + p_{out}^{A|B} \eta_{10}^B)(p_{in}^A \eta_{01}^A + p_{out}^{A|B} \eta_{01}^B), \end{aligned} \quad (2.12)$$

$$\begin{aligned} \eta_{10}^A = & (p_{in}^A \eta_{10}^A + p_{out}^{A|B} \eta_{10}^B)^2 \\ & + 2(p_{in}^A \eta_{10}^A + p_{out}^{A|B} \eta_{10}^B)(p_{in}^A \eta_{11}^A + p_{out}^{A|B} \eta_{11}^B), \end{aligned} \quad (2.13)$$

$$\begin{aligned} \eta_{01}^A = & (p_{in}^A \eta_{01}^A + p_{out}^{A|B} \eta_{01}^B)^2 \\ & + 2(p_{in}^A \eta_{01}^A + p_{out}^{A|B} \eta_{01}^B)(p_{in}^A \eta_{11}^A + p_{out}^{A|B} \eta_{11}^B). \end{aligned} \quad (2.14)$$

These consist a quadratic system of 6 equations with 6 variables. This system of equations is numerically solvable. The solutions are continuous with respect to coefficients $p_{in}^A, p_{out}^{B|A}$. The new equations of Eqs. (2.12,2.13,2.14) form a generalization of the system studied in [39].

The above procedure can also be easily generalized to system with $q > 2$ components leading to more terms on the righthand side of Eqs. (2.12,2.13,2.14). In general, we can define an abstract function g :

$$\begin{aligned} \{0, 1, 2\}^q \setminus \{(0, 0, \dots, 0)\} & \rightarrow \{0, 1\}^q \setminus \{(0, 0, \dots, 0)\} \\ (a_1, a_2, \dots, a_q) & \mapsto \\ \begin{cases} (a_1, a_2, \dots, a_q) & 2 \notin \{a_1, a_2, \dots, a_q\} \\ (\lfloor a_1/2 \rfloor, \lfloor a_2/2 \rfloor, \dots, \lfloor a_q/2 \rfloor) & 2 \in \{a_1, a_2, \dots, a_q\}, \end{cases} \end{aligned}$$

then for any $\mathbf{a} = (a_1, a_2, \dots, a_q) \in \{0, 1\}^q \setminus \{(0, 0, \dots, 0)\}$ and $1 \leq i \leq q$, we have the

equation

$$\eta_{\mathbf{a}}^{A_i} = \sum_{\substack{g(\mathbf{u}+\mathbf{v})=\mathbf{a} \\ \mathbf{u},\mathbf{v} \in \{0,1\}^q \setminus \{(0,0,\dots,0)\}}} (p_{in}^{A_j} \eta_{\mathbf{u}}^{A_i} + \sum_{j \neq i} p_{out}^{A_i|A_j} \eta_{\mathbf{u}}^{A_j}) \times \\ (p_{in}^{A_j} \eta_{\mathbf{v}}^{A_i} + \sum_{j \neq i} p_{out}^{A_i|A_j} \eta_{\mathbf{v}}^{A_j}). \quad (2.15)$$

In the above equation (Eq. (2.15)), \mathbf{a} denotes the q -dimensional incoming message composed of 0 and 1s. We introduce (\mathbf{u},\mathbf{v}) to be any pair of vectors that “sum up to” a given vector \mathbf{a} , in the sense of $g(\mathbf{u} + \mathbf{v}) = \mathbf{a}$. We numerically evaluate the order parameter $\eta(\mathbf{u})^{state}$ as a function of p_{in} .

From this, we can obtain the phase boundaries of the solvable region. Furthermore, to test whether our simulation result matches the theory, we perform the same accuracy test using our greedy algorithm on ER graphs with $\langle k \rangle = 16$ and four equal-sized clusters. Our result of Fig. 12 in [27] in Appendix A is consistent with the cavity inspired result of Fig. 6.7 in [39]. In both plots of the percentage of correctly identified nodes in terms of p_{in}/Z_{out} , the critical value of $p_{in}^{critical}$ and $Z_{out}^{critical}$ for the accuracy drops are the same if we transfer Z_{out} into p_{in} via the relation $p_{in} = \frac{16-Z_{out}}{16}$ (for the graphs considered therein with an average total coordination number per node of $\langle k \rangle = 16$). In, e.g., Fig. 6.7 of Ref. [39], the threshold value of p_{in} is given by $p_{in}^c \approx 45\%$. In Fig. 12 of [27], the critical Z_{out} obtained by our greedy algorithm is $Z_{out} \approx 9$, which corresponds to $p_{in} \approx \frac{16-9}{16} = 43\% \approx 45\%$.

2.8 Spin glass type transitions

2.8.1 Results for information theory correlation and thermodynamic quantities

With all of the preliminaries now in place, we now report our findings. The upshot of the results to be presented is evidence for the existence of *two* spin glass type transitions in general random graphs. Evidence for these transition is afforded by changes in the accuracy of the solution obtained by the “APM” in Eq. (2.1) when noise is introduced. This is shown in Fig. 2.3. The variation of information V between the test system result and the solution displays a phase transition as the noise p_{out} increases. A transition is manifest in the sudden jump of V . The variation of information V remains zero (indicating, essentially, perfect solutions) up to a threshold value of the noise where a very sharp transition is seen. We compared this transition to similar transitions that we detected via more standard methods. These are labeled, in Fig. 2.3, by “Q-opt SA” (maximization of modularity (Q), set by a comparison to a null model, [6] as solved by simulated annealing (SA)) and “RBPM” (the Potts model of [17] wherein the parameters in the Hamiltonian are also defined by a null model). As seen, our “APM” of Eq. (2.1) (which is free [27] of the so-called “resolution limit” [10, 86] that appears in systems with null models) can be used to examine graphs with high levels of noise [18]. By comparison to other models compared to null models, the APM exhibits a sharper transition as the number of nodes N is increased [27].

2.8.2 General features of the phase diagram as ascertained by numerical data

As is evident in Fig. 2.4, there are three different phases. We denote these phases by the qualifiers of (i) “easy”, (ii) “hard”, and (iii) “unsolvable”. These three phases are the analogs of the three phases (i) “SAT”, (ii) “hard”, (iii) “unSAT” in the k-SAT problem [30]. In later discussions, we elaborate on their possible physical significance of these phases in disparate arenas such as that of supercooled liquids. In what follows, we first present our results. We first discuss the zero temperature case ($T = 0$) and then explore the physics at $T > 0$.

$$T = 0$$

In Fig. 2.4, the low noise region ($p_{out} < p_1$) is seen to be in the “easy” phase. In this phase, the accuracy (V), entropy (H), and the computational susceptibility (χ) are constant. Within this regime, the algorithm is able to correctly distribute nodes into their correct communities. We test several systems with different system size N and number of communities q , and in the later section, we plot the first transition point p_1 in terms of N and q .

As the noise p_{out} is further increased beyond a threshold value of p_1 , the system enters the “hard” phase. The existence of the “hard” phase is reflected by the rapid growth (decrease) in the entropy and computational susceptibility (accuracy) curves. Even though we can increase the number of trials in order to improve the accuracy of

our solutions (as seen in panel (c) of Fig. 2.4), it is, nevertheless, still hard to obtain exact solutions.

As the noise is yet further increased and exceeds a second threshold value ($p_{out} > p_2$), the system undergoes another phase transition from the “hard” phase to an “unsolvable” phase. This “unsolvable” region is reflected, amongst other things, by the collapse of all of the curves in each panel of Fig. 2.4. In this regime, it is impossible to solve the system correctly without infinite time in the third region.

$$T \geq 0$$

A more detailed, higher dimensional perspective, that includes the effects of temperature is provided in Fig. 2.5. In this figure, summarizing our results for the computational susceptibility $\chi(T, p_{out})$, Shannon entropy $H(T, p_{out})$, normalized mutual information $I_N(T, p_{out})$ and system energy $E(T, p_{out})$ at general finite temperatures $T \geq 0$, we plot the loci of point marking the boundaries between the different phases. The “flat” phase that lies in the middle of these panels is the “easy” phase. [Within the “easy” phase, the system is easily solvable and the planted communities are perfectly detected.] This “easy” phase is separated by “ridges” of high computational susceptibility (marking the “hard” regions) from the “unsolvable” phases. As expected, the computational susceptibility/energy/entropy/ I_N exhibit a precipitous jump as the noise p_{out} exceeds some threshold value $p_1(T)$. A low temperature hard phase appear for noise levels $p_1(T) \leq p_{out} \leq p_2(T)$. We can determine the boundaries of the “hard” phase, whenever it generally exists, by seeing for which values of p_{out} and T there is

a rapid increase of χ and E . An additional high temperature bump in the computational complexity χ and E appears for noise levels $p_3(T) \leq p_{out} \leq p_4(T)$. In this phase, the minimization of Eq. (2.1) is non-trivial. At yet higher temperatures/noise levels, it is generally impossible to solve the system. Thus, the two loci of “ridges” in the computational complexity (i.e., $p_1(T) \leq p_{out} \leq p_2(T)$ or $p_3(T) \leq p_{out} \leq p_4(T)$) delineate the “hard” phases. To emphasize the appearance of these ridges and their manifestation in all measured quantities, a guide to the eye is drawn. Within the low temperature hard phase ($p_1(T) \leq p_{out} \leq p_2(T)$), the system becomes trapped in the local energy minima (panel (a) of Fig. (2.6)). At low temperatures, we find from the exact and extensive numerical calculations (as shown in panel (d) in Fig. 2.4), a very dramatic increase in complexity just at the transition p_1 followed by a much more gradual decrease up to p_2 . The convergence time for a local greedy algorithm (such as ours shown in (c) of Fig. 2.4) does not correlate with the complexity as the system. This is so as the system can easily converge to a wrong local metastable minimum (while the number of such minima is given by the complexity).

In Fig. 2.6, we provide caricatures of the underlying physics in these phases and the low temperature/low noise transitions. At low temperatures, for noise p_{out} slightly above p_1 (at zero temperature), the system becomes quenched in metastable local minima at low temperatures. This is schematically illustrated in panel (a). As the temperature is increased, the system may, as depicted in panel (b) of Fig. 2.6, veer towards its global minimum by annealing. Physically, a similar mechanism is at work in many frustrated physical systems where it goes under the name of “order by

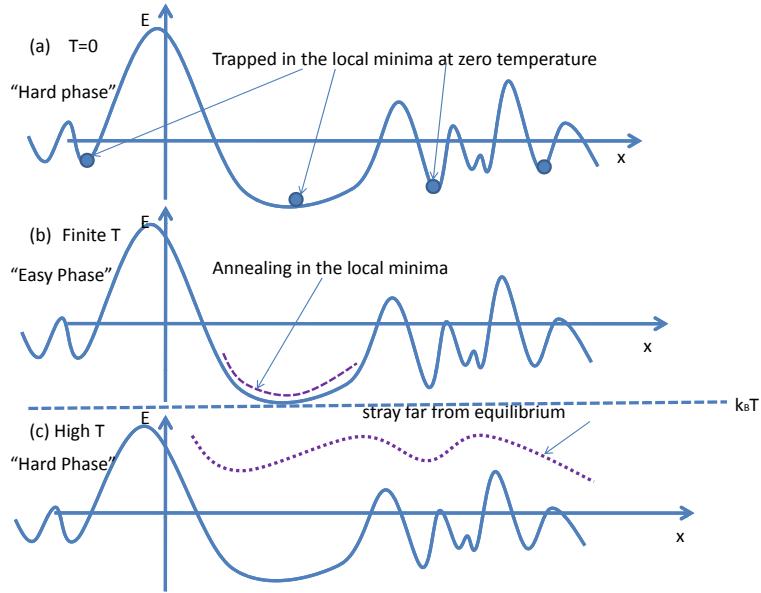


Figure 2.6: A caricature of the accessible energy landscape at different temperatures for a system, such as that examined in Fig. (2.5) with a fixed noise level p_{out} which slightly exceeds $p_1(T = 0)$. In panel(a), at zero temperature, the system is trapped in local minima. Panel(b) shows the system at temperatures that are sufficiently high for the system to anneal and better access regions in the vicinity of the lowest energy states. This situation corresponds to the intermediate region that lies between the two “ridges” in Fig. 2.5. Panel(c) shows the system in a high temperature phase where, thermal fluctuations are exceedingly large and the system does not veer towards low energy states.

disorder” . In such cases, by virtue of entropic fluctuations, quenching is thwarted and the system may probe low lying states and indeed order [87, 88, 89, 90]. Thus, the energy and computational susceptibility may remain constant (there is only one global energy minimum, i.e., one state or a finite set of such states). However, it

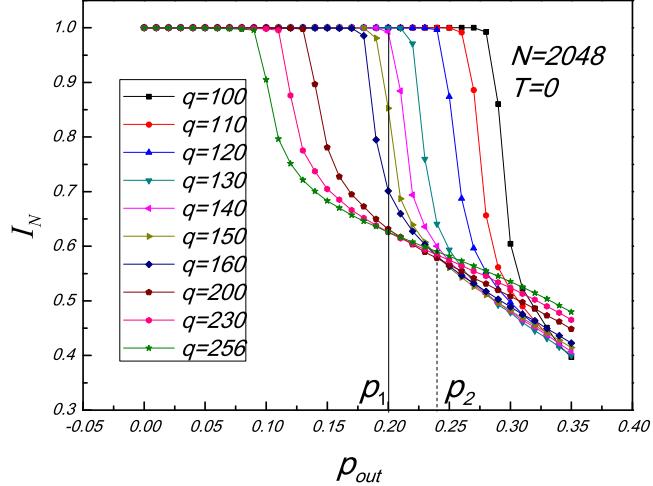


Figure 2.7: The normalized mutual information I_N as a function of p_{out} for system $N = 2048$ at temperature $T = 0$. The noise levels p_1 and p_2 are the first and the second transition points for the particular displayed system of $N = 2048$ and $q = 140$. The inferred values of $p_1 = 0.2$ and $p_2 = 0.24$ are consistent with Fig. 2.4. The normalized mutual information I_N records the overlap between the “important partitions” (the optimal partition corresponding to the lowest energy state of Eq. (2.1)) and the contending partitions found by the algorithm.

does take progressively more time to locate the global minimum state ((c) in Fig. 2.4). As the noise is further increased, the system is still ergodic. However, it takes a very long time to find the lowest energy state. On finite time scales, the system stays in the vicinity of local minima thus yielding a higher observed energy. Only on sufficiently long time scales does the system veer towards its global minimum (or minima). Within this “hard” region, there are many metastable states. This leads to a significant increase in the complexity as is made evident by the rapid growth of

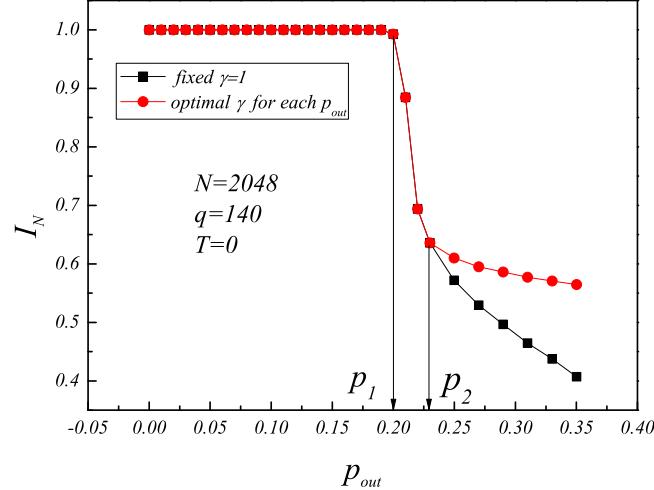


Figure 2.8: A comparison of the normalized mutual information I_N as a function of noise p_{out} between two cases: (i) one with a fixed resolution parameter $\gamma = 1$ (see Eq. (2.1)) and (ii) a computation with the optimal γ determined by the maximal I_N /minimal V (minimal variation of information). No change in the transition points p_1 and p_2 occurs by optimizing γ in this zero temperature system. Indeed, for this system $\gamma = 1$ is the optimal value of γ for noise levels $p_{out} < p_2$. The two curves start to separate for higher noise levels.

the computational susceptibility χ of Eq. (2.8). The large computational complexity marks the initial rapid climb of the complexity.

We now return to the results of Fig. 2.5 at yet higher temperatures and values of the noise p_{out} . The high temperature “ridge” in Fig. 2.5 ($p_3(T) \leq p_{out} \leq p_4(T)$) corresponds to the system being far away from the minimum energy state. As we remarked earlier, this delineates yet another “hard” phase. According to the above explanation and the corresponding caricature of Fig. 2.6, increasing the running time

and/or number of trials should help increase the accuracy of the solution in this region (the peak area of the computational susceptibility). Beyond this region, at higher temperatures, the system is unsolvable. This corresponds to panel (c) in Fig. (2.6).

At low temperatures and high noise, due to the proliferation of metastable states, (i) the convergence time τ (as seen in panel (c) of Fig. (2.4)) can be low while (ii) the increase in accuracy by performing more and more trials is, essentially, nil [as seen by the low value of χ in panel (d) of Fig. 2.5]. Similar conclusions can be arrived at finite temperatures by examining constant T slices of $\chi(T, p_{out})$.

We now examine, in further detail, several aspects of these transitions at $T = 0$. The (zero-temperature) normalized mutual information is displayed in Fig. 2.7. As evident from the figure, I_N starts to drop below its maximal value of $I_N = 1$ (which indicates perfect agreement with the optimal solution) when $p_{out} = p_1$ (i.e., at the very same value of the noise $p_{out} = p_1$ where the relaxation time is maximal and the complexity increases) and I_N levels off at a higher value of the noise $p_{out} = p_2$ (coincident with the transition value as ascertained from the energy, entropy and complexity in Fig. 2.4). Amongst other collapses that we observed, systems with differing number of communities q all collapse onto each other at $p_{out} = p_2$.

Before we turn to a more detailed analysis of spin-glass character of the transitions, we make one remark. A possible concern is that we did not examine transitions for the optimal value of γ . Indeed, the central thesis of [18] was that there are optimal values of γ that signify the natural scales in the system. In general, transitions as

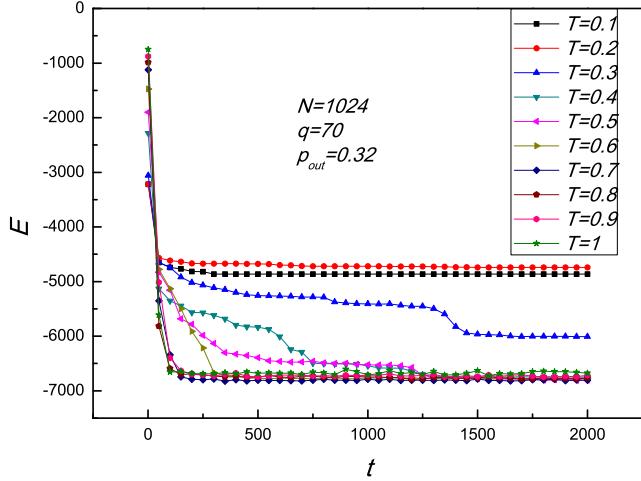


Figure 2.9: The plot of energy versus time for the system $N = 1024$, $q = 70$. We fix the noise as $p_{out} = 0.32$ which is above the first transition point $p_1 = 0.3$ (in the “hard” region). Note that the energy curves with different temperatures have a “crossover” at about $t = 1250$. Before that, the curve with low temperature is always above the one with high temperature. After that, except at temperatures of $T = 0.1, 0.2$ or 0.3 , the curves of the low temperature systems dip below those of the higher temperature ones. The “crossover” property shown here is a sign of transition from non-equilibrium to equilibrium.

a function of γ correspond to transitions in structure that appear as the system is examined on larger and larger scales as we have examined in detail in earlier works [18, 63, 64, 80]. To ascertain the changes that occur in the random systems that we investigated in this chapter for a broad spectrum of different values of γ (i.e., containing general $\gamma \neq 1$), we re-investigated these systems with γ values within the range $10^{-2} \leq \gamma \leq 100$. The “best ” values of γ are ascertained by maxima of the

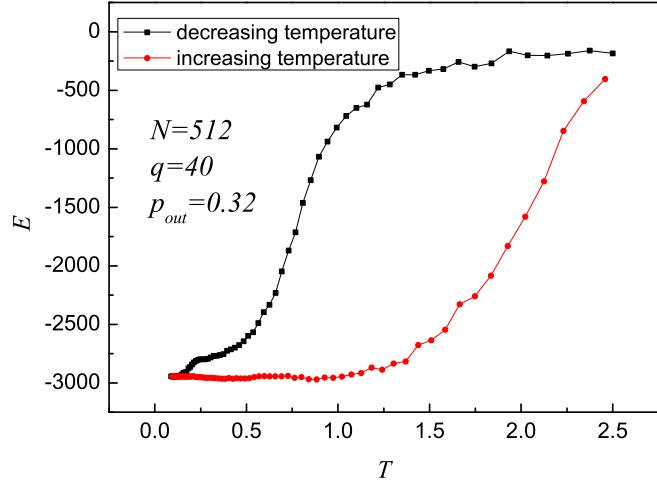


Figure 2.10: The energy E as a function of the temperature T for a system with $N = 512$ nodes, $q = 40$ communities, and noise $p_{out} = 0.32$. (This noise level exceeds the zero temperature $p_1 = 0.29$ for this system.) We perform a computational experiment at $T = 2.5$ and lower the temperature according to $T_{k+1} = 0.95T_k$ in consecutive time steps k . After a steady-state is obtained, the process is reversed. A clear hysteresis-like effect is evident.

normalized mutual information I_N [27]. In Fig. 2.8, we display I_N as the function of the noise p_{out} for both the fixed $\gamma = 1$ and the optimal γ determined by the multiresolution algorithm. The first transition point p_1 is the same in both cases, and the two curves start to separate around the second transition point p_2 . This indicates that, as it so happens to be in this case, $\gamma = 1$ is the best value of resolution parameter for noise levels below p_2 in this example system ($N = 2048, q = 140$) at zero temperature.

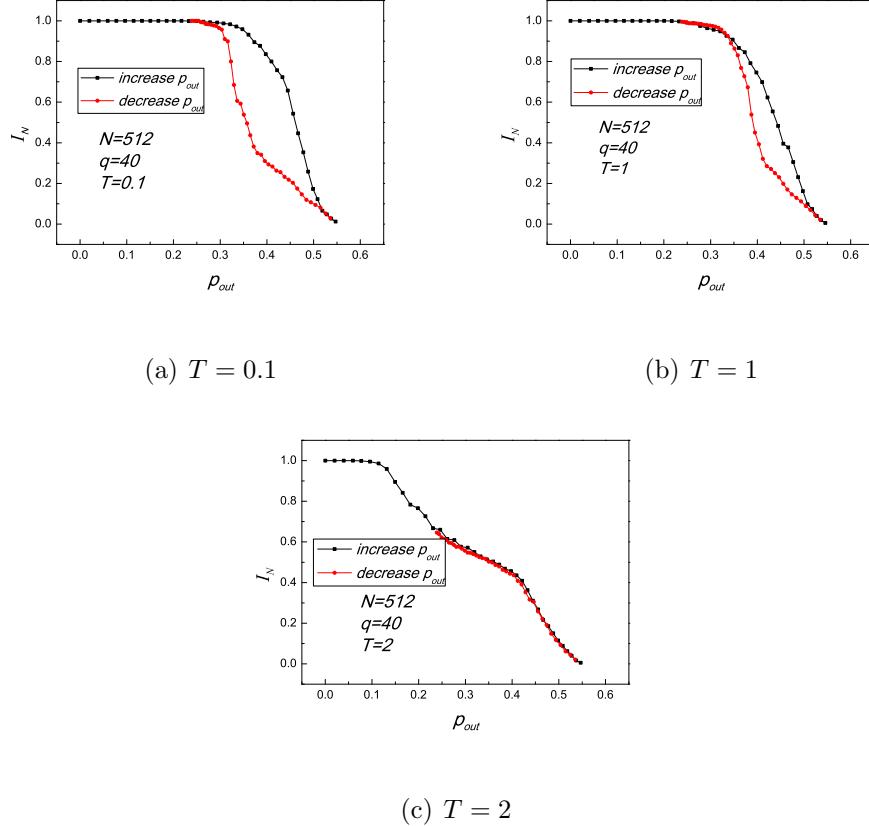


Figure 2.11: The plot of I_N in terms of p_{out} for system with $N = 512$, $q = 40$. (I_N is a normalized variant of mutual information, for detailed explanation, see Sec. 2.5.) From top to bottom, the temperature varies from $T = 0.1$ to $T = 2$. Note that the curves in panel(a) and (b) show the effect of hysteresis at low temperatures. Hysteresis disappears when the temperature is sufficiently high, e.g., $T = 2$ in panel (c).

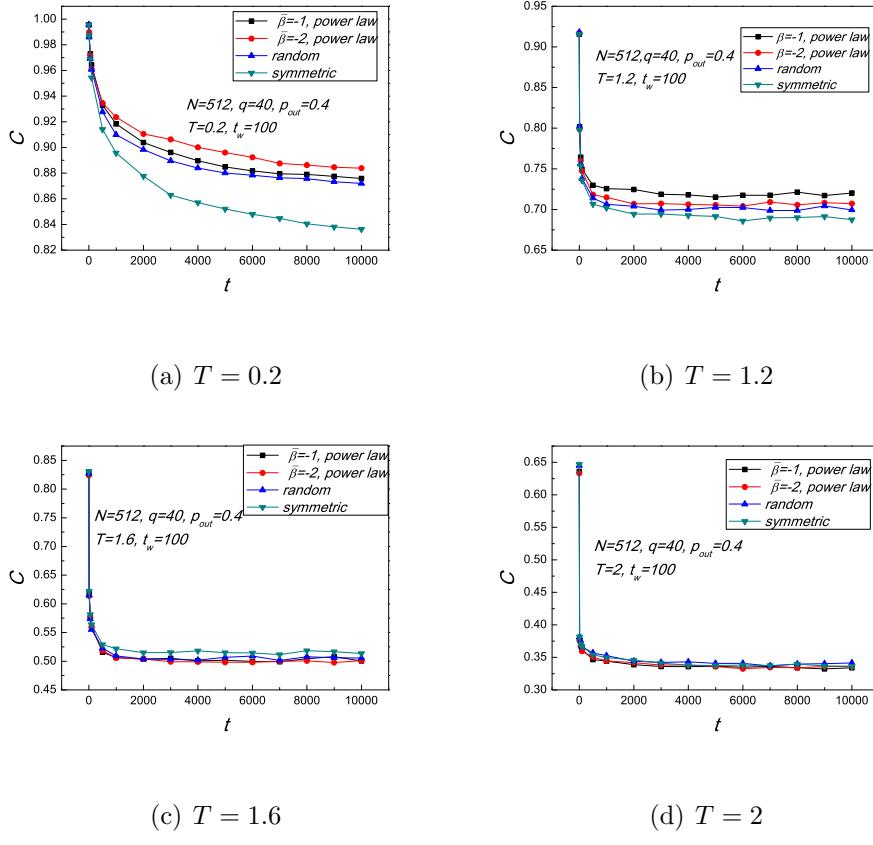


Figure 2.12: The autocorrelation function as a function of time for system $N = 512$, $q = 40$, $p_{out} = 0.4$ (above the transition point in this system). The waiting time $t_w = 100$ in all the panels. The four curves in each panel represent four different initializations for the studied system. Temperature varies from $T = 0.2$ to $T = 2$. At low temperature, all the curves with different initializations separate from each other even up to $t = 10000$ (panel(a)). Then, as T increases, all of the curves start moving towards (panel(b)), and finally overlap (panel(c) and (d)). with each other.

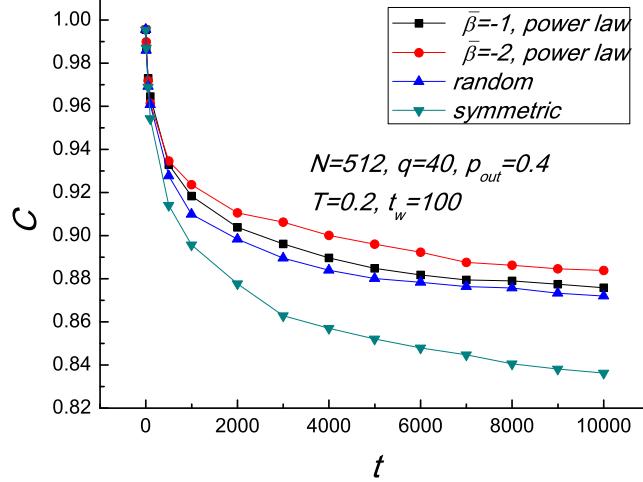


Figure 2.13: The autocorrelation function (Eq. (2.16)) as a function of time for system of $N = 512$ nodes consisting of $q = 40$ communities with a noise level of $p_{out} = 0.4$. The waiting time $t_w = 100$ and the temperature $T = 0.2$. The four displayed curves represent four different initializations for the studied system. “Symmetric” initialization means that each node forms its own community, so there are N communities as a starting point for the algorithm. “Random” means randomly filling q_0 communities with nodes, where q_0 is a random number generated between 2 and $\frac{N}{2}$. “Power law distribution” means separating N nodes into different communities, whose size satisfy power law distribution with a negative exponent, set to be $\bar{\beta} = -1, -2$. also, the maximal community size is set to be 50, the minimal community size is 8 in the above simulation results. At low temperature ($T = 0.2$), all of the curves with different initialization separate from each other even up to times of $t = 10000$ steps. As this figure makes clear, different sorts of randomness lead to different behaviors.

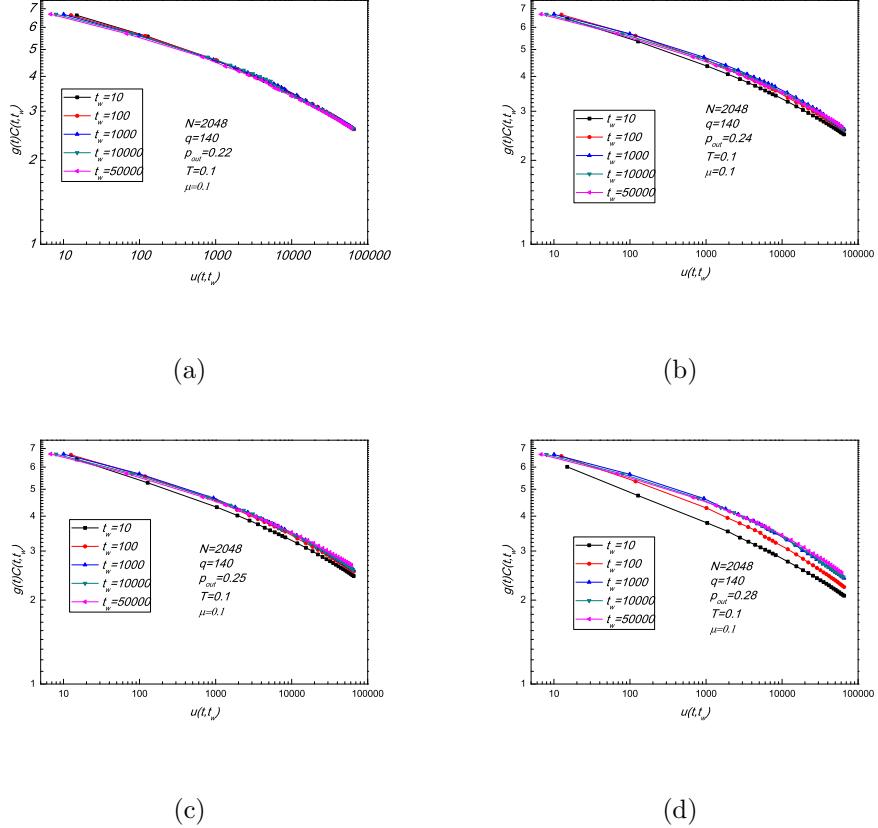


Figure 2.14: A validation of the spin glass character of the low temperature hard phase.

We show a collapse of the autocorrelation curves for the for different waiting times t_w for a system with $N = 2048$ nodes, $q = 140$ communities, and p_{out} varies from 0.22 to 0.28. The first and second transition points for this system are $p_1 = 0.2$ and $p_2 = 0.24$. The heat bath temperature is $T = 0.1$ in all these panels. The vertical axis is $g(t)C(t_w, t)$ where $g(t) = 8 - \log_{10}(t)$. The horizontal axis is $u(t_w, t) = \frac{1}{1-\mu}[(t + t_w)^{1-\mu} - t_w^{1-\mu}]$ where $\mu = 0.1$. (See text.) The noise $p_{out} = 0.22$ in panel (a) lies within the “hard” region where the collapse of correlation function is perfect. The noise values of $p_{out} = 0.24$ and $p_{out} = 0.25$ in panels (b) and (c) respectively are around the second transition point, where the collapse becomes fainter. The noise of $p_{out} = 0.28$ in panel(d) is above the second transition point p_2 -i.e.-in the “unsolvable” region, where the collapse becomes very poor. That the collapse of the correlation function starts to degrade right after the second transition point p_2 at 56 low temperature indicates that this transition is of the spin-glass type.

2.8.3 Numerical validation of the spin glass character of the two transitions

The proliferation of metastable states thwarts equilibration. A specific facet of this is detailed in Fig. 2.9 wherein, by energy measurements, the lack of equilibration at short times is evident.

In equilibrium, the energy is (of course) constant. The system energy is set by its temperature. In Fig. 2.9 we investigate, at different temperatures, the evolution of the system from an initial high energy state. In the particular results that we provide below, the energy of the disparate systems at low temperatures would, at short times, naively seem to violate thermodynamic expectations. Systems with lower temperature can have higher energies than those at higher temperature. The origin of this and similar effects is that significant time may be required to achieve thermodynamic equilibrium. Within the low temperature unsolvable phase the system is out of equilibrium. In the hard phase, equilibrium is achieved yet it requires long times.

We set the system size (number of nodes) to be $N = 1024$ with $q = 70$ communities and a value of the noise given by $p_{out} = 0.32$. As such, with this value of p_{out} which is larger than the threshold value of $p_1 = 0.3$ for this system, the system is in the “hard” phase. We examine the system evolution with the algorithm time steps in Fig. 2.9. In this plot, the system has a “crossover” at about $t = 1250$. Prior to that time, the energy always decreases as T increases. This reflects the fact times below $t = 1250$ are not sufficiently long for the system to equilibrate. After that, except for

the cases of $T = 0.1, 0.2$ or 0.3 , the energy turns to increase as T increases. Thus, $t = 1250$ constitutes sufficient time for equilibration except a few systems at very low temperature (that require yet longer times). This “crossover” property for system is a sign of the restoration of equilibrium at sufficiently long times.

All the curves show a decrease of the energy with time until a plateau is reached. When time is not sufficiently long, the system is not ergodic and out of equilibrium. As seen in Fig. 2.9, times $t > 2000$ are required for lowest temperature systems (e.g., $T = 0.1, T = 0.2$) to equilibrate.

As is well appreciated, this absence of equilibration due to multiple metastable states may lead to spin-glass-like (as well as structural glass like) properties. Amongst other traits, these include memory effects previously studied for other systems [91, 92].

When a spin glass is cooled down, a memory of the cooling process is imprinted in the spin structure, and this process will be reproduced if one heats the system up.

We conduct a similar computational “experiment”. We immerse our system (with a fixed value of the noise p_{out}) in a heat bath. We then lower the heat bath temperature T by small increments at consecutive time steps k . (Each time step corresponds to a single iteration through all nodes according to the minimization algorithm of [18, 27].) In this case, we set $T_{k+1} = 0.95T_k$. After attaining a steady-state solution, we then reverse the process and increase T after each step via $T_{k+1} = 1.05T_k$. In Fig. 2.10, we plot the long time system energy E as a function of T during this process. The energy curve as T decreases follows a different path than when T increases which strongly implies a hysteresis-like effect. This memory effect as the temperature

is cycled between high and low T reinforces the similarity between the community detection and a spin glass system.

The behavior of the energy displayed in Fig. 2.10 suggests the same three regions that we ascertained earlier: (i) When the two curves overlap at low temperatures (i.e., $T < 0.1$), the system is in its “frozen phase”. (ii) When the two curves separate in a medium temperature range (i.e., $0.1 < T < 2.5$), the system is in a “spin-glass” phase. (iii) At yet higher temperature ($T > 2.5$), the two curves overlap once again. This marks the onset of the “disordered” high temperature regime.

As illustrated in Figs.(2.4,2.5) (and as will be further discussed in Figs. (2.14,2.15)), the hard phases at both low and high temperatures do not extend over all temperatures. Rather, as we have emphasized above, the hard phases only appear in the “complexity” ridges as shown in panel (a) of Fig. 2.5. However, in Fig. 2.10, the hysteresis occurs in the temperature range $0.1 < T < 2.5$. This range is considerably larger than that of the hard phases. To understand this, we remark on the “experimental” differences between the results displayed in Fig. 2.5 and those in Fig. 2.10. In constructing the 3D plot of the “complexity” (panel (a)) in Fig. 2.5, we apply the “HBA” at each temperature. The systems at different temperatures are independent of one another. That is, each system is solved afresh from the symmetric initial state. In the hysteresis loop in Fig. 2.10 on, e.g., the decreasing temperature curve, a system at higher temperature provides the initial state for a lower temperature system. Thus, in this case, the systems at different temperatures are not independent but rather serve as “seed” states for one another.

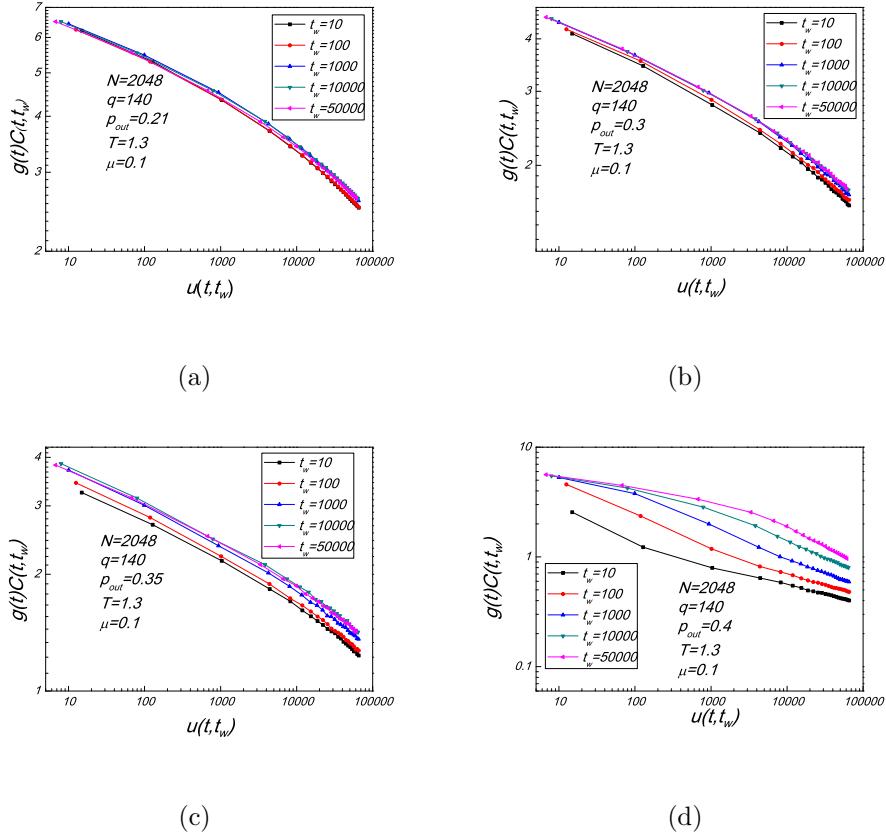


Figure 2.15: An illustration of the spin glass character of the high temperature hard phase.

Shown is a collapse of the autocorrelation curves for different waiting times t_w for the system of $N = 2048$ nodes and $q = 140$ communities. The heat bath temperature is $T = 1.3$ in all panels. In this collapse (see text), the vertical-axis is $g(t)C(t_w, t)$ where $g(t) = 8 - \log_{10}(t)$ and the horizontal-axis is $u(t_w, t) = \frac{1}{1-\mu}[(t + t_w)^{1-\mu} - t_w^{1-\mu}]$ where $\mu = 0.1$. Panel(a) of $p_{out} = 0.21$ is within the high temperature hard phase (evident as the higher temperature “bump” in the 3D plot of computational susceptibility $\chi(p_{out}, T)$ (Fig. 2.5)). Within the hard phase, the collapse is perfect. Panel(b) of $p_{out} = 0.3$ is around the boundary of the hard phase. Correspondingly, the collapse starts to lose its precision. Panel(c) corresponds to $p_{out} = 0.35-$ outside the hard phase. A poor collapse is seen. Panel(d) corresponds to $p_{out} = 0.4$ is far from the hard phase. No collapse is seen. The collapse of the autocorrelation function loses its perfection right after the second transition point p_4 at *high*

Aspects of the memory effect are evidently not limited to those of, e.g., Fig. 2.10. For instance, if we incorporate the effects of increasing and decreasing noise to the same system [93] instead of temperature, the accuracy of the solution also forms a hysteresis loop at low temperature (see Fig. 2.11). That is, we consider the effect of adding external edges between disparate communities (i.e., increasing p_{out}) and then removing these edges (i.e., decreasing p_{out}). We examine the accuracy of solutions as a function of noise and see whether the two curves coincide. The non-coincidence between the two processes will exhibit exactly the same memory effect that we earlier reported on by varying the temperature.

Fig. 2.11 shows the results of the above experiments at three temperatures: $T = 0.1, 1$ and 2 . In the $T = 0.1, 1$ systems, the curves with increasing p_{out} and decreasing p_{out} form hysteresis loops. The hysteresis loop in temperature $T = 1$ in panel (b) is less significant than its counterpart for at $T = 0.1$ in panel (a). Upon further increase of the temperature, the hysteresis disappears (as shown in panel (c)).

The plots in Fig. 2.11 have already exhibited decreasing memory effects as the temperature increased. Thus, there must exist a temperature beyond which the effect disappears. We investigated the plots at temperatures $T = 1.1, 1.2, \dots, 1.9$ (not shown here). The hysteresis loop disappeared at about $T = 1.7$ in line with our other reported results including the disappearance (at $T = 1.6$) of memory of initial conditions to which we turn to next.

A general quantitative measure of the memory, the two-time autocorrelation func-

tion between the system at times t_w and time $t + t_w$,

$$C(t_w, t) = \frac{1}{N} \sum_{i=1}^N \delta_{\sigma_i(t_w), \sigma_i(t_w+t)}, \quad (2.16)$$

can be used to explore the spin-glass-like behavior. The upshot of the below discussion is that the autocorrelation function data only within the hard phases (both at low and at high temperatures- coincident, as emphasized earlier, with the “ridges” in Fig. (2.5)) adheres to a spin-glass type collapse. This affirms, once again, the spin glass character of the transitions.

If we apply the HBA starting from *different initial configurations* at low temperature (as elaborated on in Fig. 2.12), all of the auto-correlation curves with different initializations separate from each other even up to times t [in units of the iteration through all nodes according to the algorithm of [18, 27]] as large as $t = 10000$ (Fig. 2.13). We report on the autocorrelation functions (Eq. (2.16)) for *three different types of initial configurations*. The conclusion is that the system may be sensitive to initial conditions. The three initializations are denoted as symmetric, random and power law distribution.

- “Symmetric” initialization alludes to an initialization wherein each node forms its own community, so there are N communities in the beginning (as in step (1) of the algorithm outlined in Sec. 2.4).
- “Random” refers to randomly filling in q_0 communities with nodes, where q_0 is a random number generated between 2 and $\frac{N}{2}$.
- In the “Power law distribution” N nodes are partitioned into different commu-

nities whose size adheres to a power law distribution ($\text{Prob} \sim n^{-\bar{\beta}}$) with a negative exponent $\bar{\beta}$. In the cases displayed, we set $\bar{\beta} = -1, -2$.

The maximal community size in the true solution is set to be 50 and the minimal community size is 8.

Fig. 2.12 vividly illustrates that all the curves with different initializations separate, at low temperatures, from each other even up to times of size $t = 10000$. The curve with symmetric initialization lies on the bottom in panel (a). However, as temperature increases, all of the curves veer towards each another. The symmetric curve moves form the bottom to the top at a temperature $T = 1.6$ as shown in panel (c). As temperature increases furthermore ($T = 2$), all of the curves overlap in panel (d).

At a temperature of $T = 1.6$, systems with different initial configurations start to overlapping. Beyond this temperature, there is no remaining memory of the initial conditions. Furthermore, the relative position of the curves become different, which is another indication for the lose of memory. The spin temperature at which we found the hysteresis loop to disappear in Fig. 2.11, $T = 1.7$, nearly coincides with the temperature found here.

In Fig. 2.12, the relative positions for the “random” and “power law distribution” do not persist: their positions change irregularly as temperature varies. This indicates that these two are similar to each other—there is no essential difference between them. However, the curve of symmetric initialization lies below the other two until the temperature rises up to $T = 1.6$, which happens in all the waiting times that we tested ($t_w = 100$, $t_w = 10$ and $t_w = 1000$ (not shown here)). This suggests that the

symmetric initialization differs, in an essential way, from the other two initializations.

Fig. 2.12 and Fig. 2.13 indicate that disparate sorts of randomness can, generally, lead to different results. As the temperature T is increased, all of the curves ultimately *collapse* onto one another. The temperature at which the different initial configurations overlap indicates when the respective systems start losing memory of their initial configurations directly relates to the transition temperature in the hysteresis loop for the same system. This further establishes the existence of spin glass transition in the community detection problem.

We use the HBA starting from a symmetric initial state and calculate the autocorrelation in Eq. (2.16) for different waiting times t_w and temperatures T . We further found that each auto-correlation curve $C(t, t_w)$ corresponding to longer waiting time t_w lies above those with shorter waiting times, and all the curves (with different waiting times) are non-zero for a long period of simulation time indicating a memory effect. Moreover, we can predict the long time behavior of $C(t_w, t)$ by fitting the curves using a commonly-used equation in Fig. (2.14, 2.15), for more details, see [94, 95].

Towards this end, we set

$$g(t) = a - b \log_{10}(t), \quad (2.17)$$

and

$$u(t_w, t) = \frac{1}{1-\mu} [(t + t_w)^{1-\mu} - t_w^{1-\mu}]. \quad (2.18)$$

In the above equations, a, b and μ are parameters that need to be optimized in order

to ascertain whether a generic spin glass type collapse occurs [94, 95]. In searching for a collapse of the data points at different waiting times t_w , we use $g(t)C(t_w, t)$ as a vertical-axis and $u(t_w, t)$ as a horizontal-axis. As seen in Figs. (2.14, 2.15), a collapse indeed occurs over 4 decades in values of $u(t_w, t)$ in *both the high and low temperature hard phases*.

We discuss several features of this collapse and its coincidence with the hard phase below. Fig. 2.14 corresponds to the low temperature hard phase and Fig. 2.15 corresponds to the high temperature hard phase [see, e.g., the 3D computational susceptibility plot $\chi(p_{out}, T)$ in Fig. 2.5]. As seen in Figs.(2.14, 2.15), *both the high and low temperature cases*, the autocorrelation functions with different waiting times t_w exhibit spin-glass collapse when the value of p_{out} lies within the “ridge” area of the hard phases. This collapse wanes when p_{out} veers towards the “foot” of the complexity ridge just at the onset of the hard phase. The collapse ultimately becomes non-existent when p_{out} is further away from the “ridge” area. The regime where the correlation functions satisfy the spin-glass collapse is consistent with the parameters corresponding to the hard phase (or “ridge” in the 3D computational susceptibility plot of Fig. (2.5)). Putting all of the pieces together, we see from our scaling and collapse in Figs. (2.14, 2.15), that both high and low temperature transitions of the *spin-glass type*.

In the random graphs, we reported on *spin-glass* type transitions. Although trivial, for completeness, we should however note that a graph can, obviously, also be very regular. A prototypical example is that of the two-dimensional square lattice. For

such regular unfrustrated lattice systems, the Potts model of Eq. (2.1) becomes the “standard” Potts model of lattice systems. In these instances, we generally have single *first or second order* transitions instead of spin-glass type transitions. We briefly elaborate on this point. Simple regular lattices are a particular realization of a graph (one with the fixed coordination and translational symmetry). As is well known, on, e.g., the square lattice, the Potts model, which we use throughout, exhibits as a function of the temperature T , two phases with an intervening *critical point* for small q ($q \leq 4$); for larger q ($q > 4$), a *first order transition* appears. Thus, particular realizations of our hamiltonian for these graphs display (usual) critical points and first order phase transitions. For more generic random graphs with high coordination, the system displays (as we showed above and will further elaborate on), spin-glass type transitions appear along with intervening hard phases.

We further reiterate an earlier remark and note that in systems with well defined structures on multiple scales, *additional transitions* may appear as the resolution parameter γ of Eq. (2.1) is varied. In earlier works, we reported on these transitions and further employed these in the analysis of disparate systems [18, 63, 64, 80].

2.8.4 General discussion

In this subsection, we detail general considerations directly related to the spin-glass Potts analysis thus far. In the next section, we will further discuss dynamics which further relates to aspects that we detail herein. This subsection is different from others in that here (and only in this subsection), we present a general discussion and

some speculations and not present data.

Theoretical Expectations from NP-completeness

In [96], it was shown that maximizing modularity (an earlier alluded to prominent approach for the community detection problem [6, 10, 75]) is NP-complete. Thus, as all NP complete problems may (by their very definition) be mapped onto one another, maximizing modularity on the most general graphs must span the three phases (solvable and unsolvable with the further division of the solvable problems into the “easily solvable phase” and the “hard phase”) that appear, e.g., in k-SAT problem [30] which is known to be NP-complete [97]. Similarly, if other approaches to community detection are, ultimately, equally hard as maximizing modularity, then all of these approaches may in general display three phases. It may be, as in the k-SAT problem, that for simple problems, we have only an “easy phase” and an “unsolvable” (or “unSAT”) phase. This does indeed occur for some graphs. In general, though, we find the three different phases (as expected) that we reported on in this work.

Physical content of the transition in many body systems

- *Approximate decoupling*

We briefly speculate, in this subsection alone, on potential physical consequences of the phase transition that we find in the community detection problem. As elaborated on in [63, 64] a general many body system with two particle interactions may

be regarded as a network with edge weights determined by the interactions. In the easy phase in the extreme limit of $p_{out} = 0$, the system is essentially that of disjoint non-interacting clusters. More generally, The Potts model Hamiltonian Eq. (2.1) can be written as: $H(\{\sigma\}) = \sum_{k=1}^q H_k$. Thus, the partition function becomes:

$$\begin{aligned} Z(\beta) &= \sum_{\{\sigma\}} e^{-\beta H} = \sum_{\{\sigma\}} e^{-\beta \sum_{k=1}^q H_k(\{\sigma_i\} \in k)} \\ &= \sum_{\{\Lambda\}} \prod_{k=1}^q \left(\sum_{\{\sigma_i\} \in k} e^{-\beta H_k} \right) = \sum_{\{\Lambda\}} \left(\prod_{k=1}^q Z_k \right). \end{aligned} \quad (2.19)$$

In Eq. (2.19), Z_k is the partition function as computed with the Hamiltonian of the entire system for the particles in community k , and $\{\Lambda\}$ denote partitions of the system.

A similar form was proposed for many body systems in [63] when partitioning a general interacting system into decoupled clusters. Even though, we sum over all partitions, we may have an important subset of partitions, denoted as $\{\Lambda'\}$ (each with a corresponding number of clusters equal to $q_{\Lambda'}$), which will have in general instances, high Boltzmann weights and/or frequencies and will dominate the sum. These partitions will, correspondingly, have a significant lower free energy relative to other partitions. In such cases, the partition function can be approximated as

$$Z \simeq \sum_{\{\Lambda'\}} \prod_{k=1}^{q_{\Lambda'}} Z_k \quad (2.20)$$

Eq. (2.20) is exact in the limit of $T = 0$ where Λ' denotes the ground-state(s) of our Potts type Hamiltonian. If p_{out} is small (in particular if $p_{out} = 0$), then there will

generally exist a small number of sharply defined ground-states $\{\Lambda'\}$ pertaining to partitions into completely disconnected communities. This general trend of dominant subsets may persist within the easily solvable phase.

The possible upshot of this discussion is that we might, in easy phases, approximate many body interacting systems (such as supercooled liquids that we will briefly discuss next) as effectively composed of disjoint non-interacting clusters. This picture may badly break down once transition lines between the easy phase and the hard or unsolvable phases are traversed.

- *Possible relation to structural glasses and other complex physical systems*

Glasses (according to the theories such as the random first order transition theory of glass (RFOT) in [98]) may have three phases as a function of temperature. In the intermediate phase, the system displays a large complexity (as manifest in the configurational entropy being extensive). If we replace the interacting particles in a supercooled liquid (that form a glass at low temperatures) by decoupled communities [63, 64], then the three phases found in the computational community detection problem may be manifest as three disparate phases of supercooled liquids as a function of temperature. Within RFOT, at temperatures in an intermediate region ($T_0 < T < T_A$), the system physically displays an extensive configurational entropy (which is tantamount to an extremely large complexity in the current context). This configurational entropy precipitously onsets at $T = T_A$ and gradually diminishes un-

til it no longer becomes extensive at a lower temperature ($T = T_0$) whence the system freezes into an “ideal glass” that is permanently stuck in a metastable state.

We will discuss, in Section 2.9 dynamical aspects that directly relate to the Potts model Hamiltonian. Insofar as additional generally related aspects of the results of our community detection analysis the implication of phase boundaries, we make a brief comment. When, as discussed in [63, 64], a weighted version of Eq.(2.1) is used with edge weights that are set by forces, then in overdamped viscous systems (where the total force on a particle is proportional to its velocity, $\vec{f}_i = c\vec{v}_i$), particles that experience a similar total force, will tend to move in unison. Thus, in the easy phase motion of decoupled cohesively moving particles will occur. In the unsolvable phase, the particle motion will be more complicated. In earlier work, forces were used to study community detection with similar results overall to the one afforded by our spin glass approach in this work [73]. When other weights are used (such as potentials, two-body correlations or other metrics), similar decoupling within the solvable phase signifies a tendency of the clusters not to be related insofar as the metric being used.

Image segmentation

Recently, we investigated and invoked the features of the phase diagram in order to address the computer vision problem of detecting objects in general images [80] (including notably challenging ones). As in this work for random graphs, by varying parameters such as the temperature, the (graph) resolution parameter γ and physical length scales, we explored the community detection phase diagrams for image seg-

mentation. Within the easy phase, disparate objects were clearly seen. As the system moved into the hard phase, the sharpness of the objects became more fragmented. These ultimately became very noisy in the unsolvable phase.

In summary, whenever a decomposition of an interacting many body system into nearly decoupled communities is possible (indeed, as alluded to above, such a decomposition is exact for Potts model systems wherein the exchange energy between spins in different domains is zero) then the phase transitions that we report on here for the community detection problem may carry direct physical consequences. This may afford a direct link between the phase diagrams of hard computational problems (as ascertained by physically inspired approaches) and the phase diagrams of physical systems that may be investigated via solutions to these related computational problems. It is important to note that the direct relation between complexity and glassiness is not simple as some problems that may be investigated by sub-optimal algorithms (such as physical stochastic systems) may appear to have a “hard phase” while if investigated by a more efficient algorithm do not have a “hard phase” [99]. Nevertheless, it may well be possible that the decomposition of physical systems into simple elements will no longer be simple at the onset into complex states such as those of supercooled liquids. Indeed, in recent work, we applied the community detection ideas to general many body systems (including glasses) in order to flesh out prospective important structures on all scales [63, 64, 80].

2.9 Dynamical Aspects

In the following, we also study the related dynamical transitions in both classical and quantum systems. Dynamic approaches to community detection have been suggested earlier [26, 73]. To describe the dynamical process, we need to calculate the trajectory (of community memberships) for each node as a function of time. Specifically, we use the correspondence between the q -state Potts model and a clock-type model in $(q - 1)$ dimensions. We replace the Kronecker delta $\delta(\sigma_i, \sigma_j)$ in Eq. (2.1) by a product $\vec{n}_i \cdot \vec{n}_j$ where \vec{n}_i and \vec{n}_j are the vertices of a regular $(q - 1)$ -dimensional simplex. On such simplices (e.g., an equilateral triangle ($q = 3$), tetrahedron ($q = 4$), ...), $\vec{n}_i \cdot \vec{n}_j = [1 + 1/(q - 1)]\delta_{ij} - 1/(q - 1)$. Thus, as is well known, we can cast the Hamiltonian of Eq. (2.1) into the form $H = -\sum_{ij} A'_{ij} \vec{n}_i \cdot \vec{n}_j$ where $A'_{ij} = (1+\gamma)A_{ij} - \gamma$ is the interaction weight. If we allow for an external field \vec{h}_i then the Hamiltonian reads

$$H = -\sum_{ij} A'_{ij} \vec{n}_i \cdot \vec{n}_j - \sum_i \vec{h}_i \cdot \vec{n}_i. \quad (2.21)$$

In what follows, we will first outline a very simple new general method for relating a general statistical mechanics system (such as the particular Potts model under consideration) and a dynamical system from classical mechanics. Although, this method will be specifically invoked for the Potts model, all of its steps can be replicated for other systems as well. We will then proceed to show the results of our numerical analysis. The final result of our analysis is that *the spin glass transitions relate to transitions to chaos in the dynamics of the continuous mechanical system.*

2.9.1 Relating discrete Hamiltonians to continuous dynamics

We will in this subsection illustrate how it is possible to relate the discrete Potts model Hamiltonian of Eq. (2.1) [and its clock model variant of Eq. (2.21)] to mechanical system with continuous dynamics. Many possible similar variants of the method outlined below are possible. Although our present aim is to investigate the Potts model Hamiltonians, as noted above, our method can be applied mutatis mutandis to general discrete Hamiltonians. A benefit of this mapping is that it bridges *chaos* in the more standard mechanical sense to that reported in spin glass systems.

Starting with Eq.(2.21), we perform a Hubbard-Stratonovich transformation via non-compact auxiliary fields $\vec{\eta}$ to arrive at the effective Hamiltonian (or, more precisely, free energy)

$$\begin{aligned}\beta H_{eff} &= -\ln Z \\ &= \sum_{i \neq j} \vec{\eta}_i (\beta A')_{ij}^{-1} \vec{\eta}_j - \ln \left(T r_{\vec{\eta}_i} e^{\vec{\eta}_i(\vec{\eta}_i + \beta \vec{h}_i)} \right)\end{aligned}\quad (2.22)$$

where Z is the partition function,

When we diagonalize the A'_{ij} matrix in Eq. (2.22) and use λ_α to denote its eigenvalues, the effective Hamiltonian in Eq. (2.22) becomes

$$\tilde{H} = \sum_{\alpha} (\beta \lambda_{\alpha})^{-1} |\eta_{\alpha}|^2 + O(\eta), \quad (2.23)$$

The form that we wrote involving the absolute value $|\eta_{\alpha}|^2$ in Eq. (2.23) for the bilinear in its eigenbasis was for a Hermitian symmetric matrix A'_{ij} . In general, for a

real symmetric matrix A_{ij} , no absolute value is needed. \tilde{H} in Eq. (2.23) is positive in $\{\eta_\alpha\}$, $\min\{\lambda_\alpha\} > 0$.

The dynamical equation for a node moving under the effective field is, for a damped system, given by

$$\begin{aligned}\frac{d\vec{\eta}_i}{dt} &= -\frac{\delta H_{eff}}{\delta \vec{\eta}_i} \Big|_{\vec{h}_i=0} \\ &= -\beta^{-1} \sum_j (\beta A')_{ij}^{-1} \vec{\eta}_j + \beta^{-1} \frac{\sum_{\vec{n}_i} \vec{n}_i e^{\vec{n}_i \vec{\eta}_i}}{\sum_{\vec{n}_i} e^{\vec{n}_i \vec{\eta}_i}}.\end{aligned}\quad (2.24)$$

We initialize the auxiliary field $\vec{\eta}_i$ to be some constant vector close to 0. We can solve this dynamical relation to obtain the non-compact auxiliary field $\vec{\eta}$ as a function of time [100].

We can obtain an equation of motion by differentiating the partition function Z 's [Eq. (2.22)] with respect to the source \vec{h}_i , i.e.,

$$\begin{aligned}\langle \vec{n}_i \rangle \Big|_{\vec{h}_i=0} &= \frac{\delta \ln Z\{\vec{h}_i\}}{\delta \beta \vec{h}_i} \Big|_{\vec{h}_i=0} \\ &= \frac{\sum_{\vec{n}_i} \vec{n}_i e^{\vec{n}_i \vec{\eta}_i}}{\sum_{\vec{n}_i} e^{\vec{n}_i \vec{\eta}_i}}.\end{aligned}\quad (2.25)$$

Substituting $\vec{\eta}$ in Eq. (2.24) into Eq. (2.28), we can determine the trajectory of the nodes.

2.9.2 Numerical results for the continuous dynamical analog

Eq. (2.24) describes overdamped (or Aristotelian) dynamics. It is, of course, possible to also define the system in such a way that it evolves according to Newton's equation. In overdamped systems, the energy of the system goes down with time and thus the

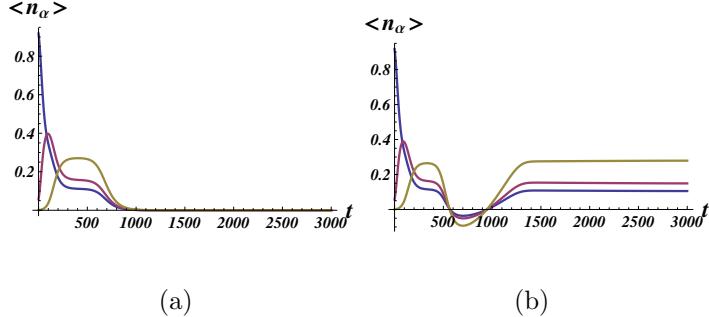


Figure 2.16: The three Cartesian components ($\alpha = 1, 2, 3$) of the node trajectories $\langle n_\alpha \rangle$ (for any node i) as a function of time t . The displayed system has $N = 24$ nodes, $q = 4$ communities, and is solved at a temperature of $T = 0.05$. In each plot, the three different Cartesian components of $\langle \vec{n}_i \rangle$ marked by different colors (shades). In panel (a), the noise $p_{out} = 0.2$ is below the transition point $p_1 = 0.28$. In panel (b), $p_{out} = 0.3$ is above p_1 . Note that panel (a) shows a convergent solution for node i where panel (b) indicates the absence of a collapse.

system veers towards a local (or global) energy minimum. The system exhibits no dynamics once it gets stuck in a local (global) minimum of the energy. For the shown system in Fig. 2.16 , in the absence of perturbing fields, at low noise within the solvable region the system, the node coordinates $\langle \vec{n}_i \rangle$ quickly collapse to the origin. Conversely, at high values of the noise (i.e., large p_{out}) the node coordinates do not converge (and indeed, as we elaborate on below, the system is not solvable). As detailed later in Sec. 2.9.3, we further applied weak perturbing fields $\{\vec{h}_i\}$ and found that they can indeed veer the system, at low noise, towards the correct solutions.

If there is no dissipation in the system (inertial system), the inertial equation

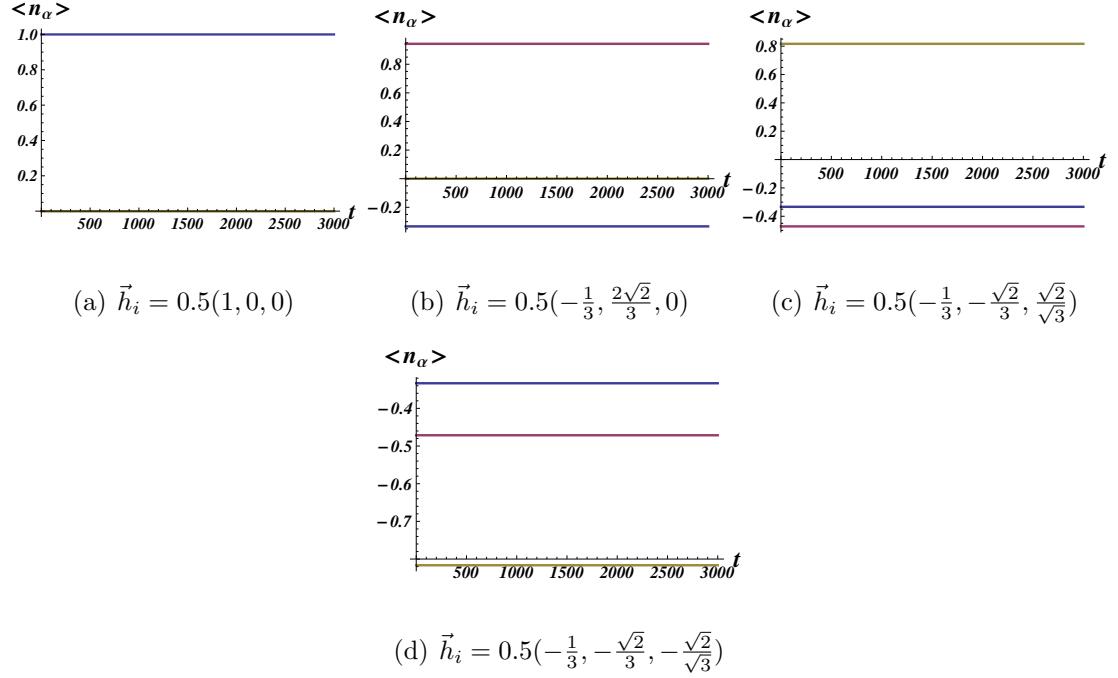


Figure 2.17: The three Cartesian components ($\alpha = 1, 2, 3$) of the node trajectories $\langle n_\alpha \rangle$ (for any node i) under applied fields \vec{h}_i (below each plot) for system of size $N = 24$ with $q = 4$ communities with a noise level of $p_{out} = 0.1$ at a temperature $T = 0.01$. Each Cartesian component is labeled by a different color (shade) in the above figure (In Panel (a), two components are overlapping in the t-axis). The field \vec{h}_i is chosen to be the same as that of the preset cluster membership for node i . The applied fields bias the node trajectories towards the solution of the system.

for the variable η_i if we chose this dynamics instead of our overdamped one that ultimately gets stuck at a random low energy state (unsolvable regime) or maintains regularity (solvable phase) is as follows,

$$M \frac{d^2\eta_i}{dt^2} = -\frac{\delta H_{eff}}{\delta \eta_i}, \quad (2.26)$$

At long times due to dissipation in our overdamped form, the system gets stuck at a given state (either a higher energy metastable state (unsolvable region) or the origin (solvable sans a perturbing field) or to a specific correct solution once weak external fields favoring a particular solution are applied). The inertial form would allow for bona fide chaos also at asymptotically long times.

The system shown in Fig. 2.16 contains only $N = 24$ nodes with $q = 4$ communities at a temperature $T = 0.05$. Prior to investigating this system using the dynamical approach outlined above, we first examined this system also using the entropy/energy/computational susceptibility measures discussed in this chapter and found that, in this system, there is no hard phase. Rather, there is a direct transition (or, more precisely, crossover in this small N system) from an easy solvable phase for $p_{out} < p_1 = 0.28$ to a disordered unsolvable system for $p > p_1$. The result of our dynamic analysis following Eq.(2.28), demonstrates the existence of a phase transition (or crossover for this finite N system) *at precisely the same values of p_{out} found by the analysis of the thermodynamic quantities associated with the Potts model for this small system.* In this case, the dynamics of the nodes illustrate that when p_{out} ex-

ceeds p_1 , the system exhibits a transition from a stable system ($p < p_1$) to one which is chaotic ($p > p_1$). Our dynamic approach to the community detection transition may generally bridge such transitions in system dynamics to thermodynamic phase transitions.

We illustrated via our dynamic approach, how ergodic behavior can arise depending on p_{out} (and, similarly, also on temperature). This relates to “*chaotic*” behavior reflecting the sensitivity to the temperature and in our case other parameters (such as p_{out}) that define the computational problem in spin-glasses [101, 102, 103] to real chaotic behavior of a dynamical system. Further, in our spin-glass approach, Fig. 2.12 illustrates that auto-correlation functions corresponding to different initial conditions (or randomness) remain different up to long times. This sensitive dependence on the initial conditions is the hallmark of chaotic systems. Although, we have not observed such an intermediate hard phase for the small N system that we investigated using this dynamic approach, we speculate the above dynamic transition from more stable orbits to “*chaos*” may, for larger systems, exhibit also indeed an intermediate region corresponding ($p_1(T) \leq p_{out} \leq p_2(T)$ for low T or also $p_3(T) \leq p_{out} \leq p_4(T)$ for higher T) where more and more branching points may appear (or period doubling, etc.) as the system transitions into chaos. Ideas from KAM analysis may, hopefully, be invoked in more sophisticated treatments.

2.9.3 Node trajectories in the presence of a biasing field

As mentioned earlier in the text, effective fields may direct the continuous dynamical system of Section (2.9) towards correct non-trivial solutions. In this brief section, we outline how this is achieved and provide some results.

The dynamical equation for a node moving under the effective field is

$$\begin{aligned} \frac{d\vec{\eta}_i}{dt} &= -\frac{\delta H_{eff}}{\delta \vec{\eta}_i} \Big|_{\vec{h}_i} \\ &= -\beta^{-1} \sum_j \frac{1}{2} (\beta A')_{ij}^{-1} \vec{\eta}_j + \beta^{-1} \frac{\sum_{\vec{n}_i} \vec{n}_i e^{\vec{n}_i(\vec{\eta}_i + \beta \vec{h}_i)}}{\sum_{\vec{n}_i} e^{\vec{n}_i(\vec{\eta}_i + \beta \vec{h}_i)}} \end{aligned} \quad (2.27)$$

Similarly to Section (2.9) yet now with general applied fields, we have

$$\langle \vec{n}_i \rangle \Big|_{\vec{h}_i} = \frac{\delta \ln Z\{\vec{h}\}}{\delta \beta \vec{h}_i} \Big|_{\vec{h}_i} = \frac{\sum_{\vec{n}_i} \vec{n}_i e^{\vec{n}_i(\vec{\eta}_i + \beta \vec{h}_i)}}{\sum_{\vec{n}_i} e^{\vec{n}_i(\vec{\eta}_i + \beta \vec{h}_i)}}. \quad (2.28)$$

As shown in Fig. 2.17, if we choose the perturbation field to favor a preset community membership for each node, i.e., let $\vec{h}_i = \alpha \vec{n}_i$, where α is a small constant value, then within the solvable phase the nodes will be biased towards the corresponding particular partition of the system.

2.9.4 The mapping from the classical overdamped system to the quantum system

The phase transitions that we reported on in the last subsection for overdamped classical systems suggest similar behavior in a corresponding quantum system. Our dynamical equation in Eq. (2.24) can be mapped to the evolution equation for the

probability distribution written as a Schrödinger equation in imaginary time following the method in [104, 105, 106].

$$\partial_t P = -H_{FP}P, \quad (2.29)$$

where H_{FP} is the Fokker-Planck operator and expressed as

$$H_{FP} = -\sum_i \frac{1}{\gamma_i} \frac{\partial}{\partial \mathbf{x}_i} [\nabla_i U_N + T \frac{\partial}{\partial \mathbf{x}_i}]. \quad (2.30)$$

The \mathbf{x}_i and U_N in Eq. (2.30) correspond to $\vec{\eta}_i$ and H_{eff} in our case in Eq. (2.24), respectively. Thus, following the same procedure in [106], the Fokker-Planck operator can be mapped into a Hermitian quantum Hamiltonian,

$$H = \sum_i \frac{\mathbf{p}_i^2}{2m_i} + \sum_i [-\frac{1}{2} \nabla_i^2 U_N + \frac{1}{4T} (\nabla_i U_N)^2]. \quad (2.31)$$

Substituting U_N with H_{eff} and taking the partial differentials with respect to $\vec{\eta}_i$ instead of \mathbf{x}_i , we obtain the quantum Hamiltonian in our case as shown below, Eq. (2.32):

$$\begin{aligned} H = & \sum_i \frac{\mathbf{p}_i^2}{2m_i} + \sum_i [\frac{1}{2}(1 - \langle \vec{n}_i \rangle^2) \\ & + \frac{1}{4T} \sum_{i \neq j, k} [(\beta A')_{ij}^{-1} (\beta A')_{ik}^{-1} \vec{\eta}_j \cdot \vec{\eta}_k + \langle \vec{n}_i \rangle^2 \\ & - \sum_{i \neq j} (\beta A')_{ij}^{-1} \vec{\eta}_j \cdot \langle \vec{n}_i \rangle)], \end{aligned} \quad (2.32)$$

where $\langle \vec{n}_i \rangle$ is given by Eq. (2.25).

Eq.(2.32) provides a mapping between a quantum many-body bosonic system and an overdamped classical systems similar to that which we investigated in the subsections 2.9.1 and 2.9.2. This quantum system may thus exhibit an analog of the transitions found for the classical system. The above mapping can be replicated mutatis mutandis for general classical systems that exhibit the transitions that we found.

We may simplify the above quantum Hamiltonian Eq. (2.32) in the low noise phase, in which case $\langle \vec{n}_i \rangle = 0$ (as in plot Fig. 2.16),

$$H = \sum_i \frac{\mathbf{p}_i^2}{2m_i} - \frac{N}{2} + \frac{1}{4T} \sum_{i \neq j, k} [(\beta A')_{ij}^{-1} (\beta A')_{ik}^{-1} \vec{\eta}_j \vec{\eta}_k]. \quad (2.33)$$

2.10 Conclusion

We reported on disparate high and low temperature spin glass type phase transitions in the community detection problem and, by extension, rather general disordered Potts spin systems. Our investigation involved several complementary approaches and was not confined to systems with a small number of Potts spin flavors or communities. In the community detection setting, similar to other computational problems, phase transitions occur between a solvable and unsolvable region. The solvable region may further split into an “easy” and a “hard” region. We illustrated how thermal “order out of disorder” may come into play in these systems and provided ample

evidence of the spin-glass character of the transitions that occur. Amongst other results, we found that different sorts of randomness can lead to different behaviors, e.g., “chaos”. We introduce a general correspondence between discrete spin systems and mechanical systems with continuous dynamics. With the aid of this mapping, we illustrated that spin glass type transitions in the disordered system correspond to transitions to chaos in the mechanical system. The mapping that we use to relate the thermodynamics to the dynamics suggests how chaotic-type behavior in thermodynamical system can indeed naturally arise in hard-computational problem and spin-glasses. We further briefly speculate on possible physical consequences (such as supercooled liquids and glasses) of the transitions that we find here. Recently, we indeed employed the transitions that we found here in the analysis of such complex physical systems [63, 64] as well as image segmentation [80].

Chapter 3

The stability to instability transition in the structure of large scale networks

3.1 Introduction

Applications of physics to networks [1] has opened fascinating doors for enhancing our understanding of these complex systems. In particular, community detection [10] endeavors to identify pertinent structures within such systems. Applications of the problem are exceptionally broad, and numerous methods have been proposed to attack the problem [3, 15, 17, 18, 19, 22, 23, 107], some of which have been compared for efficiency and accuracy [3, 108, 109, 110].

Computational “phase transitions” have been studied in many challenging prob-

lems [30, 31, 32, 33, 111, 112, 113, 114]. Practical implications of such studies abound (e.g., Refs. [33, 112, 115, 116, 117]), and understanding the behavior of algorithmic solutions to these problems is of interest because the knowledge can be leveraged to understand when a particular solution is computationally challenging, trustworthy, or perhaps not obtainable either via an inherent difficulty or required computational effort. Such knowledge may be used to in certain cases to predict the hard or unsolvable regimes of the problem *a priori* (e.g., k -SAT [30]) or perhaps, more practically in general, to dynamically adapt the solver during the onset of a phase transition [118].

Earlier work related to computational phase transitions with connections to clustering include [119, 120], and Ref. [121] reviewed some critical phenomena in complex networks. The complexity of the energy landscape in community detection was studied for a “fixed” Potts model (model parameters are not set by the network under study) [18, 40], modularity [75], and belief propagation on block models [122]. The former and latter studies explicitly identified phase transitions in the respective systems. We extend a previous analysis [40] of a Potts model where we studied the thermodynamic and complexity character resulting in two distinct transitions: an entropic stabilization transition where added complexity can result in “order by disorder” annealing and a high temperature disordered unsolvable phase. For extreme complexity (high noise) at low T , the system is again unsolvable. Additional transitions can appear between unsolvable and difficult solutions or contending partitions of natural network scales. Here, we seek to move beyond characterizing the solvable/unsolvable transition to study the transitions in terms of changes in the energy landscape and

thermodynamic functions as functions of temperature and “noise” (intercommunity edges).

We utilize overlap parameters in the form of information theory measures and a “computational susceptibility” χ . Using these measures, we monitor increases in the number of local minima corresponding to (often sharply) increased computational complexity. We apply our Potts model to solve a random graph with an embedded ground state, and we identify phase transitions between “easy” and “hard” *solvable* phases which transition into *unsolvable* regions. Specifically, the normalized mutual information (NMI) I_N , Shannon entropy H , the energy E , and χ exhibit progressively sharper changes as the system size N increases suggesting the existence of genuine thermodynamic transitions. Similar analysis can be done for other community detection approaches. Many community detection methods will agree on the best solution within the easy phase, but the hard region presents a substantially more difficult challenge.

The identified transitions may be connected to jamming [116, 117] and avalanche (cascade) transitions [123, 124, 125] in networks. Dynamic jamming transitions occur in traffic, computer network, particulate matter (e.g., sand piles), and the glassy state in amorphous solids may be caused by similar behavior. Refs. [126, 127, 128] showed relations between clustering and cascades in certain networks, and Ref. [129] relates agent dynamics to the Kuramoto oscillators model which has been used for community detection [26]. The threshold emergence of Giant Connected Components (GCC) is related to epidemic thresholds [130, 131], and by nature of the emerging global

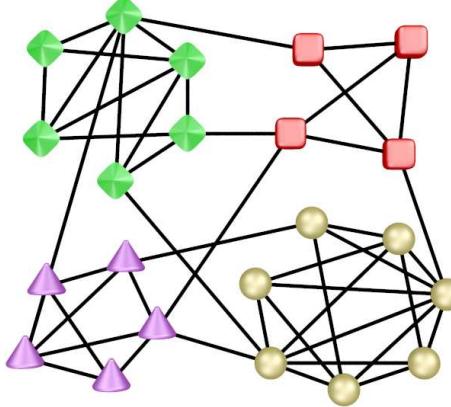


Figure 3.1: The figure illustrates a partition where nodes are separated into distinct communities as indicated by distinct shapes and colors, thus identifying relevant structure in the graph. The current work elaborates on computational transitions and disorder in terms noise (extraneous intercommunity edges) or thermal effects (high temperature T or large system size N) of solving such systems using a stochastic heat bath solver.

connectivity, the GCC is directly detectable via clustering at large-scale resolutions [*i.e.*, small γ in Eq. (3.1)]. Jones polynomials in knot theory are related to Tutte polynomials for the Potts model, so our results suggest similar transitions in random knots (see Sec. 3.10).

We will analytically investigate partition functions and free energies of several graphs in the high temperature T and large number of communities q approximations. We illustrate that increasing T emulates increasing q for a general system, and the analytical results are consistent with the computational phase diagrams.

3.2 Potts Hamiltonian

We employ a spin-glass-type Potts model Hamiltonian for solving the community detection problem

$$H(\{\sigma\}) = -\frac{1}{2} \sum_{i \neq j} [A_{ij} - \gamma(1 - A_{ij})] \delta(\sigma_i, \sigma_j) \quad (3.1)$$

which we refer to as an “Absolute Potts Model” (APM). In general, A_{ij} may be trivially extended to a weighted adjacency matrix w_{ij} (perhaps including “adversarial” relations) [18], but we utilize unweighted graphs in most of the current work (see Sec. 3.6). Each spin σ_i may assume integer values in the range $1 \leq \sigma_i \leq q$ where q is the (dynamic) number of communities where node i is a member of community k when $\sigma_i = k$. In the current work, we set the resolution parameter [27] to $\gamma = 1$ which is near an optimal value for communities with high internal edge densities (see Sec. 3.3).

Previous work [18, 27] elaborated on a “zero-temperature” ($T = 0$) community detection algorithm which we used to minimize Eq. (3.1). A depiction of community structure is shown in Fig. 3.1 where different communities are represented by different node shapes and colors. Here, we investigate the Hamiltonian at non-zero temperatures ($T > 0$) by applying a heat bath algorithm (HBA).

We further invoke s independent solutions (“trials”, see Sec. 2.3) by solving copies of the system which differ by a permutation of the order of the spin indices. This process leads to states that (perhaps locally) minimize Eq. (3.1), so we select the lowest energy trial as the best solution. We vary s in the range $4 \leq s \leq 20$ where we

employ $s = 4$ trials in general and use $s > 4$ trials for calculating the computational susceptibility in Eq. (2.8).

In our multi-scale (“multiresolution”) analysis, we solve $r = 100$ independent “replicas” (see Sec. 2.3) and examine information theory correlations between the replicas and the planted ground state solutions. We schematically show such a set of independent solvers in Fig. 2.1 where stronger agreement among the replicas indicates a more robust solution. We compute the average inter-replica information correlations among the ensemble of replicas allowing us to infer a more detailed picture of the system beyond that of a single optimized solution. Specifically, *information theory extrema* as a function of T and γ (or other scale parameters in general) correspond to most relevant scale(s) of the system.

3.3 Construction of embedded graphs and the noise test

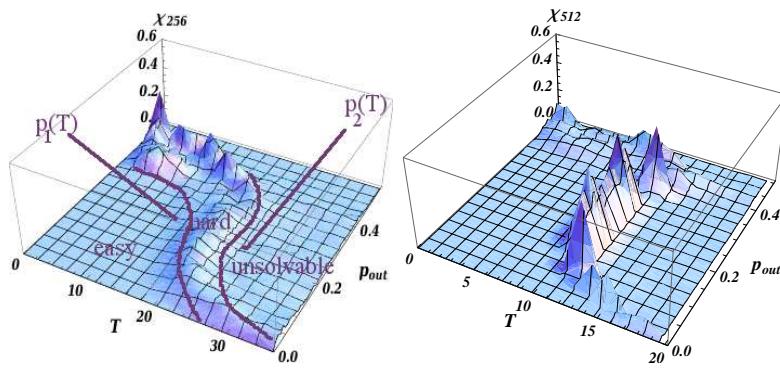
Similar to [3], we construct a “noise test” benchmark as a medium in which to study phase transitions in random graphs with embedded solutions [18, 40]. We define the system “noise” as intercommunity edges that connect a given node to communities other than its original or “best” community assignment. In general [18], it is not possible at the beginning of an attempted solution to ascertain which edges contribute to noise and which constitute edges within communities of the best partition(s).

For each benchmark graph, we divide N nodes into q communities with a power

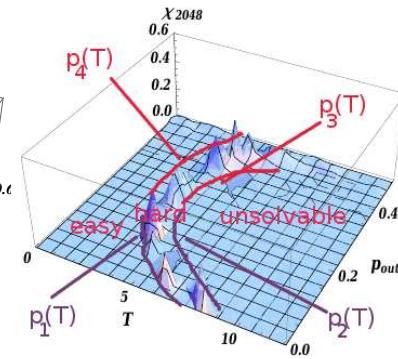
law distribution of community sizes $\{n_i\}$ given by n^β where $\beta = -1$. We then connect “intracommunity” edges at a high average edge density $p_{in} = 0.95$. Initially, the external edge density is zero, $p_{out} = 0$, so that we have perfectly decoupled clusters. To this system, we add random intercommunity edges at a density of $p_{out} < 0.5$. We define p_{in} (p_{out}) as the ratio of the number of intracommunity (intercommunity) edges over the maximum possible intracommunity (intercommunity) edges.

We define the *average external degree* of each node Z_{out} as the average number of links that a given node has with nodes in communities other than its own. Similarly, the *average internal degree* Z_{in} is defined as the average number of links to nodes in the *same* community, and $Z_{in} + Z_{out} = Z$ where Z is the average coordination number. Then we can explicitly write the internal and external edge densities as shown in Eq. (2.9) and Eq. (2.10).

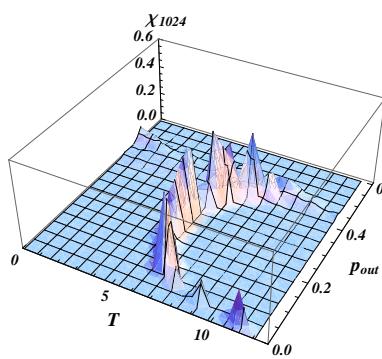
The communities in this construction are well defined, on average, at reasonable levels of noise ($p_{out} \lesssim 0.3$ depending on the typical community size n). As external links are progressively added to the system (p_{out} increases), the communities become increasingly difficult to detect. At some stage, enough noise is added and p_{out} is sufficiently high that the planted partition cannot be detected despite the fact that the optimal ground state is still well-defined. This transition often occurs sharply, particularly for large networks. We investigate the phase transition from the solvable to unsolvable phases at both low and high temperatures by means of the heat bath algorithm.



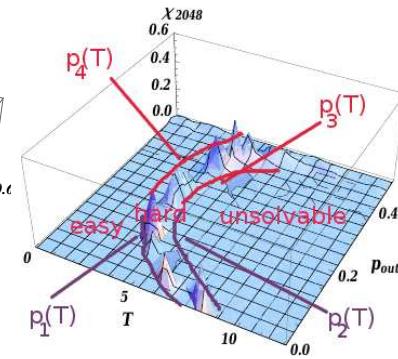
(a) $N = 256, q = 4, \alpha = 0.016$



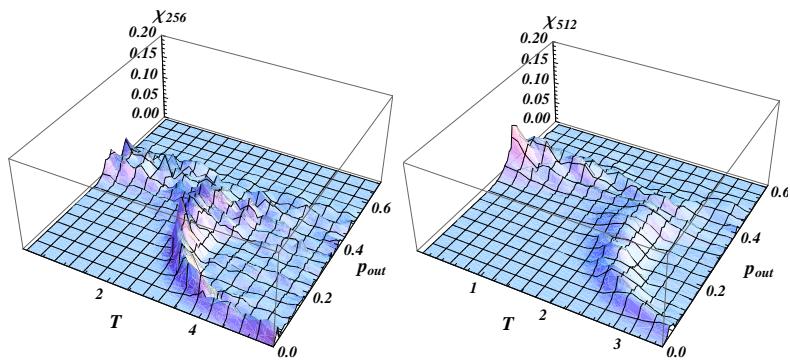
(b) $N = 512, q = 8, \alpha = 0.016$



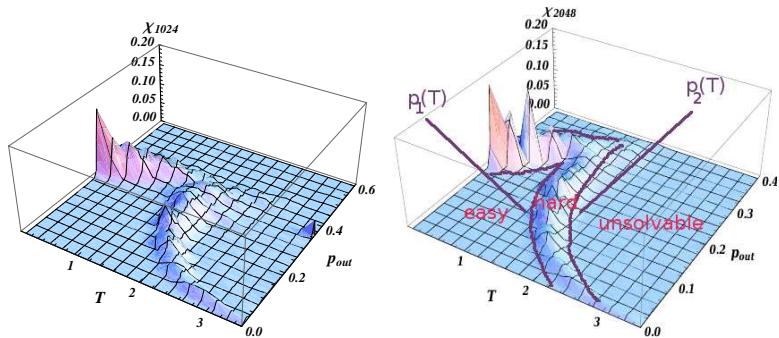
(c) $N = 1024, q = 16, \alpha = 0.016$



(d) $N = 2048, q = 32, \alpha = 0.016$



(e) $N = 256, q = 18, \alpha = 0.07$ (f) $N = 512, q = 35, \alpha = 0.07$



(g) $N = 1024, q = 70, \alpha = 0.07$ (h) $N = 2048, q = 140, \alpha = 0.07$

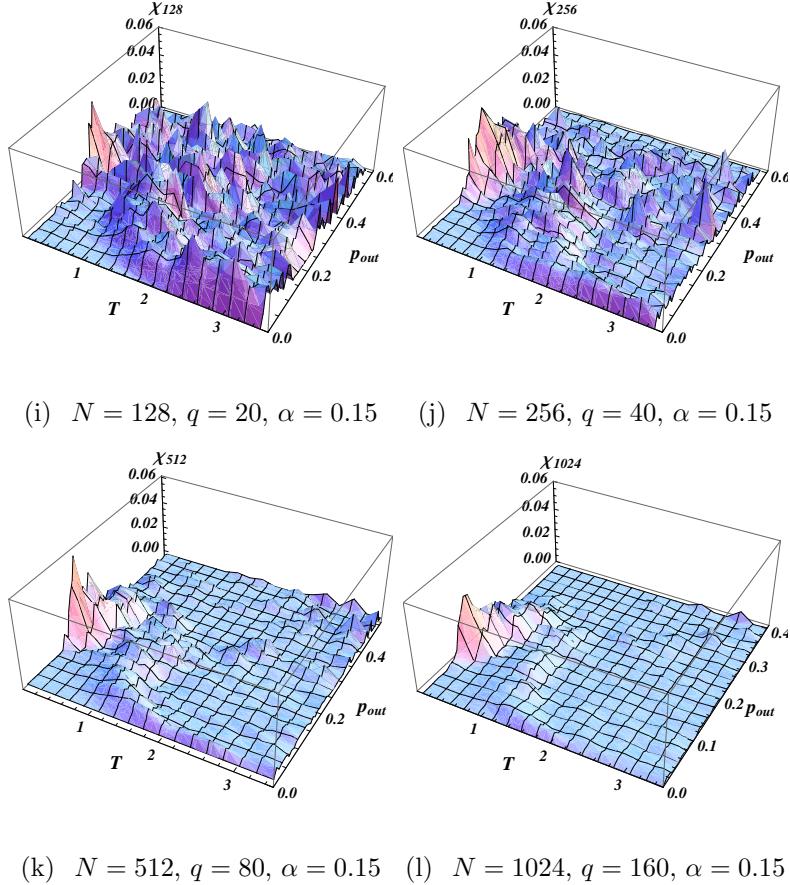


Figure 3.2: Each panel shows a 3D plot $\chi(T, p_{out})$ as a function of temperature T and noise level p_{out} for systems with the indicated number of nodes N , communities q , and $\alpha = q/N$ ratio. In panels (a–h) for $\alpha = 0.016$ and 0.07 , all plots show three clear phases, and the “ridges” at low and high temperatures mark the hard phase. In panels (a)–(h), the ridges in $\chi(T, p_{out})$ become narrower as N increases. The area of the easy (hard) regions decreases (increases) from panel (a) to (d) and (e) to (h), respectively. In panels (a–d) for $\alpha = 0.016$, the hard phase at low temperature becomes less prominent from panel (a) to (d), but it becomes more prominent at high temperature. In panels (e–h) for $\alpha = 0.07$, the hard phase at low temperature becomes more prominent from panel (e) to (h), but it remains constant at high temperature. In panels (i–l) for $\alpha = 0.15$, only the larger systems with $N \geq 512$ show clear phases. The smaller systems with $N = 128$ in panel (i) and $N = 256$ in panel (j) show very noisy phases where only the easy phase can be readily determined,

3.4 Spin glass type transitions

We previously reported [40] on the existence of *two* spin-glass-type transitions in the constructed graphs mentioned in Sec. 3.3. Evidence for the transitions are observed in several measures such the accuracy of the solution obtained by means of the APM in Eq. (3.1) (and other models [18, 27] in general), the computational effort required to converge to a solution [18, 27], entropy effects, and others. Compared to another Potts-type quality function [18] utilizing a “null model” (a random graph used to evaluate the quality of a candidate partition), the APM exhibits a somewhat sharper transition as N is increased [27]. As alluded to above, two transitions are generally encountered as the noise value (or temperature) is increased. At fixed temperature T , as p_{out} is steadily increased from zero, the first onset of spin glass behavior first appears for values $p_1 \leq p_{out} \leq p_2$.

Fig. 2.6(a) illustrates a one dimension characterization of the easy and hard phases in terms of the level of noise (extraneous intercommunity edges) encountered by a greedy solver. It is in this context that greedy algorithms are, in general, more easily trapped in local energy minima above a certain noise threshold. Stochastic solvers such a heat bath algorithm discussed in or simulated annealing (SA) enable one to circumvent noise to some extent, but excessive levels will even thwart these more robust solvers because meaningful information is eventually obscured by the complexity of the energy landscape. Fig. 2.6(b,c) depict the easy and hard phases at low and high temperatures T , respectively, for our HBA. Above a graph-dependent

threshold, the solver is insensitive to local features, and it is unable to find an accurate solution.

We showed that Eq. (3.1) is robust to noise [27] leading to exceptional accuracy even with a greedy algorithm. Some other methods and cost-functions [15] have also proven to be very accurate [3] with a greedy-oriented algorithm. While maximizing modularity [12] and a closely related cost function in [18] have proven to be accurate and productive, Refs. [3, 10, 75] have discussed problems associated with maximizing modularity in community detection. We briefly illustrated [18] a correspondence between the major transition experienced by Eq. (3.1) and a Potts model in [27]. We conjecture the existence of a related transition for random knots in Sec. 3.10.

In Secs. 3.4.1 and 3.4.2, we elaborate on the transitions using a “*computational susceptibility*” χ as defined in Eq. (2.8) (see also [18, 40]). In analogy with other physical susceptibility parameters, χ measures the response of the system to additional optimization effort. In the presence of multiple local minima [e.g., Fig. 2.2(b)], more trials may increase the accuracy. A higher χ indicates a more disordered, but navigable, energy landscape whereas a low χ indicates that additional optimization has less effect whether due to extreme disorder or a trivially solvable system. Finally in Sec. 3.4.3, we illustrate the transitions using additional stability measures.

3.4.1 $\chi(T, p_{out})$ at fixed $\alpha = q/N$

We show the phase transitions in terms of three-dimensional (3D) plots with the computational susceptibility $\chi(T, p_{out})$ for a range of system sizes N and numbers of

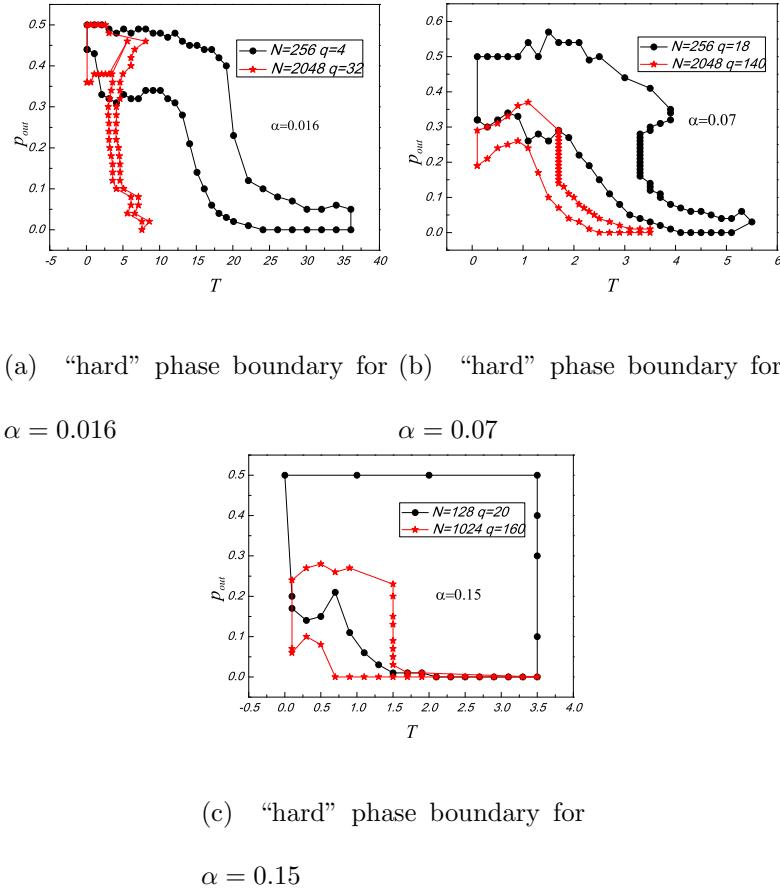


Figure 3.3: Corresponding to Fig. 3.2 and Sec. 3.4.1, each plot depicts the boundaries of the hard phase for the system series with a fixed $\alpha = q/N$ ratio. Panels (a), (b), and (c) show the results for $\alpha = 0.016$, $\alpha = 0.07$, and $\alpha = 0.15$, respectively. System sizes range from $N = 256$ to 2048 , and q varies from 4 to 160 as indicated in each plot. For each α , the area within hard phase boundary becomes progressively narrower indicating that the transitions from the easy to unsolvable phases are more clear in the thermodynamic limit.

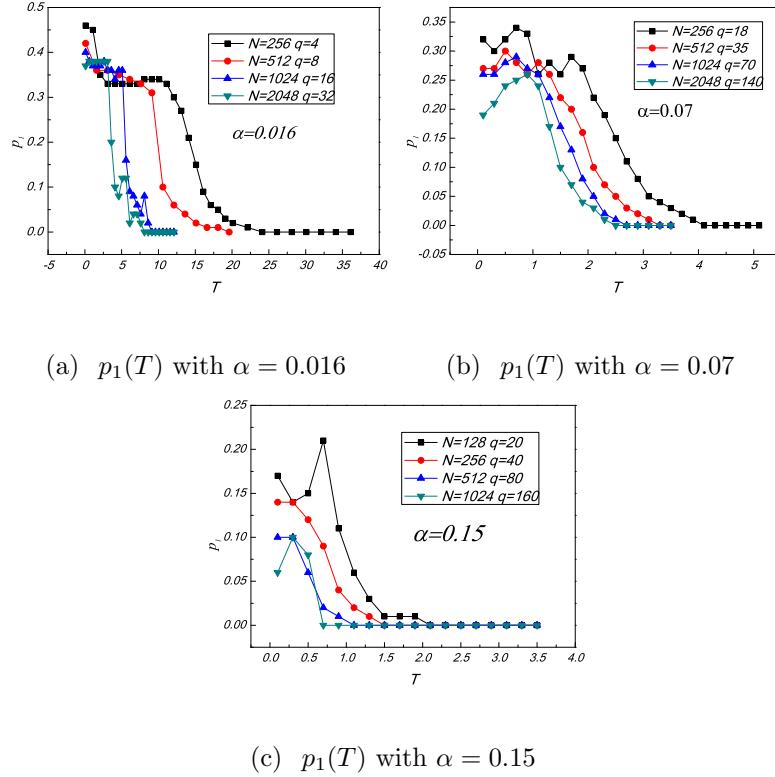


Figure 3.4: Corresponding to Fig. 3.2 and Sec. 3.4.1, each plot depicts the first phase transition point p_1 as a function of the temperature T for systems with a fixed ratio of $\alpha = q/N$. Panels (a), (b), and (c) show the results for $\alpha = 0.016$, $\alpha = 0.07$, and $\alpha = 0.15$, respectively. System sizes range from $N = 256$ to 2048 , and q varies from 4 to 160 as indicated in each plot. All panels show that when α is fixed, the value of the first transition point p_1 decreases as the system size increases. This behavior further indicates that the system becomes more complex to solve in the thermodynamic limit.

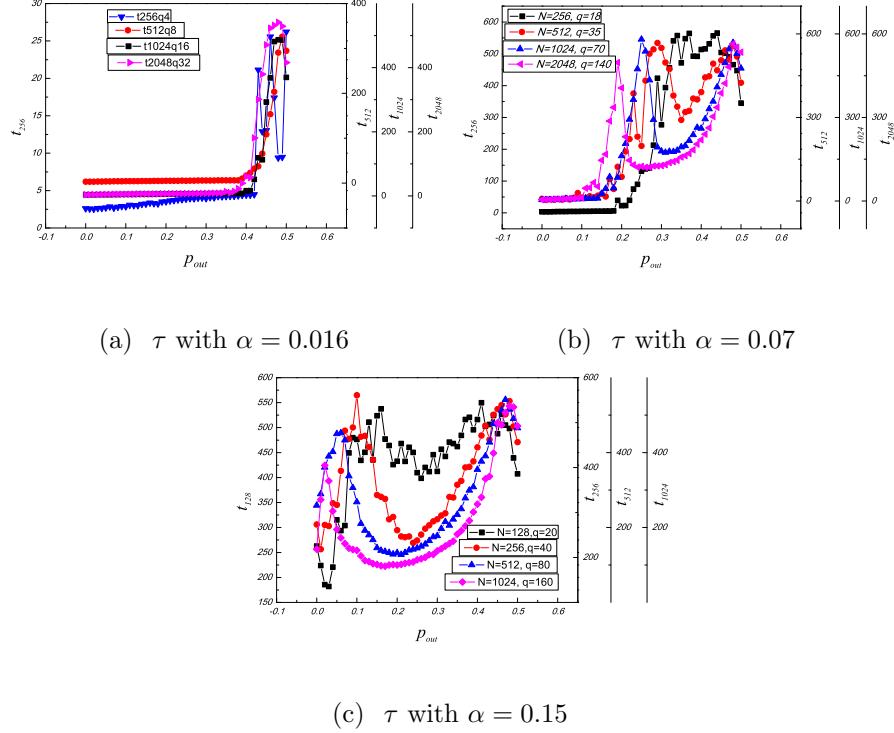
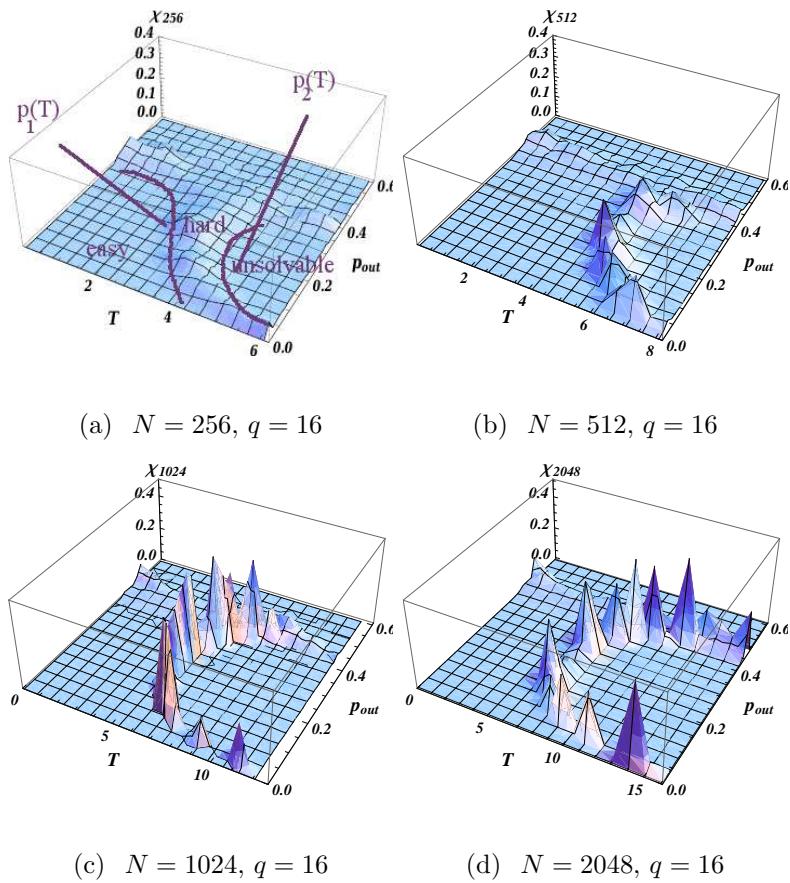
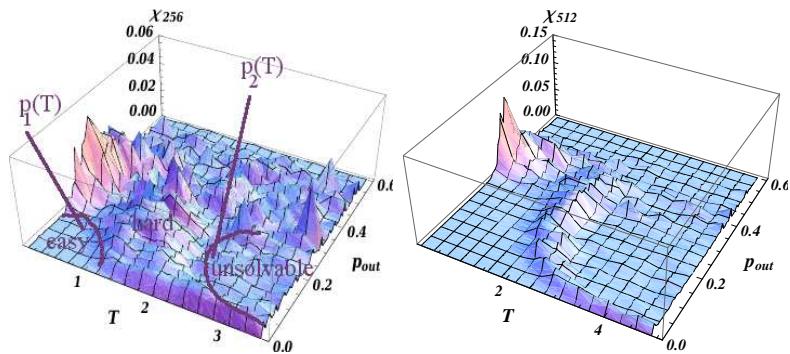
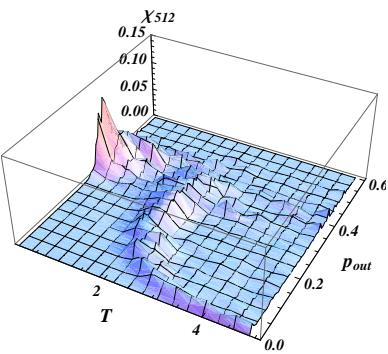


Figure 3.5: Corresponding to Fig. 3.2 and Sec. 3.4.1, the convergence time τ [see Fig. 2.2(a)] as a function of noise p_{out} at zero temperature for the systems with a fixed ratio of $\alpha = q/N$. Panels (a), (b), and (c) show the results for $\alpha = 0.016$, $\alpha = 0.07$ and $\alpha = 0.15$, respectively. System sizes range from $N = 256$ to $N = 2048$, and q varies from 4 to 160 as indicated in each plot. The noise level p_{out} at the first peak of the convergence time corresponds to the first transition point p_1 in Fig. 3.4 at zero temperature. As the system size increases, the first peak in the convergence time moves to the left. They share the same trend as in Fig. 3.3 and Fig. 3.4.

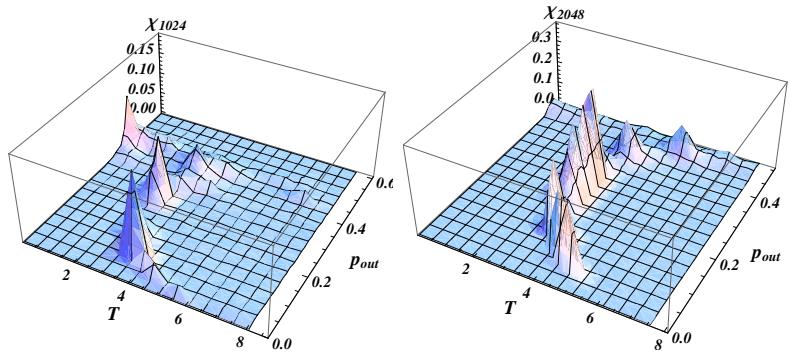




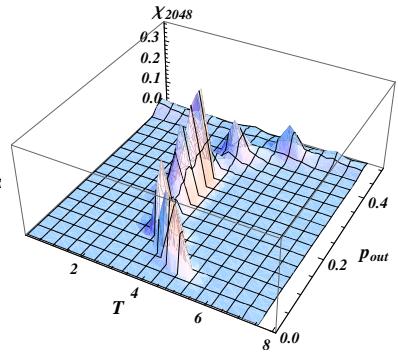
(e) $N = 256, q = 40$



(f) $N = 512, q = 40$



(g) $N = 1024, q = 40$.



(h) $N = 2048, q = 40$.

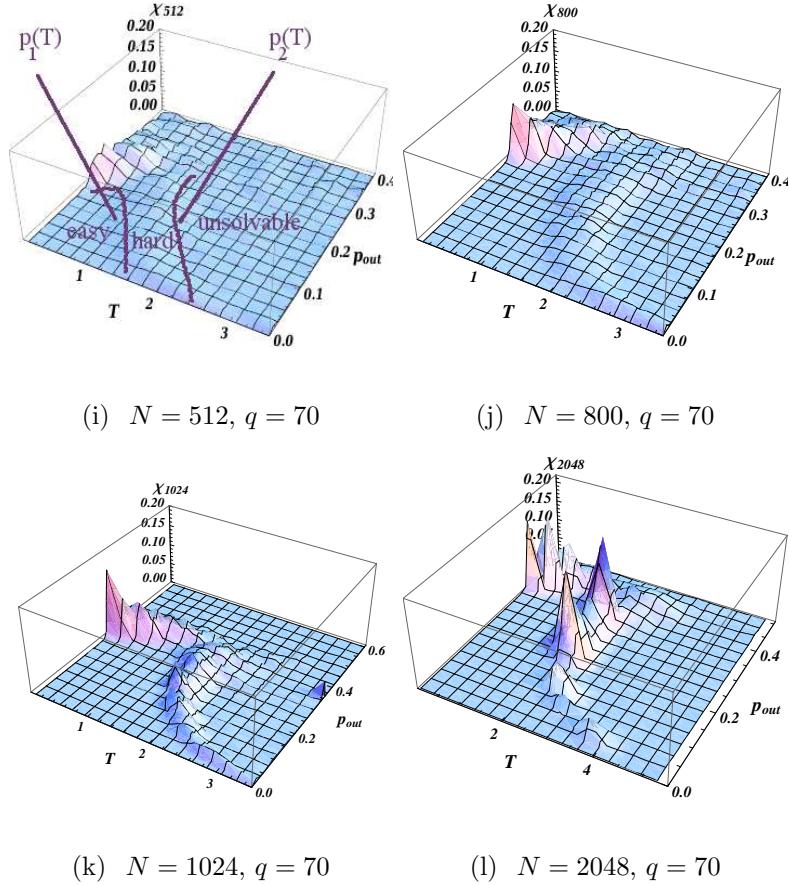


Figure 3.6: Similar to Fig. 3.2, we plot of $\chi(T, p_{out})$ as a function of temperature T and noise level p_{out} for systems with the indicated number of nodes N , communities q , and $\alpha = q/N$ ratio. Here, q is fixed for each row series, and we vary α (rows) to examine the behavior as N increases (columns). The heights of the susceptibility peaks at higher T increase across each series as N increases whereas the heights at low T are relatively constant. The width of the hard phase in high temperature also decreases as N increases, as shown in panels (j) (width $\Delta T^j = 1$) and (l) ($\Delta T^l = 0.5$), also in Fig. 3.7. The $N = 256$ node systems do not show clear hard or unsolvable phases, but the transitions are strong at high temperature for most panels in the second and third columns of plots.

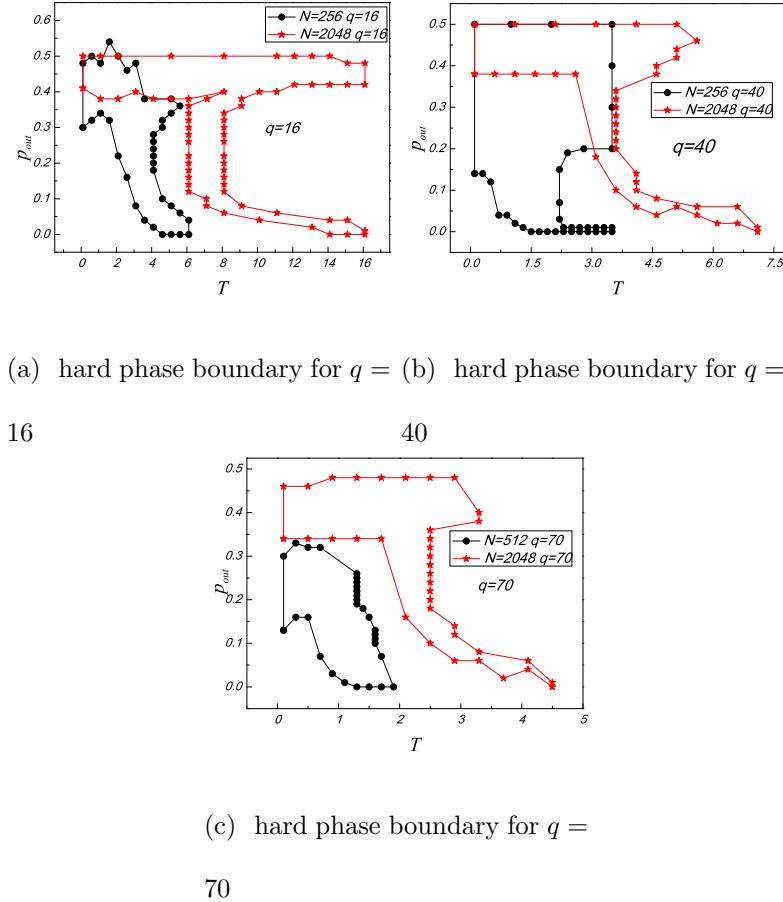


Figure 3.7: Corresponding to Fig. 3.6 and Sec. 3.4.2, each plot depicts the boundaries of the hard phase for the system series with a fixed number of communities q where panels (a), (b), and (c) correspond to $q = 16, 40$, and 70 , respectively. System sizes range from $N = 256$ to 2048 as indicated. For each q , the area of the hard phase becomes progressively narrower which indicates clearer transitions from the easy to unsolvable phases in the thermodynamic limit.

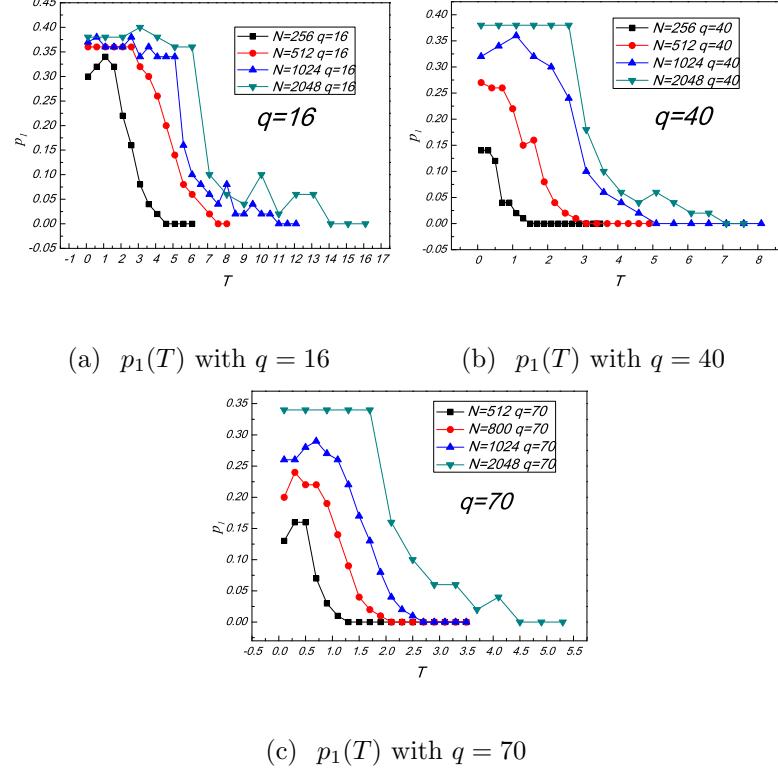


Figure 3.8: Corresponding to Fig. 3.6 and Sec. 3.4.2, each plot depicts the first phase transition point p_1 as a function of temperature T for systems with a fixed q . Panels (a), (b), and (c) show the results for $q = 16$, 40, and 70, respectively. System sizes range from $N = 256$ to 2048 as indicated in each plot. All panels show that the first transition point increases as the system size increases which is consistent with the complexity trend of the system series.

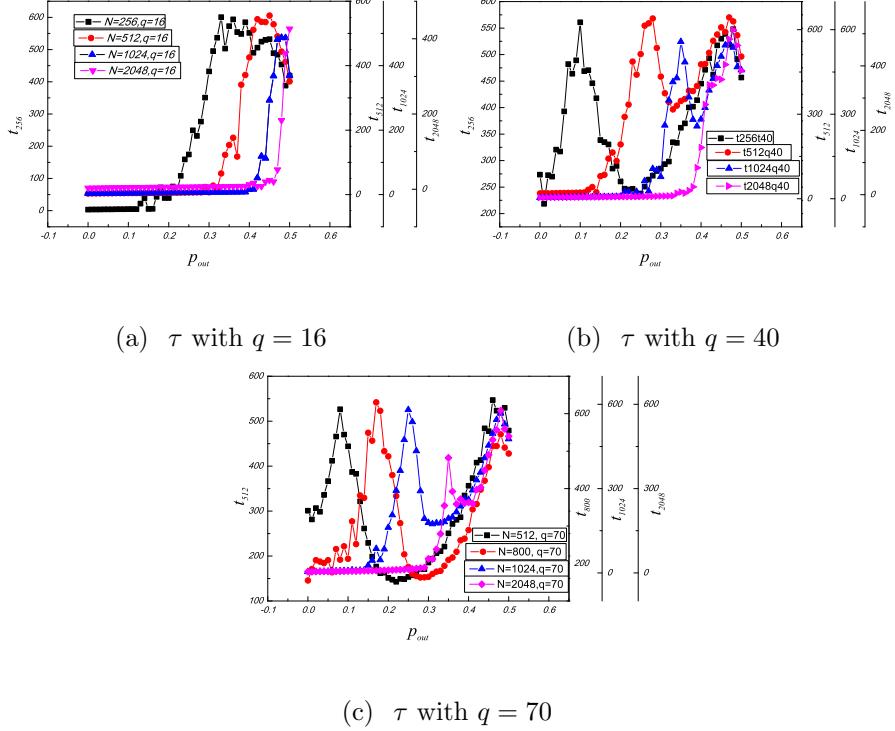


Figure 3.9: Corresponding to Fig. 3.6 and Sec. 3.4.2, the convergence time τ [Fig. ??(a)]

as a function of noise p_{out} for systems with fixed q . Panels (a), (b) and (c) show the results for $q = 16$, 40, and 70, respectively. System sizes vary from $N = 256$ to 2048 as indicated in each plot. The noise level p_{out} at the first peak of the convergence time corresponds to the initial transition point p_1 in Fig. 3.8 at zero temperature. As the system size increases, the first peak in the convergence time moves to the right. They share the same trend as in Figs. 3.7 and 3.8.

communities q . First, we fix the ratio $\alpha = q/N$ and study the phase transitions as N increases. Then we test a range of systems with fixed q as N increases.

$\chi(T, p_{out})$ at $\alpha = 0.016$

In Fig. 3.2 panels (a) through (d), we begin the analysis at a small $\alpha = q/N = 0.016$ ratio. The results for four system sizes are shown: $N = 256$, $N = 512$, $N = 1024$ and $N = 2048$ which maintain a fixed ratio of α across the respective rows. Each plot shows the “easy”, “hard”, and “unsolvable” phases.

The two “ridges” in each plot denote the hard phases. Generally, we may have two sets of spin-glass-type transitions into the hard phase as p_{out} or T is increased. For a given temperature T , hard phases generally appear for $p_1 \leq p_{out} \leq p_2$ (as marked in panels (a), (d), and (h) of Fig. 3.2). As we will re-iterate and allude to later, in some cases, these transitions correspond to an annealing effect wherein increasing the noise and/or temperature improves the accuracy of the solutions. Generally, higher temperature transitions may also appear [see, e.g., the V shaped ridge in panel (h) of Fig. 3.2 wherein p_1 and p_2 are non-monotonic in T]. As is similarly vividly seen in panel (d), the hard phase boundaries may become multi-valued and “bow” as a function of p_{out} ; such a situation corresponds to a re-entrant transition whence the system goes from the unsolvable disordered phase via a hard phase extending from $p_1(T) < p_{out} < p_2(T)$ to the ordered easy phase and then, via a second higher noise hard phase ($p_3(T) < p_{out} < p_4(T)$) back to the unsolvable disordered phase as p_{out} is increased at a fixed temperature T . Later on, we will corroborate the location of these

phase boundaries by the use of other metrics including disparate information theoretic measures and thermodynamic quantities such as the entropy and energy. For a fixed value of the ratio of q/N , the height of the first ridge at low temperature decreases as the system size N increases while the height of the second ridge at high temperature increases in the same process. This finite size scaling behavior for the hard phase at high temperature indicates that the phase transition at high temperature exists in the thermodynamic limit. However, the phase transition at low temperature will disappear in the same limit. In the meantime, the ridge in the high temperature will gradually expand into the low temperature region as the system size increases. Thus, for the systems with the small ratio of α , the phase transition will exist in almost the entire temperature range in the thermodynamic limit (see Sec. 3.5).

The “easy” phase shrinks and the unsolvable phase expands as q increases for fixed $\alpha = \frac{q}{N}$. The approximate area of the easy phase on the left corner in panel (a) is in the range of $T \in (0, 20)$ and $p_{out} \in (0, 0.4)$. The area of the unsolvable phase on the right upper corner is in the range of $T \in (20, +\infty)$ and $p_{out} \in (0, 0.4)$. As the system size increases from $q = 4$ in panel (a) to $q = 16$ in panel (c) for a fixed $\alpha = \frac{q}{N}$, the area of the easy phase shrinks to the range of $T \in (0, 5)$ and $p_{out} \in (0, 0.4)$ while the unsolvable phase expands to $T \in (5, +\infty)$ and $p_{out} \in (0, 0.4)$. As the number of communities further increases to $N = 32$ in panel (d), the easy phase further shrinks to the range of $T \in (0, 4)$ and $p_{out} \in (0, 0.4)$ while the unsolvable phase expands to $T \in (4, +\infty)$ and $p_{out} \in (0, 0.4)$. We note that the range of p_{out} for the easy phase does not decrease as the number of communities increases.

Low temperature transitions correspond to an “order by disorder” type effect where fluctuations render the system ordered or solvable. As the temperature T increases, the system may veer towards its global minimum by annealing. In such cases, by virtue of entropic fluctuations, quenching is thwarted and the system may probe low lying states and achieve an ordered configuration.

In order to track the range of the hard phases, we further display a set of “boundary” plots in Fig. 3.3 as well as the first transition point p_1 as the function of temperature in Fig. 3.4. For the system series with the fixed $\alpha = 0.016$ discussed above, the 2D “hard phase” boundaries and the values of the first transition points are in panel (a) of Fig. 3.3 and Fig. 3.4, respectively.

In Fig. 3.3(a), the area of the hard phase shrinks, and its area at high temperature becomes narrower as the system size increases. Specifically, the width of the hard phase for $N = 256$ is about $\Delta T = 6$, while it only extends to $\Delta T = 1$ for the $N = 2048$. Together with the 3D phase diagrams in panels (a)–(d) of Fig. 3.2, we conclude that the hard phase at the high temperature becomes sharper in the thermodynamic limit.

The boundaries of the hard phase at low temperature are more easily seen in Fig. 3.4(a) where we plot the first transition point p_1 as the function of temperature T for a range of systems. The plots confirm the observations in Fig. 3.2(a)–(d) regarding the constant p_{out} range. That is, the range of p_{out} for the easy phase does not decrease as the system size increases [in Fig. 3.4(a), p_1 collapses before $T \leq 5$ for all the systems]. This behavior hints that the first transition point p_1 at low temperature and small α

remains constant in the thermodynamic limit.

As depicted in Fig. 2.2(a), the convergence time τ provides another view of the phase transition. We plot τ as a function of noise level p_{out} in Fig. 3.5(a) for systems with a fixed ratio of $\alpha = q/N = 0.016$. The value p_{out} at the first peak of the convergence time in each system is consistent with the first transition point p_1 observed in Fig. 3.4(a). As the number of communities q increases, the peak convergence time shifts to the left, which corresponds to the lower value of p_1 .

For a given noise level p_{out} , the highest temperature at which the system is still solvable tends to zero as the number of communities q increases. As we will demonstrate in this work for myriad examples (see also [132]), the critical temperature beyond which the system is unsolvable is $T^* \approx 1/\log q$. As the number of communities $q \rightarrow \infty$, the entire system becomes unsolvable.

$\chi(T, p_{out})$ at $\alpha = 0.07$

For $\alpha = 0.07$, the phase transitions are presented in Fig. 3.2 panels (e) through (h). The phases in panel (e) are noisy compared to panels (f) through (h), and all of the systems are more complicated than the plots with $\alpha = 0.016$. As N increases, the phase transitions become more clear. However, contrary to panels (a) through (d), the phase transition at low temperature becomes more prominent as N increases, and the transition at high temperature stays roughly constant. Specifically, the height of the susceptibility peak at low temperature increases from $\chi = 0.01$ at $N = 256$ in panel (e), $\chi = 0.05$ at $N = 512$ in panel (f), $\chi = 0.1$ for $N = 1024$ in panel (g), and

finally reaches $\chi = 0.2$ in panel (h) with $N = 2048$. The phase transitions in this series appear to be persistent.

The easy phase (lower left of each panel) decreases in area as the system size increases. This is the same trend that was observed in the previous $\alpha = 0.016$ series implying that the easy phase will tend to decrease in the thermodynamic limit up to a threshold (see Sec. 3.5). Specifically, the easy phase in the smallest system in panel (e) covers the range of $T \in (0, 3)$ and $p_{out} \in (0, 0.3)$ while in the large system in panel (h) covers $T \in (0, 1.5)$ and $p_{out} \in (0, 0.2)$. The range for p_{out} in the easy phase decreases as the N increases which differs from the $\alpha = 0.016$ data where the noise p_{out} stayed at a roughly constant range of $p_{out} \in (0, 0.4)$. In both series for $\alpha = 0.016$ and 0.07, the value of the initial transition point p_1 decreases in the thermodynamic limit.

The corresponding 2D plots of the hard phase boundaries and the first transition points p_1 are displayed in Fig. 3.3(b) and Fig. 3.4(b), respectively. For the series with $\alpha = 0.07$ in Fig. 3.3(b), the area of the hard phase becomes narrower at both low and high temperatures as the system size increases. In detail, the width of the hard phase for $N = 256$ is about $\Delta T = 1.3$, while the width shrinks to about $\Delta T = 0.3$ at $N = 2048$. Together with the 3D phase diagrams in Fig. 3.2(e)–(h), the phase transitions become sharper in the thermodynamic limit.

As shown in Fig. 3.4(b), the first transition point p_1 decreases as the system size increases, even in the low temperature limit. This is consistent with the first peak of the convergence time τ at zero temperature in Fig. 3.5(b). This indicates that the

system becomes progressively harder to solve in the thermodynamic limit over the whole temperature range.

$\chi(T, p_{out})$ at $\alpha = 0.15$

In panels (i) through (l) of Fig. 3.2, $\alpha = 0.15$ and the clusters are smaller on average resulting in systems that are more difficult to solve. In panels (i) and (j), almost the entire region is covered by small peaks which indicates mixing of the hard and unsolvable phases thus making the phase boundaries hard to detect.

The flat easy regions are recognizable in all panels, but the area is small relative to the previous cases and becomes even smaller as N increases into panel (l). In panel (i), the flat easy region is roughly triangular with legs along $T \in (0, 1.5)$ and $p_{out} \in (0, 0.2)$. The easy region shrinks to a smaller triangle along $T \in (0, 0.2)$ and $p_{out} \in (0, 0.2)$ in panel (j) and (k). In panel (l), it further shrinks to $T \in (0, 1)$ and $p_{out} \in (0, 0.1)$. The easy phase shrinks for both p_{out} and T as N increases which further indicates that the initial transition point p_1 decreases substantially in the thermodynamic limit.

The corresponding plots of the hard phase boundaries and the first transition points p_1 are displayed in Figs. 3.3(c) and 3.4(c), respectively. From Fig. 3.3(c), the area of the hard phase shrinks in the thermodynamic limit. The hard phase is more identifiable relative to the unsolvable region as N increases. The initial transition point p_1 drops as N increases as shown in Fig. 3.4(c). The convergence time τ for the systems with the fixed ratio of $\alpha = q/N = 0.15$ at zero temperature is shown

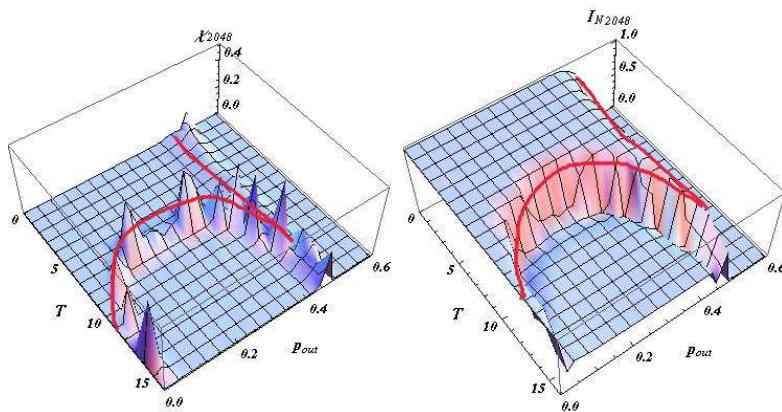
in Fig. 3.5(c) where the first peak of τ shifts to the left as the system size increases. This is consistent with the trend observed in Fig. 3.4(c). We further show in Fig. 3.11 and Fig. 3.12 that the first transition points in “computational susceptibility”, energy, entropy, convergence time and normalized mutual information are consistent with each other.

In Fig. 3.13, we provide plots of scaled waiting correlation function data which clearly indicate *spin glass type collapse*. The collapse is best at the center of the computational susceptibility ridge Fig. 3.13(b). The collapse persists up to the ends of the susceptibility ridge (e.g., $p_{out} = p_1$ in Fig. 3.13(a)) and is no longer valid outside the susceptibility ridge (e.g., $p_{out} = 0.26 > p_2 = 0.24$ in Fig. 3.13(c)).

3.4.2 $\chi(T, p_{out})$ at fixed q

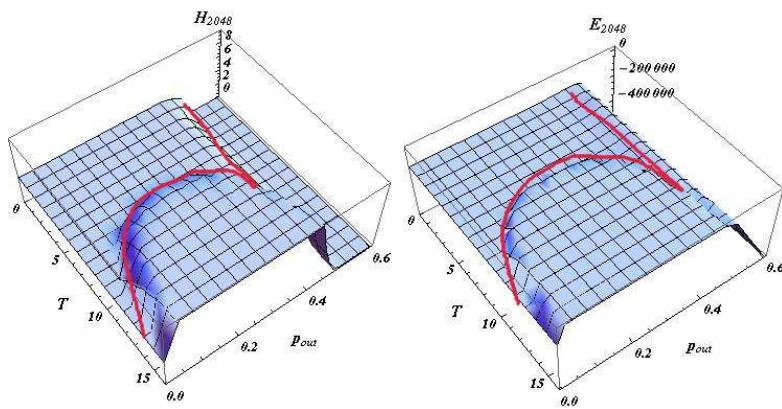
We fix the number of communities at $q = 16, 40$, or 70 and increase the system size N from 256 to 2048 . The plots of computational susceptibility $\chi(T, p_{out})$ for $q = 16$ series of systems are shown in panels (a) through (d) of Fig. 3.6. As in Sec. 3.4.1, the ridges indicate hard phases which become more prominent as N increases while the ridges at low temperature remain at relatively low constant values.

The areas of the easy phases on the lower left corner expand as the system size increases from panel (a) to (d). This trend of increasing area is the reverse of the behavior in the fixed α systems systems in Sec. 3.4.1. This is easy to understand since, q increases with N here, and the high internal edge density p_{in} causes the larger clusters to be more strongly defined.



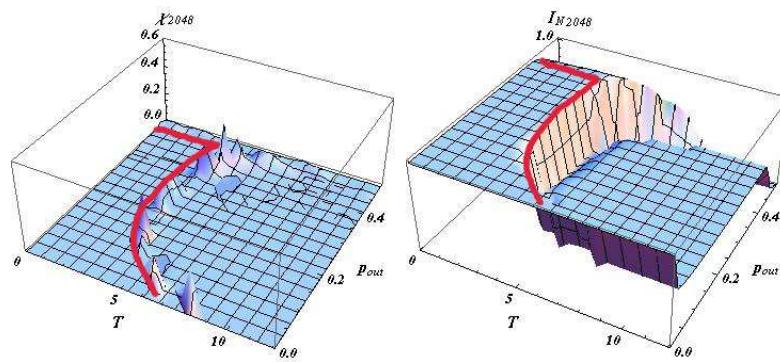
(a) $\chi(T, p_{out})$ for $q = 16$

(b) $I_N(T, p_{out})$ for $q = 16$



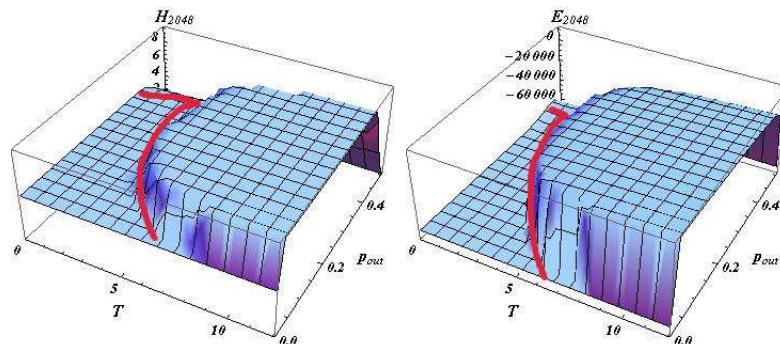
(c) $H(T, p_{out})$ for $q = 16$

(d) $E(T, p_{out})$ for $q = 16$



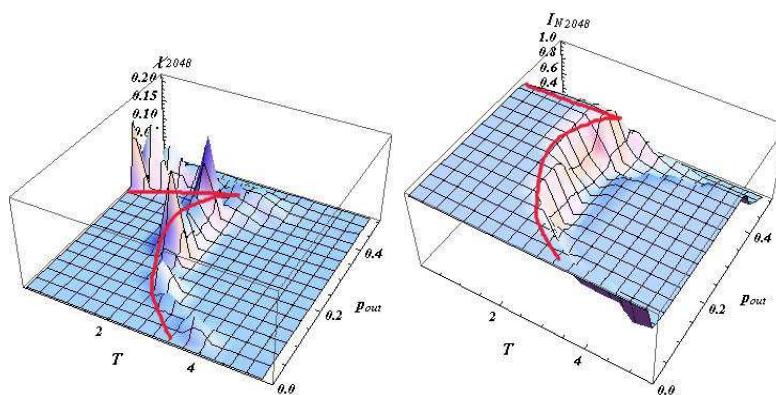
(e) $\chi(T, p_{out})$ $q = 32$

(f) $I_N(T, p_{out})$ for $q = 32$



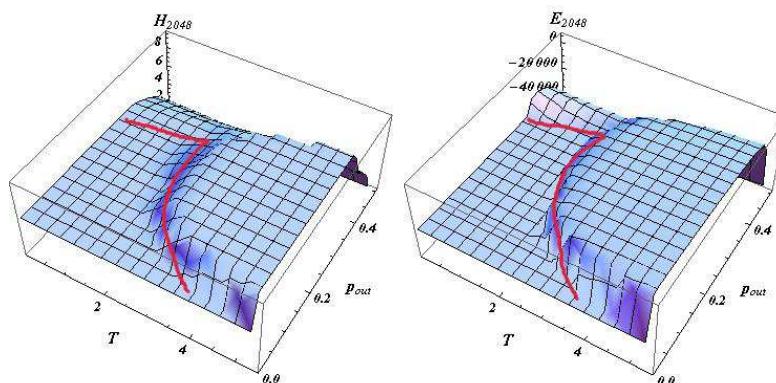
(g) $H(T, p_{out})$ for $q = 32$

(h) $E(T, p_{out})$ for $q = 32$



(i) $\chi(T, p_{out})$ for $q = 70$

(j) $I_N(T, p_{out})$ for $q = 70$



(k) $H(T, p_{out})$ for $q = 70$

(l) $E(T, p_{out})$ for $q = 70$

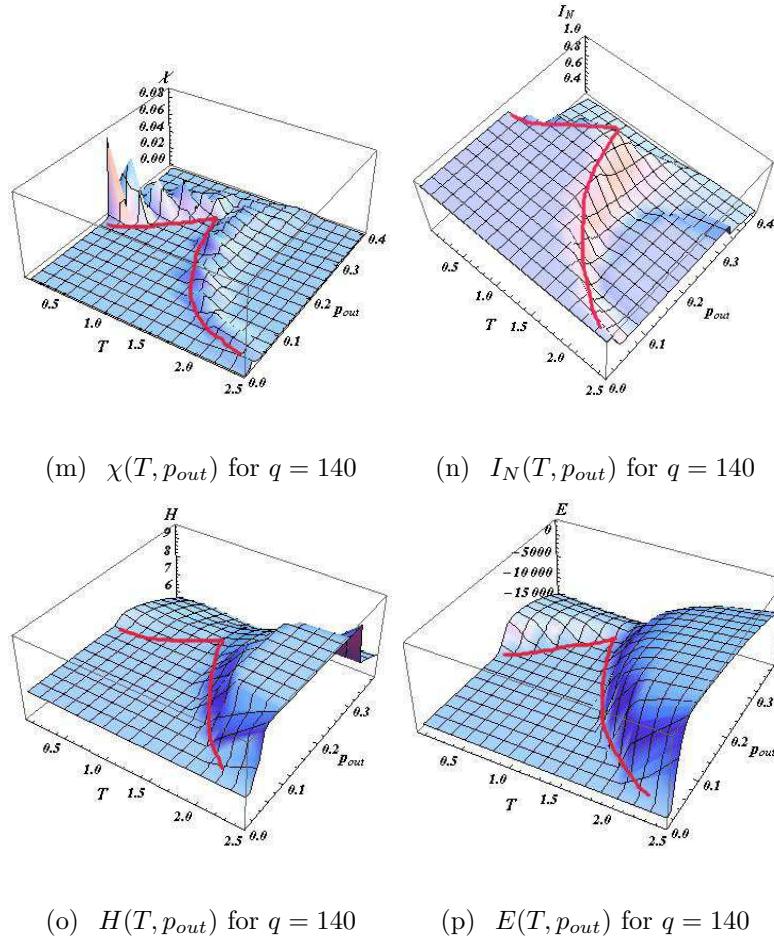


Figure 3.10: Plots of the computational susceptibility χ (column one), NMI I_N (column two), Shannon entropy H (column three), and energy E (column four) as functions of temperature T and intercommunity noise p_{out} . System sizes all use $N = 2048$, and q varies from 16 to 140 in different rows. All plots show the easy, hard, and unsolvable phases often by rapid shifts in the respective measures. The marked (red, color online) line highlights the onset of the hard phase at $p_{out} = p_1(T)$ in each measured quantity where we note that the boundaries match well across each row.

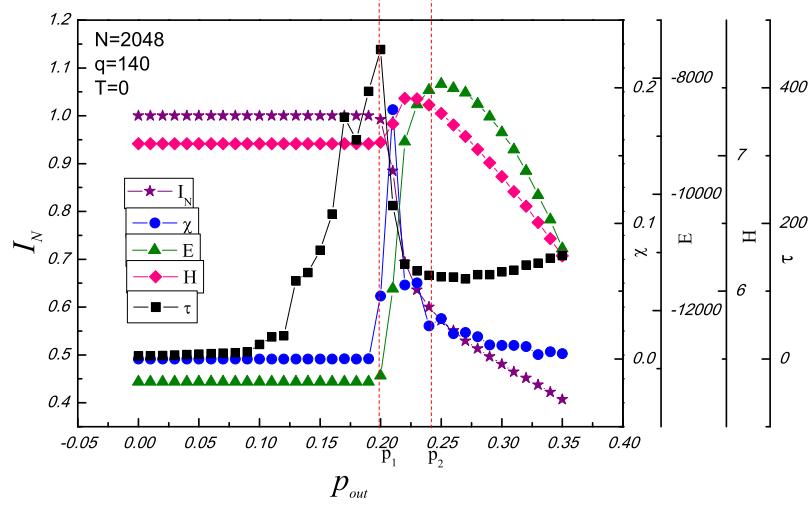
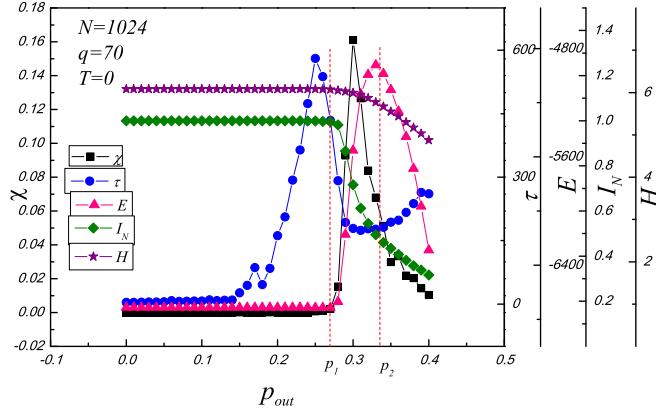
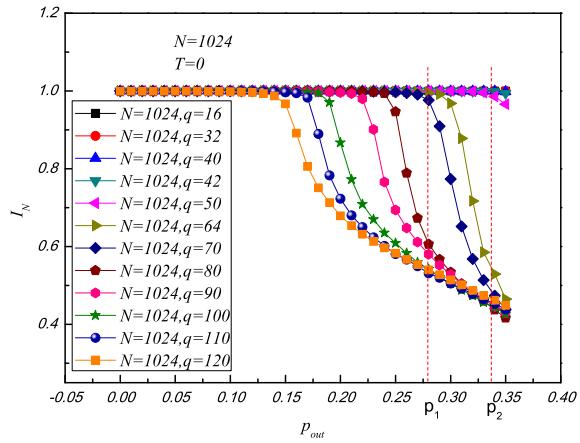


Figure 3.11: The plots of susceptibility χ , convergence time τ , energy E , accuracy I_N and the Shannon entropy H in terms of noise p_{out} for the system $N = 2048$ and $q = 140$ at a zero temperature. All the plots show three phases as noise varies: (1) Below $p_1 = 0.2$, the system can be solved in this “easy” region (e.g., the accuracy is $I_N = 1$); (2) When $0.2 < p_{out} < 0.24$, where the benefit of extra trials is the largest, it is “hard” to solve the system without misplacing nodes (e.g., χ , E and H achieve the peak) ; (3) Above $p_2 = 0.24$, the system is “impossible” to be perfectly solved. $[p_1, p_2]$ are generous bounds in transition crossover regions. Note that the two transitions are demonstrated to be of spin-glass-type by observing the scaling of the correlation function between $[p_1, p_2]$ in Fig. 3.13.

We increase the number of communities to $q = 40$ for the systems in panels (e) through (h). N varies from 256 to 2048, and $\alpha = q/N$ decreases as N increases so that the systems again become less complicated because the communities become more strongly defined. The hard and unsolvable phases in the small $N = 256$ system in panel (e) are difficult to distinguish. Only the easy phase can be easily identified


 (a) χ, τ, E, H, I_N for the system of size $N = 1024$

 (b) I_N for the system of $N = 1024$ with different

 q

Figure 3.12: (a), The plots of susceptibility χ , convergence time τ , energy E , accuracy I_N , Shannon entropy H in terms of noise p_{out} for the system $N = 1024$ and $q = 70$ at a zero temperature. (b), The normalized mutual information I_N in terms of noise p_{out} for a series of systems with the size of $N = 1024$ but different number of communities q . From both plots, we are able to detect the first and second transition point p_1 and p_2 . p_1 is the point where the I_N drops from 1, χ increases from 0, τ achieves the peak, E and H increases from some constant value. p_2 is the position where the I_N curves with different number of communities collapse shown in (b). p_2 also corresponds to the peak of energy and entropy as shown in (a).

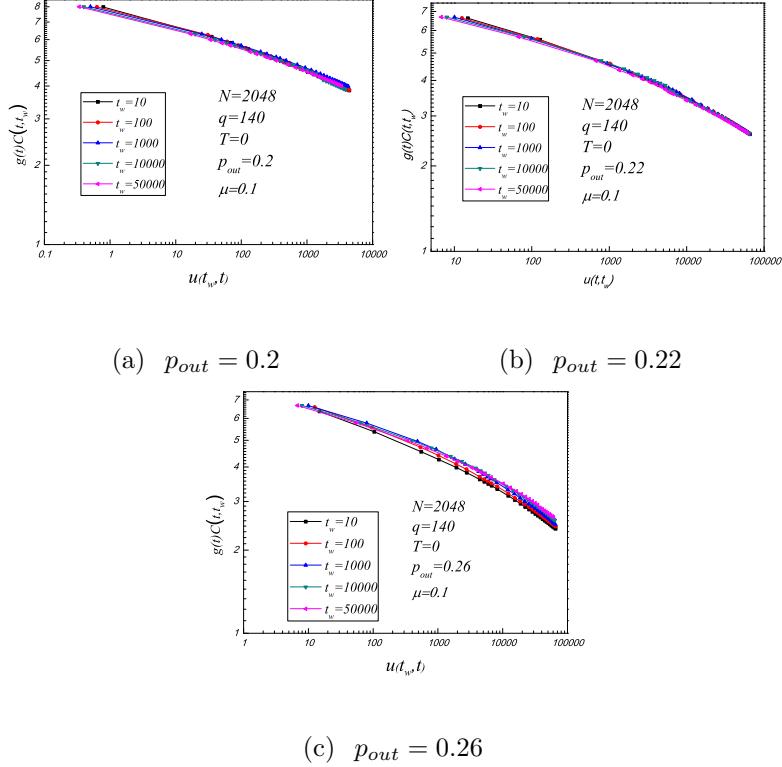


Figure 3.13: We show a collapse of the correlation curves for different waiting times t_w for a system with $N = 2048$ nodes, $q = 140$ communities. p_{out} varies from 0.2 in panel (a) to 0.26 in panel (c). The first and second transition points for this system are $p_1 = 0.2$ and $p_2 = 0.24$. The temperature is $T = 0$. The vertical axis is $g(t)C(t_w, t)$ where $g(t) = 8 - \log_{10}(t)$, $C(t_w, t) = \frac{1}{N} \sum_{i=1}^N \delta_{\sigma_i(t_w), \sigma_i(t_w+t)}$ is the correlation function. The horizontal axis is $u(t_w, t) = \frac{1}{1-\mu}[(t + t_w)^{1-\mu} - t_w^{1-\mu}]$ where $\mu = 0.1$. The noise $p_{out} = 0.2$ (a) and $p_{out} = 0.22$ (b) lie within the “hard” region where the collapse of correlation function is perfect. The noise $p_{out} = 0.26$ (c) is above the second transition point p_2 in the “unsolvable” region, where the collapse becomes poor. That the collapse of the correlation function starts to degrade right after the second transition point p_2 at zero temperature indicates that this transition is of the spin-glass type.

by noting the flat region on the lower left of each panel. The “computational susceptibility” $\chi(T, p_{out})$ peaks at increasing heights at both the low and high temperatures from panels (f) to (h) indicating that the phase transitions become more prominent as the system size increases.

We further increase the number of communities to $q = 70$ and study the phase transitions for the same range of system sizes. The hard phase at high temperature in panel (i) is difficult to detect. $\chi(T, p_{out})$ clearly shows the three phases in panels (j) and (k). The easy phases again become larger as the system size increases. $\chi(T, p_{out})$ in the hard phase increases as N increases indicating that the phase transitions at both low and high temperatures are more obvious from panel (i) to (k).

From panel (a) to (d) in Fig. 3.6, the number of communities is fixed at $q = 16$, the community size follows the power law distribution. As the system size N increases, the width of the hard phase becomes narrower and the height increases at high temperature. This is also confirmed by panel (a) of Fig. 3.7. From panel (e) to (h) in Fig. 3.6, the number of communities is $q = 40$. The hard phase becomes sharper as the system size increases as well. The same behavior happens from panel (i) to (l) in Fig. 3.6 when the number of communities is fixed at $q = 70$.

In Figs. 3.7 and 3.8, we also show corresponding 2D plots for the boundaries of the hard phase and the first transition point p_1 as the function of temperature T . In Fig. 3.7, the area of the hard phase becomes narrower as the system size increases. At $q = 40$, for example, the width of the hard phase for the smallest system at $N = 256$ is about $\Delta T = 1.5$. As N increases, the hard phase width shrinks to $\Delta T = 1$ at

$N = 512$ and down to $\Delta T = 0.5$ for $N = 2048$ which further indicates that the phase transition becomes sharper in the thermodynamic limit. In Fig. 3.8, the first transition point p_1 increases over the entire temperature range as N increases. This behavior is consistent with the system complexity trend as previously mentioned.

In Fig. 3.9, we further plot the convergence time τ as the function of noise p_{out} for a fixed number of communities q at zero temperature. p_{out} for the first peak of the convergence time matches the first transition point p_1 in Fig. 3.8. As the system size increases, the peak moves to the right. This is also consistent with Fig. 3.8 where the system becomes less complicated as N increases.

3.4.3 Other information theoretic and thermodynamic quantities

We further fortify and provide our results of the phase diagram of our systems as ascertained via other information theoretic and thermodynamics quantities. These measures include the average normalized mutual information I_N between replica pairs, Shannon entropy H , and energy E as shown in Fig. 3.10. We additionally show the corresponding computational susceptibility χ from Fig. 3.2 or 3.6 for comparison. All panels are for a system of size $N = 2048$. In panels (a) through (d), $q = 16$ which corresponds to Fig. 3.2(d). Panels (e) through (h) plot results for $q = 32$ with $\alpha = 0.015$ which corresponds to Fig. 3.6(d). Panels (i) through (l), display the results for $q = 70$ which corresponds to Fig. 3.2(l). Finally, Panels (m) through (p) display

results for $q = 140$ and $\alpha = 0.07$ corresponding to Fig. 3.6(h).

All panels consistently display the three different complexity phases: the “easy” (flat region, lower left), “hard” (varied central regions), and “unsolvable” phases (far right or top). The existence of the hard phase is reflected by the ridges at both low and high temperatures in the susceptibility χ plot which often corresponds rapid shifts (up or down) in the other measures. In each plot, the red line serves as a guide to the eye to emphasize the boundaries between different phases. The phase boundaries adduced from the different measures agree with each other. .

In Ref. [40], we also demonstrated the spin glass character of the phase transition by observing the exceptional collapse of time autocorrelation curves (over four orders of magnitude of time at high and low temperatures) in the vicinity of the hard phase. We further elucidated on evidence regarding phase transitions [40] in identifying community structure via a dynamical approach (some other dynamical methods include [23, 26]) where “chaotic-type” transitions that we speculated upon may extend into the node dynamics for large systems.

3.5 non-interacting cliques

As depicted in Fig. 3.14, we analytically estimate a minimum transition temperature by examining a system with q non-interacting cliques. In panel (a), each of the q communities consists of l nodes which are maximally connected, but no noise exists between these cliques. The presence of noise will, in general, lower the temperature

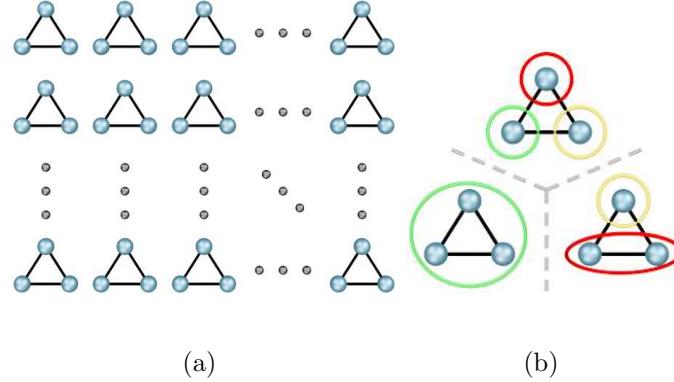


Figure 3.14: Panel (a) depicts q independent cliques (maximally connected clusters).

Panel (b) indicates the different combinations of $l = 3$ nodes which must be summed (including three copies of the 2-1 configuration) in order to determine the partition function for a single clique.

T_x of the transition point which manifests as departure from the easy phase in certain regions of Figs. 3.2 and 3.6.

Within our algorithm and model, communities do not interact in an explicit sense. In addition, with this model problem the situation is greatly simplified because no edges are assigned between cliques, so we use Eq. (3.1) to calculate the partition function of the system by counting the energy contribution of all edges within each cluster over the number of combinations for partitioning the clusters. As a further simplification, we also set the energy contribution for a single edge to be -2 so that the Hamiltonian gives an energy of -1 for each edge.

3.5.1 Partition function

First, we investigate the smallest non-trivial clique size with $l = 3$ nodes. The partition function for the decoupled cliques is,

$$\mathbf{Z} = (\mathbf{Z}_l)^q = \sum_{\sigma_i, \sigma_j} e^{-\beta H_{i,j}} \quad (3.2)$$

where \mathbf{Z}_l is the partition function for a single clique and $\beta = 1/T$ is the inverse temperature. Considering the $l = 3$ cluster combinations depicted in Fig. 3.14(b), \mathbf{Z}_3 is

$$\mathbf{Z}_3 = qe^{6\beta} + 3q(q-1)e^{2\beta} + q(q-1)(q-2). \quad (3.3)$$

The first term represents the optimal local cluster solution, and the sum of the remaining terms accounts for the remaining sub-optimal local partitions. We define ω_l as the ratio of Boltzmann weights of the sub-optimal partitions to the optimal solution. For \mathbf{Z}_3 , the ratio ω_3 is

$$\omega_3 = \frac{q(q-1) [3e^{2\beta} + (q-2)]}{qe^{6\beta}}. \quad (3.4)$$

$\omega_l < 1$ indicates that the optimal solution is dominant, while $\omega_l \rightarrow \infty$ means the system is disordered. We can define $\omega_l = 1$ as the transition point from the ordered phase to the disordered phase, and the corresponding ‘‘crossover’’ temperature T_x is found by solving the transcendental equation

$$3(q-1)e^{-4/T_x} + (q-1)(q-2)e^{-6/T_x} = 1. \quad (3.5)$$

In the limit of large q , this equation simplifies to

$$q^2 e^{-6/T_x} \simeq 1, \quad (3.6)$$

which yields our estimate for the crossover temperature

$$T_x \simeq \frac{3}{\log q} \quad (3.7)$$

for the $l = 3$ clique system. This is in agreement to general trend on accuracy in [132].

If we generalize to arbitrary clique size l , the corresponding partition function for a single clique becomes

$$\begin{aligned} Z_l &= q e^{2\beta \binom{l}{2}} + lq(q-1)e^{2\beta \binom{l-1}{2}} \\ &\quad + \frac{l(l-1)}{2}q(q-1)(q-2)e^{2\beta \binom{l-2}{2}} \\ &\quad + \cdots + q(q-1)(q-2)\cdots(q-l). \end{aligned} \quad (3.8)$$

Again, the first term in Eq. (3.8) is the Boltzmann weight of the optimal clique partition, and the other terms sum the weights of the incorrect partitions. ω_l is

$$\omega_l = \frac{\sum_{k=1}^l \binom{l}{k} \binom{q}{k+1} (k+1)! e^{2\beta \binom{l-k}{2}}}{qe^{2\beta \binom{l}{2}}}, \quad (3.9)$$

and $\omega_l = 1$ returns the cross-over temperature T_x for arbitrary cliques of size l . We summarize the crossover temperature relations in column one of Table 3.5.1 where we express e^{2/T_x} in terms of powers of q for several values of l . The general relation is

$$T_x \simeq \frac{l}{\log q}. \quad (3.10)$$

3.5.2 Symmetry Breaking

We can inquire about the crossover temperature T_x from another perspective. Take two nodes i and j in the same clique. If the probability that a solution assigns them

l	e^{2/T_x}	$e^{2/T_x^{(1/q)}}$	$e^{2/T_x^{(p)}}$
2	q	q	pq^2
3	$q^{2/3}$	$q^{2/3}$	$p^{1/3}q$
4	$q^{1/2}$	$q^{1/2}$	$p^{1/6}q^{2/3}$
5	$q^{2/5}$	$q^{2/5}$	$p^{1/10}q^{1/2}$
6	$q^{1/3}$	$q^{1/3}$	$p^{1/15}q^{2/5}$
\vdots	\vdots	\vdots	\vdots
l	$q^{2/l}$	$q^{2/l}$	$p^{2/(l^2-l)}q^{2/(l-1)}$

Table 3.1: In column one, the crossover temperature T_x from an “ordered” to a “disordered” state is determined by defining the ratio $\omega_l = 1$ of the sum of Boltzmann weights of sub-optimal node assignments to the weight of the optimal assignment into clique communities as a function of the cluster size l and the number of communities q in the large q limit. In column two, we estimate $T_x^{(1/q)} \simeq T_x$ through different means by calculating the probability $p = 1/q$ that two nodes (in the same clique ideally) are determined to be in the same cluster. In the last column, we generalize column two for an arbitrary probability p .

to the same community is high, then the system is in the “ordered” state. If this probability is $1/q$, the system is in its “disordered” phase. We can define a crossover temperature $T_{\times}^{(1/q)}$ at which the probability of node i and j being in the same cluster exceeds $1/q$ and thus symmetry between Potts spins is broken. This probability

$P(\sigma_i = \sigma_j) = \langle \delta_{\sigma_i, \sigma_j} \rangle$ is

$$P(\sigma_i = \sigma_j) = \frac{\text{Tr}_{\sigma_i} \delta_{\sigma_i, \sigma_j} e^{-\beta H}}{\text{Tr}_{\sigma} e^{-\beta H}}, \quad (3.11)$$

where σ_i and σ_j denote the cluster memberships for nodes i and j , respectively.

Expressing the numerator and in terms of l and q , Eq. (3.11) becomes,

$$\begin{aligned} P(\sigma_i = \sigma_j) &= \left\{ qe^{2\beta \binom{l}{2}} + (l-2)q(q-1)e^{2\beta \binom{l-1}{2}} \right. \\ &\quad \left. + \cdots + q(q-1)\cdots(q-l+2) \right\} \\ &/ \left\{ qe^{2\beta \binom{l}{2}} + lq(q-1)e^{2\beta \binom{l-1}{2}} \right. \\ &\quad \left. + \cdots + q(q-1)(q-2)\cdots(q-l) \right\}. \end{aligned} \quad (3.12)$$

In the limit of large q , Eq. (3.12) simplifies to

$$P(\sigma_i = \sigma_j) \simeq \frac{qe^{2\beta \binom{l}{2}} + \sum_{k=1}^{l-2} \binom{l-2}{k} q^{k+1} e^{2\beta \binom{l-k}{2}}}{qe^{2\beta \binom{l}{2}} + \sum_{k=1}^l \binom{l}{k} q^{k+1} e^{2\beta \binom{l-k}{2}}}. \quad (3.13)$$

Choosing $P(\sigma_i, \sigma_j) = 1/q$ yields in a crossover temperature $T_{\times}^{(1/q)}$ at which the system goes from being unbroken q -state symmetry to ordered. When $l = 3$, Eq. (3.13) becomes,

$$q^2 e^{6\beta} + q^3 e^{2\beta} \simeq q e^{6\beta} + 3q^2 e^{2\beta} + 3q^3 + q^4. \quad (3.14)$$

In the large q limit, $e^{2\beta} \simeq q^{2/3}$, and the crossover temperature is $T_{\times}^{(1/q)} = 3/\log q$.

The asymptotic expressions for several values of q and l are summarized in column

two of Table 3.5.1. For general q and l , the relation is

$$T_{\times}^{(1/q)} \simeq \frac{l}{\log q}. \quad (3.15)$$

Eq. (3.15) is consistent with Eq. (3.10) and also [132].

For a general crossover probability $P(\sigma_i, \sigma_j) = p$ with $l = 3$, the crossover temperature $T_{\times}^{(p)}$ is determined by solving

$$e^{6\beta} + qe^{2\beta} \simeq pe^{6\beta} + 3qpe^{2\beta} + 3q^2p + q^3p. \quad (3.16)$$

In the large q limit, Eq. (3.16) is $e^{2\beta} \simeq p^{1/3}q$, where $T_{\times}^{(p)} \simeq 2/(\log q + 1/3 \log p)$.

Results for $T_{\times}^{(p)}$ for several values of q and l are shown column three of Table 3.5.1.

For general q and l , the relation is

$$T_{\times}^{(p)} \simeq \frac{1}{p^{1/l}q}. \quad (3.17)$$

3.5.3 Simulated crossover temperature

We can also simulate the crossover temperature T_{\times} or $T_{\times}^{(p)}$ as a function of system size N by solving the non-interacting clique problem using our heat bath community detection algorithm. As seen in Fig. 3.15, the simulated and analytic asymptotic behaviors agree well in the large N limit, so the crossover temperature for this trivial system is $T_{\times} = 0$.

The crossover temperature derived in this section deals with a *heat-induced* disorder. That is, it marks the onset of a “liquid” phase that transitions at a lower heat bath temperature as the system size grows. In practice, one uses a SA algorithm that

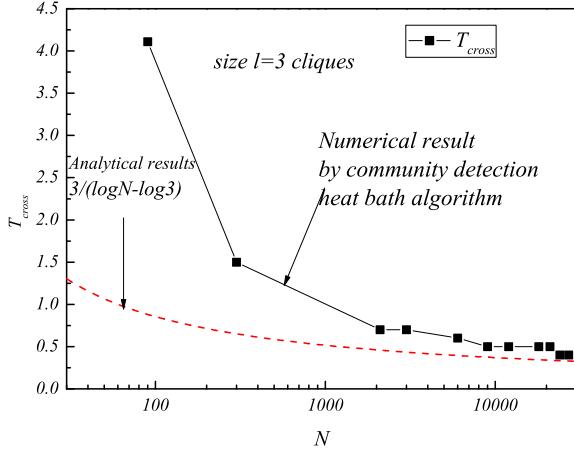


Figure 3.15: The crossover temperature at which the system cannot be perfectly solved as the function of the system size N . The data here uses cliques of size $l = 3$. The dashed line is the analytical result and the solid line is determined by the heat bath community detection algorithm optimizing the Hamiltonian of Eq. (3.1).

applies a cooling scheme (as opposed to a constant temperature HBA) to improve the attempt at locating the ground state of the system. That is, it applies a high temperature exploration of the general landscape finished by low temperature “fine tuning” of the solution. For the non-interacting cliques in this section, SA would obviously still identify the ground state because the energetic fluctuations would trivially diminish as the system is cooled toward $T = 0$.

With increasing p_{out} at low T , disorder imposed by the glass-type transition is induced by the complexity of the energy landscape, but the transition is qualitatively comparable in the sense of the induced disorder in the solutions found by the HBA. The glass phase also experiences a transition to a liquid-like disordered state at a

temperate that increases slowly with the level of noise, but here, a SA solver will not necessarily transition readily to the ideal solution as the system is cooled because of the inherent complexity of the energy landscape. The greedy algorithm used in [18] (equivalent to the HBA at $T = 0$) applied to the Potts model of Eq. (3.1) is already very accurate [3, 18, 27], so we expect that the greatest benefit of SA over a greedy-oriented solver using Eq. (3.1) will manifest in the hard region near the onset of the “glassy” transition.

3.5.4 A discussion of the crossover temperature

For a spin system with fixed size N , a larger number of spin states q corresponds to a more disordered system. If we expand the partition function of the Potts model in terms of $1/q$, it is explicitly represented as a sum over configurations with progressively larger clusters of identical spins [133]. That is, two spins with the same index $\sigma_i = \sigma_j$ are connected. Then three spins $\sigma_i = \sigma_j = \sigma_k$ are connected, etc. The resulting terms illustrate that increasing q emulates increasing temperature T .

Our analysis in this section applies to general graphs with ferromagnetic interactions (equivalent to the “label propagation” community detection algorithm [134]) on regular, fixed-coordinate lattices [135, 136, 137]. Increasing the number of system states q causes the system to be increasingly disordered. Thus, in the community detection problem, increasing number of communities q linearly with the system size N (such that the average community size remains constant), the solvable (easy) phase shrinks to a “small” region as $N \rightarrow \infty$.

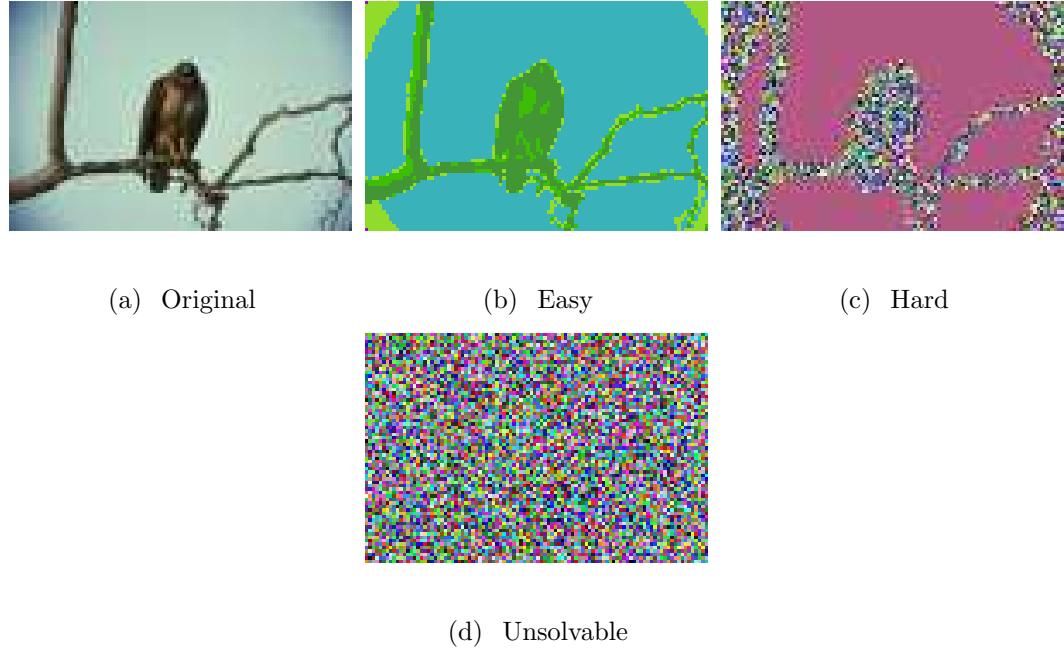


Figure 3.16: (Reproduced from Ref. [80]) We show an image where we apply our community detection algorithm to detect the relevant structures. This case seeks to identify a bird and tree against a sky background. The original images is in panel (a), and the segmentation results are shown in panels (b–d) corresponding to the easy, hard, and unsolvable regions of the community detection problem, respectively. Figure 3.17 shows the phase diagram identifying these respective regions.

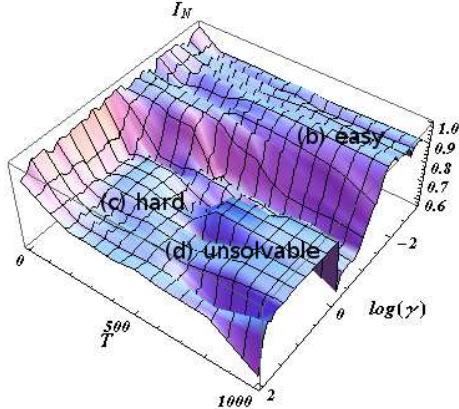


Figure 3.17: (Reproduced from Ref. [80]) We show a three-dimensional phase diagram of NMI (I_N) versus $\log(\gamma)$ and $\log(T)$ for the image segmentation of the bird in Fig. 3.16. T is heat bath temperature for a stochastic community detection solver, and γ is the model weight in Eq. (3.1). We note that the optimal values in the easy and hard regions correspond to the “physical” segmentations of the bird and tree against the background, but the bird is undetectable in the unsolvable region.

Figures 3.10(m–p) illustrate the distinction in the different regions or types of disorder: entropic (high complexity) and energetic (high T). Interestingly, in some cases, additional noise emulates a higher temperature solution process in the sense that it provides additional avenues to explore different configurations. Such an effect may occur in Fig. 3.10(a-d) where the accuracy [I_N in panel (b) *increases* for a short time with increasing noise p_{out}].

Fig. 3.10(n) further shows a crossover region $0.24 \lesssim p_{out} \lesssim 0.32$ where mid-range temperatures improve the solution accuracy (higher I_N). Although this data uses a constant temperature heat bath (no cooling schedule), this is the effect of a stochastic

solver, allowing it to navigate the difficult energy landscape more accurately than a greedy solver. On the left (lower T), the more greedy nature of the solver prevents an accurate solution in the presence of high noise. On the right, the higher temperature of the heat bath itself hinders an accurate solution. In effect, the HBA “wanders” at energies above the meaningful, but locally complex, features of the energy landscape resulting in more random solutions.

The results here incorporate a “global” model parameter γ in Eq. (3.1). That is, the model asserts globally optimal $\gamma(s)$ for the entire graph. For large graphs, this condition is less likely to be true across the full scope of the network, but one can explore methods to obtain locally optimal γ_ℓ (in time or space) for each region or cluster ℓ [80]. Utilizing locally optimal γ_ℓ s will likely work to circumvent the temperature transition at low levels of noise. The successful selection of a local γ_ℓ in the glassy (high noise) region is more difficult because of the complex nature of the local energy landscape.

In the following section, we study the free energy of several systems for ferromagnetic Potts models and then generalize to *arbitrary* weighted Potts models, including antiferromagnetic interactions, on *arbitrary* graphs [132].

3.6 An example of a Phase transition in an image segmentation problem

We illustrate the phase transition effect with an realistic image segmentation example [80]. In Fig. 3.16, we apply our community detection algorithm to detect a bird and tree against a sky background. We display the results in Fig. 3.17 where we plot NMI (I_N) versus $\log(\gamma)$ in Eq. (3.1) and $\log(T)$ where T is the temperature for our stochastic community detection solver. For this problem, we apply edge weights by replacing the A_{ij} elements in Eq. (3.1) with “attractive” and “repulsive” weights w_{ij} which are defined by regional intensity differences within the image [80].

We label the easy (b), hard (c), and unsolvable (d) regions in the phase plot for the bird image in panel (a). Panel (b) shows that our algorithm clearly detects the bird and tree against the background, meaning that the NMI information measure identifies the physically relevant clusters in the problem. In panel (c), the background is segmented separately, but the bird and tree are composed of many small clusters. Panel (d) shows that the bird is undetectable in the unsolvable region.

3.7 free energy: Simple results

In the following analysis, we explicitly show the large q and large T expansions for the free energy per site in three example systems (a non-interacting clique system, simple interacting clique system, and a random graph) before generalizing the analysis

to arbitrary unweighted and weighted graphs. Previous works examined disorder transitions for random-bond Potts models [137, 138] and Ref. [139] studied zeros of the partition function in the large q limit. Large q behavior was shown to approach mean-field theoretical results on fixed lattices [140, 141]. For the unweighted systems, we use a binary distribution for the interaction strength $J = 1$ or 0 (*i.e.*, the energy contribution of an edge is either “on” or “off”).

3.7.1 Free energy of a non-interacting clique system under a large q expansion

If we generalize the non-interacting clique system in Fig. 3.14 to cliques of size l , the partition function is

$$\begin{aligned} \mathbf{Z} = & \left[q e^{\beta J \binom{l}{2}} + lq(q-1) e^{\beta J \binom{l-1}{2}} \right. \\ & + \frac{l(l-1)}{2} q(q-1)(q-2) e^{\beta J \binom{l-2}{2}} \\ & \left. + \cdots + q(q-1)(q-2) \cdots (q-l) \right]^q. \end{aligned} \quad (3.18)$$

When $q \rightarrow \infty$,

$$\begin{aligned} \mathbf{Z} \approx & \left[q e^{\beta J \binom{l}{2}} + lq^2 e^{\beta J \binom{l-1}{2}} + \frac{l(l-1)}{2} q^3 e^{\beta J \binom{l-2}{2}} \right. \\ & \left. + \cdots + q^{l+1} \right]^q. \end{aligned} \quad (3.19)$$

The free energy per site, $f = -\frac{k_B T}{N} \log \mathbf{Z}$, (with the Boltzmann constant set to $k_B = 1$) is

$$f \approx -T \log q - T \sum_{k=0}^{l-2} \left[\binom{l-1}{k} \frac{1}{k+1} e^{\beta J \binom{l-1}{k-1}} \right] q^{-(k+1)}. \quad (3.20)$$

From Eq. (3.20), we further simply the free energy per site

$$\begin{aligned} f &\approx -T \log q - T \sum_{k=0}^{l-2} a(k) e^{\beta J \binom{l-1}{k-1}} q^{-(k+1)} \\ f &\approx -T \log q - T a(0) \frac{e^{\beta J}}{q} \end{aligned} \quad (3.21)$$

where $a(k) = \binom{l-1}{k} \frac{1}{k+1}$. We will compare Eq. (3.21) with the high T expansion in the next section. Despite the functional dependence of $\exp(\beta J)$, the large q limit dominates the expansion, forcing the system to be approximately equivalent to a large temperature limit.

3.7.2 Free energy of a non-interacting clique system as ascertained from a high temperature expansion

Note that the *most* ordered Potts graph is a system of non-interacting cliques (maximally connected sub-graphs). That is, the presence of noise (extraneous intercommunity edges) will only serve to *increase* the overall disorder in the system. One exception is that increased disorder can emulate increased temperature T for both greedy and stochastic community detection solvers (see also Sec. 3.5.4).

We can construct the high T expansion easily by means of Tutte polynomials [142] (see Sec. 3.8.1) where we again solve a system of q cliques of size l . Equation (3.1) and a ferromagnetic Potts model have the same *ground state* energy for this clique system (see also Secs. 3.7.5, 3.7.7, and 3.7.8 for more general derivations), so the partition function in terms of the Tutte polynomial $t(G; x, y)$ for a graph G is

$$\mathbf{Z} = q^{k(G)} v^{|V|-k(G)} t(G; x, y) \quad (3.22)$$

where q is the number of clusters or states, $v = \exp(\beta J) - 1$, G denotes the graph, $k(G)$ is the number of connected components in G , $|V| = N$ is the number of vertices, $x = (q + v)/v$ and $y = v + 1$. For the non-interacting clique system, $k(G) = q$ and $N = lq$. We denote the Tutte polynomial of a single clique of size l as $K_l(G; x, y)$.

$K_2(G; x, y) = x$, so the partition function is

$$\begin{aligned} \mathbf{Z} &= q^q v^q x^q, \\ \mathbf{Z} &= q^q v^q \left(\frac{q}{v} + 1 \right)^q, \end{aligned} \quad (3.23)$$

where we used $N = 2q$. In a high T approximation, $x \approx q/v \gg 1$, so the partition function becomes $Z \approx q^{2q}$, and the free energy is

$$f \approx -T \log q, \quad (3.24)$$

which simply states that the system is completely random in the large T limit.

For triangle cliques, $K_3(G; x, y) = x^2 + x + y$. The graph G is composed of disjoint triangles, so the Tutte polynomial is $t(G; x, y) = (x^2 + x + y)^q$, and the partition function becomes

$$\mathbf{Z} \approx q^q v^{2q} (x^2 + x + y)^q. \quad (3.25)$$

In a high T approximation $y \approx 1$, but $x \approx q/v \gg 1$ in either the large q or large T limits, so we make a further approximation of $y \approx 0$. Then, $K_3(G; x, y = 0) = x^q(x + 1)^q \approx x^{2q}$. The partition function simplifies to $\mathbf{Z} \approx q^{3q}$, so the free energy per site for $l = 3$ is again

$$f \approx -T \log q \quad (3.26)$$

which is identical to the $l = 2$ result because we consistently applied the approximation $q/v \gg 1$ to $x = (q/v + 1) \approx q/v$ and $(x + 1) = (q/v + 2) \approx q/v$.

Generalizing to an arbitrary clique size l in the high T approximation, the Tutte polynomial $K_l(G; x, y = 0)$ is

$$K_l(G; x, y = 0) = \frac{\Gamma(x + l - 1)}{\Gamma(x)}, \quad (3.27)$$

The partition function is

$$\mathbf{Z} \approx q^{lq} \left(\frac{v}{q}\right)^{(l-1)q} \frac{\Gamma\left(\frac{q}{v} + l - 1\right)}{\Gamma\left(\frac{q}{v}\right)}, \quad (3.28)$$

and $v = e^{\beta J} - 1 \approx \beta J$, so the free energy per site yields

$$f \approx -T \log q - \frac{l-1}{l} T \log \left(\frac{\beta J}{q}\right) - \frac{T}{lq} \log \left[\frac{\Gamma\left(\frac{q}{\beta J} + l - 1\right)}{\Gamma\left(\frac{q}{\beta J}\right)} \right]. \quad (3.29)$$

The leading $\log q$ term represents the infinite T limit which is approximately constant in large systems for any clique size l . That is, the partition function $\mathbf{Z}_{T \rightarrow \infty} \approx q^N$ for every system. The $l = 2$ and 3 results above illustrate that when $l \ll q$, the ratio of gamma functions in Eq. (3.29) simplifies to x^{lq} , and the free energy for the non-interacting clique system is approximately $f \approx \log q$ in the large T limit.

The second term in Eq. (3.29) gives the leading order correction for high T . It is absent in the explicit $l = 2$ and 3 results above because we applied the approximation $q/v \gg 1$. Together, the last two terms imply that increasing the temperature T (decreasing β) emulates increasing the number of communities q for a ferromagnetic Potts model.

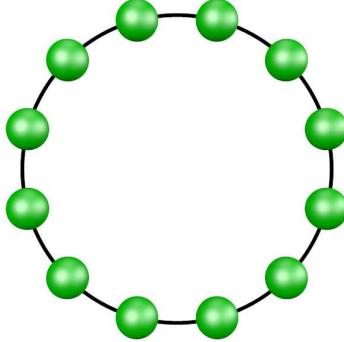


Figure 3.18: A depiction of a circle of cliques (maximally connected clusters) of size l connected by single edges. In contrast to Fig. 3.14, this system adds a simple interaction between cliques. We analyze the configuration in Sec. 3.7.3 and show that a ferromagnetic Potts model behaves the same in the large q and large T limits.

3.7.3 Free energy for the “circle of cliques” in the high q or the high T expansion

We now investigate the slightly more complicated system depicted in Fig. 3.18: a “circle of cliques” where each complete sub-graph cluster is connected to its neighbors by a single edge. We construct q cliques of size $l = 3$ and apply the Tutte polynomial method [142] to solve the system. As in the previous sub-section, the ground state of Eq. (3.1) and a ferromagnetic Potts model have the same energy, so we use a ferromagnetic model. In terms of the Tutte polynomial $t(G; x, y)$ for a graph G , the partition function is given by Eq. (3.22).

Eq. (3.67) in Sec. 3.9 derives the exact Tutte polynomial for Fig. 3.18 with $l = 3$, and Eq. (3.69) gives the high T expansion $t(G; x, y = 0) = (1 + x)^{q+1} x^{2q-3}$. Sub-

stituting $N = 3q$ and the approximation $x \approx q/v$ (in either the large q or large T limits), the partition function becomes

$$\mathbf{Z} \approx q^{2q-2} v^{q+2} (1+x)^{q+1} x^{2q-3} \quad (3.30)$$

We factor out q^{3q} , and then apply the approximations: $v = \exp(\beta J) - 1 \approx \beta J$, $x \approx q/v \approx q/(\beta J) \gg 1$, and $q \gg 1$. The free energy per site is then

$$f \approx -T \log q - \frac{2T}{3} \log \left(\frac{q}{\beta J} \right) \quad (3.31)$$

As in the previous sub-section, the leading $\log q$ term represents the infinite T limit. Equation (3.31) affirms the implication of Eq. (3.29) regarding the corresponding behavior of large q or T . Specifically, increasing the temperature (decreasing β) emulates increasing the number of communities q for a ferromagnetic Potts model.

3.7.4 Free energy of a random graph in a large q or a large T expansion

We apply the Tutte polynomial method of Sec. 3.8.1 to determine the high T and high q partition function for a random graph. For calculation purposes, we begin with a complete graph of size N . Then we randomly remove edges to construct a random graph such that any two nodes are connected by an edge with a probability p . The derivation repeatedly applies lemma 1 stated in Sec. 3.8.1.

We denote the Tutte polynomial of a complete graph (clique) of size l as K_l . $t(G)$ for a clique with d duplicated edges (multiply defined edges between two nodes) or

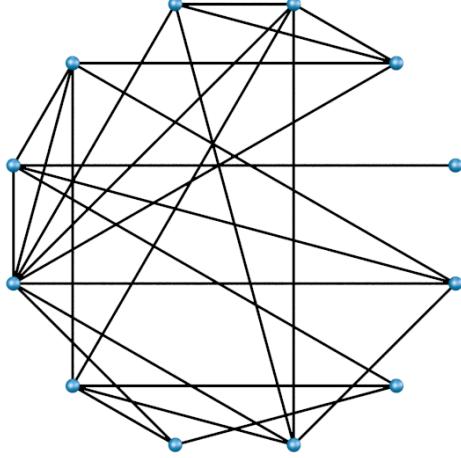


Figure 3.19: A sample depiction of a random graph with N nodes. In Sec. 3.7.4, we analyze such a system by randomly removing edges from a clique configuration of N nodes under the assumption that we maintain a connected graph. We show that a ferromagnetic Potts model on a random graph behaves the same in the large q and large T limits.

loops (self-edges) is defined as $K_l^{(d)}$. For economy of notation, we also define $G_l^{[m]}$ as the Tutte polynomial of a graph with m missing edges (*i.e.*, not a clique). Note that $K_l^{(0)} \equiv G_l^{[0]} \equiv K_l$. For the following derivation, we work under the *assumption* that when we delete or contract any edge, the random graph remains connected.

Under the high temperature T or high number of clusters q approximations, $y \ll x$ and $y \simeq 0$. Equation (3.27) gives the exact expression of the Tutte polynomial $K_l(G; x, y = 0)$ for a clique at $y = 0$. If we cut one edge from the complete graph K_N , we obtain the recursion formula

$$\begin{aligned} K_N &= G_N^{[1]} + K_{N-1}^{(N-1)}, \\ K_N &= G_N^{[1]} + K_{N-1}. \end{aligned} \tag{3.32}$$

where we applied lemma 1 to obtain Eq. (3.32). From henceforth, we assume the application of lemma 1. We are interested in the graph with missing edges, so we solve Eq. (3.32) for $G_N^{[1]}$.

$$G_N^{[1]} = K_N - K_{N-1}. \quad (3.33)$$

Note that the reduced graph is represented as a summation over complete graphs.

Now we apply the Tutte recursion formula to *both sides* of Eq. (3.33).

$$G_N^{[2]} + G_{N-1}^{[1]} = G_N^{[1]} + K_{N-1} - G_N^{[1]} - K_{N-1}. \quad (3.34)$$

We can choose the deleted and contracted edges in the corresponding terms to be identical because the resulting Tutte polynomial is in general independent of the operation order. After collecting terms and substituting the previous $G_N^{[1]}$ result, we solve for $G_N^{[2]}$ to obtain

$$G_N^{[2]} = K_N - 2K_{N-1} + K_{N-2}, \quad (3.35)$$

for this particular random graph. Again, the right-hand-side of Eq. (3.35) is a summation over complete graphs. This recursive relation for $G_N^{[k]}$ continues until we obtain

$$G_N^{[k]} = \sum_{i=0}^k (-1)^i \binom{k}{i} K_{N-i}. \quad (3.36)$$

We insert this into Eq. (3.27) with the pre-factor qv^{N-1} to generate the partition function at high T

$$\mathbf{Z} = q^N \left(\frac{v}{q} \right)^{N-1} \left[\sum_{i=0}^k (-1)^i \binom{N}{i} \frac{\Gamma(x + N - i - 1)}{\Gamma(x)} \right]. \quad (3.37)$$

We substitute $x = (q + v)/v \approx q/v$ when $v \ll q$ (high T or high q approximations) and again utilize $v = e^{\beta J} - 1 \approx \beta J$ in the high T approximation to obtain the free

energy per site

$$\begin{aligned} f &= -T \log q - \frac{N-1}{N} T \log \left(\frac{\beta J}{q} \right) \\ &\quad - \frac{T}{N} \log \left[\sum_{i=0}^k (-1)^i \binom{N}{i} \frac{\Gamma \left(\frac{q}{\beta J} + N - i - 1 \right)}{\Gamma \left(\frac{q}{\beta J} \right)} \right]. \end{aligned} \quad (3.38)$$

Note that the first two terms become $\log(q)/N \log(\beta J)$ as $N \rightarrow \infty$. From Eq. (3.38), we obtain the same conclusion for this random graph as for the previously analyzed clique systems. While Secs. 3.7.1, 3.7.2, and 3.7.3 result in free energies with different functional forms, in each case, q and T have the same functional form in the arguments of the functions in the high T limit.

3.7.5 Free energy of an arbitrary graph G in the large T expansion

We can construct the explicit high T expansion for an arbitrary (unweighted) graph G by means of the Tutte polynomial method [142]. Factoring out q^N and substituting $|V| = N$, $x = q/v + 1$, and $y = v + 1$ in Eq. (3.22), we write a trivially modified form of the partition function

$$\mathbf{Z} = q^N \left[\left(\frac{v}{q} \right)^{N-k(G)} t \left(G; \frac{q}{v} + 1, v + 1 \right) \right]. \quad (3.39)$$

At this point, the equation is completely general, but the corresponding behavior for temperature T and number of clusters q is almost apparent in the reciprocal relationship of q and v .

Again, $x \approx q/v$ in either the large q or large T limits. In a high T approximation, $v \approx \beta J = T/J$ and $y \approx 0$ or 1 ($y = 0$ is a common approximation since $x \gg y$ in the same limit).

$$\mathbf{Z} \approx q^N \left[\left(\frac{J}{qT} \right)^{N-k(G)} t \left(G; \frac{qT}{J}, y_{T'} \right) \right] \quad (3.40)$$

where $y_{T'} = 0$ or 1 . The free energy per site is then

$$f \approx -T \log q - \frac{N - k(G)}{N} T \log \left(\frac{J}{qT} \right) - \frac{T}{N} \log \left[t \left(\frac{qT}{J}, y_{T'} \right) \right]. \quad (3.41)$$

The leading $\log q$ term appears in our previous calculations. Again, it represents the infinite T limit for an arbitrary system which is approximately constant in large systems.

From the perspective of increasing q , the similarity to the large T behavior is more apparent if we fix the temperature $T = T'$ and define an effective interaction constant $J_q \equiv e^{J/T'} - 1$. We then rewrite Eq. (3.41) as

$$f \approx -T \log q - \frac{N - k(G)}{N} T \log \left(\frac{J_q}{q} \right) - \frac{T}{N} \log \left[t \left(\frac{q}{J_q}, y_q \right) \right]. \quad (3.42)$$

where $y_q \equiv e^{J/T'}$ is a constant. When $N \rightarrow \infty$ and $k(G) \ll N$, the first two terms become $T \log(q)/N \log(\beta J)$. Comparing Eq. (3.41), Eq. (3.42) shows the close correspondence between increasing q (at fixed T') and increasing T . J_q grows exponentially faster than q with decreasing T' , so a finite (perhaps small) stable or solvable region is likely except in the presence of high noise.

3.7.6 Annealed versus quenched averages

The above proofs apply to quenched averages because the binary distribution is constant with respect to the distribution integration. That is, using Eq. (3.42), we assume a probability distribution $P(\{J_{ij}\})$ and integrate over it to obtain the quenched average free energy per site

$$\begin{aligned} f[\{J_{ij}\}] &= \int D J_{ij} \prod_{i \neq j} P(\{J_{ij}\}) \left\{ \log q \right. \\ &\quad + \frac{N - k(G)}{N} \log \left(\frac{J}{qT} \right) \\ &\quad \left. + \frac{1}{N} \log \left[t \left(\frac{qT}{J}, y_{T'} \right) \right] \right\}, \end{aligned} \quad (3.43)$$

but the integrand (f_0) is a constant because J is independent of $\{J_{ij}\}$, so the integral trivially simplifies to

$$f[\{J_{ij}\}] = f_0 \int D J_{ij} \prod_{i \neq j} P(\{J_{ij}\}). \quad (3.44)$$

where the integral is unity. In a more general model with a defined $\{J_{ij}\}$ probability distribution, the leading order $\log q$ contribution would remain unchanged, but we would obtain correction terms from the integration over the quenched interaction distribution $\{J_{ij}\}$.

3.7.7 Free energy of non-interacting cliques for an *arbitrary* weighted Potts model under a large T expansion

We can represent an *arbitrary* weighted Potts model with ferromagnetic and antiferromagnetic interactions. That is, we can generally write

$$H(\{\sigma\}) = -\frac{1}{2} \sum_{i \neq j} [a_{ij} A_{ij} - b_{ij} (1 - A_{ij})] \delta(\sigma_i, \sigma_j). \quad (3.45)$$

where a_{ij} and b_{ij} are arbitrary “attractive” and “repulsive” edge weights. This summarization includes modularity [12], a Potts model incorporating a “configuration null model” (CMPM) comparison [17] (the most common variation in [17] is effectively generalizes modularity), CMPM allowing antiferromagnetic relations, “label propagation” [107, 134], an Erdős-Rényi Potts model [17], a “constant Potts model” [143], the weighted form of the APM [18, 27], or a “variable topology Potts model” suggested in [18].

Note that the repulsive weights b_{ij} are important in that they provide a “penalty function” which enables a well-defined ground state for the Hamiltonian for an arbitrary graph. That is, the ground state of a purely ferromagnetic Potts model in an arbitrary graph is trivially a fully collapsed system (perhaps with disjoint subgraphs). Several of the above models incorporate a weighting factor γ of some type on the penalty term which allows the model to span different scales of the network in qualitatively similar ways.

We denote the partition function of a graph G^* with l nodes and weighted edges $\{e\}$ by $Z(G^*; q, \mathbf{v}) \equiv \mathcal{K}_l$. We assume that $J_e \ll T$ for all edges e , and all

pairs of nodes in G^* are connected by a weighted edge J_e (either ferromagnetic or antiferromagnetic). From Sec. 3.8.2, a recurrence relation for the multivariate Tutte polynomial of a general weighted clique is

$$\mathcal{K}_l \approx \left(q + \sum_{k=1}^{l-1} v_k \right) \mathcal{K}_{l-1} + O(y_e), \quad (3.46)$$

The partition function for \mathcal{K}_l at high T is

$$\mathcal{K}_l \approx q^N \prod_{j=2}^l \left(1 + \sum_{k=1}^{j-1} \frac{v_k}{q} \right), \quad (3.47)$$

Now, we generate a graph consisting of a set of q non-interacting cliques of size l_i where $i = 1, 2, \dots, q$.

$$\mathcal{K}_l \approx q^N \prod_{i=1}^q \prod_{j=2}^{l_i} \left(1 + \sum_{k=1}^{j-1} \frac{\beta J_k}{q} \right). \quad (3.48)$$

where we used $v_e \approx \beta J_e$ at high T for general edge weights J_e (even if $J_e < 0$ as long as $J_e \ll T$).

The free energy is

$$\begin{aligned} f &\approx -T \log q - \frac{T}{N} \sum_{i=1}^q \sum_{j=2}^{l_i} \sum_{k=1}^{j-1} \frac{\beta J_k}{q} \\ &\approx -T \log q - \frac{1}{N} \sum_{i=1}^q \frac{E_i}{q} \\ &= -T \log q - \frac{E}{qN} \end{aligned} \quad (3.49)$$

where we invoked $\log(1 + x) \approx x$ for $x \ll 1$ there. E_i is the energy of cluster i according to the weighted Potts model of Eq. (3.45), and E is the total energy of the graph. Eq. (3.48) and Eq. (3.49) both imply that large q emulates large T for an *arbitrary* Potts model on a weighted graph G . That is, if a community detection

quality function can be expressed in terms of the general Potts model in Eq. (3.45), then large q and large T are essentially equivalent.

3.7.8 Free energy of non-interacting cliques for an *arbitrary* weighted Potts model under a large q expansion

The multivariate Tutte polynomial [144] (see also Sec. 3.8.2 and Ref. [132]) appears in a subgraph expansion over the subset of edges $\mathcal{A} \subseteq \mathcal{E}$ in a graph $G = (V, \mathcal{E})$ with a set of V vertices and \mathcal{E} edges

$$Z(G; q, \mathbf{v}) = q^N \left[\left(1 + \sum_{e'=1}^{|\mathcal{E}|} \frac{v_{e'}}{q} \right) + \cdots + q^{k(G)-N} \prod_{f'=1}^{|\mathcal{E}|} v_{f'} \right] \quad (3.50)$$

$k(\mathcal{A})$ is the number of connected components of $G_A = (V, \mathcal{A})$ and $v_e = \exp(\beta J_e) - 1$.

For our purposes, Eq. (3.50) serves as an alternate representation of Z_G to facilitate the calculation of the large q expansion.

For large q , when $q^N \gg |v_e|^L$, the last term may neglect, and for a system of non-interacting cliques of size l_i with $i = 1, 2, \dots, q$, the leading order terms in large q are

$$Z(G; q, \mathbf{v}) \approx q^N \prod_{i=1}^q \prod_{j=2}^{l_i} \left(1 + \sum_{k=1}^{j-1} \frac{v_k}{q} \right). \quad (3.51)$$

The approximation is identical to Eq. (3.47) at high T . Ref. [132] calculates an explicit crossover temperature including the last subgraph $\mathcal{A} = \mathcal{E}$ that competes with the large q terms as $T \rightarrow 0$. The free energy corresponding to Eq. (3.51) becomes

$$f \approx -T \log q - \frac{T}{N} \sum_{i=1}^q \sum_{j=2}^{l_i} \sum_{k=1}^{j-1} \frac{v_k}{q} \quad (3.52)$$

where we applied the small x approximation $\log(1 + x) \approx x$.

In order to illustrate the correspondence in large q and T , we fix $T = T'$, define

$J_e^{(q)} \equiv \exp(\beta' J_e) - 1$, and rewrite the free energy per site

$$f \approx -T' \log q - \frac{T'}{N} \sum_{i=1}^q \sum_{j=2}^{l_i} \sum_{k=1}^{j-1} \frac{J_k^{(q)}}{q}. \quad (3.53)$$

Large q in Eq. (3.50) emulates large T in Eq. (3.48). As with the unweighted case in Eq. (3.42) in Sec. 3.7.5, $J_e^{(q)}$ is exponentially weighted in $\beta' = 1/T'$, so a non-zero (perhaps small) region of stability is essentially ensured except in the presence of high noise [132]. We can additionally determine a rigorous bound using methods in [132, 145, 146]

$$T_x^{\text{UB}} = \frac{\bar{J}_0}{\log \left[\frac{p(q-1)}{(1-p)} \right]}, \quad (3.54)$$

where $\bar{J}_0 = \frac{1}{2} \sum_j J_{j0} [1 + \text{sgn}(J_{j0})]$ is a generous upper bound summing only positive energy contributions and p is the probability for finding a given spin σ_0 in a specific spin state $\bar{\sigma}$. This result further agrees with our conclusions. Note that as $p \rightarrow 1/q$, the system is completely disordered, so $T_x \rightarrow \infty$. As $p \rightarrow 1$, the system is perfectly ordered, so $T \rightarrow 0$.

3.8 Tutte polynomials

We give a very brief introduction to Tutte polynomials consisting of the essential facts necessary for the derivations presented in this chapter. The notation used here is mostly standard, but the notation elsewhere in the text deviates from standard

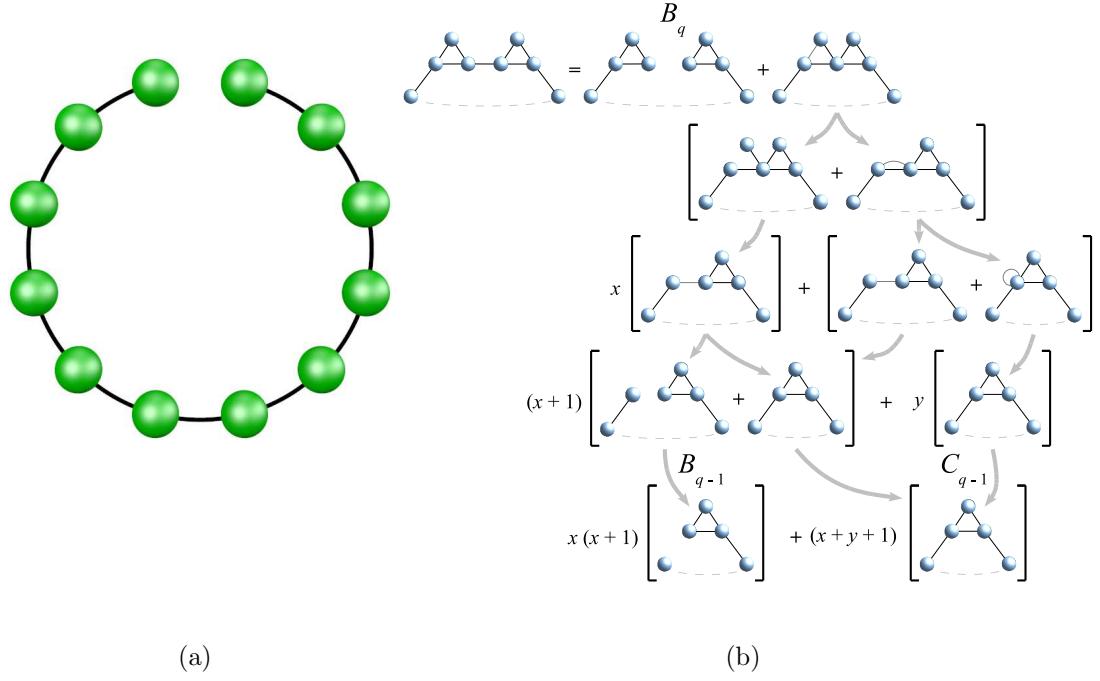


Figure 3.20: In panel (a), we depict a chain B_q of q cliques (complete sub-graphs of maximally connected clusters) of size l connected by single edges. The corresponding circle of q cliques C_q is depicted in Fig. 3.18. In panel (b), we show the derivation of the Tutte polynomial in Eq. (3.65) for size $l = 3$ cliques. We iteratively break edges and merge nodes according to the Tutte polynomial recurrence relation [142] in Sec. 3.8.1 until we arrive at configurations that are reduced clique circle components. For presentation purposes, gray edges are cut in the next line of the derivation. The dashed gray line at the bottom of each sub-graph represents the remainder of the clique circle which is not touched or affected by the operations on the displayed subgraph.

notation in order to facilitate the partition function derivation in Sec. 3.7.4. For an undirected graph G , we denote the *deletion* (removal) of an edge e by G' and a *contraction* of the edge by G'' where a contraction consists of removing the edge e and merging the corresponding vertices.

3.8.1 Unweighted graph G

If G has no edges, the Tutte polynomial is $t(G; x, y) = 1$. If G is a disjoint graph of partitions, then A and B $t(G; x, y) = t(A; x, y) t(B; x, y)$. When an edge e in an unweighted graph G is “*cut*,” the recurrence relations are:

- For a general edge, $t(G; x, y) = t(G'_e; x, y) + t(G''_e; x, y)$ which is the sum of two graphs where e is deleted and contracted.
- If edge e is an isthmus between two otherwise disconnected regions of G , then $t(G; x, y) = x t(G''_e; x, y)$ where the edge e is contracted.

- If edge e is a loop (a vertex self-edge), then $t(G; x, y) = y t(G'_e; x, y)$ where the edge e is deleted.

The resulting Tutte polynomial is a function of two variables (x, y) , and it is independent of the construction order. Different graphs G and H may be described by the same function $t(G; x, y) = t(H; x, y)$. A sample calculation is performed Sec. 3.9 for a circle of complete sub-graphs (cliques) as shown in Fig. 3.20(b).

Tutte polynomials are related to the partition function of a ferromagnetic ($J > 0$) or antiferromagnetic ($J < 0$) Potts model given by

$$H(\{\sigma\}) = - \sum_{i \neq j} J \delta(\sigma_i, \sigma_j) \quad (3.55)$$

for any connected pair of nodes i and j with an interaction strength J . The corresponding partition function is

$$Z = q^{k(G)} v^{|V|-k(G)} t(G; x, y) \quad (3.56)$$

where q is the number of clusters or states, $v = \exp(\beta J) - 1$, G denotes the graph, $k(G)$ is the number of connected components in G , $|V|$ is the number of vertices, $x = (q + v)/v$ and $y = v + 1$.

In Sec. 3.7.4, we use the following lemma to derive high temperature T approximation for a constructed random graph. We denote K_l as the Tutte polynomial for a complete graph, and $K_l^{(d)}$ denotes that the graph has d duplicated (possibly redundant) edges.

Lemma 1 *For a clique $K_l^{(d)}$ of size l with d duplicate edges between any pair of nodes, the Tutte polynomial at $y = 0$ is K_l .*

Let G be a complete graph with l vertices and $d = 1$ redundant edge. If we delete and contract the duplicate edge, the Tutte polynomial $t(G) \equiv K_l^{(d=1)}$ is

$$K_l^{(1)} = K_l + K_{l-1}^{(l-1)}$$

The contracted vertex in the second term contains $r = 1$ loop. We cut the loop and have

$$\begin{aligned} K_l^{(1)} &= K_l + yK_{l-1}^{(l-2)} \\ K_l^{(1)} &= K_l \end{aligned} \tag{3.57}$$

where we used $y = 0$ in the second line.

Now, assume that we can reduce $K_l^{(d)} = K_l$. Let G be a complete graph with l vertices and $d + 1$ duplicate edges. If we cut one duplicate edge, the resulting Tutte polynomial $t(G) \equiv K_l^{(d+1)}$ is

$$K_l^{(d+1)} = K_l^{(d)} + K_{l-1}^{(d+l-1)}$$

The contracted vertex in the second term contains $r \geq 1$ loops. We cut each loop in sequence and obtain

$$\begin{aligned} K_l^{(d+1)} &= K_l^{(d)} + y^r K_{l-1}^{(d+l-r-1)}. \\ K_l^{(d+1)} &= K_l^{(d)} \end{aligned} \tag{3.58}$$

Since $K_l^{(d)} = K_l$, we also equate $K_l^{(d+1)} = K_l$ by Eq. (3.58). Equation (3.57) shows that the relation holds for $d = 1$; therefore, by mathematical induction $K_l^{(d)} = K_l$ holds true for any integer $d \geq 1$.

3.8.2 Weighted graph G

An excellent summary of multivariate Tutte polynomials (MVTP) is found in Ref. [147]. The MVTP allows for arbitrary weights $\mathbf{v} = [v_e]$ for the edges $\{e\}$ of G . If G

has no edges, the MVTP is $Z(G; q, \mathbf{v}) = q$. For an undirected graph G , the weighted Potts Hamiltonian is

$$H(\{\sigma\}) = - \sum_{i \neq j} J_{ij} \delta(\sigma_i, \sigma_j). \quad (3.59)$$

When an edge e in G is “cut,” the recurrence relation is

$$Z(G; q, \mathbf{v}) = Z(G'; q, \mathbf{v}) + v_e Z(G''; q, \mathbf{v}) \quad (3.60)$$

where J_e corresponds to the edge weight between two nodes i and j and $v_e = \exp \beta J_e - 1$.

As with the unweighted case, if G is a disjoint graph of partitions A and B , then $Z(G; x, y) = Z(A; q, \mathbf{v}) Z(B; q, \mathbf{v})$. If partitions A and B are joined at a single vertex, then $Z(G; x, y) = Z(A; q, \mathbf{v}) Z(B; q, \mathbf{v})/q$. Unlike Eq. (3.56) for unweighted graphs, Eq. (3.60) holds for loops or bridges, but for concreteness, cutting an isthmus e yields

$$Z(G; q, \mathbf{v}) = (1 + v_e/q) Z(G'_e; x, y) \quad (3.61)$$

$$Z(G; q, \mathbf{v}) = (q + v_e) Z(G''_e; x, y) \quad (3.62)$$

where e is deleted or contracted, respectively. If e is a loop, then

$$Z(G; q, \mathbf{v}) = (1 + v_e) Z(G'_e; x, y). \quad (3.63)$$

Note that the MVTP *is* the partition function. That is, there are no prefactors of q or v_e . Finally, if two parallel edges connect the same pair of nodes i and j with weights J_1 and J_2 , then Z_G is unchanged if we replace the parallel edges by a single edge with a weight $J' = J_1 + J_2$ (this negates the need for lemma 1 above).

3.9 Derivation of the Tutte polynomial for a circle of cliques

As depicted in Fig. 3.18, we define C_q as a circle of q cliques where we focus those of size $l = 3$ for the current derivation. The Tutte polynomial for a triangle is $\Delta \equiv (x^2 + x + y)$. For convenience, we also define, $\Delta' \equiv (\Delta + x + 1) = [(x + 1)^2 + y]$ and $y' \equiv (x + y + 1)$.

We define B_q to be the Tutte polynomial for a clique chain as depicted in Fig. 3.20(a). In this case, it is trivial to construct B_q

$$B_q = x^{q-1} (x^2 + x + y)^q. \quad (3.64)$$

With Eq. (3.64), we construct a recurrence relation for the clique circle configurations as shown in Fig. 3.20(b)

$$C_q = B_q + x(x+1)B_{q-1} + (x+y+1)C_{q-1}. \quad (3.65)$$

From this relation, we can sum the series exactly.

$$\begin{aligned} C_q &= B_q + \Delta' B_{q-1} + x(x+1)(x+y+1)B_{q-2} \\ &\quad + (x+y+1)^2 C_{q-2} \\ &\quad \vdots \quad \vdots \\ C_q &= B_q + \Delta' \sum_{i=0}^{q-4} (x+y+1)^i B_{q-i-1} \\ &\quad + \Delta(x+y+1)^{q-3} B_2 + (x+y+1)^{q-2} C_2. \end{aligned} \quad (3.66)$$

Note that the last B_j term uses Δ not Δ' . Also, it can be shown that $C_2 = (x+1)^2(x^3+\Delta) + y(x+1)\Delta$. Substituting these values into the equation, we arrive at

$$\begin{aligned} C_q &= x^{q-1}\Delta^q + \Delta' \sum_{i=0}^{q-4} y'^i x^{q-i-2} \Delta^{q-i-1} + xy'^{q-3} \Delta^3 \\ &\quad + xy'^q (x^2 + x + 1) + yy'^{q-1} \Delta'. \end{aligned} \quad (3.67)$$

In the high temperature T limit, $y \ll x$, so we approximate $y \simeq 0$, and the equation simplifies to

$$\begin{aligned} C_q^{(T)} &\simeq x(x+1)^q \left[x^{2q-2} + \dots + x^2 + x + 1 \right] \\ &= x(x+1)^q \left[\frac{1-x^{2q-1}}{1-x} \right], \end{aligned} \quad (3.68)$$

We make a final high T approximation

$$C_q^{(T)} \simeq (x+1)^{q+1} x^{2q-3} \quad (3.69)$$

using $(x^{2q-1} - 1) \simeq x^{2q-1}$ and $(1-x)^{-1} \simeq (1+x)/x^2$.

3.10 Random knot “transitions”

A general 3D knot may be represented as a 4-valent planar graph [147] [*i.e.*, corresponding to a two-dimensional (2D) square lattice connectivity allowing self-loops]. This relation connects the Tutte polynomial to the Jones polynomial in knot theory. Conversely, all connected, signed planar graphs have a corresponding link diagram representation (2D knot projection). Alternating over-under crossings result in unsigned planar graphs [147] (e.g., the trefoil knot in Fig. 3.21). Ref. [148] provides an

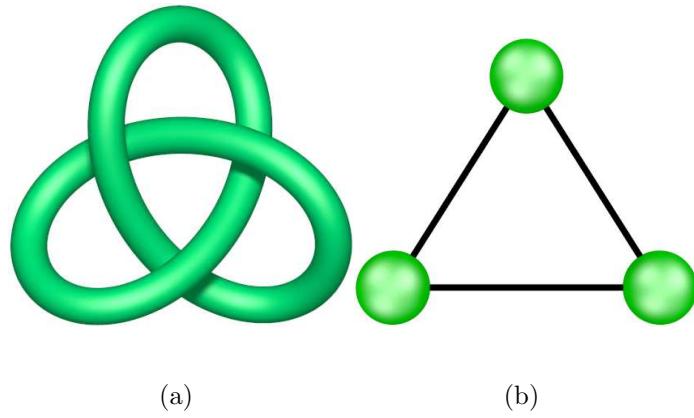


Figure 3.21: Panel (a) depicts the trefoil knot, and panel (b) shows the corresponding graph G constructed from the distinct knot regions and crossings [147]. That is, nodes correspond to “checkerboard-shaded” regions (shade the outside lobes of the trefoil knot leaving the interior region unshaded), and edges correspond to knot crossings. Jones polynomials $V_J(x)$ in knot theory are related to Tutte Polynomials, and Eq. (3.70) represents the trefoil knot corresponding to the triangle subgraph in panel (b).

introduction to the mathematics and physics of knot theory. The Jones polynomial of a given knot is intimately related to quantum field theories [149], via its connection to [an SU(2) type] Wilson loop associated the same knot.

As a concrete example, Fig. 3.21(a) depicts a simple trefoil knot which is related to the triangle clique depicted in Fig. 3.21(b) [147]. The Tutte polynomial of Fig. 3.21(b) is $K_3(G; x, y) = x^2 + x + y$. Then we generate the Jones polynomial

$$V_J(x) = x^2 + x + \frac{1}{x} \quad (3.70)$$

where we used $xy = 1$ because the trefoil knot has alternating crossings [150]. While the trefoil knot is clearly not random, we conjecture that the transitions detected in random graphs with embedded ground states in the current work can have similar transition repercussions in random knots.

3.11 Conclusions

We systematically examined the phase transitions for the community detection problem via a “noise test” across a range of parameters. The noise test consists of a structured graph with a strongly-defined ground state. We add increasing numbers of extraneous intercommunity edges (noise) and test the performance of a stochastic community detection algorithm in solving for the well-defined ground state. Specifically, we studied two types (sequences) of systems. In the first such sequence of systems in Fig. 3.2, we fixed the ratio $\alpha = q/N$ of the number of communities q to the number of nodes N . We fixed q at different values and varied N in the second se-

quence of systems in Fig. 3.6. In Fig. 3.10, we explored the largest tested systems with $N = 2048$ nodes in more detail where we depicted additional measures to illustrate the transitions. All of these systems showed regions with distinct phase transitions in the large N limit. Deviations occurred most often in smaller systems indicating a definite finite-size effect.

The spin-glass-type phase transitions in our noise test occurred between solvable and unsolvable regions of the community detection problem. A hard, but solvable, region lies at the transition itself where it is difficult, in general, for any community detection algorithm to obtain the correct solution. We analyzed a system of non-interacting cliques and illustrated that in the large q limit, the system experiences a thermal disorder in the thermodynamic limit for any non-zero temperature. When in contact with a heat bath, the asymptotic behavior of the temperatures beyond which the system is permanently disordered varies slowly with the number of communities q , specifically, $T_x \simeq O[1/\log q]$. This implies that problems of practical size maintain a definite region of solvability. Given the connection between Jones polynomials of knot theory and Tutte polynomials for the Potts model, our results imply similar transitions in large random knots (see Sec. 3.10).

We further studied the free energy of arbitrary graphs arriving at the same conclusion. Increasing number of communities q emulates increasing T in arbitrary graphs for a general Potts model. The effective interaction strength for increasing q scales such that this disorder is circumvented by the often standard use of a simulated annealing algorithm, but the “glassy” (high noise) region remains a challenge for any

community detection algorithm.

Chapter 4

Image segmentation

4.1 Introduction

“Image segmentation” refers to the process of partitioning a digital image into multiple segments based on certain visual characteristics [42, 43, 44]. Image segmentation is typically used to locate objects and boundaries in images. The result of image segmentation is a set of segments that collectively cover the entire image or a set of extracted contours of the image. This problem is challenging (see, e.g., Fig. (4.1)) and important in many fields. Examples of its omnipresent use include, amongst many others, medical imaging [46] (e.g., locating tumors and anatomical structure), face recognition [47], fingerprint recognition [48], and machine vision [50]. Numerous algorithms and methods have been developed for image segmentation. These include thresholding [51], clustering [52], compression [53] and histogram based [54] approaches, edge detection [55], region growing [56], split and merge [57], gradient

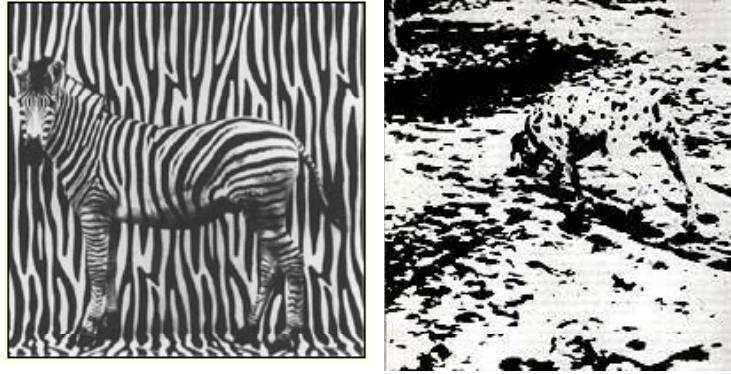


Figure 4.1: Examples of currently challenging problems in image segmentation. Left: The left image is that of zebra (courtesy of Ref.[161]) with the a similar “stripe” background. Right: The image on the right is that of a dalmatian dog [162]. Most people do not initially recognize the dog before given clues as to its presence. Once the dog is seen it is nearly impossible to perceive the image in a meaningless way. [162]

flows and partial differential equation based approaches [151, 152], graph partitioning methods and normalized cuts [58, 153], Markov random fields and mean field theories [154, 155, 156, 157], watershed transformation [59], random walks [158], isoperimetric methods [159], neural networks [160], and a variety of other approaches, e.g., [60, 61, 62].

In this chapter, we will apply a “community detection” algorithm to image segmentation. This method belongs to the graph partitioning category. Community detection [1, 10, 12, 23, 26] seeks to identify groups of nodes densely connected within their own group (“community”) and more weakly connected to other groups. A so-

lution enables the partition of a large physically interacting system into optimally decoupled communities. The image is then divided into different regions (“communities”) based on a certain criterion, and each resulting region corresponds to an object in the original image.

It is notable that by virtue of its graph theoretical nature, community detection is suited for the study of arbitrary dimensional data. However, unlike general high dimensional graphs, images are two (or three) dimensional. Thus, real images are far simpler than higher dimensional data sets as, e.g., evinced by the four color theorem stating that four colors suffice to mark different neighboring regions in a segmentation of any two dimensional image. Thus, geometrical (and topological) constraints can be used to further improve the efficiency of the bare graph theoretical method. In [63, 64], in the context of analyzing structures of complex physical systems such as glasses, we used geometry dependent physical potentials to set the graph weights in various two and three dimensional systems. In the case of image segmentation, in the absence of underlying physics, we will invoke geometrical cut-off scales.

In this work, we will discuss “unsupervised” image segmentation. By this term, we allude to a general multi-purpose segmentation method based on a general physical intuition. The current method does not take into account initial “training” of the algorithm- i.e., provide the system with known examples in which specific patterns are told to correspond to specific objects. We leave the study of supervised image segmentation and more sophisticated extensions of our inference procedure to a future work. One possible avenue which can be explored is the use of inference beyond that

relating to different “replicas” in the simple form discussed in this chapter that is built on prior knowledge (and prior probabilities in a Bayesian type analysis) of expected patterns in the studied images.

We will heavily employ multi-resolution community detection [10, 63, 64, 26, 27, 163]. Specifically, we will apply a particular multi-resolution scheme of the type introduced in [27] to investigate the overall structure at different resolutions in the test images. Similar to [27], we will employ information based measures (e.g., the normalized mutual information and the variation of information) to determine the significant structures at which the “replicas” (independent solutions of the same community detection algorithm) are strongly correlated. With the aid of these information theory correlations, we illustrate how we may discern structures at different pertinent resolutions (or spatial scales). An image may be segmented at different levels of detail and scales by setting the resolution parameters to these pertinent values. We demonstrate in a detailed study of various test cases, how our method works in practice and the resulting high accuracy of our image segmentation method.

4.2 Potts Models

In what follows, we will briefly elaborate on our particular Potts model representations of images and the corresponding Hamiltonians (energy or cost functions), also in Sec.2 in Chapter 2.

4.2.1 Representation

As is well appreciated, different objects in an image or more general communities in complex graph theoretical problems are ultimately denoted by a “Potts type” [164] variable σ_i . That is, if node i lies in a community number w then $\sigma_i = w$. If there are q communities in the graph then σ_i can assume values $1 \leq \sigma_i \leq q$. A state $\{\sigma_s\}_{s=1}^N$ corresponds to a particular partition (or segmentation) of the system into q communities (or objects). In the context of image segmentation, Potts model representations can, e.g., also be found in [165, 166, 167, 168]. Our algorithm shares certain similarities to some other multi-scale approaches [10, 17, 26, 163] such as “superparamagnetic clustering” [169] and scaling modularity-like quality functions [26]. Some other multi-scale approaches operate via more diverse methods.

4.2.2 Potts model Hamiltonian for unweighted graphs

In Chapter 2, a particular Potts model Hamiltonian was introduced for community detection. The ground states of this Hamiltonian (or lowest energy states) correspond to optimal partition of the nodes into communities. This Hamiltonian *does not involve a comparison relative to random graphs (“null models”)* [10] and as such was free of the “resolution limit” problems [10, 86, 170] wherein the number of found communities or objects scaled with the system size in a way that was independent of the actual system studied. In what follows below, there are N elementary nodes in a graph (or pixels in an image), we consider general *unweighted* graphs in which any pair of nodes

may be either linked with a uniform weight or not linked at all. Specifically, a link between sites i and j is associated with edge weights A_{ij} and J_{ij} . In these unweighted graphs, A_{ij} is an element of the adjacency matrix. That is, $A_{ij} = 1$ if nodes i and j are connected by an edge and $A_{ij} = 0$ otherwise. The weights $J_{ij} = (1 - A_{ij})$.

The goal of the general (or “absolute”) Potts model Hamiltonian [18] was to energetically favor any pair of linked nodes to be in the same community, to penalize for a pair of unlinked nodes to be in the same community and conversely for nodes in different communities (penalize for having two linked nodes be in different communities and favor disjoint nodes being in different communities). Putting all of these bare energetic considerations together (sans any comparisons to random graphs), the resulting Potts model Hamiltonian (or energy function) for a system of N nodes simplifies to [18, 171]

$$\mathcal{H}(\{\sigma_s\}_{s=1}^N) = -\frac{1}{2} \sum_{i \neq j} [A_{ij} - \gamma(1 - A_{ij})] \delta(\sigma_i, \sigma_j). \quad (4.1)$$

In Eq.(4.1), we emphasize the dependence of the Hamiltonian on the N different variables $\{\sigma_s\}$ at each lattice site s (each of which can assume q values). In what follows, the dependence of the Hamiltonian on $\{\sigma_s\}_{s=1}^N$ will always be understood.

The Kronecker delta $\delta(\sigma_i, \sigma_j) = 1$ if $\sigma_i = \sigma_j$ and $\delta(\sigma_i, \sigma_j) = 0$ if $\sigma_i \neq \sigma_j$. In this Hamiltonian, by virtue of the $\delta_{\sigma_i \sigma_j}$ term, each spin σ_i interacts only with other spins in its own community. As such, the resulting model is *local*—a feature that enables high accuracy along with rapid convergence [18].

As noted above, minimizing this Hamiltonian corresponds to identifying tightly

connected clusters of nodes. The parameter γ is the so called “resolution parameter” which adjusts the relative weights of the linked and unlinked edges, as in Eq. (4.1). This is easily seen by inspecting Eq.(4.1). A high value of γ leads to forbidding energy penalties unless all intra-community nodes “attract” one another and lie in the same community. By contrast, $\gamma = 0$ does not penalize the inclusion of any additional nodes in a given community and the lowest energy solution generally corresponds to the entire system.

4.2.3 Potts Model Hamiltonian for Weighted Graphs

In *weighted* graphs, we assign edges between nodes with the respective weights based on the interaction strength (e.g., the (dis-) similarity of the intensity or color defines the edge weight in image segmentation problem). Specifically, in image segmentation problems, we determine (based on, e.g., color or intensity differences) the weight V_{ij} between each member of a node pair. We then shift each such value by an amount set by a background \bar{V} , i.e., $V'_{ij} = (V_{ij} - \bar{V})$. The subtraction relative to the background of \bar{V} allows for our community detection algorithm to better partition the network of pixels. In principle, this background can be set to be spatially non-uniform. However, in this work we set \bar{V} to be a constant. Thus, we generalize the earlier model of [27] in Eq. (4.1) by the inclusion of a background \bar{V} and by allowing for continuous weights V_{ij} instead of discrete weights that are prevalent in graph theory. The resulting Hamiltonian [63, 64] reads

$$\mathcal{H} = \frac{1}{2} \sum_{a=1}^q (V_{ij} - \bar{V}) [\Theta(\bar{V} - V_{ij}) + \gamma \Theta(V_{ij} - \bar{V})] \delta(\sigma_i, \sigma_j). \quad (4.2)$$

The form of this Hamiltonian and that of Eq. (4.1) was inspired by positive and negative energy terms that favor the formation of tightly bound clusters (or “solutes”) that are more weakly coupled to their surroundings [171]. Similar to the important effects of the solute found in physical systems [172], the Hamiltonian of Eq. (4.2) captures all interactions in the system [171]. Earlier [63, 64]), we invoked the Hamiltonian of Eq. (4.2) to analyze static and dynamic structures in glasses.

We will discuss our minimization procedure in Sec. 4.4. When running and extending the initial community detection algorithm of [27] with the use of Hamiltonian of Eq. (4.2), we effectively view the pertinent segments of the image as initial nuclei and use energy comparison conditions to see if more vertices should be added to the existing nuclei. This nucleation type approach might be of use in other computational problems when pertinent features of systems are to be found and segmented accordingly.

In Eq. (4.2), the number of communities q may be specified from the input or left arbitrary and have the algorithm decide by steadily increasing the number of communities q for which we have low energy solutions. The Heavyside functions $\Theta(x)$ “turns on” or “off” the edge designation [$\Theta(x > 0) = 1$ and $\Theta(x < 0) = 0$] relative to the aforementioned background \bar{V} . As before, minimizing the Hamiltonian of Eq. (4.2) corresponds to identifying strongly connected clusters of nodes.

While in Eq. (4.1 (or Eq. 4.2),) the input concerns two-point ($p = 2$) edge weights

V_{ij} (or A_{ij}) , it is, of course, possible to extend these Hamiltonian to allow for more general motifs (such as $p = 3$ node triangles) and include $p \geq 3$ point weights V_{ijk} (and extensions thereof). These correspond to p spin interactions. In the current study, however, we limit ourselves to $p = 2$ node weights.

4.3 Casting Images as Networks

We will now detail how we translate images into networks with general edge weights that appear in Eqs.(4.1, 4.2). We will represent pixels as the nodes in a graph. Edge weights define the (dis-) similarity between the neighborhood pixels.

Images may be broadly divided into two types: (a) those with the uniform and (b) those with varying intensity. “Uniform intensity” means that the entire object or each component is colored by one intensity or color. By the designation of “varying intensity”, we allude to objects or components that exhibit alternating intensities or colors, e.g., the stripes and spots seen in Fig. (4.1).

Regarding the above two types of images, two different methods may be employed to define the edge weights: (i) The intensity/color difference between nodes is defined as the edge weight in images with uniform intensity. (ii) The overlap between discrete Fourier transforms of blocks is defined as the edge weight in images with varying intensity. The second method is designed to distinguish the target and the background by their specific frequencies. We will detail both methods below in Sec. 4.3.1 (where we discuss images with uniform intensities) and Sec. 4.3.2 (spatially

varying intensities).

4.3.1 Edge definition for images with uniform intensity

For images of uniform intensity, we will define edges based on the color (dis-) similarity. For the unweighted Potts model of Eq. (4.1), we will assign an edge between two pixels (i and j) if the “color” difference (D_{ij}) between them is less than some threshold (\bar{V}). That is,

$$A_{ij} = \Theta(\bar{V} - D_{ij}). \quad (4.3)$$

For weighted Potts model in Eq. (4.2), we will, as we will elaborate on momentarily, set the weights V_{ij} to be the “color” difference (D_{ij}) between nodes i and j , i.e.,

$$V_{ij} = D_{ij}. \quad (4.4)$$

As seen from the energy functions of Eqs. (4.1, 4.2), a large dis-similarity V_{ij} favors nodes i and j being in different clusters.

A grey scale image is an image that in which the value of each pixel carries only intensity information. Images of this sort are composed exclusively of shades of gray, varying from black at the weakest intensity ($I = 0$) to white at the strongest ($I = 255$). For a grey-scale image, the “color” difference is the absolute value of the intensity difference, i.e.,

$$D_{ij} = |I_i - I_j|. \quad (4.5)$$

A “color image” is an image that includes color information for each pixel. Each pixel contains three color components: red, green and blue (or RGB). The value of the

intensity of each of these three components may attain any of 2^8 values (any integer in the interval $[0, 255]$). For a color-scale image, we define the “color” difference as the average of the differences between the color components red, green and blue. That is, with R_i , G_i , and B_i respectively denoting the strengths of the red, green, and blue color components at site i , we set

$$D_{ij} = \frac{1}{3}(|R_i - R_j| + |G_i - G_j| + |B_i - B_j|). \quad (4.6)$$

We do not store edges between every pair of nodes. Rather, edges connect nodes whose distance is less than a tunable value Λ .

4.3.2 Edge definition for images with varying intensities

Typically, images with varying intensities contain different patterns. To separate these patterns, we construct a “block-structure” containing the quintessential pattern information. We next introduce a method to divide blocks and then elaborate on two different ways to connect edges between blocks.

General contending pertinent scales may be determined by, e.g., examining the peaks of the Fourier transform of an entire image (whose location yields the inverse wave-length and whose width reveals the corresponding spatial extent of these modulations). While such simple transforms may aid optimization in determining candidate parameter scales, our algorithm goes far beyond such simple global measures.

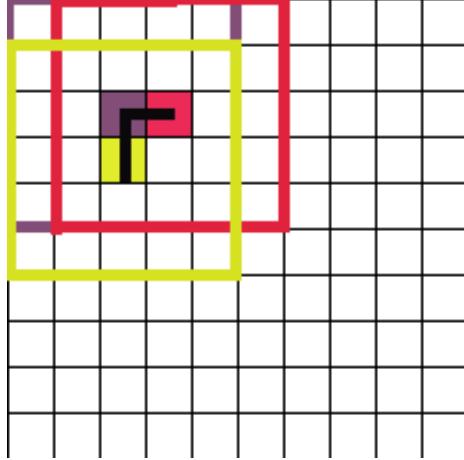


Figure 4.2: An example of overlapping blocks. The block size is $L_x \times L_y = 5 \times 5$. The nearest neighbor of the block enclosed by “purple” in x-direction is the one enclosed by “red”, and its nearest neighbor in y-direction is the one enclosed by “yellow”. They are connected due to the nearest neighbor condition.

Overlapping blocks

We will divide an entire image of size $N = N_x \times N_y$ into N overlapping blocks. These blocks are centered about each (of the N) pixels and are of size $L_x \times L_y$. The dimensions of the individual blocks are, generally, far smaller than that of the entire system, $L_{x,y} \ll N_{x,y}$. General scales can be gleaned from a Fourier transform of the entire image.

To construct the connection matrix between the blocks, we connect edges between each pair of blocks and set the distance between the nearest block pair to be 1. This choice has the benefit of overlapping the nearest neighbor blocks, which share more commons. Fig. 4.2 gives a schematic plot of the “overlapping block” structure.

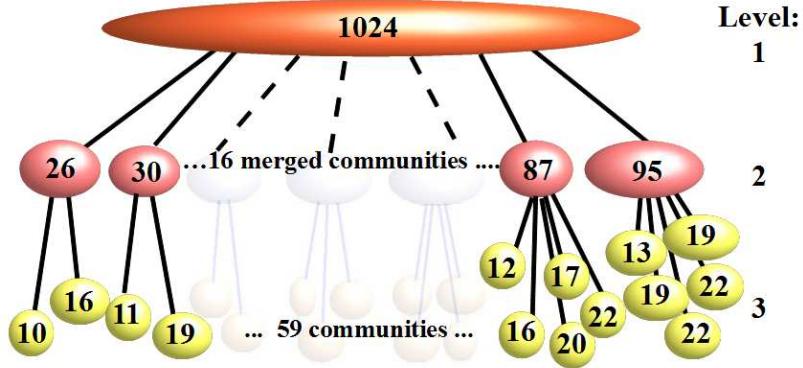


Figure 4.3: Heterogeneous hierarchical system corresponding to the plots in Fig. 4.4. In this figure, the 1024 node system is divided into a three-level hierarchy. Level 3 has 59 communities with sizes from 10 to 24 nodes. Level 2 has 16 communities with sizes from 26 to 95 nodes. Level 1 is the completely merged system of 1024 nodes. The average edge density is $p = 0.054$. This system has 28185 edges.

Average intensity difference between blocks

Following the construction of the overlap blocks structure, we next compute the average intensity of each block and connect the edges between blocks based on the difference. In this case, each block can be treated as a “super-node” which contains the pattern information of the studied image.

To further incorporate geometrical scales, we multiply the edge weights by $\exp(-|\mathbf{r}_m - \mathbf{r}_n|/\ell)$ (where ℓ is a tunable length scale and the vectors \mathbf{r}_m and \mathbf{r}_n denote the spatial locations of points m and n). We remind the reader that there are N basic blocks

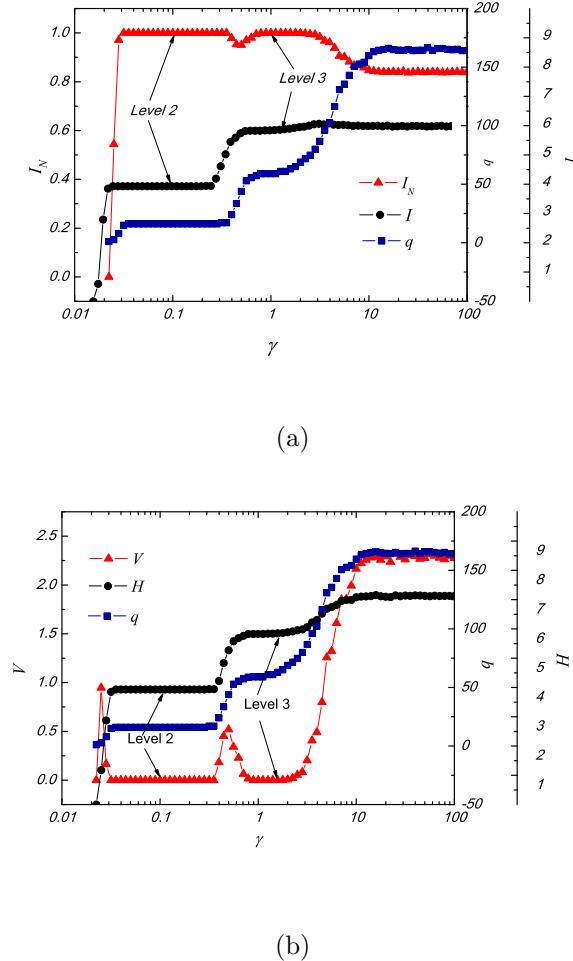


Figure 4.4: Plot of information measures I_N , V , H and I vs the Potts model weight γ in Fig. 4.3. In panel(a), the peak (plateau) I_N denoted by the arrows correspond to levels 2 and 3 of the hierarchy depicted in Fig. 4.3. Similarly in panel (b), the minimal V values, indicated by arrows, accurately correspond to levels 2 and 3 of the hierarchy. The number of communities q is 16 and 59 in disparate plateau regions (denoted by the arrows) in both panels. These communities assignments (and, obviously, also their numbers) are exactly the same as those of the communities in levels 2 and 3 of the original hierarchical graph of Fig. 4.3. In panels (a) and (b), both the mutual information I and the Shannon entropy H display a plateau behavior corresponding to the correct solutions.

and thus N possible values of m (and N possible values of n). We will set in Eq. (4.2), the weights V_{mn} between block m and n to be

$$V_{mn} = \frac{D_{mn} \exp(-|\mathbf{r}_m - \mathbf{r}_n|/\ell)}{L_x L_y}, \quad (4.7)$$

where

$$D_{mn} = (1 - \delta(m, n)) \left| \sum_{i=0}^{i=L_x-1} \sum_{j=0}^{j=L_y-1} (I^m(i, j) - I^n(i, j)) \right|. \quad (4.8)$$

As seen in Eq. (4.8), D_{mn} is the sum of the absolute values of the intensity differences between blocks m and n with each of these blocks being of size $L_x \times L_y$. In Eq. (4.7), $|\mathbf{r}_m - \mathbf{r}_n|$ is the physical distance between block m and n (i.e., the distance between the central nodes of each block).

The geometrical factor of $(\exp(-|\mathbf{r}_m - \mathbf{r}_n|/\ell))$ in Eq. (4.7) with a tunable length scale ℓ can be set to prefer (and, as we will illustrate also to detect) certain scales in the image. This enables the algorithm to detect clusters of varying sizes that contain rich textures.

Fourier amplitude derived weights

As it is applied to image segmentation, the utility of Fourier transformations is well appreciated. We next discuss how to invoke these in our Potts model Hamiltonian. To highlight their well known and obvious use, we note that, e.g., the stripes of the

zebra in Fig. 4.1 contain wave-vectors which are different from those of the more uniformly modulated background. Thus, a spatially local Fourier transform of this image may distinguish the zebra from the background. We will now invoke Fourier transforms in a general way in order to determine the edge weights in our network representation of the image.

With the preliminaries of setting up the block structure in tow, we now apply a discrete Fourier transform inside each block. Rather explicitly, excluding the spatial origin, the local discrete $2 - D$ Fourier transform of a general quantity f_m within block m with internal Cartesian coordinates (a, b) is

$$F_m(k, l) = \sum_{a=0}^{L_x-1} \sum_{b=0}^{L_y-1} f_m(a, b) e^{-2\pi i (\frac{ka}{L_x} + \frac{lb}{L_y})} - f_m(0, 0). \quad (4.9)$$

The wave-vector components $k = 0, 1, \dots, L_y - 1$ and $l = 0, 1, \dots, L_x - 1$. In applications, we set, for grey-scale images, $f_m(a, b)$ to be the intensity I at site (a, b) in block m (a whose location relative to the origin of the entire image we denote by $\mathbf{r}_{ab;m}$). That is, $f_m(a, b) = I(\mathbf{r}_{ab;m})$. In color images, we set f to be the average over the intensity of the red, green and blue components: $f(a, b) = \frac{1}{3}(R(a, b) + G(a, b) + B(a, b))$. We fix the couplings J_{mn} between blocks m and n to be

$$J_{mn} = \sum_{k=0}^{L_x-1} \sum_{l=0}^{L_y-1} |F_m^*(k, l) F_n(k, l)|. \quad (4.10)$$

We connect blocks whose spatial separation is less than the aforementioned tunable cutoff distance Λ by links having edge weights V_{mn} . In practice, we fixed Λ . With Eq. (4.10) in hand, we set

$$V_{mn} = (\delta(m, n) - 1) J_{mn} \exp(-|\mathbf{r}_m - \mathbf{r}_n|/\ell) \quad (4.11)$$

in Eq. (4.2). In this case, the background \bar{V} would be negative.

When inverting the sign of the left hand side of Eq.(4.11) (shown in later section), our algorithm will be also suited for the detection of changing objects against a more uniform background.

We now briefly comment on the relation between the Fourier space overlaps and weights in Eqs.(4.10,4.11) and the real space overlaps and weights in Eqs.(4.7,4.8). It is notable that in Eq.(4.10), we sum over the *modulus* of the products of the Fourier amplitudes. By Parseval's theorem, sans the modulus in the summation in Eq.(4.10), J_{mn} would be identical to the overlap in real space between f_m and f_n . Such real space overlaps directly relate to the real-space overlaps in Eq.(4.8) [following a replacement of the absolute value in Eq.(4.8) by its square and an overall innocuous multiplicative scale factor]. Thus, without the modulus in Eq.(4.10), the Fourier space calculation outlined above affords no benefit over its real space counterpart. Physically, the removal of the phase factors when performing the summation in Eq.(4.10) avoids knowledge of the relative location of the origins between different blocks. This allows different regions of a periodic pattern to be strongly correlated and clumped together. By contrast, for a periodic wave of a particular wave-vector, the real space overlaps between blocks m and n may vanish when the origins of blocks m and n are displaced by a real space distance that is equal to half of the wave-length of the periodic wave along the modulation direction. Thus, the real space weights as derived from Eqs.(4.7,4.8) may vanish when the corresponding Fourier space derived

weights (Eqs.(4.10,4.11)) are sizable.

It is possible to improve on the simple Fourier space derived weights by a general wavelet analysis.

4.3.3 Edge definition for images with “mixed type” intensities

For images with intensities of “mixed type,” i.e., those which harbor both (i) uniform and (ii) varying intensity regions, we may determine edge weights by invoking the Fourier analysis method discussed in Sec. 4.3.2. In this way, we are able to detect objects with varying intensity as one cluster, as well as boundaries of regions with uniform intensity. Varying intensity regions have a different Fourier space structure than others and are, consequently, easily detectable by this method. That is, the edge weights within these regions are very different from the outside, and this difference will lead to the detection of these regions. Furthermore, boundaries of individual uniform intensity regions are formed by intersections with the background or other uniform regions. These boundaries can be treated as the object of the “varying intensity”, thus can be detected by Fourier analysis.

4.4 The community detection algorithm

Our community detection algorithm for minimizing Eqs. (4.1, 4.2) follows four steps [18].

(1) We partition the nodes based on a “symmetric” or “fixed q ” initialization where q is the number of communities.

- Symmetric initialization alludes to an initialization where each node forms its own community (i.e., initially, there are $q_0 = N$ communities).
- Fixed q initialization corresponds to a random initial distribution of all nodes into q communities.

For image segmentation, a symmetric initialization is used for the “*unsupervised*” case. In this case, the algorithm does not know what the number of communities are, so the symmetric initialization provides the advantage of no bias towards a particular community. The algorithm will decide the number of community q by means of the lowest energy solution.

Fixed q initialization may be used in a “*supervised*” image segmentation. The community membership of individual node will be changed to lower the solution energy. One has to decide how much information is needed by observing the original image and enter the number of communities q as an input. Different levels of information correspond to different number of communities q . For instance, if only one target needs to be identified, $q = 2$ is enough. The q communities include the target and background.

In the following sections, we will use the unsupervised image segmentation and let the algorithm decide the community number q .

(2) Next, we “pick up” [173] each node and place it in the community that best lowers the energy of Eqs. (4.1) and (4.2) based on the current state of the system.

(3) We repeat this process for all nodes and continue iterating until no energy lowering moves are found after one full cycle through all nodes.

(4) We repeat these processes for t trials and select the lowest energy result as the best solution. As discussed in Sec. 4.2, different trials differ solely by the permuted node order of the initial state. In the multi-resolution application discussed below, we further use different “replicas” for the “permuted” (randomized) variants and solve the system at a finite temperature T algorithm ([40]).

4.5 Multi-scale networks

After determining for the adjacency matrix in Sec. 4.3.1 and Sec. 4.3.2, we now turn to the so-called “resolution parameter” (γ) in Eq. (4.1)/Eq. (4.2). In [27], we introduced the multiresolution algorithm to select the best resolution. Our multiscale community detection was inspired by the use of overlap between replicas in spin-glass physics. In the current context, we employ information-theory measures, to examine contending partitions for each system scale. Decreasing γ , the minima of Eqs. (4.1, 4.2) lead to solutions progressively lower intra-community edge densities,

effectively “zooming out” toward larger structures. We determine all natural graph scales by identifying the values of γ for which the earlier mentioned “replicas” exhibit extrema in the average of information theory overlaps such as the normalized mutual information (I_N) and the variation of information (V) when expressed as functions of γ , ℓ . The extrema and plateau of the average information theory overlaps as a function of γ , ℓ over all replica pairs indicate the natural network scales [27]. The *replicas* can be chosen to be identical copies of the same system for the detection of static structures, e.g., the image segmentation.

We have introduced the information theory measures Sec. 2.5 in Chapter 2. Now, here is a key idea employed in [27] which will be of great use in our image segmentation analysis: *when taken over an entire ensemble of replicas*, the average I_N or V indicates how strongly a given structure dominates the energy landscape. High values of I_N (or low values of V) corresponds to more dominant and thus more significant structure. From a local point of view, at resolutions where the system has well-defined structure, a set of independent replicas should be highly correlated because the individual nodes have strongly preferred community memberships. Conversely, for resolutions “in-between” two strongly defined configurations, one might expect that independent replicas will be less correlated due to “mixing” between competing divisions of the graph.

4.5.1 The application of the multiresolution algorithm for a hierarchical network example

We will shortly illustrate how the multiresolution algorithm [27] works in practice by presenting an example of the multiresolution algorithm as it is applied to a hierarchical test system of $N = 1024$ nodes.

To begin the multiresolution algorithm, we need to specify the *number of replicas* r at each test resolution, the *number of trials per replica* t , and the starting and ending resolution $[\gamma_0, \gamma_f]$. Usually, the number of replicas is $8 \leq r \leq 12$, the number of trials is $2 \leq t \leq 20$. As detailed in Section 2.3, we select the lowest energy solution among the t trials for each replica. The initial states within each of the replicas and trials are generated by reordering the node labels in the “symmetric” initialized state of one node per community. These permutations P simply reorder the node numbers $(1, 2, 3, \dots, i, \dots, N) \rightarrow (P1, P2, \dots, PN)$ (with Pi the image of i under a permutation) and thus lead to a different initial state.

- (1) The algorithm starts from the initialization of the system described in item (1) of Section 4.4.
- (2) We then minimize Eq. (4.1) independently for all replicas at a resolution $\gamma = \gamma_i \in [\gamma_0, \gamma_f]$ as described in Section 4.4. Initially $i = 0$ (i.e., $\gamma = \gamma_0$).
- (3) The algorithm then calculates the average inter-replica information measures like I_N and V at that value of γ .

(4) The algorithm then proceeds to the next resolution point $\gamma_{i+1} \in (\gamma_0, \gamma_f]$ (with $\gamma_{i+1} > \gamma_i$).

(5) We then return to step number **(3)**.

(6) After examining the case of $\gamma = \gamma_f$, the algorithm outputs the inter-replica information theory overlaps for entire the range of the resolutions studied (i.e., γ on the interval $[\gamma_0, \gamma_f]$).

(7) We examine those values of γ corresponding to extrema in the average inter-replica information theory overlaps. Physically, for these values the resulting image segmentation is locally insensitive to the change of scale (i.e., the change in γ) and generally highlights prominent features of the image.

With A and B denoting graph partitions in two different replicas and $Q(A, B)$ their overlap, these average inter-replica overlaps for a general quantity Q [27] are explicitly

$$\langle Q \rangle = \frac{1}{r(r-1)} \sum_{A \neq B} Q(A, B). \quad (4.12)$$

Similarly, for a single replica quantity (such as the Shannon entropy H for partitions A in different replicas), the average is, trivially, $\langle Q \rangle = \sum_A H(A)/r$. (Averages for higher order inter-replica cumulants may be similarly written down with a replica symmetric probability distribution function [27].)

Fig. 4.4 shows the result of multiresolution algorithm applied to the three-level hierarchy system in Fig. 4.3. The system investigated is that of a standard simple

graph with unweighted links (i.e., in Eq.(4.1), $A_{ij} = 1$ if nodes i and j share an edge and is zero otherwise). In Fig. 4.3, “level 3” communities exhibit a density of links $p_3 = 0.9$ (i.e., a fraction p_3 of the intra-community node pairs are connected by a link ($A_{ij} = 1$)). The individual communities in level 3 have sizes that range between 5 to 24 nodes. The less dense communities in level 2 harbor a density of links $p_2 = 0.3$; the nodes in this case, are divided into five groups with sizes that vary from 26 to 95. Highest up in the hierarchy is the trivial level 1 “partition”—that of a completely merged system of 1024 nodes—with $p_1 = 0.025$ between nodes in different subcommunities. Thus, as a function of γ , this easily solvable system exhibits “transitions” between different stable solutions corresponding to different regions (or basins) of γ . In Sections(4.8, 4.9.6), we will further discuss additional transitions between easy solvable regions and regions of parameter space which are very “hard” or impossible (“unsolvable”).

Fig. 4.4(a) depicts the averages of I_N (on the left axis) and I (right axis) over all replica pairs. (We further provide in this figure the number of communities q .) A “cascade” composed of three plateaus is evident in these information theory measures. Similarly, Fig. 4.4(b) shows the V in left axis and H in right axis average over all replica pairs. The extrema denoted by the arrows in both panel (a) and (b) are the correctly identified levels 2 and 3 respectively of the hierarchy depicted in Fig. 4.3. The two plateaus with the peak values in panel(a) correspond to a normalized mutual information of size $I_N = 1$ (the highest theoretically possible) and similarly the corresponding minima in panel(b) have a variation of information $V = 0$ (the

smallest value possible) for the same range of γ values. These extreme values of I_N and V indicate perfect correlations among the replicas for both levels of the hierarchy. The “plateaus” in H , I and q are also important indicators of system structure. These plateau (and more general extrema elsewhere) illustrate when the system is insensitive to parameter changes and approaches stable solutions. In Section 4.8 (and in Eq.(4.17) in particular), we will discuss this more generally in the context of the phase diagram of the community detection problem.

4.6 Replica correlations as weights in a graph

Within the multiresolution method, significant structures are identified by strongly-correlated replicas (multiple copies of the studied system). Thus, if a node pair is always in the same community in all replicas, the two nodes must have strong preference to be connected or have a large edge weight. Similarly, if a node pair is not always in the same community in all replicas, they must have preference not to be connected or have a small edge weight. We re-assign edge weights based on the correlations between replicas.

Specifically, we first generate r replicas by permuting the “symmetric” initialized state of one node per cluster of the studied system, then apply our community detection algorithm to each replica and record the community membership for each node. We then calculate the probability of each node pair based on the statistics of replicas. The probability is defined as follows

$$p_{ij} = \frac{1}{r} \sum_{\alpha=1}^r \omega_{ij}^\alpha, \quad (4.13)$$

where

$$\omega_{ij}^\alpha = \delta_{\sigma_i^\alpha, \sigma_j^\alpha} + (1 - \delta_{\sigma_i^\alpha, \sigma_j^\alpha}) \frac{\exp(-|\mathbf{r}_i^\alpha - \mathbf{r}_j^\alpha|/\ell)}{N_A^\alpha N_B^\alpha}. \quad (4.14)$$

In Eq. (4.14), when node i and j belongs to the same community in replica α , i.e., $\delta_{\sigma_i^\alpha, \sigma_j^\alpha} = 1$, $\omega_{ij}^\alpha = 1$. When node i and j are not in the same community in replica α , i.e., $\delta_{\sigma_i^\alpha, \sigma_j^\alpha} = 0$, $\omega_{ij}^\alpha = \frac{\exp(-|\mathbf{r}_i^\alpha - \mathbf{r}_j^\alpha|/\ell)}{N_A^\alpha N_B^\alpha}$ (we use A and B to represent these two different communities, where $i \in A$ and $j \in B$. N_A^α and N_B^α denote the sizes of clusters A and B in replica α). As throughout, $|\mathbf{r}_i^\alpha - \mathbf{r}_j^\alpha|$ is the distance between node i and j in replica α . In Eq. (5.5), we sum the probability in each replica to define the edge weight. The assigned weights given by Eq. (5.5) are based on a frequency type inference. Although we will not report on it in this work, it is possible to perform Bayesian analysis with weights (“priors”) that are derived from a variant of Eq. (4.14); this enables an inference of the correlations from the sequence of results concerning the correlations between nodes i and j in a sequence of different replicas.

In unweighted graphs, we connect nodes if the edge weight between the node pair is larger than some threshold value \bar{p} in Eq. (4.1), i.e.,

$$A_{ij} = \Theta(p_{ij} - \bar{p}). \quad (4.15)$$

In weighted graph, the analog of Eq. (4.2) is the Hamiltonian given by

$$\mathcal{H} = \frac{1}{2} \sum_{a=1}^q (\bar{p} - p_{ij}) \left[\Theta(p_{ij} - \bar{p}) + \gamma \Theta(\bar{p} - p_{ij}) \right] \delta(\sigma_i, \sigma_j). \quad (4.16)$$

That is, when there is a high probability p_{ij} , relative to a background threshold \bar{p} , that nodes i and j are linked, we assign a positive edge weight to the link (ij) of size $(p_{ij} - \bar{p})$. Similarly, if the probability of a link (ij) is low, we assign a negative weight of size $\gamma(p_{ij} - \bar{p})$.

Armed with Eq. (4.16) (after the reassignment of the edge weights), the simulation is repeated with one replica and that the final clustering is obtained from this. That is, we then minimize in an identical fashion to the minimization of Eq. (4.2) that we discussed earlier. Specifically, we follow the 4 steps outlined in Section 4.4 for non multi-scale images and the 7 steps of Section 4.5.1 in the analysis of general multi-scale systems.

4.7 Summary of parameters

We now very briefly collect and list anew the parameters that define our Hamiltonians and appear in our methods.

- The resolution parameter γ in Eqs.(4.1, 4.2, 4.16). This parameter sets the graph scale over which we search for communities. This parameter is held fixed (typically with a value of $\gamma = \mathcal{O}(1)$) in the community detection method and varies within our multi-scale analysis. We determine the optimal value of γ by determining the local extrema of the average information theory overlaps between replicas.

- The spatial scale ℓ in Eq.(4.7). Similar to the more general graph scale set by γ , we may determine optimal ℓ by examining extrema in the average inter-replica information theory correlations. In practice, in all but the hardest cases (i.e., the case of the dalmatian dog in Fig.(4.1)), we ignored this scale and fixed ℓ to be infinite. Fixing $\ell = 1$ led to good results in the analysis of the dalmatian dog.

- The spatial cutoff scale Λ for defining link weights- see the brief discussions after Eqs.(4.6, 4.10). Whenever the spatial distance between two sites or blocks exceeded a threshold distance Λ we set the link weight to be zero. We did not tune this parameter in any of the calculations. It was fixed to the value of $\Lambda = 30$.

- The scale of the block size $L_{x,y}$ introduced in Section 4.3.2. This parameter is far smaller than the image size $N_x \times N_y$, yet large enough to cover the image features. We usually set $L_x \times L_y$ to be 9×9 for an image size $N_x \times N_y$ of around 200×200 .

- The background intensities \bar{V} in Eq.(4.2) and \bar{p} in Eq.(4.16). Similar to the graph scale set by γ and ℓ , we may determine the optimal \bar{V} and \bar{p} by observing the local extrema of the average information correlations between replicas.

As we will discuss on briefly, all optimal parameters may be found by determining the local extrema of the information theory correlations that signify no change in structure over variations of scale. Our determination of the optimal parameters is different from other methods, e.g., [174]. The parameter $\alpha_{i,s}$ in [174] controls the weight of the relative contribution of each gradient signal, which is similar to γ in our algorithm. The value of $\alpha_{i,s}$ is learned by gradient ascent on the F-measure [Eq. (4.18)] using the training images and corresponding to the ground-truth of the BSDS

[174]. In reality, we may fix some parameters and vary others—usually, Λ fixed as 30, $L_x \times L_y$ in the range of 7×7 to 11×11 , and γ , \bar{V} and \bar{p} been changed.

As an aside, in this brief paragraph, we briefly note for readers inclined towards spin glass physics and optimization theory that, in principle, in the large N limit (images with a large number of pixels) the effective optimal values for the kinds of the parameters listed above may be derived by solving the so-called “cavity” equations [30] that capture the maximal inference possible (in their application without the aid of replicas that we introduced here) [19, 122]. In the current context, in applying these equations anew to image segmentation, we arrive at the maximal inference possible of objects in an image. While these equations are tractable for simple cases, solving these equations is relatively forbidding for general cases. In practice, we thus efficiently directly examine our Potts model Hamiltonians of Eqs. (4.1, 4.2, 4.16) and, when needed, directly infer optimal values of the parameters by examining inter-replica correlations as described in the earlier sections. This will be expanded on in the next section (specifically, in Eq.(4.17)). [Detailed applications of this method are provided in Section 4.9.6.]

4.8 Computational Complexity, the Phase Diagram, and the determination of optimal parameters

Our community detection algorithm is very rapid. For a system with L links, the typical convergence time scales as $\mathcal{O}(L^{1.3})$ [18]. (The convergence time is the number

of steps that the algorithm uses to find a local minimum via our greedy algorithm.) In an image with N pixels, with all of the constants $\Lambda, L_{x,y} = \mathcal{O}(1)$ (i.e., not scaling with the system size), the number of links $L \sim N$.

Our general multi-scale community detection algorithm (that with varying γ) has a convergence time $\tau \sim L^{1.3} \ln N$ [27]. Thus, generally, for an image of size N , the convergence time $\tau \sim N^{1.3} \ln N$. Rapid convergence occurs in all but the “hard phase” of the community detection problem.

Specifically, we numerically investigated the phase diagram as a function of noise and temperature (i.e., when different configurations are weighted with a Boltzmann factor $\exp(-\beta \mathcal{H})$ with $\beta = 1/T$ at a temperature T for general graphs with an arbitrary number of clusters in [40].) Related analytic calculations were done for sparse graphs in [122]. In particular, in these and earlier works [18, 27] it was found that there is a phase transition between the detectable and undetectable clusters. The detectable phase further splinters into an “easy” and a “hard” phase.

These three phases in the community detection problem constitute analogs of three related phases in the “SAT-hard-unSAT” in the k-SAT problem [30]. The determined phase diagram [40] exhibits universal features. [We add or subtract “noise” (randomly insert or remove links between segments) or increase the temperature to a high enough value to force the system into its disordered phase]. Increasing the temperature can aid the detection of clusters [40]. The universal features of the phase diagram and the known cascade of transitions that appear on introducing temperature enable better confidence in the results of the community detection algorithm.

In some hard problems, accurate community detection is more readily achieved by using an “annealed” heat bath algorithm to achieve a thermal “order by disorder” type effect. By varying the temperature (as well as the “disorder” of the current state), a clearer understanding of the system is in its solvable phases is obtained. In such a case, the solution is physically meaningful. In its disordered (or unsolvable) phases, where the correlations are low between replicas, the proposed image segmentation solutions are not physically meaningful and are not trustworthy. One of the central results of Ref.[40] is that the “easy” solvable phase(s) of the community detection problem which leads to correct relevant solutions (i.e., not noisy partitions of a structureless system) universally appear in a “flat” [27, 40] phase(s) [see also the flat information theoretic curves in Fig. (4.4) and related discussion in Section(4.5.1)] as ascertained by the *inter-replica averages of all thermodynamic and information theoretic quantities* $\{\langle Q \rangle\}$. These may correspond to the internal energy ($Q = \mathcal{H}$), average Shannon entropy ($Q = H$), average inter-replica normalized mutual information and variation of information ($Q = I_N, V$), the complexity ($Q = \Sigma$) [30] or an associated “susceptibility” ($Q = \chi$) [27, 40] that monitors the onset of large complexity. [This susceptibility will be defined with the aid of the change in the average normalized mutual information I_N as a function of the number of trials t . It is defined as $\chi(n) = [I_N(t = n) - I_N(t = 4)]$.] That is, with z denoting a set of generalized parameters (e.g., artificially added additional noise in networks ($z = p_{out}$) [40], temperature ($z = T$) [40], or resolution parameter ($z = \gamma$) [27]), pertinent partitions

appear for those values of the parameters z for which

$$\frac{\partial \langle Q \rangle}{\partial z} = 0. \quad (4.17)$$

As alluded to above, a particular realization of Eq.(4.17) appears in the hierachal system discussed in Section 4.5.1 wherein $z = \gamma$ and $Q = I_N, V$. In that case, Eq.(4.17) was satisfied in well defined plateaus.

When present, crisp solutions are furthermore generally characterized by relatively high values of I_N , and these correspond to the “easy phase” of the image segmentation problem. In Sec. 4.9.6, we will provide explicit analysis of the phase diagram and optimal parameters as they pertain to several example images.

All of the results (except the ones in Sec. 4.9.6) presented below in the current chapter were attained at zero temperature and may be improved by the incorporation of thermal annealing as the results of [40] illustrate for general systems.

4.9 Results

4.9.1 Brain Image

Unweighted graphs

We start the review of the results of our methods by analyzing an *unweighted* graph (Eq. (4.1)) for the grey-scale brain image as shown in Fig. 4.5. We assign edges between pixels if the intensity difference is less than the threshold $\bar{V} = 16$ as denoted in panel (a) of Fig. 4.5. The algorithm uses Eq. (4.1) to solve for a range of resolution

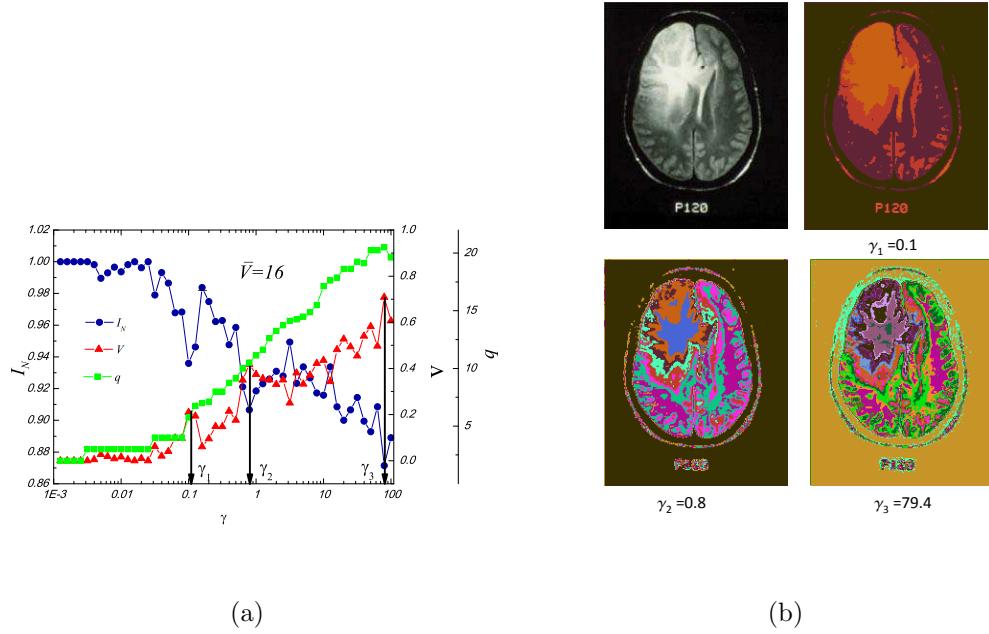


Figure 4.5: The plot of the normalized mutual information I_N , variation of information V and the number of communities q as a function of γ for the brain image (upper left of panel (b), reproduced with permission from Dr. Edward Aul and the Iowa Neuroradiology Library). The axis for γ is on a logarithmic scale. We objectively choose three local peaks in V (or equivalently local minima in I_N) and apply our community detection algorithm to the grey-scale brain image at the corresponding three values of γ 's. The corresponding results are shown in panel (b). Note that the results show a three-level hierarchy as γ varies.

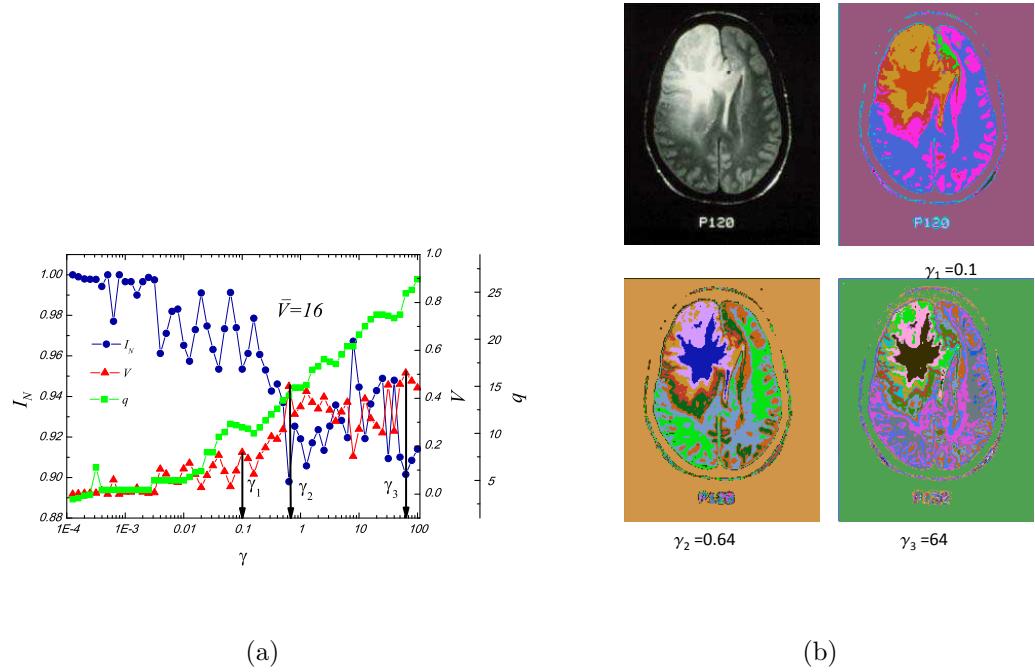


Figure 4.6: The brain image is reproduced with permission from Dr. Edward Aul and the Iowa Neuroradiology Library. In panel (a), the multiresolution result here behaves different from Fig. 4.5, but it maintains the same trend. The structure is only stable in the resolution range of $\gamma < 0.01$, compared to the wider range of $\gamma < 0.1$ in Fig. 4.5. We objectively choose three peaks of I_N and V and apply our community detection algorithm to the grey-scale brain image at the corresponding three values of γ s in panel (b).

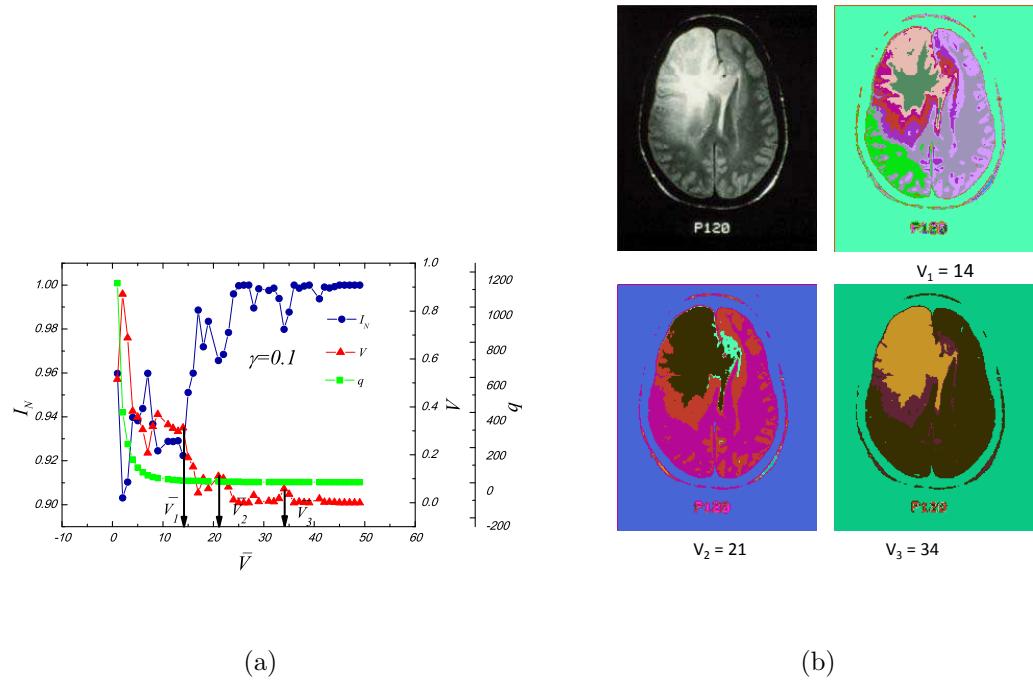


Figure 4.7: The brain image is reproduced with permission from Dr. Edward Aul and the Iowa Neuroradiology Library. The “multiresolution” result also shows the hierarchy structure as the threshold \bar{V} varies, as in Fig. 4.6. Higher “ \bar{V} ” corresponds to the lower “ γ ”, which means pixels intend to merge in higher “ \bar{V} ”. The structure is stable in the range of $\bar{V} > 25$, below which the structure is sensitive.

parameters γ in the interval $[\gamma_0, \gamma_f]$. In the particular case in Fig. 4.5, $\gamma_0 = 10^{-3}$ and $\gamma_f = 100$. Two additional input parameters are required: the number of independent replicas r that will be solved at each tested resolution and the number of optimization trials t per replica. We use $r = 10$ and $t = 4$ in Fig. 4.5.

As noted earlier (see Section 2.3), for each *replica*, we select the lowest energy solution among the t *trials*. The r replicas are generated by reordering the symmetric initialized state of one node per community. We then use the information based measures (i.e., I_N or V) to determine the multiresolution structure.

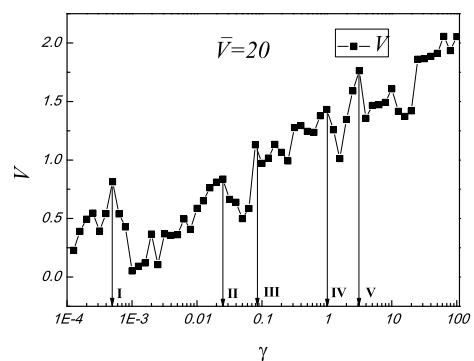
The plots of I_N , V and q as a function of γ in Fig. 4.5 exhibit non-trivial behaviors. Extrema in I_N and V correspond to jumps in the number of communities q . In the low γ region, i.e., $\gamma < 0.1$, the number of communities is stable. However, when $\gamma > 0.1$, the number of communities q sharply increase. This indicates that the structure changes rapidly as the resolution γ varies. There are three prominent peaks in the V (variation of information) curve. We show the corresponding images at these resolutions, that is in panel(b) in Fig. 4.5. These corresponding segmented images show more and more sophisticated structures. The lower right image at a resolution of $\gamma = 79.4$ shows the information in detail. Different colors in the image correspond to different clusters. There are, at least, five contours surrounding the tumor, that denote the degree by which the tissue was pushed by the tumor. The lower left image at the resolution $\gamma = 0.8$ is less detailed than the one on the right. Nevertheless, it retains the details surrounding the tumor. If we further decrease γ , the upper right image at the resolution $\gamma = 0.1$ will not keep the details of the tumor boundary, only

the rough location of the tumor. Thus, neither too large nor too small resolutions are appropriate for tumor detection in this image. The resolution around $\gamma = 0.8$ is the most suitable in this case. This is in accord with our general found maxim in Section 4.7 concerning a value of $\gamma = 1$. We re-iterate that, in general, the optimal value of γ is found by Eq. (4.17) (an example of which is manifest in information theory plateaus [18, 27] discussed in Sec. 4.5.1). In Sec. 4.9.6, we discuss how the optimal values of γ are determined in weighted example systems.

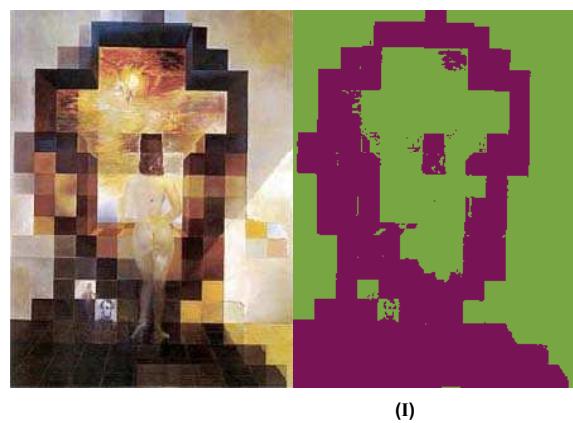
Weighted graphs

In Figs. (4.6, 4.7), we provide the “multiresolution” results for the *weighted* graph (Eq. (4.2)) of γ and \bar{V} for the brain image. Both the resolution γ and the threshold \bar{V} control the hierarchy structures. The jumps in q correlate with the changes in hierachal structures on different scales. We can combine both parameters to obtain the desirable results in the test images. See, e.g., the 3D plot of $I_N(\bar{V}, \gamma)$.

According to Eq. (4.2), edges (ij) with small (or large) difference $|D_{ij}|$ will decrease (or increase) the energy by $|\bar{V} - D_{ij}|$ (or $\gamma|\bar{V} - D_{ij}|$). However, if the unweighted graphs and the Potts model with discrete weights (Eq. (4.1)) are applied, the edges with small or large “color” difference will decrease or increase the energy by the amount of 1 or γ . Thus, considerable information (e.g., the “color” of each pixel) is omitted when using an unweighted graph approach.



(a)



(b)



(c)

Figure 4.8: The specific image is reproduced with the permission from [175]. At close distance, this is “Gala contemplating the Mediterranean sea” while at larger distance is “a portrait of Abraham Lincoln”. Panel(a) shows the variation of information as a function of resolution. We pick the resolution at each “peak” position and apply our algorithm at these particular resolutions. Panels (b) and (c) show the resulted images at the corresponding resolutions marked in panel (a). Note that at low resolution, the resulting segmentation clearly depicts “the portrait of Abraham Lincoln” as shown in panel (b) on the right. In particular, notwithstanding noise, as γ increases, the segmentation results show more details and we could detect the lady in the middle in (II)-(V) of Panel(c).

4.9.2 A painting by Salvador Dali

We next apply our multiresolution community detection algorithm to the images that are by construction truly multi-scale. The results at different resolutions are shown in Fig. 4.8. The original image is that of Salvador Dali’s famous painting “Gala contemplating the Mediterranean sea which at twenty meters becomes a portrait of Abraham Lincoln”. Our algorithm perfectly detected the portrait of Lincoln at low resolution as shown in Fig. 4.8 in the segmentation result appearing in panel (I) of (b). As the resolution parameter γ increases, the algorithm is able to detect more details. However, due to the non-uniform color and the similarity of the surrounding colors to those of the targets, the results lead to highly fractured partitions as seen in Fig. 4.8. At the threshold of $\bar{V} = 20$, the algorithm has difficulty in merging pixels to reproduce the lady in the image. For example, in image (II) in Panel (c), the lady’s legs are merged into the background. In image (III), only one leg is detected. In images (IV) and (V), both legs can be detected but belong to different clusters.

4.9.3 Benchmarks

In order to assess the success of our method and ascertain general features, we applied it to standard benchmarks. In particular, we examined two known benchmarks: (i) The Berkeley image segmentation benchmark and (ii) that of Microsoft Research.

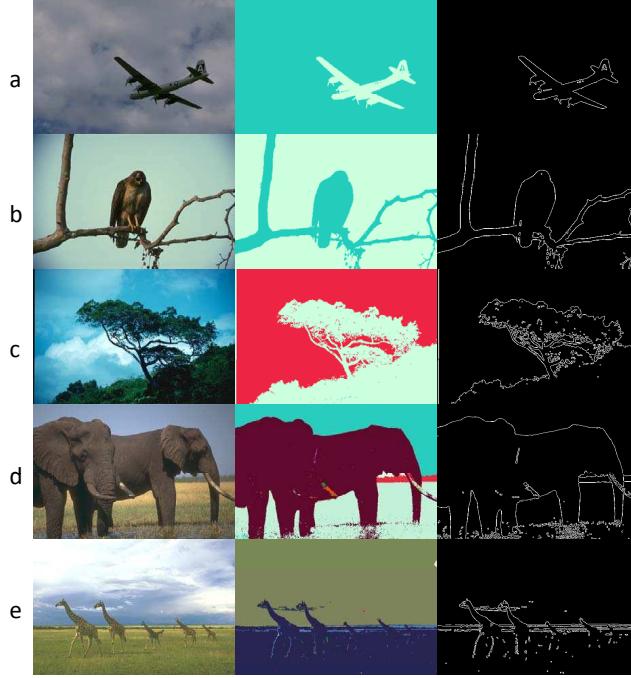


Figure 4.9: In the left column, the original images are depicted. The central column contains the results of our method, and the right column provides the boundaries of the images by running “EdgeDetect” in Mathematica on the results of our run in the central column. We performed the boundary detection on the results of our community detection algorithm (i.e., the central column) and employed the “F-measure” accuracy parameter in order to compare the results of our algorithm with earlier results reported for the Berkeley image segmentation benchmark (shown in Table. 4.1).

Berkeley Image Segmentation Benchmark

We were able to accurately detect the targets in test images, as in Figs.(4.9, 4.11). The original images in Fig. 4.9 were downloaded from the Berkeley image segmentation

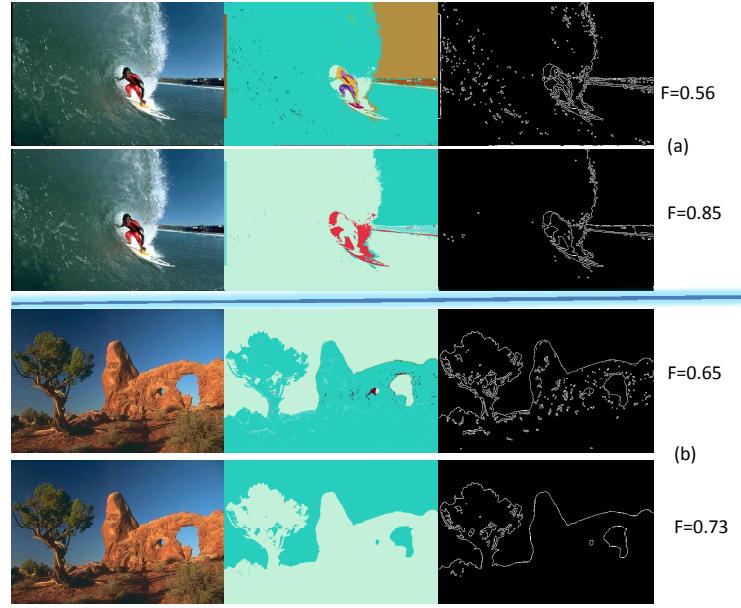


Figure 4.10: The image segmentation results by our algorithm. The original images in the left most column are downloaded from Berkeley Image segmentation benchmark. The central image in the first row/ the third row is the result of our algorithm at $\gamma = 0.01$ and $\bar{V} = 20$. The right image in the first/the third row is the boundary detection result of the corresponding central image by the software Mathematica. There are many dots/circles which denote the white spray in original image in the first row. The small dots/circles in the third row denote the shadow in the original image. We merge these small dots in the first and third row into the background and the results shown in the second and fourth row are more smooth and close to the groundtruth. This is confirmed by the larger F value shown in Table. 4.2.

benchmark BSDS300 [174], and those of Fig. 4.11 are downloaded from the Microsoft Research [176]. We will now compare our results with the results by other algorithms.

The Berkeley image segmentation benchmark provides the platform to compare the boundary detection algorithms by an “F-measure”. This quantity is defined as

$$\text{F-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (4.18)$$

“Recall” is computed as the fraction of correct pixels among all pixels that actually belong to the relevant subset, while “Precision” is the fraction of correct pixels among those that the algorithm believes to belong to the relevant subset. Thus, we have to draw the boundaries in our results and compute the F-measure. We use the tool “EdgeDetect” of Mathematica software to draw the boundaries within our region detection results, as shown in the right column in Fig. 4.9. The comparison of the “F-measure” of our algorithm (“F-Absolute Potts Model”) and the best algorithm in the benchmark (“F-Global Probability of boundary”)[177, 178] is shown in Table. 4.1. On the whole, our results are better than the algorithm of the Berkeley group.

Improved F-value by removing small high precision features

As seen in Sec. 4.9.3, our results in the first three images except the last one are better than the corresponding ones by the best algorithm in the Berkeley Image Segmentation Benchmark. One possible reason to cause the worse result in the last image is that our algorithm is too accurate. For example, the top image in Fig. 4.10, our result could detect the small white spray, which becomes the dots in the background. These small dots will form small circles in the boundary image shown in the right column, which are unexpected from the groundtruth, thus will reduce the value of precision and F . (In this case, $F = 0.56$.)

	F-Absolute Potts Model	F-Global Probability of boundary
a	0.79	0.78
b	0.94	0.91
c	0.82	0.74
d	0.79	0.83
e	0.75	0.60

Table 4.1: The comparison of “F measure” for Fig. 4.9 by our community detection algorithm (“F-Absolute Potts Model”) with the algorithm “Global Probability of boundary” (“gPb”) [177, 178] which has the highest score in the Berkeley image segmentation benchmark (“F-Global Probability of boundary”). The higher F-value corresponds to the better detection. Note that our algorithm is performing better than the “gPb algorithm” in almost all images except the fourth one. Our fourth (d) image gets lower score is mostly because there are dots in the lower grass place. These small dots will lead to small high accuracy features. These features are unexpected in the ground truth and thus lower the F-value.

	F-Our algorithm	F-Our algorithm without noise	F-gPb
a	0.56	0.85	0.82
b	0.65	0.73	0.74

Table 4.2: The F-measure of the images shown in Fig. 4.10. We provide the comparison with the results by algorithm Global Probability of Boundary (gPb). Note that after removing the small dots/noise in both images, the value of F increase significantly. After this merger, our results become equivalent to (or even better than) the best results to date.

Merging these high precision small dots with the background as, e.g., fleshed out in the second row in Fig. 4.10, leads to results that are equivalent to or better than those determined by the algorithm of global probability of boundary (gPb). A summary is presented in Table. 4.2.

Microsoft Research Benchmarks

In Fig. 4.11, we compare our results (in the rightmost column) with the ground truths provided by Microsoft Research (the central column). By adjusting the γ and \bar{V} values, we can merge the background pixels and highlight the target. In the segmentation of the image of the flower in the first row, $\gamma = 0.001$ and $\bar{V} = 20$. For both the picnic table in the middle row and that of the two sheep in the bottom row, we set $\gamma = 0.01$ and $\bar{V} = 15$.

Detection of quasi-periodic structure in quasicrystals

The ideas presented in our work may be applied to the analysis of images describing physical structures such as those of crystals as well as those of more complex systems [63, 64]. In this brief section, we outline the results of an iterative application of our method to tease out the multi-scale structure of "quasi-crystals" when these are viewed as images to be segmented.

Quasicrystals [179] are ordered but not periodic (hence the name "quasi"). In Fig. 4.12, the image in row (a) is such a quasi-crystal formed by "Penrose tiling". We applied the Fourier transform method to reveal the corresponding underlying struc-

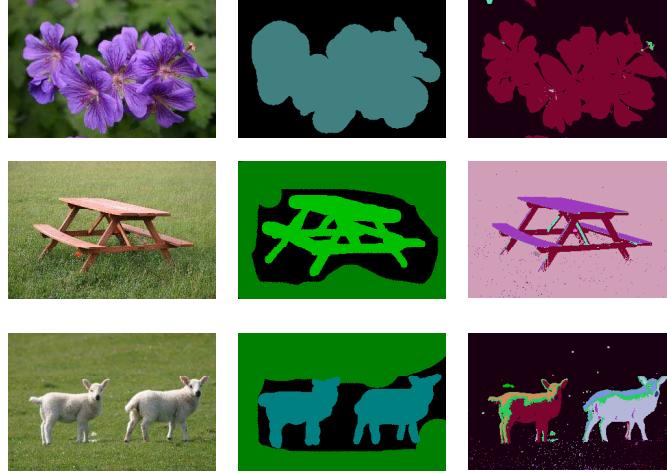


Figure 4.11: The results of some image segmentations by our Potts model in Eq. (4.2) and community detection algorithm [27]. The images are downloaded from the website of Microsoft Research database [176]. The left column contains the original images. The central column depicts the “ground truths” as defined by Microsoft Research [176], which are the desirable image segmentation results. These ground truth images were not generated by algorithm, but by humans. The right column shows the segmentation results of our algorithm. The parameters used for each image are: (1) $\gamma = 0.001$, $\bar{V} = 20$ for the flower image. (2) $\gamma = 0.01$, $\bar{V} = 15$ for the image of the picnic table. (3) Similarly, $\gamma = 0.01$, $\bar{V} = 15$ for the image of the two sheep. Our algorithm generally appears to find the “ground truth,” and it appears to work very well for these kinds of images where the color is nearly uniform within each object.

tures. In row (a), the image marked by (I) is the original image (downloaded from [180]), the one with notation (II) is the result of our algorithm, and (III) is the image

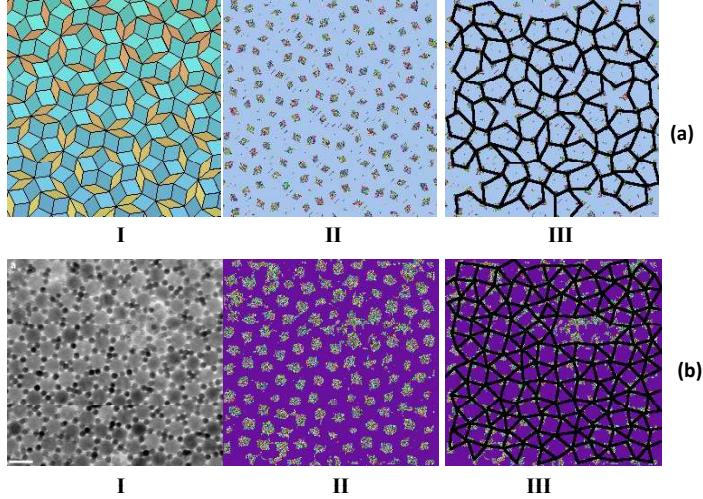


Figure 4.12: Quasicrystal images are displayed in panels (I). The corresponding image segmentation results by our algorithm are shown in (II). In (III), we connect the basic object by line, resulting in large basic blocks. This process can be repeated recursively leading to larger and larger scale structures. Note that we are able to reveal the underlying quasi-periodic structures in both row (a) (the original image in (I) is reproduced with the permission from the website of [180] by Professor John Baez and also the website of [181] by Greg Egan) and (b) (the original image in (I) is from [182], adapted by permission from Macmillan Publishers Ltd: [NATURE] ([182]), copyright (2009)). We show the first Penrose tiling (tiling ‘‘P1’’) in (a), and the structural motif of the (3².4.3.4) Archimedean tiling in (b).

of (II) with the connections of the nearest neighbor nodes. The images marked by (II) and (III) show the first Penrose tiling (tiling P1). Penrose’s first tiling employs a five-pointed pentagram, 3/5 pentagram shape and a thin rhombus. Similarly, the

result images of panels (II) and (III) in row (b) reveal the underlying structure of the superlattice with AB_4 stoichiometry and the structural motif of the (3².4.3.4) “Archimedean tiling” of the original image (I) (from [182]). The Archimedean tiling displayed in image (III) of row (b) of Fig. 4.12 employs squares and triangles. It is straightforward to analyze the quasi-periodic structure by applying our image segmentation algorithm as shown in Fig. 4.12. By iterating the scheme outlined herein, structure on larger and larger scales was revealed.

4.9.4 Images with spatially varying intensities

If the target is similar to the background (as in, e.g., animal camouflage), then the simplest initialization of edges with linear weights will, generally, not suffice. For example, in Fig. 4.13 the zebra appears with black and white stripes. It is hard to directly detect the stripes of the zebra because of the large “color” difference between the black and white stripes of the zebra. Fig. 4.15 has the similar stripe-shaped background which is very difficult to distinguish from the zebra itself by using the weights of Eqs.(4.7, 4.8) for the edges. Towards this end, we will next employ the Fourier transform method of Sec. 4.3.2.

As seen in Fig. 4.13, the original images are not uniform. Rather, these images are composed of different basic components such as stripes or spots, etc. With the aid of Fourier transform within each block, as discussed in Section 4.3.2, we are able to detect the target. For some of the images such as the second one in Fig. 4.13, when the target is composed of more than one uniform color or style, our community detection



Figure 4.13: The image segmentation results by the community detection algorithm with Fourier weights as described in Section 4.3.2. The top three images are downloaded from the Microsoft Research database ([176]). The fourth one is download from the Berkeley image segmentation benchmark ([174]). The fifth image is reproduced with the permission from the website of [183]. The left column contains the original images. The central column (apart from the last two rows) provides the “ground truths”. The right images on the right are our results. Even though the color is not uniform inside the targets, we can nevertheless easily detect the targets by this method.

algorithm is able to detect the boundaries, but the regions inside the boundary are hard to merge. This is because the block size is smaller than that needed to cover

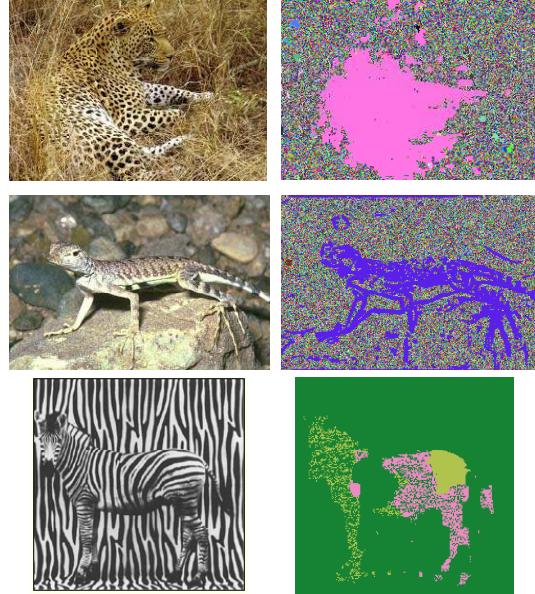
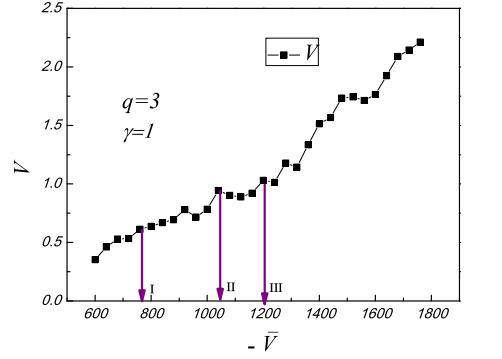
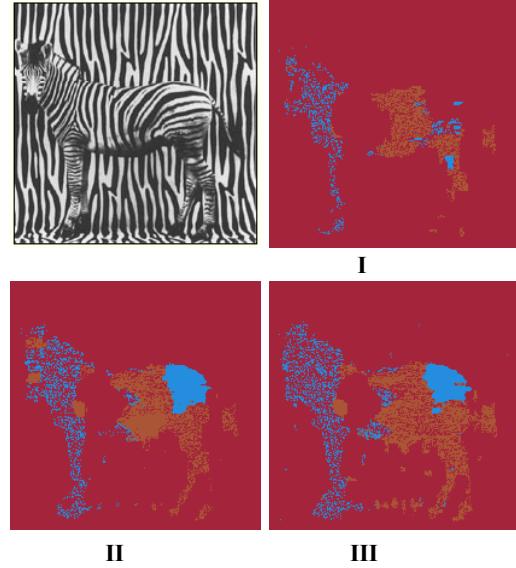


Figure 4.14: The results of the image segmentation for a “camouflaged image”. The image of the leopard is from ([184], A great male Leopard, made in Sabi Sand Private Game Reserve, South Africa, by Lukas Kaffer). The lizard is provided in the Berkeley image segmentation benchmark([174]). The last image is reproduced with the permission from the website of the EECS department of Berkeley ([161]). The parameters for the shown segmentations are: (1) $\gamma = 1$, $\bar{V} = -700$ for the image of the leopard, (2) $\gamma = 0.1$, $\bar{V} = -500$ in the image of the lizard, and (3) $\gamma = 1$, $\bar{V} = -1100$ for the zebra image.

both the target and the background. That is, block size of $L_x \times L_y = 5 \times 5$ is much smaller than the image size of $N_x \times N_y = 320 \times 213$ in the car image in the second row, so most of the blocks are within one color of the target (car) or the background (ground). However, the dominant Fourier wave-vector of the region within one color component of the car is similar to that of the ground. Therefore, the algorithm always



(a)



(b)

Figure 4.15: The “multiresolution” result of zebra (with permission from [161]) with resolution $\gamma = 1$. In panel(a), we plot the variation of information V as a function of negative threshold $-\bar{V}$. The peaks in V correspond to the changes of structures. We choose three peaks and run the algorithm at these three particular thresholds, and the result images are shown in panel (b). As $|\bar{V}|$ increases, less regions in the zebra merge to the background, and the boundary becomes more clear. If we increase the threshold further, the result is more noisy as the last image of $\bar{V} = -1200$ (III) shows.

treats them as the same cluster, rather than merging the region inside the car with the boundary.

In other instances (e.g., all the other rows except the second in Fig. 4.13), the targets are markedly different from the backgrounds. Following the scheme discussed in Section 4.8 (that will be fleshed out in Section 4.9.6), we may always optimize parameters such as the resolution, threshold, or the block size to obtain better segmentation.

4.9.5 Detection of camouflaged objects

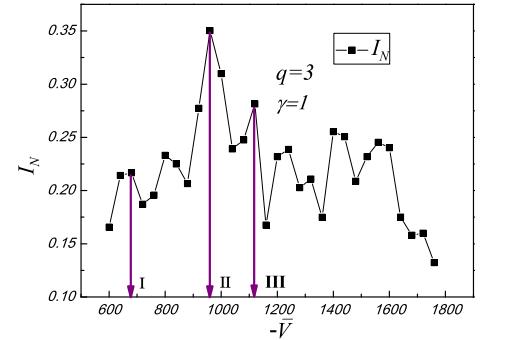
In the images of Fig. 4.13, the target objects are very different from their background. However, there are images wherein (camouflaged) objects are similar to their background. In what follows, we will report on the results of our community detection algorithm when these challenging images were analyzed. In all of the cases below in Section 4.9.5, the edge weights were initialized by the Fourier amplitudes discussed earlier (Section 4.3.2). In the case of the dalmatian dog image in Section 4.9.5, we further employed the method of average intensity difference between blocks discussed in Section 4.3.2. In all cases but this last one of the dalmatian dog, we fixed the length scale parameter ℓ of Section 4.3.2 to be infinite.

Images of a leopard, a lizard, and a zebra

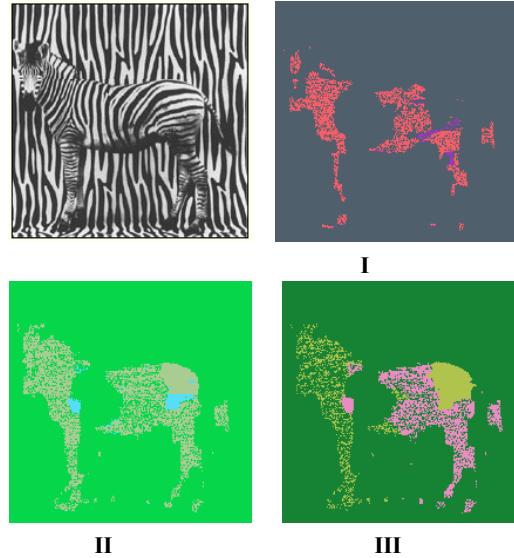
“Camouflage” refers to a method of hiding. It allows for an otherwise visible organism or object to remain unnoticed by blending with its environment. The leopard in the

first row of Fig. 4.14 is color camouflaged. With our algorithm, we are able to detect most parts of the leopard except the head. The lizard in the second row uses not only the color camouflage but also the style camouflage, both the lizard and the ground are composed of grey spots. We can detect the lizard. The zebra in the bottom row uses the camouflage— both the background and the zebra have black-and-white stripes. Our result is very accurate, even though the algorithm treats the middle portion of the zebra (the position of the “hole”) as the background by mistake. This is because, in this region, the stripes within the zebra are very hard to distinguish from the stripes in the background, they are both regular and vertical.

We applied the “multiresolution” algorithm to the zebra image in the last row of Fig. 4.14 as shown in Fig. 4.15. The number of communities is $q = 3$, the resolution parameter $\gamma = 1$ and the threshold \bar{V} was varied from $\bar{V} = -600$ to $\bar{V} = -1800$. In the low $|\bar{V}|$ area, some regions inside the zebra tend to merge into the background (the image with the threshold $\bar{V} = -760$). As the background threshold $|\bar{V}|$ increases in magnitude, the boundary of the zebra becomes sharper (the shown segmentation corresponds to a threshold of $\bar{V} = -1040$). For yet larger values of $|\bar{V}|$, the results are noisy (the image with the threshold $\bar{V} = -1200$). Thus, in the range $760 \leq |\bar{V}| \leq 1200$, we obtain the clear detection seen in the last row of Fig. 4.14.



(a)



(b)

Figure 4.16: The “multiresolution” result of zebra with fixed community number $q = 3$ and resolution $\gamma = 1$. In panel(a), we plot the normalized mutual information I_N as a function of negative threshold \bar{V} . The peaks in I_N also correspond to the changes of structures. We choose three peaks and run the algorithm at these three particular thresholds, and the result images are shown in panel (b). As $|\bar{V}|$ increases, less regions in the zebra merge to the background, and the boundary becomes more clear.

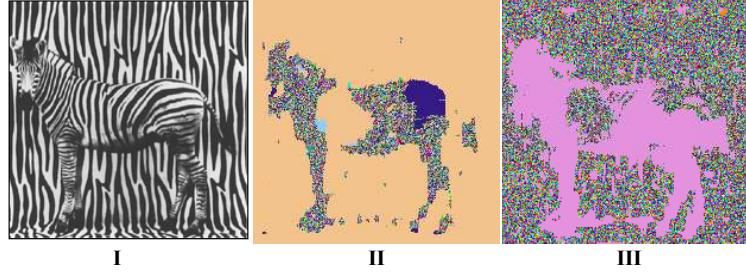
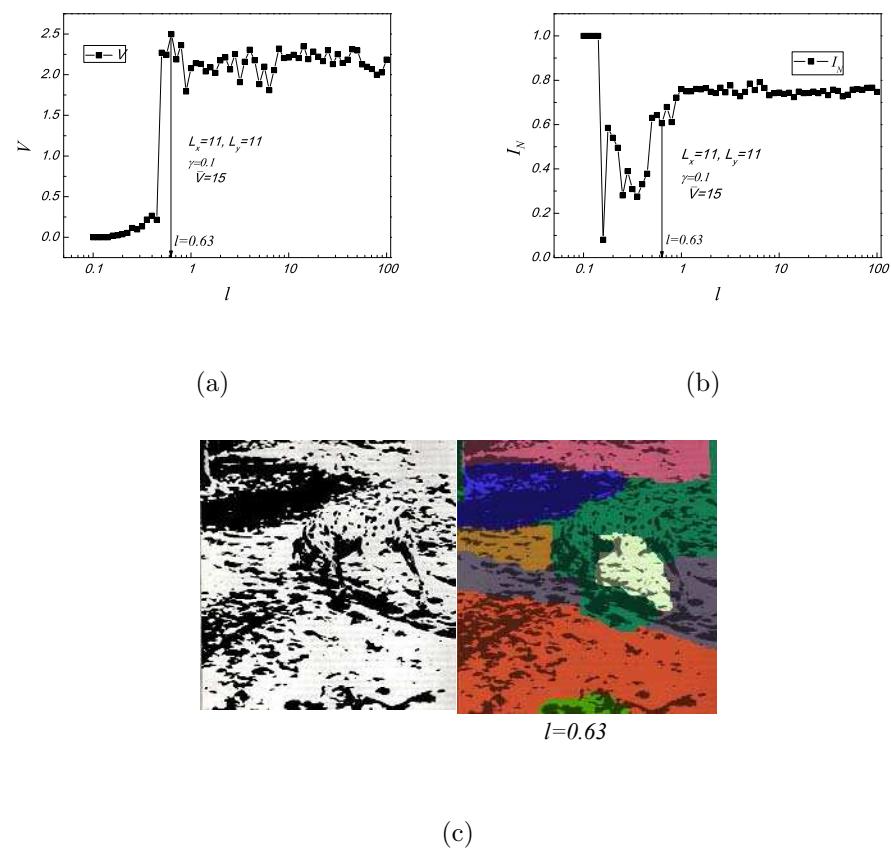


Figure 4.17: The image segmentation results (II and III) of the original camouflaged zebra in (I). In panel (II), we used the Fourier based edge weights of Eq. (4.11) and with a negative background $\bar{V} = -1200$ (Other parameters are $\gamma = 1$, block size $l_x \times l_y = 11 \times 11$). (III) The resulting segmentation when the sign on the right hand side of Eq. (4.11) is flipped. Here, we applied a positive background $\bar{V} = 900$ (Other parameters are $\gamma = 1$, block size $l_x \times l_y = 7 \times 7$). Both of the results shown here (i.e., II and III) are able to detect the zebra.

The image segmentation corresponding to the mutual information (I_N) peak

As emphasized throughout this work, we focus on inter-replica information theory overlap extrema. In some of the earlier examples, we discussed the results pertaining to variation of information maxima (often correlating with normalized mutual information minima). We now briefly discuss sample results for the normalized mutual information maxima. We provide one such example in Fig. 4.16. Herein, we plot I_N as a function of γ and provide the corresponding segmented images at the peaks of I_N . As shown before, in panels I-III of Fig. 4.15, we provide the image segmentation that correspond to the values of γ for which the variation of information V exhibits a



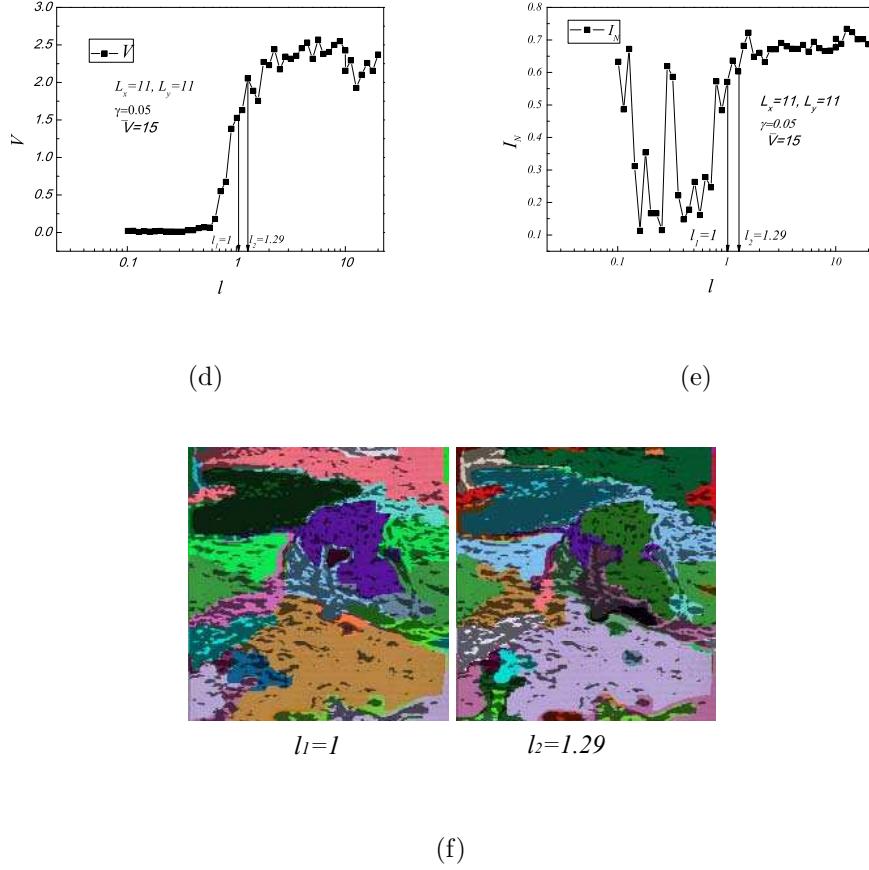


Figure 4.18: Results of our algorithm as a function of the length scale ℓ in Eq. (4.7) for the dalmatian dog image (reproduced with permission from [162]). Plots of V , I_N as a function of the length scale ℓ appear in panels (a, b)(at $\gamma = 0.1$) and (d, e) ($\gamma = 0.05$). Panel (c) shows the original image. As seen in panels (a,b), a coincident local maximum of V and local minimum of I_N appears (for $\gamma = 0.1$) at $\ell = 0.63$. Similarly, panel (f) shows the images corresponding to the peak of V (coincident with a local minimum of I_N) in panel (d) (and (e)) at $\ell_2 = 1.29$ (and $\gamma = 0.05$). We examine the results for $\ell_1 = 1$ in panel (f). We are able to detect the body and the back two legs of the dog, even though with some “bleeding” in panel (f). In (c), we are detecting well except for the inclusion of some “shade” noise under the body.

local maximum. In Fig. 4.16, we do the same for the normalized mutual information I_N .

The image segmentation with negative and positive Fourier weight

In this brief section, we wish to compare results obtained with the weights given by those of Eq. (4.11) to those obtained when V_{ij} is set to be of the same magnitude as in Eq.(4.11) but of opposite sign (referred to below as its “negative counterpart”).

In the latter case, a large weight V_{ij} corresponds to a large overlap between patterns in blocks. Thus, minimizing the Hamiltonian will tend to fragment a nearly uniform background (for which the overlap between different blocks within is large) and will tend to group together regions that change. The results of the application of Eq. (4.11) and that of its negative counterpart are shown side by side in Fig. 4.17 II and III. In both cases, the zebra is successfully detected from the similar stripe-shaped background, as long as using the right parameters. In (II), the parameters used are as follows: the background $\bar{V} = -1200$, the resolution parameter is $\gamma = 1$ and the block size is $l_x \times l_y = 11 \times 11$. In (III), we use a positive background $\bar{V} = 900$ but with a negative Eq. (4.11), resolution $\gamma = 1$ and block size $l_x \times l_y = 7 \times 7$. The difference shown in Fig. 4.17 between result (II) and (III) due to different fourier weights is that: In (II), the background forms a large cluster and the zebra is composed of lots of small clusters. In (III), the zebra forms a large single community while the background is composed of many small communities. For the images in Fig. 4.13, we substitute in Eq. (4.2), the weights of Eq. (4.11) along with a negative background \bar{V} .

Dalmatian dog

The camouflaged dalmatian dog in panel (c) of Fig. 4.18 (and Fig. 4.1) is a particularly challenging image. We invoke the method detailed in Sec. 4.3.2 to assign edge weights. We then apply the multiresolution algorithm to ascertain the length scale ℓ in Eq. (4.7). The inter-replica averages of the variation of information V and the normalized mutual information I_N are, respectively, shown in panels (a,b) and panels (d,e) of Fig. 4.18. These information theory overlaps indicate that, as a function of ℓ , there are, broadly, two different regimes separated by a transition at $\ell \sim 1$. We determine the value of ℓ at the local information theory extremum that is proximate to this transition and determine the edge weights set by this value of ℓ . (See Eq. (4.7).) In Section 4.9.6, we will illustrate how we may determine an optimal value of ℓ .

We segment the original image of the dalmatian dog via our community detection algorithm as shown in panels (c) and (f) in Fig. 4.18. The result in panel (c) corresponds to a resolution of $\gamma = 0.1$. The image on the right in panel (c) is the superimposed image of our result and the original image (on the left) at the particular length $\ell = 0.63$. The “green” color corresponds to the dalmatian dog. The method is able to detect almost all the parts of the dog except the inclusion of “shade” noise under the body. The results in panel (f) correspond to a resolution $\gamma = 0.05$. The image on the left in panel (f) is the superimposed image of our immediate running result and the original one at the length $\ell_1 = 1$, which is close to the maximum of V (and the local minimum of I_N). On the right, we provide the result for $\ell_2 = 1.29$ (a

value of ℓ corresponding to a maximum of V and a minimum of I_N). The “purple” color in the segmented image corresponds to the dalmatian dog. We are able to detect the body and the two legs in the back, even though with some “bleeding”. As we will discuss in the next subsection, it is possible to relate the contending solutions found in Fig.(4.18) for different values of γ and ℓ to the character of the phase diagram.

4.9.6 Phase Diagram

As previously alluded to in Sec. 4.8, we investigated numerically the phase diagram and the character of the transitions of the community detection problem for general graphs in [40]. From this, we were able to distinguish between the “easy”, “hard” and “unsolvable” phases as well as additional transitions within contending solutions within these phases (e.g., our discussion in Section 4.5.1). Strictly speaking, of course, different phases appear only in the thermodynamic limit of a large number of nodes (i.e., $N \rightarrow \infty$). Nevertheless, for large enough systems ($N \gg 1$), different phases are, essentially, manifest. As we will now illustrate, the analysis of the phase diagram enables the determination of the optimal parameters for the image segmentation problem. To make this connection lucid, we will, in this section, detail the phase diagrams of several of the images that we analyzed thus far.

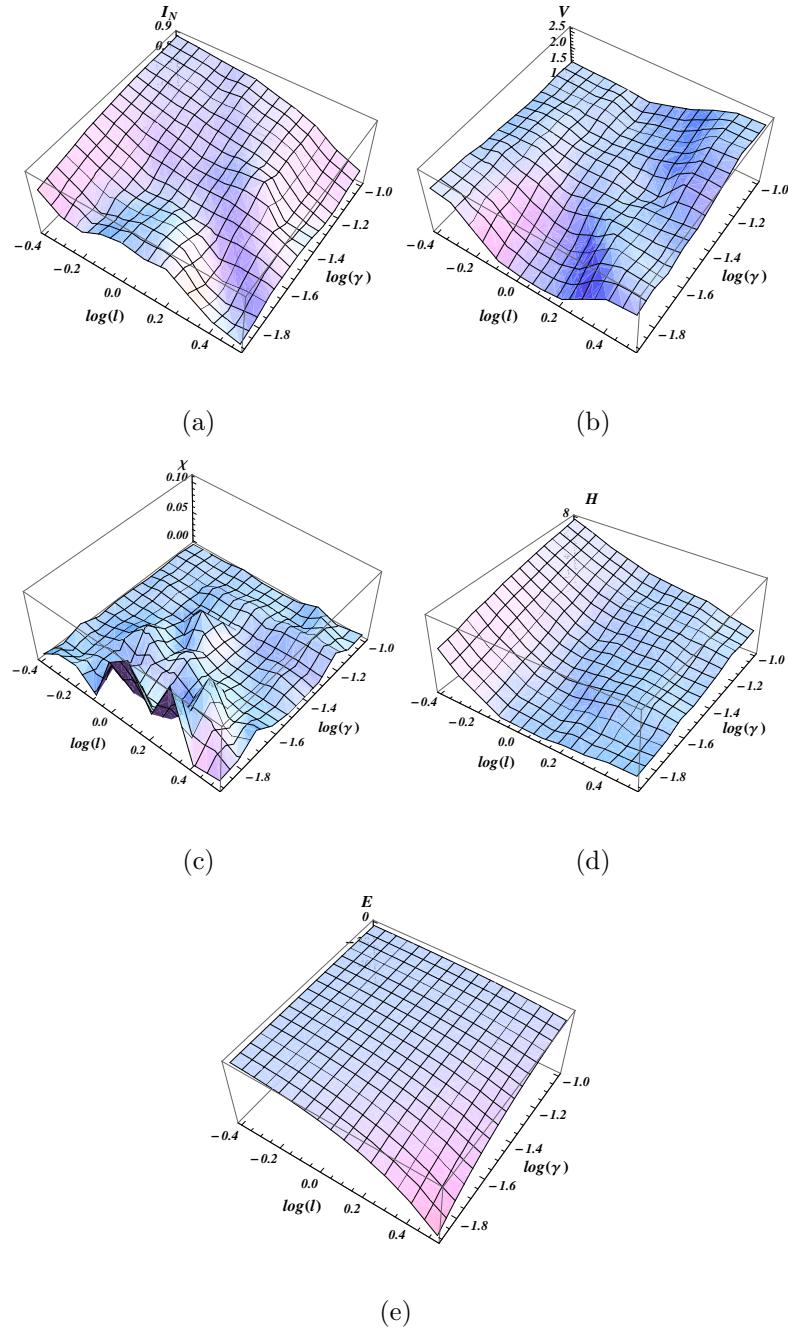


Figure 4.19: Plots of I_N , V , χ , H and energy E as the function of $\log(\ell)$ and $\log(\gamma)$ for the “dalmatian dog” image in Fig. 4.18.

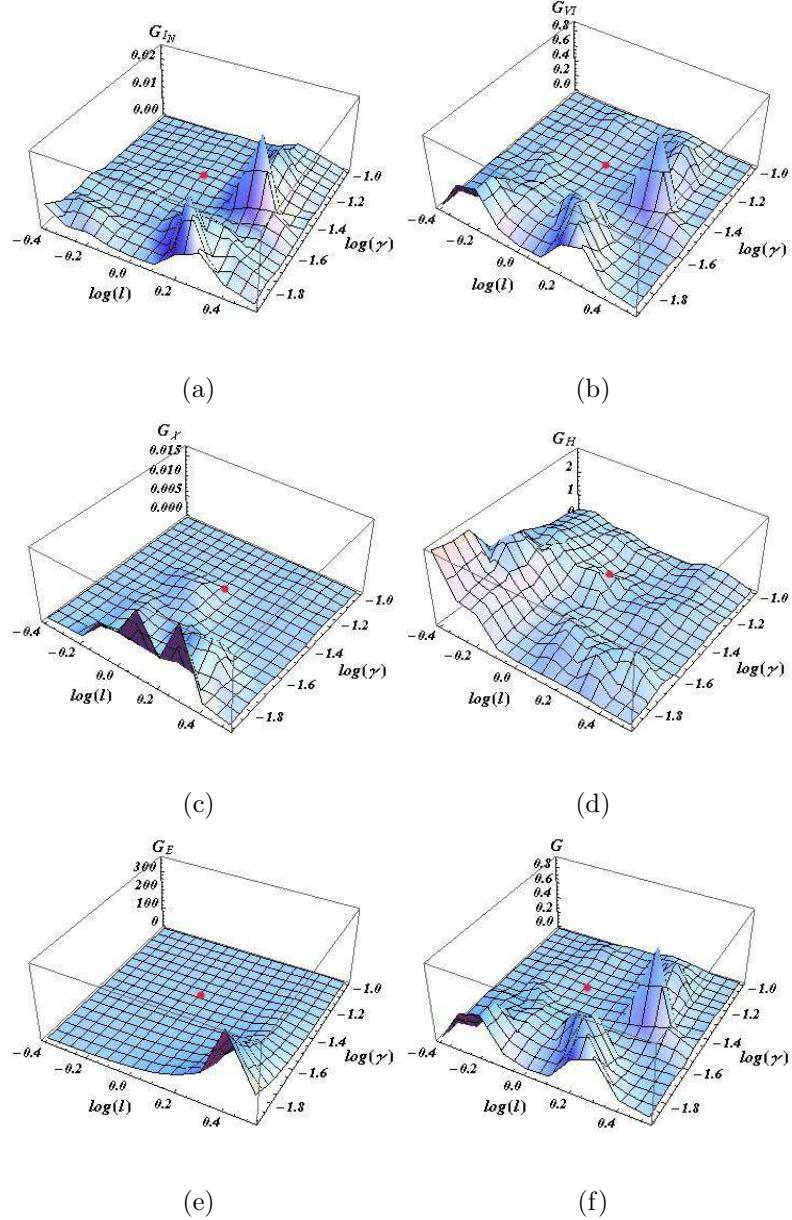


Figure 4.20: Information theory and thermodynamic measures relating to the dalmatian dog image of Fig. 4.18. The squares of the gradient of I_N , V , χ , H , E (panel (a)-(e)) and the sum of the squares of the gradients of I_N , V and χ (panel (f)) as the function of $\log(\ell)$ and $\log(\gamma)$. The red dot in each panel denotes the location of the parameters $((\log(\ell), \log(\gamma)) = (0, -1.3)$ (i.e., $(\ell, \gamma = (1, 0.05))$) of the results in Fig. 4.18. This good segmentation found for these parameters correlates with a local minimum within each panel.

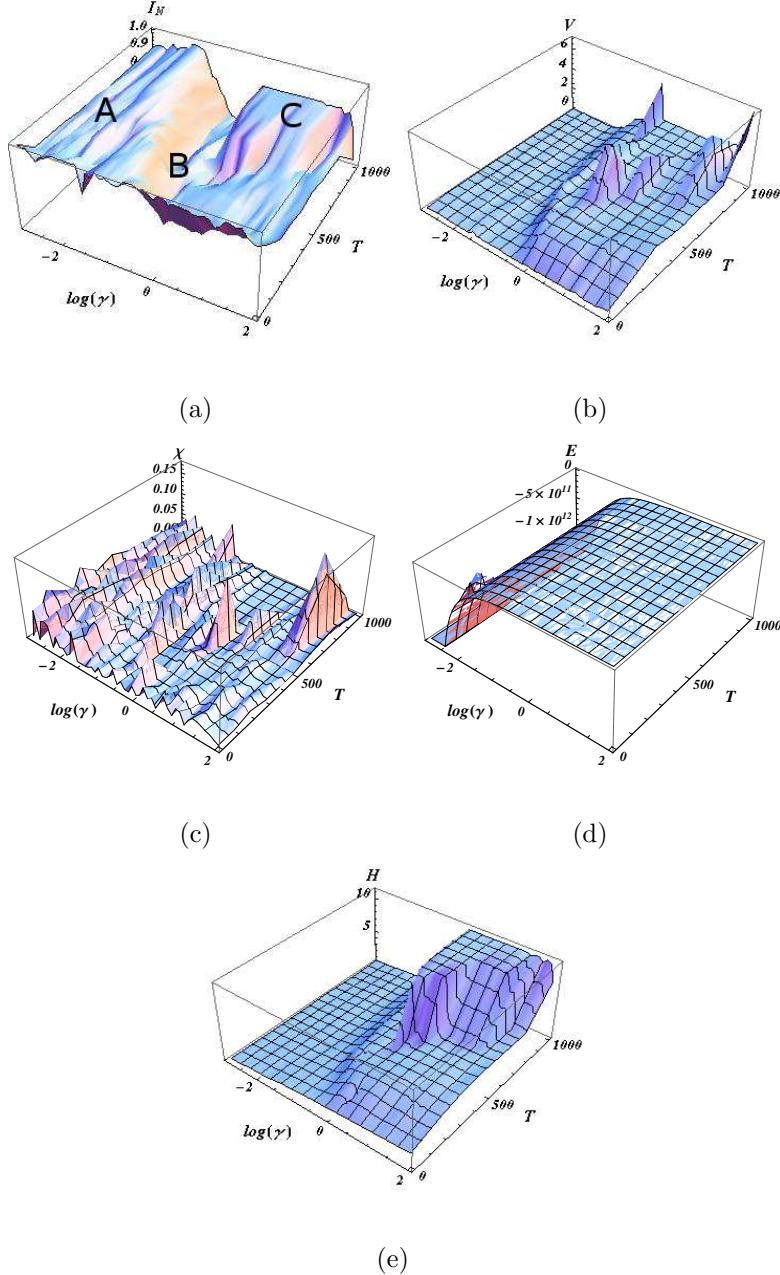


Figure 4.21: The normalized mutual information I_N , variation of information V , susceptibility χ , energy E and Shannon entropy H as the function of the resolution $\log(\gamma)$ and temperature T for the “bird” image in Fig. 4.22. In panel (a), we mark (i) the “easy” phase (where I_N is almost 1) as “A”, (ii) the “hard” phase (where I_N decreases) by “B”, and (iii) denote the “unsolvable” phase (where I_N forms a plateau whose value is less than 1) by “C”. The physical character of the “easy”, “hard”²²¹ and “unsolvable” phases is further evinced by the corresponding image segmentation results in Fig. 4.22. We can determine the signatures

Phase diagram of the Potts model corresponding to the dalmatian dog image

We will now analyze the thermodynamic and information theory measures as they pertain to the dalmatian dog image (Fig. 4.18) for a range of parameters. In a disparate analysis, in subsection 4.9.6, we will extend this approach also to finite temperature (i.e., $T > 0$) where a heat bath algorithm was employed. Here, we will content ourselves with the study of the zero temperature case that we have focused on thus far.

Plots of the normalized mutual information I_N , variation of information V , susceptibility χ , entropy H , and the energy E are displayed in Fig. 4.19. We set the background intensity to $\bar{V} = 15$. The block size is $L_x \times L_y = 11 \times 11$. We then varied the resolution γ and the spatial scale ℓ within a domain given by $\gamma \in [0.01, 0.1]$ and $\ell \in [0.4, 4]$. In Fig. 4.19, all logarithms are in the common basis (i.e., \log_{10}).

Several local extrema are manifest in Fig. 4.19. In the context of the data to be presented below, the quantity Q of Eq. (4.17) can be I_N , V , χ , H or E , and z may be γ or ℓ . Examining the squares of the gradients of these quantities, as depicted in Fig. 4.20, aids the identification of more sharply defined extrema and broad regions of the parameter space that correspond to different phases.

In Fig. 4.20, we compute the squares of the gradients of I_N , V , χ , H and E in panels (a) through (e). Panel (f) shows the sum of the squares of the gradients of I_N , V and χ . A red dot denotes parameters for a “good” image segmentation with the

parameter pair being $(\gamma, \ell) = (0.05, 1)$ (or $(\log(\gamma), \log(\ell)) = (-1.3, 0)$) corresponding to the left hand segmentation in panel (f) of Fig. 4.18). Clearly, the red dot is located at the local minimum in each panel. This establishes the correspondence between the optimal parameters and the general structure of the information theoretic and thermodynamic quantities.

As evinced in Fig. 4.20, there is a local single minimum which is surrounded by several peaks in the 3D plots of the squares of the gradients of I_N , V (panel(a),(b)) and their sum (panel (f)). For the dalmatian dog image (Fig. 4.18, setting Q in Eq. (4.17) to be the square of the gradients efficiently locates optimal parameters. Note that the other contending solutions in Fig. (4.18) relate naturally to the one at $\gamma = 0.05$ and $\ell = 1$. The $\ell = 1.29$ (i.e., $\log(\ell) = 0.11$) solution on the right hand side of panel (f) appears in the same “basin” as that of the $\ell = 1$ solution. Indeed, both segmentations of panel (f) of Fig. (4.18) share similar features. By contrast, the $\gamma = 0.1$ and $\ell = 0.63$ (i.e., $(\log(\gamma) = -1, \log(\ell) = -0.2)$) segmentation result of panel (c) in Fig.(4.18) relates to a different region. The chosen parameters lie in a solvable phase. Other “nearby” parameters within the same basin lead to similar partitions. Many complex problems have more than a single solvable phase (e.g., hierachal or more general multi-scale problems [27, 63, 64]), so there may be multiple meaningful solutions.

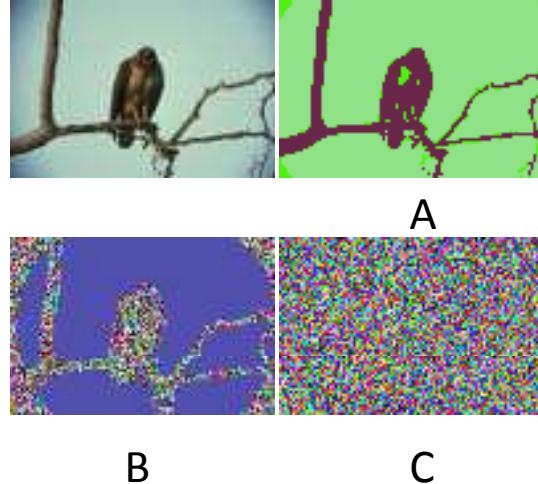


Figure 4.22: The image segmentation results of the “bird” image (reproduced from the Berkeley image segmentation benchmark). The original image is on the upper left. The segmentations denoted by “A”, “B” and “C” correspond to results with different parameter pairs ($\log(\gamma)$, T) that are marked in panel (a) of Fig. 4.21. Both results “A” and “B” are able to distinguish the “bird” from the “background”. However, in panel (b), the “bird” is composed of numerous of small clusters. The segmentation “C” does not detect the “bird”. The results shown here at points A, B, C correlate with the corresponding “easy-hard-unsolvable” phases in the phase diagram in Fig. 4.21.

A finite temperature phase diagram

Fig. 4.21 depicts the finite temperature ($T > 0$) phase diagram of the image of the bird of Fig. 4.22. We will find that for this easy image, the phase boundaries between the easy, hard, and unsolvable phases of the image are relatively sharply defined.

In the context of the data to be presented, we fixed the background intensity

$\bar{V} = 15$, set the block size to be $L_x \times L_y = 1 \times 1$ and took the spatial scale $\ell \rightarrow \infty$. The varying parameters are the resolution γ and temperature T . Instead of applying our community detection algorithm at zero temperature, we will incorporate the finite temperature [40] in this section. The ranges of the γ and T values are $[0.001, 100]$ and $[0, 1000]$ respectively. In the panels of Fig. 4.21, we show the normalized mutual information I_N , variation of information V , susceptibility χ , energy E and Shannon entropy H as the function of the temperature T and the logarithm of the resolution $\log(\gamma)$.

We can clearly distinguish the “easy”, “hard” and “unsolvable” phases from the 3D plots of I_N (panel (a)), V (panel (b)) and H (panel (e)). The label “A” in panel (a) marks the “easy” phase, where $\gamma \in [0.001, 0.3]$ for $T \in [0, 500]$ and $\gamma \in [0.001, 0.01]$ for $T \in [500, 1000]$. The “easy” phase becomes narrower as temperature increases. The corresponding image segmentation result shown in Fig. 4.22 validates the label of the “easy” phase. The “A” image in Fig. 4.22 is obtained by running our community detection algorithm with the parameter pairs located in the area labeled by “A” in Fig. 4.21. The image segmentation denoted by “A” can perfectly detect the bird and the background. The bird is essentially composed of two clusters and the background forms one contiguous cluster. This reflects the true composition of the original image on the upper left. Thus, the bird image can be perfectly segmented in an unsupervised way when choosing parameters to be in the “A” region (corresponding to the computationally “easy” phase).

The region surrounding point “B” in panel (a) in Fig. 4.21 denotes the “hard”

phase, where γ is in the range of $[0.3, 100]$ and T in the range of $[0, 500]$. Within the “hard” phase, as the corresponding image labeled by “B” in Fig. 4.22 illustrates, the bird is composed of numerous small clusters with the background still forming one cluster. In this phase, the image segmentation becomes harder and some more complicated objects cannot be detected.

The label “C” in panel (a) in Fig. 4.21 denotes the “unsolvable” phase, where the range for γ and T is about $[0.1, 100]$ and $[500, 1000]$ respectively. The corresponding image in Fig. 4.22 labeled by “C” is composed of numerous small clusters for which it is virtually impossible to distinguish the bird from the background. In this phase, the normalized mutual information I_N is far less than 1 (indicating, as expected, the low quality of segmentations).

Other 3D plots in Fig. 4.21 generally show similar phase transitions. Especially, the 3D entropy plot (panel (e)) vividly depicts accurate three phases and their clear boundaries.

4.10 Conclusions

In summary, we applied a multi-scale replica inference based community detection algorithm to address unsupervised image segmentation. The resolution parameters can be adjusted to reveal the targets in different levels of details determined by extrema and transitions. The underlying intuition behind this condition is that a regime of parameters often exists which all lead to observations consistent with a potential so-

lution when parameters are changed locally. This condition indicates a locally stable phase, and the behavior is most evident in hierachal network structures (see Sec. 4.5.1 in the current work and also[27]) where plateaus in the information measures exist (i.e., extended regions where Eq. (4.17) is satisfied). In the images with uniform targets, we distributed edge weights based on the color difference. For images with non-uniform targets, we applied a Fourier transformation within blocks and assigned the edge weights based on an overlap. Our image segmentation results were shown to be, at least, as accurate as some of the best to date (see, e.g., Table. 4.1) for images with both uniform and non-uniform targets. The images analyzed in this work cover a wide range of categories: animals, trees, flowers, cars, brain MRI images, etc. Our algorithm is specially suited for the detection of camouflage images. We illustrated the existence of the analogs of three computational phases (“easy-hard-unsolvable”) found in the satisfiability (k -SAT) problem [30] in the image segmentation problem as it was formulated in our work. When the system exhibits a hierachal or general multi-scale structure, transitions further appear between different contending solutions. With the aid of the structure of the general phase diagram, optimal parameters for the image segmentation analysis may be discerned. This general approach of relating the thermodynamic phase diagram to parameters to be used in an image segmentation analysis is not limited to the particular Potts model formulation for unsupervised image segmentation that was introduced in this work. In an upcoming work, we will illustrate how supervised image segmentation with edge weights that are inferred from a Bayesian analysis with prior probabilities for various known pat-

terns (or training sets), can be addressed along similar lines [185]. We conclude with a speculation. It may well be that, in real biological neural networks, parameters are adjusted such that the system is solvable for a generic expected input and critically poised next to the boundaries between different contending solutions [186].

Chapter 5

Segmenting Intracellular Distribution Images

5.1 Introduction

Image segmentation plays a crucial role in many medical imaging applications by detecting anatomical structures and other regions of interest. Examples of the medical image segmentation methods are listed in [65], which include the graph partitioning methods and normalized cuts [58, 153], and mixture Gaussian distributions (MGD) method [67], to name a few.

In this chapter, we propose a multiresolution “community detection” (CD) approach based on graph partitioning theory [58] to segment complex intracellular signals derived using fluorescent dyes. CD [1, 10, 12, 27, 18, 40] seeks to divide groups of nodes with dense connections internally and with sparser connections between the

groups. Moreover, it partitions a large physically interacting system into optimally decoupled communities. To demonstrate our approach, we have used fluorescence lifetime imaging microscopy (FLIM) data captured using near-infrared (NIR) fluorescent dyes [189], where the underlying signal describing intra-cellular distribution is complex in nature. FILM, a promising technique for imaging molecular process, uses time-resolved measurements of fluorescence from cells and thin tissue sections to generate images of the characteristic fluorescence lifetimes (FLTs) within a pixel or voxel. The FLT is the average time a molecule resides in the excited state before returning to the ground state through fluorescence emission [190].

Segmentation of the FLIM data is challenging due to the high amount of spatial and temporal noise attached with such data. Such problem is severe for images captured using NIR dyes because most of the current organic NIR dyes are not highly photostable. Hence, their resulting images are mostly low-signal-to-noise ratio (SNR) and low-resolution in nature, and there exists a definite niche to efficiently segment such data to mine the spatial structures hidden in them. Note that although our presentation herein focuses on the NIR FLIM data, however, it is in general applicable to any fluorescence microscopy imaging data.

To perform the segmentation, our multiresolution CD first investigates the optimal structure at different resolutions in the input image data. It then analyzes for each of these resolutions the respective number of estimated communities and information measurement between the independent solutions (“replica”) attained using it from different starting points, to determine the significant structures at which the replicas

are strongly correlated. The outcome of the proposed method for the case of FLIM images is a segmented image containing distinct average FLTs in each of its segments. It is noteworthy to mention that our proposed method as discussed in this chapter is “unsupervised” in nature. Namely, it does not employ any ground-truth as a prior knowledge to train the algorithm-i.e., the algorithm does not assume any prior knowledge of specific spatial patterns corresponding to any specific community hidden in the input data.

We have compared the performance of our proposed CD method with another image segmentation method based on the mixture of Gaussian distributions (MGD) [67] for segmenting the NIR FLIM images. Our method proves to be more efficient in finding diverse spatial regions with distinct average FLTs in the NIR FLIM data. While a detailed biological validation of the resulting segments to understand their biological roles is beyond the scope of this work, since such validation varies from one biological study to another, we nevertheless argue that our proposed method is general, and it can be used to generate hypothesis for future biological validation.

The chapter is presented as follows. In Section 5.2, we discuss briefly on our NIR FLIM system and the structure of the data acquired from such system. Section 5.3 presents our proposed multiresolution CD approach for performing segmentation of images based on their spatial information. Section 5.4 extends our approach for images described by both spatial and temporal information. In both of these sections, we also develop image segmentation methods based on MGD. In Section 5.5, we present the performance of our proposed method using NIR FLIM image of an *ex*

vivo liver tissue sample of mice treated with an NIR fluorescent dye. Section 5.5 also presents a comparison of this performance with that attained using the MGD method. We conclude in Section 5.6.

5.2 Near-Infared Fluorescence Lifetime Imaging Microscopy

Before we present our proposed segmentation method in detail, we present a brief overview of our NIR FLIM system and the structure of the data acquired from this system [189]. Our NIR FLIM system consists of a fiber-coupled laser diode (BDL-785-SMC, Becker-Hickl, Germany), a confocal laser scanning fluorescence microscope (FV1000, Olympus, Center Valley, PA), a thermoelectrically cooled, red-enhanced PMT (PMC-100-20, Becker-Hickl, Germany), and a TCSPC card (SPC-730, Becker-Hickl, Germany). The laser diode operates at TEM₀₀ mode [198] and provides 785 nm excitation light. Its pulsed wave (PW) duration is nominally 60-80 ps with frequency of 50 MHz. The PMT offers a minimum photon-count rate of 5 MHz. The TCSPC card has a transit-time spread of 180 ps and a dead time of 125 ns. The single photon counting operation employs equally spaced 256 time gates of duration 16.6 ns, with an initial delay of 1.4 ns for eliminating photons from the excitation pulse in the measurement.

To capture the FLIM images, the laser light was collimated, passed through the confocal system, and focused onto the sample using a 20X, 0.95-NA objective. Single

photon fluorescence from the sample was collected through the same objective and directed by a dichroic mirror toward the confocal pinhole and detected by the PMT. Residual excitation light was removed using a bandstop filter with cutoff wavelengths of 765 nm and 805 nm before the PMT. Data acquisition was performed using the TCSPC card, triggered via a synchronization signal generated by the laser driver for each laser pulse. Images were acquired by unidirectional scanning with the excitation beam using a galvanometric mirror pair as embedded in the confocal microscope system.

To generate the input data for the segmentation methods discussed in this chapter, collected time traces captured from the NIR FLIM system are first analyzed in the SPCImage software (Becker-Hickl, Germany). It replaces time traces per pixel with the cumulative time traces computed using the target pixel and its neighboring eight pixels to increase the SNR. Maximum SNR here is typically defined by the square root of the number of photons acquired in the peak channel of the 256 time gates used to perform TCSPC acquisition [199]. The resulting data of size $128 \times 128 \times 256$ describes spatio-temporal FLT information of the target field-of-view (FOV) of the imaging sample of interest. The first two dimensions of size 128×128 describes the spatial pixel locations in this FOV. The third dimension of length 256 samples for any of these pixels describes a temporal convolution between the instrument response function (IRF) and the fluorescent decay trace for this pixel location. To generate a spatial 2D data for performing the segmentation, the above-mentioned 3D data is analyzed further using the SPCImage software. Here, in each pixel location of

the imaging sample, the earliest time points allow fitting of the IRF, while a single-exponential curve is sufficient to fit the subsequent falloff with time along the third dimension of length 256. A χ^2_r fitness test determined the validity of the fit, providing χ^2_r values < 1.5 for all pixels. The resulting data of size 128×128 describes a spatial FLT information of the target FOV of the imaging sample of interest. This chapter describes methods for performing spatial segmentation for both 2D and 3D described herein.

To conduct specific examples for analyzing performances of the segmentation methods described in this chapter, we imaged *ex vivo* liver tissue sample of a mice treated with an NIR fluorescent dye. The NIR FLIM data was captured using the system and methodology as described above. All animal studies were performed in compliance with the Washington University School of Medicine Animal Studies Committee requirements for the humane care and use of laboratory animals in research.

5.3 Segmentation of Imaging Described by 2D Spatial Information

This section first discusses our proposed CD method for spatially segmenting images described by 2D spatial information. We also briefly review the MGD based method for segmenting such images.

5.3.1 Community Detection Method

We employ the multiresolution CD algorithm to investigate the optimal structure at different resolutions in the input images. We use the number of estimated communities and the information based measures to determine the significant structures at which the “replicas” are strongly correlated. (“replicas” are defined by the independent solutions of the multiresolution CD algorithm attained from different starting points.) We determine different levels of detail and resolutions by setting the resolution parameters.

We minimize a Potts model Hamiltonian for solving the CD problem [40],

$$\mathcal{H} = \frac{1}{2} \sum_{a=1}^q \sum_{i \neq j} (V_{ij} - \bar{V}) [\Theta(\bar{V} - V_{ij}) + \gamma \Theta(V_{ij} - \bar{V})] \delta(\sigma_i, \sigma_j). \quad (5.1)$$

Here the number of communities (segments) q can be specified from the input or left arbitrary and have the algorithm decide by steadily increasing the number of segments q for which we have low energy solutions. The weight V_{ij} denotes the absolute FLT difference between a pixel pair formed by the i^{th} and j^{th} ($\{i, j\} \in \{1, 2, \dots, n\}$) pixels in the input image with n pixels. The Heavyside functions $\Theta(x)$ “turns on” or “off” the edge designation [$\Theta(x > 0) = 1$ and $\Theta(x < 0) = 0$] relative to the background \bar{V} . The Kronecker delta $\delta(\sigma_i, \sigma_j) = 1$ if $\sigma_i = \sigma_j$ and $\delta(\sigma_i, \sigma_j) = 0$ if $\sigma_i \neq \sigma_j$. In this Hamiltonian, by virtue of the $\delta_{\sigma_i \sigma_j}$ term, each spin σ_i interacts only with other spins in its own segment. The spin σ_i ($\forall \sigma_i \in \{1, 2, \dots, q\}$) defines the segment identity for the i^{th} ($i \in \{1, 2, \dots, n\}$) pixel, and the algorithm searches

for it in minimizing the energy defined by Eq. (5.1). As such, the resulting model is *local*—a feature that enables high accuracy along with rapid convergence [18]. As before, minimizing the Hamiltonian of Eq. (5.1) corresponds to identifying strongly connected segments of pixels.

For the 2D NIR FLIM images, we define the weight of any edge formed by two pixels as the absolute difference between their FLTs. We then apply the multiresolution CD algorithm to segment the network formed using the image pixels as nodes. We then analyze as the function of the resolution γ the number of the respective estimated segments \hat{q} and the replicas' information theoretic normalized mutual information I_N and variation of information V . Decreasing γ , the minima of Eq. (5.1) leads to solutions progressively lower intra-community edge densities, effectively “zooming out” toward larger structures. We determine all natural network resolutions by identifying the values of γ for which the replicas exhibit extrema in the average of each of their pairs' information theoretic overlaps when expressed as functions of γ [27]. For more details see [41].

5.3.2 Mixture of Gaussian Distribution Method

The MGD based segmentation method models the data to be a mixture of multivariate Gaussian distributions of unknown means and covariances [67]. For 2D FLIM data, FLTs of all the pixels are assumed to be a mixture of univariate Gaussian distributions of unknown means and variances, and it is segmented using the expectation maximization based optimization method for segmenting mixture of Gaussian distri-

butions developed originally by Hastie et. al [67]. We discuss below the MGD method for the general case of mixture of multivariate Gaussian distributions, assuming that the number of models is known. We then present how we have used this method specifically for the 2D data as described in Section 5.2.

The MGD method models the probability density function (pdf) of the k^{th} ($\forall k \in \{1, 2, \dots, q\}$) component of the data \mathbf{x} $\mathbf{x} \in \mathcal{R}^d$) as Eq. (5.2), where q is the number of models defined by the number of Gaussian distributions in the mixture, d is the dimension of the data, and $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ are the mean and covariance of the k^{th} component. Assuming the weight of the k^{th} component as a_k , the mixture pdf is given by, Eq. (5.3).

$$f_k(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}_k|}} \exp\left(-\frac{(\mathbf{x} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k)}{2}\right) \quad (5.2)$$

$$f(x) = \sum_{k=1}^K a_k f_k(x) \quad (5.3)$$

To estimate the segment identity for the observed incomplete data $\mathbf{x}' = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$, where n is the sample size, we define complete data as $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)\}$, where y_i ($\forall y_i \in \{1, 2, \dots, q\}$) is the unknown segment identity of sample x_i ($i \in \{1, 2, \dots, n\}$). The other unknown parameters $\boldsymbol{\theta}$ includes a_k , $\boldsymbol{\mu}_k$, and $\boldsymbol{\Sigma}_k$ ($\forall k \in \{1, 2, \dots, q\}$). The maximum likelihood function is given by, Eq. (5.4). A classification expectation-maximization algorithm developed by Hastie *et al.* maximizes Eq. (5.4).

$$L(\mathbf{x}'|\boldsymbol{\theta}) = \sum_{i=1}^n \log\left(\sum_{k=1}^q a_k f_k(\mathbf{x}_i)\right) \quad (5.4)$$

To segment a 2D NIR FLIM image, we segment the collection of the FLTs $\{x_1 = \tau_1, x_2 = \tau_2, \dots, x_n = \tau_n\}$ from all the pixels of this image using the univariate version ($d = 1$) of the MGD algorithm. For simplicity, we assume q to be known, and perform the segmentation for increasing number of segments q . The final segmented image is generated using the q for which each of the resulting segments contain at least a predefined number of pixels.

5.4 Segmentation of Images Described by 3D Spatio-Temporal Information

This section extends the methods developed in Section 5.2 for spatially segmenting images described by 3D spatio-temporal information.

5.4.1 Community Detection Method

For spatially segmenting images described by 3D spatio-temporal signatures, we propose to apply multiresolution CD method in two steps. In the first step, the multiresolution CD method identifies strongly-correlated replicas in the images based on their temporal signature, and finds hidden spatial structures in them. Here if a pixel pair is always in the same community (segment) in all the replicas, they must have a strong

preference to be part of the same segment or have a large edge weight. Similarly, if a pixel pair is not always in the same segment in all the replicas, they must have a preference not to be part of the same segment in all replicas or have a small edge weight. The resulting edge weights are in turn used in the second step to perform another multiresolution CD to spatially segment the target image. This strategy allows incorporate both the intensity and FLT information to determine spatial structures automatically in 3D FLIM data described by a spatio-temporal information.

To differentiate the proposed approach with the general image segmentation method developed in our earlier work [41], we note that the method described in [41] first generates r replicas by permuting the “symmetric” initial state of one pixel per segment of the studied system. It then applies the CD algorithm to each replica, and records the segment membership for each pixel. In contrast, for the images described by 3D spatio-temporal signatures in this chapter, the multiresolution CD method as described in Section 5.3.1 considers the spatial 2D frames of this 3D data as series of replicas, and segments each of them in its first step. That is, each replica is represented by a frame corresponding to each temporal location of the 3D FLT data. Then the weight of each pixel pair is calculated based on the statistics of replicas, details of which is described below.

We define the edge weight p_{ij} between the i^{th} and j^{th} ($\{i, j\} \in \{1, 2, \dots, n\}$) pixels of the input image as follows, for performing the multiresolution CD in another second

step. Namely,

$$p_{ij} = \frac{1}{r} \sum_{\alpha=1}^r \delta_{\sigma_i^\alpha, \sigma_j^\alpha}, \quad (5.5)$$

where $\delta_{\sigma_i^\alpha, \sigma_j^\alpha} = 1$, when pixels i and j belong to the same segment in replica α .

Using Eq. (5.5), we redefine the analog Hamiltonian of Eq. 5.1 for the next step multiresolution CD as

$$\mathcal{H} = \frac{1}{2} \sum_{a=1}^q \sum_{i \neq j} (\bar{p} - p_{ij}) \left[\Theta(p_{ij} - \bar{p}) + \gamma \Theta(\bar{p} - p_{ij}) \right] \delta(\sigma_i, \sigma_j). \quad (5.6)$$

We next perform multiresolution CD by minimizing Eq. (5.6).

5.4.2 Mixture of Gaussian Distribution Method

For performing the MGD based segmentation using the 3D spatio-temporal data, we normalize each temporal data corresponding to pixel i ($i \in \{1, 2, \dots, n\}$) in 2D spatial location to form $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{id}]$, where $d = 256$ in our case, see Section 5.2. The normalization here ensures $\max\{x_i | i \in \{1, 2, \dots, d\}\} = 1$. The resulting \mathbf{x}_i ($i \in \{1, 2, \dots, n\}$) are used to perform the multivariate MGD as discussed in Section 5.3.2.

5.4.3 The Classification Expectation Maximization Algorithm

To maximize Eq. (5.4), we use a classification expectation maximization algorithm [67], steps of which we briefly review below. For more discussion, enthusiastic readers might consult the classical book [67] on machine learning written by Hastie *et al.*

1. Initialize parameters
2. E-step: Compute the posterior probabilities for all $i = 1, 2, \dots, n$, $k = 1, 2, \dots, q$, for the iteration l .

$$p_{i,k} = \frac{a_k^{(l)} f_k(\mathbf{x}_i | \boldsymbol{\mu}_k^{(l)}, \boldsymbol{\Sigma}_k^{(l)})}{\sum_{k=1}^q a_k^{(l)} f_k(\mathbf{x}_i | \boldsymbol{\mu}_k^{(l)}, \boldsymbol{\Sigma}_k^{(l)})} \quad (5.7)$$

3. Classification:

$$J_i^{l+1} = \operatorname{argmax}_k \rho_{i,k} \quad (5.8)$$

$$\hat{\rho}_{i,k'} = \begin{cases} 1 & \text{if } k' = \operatorname{argmax}_k \rho_{i,k} \\ 0 & \text{otherwise} \end{cases}$$

4. M-step

$$a_k^{l+1} = \frac{\sum_{i=1}^n \hat{\rho}_{i,k}}{n} \quad (5.9)$$

$$\boldsymbol{\mu}_k^{(l+1)} = \frac{\sum_{i=1}^n \hat{\rho}_{i,k} \mathbf{x}_i}{\sum_{i=1}^n \hat{\rho}_{i,k}} \quad (5.10)$$

$$\boldsymbol{\Sigma}_k^{(l+1)} = \frac{\sum_{i=1}^n \hat{\rho}_{i,k} (\mathbf{x}_i - \boldsymbol{\mu}_k^{(l+1)}) (\mathbf{x}_i - \boldsymbol{\mu}_k^{(l+1)})^T}{\sum_{i=1}^n \hat{\rho}_{i,k}} \quad (5.11)$$

5. Repeat Steps 2-4, until convergence.

5.5 Results

This section describes representative examples of the proposed multiresolution CD method using an *ex vivo* liver tissue sample treated using an NIR fluorescent dye. We also compare the performance of our proposed method with the MGD method.

For the FLIM image of the liver tissue shown in Fig. 5.1A, we define the edge weight between two pixels as the absolute FLT difference between them. We then apply the multiresolution CD to solve the resulting network formed by the image pixels as nodes. Fig. 5.1B shows the plots of the information theoretic overlaps between the replicas of the multiresolution CD, such as their normalized mutual information I_N and variation of information V , together with the number of estimated segments \hat{q} as a function of the resolution γ . Decreasing γ , the minima of Eq. (5.1) leads to solutions progressively lower intra-segment edge densities, effectively “zooming out” toward larger structures. We determine all natural network resolutions by identifying the values of γ for which the replicas exhibit extrema in the average of their information theoretic overlaps when expressed as a function of γ . The extrema and plateau of the average information theoretic overlaps as a function of γ over all replica pairs indicate the natural network resolutions [27].

Figs. 5.1C-5.1G show the image segmentation results by our multiresolution CD algorithm at different resolutions. As the resolution decreases from Fig. 5.1C to 5.1G, the images show decreasingly less detailed structures. In Fig. 5.1C, three segments are clearly visible (light blue, green, and orange), in addition with the background

segment (brown). In Fig. 5.1G, only one segment (light orange) appears against the background (light blue). Results shown in Fig. 5.1D-5.1F show the results produced using the multiresolution CD algorithm with the ranges of γ that are in between of the ones that are used for generating Figs. 5.1C and 5.1G. The segmented images show two distinguished segments located in the background. Thus by using different resolutions γ , the proposed multiresolution CD method is able to detect the structures in different scales.

We then compare in Fig. 5.2 the multiresolution CD method with the MGD method for the FLIM image shown in Fig. 5.1A. In Figs. 5.2A and 5.2C, we present the image segmentation results by the MGD and multiresolution CD, respectively. Note that we evaluated the former method here for increasing number of segments K , and here for $K > 3$, the method started producing segments with pixels fewer than 100. In contrast, the multiresolution CD method selects the number of segments automatically as discussed above. In the current example we used $\gamma \in (0.6, 0.7)$ for performing the segmentation. To measure the performances of these two methods, we plot in Figs. 5.2B and 5.2D the respective normalized average decay traces as the function of photon arrival time in the PMT. The proposed multiresolution CD method is able to identify segments with distinct and diverse decay traces as evident in these two plots. This method allows the user to adjust the resolution γ , and thus to obtain the segments with distinct and diverse decay traces.

We finally compare in Fig. 5.3 the multiresolution CD with the MGD method for the 3D spatio-temporal version of the FLIM image shown in Fig. 5.1A. In Figs. 5.3A

and 5.3C, we present the image segmentation results by the MGD and multiresolution CD, respectively. The optimal number of segments were selected respectively by the MGD method and the CD method using the similar procedures as stated above. To conduct the segmentation using the multiresolution CD developed for segmenting 3D spatio-temporal data, we used single resolutions in both of its steps; particularly, we used $\gamma_1 = 1$ and $\bar{V}_1 = 2.5$ in its first step and $\gamma_2 = 10$ and $\bar{V}_2 = 2.5$ in its second step. Both of the methods are able to produce more connected structures for the 3D spatio-temporal data than that was obtained for the 2D spatial data. The resulting normalized average decay traces of the estimated segments are shown in Fig. 5.3B and 5.3D for the respective methods. For a clearer depiction, Fig. 5.3D here shows the decay traces only for segments with pixels more than 300. Our results indicate that both of the methods here appear to be performing similarly in terms of generating segments with distinct and diverse decay traces.

5.6 Conclusion

We have developed a multiresolution community detection algorithm to segment two-dimensional and three-dimensional fluorescence lifetime imaging microscopy (FLIM) data. The proposed method is able to identify structures in different scales in the input FLIM images based on an information theoretic measures. It performs better than or equivalent to a state-of-the art method for image segmentation based mixture of Gaussian distributions, in producing distinct and diverse decay traces in the

segmented images.

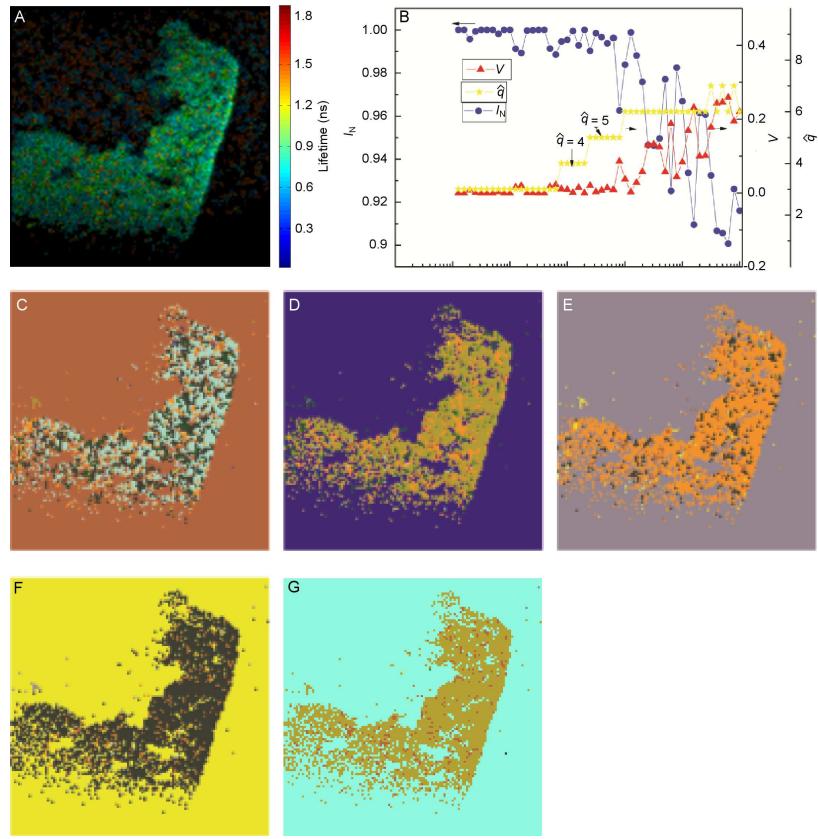


Figure 5.1: Segmentation results for an *ex vivo* liver tissue image with two-dimensional (2D) spatial information: Panel A shows the result generated by the multiresolution community detection (CD) method for segmenting images described by the 2D fluorescence lifetime imaging microscopy data of an *ex vivo* liver tissue sample. Panel B shows the normalized mutual information I_N and the variation of information V of the replicas of the multiresolution CD method and its number of estimated segments \hat{q} as the function of resolution parameter γ . We selected the values of γ at the peaks of the curve described using V , and obtained the segmented images as shown in Panels C-G by our proposed multiresolution CD method. Images with the higher value of γ have more precise structures, see Panels C and D. Images with the lower value of γ have rough structures, see Panels E, F, and G.

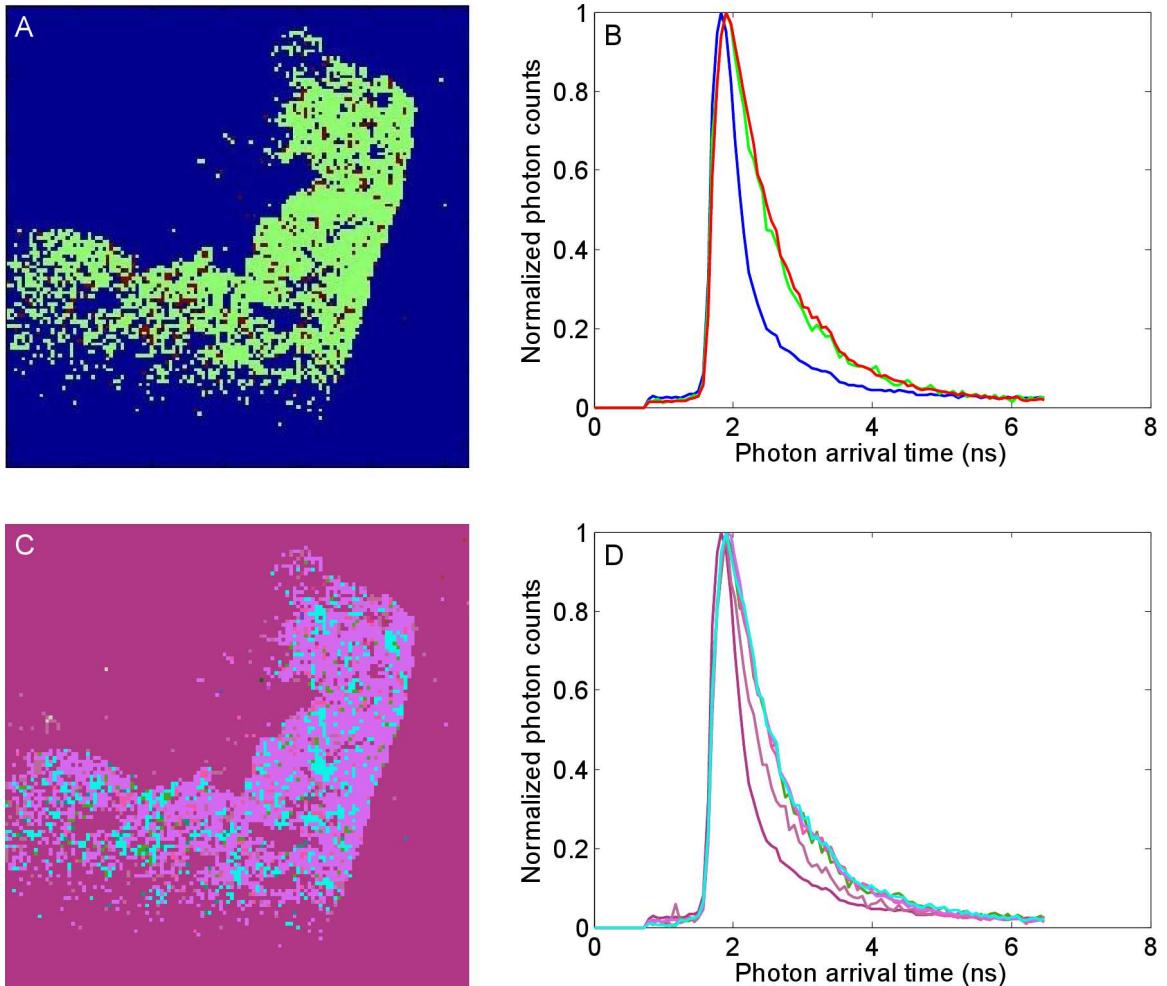


Figure 5.2: Comparison between the multiresolution community detection (CD) method and the mixture of Gaussian distributions (MGD) method for images described by two-dimensional spatial information: Panel A is the segmentation result by the MGD method. Panel B shows the normalized average decay traces corresponding to the estimated segments shown in Panel A. Panel C is the segmentation result by the multiresolution CD method using $\gamma = 0.7$. Panel D shows the normalized average decay traces corresponding to the estimated segments shown in Panel C. Multiresolution CD is able to identify more distinct and diverse decay traces than the MGD method.

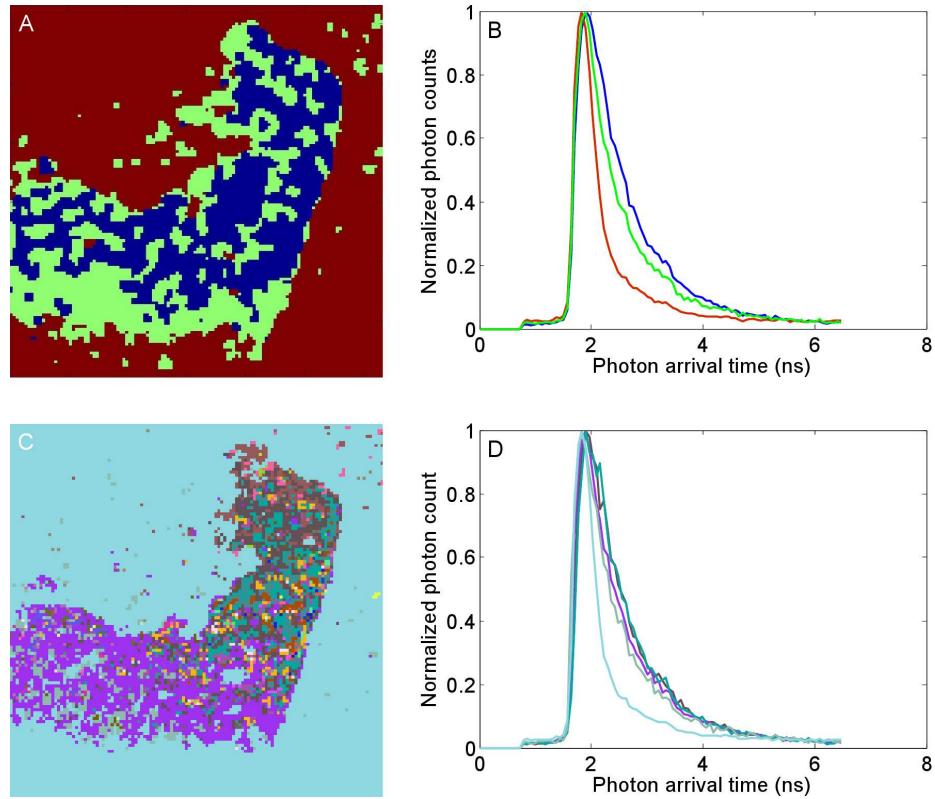


Figure 5.3: Comparison between the multiresolution community detection (CD) method inspired by the replica correlation method and the mixture of Gaussian distributions (MGD) method for images described by three dimensional (3D) spatio-temporal information. Panel A is the result obtained by the MGD method. Panel B shows the normalized average decay traces corresponding to the estimated segments shown in Panel A. Panel C is the segmentation result by the multiresolution CD method developed for images described by 3D spatio-temporal information, using $\gamma_1 = 1$ and $\bar{V}_1 = 2.5$ in its first step and $\gamma_2 = 10$ and $\bar{V}_2 = 2.5$ in its second step. Panel D shows the normalized average decay traces corresponding to the estimated segments shown in Panel C. Both of the methods are able to produce more connected structures for the 3D spatio-temporal data than that was obtained for the 2D spatial data. Our results indicate that both of the methods here appear to be performing similarly in terms of generating segments with distinct and diverse decay traces.

Bibliography

- [1] M.E.J.Newman. *The physics of networks*, Phys. Today, 61(11): 33-38, (2008).
- [2] Leon Danon, Albert Diaz-Guilera, and Alex Arenas. *The effect of size heterogeneity on community identification in complex networks*. J. Stat. Mech.: Theory Exp., 11(11):P11010, (2006)
- [3] Andrea Lancichinetti, Santo Fortunato, and Janos Kertesz. *Detecting the overlapping and hierarchical community structure of complex networks*. New J. Phys. 11:033015, (2009).
- [4] Aaron Clauset, *Finding local community structure in networks*. Phys. Rev. E, 72:026132, (2005).
- [5] Vincent D. Blondel, Jean-Loup Guillaume, Renaude Lambiotte, and Etienne Lefebvre. *Fast unfolding of communities in large networks* J. Stat. Mech.: Theory Exp., 10(10):P10008, (2008).
- [6] M. Girvan and M.E.J.Newman. *Community structure in social and biological networks* Proc. Natl. Acad. Sci. U.S.A., 99(12):7821-7826, (2002).

Bibliography

- [7] Gergely Palla, Imre Derenyi, Illes Farkas, and Tamas Vicsek. *Uncovering the overlapping community structure of complex networks in nature and society.* Nature(London), 435:814-818, (2005).
- [8] Roger Guimera and Luss A. Nunes Amaral. *Functional cartography of complex metabolic networks.* Nature(London), 433:895-900, (2005).
- [9] Reka Albert and Albert-Laszlo Barabasi. *Statistical mechanics of complex networks,* Rev. Mod. Phys., 74(1):47-97, (2002).
- [10] Santo Fortunato and Claudio Castellano. in *Encyclopedia of Complexity and Systems Science*, edited by R.A.Meyers. Springer, (2009).
- [11] Santo Fortunato. *Community detection in graphs.* Phys. Rep., 486:75-174, (2010).
- [12] M.E.J.Newman. *Detecting community structure in networks.* Eur. Phys. J. B. 38(2): 321-330. (2004).
- [13] M.E.J.Newman. *Fast algorithm for detecting community structure in networks,* Phys. Rev. E 69 (6): 066133 (2004).
- [14] Rodrigo Aldecoa and Ignacio Marin. *Deciphering network community structure by Surprise,* PLoS ONE 6(9):e24195 (2011).
- [15] V.D.Blondel, J-L.Guillaume, R.Lambiotte and E.Lefebvre *Fast unfolding of community hierarchies in large networks,* J. Stat. Mech 2008(10):P10008 (2008).

Bibliography

- [16] T.S.Evans *Clique Graphs and overlapping communities*, J.Stat.Mech: P12037 (2010).
- [17] Jorg Rerichardt and Stefan Bornholdt. *Detecting fuzzy community structures in complex networks with a Potts model*, Phys. Rev. Lett., 93:218701 (2004).
- [18] P. Ronhovde and Z. Nussinov, *Local resolution-limit-free Potts model for community detection*, Phys. Rev. E **81**, 046114 (2010).
- [19] M.B.Hastings, *Community detection as an inference problem*, Phys. Rev. E, 74:035102(R) (2006).
- [20] Jussi M. Kumpula, Mikko Kivela, Kimmo Kaski, and Jari Saramaki. *A sequential algorithm for fast clique percolation*, Phys. Rev. E, 78:026109 (2008).
- [21] M.E.J.Newman. *Finding community structure in networks using the eigenvectors of matrices*, Phys.Rev.E, 74:036104, (2006).
- [22] Martin Rosvall and Carl T. Bergstrom. *Maps of random walks on complex networks reveal community structure*, Proc. Natl. Acad. Sci. U.S.A., 105(4):1118-1123, (2008).
- [23] V.Gudkov, V.Montealegre, S. Nussinov, and Z. Nussinov. *Community detection in complex networks by dynamical simplex evolution*, Phys. Rev. E, 78:016113 (2008).

Bibliography

- [24] Aaron Clauset, M.E.J.Newman, and Christopher Moore. *in Proceedings of the 23th International Conference on Machine Learning*, Association of Computing Machinery, New York (2006).
- [25] M.E.J.Newman. *Analysis of weighted networks*, Phys. Rev. E, 70(5):056131, (2004).
- [26] A.Arenas, A. Fernandez, and S. Gomez. *Analysis of the structure of complex networks at different resolution levels*, New. J. Phys., 10:053039, (2008).
- [27] P. Ronhovde and Z. Nussinov, *Multiresolution community detection for megascale networks by information-based replica correlations*, Phys. Rev. E **80**, 016109 (2009).
- [28] Hristo N. Djidjev. *A scalable multilevel algorithm for graph clustering and community structure detection*. In Algorithms and Models for the Web-Graph: Fourth International Workshop, WAW 2006, Revised Papers, volume 4936, pages 117-128, Berlin, Heidelberg, 2007. Springer-Verlag.
- [29] Marina Meila. *Comparing clusterings by the variation of information*, Learning Theory and Kernel Machines: 173-187, (2003).
- [30] M.Mezard, G. Parisi, and R. Zecchina, *Analytic and algorithmic solution of random satisfiability problems*, Science, 297:812-815, (2002).

- [31] Florent Krzakala and Lenka Zdeborova, *Phase transitions and computational difficulty in random constraint satisfaction problems*, Journal of Physics: Conference Series, 95(1):012012, (2008).
- [32] Remi Monasson, Riccardo Zecchina, Scott Kirkpatrick, Bart Selman, and Lidror Troyansky, *Determining computational complexity from characteristic ‘phase transitions’*, Nature, 400:133-137, (1999).
- [33] Tad Hogg, Bernardo A. Huberman, and Colin P. Williams, *Phase transitions and the search problem*, Artificial Intelligence, 81(1-2):1-15, (1996).
- [34] M.Bayati, C. Borgs, A. Braunstein, J. Chayes, A. Ramezanpour, and R. Zecchina, *Statistical mechanics of steiner trees*, Phys. Rev. Lett., 101(3):037208, (2008).
- [35] Jie Zhou and Haijun Zhou, *Ground-state entropy of the random vertex-cover problem*, Phys. Rev. E, 79(2):020103, (2009).
- [36] Peter Cheeseman, Bob Kanefsky and William M. Taylor, *Where the really hard problems are*, In Conference on Artificial intelligence archive. Proceedings of the 12th international joint conference on Artificial intelligence, volume 1, pages 331-337, (1991).
- [37] Hideyoshi Nishimori and K.Y.Michael Wong, *Statistical mechanics of image restoration and error-correcting codes*, Phys. Rev. E, 60(1):132-144, (1999).

Bibliography

- [38] Valmir C. Barbosa and Rubens G. Ferreira, *On the phase transitions of graph coloring and independent sets*, <http://arxiv.org/abs/cond-mat/0309518>, (2002).
- [39] J. Reichardt, *Structure in Complex Networks*, Lecture Notes in Physics Vol. **766**, Springer-Verlag, Berlin, (2009).
- [40] Dandan Hu, Peter Ronhovde and Zohar Nussinov, *Phase transitions in random Potts systems and the community detection problem: spin-glass type and dynamic perspectives*, Philosophical Magazine 92, 4 (2012).
- [41] Dandan Hu, Peter Ronhovde and Zohar Nussinov, *The stability to instability transition in the structure of large scale networks*, <http://arxiv.org/abs/1204.4167>, submitted to Phys.Rev.E.
- [42] Y-J Zhang, *Advances in Image and Video Segmentation*, Editor IRM Press, (2006).
- [43] J. Bigun, *Vision with Direction: A Systematic Introduction to Image Processing and Computer Vision*, Springer, (2010).
- [44] L. O'Gorman, M. J. Sammon, and M. Seul, *Practical Algorithms for Image Analysis*, Cambridge University Press, (2008).
- [45] William E. Lorensen, Harvey E. Cline, *Marching Cubes: A high resolution 3D surface construction algorithm*, Computer Graphics, Vol. 21, Nr. 4, July 1987.

- [46] D. L. Pham, C. Xu, and J. L. Prince, *Current Methods in Medical Image Segmentation*, Annual Review of Biomedical Engineering, **2**, 315-337 (2000).
- [47] R. Brunelli and T. Poggio, *Face Recognition: Features versus Templates*, IEEE transaction on pattern analysis and machine intelligence, **15**, 10 (1993).
- [48] B. Bhanu, X. Tan, *Computational algorithms for fingerprint recognition*, Boston : Kluwer Academic Publishers (2004).
- [49] G. McDougall and A. S. Roberts, Commercializing Air Traffic Control: Have the Reforms Worked, Canadian Public Administration. **51**, No. 1, 45-69 (2009).
- [50] S. Carsten, M. Ulrich, and C. Wiedemann, *Machine Vision Algorithms and Applications*. Weinheim: Wiley-VCH. p. 1. ISBN 9783527407347 (2008).
- [51] M. Sezgin and B. Sankur, *Survey over image thresholding techniques and quantitative performance evaluation*, Journal of Electronic Imaging **13** (1): 146-165 (2003).
- [52] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, A. Y. Wu, *An efficient k-means clustering algorithm: Analysis and implementation*. IEEE Trans. Pattern Analysis and Machine Intelligence **24**: 881-892 (2002).
- [53] H. Mobahi, S. Rao, A. Yang, S. Sastry and Y. Ma, *Segmentation of Natural Images by Texture and Boundary Compression*, <http://arxiv.org/abs/1006.3679> (2010).

- [54] O. Chapelle, P. Haffner, and V. N. Vapnik, *Support Vector Machines for Histogram-Based Image Classification*, IEEE Transactions on neural networks, **10**, 5 (1999).
- [55] P. Meer, S. Wang, H. Wechsler, *Edge detection by associative mapping*, Pattern Recognition, **22**, 491-504 (1989).
- [56] S. A. Hojjatoleslami and J. Kittler, *Region Growing: A New Approach*, IEEE Transaction on image processing, **7**, No. 7, 1057-7149, p. 1079 (1998).
- [57] C. Wojtan, N. Thürey, M. Gross, G. Turk, *Deforming Meshes that Split and Merge*. ACM Trans. Graph. **28**, 3, Article 76 (2009).
- [58] S. Fortunato, *Community detection in graphs*, Physics Reports **486**, 75-174 (2010).
- [59] S. Beucher, *The Watershed Transformation Applied To Image Segmentation*, Scanning Microscopy International, supp. **6**, 299-314 (1992).
- [60] A. Kelemen, G. Szekely, G. Gerig, *Three-dimensional model-based segmentation of brain MRI*, Workshop on Biomedical image analysis, 4-13 (1998).
- [61] H. Trinh, *Efficient Stereo Algorithm using Multiscale Belief Propagation on Segmented Images*. In British Machine Vision Conference (BMVC) (2008).

- [62] B. Micusik and A. Hanbury, *Steerable Semi-automatic Segmentation of Textured Images*, Scandinavian Conference on Image Analysis (SCIA), Joensuu, Finland (2005).
- [63] P. Ronhovde, S. Chakrabarty, M. Sahu, K. K. Sahu, K. F. Kelton, N. Mauro, and Z. Nussinov, *Detection of hidden structures on all scales in amorphous materials and complex physical systems: basic notions and applications to networks, lattice systems, and glasses*, arXiv:1101.0008 (2011).
- [64] P. Ronhovde, S. Chakrabarty, D. Hu, M. Sahu, K. F. Kelton, N. A. Mauro, K . K. Sahu, and Z. Nussinov, *Detecting hidden spatial and spatio-temporal structures in glasses and complex physical systems by multiresolution network clustering*, Eur. J. of Phys. E **34**, 105 (2011).
- [65] Dzung L. Pham, Chenyang Xu, Jerry L. Prince *A survey of current methods in medical image segmentation*, Annual Review of Biomedical Engineering (1998).
- [66] S. Zulaikha Beevi *An effective approach for segmentation of MRI images: Combining spatial information with fuzzy c-mean clustering*, European Journal of Scientific Research, **41**, 437, (2010).
- [67] Lalit Gupta, Thotsapon Sortrakul, *A gaussian-mixed-based image segmentation algorithm*, Pattern Recognition, **31**, 3, (1998).

- [68] VK Tidwell, JR Garbow, AS Krupnick, JA Engelbach, A Nehorai, *Quantitative analysis of tumor burden in mouse lung via MRI*, Magn Reson Med, **67(2)**, 572, (2012).
- [69] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, *Defining and identifying communities in networks*, Proc Natl Acad Sci USA **101**, 2658-2663 (2004).
- [70] A. Lancichinetti, F. Radicchi and J. Ramasco, *Statistical significance of communities in networks*, Phys. Rev. E **81**, 046110 (2010).
- [71] S. Gregory, *Finding overlapping communities in networks by label propagation*, New J. Phys. **12**, 103018 (2010).
- [72] J. Duch and A. Arenas, *Community detection in complex networks using extremal optimization*, Phys. Rev. E **72**, 027104 (2005).
- [73] V. Gudkov, V. Montealegre, S. Nussinov, and Z. Nussinov, *Community detection in complex networks by dynamical simplex evolution*, Phys. Rev. E **78**, 016113 (2008).
- [74] J. Šíma, S. Schaeffer, *On the NP-Completeness of some graph cluster measures*, Proceedings of the Thirty-second International Conference on Current Trends in Theory and Practice of Computer Science (Sofsem 06), in: Lecture Notes in Computer Science, Vol. **3831**, p. 530 (2006).

- [75] B. Good, Y. de Montjoye, and A. Clauset, *Performance of modularity maximization in practical contexts*, Phys. Rev. E **81**, 046106 (2010).
- [76] M. Mézard, G. Parisi, and M. Virasoro, *Spin Glass Theory and Beyond*, World Scientific Lecture Notes in Physics, vol. 9 (1986).
- [77] A. Braunstein, M. Mézard, and R. Zecchina, *Survey Propagation: An algorithm for satisfiability*, Random Structures and Algorithms **27**, 201-226 (2005).
- [78] R. Gallager, *Low Density Parity Check Codes*, Monograph, M.I.T. Press (1963).
- [79] J. Pearl, *Reverend Bayes on inference engines: A distributed hierarchical approach*, Proceedings of the Second National Conference on Artificial Intelligence. AAAI-82: Pittsburgh, PA. Menlo Park, California: AAAI Press, 133 (1982).
- [80] D. Hu, P. Ronhovde, and Z. Nussinov, *A Replica Inference Approach to Unsupervised Multi-Scale Image Segmentation*, Phys. Rev. E 85, 016101, (2012).
- [81] M. Mézard and G. Parisi, *The cavity method at zero temperature*, Journal of Statistical Physics **111**, Nos. 1/2 (2003).
- [82] J. Reichardt and M. Leone, *(Un)detectable Cluster Structure in Sparse Networks*, Phys. Rev. Lett. **101**, 078701 (2008).
- [83] A. Allahverdyan, G. Ver Steeg and A. Galstyan, *Community detection with and without prior information*, EuroPhysics Letters **90**, 18002 (2010).

- [84] S. Kirkpatrick, C.D.Gelatt and M.P.Vecchi *Optimization by Simulated Annealing*, Science. **220**: 4598 (1983).
- [85] A. Lancichinetti, S. Fortunato, and F. Radicchi, *Benchmark graphs for testing community detection algorithms*, Phys. Rev. E **78**, 046110 (2008).
- [86] S. Fortunato and M. Barthélemy, *Resolution limit in community detection*, Proc. Natl. Aca. Sci. U.S.A. **104**, 36 (2007);
J. M. Kumpula, J. Saramäki, K. Kaski, and J. Kertész, *Limited resolution in complex network community detection with Potts model approach*, Euro. Phys. J. B **56**, 41 (2007).
- [87] J. Villain, R. Bidaux, J. P. Carton and R. Conte, *Order as an Effect of Disorder*, J. Phys. (Paris) 41, no.11, 1263C1272 (1980).
- [88] E. F. Shender, *Antiferromagnetic Garnets with Fluctuationally Interacting Sub-lattices*, Sov. Phys. JETP 56 178C184 (1982).
- [89] C. L. Henley, *Ordering Due to Disorder in a Frustrated Vector Antiferromagnet*, Phys. Rev. Lett. 62 2056C2059 (1989).
- [90] Z. Nussinov, M. Biskup, L. Chayes, and J. van den Brink, *Orbital order in classical models of transition-metal compounds*, Europhysics Letters, 67, 990-996 (2004).
- [91] K. Jonason, E. Vincent, J. Hammann, J. Bouchaud, and P. Nordblad, *Memory and Chaos Effects in Spin Glasses*, Phys. Rev. Lett. **81**, 3243 (1998).

- [92] K. Jonason, P. Nordblad, E. Vincent, J. Hammann, and J. Bouchaud, *Memory interference effects in spin glasses*, Eur. Phys. J. B **13**, 99 (2000).
- [93] D. Foster, J. Foster, M. Paczuski, and P. Grassberger, *Clustering Phase Transition and Hysteresis: Pitfalls in Constructing Network Ensembles*, Phys. Rev. E **81**, 046115 (2010).
- [94] D. Stariolo, M. Montemurro, and F. Tamarit, *Aging dynamics of $\pm J$ Edwards-Anderson spin glasses*, Eur. Phys. J. B **32**, 361 (2003).
- [95] L. Berthier and A. Young, *Aging dynamics of the Heisenberg spin glass*, Phys. Rev. B **69**, 184423 (2004).
- [96] U. Brandes, D. Delling, M. Gaertler, R. Gorke, M. Hoefer, Z. Nikoloski, and D. Wagner, *Maximizing Modularity is hard*, <http://arxiv.org/pdf/physics/0608255> (2006).
- [97] S. Cook, *The complexity of theorem proving procedures*, Proceedings, Third Annual ACM Symposium on the Theory of Computing, ACM, New York, 151-158 (1971)
- [98] V. Lubchenko and P. Wolynes, *Theory of Structural Glasses and Supercooled Liquids*, Annu Rev Phys Chem **58**, 235-266 (2006).
- [99] F. Ricci-Tersenghi, *Being Glassy Without Being Hard to Solve*, Science **330**, 1639 (2010).

Bibliography

- [100] Similar expressions and analysis appear for an inertial system (wherein the force is given by $(d^2\vec{\eta}_i/dt^2)$). The damping accentuates the difference between converging and diverging trajectories as seen on the experimental time scale as discussed in the text.
- [101] D. Fisher and D. Huse, *Equilibrium behavior of the spin-glass ordered phase*, Phys. Rev. B **38**, 386 (1998).
- [102] A. Bray and M. Moore, *Chaotic Nature of the Spin-Glass Phase*, Phys. Rev. Lett. **58**, 57 (1987).
- [103] C. Thomas, D. Huse, and A. Middleton, *Chaos and universality in two-dimensional Ising spin glasses*, <http://arxiv.org/abs/1012.3444v1> (2010)
- [104] G. Parisi, *Statistical Field Theory*, (Addison-Wesley, New York, 1988).
- [105] J. Zinn-Justin, *Quantum Field Theory and Critical Phenomena*, (Oxford University Press, New York, 2002).
- [106] G. Biroli, C. Chamon, and F. Zamponi, *Theory of the superglass phase*, Phys. Rev. B **78**, 224306 (2008).
- [107] M.J.Barber and J.W.Clark, Phys.Rev. E **80**, 026129 (2009).
- [108] L.Danon, A.Diaz-Guilera, J.Duch, and A.Arenas, J. Stat. Mech: Theory Exp. **9**, P09008 (2005).

Bibliography

- [109] A. Noack and R. Rotta, in Experimental Algorithms, edited by J. Vahrenhold (Springer-Verlag Berlin, Heidelberg, 2009), vol.5526, pp.257-268.
- [110] H.W.Shen and X.Q.Cheng, J.Stat.Mech: Theory Exp, P10020 (2010).
- [111] S. Mertens, Phys. Rev. Lett. **81**, 4281 (1998).
- [112] I.P.Gent and T.Walsh, Artificial Intelligence **88**, 349 (1996).
- [113] M. Weigt and A.K.Hartmann, Phys. Rev. Lett. **84**, 6118 (2000).
- [114] L.Lacasa, B.Luque, and O.Miramontes, New J. Phys. **10**, 023009 (2008).
- [115] H.Bauke, S. Mertens, and A.Engel, Phys. Rev. Lett. **90**, 158701 (2003).
- [116] G.Mukherjee and S.S Manna, Phys. Rev. E **71**, 066108 (2005).
- [117] R. Arevalo, I. Zuriguel and D. Mara, Phys. Rev. E **81**, 041302 (2010).
- [118] B. Ashok and T.K.Patra, Pramana **75**, 549 (2010).
- [119] K.Rose, E.Gurewitz, and G.C.Fox, Phys. Rev. Lett. **65**, 945 (1990).
- [120] T. Graepel, M. Burger, and L. Pbermaner, Phys. Rev. E **56**, 3876 (1997).
- [121] S.N.Dorogovtsev, A.V.Goltsev, and J.F.F.Mendes, Rev. Mod. Phys. **80**, 1275 (2008).
- [122] A. Decelle, F. Krzakala, C. Moore, L. Zdeborová, *Phase transition in the detection of modules in sparse networks*, arXiv:1102.1182 (2011).

- [123] E.J.Lee, K.I.Goh, B. Kahng, and D.Kim, Phys. Rev. E **71**, 056108 (2005).
- [124] Y.Moreno, R.Pastor-Satorras, A. Vzquez, and A. Vespignani, Europhys. Lett. **62**, 292 (2003).
- [125] W.X.Wang and G. Chen, Phys. Rev. E **77**, 026101 (2008).
- [126] J.F.Zheng, Z. Y. Gao and X.M.Zhao, Europhys.Lett. **79**, 58002 (2007).
- [127] J.Wu, Z.Gao and H.Sun, Phys. Rev. E **74**, 066111 (2006).
- [128] Y. Ikeda, T. Hasegawa and K.Nemoto, Journal of Physics: Conference Series **221**, 012005 (2010).
- [129] A.Tahbaz-Salehi and A. Jadbabaie, in Proceedings of the 2007 American Control Conference (IEEE, 2007), pp. 699-704.
- [130] M.A.Serrano and M.Boguna, Phys. Rev. Lett. **97**, 088701 (2006).
- [131] C. Moore and M.E.J.Newman, Phys. Rev. E **61**, 5678 (2000).
- [132] P.Ronhovde, D.Hu and Z.Nussinov, e-print arXiv:1204.3649, (2012).
- [133] P.W.Kasteleyn and C.M.Fortuin, in Proceedings of the International Conference on Statistical Mechanics, September 9-14, Koyto (1969), vol. 26, p11.
- [134] U.N.Raghavan, R.Albert and S. Kumara, Phys. Rev. E **76**, 036106 (2007).
- [135] M.T.Mercaldo, J.C.Angles dAuriac and F. Igloi, Phys. Rev. E **69**, 056112 (2004).

Bibliography

- [136] J.C.A.dAuriac, F.Igloi, M.Preissmann and A.Sebo, *J. Phys.A* **35**, 6973 (2002).
- [137] R.Juhasz, H.Rieger and F. Igloi, *Phys.Rev.E* **64**, 056122 (2001).
- [138] M.T.Mercaldo, J.C.A.dAuriac and F.Igloi, *Europhys. Lett.* **70**, 733 (2005).
- [139] S.C.Chang and R.Shrock, *Int.J.Mod.Phys.B* **21**, 979 (2007).
- [140] L.Mittag and M.J.Stephan, *J.Phys.A:Math.Nucl.Gen.* **7**, L109 (1974).
- [141] P.A.Pearce and R.B. Griffiths, *J.Phys.A: Math. Gen.* **13**, 2143 (1980).
- [142] D.J.A.Welsh and C. Merino, *J. Math. Phys.* **41**, 1127 (2000).
- [143] V.A.Traag, P.Van Dooren, and Y. Nesterov, *Phys. Rev. E* **84**, 016114 (2011)
- [144] B. Jackson and A.D.Sokal, *J. Combinatorial Theory, Series B* **99**, 869 (2009).
- [145] C.D.Batista and Z. Nussinov, *Phys. Rev. B* **72**, 045137 (2005).
- [146] Z. Nussinov, G.Ortiz and E. Cobanera, e-print arXiv:1110.2179 (2011).
- [147] L.H.Kauffman, *Discrete Appl.Math* **25**, 105 (1989).
- [148] L.H.Kauffman, *Rep.on Progress in Phys.* **68**, 2829 (2005).
- [149] E.Witten, *Comm. Math. Phys.* **121**, 351 (1989).
- [150] M.B.Thistlethwaite, *Topolpgy* **26**, 297 (1987).
- [151] J. Kim, J. W. Fisher, A. Yezzi, M. Cetin, and A. S. Willsky, *A Nonparameteric Statistical Method for Image Segmentation Using Information Theory and Curve*

- Evolution*, IEEE Transitions on Image Processing, **14**, No. 10, 1057-7149, p. 1486 (2005).
- [152] S. Osher and N. Paragios, *Geometric Level Set Methods in Imaging Vision and Graphics*, Springer Verlag, ISBN 0387954880 (2003).
- [153] J. Shi, J. Malik, *Normalized cuts and image segmentation*, IEEE Transactions on Pattern Analysis and Machine Intelligence, **22**, Issue 8, 888-905 (2000).
- [154] S. German and D. German, *Stochastic Relaxation, Gibbs distributions and the Bayesian restoration of images*, IEEE Transactions on Patterns, Anal. Mach. Intell. PAMI-6: 721 (1984).
- [155] J. L. Marroquin, *Deterministic Bayesian Estimation of Markovian random fields with applications to computer vision*, Proc. 1st Inter. Conf. Comput. Vision, London (1987).
- [156] E. B. Gamble and T. Poggio, *Visual integration and detection of discontinuities: The key role of intensity edges*, A. I. Memo No. 970, Artificial Intelligence Laboratory, MIT (1987).
- [157] D. Geiger and F. Girosi, *Parallel and deterministic algorithms for MRFs: surface reconstruction and integration*, PAMI-13(5) (1991).
- [158] L. Grady, *Random Walks from Image segmentation*, IEEE Transaction on Pattern Analysis and Machine Intelligence, **28**, No. 11 (2006).

- [159] L. Grady, E. L. Schwartz, *Isoperimetric graph partitioning for image segmentation*, IEEE Transactions on Pattern Analysis and Machine Intelligence, **28**, Issue 3 (2006).
- [160] S.C. Amartur, D. Piraino and Y. Takefuji, *Optimization neural networks for the segmentation of magnetic resonance images*, IEEE Transactions on Medical Imaging, **11**, No.2 (1992).
- [161] <http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/>
- [162] <http://www.gifford.co.uk/principia/Illusions/dalmatian.htm>
- [163] J. M. Kumpula, J. Saramäki, K. Kaski, and J. Kertész, page 660116 *Noise and Stochastics in Complex Systems and Finance*, vol. **6601** of SPIE Conference Series (2007).
- [164] R. B. Potts, *Some Generalized Order-Disorder Transformations*, Proc. Camb. Philos. Soc. **48**, 106C109 (1952).
- [165] S. Peng, B. Urbanc, L. Cruz, B. T. Hyman, and H. E. Stanley, *Neuron recognition by parallel Potts segmentation*, Proc. Natl. Acad. Sci. USA **100**, 3847C3852 (2003).
- [166] X. Descombes, M. Moctezuma, H. Maître, and J. P. Rudant, *Coastline detection by a Markovian segmentation on SAR images*, Signal Process. **55**, 123-132 (1996).

- [167] K. Tanaka, *Statistical-mechanical approach to image processing*, J. Phys. A: Math. Gen. **35**, R81CR150 (2002).
- [168] F. W. Bentrem, *A Q-Ising model application for linear-time image segmentation*, Central European Journal of Physics, vol. **8**, pp. 689-698 (2010).
- [169] M. Blatt, S. Wiseman and E. Domany, *Superparamagnetic clustering*, Phys. Rev. Lett **76**, 3251 (1996).
- [170] J. M. Kumpula, J. Saramäki, K. Kaski, and J. Kertész, *Limited resolution in complex network community detection with Potts model approach*, Euro. Phys. J. B **56**, 41 (2007).
- [171] Rather explicitly, there are four different cases to consider: (i) It is desirable to have nodes i and j in the same community if $A_{ij} = 1$ (i.e., if nodes i and j are linked). Thus, a negative energy ($-A_{ij}/2$) is introduced for having nodes i and j in the same community ($\delta(\sigma_i, \sigma_j) = 1$). (ii) It is undesirable to have nodes i and j in different communities if they linked ($A_{ij} = 1$). Thus, a positive energy contribution A_{ij} is added in such a case. This is tantamount to augmenting the energy by $(1 - \delta(\sigma_i, \sigma_j))A_{ij}/2$. (iii) If nodes i and j are not linked ($A_{ij} = 0$) then an energy penalty should appear when putting them within the same community. This is captured by the term $\gamma J_{ij}\delta(\sigma_i, \sigma_j)/2$ with $J_{ij} = (1 - A_{ij})$. (iv) If nodes i and j are not linked then it is energetically favorable to have them in different communities. This leads to a negative contribution $(-\gamma J_{ij}\delta(\sigma_i, \sigma_j)/2)$ in the energy. Putting all of these pieces together and summing over all node pairs

(ij) leads, up to an innocuous additive constant, to the Hamiltonian of Eq. (??).

All energy contributions are “absolute” and set by the existence of a link (or its absence) between the two members of each node pair without global (“null model” type) comparisons to other nodes in the system. An identical analysis in the case of more general weighted graphs leads to the Hamiltonian of Eq.(4.2) that will be employed in the current work for weighted images.

[172] S. Mossa and G. Tarjus, *Locally preferred structure in simple atomic liquids*, J. Chem. Phys **119**, 8069 (2003).

[173] The nodes are sampled in the (effectively random) order that they appear in the communities given the current state of the system. One may also apply a randomized algorithm with an essentially identical result.

[174] D. Martin and C. Fowlkes and D. Tal and J. Malik, *A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics* , Proc. 8th Int'l Conf. Computer Vision, **2**, 416–423 (2001).

P. Arbelaez, M. Maire, C. Fowlkes and J. Malik, *Contour Detection and Hierarchical Image Segmentation*, IEEE TPAMI, **33**, No. 5, 898 (2011).

[175] C. Granell, S. Gómez, and A. Arenas, *Mesoscopic analysis of networks: applications to exploratory analysis and data clustering*, Chaos **21**, 016102 (2011).

[176] <http://research.microsoft.com/en-us/projects/objectclassrecognition/>

- [177] D. Martin, C. Fowlkes and J. Malik, *Learning to Detect Natural Image Boundaries Using Local Brightness, Color, and Texture Cues*, IEEE Transactions on Pattern Analysis and Machine Intelligence, VOL. **26**, No.5, (2004).
- [178] M. Maire, P. Arbelaez, C. Fowlkes and J. Malik, *Using Contours to Detect and Localize Junctions in Natural Images*, Computer Vision and Pattern Recognition, (2008).
- [179] D. Shechtman, I. Blech, D. Gratias, J. Cahn, *Metallic Phase with Long-Range Orientational Order and No Translational Symmetry*, Phys. Rev. Lett. **53**, 1951 (1984).
- [180] <http://math.ucr.edu/home/baez/week271.html>
- [181] <http://www.gregegan.net/APPLETS/12/12.html>
- [182] A. van Blaaderen, *Quasicrystals from nanocrystals*, NATURE **461**, 15 (2009).
- [183] <http://alumnus.caltech.edu/~kantner/zebras/pictures.html>
- [184] <http://en.wikipedia.org/wiki/Camouflage>
- [185] D. Hu, P. Ronhovde, and Z. Nussinov, *in preparation*
- [186] M. S. Caudill, S. F. Brandt, Z. Nussinov, and R. Wessel, *Parameter space analysis of delayed neural feedback triads*, Phys. Rev. E **80**, 051923 (2009)
- [187] Tim McInerney and Demetri Terzopoulos *Deformable models in medical image analysis: a survey*, Medical Image Analysis, **1(2)**, 91, (1996).

Bibliography

- [188] Akselrod-Ballin, A., M. Galun *Atlas guided identification of brain structures by combining 3D segmentation and SVM classification*, Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv, **9** 209, (2006).
- [189] R. Nothdurft, P. Sarder, S. Bloch, J. Culver, and S. Achilefu, “Fluorescence lifetime imaging microscopy using near-infrared contrast agents,” to appear in *Journal of Microscopy*, 2012.
- [190] M. Y. Berezin and S. Achilefu, “Fluorescence lifetime measurements and biological imaging,” *Chemical Reviews*, vol. 110, no. 5, pp. 2641-2684, 2010.
- [191] D. S. Elson, et al., “Wide-field fluorescence lifetime imaging with optical sectioning and spectral resolution applied to biological samples,” *Journal of Modern Optics*, vol. 49, no. 5-6, pp. 985-995, 2002.
- [192] S. E. D. Webb, et al., “Single-Molecule Imaging and Fluorescence Lifetime Imaging Microscopy Show Different Structures for High- and Low-Affinity Epidermal Growth Factor Receptors in A431 Cells,” *Biophysical Journal*, vol. 94, no. 3, pp. 803-819, 2008.
- [193] P. J. Verveer et al. “Quantitative imaging of lateral ErbB1 receptor signal propagation in the plasma membrane,” *Science*, vol. 290, no. 5496, pp. 1567-1570, 2000.
- [194] M. Lahn et al., “Two-photon microscopy and fluorescence lifetime imaging reveal stimulus-induced intracellular Na⁺ and Cl⁻ changes in cockroach salivary

Bibliography

- acinar cells,” *American Journal of Physiology*, vol. 300, no. 6, pp. C1323-1336, 2011.
- [195] M. Kneen et al., “Green fluorescent protein as a noninvasive intracellular pH indicator,” *Biophysical Journal*, vol. 74, no. 3, pp. 1591-1599, 1998.
- [196] D. S. Ushakov et al., “Response of rigor cross-bridges to stretch detected by fluorescence lifetime imaging microscopy of myosin essential light chain in skeletal muscle fibers,” *The Journal of Biological Chemistry*, vol. 286, no. 1, pp. 842-850, 2011.
- [197] M. Keese et al., “Fluorescence lifetime imaging microscopy of chemotherapy-induced apoptosis resistance in a syngenic mouse tumor model,” *International Journal of Cancer*, vol. 126, no. 1, pp. 104-113, 2010.
- [198] D. K. Cheng, “Field and Wave Electromagnetics. Boston”, Addison-Wesley, 1989.
- [199] M. Kollner and J. Wolfrum, “How Many Photons Are Necessary for Fluorescence-Lifetime Measurements” *Chemical Physics Letters* vol. 200, no. 1-2, pp. 199-204, 1992.