A higher-order Monte Carlo cluster dynamics for community detection in spatially embedded graphs

Abstract. Community detection in spatially embedded graphs is notoriously difficult when edges may be attractive or repulsive, because the resulting GIBBS energy landscape is highly frustrated.

Motivated by methods from statistical physics, we develop a GIBBS-based framework for community detection on weighted signed graphs. Within this framework we define GIBBS-weak recovery, which requires a sampling algorithm to correlate with the GIBBS measure. We then introduce a higher-order cluster-dynamics sampler for two balanced communities that augments the classical SWENDSEN–WANG dynamics with triangular bonds. We eliminate local frustration while rigorously preserving the target GIBBS distribution. Using a coupling argument, we prove that *percolation* is a necessary condition for GIBBS-weak recovery for any cluster dynamics.

On the Euclidean random-graph model of SANKARARAMAN & BACCELLI, our framework coincides with theirs. In this setting, our notion of weak recovery matches the standard definition. The theory reproduces their results when the edge-based SWENDSEN–WANG dynamics is used. We strictly improve their recovery bound when our triangle-based dynamics is applied to the triangular-grid stochastic block model.

Empirically, on their challenging Euclidean random-graph benchmark, our proposed algorithm substantially outperforms current state-of-the-art methods.

KEYWORDS: Markov Chain Monte Carlo, Gibbs energy, spin configuration, Swendsen-Wang dynamics, higher-order interaction, percolation.

1 Introduction Graphs provide a convenient language for modeling a wide range of real-world systems, from social networks to biological interaction maps. A key problem is to cluster vertices that share common properties, a task known as *community detection* [8]. Algorithms inspired by statistical physics offer intuitive, mathematically tractable solutions [13, 18].

Problem setting. We consider a simple graph $G = (V, E)$ with $|V|$ vertices, each belonging to one of K unknown communities. Although our framework applies to general graphs, we focus on *spatially embedded* graphs, *e.g.*, $V \subseteq \mathbb{R}^d$.

The algorithm receives as input

- the graph structure (V, E);
- and a list of non-zero edge weights $\big(W_e\big)_{e\in E}$ whose sign indicates whether the corresponding vertices tend to belong to the same class ($W_e > 0$) or different ones ($W_e < 0$), and whose modulus $|W_e|$ reflects the interaction strength.

Following the terminology in statistical physics, we refer to edges with positive weights as *ferromagnetic*, and those with negative weights as *antiferromagnetic*. The coexistence of both types of edges is the root cause of *frustration*, a phenomenon that significantly complicates the sampling process, and on which we will elaborate later.

Contributions.

(i) Model. To tackle the problem of community detection on the weighted graph, we formalize a binary spin model in which the GIBBS distribution built on the given graph plays a fundamental role (§2.2). This choice is not arbitrary: we show that the Euclidean random-graph model of SANKARARAMAN & BACCELLI [14] falls exactly within our GIBBS framework (§2.3, Prop. 2.2). On this basis we introduce a GIBBS-based notion of *weak recovery* (Def. 2.4) coinciding with the standard definition of weak recovery when the GIBBS assumption fits, as it is the case in the SANKARARAMAN & BACCELLI model.

(ii) Sampler. We design a higher-order SWENDSEN–WANG dynamics (§3) that leaves the GIBBS distribution invariant (Prop. 6.3) yet greatly reduces frustration by freezing the minimum number of *bonds*.

(iii) Evaluation. The notion of GIBBS-weak recovery we introduced quantifies the discrepancy between the output of a proposed community detection algorithm and the GIBBS measure associated with the input weighted graph. We provide necessary conditions for this GIBBS-weak recovery to occur (Prop. 4.2 and Theorem 4.3). Considering the SANKARARAMAN & BACCELLI model, we improve the theoretical bound on the critical threshold for (GIBBS-)weak recovery (§4.2) on the triangular grid. Empirically, our method achieves better performance than state-of-the-art algorithms (§4.3).

Paper organization. The remainder of this paper is structured as follows. §2 is dedicated to the preliminaries, recalling the background of statistical physics, reviewing the SANKARARAMAN & BACCELLI geometric random-graph model that motivates our work, and introducing the notion of GIBBS-weak recovery. In §3, we introduce the higher-order SWENDSEN–WANG-triangle dynamics. §4 presents both theoretical and empirical results: §4.1 establishes a necessary condition for GIBBS-weak recovery, §4.2 proves sharper bounds on the triangular grid using our triangle dynamics, and §4.3 reports experiments on the Euclidean random graph benchmark. Finally, §5 is dedicated to the conclusion and discussion for future research.

2 Preliminaries

2.1 Model definition The *input* in our general community detection framework consists of a simple[1] graph $G = (V, E)$, a list of non-zero edge weights $(W_e)_{e\in E}$, whose signs indicate tendency or aversion to belong to the same community (positive or negative weights), and whose absolute value represents its strength. The number K of communities to recover is given. For the sake of simplicity, throughout this paper, we only consider the case with $K = 2$ communities. An *algorithm* is a map from the set of possible input data to $\{-,+\}^V$, possibly depending on some exogenous randomness.

2.2 Statistical physics: spin configuration, GIBBS energy and GIBBS sampler In the terminology of statistical physics, an estimator of the communities can be viewed as a *spin configuration*

$$\sigma : V \to \{-,+\}.$$

We partition the edge set E into two subsets: the subset F of *ferromagnetic* edges (attractive, with positive weights), and the subset $\widetilde{F}$ of *antiferromagnetic* (repulsive, negative weights) edges. The associated ferromagnetic $(\beta_e)_{e\in F}$ and antiferromagnetic $(\gamma_e)_{e\in\widetilde{F}}$ strengths are the absolute values $|W_e|$ of the weights:

$$\beta_e = W_e \text{ for } e \in F \quad \text{and} \quad \gamma_e = -W_e \text{ for } e \in \widetilde{F}.$$

[1] Undirected, no loops, no multiple edges.

We proceed to define the associated *energy* function or Hamiltonian. The GIBBS energy associated with σ is defined as

$$U(\sigma) = \sum_{e=\{x,y\}\in F} \beta_e \mathbf{1}_{\sigma_x \neq \sigma_y} + \sum_{e=\{x,y\}\in \tilde{F}} \gamma_e \mathbf{1}_{\sigma_x = \sigma_y}.$$

This energy can be interpreted as follows: for each ferromagnetic edge $e \in F$, we add a penalty $\beta_e > 0$ if the configuration σ does not satisfy ferromagnetism, *i.e.*, if the corresponding spins belong to different communities, and similarly we add a penalty γ_e for each antiferromagnetic edge $e \in \tilde{F}$ unsatisfied, *i.e.*, when the corresponding spins belong to the same community.

The GIBBS distribution is the probability measure over the set of configurations $\Sigma := \{-,+\}^V$, which assigns to a configuration $\sigma \in \Sigma$ the probability mass

$$\mu(\sigma) = \frac{1}{Z} e^{-U(\sigma)},$$

where the normalization term Z is called the *partition function*. The lower the energy $U(\sigma)$ of a configuration σ, the more probable that configuration is.

The statistical physics model we use has the triple advantage of being simple, intuitive, and general. Interestingly, the *modularity* of a graph [4], widely used for community detection in graphs [1, 19], can be written as a particular case of a GIBBS energy [13]. Note also that our model generalizes the Constant POTTS Model [18].

2.3 The SANKARARAMAN & BACCELLI generative model To evaluate our approach, we chose a hard instance of the SANKARARAMAN & BACCELLI model [14], the *geometric stochastic block model*, recently investigated by GAUDIO & GUAN [9] in the $\log n$ degree regime. This model serves as a realistic model for haplotype assembly [15].

This case is challenging because it introduces substantial frustration that spreads through the geometric structure of the space. Not clear if there are existing algorithms solving this problem. Say explicitly that there is a version of spectral clustering that works as well as ours, in the sense of overlap, but that it is of order n^3, and that our algorithm outperforms this order. Consequently, many standard algorithms fail on it (*e.g.*, spectral clustering [16] and single-spin-update GLAUBER dynamics [5] fail in our tests).

Given $n \in \mathbb{N}$, $d \in \mathbb{N}$, and $\lambda > 0$. Let $f_{\text{in}}, f_{\text{out}} : \mathbb{R}_+ \to [0,1]$ be two functions, referred to as *connection functions*. Let $\mathcal{X}_\lambda$ be a (homogeneous) POISSON point process on $\mathbb{R}^d$ with intensity λ. Order the points $x_1, x_2, \ldots \in \mathcal{X}_\lambda$ by their ℓ^∞–norm. Next, consider an unweighted, undirected graph $G(\mathcal{X}_\lambda, E)$ defined on $\mathcal{X}_\lambda$ (*e.g.*, G is the geometric graph in the sense of GILBERT-PENROSE [12], or regular grid as in §4.2). What are precise hypotheses over the set of vertices? E.x. a locally set of points.

Each node $x \in \mathcal{X}_\lambda$ is independently assigned to one of the classes $+$ or $-$, with equal probability. Let $\sigma^\star : V \to \{-,+\}$ denote the resulting *Ground-Truth* configuration. Independently sample the potential edges of G to build a subgraph $H \subseteq G$. Specifically, for every pair of nodes $\{x,y\} \in E$, we include the edge $e = \{x,y\}$ in H with probability

$$\begin{array}{ll} f_{\text{in}}(\|x-y\|) & \text{if} \quad \sigma^\star_x = \sigma^\star_y, \\ f_{\text{out}}(\|x-y\|) & \text{if} \quad \sigma^\star_x \neq \sigma^\star_y. \end{array}$$

Throughout we assume the graph is *assortative*, *i.e.*, $f_{\text{in}}(r) \geq f_{\text{out}}(r)$ for all r. Therefore, the presence of an edge $e \in H$ implies ferromagnetism while its absence makes the edge antiferromagnetic. We focus on the n-volume window $B_n = \left[-\left(\frac{n}{2}\right)^{\frac{1}{d}}, \left(\frac{n}{2}\right)^{\frac{1}{d}}\right] \subseteq \mathbb{R}^d$ and set $V^{(n)} = \mathcal{X}_\lambda \cap B_n$. Hence $\left|V^{(n)}\right|$ is a POISSON

random variable with mean $n\lambda$. Let $G^{(n)}$ and $H^{(n)}$ be respectively the graphs G and H restricted to the set of nodes $V^{(n)}$.

We show in the next section that providing $H^{(n)}$ and $G^{(n)}$ with the knowledge of f_{in} and f_{out} enables the Sankararaman & Baccelli model to be subsumed within our framework.

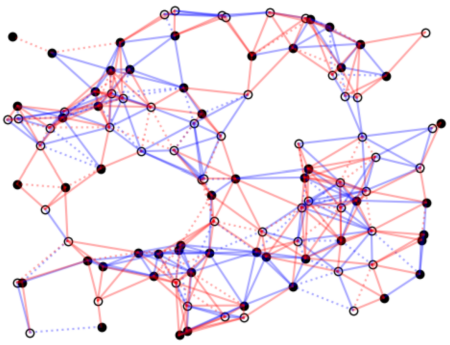

Figure 2.1: A realization of the geometric stochastic block model of Sankararaman & Baccelli ($p_{\text{in}} = \frac{3}{4} = 1 - p_{\text{out}}$). The two Ground-Truth communities are denoted by $\bullet$ and $\circ$. Ferromagnetic edges F are drawn in blue, and antiferromagnetic ones $\widetilde{F}$ in red. Edges in contradiction with Ground-Truth appear as dashed lines (in proportion $\frac{1}{4} = 1 - p_{\text{in}} = p_{\text{out}}$, same as in our experiments §4.3).

A particularly interesting case of the Sankararaman & Baccelli model is the geometric stochastic block model [9]. In this model, G is a (Gilbert-Penrose) geometric graph [12] (*i.e.*, fix a radius r and connect every pair of nodes $x, y \in V$ such that $\|x - y\| \leq r$), and $f_{\text{in}} \equiv p_{\text{in}} \geq p_{\text{out}} \equiv f_{\text{out}}$ do not depend on the distance. This case is of particular interest because a lot of false information is introduced, producing *frustration* and making the community detection problem challenging. See the example of such a graph in Fig. 2.1 with $p_{\text{in}} = \frac{3}{4} = 1 - p_{\text{out}}$ (edges in contradiction with Ground-Truth are dashed).

The Sankararaman & Baccelli model as a particular case We show that the Sankararaman & Baccelli model cast within our framework, and hence it can be used as a generative model to evaluate our method. We consider the parameters d and λ to be fixed throughout. Recall that $\mathcal{X}_\lambda$ denotes the Poisson point process, and that $G(\mathcal{X}_\lambda, E)$ refers to the corresponding random graph. We further assume the knowledge of the connection functions f_{in} and f_{out}.

For the (finite) windows of size n, we look at the (random) joint realization of the Ground-Truth $\sigma^\star : V^{(n)} \to \{-, +\}$ and $H^{(n)}$ (everything else being fixed). Let $\mu_{\text{joint}}(\cdot, \cdot)$ be the associated (finite) joint probability measure. We can explicitly compute μ_{joint}:

$$\mu_{\text{joint}}(\sigma^\star, H^{(n)}) = \frac{1}{2^{|V^{(n)}|}} \prod_{\substack{e=\{x,y\}\in G^{(n)} \\ \sigma_x^\star = \sigma_y^\star}} f_{\text{in}}(\|x-y\|)^{\mathbf{1}\{e\in H^{(n)}\}} \left(1 - f_{\text{in}}(\|x-y\|)\right)^{\mathbf{1}\{e\notin H^{(n)}\}}$$
$$\times \prod_{\substack{e=\{x,y\}\in G^{(n)} \\ \sigma_x^\star \neq \sigma_y^\star}} f_{\text{out}}(\|x-y\|)^{\mathbf{1}\{e\in H^{(n)}\}} \left(1 - f_{\text{out}}(\|x-y\|)\right)^{\mathbf{1}\{e\notin H^{(n)}\}}.$$

$f_{\text{in}}, f_{\text{out}}$ fixed. Assume them assertative by simplicity.

They determine the functions $\beta, \gamma : [0, \infty) \to [0, \infty]$ defined as

$$\beta(r) = -\log\left(\frac{f_{\text{out}}(r)}{f_{\text{in}}(r)}\right)$$
$$\gamma(r) = -\log\left(\frac{1 - f_{\text{in}}(r)}{1 - f_{\text{out}}(r)}\right).$$

For $V \subseteq \mathbb{R}^d$ a finite, fixed, and E a fixed set of edges, we define the following procedure:

- sample σ uniformly in $\{-1, 1\}^V$;
- given σ, sample $F \in 2^E$ using f_{in} and f_{out}, and an independent source of randomness;
- call (X, Y) be a random vector with this distribution.

PROPOSITION 2.1. *(V, E) given as before. For $F \subseteq E$ fixed, we define the Hamiltonian*

$$H_F(\sigma) = \sum_{e=\{x,y\} \in F : \sigma_x \neq \sigma_y} \beta_{|x-y|} + \sum_{e=\{x,y\} \in E \setminus F : \sigma_x = \sigma_y} \gamma_{|x-y|},$$

and let μ_F be the associated Gibbs distribution. Then

$$\mathbb{P}[X = \sigma | Y = F] = \mu_F(\sigma).$$

PROPOSITION 2.2 (Ground-Truth of SANKARARAMAN & BACCELLI is GIBBS). *Conditioned on the realization of $H^{(n)}$ (observation), consider the Ground-Truth $\sigma^\star$ as a random variable. Then the* posterior probability $\mu_{joint}(\sigma^\star, H^{(n)})/\mu_{joint}(\Sigma, H^{(n)})$ *of $\sigma^\star$ given the observation $H^{(n)}$ is exactly $\mu(\sigma^\star)$:*

$$\mu_{joint}(\sigma^\star, H^{(n)})/\mu_{joint}(\Sigma, H^{(n)}) = \mu(\sigma^\star).$$

Here $\mu(\cdot)$ is the GIBBS distribution (see §2.2) associated with the graph $G^{(n)}$ on the vertices $V^{(n)}$, ferromagnetic edges F are the edges present in $H^{(n)}$ and the antiferromagnetic ones $\widetilde{F}$ are those appearing in $G^{(n)}$ but not in $H^{(n)}$. For every interaction $\{x, y\} \in F \cup \widetilde{F}$ with $r = \|x - y\|$, the strengths are given by

$$\beta_r = -\log\left(\tfrac{f_{out}(r)}{f_{in}(r)}\right) \in (0, \infty] \quad \text{if} \quad \{x, y\} \in F,$$
$$\gamma_r = -\log\left(\tfrac{1-f_{in}(r)}{1-f_{out}(r)}\right) \in (0, \infty] \quad \text{if} \quad \{x, y\} \in \widetilde{F}.$$

Proof. See the Appendix, §6.1. ∎

We define algorithm which provides a random guess given a weighted graph. A weighted graph is a triple (V, E, W). Here V is a finite set (not necessarily a subset of $\mathbb{R}^d$), E is a set of edges, and $W : E \to \mathbb{R}$. For a fixed weighted graph (V, E, W), the guessed configuration is a random one drawn using the Gibbs distribution associated to the Hamiltonian

$$H(\sigma) = \sum_{e=\{x,y\} \in E : W_e \geq 0, \sigma_x \neq \sigma_y} W_e - \sum_{e=\{x,y\} \in E : W_e < 0, \sigma_x = \sigma_y} W_e.$$

2.4 Gibbs-weak recovery Our approach focuses on *clustering* rather than *classification*, *i.e.*, we look at the communities up to any permutation on the set of communities $\{-,+\}$. This amounts, for example in the case of the stochastic block model, to assuming that the communities are chosen uniformly. Our measure of *accuracy* will therefore be the *overlap* with Ground-Truth:

Definition 2.3 (Overlap). *Let σ and σ' be two spin configurations. The* overlap *between these two configurations, denoted $\sigma \cap \sigma'$, is given by*

$$\sigma \cap \sigma' := \max_{\pi} \frac{1}{|V|} \big|\{x \in V \mid \sigma_x = \pi(\sigma'_x)\}\big|,$$

where the maximum is taken over the two possible permutations on $\{-,+\}$ (the identity and the transposition).

Note that $\forall \sigma, \sigma' \in \Sigma$, $\sigma \cap \sigma' = \sigma' \cap \sigma \in [\frac{1}{2}, 1]$. Overlap is the maximization over spin-label permutations of the (relative) Hamming similarity.

Inspired by Def. 2.2 of Sankararaman & Baccelli [14], we propose the following definition for Gibbs-*weak recovery* on a family of weighted graphs.

Definition 2.4 (Gibbs-Weak recovery). *Let*

$$\left\{\Lambda^{(n)}\right\}_{n\in\mathbb{N}} = \left\{G^{(n)}\left(V^{(n)}, F^{(n)} \dot\cup \widetilde{F}^{(n)}\right), \left(\beta_e^{(n)}\right)_{e\in F^{(n)}}, \left(\gamma_e^{(n)}\right)_{e\in \widetilde{F}^{(n)}}\right\}_{n\in\mathbb{N}}$$

be a sequence of weighted-graphs such that $\left|V^{(n)}\right| \to \infty$ when $n \to \infty$. We say that the model is Gibbs-weakly recoverable *if there exists a sequence of (potentially random) spin configurations $\tau^{(n)} : V^{(n)} \to \{-,+\}$, $\tau^{(n)}$ depending only on the n-th graph $\Lambda^{(n)}$, and a positive constant $\varepsilon > 0$, such that*

$$\liminf_{n\to\infty} \mathbb{P}_{\sigma^\star \sim \mu^{\Lambda^{(n)}}}\left[\tau^{(n)} \cap \sigma^\star \geq \frac{1}{2} + \frac{1}{2}\varepsilon\right] = 1.$$

Here $\mu^{\Lambda^{(n)}}$ denotes the Gibbs distribution associated with the graph $\Lambda^{(n)}$ and $\sigma^\star$ must be independent of $\tau = \left\{\tau^{(n)}\right\}_{n\in\mathbb{N}}$.

In this definition, $\left\{\tau^{(n)}\right\}_{n\in\mathbb{N}}$ represents the outcomes of the concerning algorithm, and the limit (2.4) means that with high probability the algorithm achieves an overlap of at least $\frac{1}{2} + \frac{1}{2}\varepsilon$. The maximum $\varepsilon^{\max}$ satisfying Eq. 2.4 is the fraction of nodes we can hope to recover with high probability.

Note that standard notion of *weak recovery* in the field of community detection is associated with generative models. Here, we "reverse" the axioms, considering Ground-Truth as a posterior distribution given the observation. Nevertheless, our more general framework fits with some particular generative models, such as the one of Sankararaman & Baccelli (see Prop. 2.2).

Coupling $\sigma^\star$ with well-chosen $\sigma'^\star$ will give us necessary conditions for weak recovery (Prop. 4.2 and Theorem 4.3). For example, we will soon explain that if $(\sigma^\star, \sigma'^\star)$ is the coupling produced by standard Swendsen-Wang dynamics, we recover Theorem 6.6 of Sankararaman & Baccelli on their generative model. But we can choose for the coupling whatever MCMC having Gibbs as stationary distribution. We will show for example that we obtain better bounds on the triangular grid (§4.2) with our higher-order dynamics than Sankararaman & Baccelli.

2.5 Gibbs sampler and Swendsen-Wang dynamics Regarding the Ground-Truth as being drawn from a Gibbs distribution is not an overreach on our part; indeed, we have shown that this assumption matches exactly the generative model of Sankararaman & Baccelli (Prop. 2.2). Having

established this correspondence, the central task becomes the design of an efficient GIBBS sampler. This task is particularly critical because the state space Σ grows exponentially with the size of V. As a result, an algorithm that can efficiently sample from the GIBBS distribution is essential.

Many MARKOV Chain Monte Carlo (MCMC) algorithms have been proposed to do so. Some of them have a local dynamics using *single-spin update*, *e.g.*, GLAUBER or METROPOLIS-HASTINGS dynamics [5]. When vertices are embedded in geometric spaces, as is the case in our work, these local dynamics are blocked in local minima of energy, and their mixing time is usually exponential in the size of the system. Thus, more global-update dynamics are needed in this kind of situation.

The SWENDSEN–WANG dynamics [17] introduces a more general update mechanism that allows for simultaneous updates of entire clusters of spins. At each step, it randomly bonds neighboring spins that share the same state, forming clusters. All spins within a cluster are then flipped collectively with a certain probability. This dynamic has proven to be highly efficient for spin models with an underlying geometric structure, as in the original work in which the nodes were arranged on a two-dimensional grid.

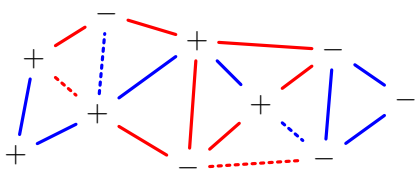

Figure 2.2: Ferromagnetic F and antiferromagnetic edges $\tilde{F}$. Edges unsatisfied under the spin configuration are shown with dashed lines.

The EDWARDS & SOKAL formalism [6] makes it possible to incorporate into the SWENDSEN–WANG dynamics both ferromagnetic and antiferromagnetic interactions, as well as higher-order *bonds* involving more than two nodes.

A single step of the SWENDSEN–WANG dynamics moves the system from a configuration σ to a new configuration σ' in the following way:

1. Keep only the edges $e \in F \cup \tilde{F}$ that are satisfied by the configuration σ and remove the unsatisfied edges (Fig. 2.2, solid and dashed lines, respectively).

2. Apply an independent percolation model to retained edges: an edge e which has been retained in the previous step is "frozen" with probability $p_e = 1 - e^{-w_e}$, where $w_e = \begin{cases} \beta_e & \text{if } e \in F \\ \gamma_e & \text{if } e \in \tilde{F} \end{cases}$.

3. For each cluster C of the resulting frozen graph, flip all its spins or keep all of them unchanged with equal probability.

See Fig. 2.3 for an illustration.

Properties of the SWENDSEN–WANG dynamics The SWENDSEN–WANG dynamics induces a *reversible* MARKOV chain on the set of configurations $\Sigma := \{-,+\}^V$ whose stationary distribution is the GIBBS distribution (Eq. 2.2, see §6.3 and [6]). Let T be the transition matrix on Σ after one step of SWENDSEN-WANG. The two properties mentioned (reversibility and GIBBS as stationary distribution) are direct consequences of the following property verified by T:

$$\forall \sigma, \sigma' \in \Sigma,\ T[\sigma, \sigma'] = e^{-\Delta U} T[\sigma', \sigma] \quad \text{where } \Delta U = U(\sigma') - U(\sigma).$$

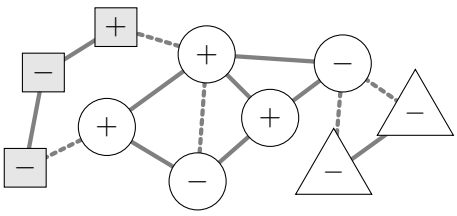

Figure 2.3: Percolation on the satisfied edges of Fig. 2.2. Not retained edges are shown with dashed lines. Three clusters (squares, circles, triangles) are formed. The spins in each cluster are flipped. Compared to Fig. 2.2, the squares are inverted, whereas the other two clusters are left unchanged.

This last property is itself a consequence of independency and accurate choice of freezing mechanisms. When defining our own "Swendsen-Wang" dynamics in §3, we have to check carefully that we recover this fundamental property (see §6.3 in Appendix).

2.6 Frustration and KBD algorithm (higher-order MCMC cluster dynamics) The Swendsen-Wang dynamics is particularly effective in purely ferromagnetic systems. In this case, the correlation between two spins σ_x and σ_y fits with the probability that the two nodes $x, y \in V$ are connected in the Random-Cluster model [7, 11]:

$$\left|\mathbb{E}[\sigma_x \sigma_y]\right| = \mathbb{P}\left[x \leftrightarrow y\right].$$

However, when antiferromagnetism induces contradictory information (*frustration*, see below Def. 2.5), the temperature at which the system exhibits spin correlations (and so communities are recoverable) is not the same at which percolation occurs for the cluster dynamics. In short: too many edges are satisfied, freezing the dynamics.

Definition 2.5 (Frustration). *A model* $(V, F, \widetilde{F})$ *is* frustrated *if no configuration* σ *simultaneously satisfies all edge constraints.*

It is easy to see that we have the following characterization of frustration:

Proposition 2.6 (Characterization of frustration). *A model* $(V, F, \widetilde{F})$ *is* not *frustrated if and only for all pairs of vertices* x, y *and all paths* $\omega \subseteq F \cup \widetilde{F}$ *linking* x *and* y*, the parity of antiferromagnetic edges* $e \in \omega \cap \widetilde{F}$ *in the path* ω does not *depend on the choice of* ω*.*

With this alternative characterisation of frustration, we now understand why Eq. 2.6 is no longer true when the system is frustrated: for a pair $x, y \in V$ of vertices, a path present in the cluster dynamics can bring positive or negative correlation, according to the parity of antiferromagnetic edges inside.

Kandel, Ben-Av & Domany (KBD) [10] proposed a more elaborate solution than standard Swendsen-Wang for the fully frustrated Ising model on the 2D grid, consisting of using higher-order *bonds* (namely "plaquettes"). Their dynamics freezes far fewer bonds yet retaining the same Gibbs energy for each configuration σ. Theoretically speaking, the KBD dynamics has the strong attribute of keeping the property in Eq. 2.6 locally. Precisely, at the scale of a plaquette [2], we have

$$\left|\mathbb{E}_l[\sigma_x \sigma_y]\right| = \mathbb{P}_l\left[x \leftrightarrow y\right],$$

where the integration is made w.r.t. to the Gibbs distribution reduced to a plaquette l (a square of four edges, three ferromagnetic and one antiferromagnetic, in the case of the KBD algorithm). In other words, considering cluster dynamics with correlations on the freezing mechanism can remove the frustration at local scale.

In the next section, we will generalize the ideas of the KBD algorithm to more general graphs.

3 Proposed method We introduce triangle dynamics instead of the classical Swendsen-Wang-edge dynamics. The key idea is to move beyond the edge scale: analyzing triangles eliminates edge-level frustration.

3.1 Energy transfer Edwards & Sokal framework allows us to study dynamics on graphs with multi-edges (see our Prop. 6.3, there is no assumption on the overlap of bonds). Indeed, let $e = \{x, y\} \in F$ be a ferromagnetic (for example) edge with weight β_e. The same system, but replacing e by two parallel ferromagnetic edges e_1 and e_2 between x and y and with respective weights $\beta_{e_1} + \beta_{e_2} = \beta_e$ produces the same Hamiltonian (seen as a function from Σ to $\mathbb{R}$), and consequently the same Gibbs distribution. Moreover, the Swendsen-Wang dynamics are the same: i) on a configuration σ with $\sigma_x \neq \sigma_y$, $\{x, y\}$ is never frozen in both dynamics; ii) on a configuration with $\sigma_x = \sigma_y$, the edge is not frozen in the first dynamics with probability $q_e = e^{-\beta_e}$ and in the second dynamics with probability $q_{e_1} \times q_{e_2} = e^{-\beta_{e_1}} e^{-\beta_{e_2}} = e^{-(\beta_{e_1} + \beta_{e_2})} = q_e$.

Energy transfer consists of distributing an edge's weight equally among all triangles that contain it. Formally, we define $w_e = \begin{cases} \beta_e & \text{if } e \in F \\ \gamma_e & \text{if } e \in \widetilde{F} \end{cases}$. For each edge $e \in F \cup \widetilde{F}$, let T_e be the set of triangles in which e appears. Let $w_{e,t}$ denote the weight transferred from the edge e to triangle t:

$$w_{e,t} := \frac{1}{|T_e|} W_e.$$

Edges which do not appear in any triangles are left unchanged.

In the formalism of Edwards & Sokal, the *bonds* B are now constituted by all the triangles $t \in T$ weighted by their three edges $w_{e_1,t}$, $w_{e_2,t}$ and $w_{e_3,t}$, $e_i \in t$ and the remaining edges not appearing in any triangles

$$B = T \cup \left\{ e \in F \cup \widetilde{F} \mid \nexists t \in T, e \in t \right\}.$$

On these remaining edges, the Swendsen-Wang-edge dynamics remains unchanged.

We now define our cluster dynamics on triangles.

3.2 Cluster dynamics on one triangle

Two categories of triangles Let us first look at the triangles formed by edges of $F \cup \tilde{F}$. There are four possible configurations, which can be divided into two groups:

- the *potentially contradictory* triangles (△ and △), without edge-frustration if we only look at the scale of a triangle;
- the *inherently contradictory* triangles (△ and △), bringing edge-frustration.

The inherently contradictory triangles bring frustration at the scale of edges; this means that even on a system made of only one such triangle, the standard Swendsen-Wang-edges dynamics does not verify the property in Eq. 2.6.

For the definition of the cluster dynamics, two triangles belonging to the same category are essentially the same (flip virtually one spin and the satisfied edges match), we thus only need to consider one representative of each category, the fully ferromagnetic △ and the fully antiferromagnetic △. Moreover, by symmetry to a flip, we can fix the value, say $+$, of one of the spin of the triangle.

Procedure for constructing cluster dynamics on triangles What we call a cluster dynamics on a triangle can be seen as a function which takes as input a weighted triangle t and a spin configuration σ and returns a probability measure on the subsets $\mathcal{P}(t)$ of t, where $\mathcal{P}(t)$ denotes the power set of t. We denote $\mathbb{P}_t^\sigma$ this distribution. $\mathbb{P}_t^\cdot$ can be represented by a $r \times s$ matrix P, r being the number of possible configurations (8 reduced to 4 by symmetry if you fix a spin) and s being $\left|\mathcal{P}(t)\right|$ (8 reduced to

5 because freezing 2 or 3 edges of a triangle leads to the same clustering). The index (i, j) of P contains the probability to freeze $\omega_t^j \subseteq t$ given the configuration σ^i

$$P[i,j] = \mathbb{P}_t^{\sigma^i}\left[\omega_t^j\right].$$

After "repainting" the clusters within the triangle, the cluster dynamics induces a MARKOV chain on the set of spin configurations on the triangle.

Combining the intuition of the authors of KBD [10] and the theoretical property to verify (Eq. 2.6 when the plaquette l is the triangle t, see [2]), we have to choose carefully the cluster dynamics such that:

- The induced MARKOV chain T_t on the triangle verifies the balance equation Eq. 2.5:

  $$T_t[\sigma, \sigma'] = e^{-\Delta U_t} T_t[\sigma', \sigma].$$

  Here, the balance equation is only verified at the scale of a triangle, but thanks to independence of the freezing mechanism, the whole MARKOV chain will also verify it.

- The spin-correlation/cluster property Eq. 2.6 is satisfied for all pairs of spins x and y of the triangle:

  $$\left|\mathbb{E}_t\left[\sigma_x \sigma_y\right]\right| = \mathbb{P}_t\left[x \leftrightarrow y\right].$$

- Among all the possible dynamics, choose the one freezing the fewest possible number of edges.

Cluster dynamics on an isotropic *triangle* We begin with *isotropic* triangles, *i.e.*, triangles whose edges share a common weight w. The cluster dynamics for both categories of triangles can be found in Tab. 3.1.

Table 3.1: Cluster dynamics on a potentially contradictory triangle (Left) and inherently contradictory triangle (Right). We suppose the *isotropic* case: each edge brings the same weight w. Let $p = 1 - e^{-2w}$ and $q = 1 - p = e^{-2w}$. Interpretation of the tables: $P[i,j] = \mathbb{P}\left[\omega^j \text{ frozen} \,\middle|\, \sigma^i\right]$.

w △ w, w	∅	⎯	/	\	△
+ + +	q	0	0	0	p
+ + −	1	0	0	0	0
− + +	1	0	0	0	0
− + −	1	0	0	0	0

w △ w, w	∅	⎯	/	\	△
+ + +	1	0	0	0	0
+ + −	q	$\frac{1}{2}p$	0	$\frac{1}{2}p$	0
− + +	q	0	$\frac{1}{2}p$	$\frac{1}{2}p$	0
− + −	q	$\frac{1}{2}p$	$\frac{1}{2}p$	0	0

Because only two energy levels occur (0 and $2w$ for fully ferromagnetic triangles △, w and $3w$ for fully antiferromagnetic ones △), the corresponding dynamics are straightforward to derive. We now summarize the intuition:

- For the fully ferromagnetic Δ, three states maximize the energy at $2w$, freeze nothing in these "worst" configurations (that is, freeze the empty set $\emptyset$ with probability 1). This strategy ensures, by construction, the minimization of frozen bonds by the dynamics. Then, to satisfy the required properties, the "best" freezing mechanism in the configuration $\overset{+}{+}\,{}_{+}$ is to freeze $\emptyset$ with probability $q = e^{-2w}$ and the full triangle with probability $p = 1 - q$. A potentially contradictory triangle always satisfies at least one edge, so even in the "worst" configurations (maximizing the energy), the classical Swendsen-Wang edges can lead to freeze something. It is not the case with our dynamics.

- For the fully antiferromagnetic Δ, the state $\overset{+}{+}\,{}_{+}$ is maximizing the energy; we therefore freeze $\emptyset$ with probability 1. The only possibility in the other states to attain $\overset{+}{+}\,{}_{+}$ is to freeze the empty set $\emptyset$; the balance equation constrains the other states to have a probability q to freeze $\emptyset$. Remark that freezing only one edge is enough to stay in the "best" low energy states. Consequently, set the probability of freezing the whole triangle to 0; the rest of the table follows directly.

Remark that this dynamics satisfies the spin-correlation/cluster property (Eq. 2.6). For example, for $t = {}_{x}\overset{z}{\Delta}_{y}$, we verify the property on x and y

$$\mathbb{E}_t\left[\sigma_x\sigma_y\right] = \frac{1}{Z}\left(q - 1 + 1 - 1\right) = \frac{q-1}{Z} = -\frac{p}{Z} \quad \text{with } Z = q + 1 + 1 + 1$$

and

$$\mathbb{P}_t\left[x \leftrightarrow y\right] = \mathbb{P}_t\left[\_ \text{ frozen }\right] = \frac{1}{Z}\left(q \times 0 + 1 \times \tfrac{1}{2}p + 1 \times 0 + 1 \times \tfrac{1}{2}p\right) = \frac{p}{Z} = \left|\mathbb{E}_t\left[\sigma_x\sigma_y\right]\right|.$$

Cluster dynamics on an anisotropic *triangle*

- Let ${}^{w_3}_{w_2}\Delta^{w_1}$ be an *anisotropic* fully ferromagnetic triangle. We suppose without loss of generality that $0 \le w_1 \le w_2 \le w_3$. By the same method as in the isotropic case, we obtain Tab. 6.1. Note that ${}^{w_3}_{w_2}\Delta^{w_1} = {}^{w_1}_{w_1}\Delta^{w_1} + {}^{w_3-w_1}_{w_2-w_1}\Delta^{0}$. The dynamics given in Tab. 6.1 is exactly the same as if we superposed the dynamics of an isotropic triangle weighted by w_1 and the classical Swendsen-Wang-edges dynamics on the remaining two edges weighted by $w_2 - w_1$ and $w_3 - w_1$.

- The ${}^{w_3}_{w_2}\Delta^{w_1}$ *anisotropic* fully antiferromagnetic case is slightly more complicated. The correct way is to decompose as in the ferromagnetic case ${}^{w_3}_{w_2}\Delta^{w_1} = {}^{w_1}_{w_1}\Delta^{w_1} + {}^{w_3-w_1}_{w_2-w_1}\Delta^{0}$. As previously, the (unique) interesting dynamics is the superposition of two dynamics, the isotropic antiferromagnetic triangle ${}^{w_1}_{w_1}\Delta^{w_1}$ with weight w_1 and the anisotropic ${}^{w_3-w_1}_{w_2-w_1}\Delta^{0}$. But now, the appropriate dynamics on ${}^{w_3-w_1}_{w_2-w_1}\Delta^{0}$ is not the Swendsen-Wang-edges. Depending on whether σ_y and σ_z are positively or negatively correlated, the dynamics differs. See Tab. 6.2 for the dynamics on ${}^{w_3-w_1}_{w_2-w_1}\Delta^{0}$.

Again, the reader can check that our dynamics satisfies the spin-correlation/cluster property (Eq. 2.6) also on the anisotropic triangles.

3.3 One step of Swendsen-Wang-triangles After transferring the energy (see §3.1), the proposed dynamics is similar to the classical Swendsen–Wang algorithm (§2.5), except that we now consider *bonds* $b \in B = T \cup F \cup \tilde{F}$, *i.e.*, either triangles $t \in T$ or edges $e \in F \cup \tilde{F}$.

The Swendsen-Wang triangle dynamics can be synthesized as follows:

1. Starting from a configuration $\sigma : V \to \{-, +\}$, apply an independent percolation model, that is for all $b \in B$ toss independently a frozen cluster $\omega_b \sim \mathbb{P}_b^\sigma$ according $\mathbb{P}_b^\sigma$. (If b is an edge, apply

the classical SWENDSEN-WANG-edges on it; if b is a triangle, freeze the cluster according to the dynamics described in the previous section §3.2.) We obtain the resulting clustering

$$\kappa = \bigcup_{b \in B} \omega_b \subseteq E.$$

2. For each cluster $C \subseteq V$ that is formed by κ, flip the spins with probability $\frac{1}{2}$, and keep them unchanged with probability $\frac{1}{2}$.

We denote $\mathbb{P}^{\sigma \to \kappa}$ the probability that the clustering-dynamics in the state σ results in the clustering κ and denote

$$\mathbb{P}^{\kappa} = \sum_{\sigma \in \Sigma} \mu(\sigma) \mathbb{P}^{\sigma \to \kappa}$$

the marginal on κ.

3.4 Our complete algorithm Our complete algorithm can be summarized in four steps. The first two steps are the energy transfer and SWENDSEN-WANG-triangles dynamics, the heart of our research and were described in the previous sections. The last two steps are more of a "recipe" we propose to finish the clustering algorithm. Depending on the constraints (*e.g.*, memory, computational capabilities), they could be adapted. Here are the four steps:

1. Energy transfer: for each edge appearing in at least one triangle, split its weight into the triangles that share it and "transfer" those weights to the triangles.

2. Start from a random configuration σ^0 and run the SWENDSEN-WANG-triangles dynamics several times, storing the community labels to estimate spin-spin correlations later.

3. Estimate spin-spin correlations – or, equivalently, the probability for two vertices $x, y \in V$ to be in the same community. Because the MCMC with SWENDSEN-WANG-triangles has the GIBBS measure as its stationary distribution. The empirical probability we compute is an estimator of

$$\mathbb{P}[\sigma_x = \sigma_y]$$

where the probability measure is the GIBBS one.

4. We interpret the $\left|V\right| \times \left|V\right|$ matrix we obtain as a complete, positively weighted and symmetric graph on which we can apply standard algorithms, *e.g.*, a spectral clustering or LEIDEN/LOUVAIN algorithms [1, 19].

Remark on K The method we propose works only with $K = 2$ communities. Adapting it for $K = 3, 4, \ldots$ is possible but would require to look at structures with $K+1$ vertices, making the complexity explode exponentially with K. Nevertheless, our methodology remains tractable for reasonable values of K.

4 Results

4.1 Theoretical results

4.1.1 Necessary condition for Gibbs-weak recovery Recall Def. 2.4 that the system is said to be *Gibbs-weakly* (or *partially*) recoverable when $\varepsilon^{\max}$ the maximal ε satisfying

$$\liminf_{n\to\infty} \mathbb{P}_{\sigma^\star\sim\mu^{\Lambda^{(n)}}} \left[\tau^{(n)} \cap \sigma^\star \geq \tfrac{1}{2} + \tfrac{1}{2}\varepsilon\right] = 1$$

is strictly positive $\varepsilon^{\max} > 0$. ($\sigma^\star$ is taken independent of τ and $\tau^{(n)}$ is a (possibly random) function depending only on $\Lambda^{(n)}$.)

Compared to the definition of Sankararaman & Baccelli, the Ground-Truth $\sigma^\star$ is seen as a random variable (tossed with respect to Gibbs $\mu^{\Lambda^{(n)}}$) and the algorithm τ is completely agnostic on the realization of the edges 'in' or 'out'.

Coupling $(\sigma^\star, \sigma'^\star)$ Applying a coupling $(\sigma^\star, \sigma'^\star)$ where $\sigma'^\star = \text{S.-W.}(\sigma^\star)$ is one step of Swendsen-Wang dynamics applied on σ, we can show that *percolation* on the random cluster κ is a necessary condition for weak recovery. In fact, if we make the coupling with the standard Swendsen-Wang-edges dynamics, we recover Theorem 6.6 of Sankararaman & Baccelli. We show in the next section that applying our Swendsen-Wang-triangles dynamics improves the theoretical bounds of their theorem on the stochastic block model on the triangular-grid.

Let us first define what is *percolation* in our framework.

Definition 4.1 (Percolation). *Same hypothesis as in Def. 2.4 for the system* $\left\{\Lambda^{(n)}\right\}_{n\in\mathbb{N}}$. *Recall that* $\mathbb{P}^\kappa$ *denotes the probability distribution on the random cluster* κ *of the dynamics when assuming that the configuration* σ *was sampled according Gibbs* $\mu^{\Lambda^{(n)}}$. *Denote* $\mathcal{C}(\kappa) = \left\{C_1, C_2, \ldots\right\}$ *the resulting clustering on the nodes* V. *(For the sake of clarity, we sometimes skip the dependence on* n.*) Let* $\varepsilon > 0$ *and denote* $\theta^\varepsilon(\kappa)$ *the proportion of points of* V *lying in a cluster of size at least* $\varepsilon\left|V\right|$, *that is*

$$\theta^\varepsilon(\kappa) = \frac{1}{\left|V\right|} \sum_{C\in\mathcal{C}(\kappa)} \left|C\right| \mathbf{1}_{\left|C\right|\geq\varepsilon\left|V\right|}.$$

(Remark that we do not make any assumption about the uniqueness *of "giant" clusters.) Let* $\theta \geq 0$ *and denote*

$$\eta(\theta) = \lim_{\varepsilon\to 0^+} \limsup_{n\to\infty} \mathbb{P}^\kappa \left[\theta^\varepsilon(\kappa) \geq \theta\right].$$

$\eta(\cdot)$ *is a non-increasing function and* $\eta(\theta) > 0$ *means that there is (asymptotically) a non-null probability for giant cluster(s) to represent at least a proportion* θ *of the points. We call* percolation probability $\theta^{\max}$ *the quantity*

$$\theta^{\max} = \inf\left\{\theta \in [0,1] \;\middle|\; \eta(\theta) = 0\right\}.$$

When $\theta^{\max} = 0$, *we say that the system* $\left\{\Lambda^{(n)}\right\}_{n\in\mathbb{N}}$ does not *percolate.*

Note that our definition applied for example to the generative model of Sankararaman & Baccelli fits with the usual definition of *percolation probability.*

Percolation is a necessary condition for Gibbs-weak recovery:

Proposition 4.2 (Percolation necessary for Gibbs-weak recovery). *Same hypothesis as in Def. 2.4 for the system* $\left\{\Lambda^{(n)}\right\}_{n\in\mathbb{N}}$. *Let* $\theta^{\max}$ *be its percolation probability. If the system does not percolate (*$\theta^{\max} = 0$*), then the system is not Gibbs-weakly recoverable.*

Proof. The proof is a consequence of the following theorem. ∎

Theorem 4.3 (Upper bound on $\varepsilon^{\max}$ in the Gibbs-weak recovery). *Same hypothesis as the previous proposition. Let* $\varepsilon^{\max}$ *be the maximum of the* ε *satisfying the hypothesis of Gibbs-weak recovery:*

$$\varepsilon^{\max} = \max_{\varepsilon\in[0,1]} \liminf_{n\to\infty} \mathbb{P}_{\sigma^\star\sim\mu^{\Lambda^{(n)}}} \left[\tau^{(n)} \cap \sigma^\star \geq \frac{1}{2} + \frac{1}{2}\varepsilon\right] = 1.$$

Then

$$\varepsilon^{\max} \le \theta^{\max}.$$

Proof. See the Appendix. ∎

4.2 Swendsen-Wang-triangles > Swendsen-Wang-edges on the triangular grid Let us first show why Theorem 6.6 of Sankararaman & Baccelli is a particular case of our Theorem 4.3 with the Swendsen-Wang-edges dynamics.

We take the same notations as in §2.3. Let $\sigma^\star \sim$ Gibbs be a sample of the Ground-Truth. We use Theorem 4.3 with a coupling $(\sigma^\star, \sigma'^\star)$ where $\sigma'^\star =$ S.-W.$(\sigma^\star)$ is one step of Swendsen-Wang-edges of $\sigma^\star$. An edge $x, y \in E$ such that $\sigma^\star_x = \sigma^\star_y$ appears in F with probability $f_{\text{in}}(r)$. It is then retained in the percolation step with probability $1 - e^{-\beta_r} = \frac{f_{\text{in}}(r) - f_{\text{out}}(r)}{f_{\text{in}}(r)}$.

Hence the probability that the edge survives to the random-clustering phase equals $f_{\text{in}}(r) - f_{\text{out}}(r)$; the very same probability holds for the antiferromagnetic edges of $\tilde{F}$.

This situation is equivalent to observing bond percolation on the random graph $H_{G,\, f_{\text{in}} - f_{\text{out}}}$, whose vertex set is $\mathcal{X}_\lambda$ and whose edges are a subset of those of G, each included independently with probability $f_{\text{in}} - f_{\text{out}}$. The Ground-Truth is not Gibbs-weakly recoverable whenever $H_{G, f_{\text{in}} - f_{\text{out}}}$ does *not* percolate. In this model, if $H_{\mathcal{X}_\lambda, f_{\text{in}} - f_{\text{out}}}$ percolates, the giant component is unique and contains a positive fraction of the points, denoted $\theta\big(H_{\mathcal{X}_\lambda,\, f_{\text{in}} - f_{\text{out}}}\big)$. This θ is exactly our $\theta^{\max}$ (Def. 4.1). Moreover, in this generative model, our notion of Gibbs-weak recovery matches with the usual one, and we recover easily Theorem 6.6 of Sankararaman & Baccelli.

The previous results relied on a coupling with the "classical" Swendsen-Wang dynamics. Exactly the same argument works for *any* transformation that preserves the Gibbs law. We shall therefore invoke our Swendsen-Wang-triangles dynamics to prove that we obtain better bounds for weak recovery on the triangular grid.

Let $G(\mathcal{X}, E)$ be the triangular grid in the plane $\mathbb{R}^2$. Fix $p \in [\frac{1}{2}, 1)$ and put, for all $r \ge 0$,

$$f_{\text{in}}(r) = p, \qquad f_{\text{out}}(r) = 1 - p, \qquad \beta_r = \gamma_r = -\log\Big(\tfrac{1-p}{p}\Big).$$

By symmetry, each edge $e = \{x, y\} \in E$ agrees with $\sigma^\star$ with probability p (ferromagnetic if $\sigma^\star_x = \sigma^\star_y$ or antiferromagnetic if $\sigma^\star_x \ne \sigma^\star_y$). We simply write $e \sim \sigma^\star$. Conditioned on $\sigma^\star$ the indicators $\mathbf{1}_{e \sim \sigma^\star}$ are i.i.d. $\mathcal{B}$ernoulli(p).

4.2.1 Swendsen–Wang edges Bond percolation is performed on the agreeing edges with retention factor

$$1 - e^{-\beta} = 1 - e^{-\gamma} = \tfrac{2p-1}{p},$$

so that, given $\sigma^\star$, the random-cluster measure is independent bond percolation on the triangular lattice with parameter

$$p_{\text{bond}} = p \cdot \tfrac{2p-1}{p} = 2p - 1 = 2\varepsilon.$$

With $\varepsilon = \frac{1}{2}(p_{\text{in}} - p_{\text{out}}) = p - \frac{1}{2}$, planar duality yields the critical value

$$p_c^{\text{edge}} = \tfrac{1}{2} + \sin\big(\tfrac{\pi}{18}\big) = 0.674\ldots$$

Hence, for $p \in \big[\frac{1}{2},\, p_c^{\text{edge}}\big)$, the Swendsen-Wang-edges dynamics fails to percolate on $\sigma^\star$ and the model is *not (Gibbs-)weakly recoverable.* No conclusion can be drawn for $p \ge p_c^{\text{edge}}$.

4.2.2 Swendsen-Wang-triangles

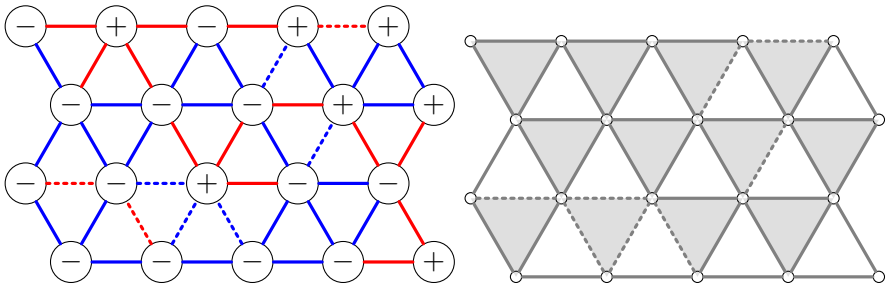

Figure 4.1: Sankararaman & Baccelli model on the triangular grid. (Left) Ferromagnetic F and antiferromagnetic $\widetilde{F}$ edges are represented. Unsatisfied edges by $\sigma^\star$ are dashed. (Right) The triangles considered for Swendsen-Wang-triangles dynamics (the whites).

Triangle dynamics.

- We use only the white triangles (Fig. 4.1); each edge $e \in F \cup \tilde{F}$ transfers its energy to its unique adjacent white triangle; boundary effects are negligible.
- In the isotropic case, each triangle inherits a common edge weight w with
$$w = \beta = \gamma = -\log\Big(\tfrac{1-p}{p}\Big).$$
  - A *potentially contradictory* triangle in its (up to a flip) unique ground state (minimizing energy) is frozen entirely with probability $1 - e^{-2\omega}$; otherwise it is left unfrozen.
  - An *inherently contradictory* triangle in one of its three ground states (two edges satisfied) freezes exactly one of its two satisfied edges with the same probability; each edge is thus frozen, in a *correlated* way, with probability $\frac{1}{2}(1 - e^{-2\omega})$.
- The resulting clusters are flipped independently with probability $\frac{1}{2}$.

A single Swendsen-Wang-triangles step from $\sigma^\star$. We now ignore the sign of an edge and only record whether it matches $\sigma^\star$: solid $-$ if $e \sim \sigma^\star$, dashed $\cdots$ otherwise. Given $\sigma^\star$, edges match independently according to Bernoulli(p).

The four triangle types are therefore

- Potentially contradictory: △ and ⩚ with probabilities p^3 and $3p(1-p)^2$. Only the first may be frozen, with probability $1 - e^{-2\omega} = \frac{2p-1}{p^2}$.
- Inherently contradictory: ⩚ and ∴ with probabilities $3p^2(1-p)$ and $(1-p)^3$. Only the first can have one of its edges frozen with probability $1 - e^{-2\omega} = \frac{2p-1}{p^2}$.

Because the triangles are independent and isotropic with positive edge-correlations, the critical threshold occurs when the probability of a full triangle equals that of an empty triangle [3]; solving yields

$$p_c^{\text{triangle}} = \frac{7}{4} - \frac{\sqrt{17}}{4} = 0.7192\ldots$$

Thus, for

$$p < p_c^{\text{triangle}} \quad \big(\text{notably } p \in [p_c^{\text{edge}},\, p_c^{\text{triangle}})\big),$$

the Swendsen-Wang-triangles dynamics fails to percolate on $\sigma^\star$ and we know that the system is *not* (Gibbs-)weakly recoverable (while we cannot conclude with Swendsen-Wang-edges or Theorem 6.6 of Sankararaman & Baccelli for $p \in [p_c^{\text{edge}},\, p_c^{\text{triangle}})$).

4.3 Empirical results We implemented a version of our algorithm for *isotropic* triangles[2].

We work on a hard instance of the Sankararaman & Baccelli [14] geometric stochastic block model (see §2.3) with parameter $p = \frac{3}{4}$, $p_{\text{in}} = p$ and $p_{\text{out}} = 1 - p$. We fixed the other parameters to $d = 2$, $n = 200$, $\lambda = 5$ (making $\left|V\right| \approx 1000$). The radius is chosen such that each node has, on average, 20 neighbors. To skip boundary effects, we take for distance the torus metric.

$p = \frac{3}{4}$ means that one edge out of four is in contradiction with the Ground-Truth $\sigma^\star$; the model exhibits huge frustration (see Fig. 2.1) and several classical algorithms we tested like spectral clustering [16] or single-spin-update Glauber dynamics [5] do not work at all.

For the final step of our algorithm, we ran the Leiden algorithm coupled with Constant Potts Model [19, 18]. We compare our algorithm with: i) the same algorithm but with the classical Swendsen-Wang-edges dynamics (in both algorithms, 1000 iterations of the dynamics were performed); ii) Leiden + Constant Potts Model directly on the input graph.

In Fig. 4.2, the distribution (over 1000 realizations) of the accuracy (*i.e.*, overlap with Ground-Truth, Def. 2.3) is shown for these three algorithms.

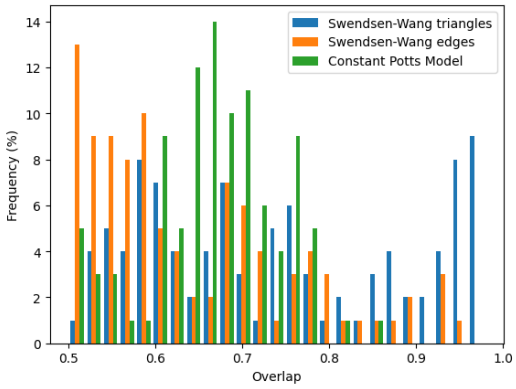

Figure 4.2: Overlap with Ground-Truth distribution (over 100 realizations) resulting from the three algorithms: our Swendsen-Wang-triangles algorithm, the Swendsen-Wang edges version and the Leiden + Constant Potts Model applied to the input graph.

Our algorithm Swendsen-Wang-triangles is the only one to achieve a very high accuracy ($\geq 90\%$) reasonably often ($\approx 25\%$ of the realizations).

5 Conclusion and discussion We have proposed a framework for community detection on weighted graphs based on the associated Gibbs distribution. We defined our own notion of Gibbs-weak recovery (Def. 2.4). We showed that coupling the Gibbs with cluster dynamics, percolation is a necessary condition for Gibbs-weak recovery (Prop. 4.2 and Theorem 4.3).

The question of cluster dynamics thus arises: we have devised a higher-order Swendsen–Wang dynamics. Defined on triangular plaquettes, this dynamics is optimal with respect to our theoretical requirements. Specifically: i) it induces a reversible Markov chain w.r.t. Gibbs (§6.3) ii) it satisfies Eq. 2.6, which links spin correlations to the probability that two spins belong to the same cluster and iii) it minimizes the number of frozen bonds.

Our definition of Gibbs-weak recovery fits with standard definition on the generative model of Sankararaman & Baccelli (Prop. 2.2). We can readily recover their results and even generalize them. Their results corresponding to the particular case of the classical Swendsen-Wang dynamics in

[2]In this case, the weight transfer is made solving a linear constraint problem. Also a relaxation factor $U(\sigma) \leftarrow \frac{1}{4}U(\sigma)$ is applied on the Hamiltonian.

our framework; we apply ours results with the SWENDSEN-WANG-triangles we developed (§3), improving theoretical bound for (GIBBS-)weak recovery on the triangular-grid geometric stochastic block model (§4.2).

On the challenging geometric stochastic block model benchmark of SANKARARAMAN & BACCELLI, our method consistently outperforms both i) a straight edge-based SWENDSEN-WANG variant and ii) the Constant POTTS Model + LEIDEN baseline.

We hope that the ideas presented here will stimulate further research at the intersection of statistical physics, sampling algorithms, and graph inference.

Future directions We propose two research directions that we regard as particularly important.

- Extending the construction to $K > 2$ communities. One promising direction is to employ simplicial complexes of order greater than two. A general procedure that, given K as an input parameter, automatically produces the corresponding dynamics would be highly desirable.

- Developing scalable alternatives for large K. At present, the cluster-repainting phase selects a permutation of the K communities, yielding a factorial cost of $K!$. Alternative strategies with lower computational complexity are therefore needed when K is large.

6 Appendix

6.1 Proof of Prop. 2.2

We can write in an exponential way:

$$\mu_{\text{joint}}(\sigma^\star, H) = \frac{1}{2^{|V|}} \exp\left[\sum_{\{x,y\}\in F} \log f_{\text{in}}(\|x-y\|)\mathbf{1}_{\sigma_x=\sigma_y} + \log f_{\text{out}}(\|x-y\|)\mathbf{1}_{\sigma_x\neq\sigma_y}\right]$$
$$\exp\left[\sum_{\{x,y\}\in \widetilde{F}} \log(1-f_{\text{in}}(\|x-y\|))\mathbf{1}_{\sigma_x=\sigma_y} + \log(1-f_{\text{out}}(\|x-y\|))\mathbf{1}_{\sigma_x\neq\sigma_y}\right]$$

Multiply by

$$2^{|V|}\exp\left[-\sum_{\{x,y\}\in F}\log f_{\text{out}}(\|x-y\|) - \sum_{\{x,y\}\in\widetilde{F}}\log(1-f_{\text{in}}(\|x-y\|))\right]$$

and you obtain

$$\mu_{\text{joint}}(\sigma^\star, H) \propto \exp\left[\sum_{\{x,y\}\in F}\log\frac{f_{\text{in}}(\|x-y\|)}{f_{\text{out}}(\|x-y\|)}\mathbf{1}_{\sigma_x=\sigma_y} + \sum_{\{x,y\}\in\widetilde{F}}\log\frac{1-f_{\text{out}}(\|x-y\|)}{1-f_{\text{in}}(\|x-y\|)}\mathbf{1}_{\sigma_x\neq\sigma_y}\right]$$
$$\propto \exp\left[-\sum_{e=\{x,y\}\in F}\beta_e\mathbf{1}_{\sigma_x=\sigma_y} - \sum_{e=\{x,y\}\in\widetilde{F}}\gamma_e\mathbf{1}_{\sigma_x\neq\sigma_y}\right] = \exp\left[-U(\sigma^\star)\right].$$

6.2 Cluster dynamics on general triangles (anisotropic case) The cluster dynamics on anisotropic triangles is presented in Tab. 6.1 and Tab. 6.2. For the antiferromagnetic triangle, only the case with an edge with zero weight is presented; the general dynamics can be recovered superposing two dynamics: ${}^{w_3}_{w_2}\triangle^{w_1} = {}^{w_1}_{w_1}\triangle^{w_1} + {}^{w_3-w_1}_{w_2-w_1}\triangle^{0}$ one on an isotropic antiferromagnetic triangle plus the remaining weights on an anisotropic antiferromagnetic triangle.

Table 6.1: Dynamics on the ferromagnetic anisotropic triangle. Suppose $\omega_1 \le \omega_2 \le \omega_3$.

w_3 △ w_1 w_2	∅	—	/	\	△
+ + +	$e^{-(\omega_3+\omega_2)}$	$e^{-(\omega_1+\omega_3)} - e^{-(\omega_2+\omega_3)}$	$e^{-(\omega_2+\omega_1)} - e^{-(\omega_2+\omega_3)}$	0	$1 - e^{-(\omega_1+\omega_3)} - e^{-(\omega_2+\omega_1)} + e^{-(\omega_3+\omega_2)}$
+ + −	$e^{\omega_1-\omega_3}$	0	$1 - e^{\omega_1-\omega_3}$	0	0
− + +	$e^{\omega_1-\omega_2}$	$1 - e^{\omega_1-\omega_2}$	0	0	0
− + −	1	0	0	0	0

Table 6.2: Dynamics on the anisotropic antiferromagnetic triangle. Let $w' = w_2 - w_1$, $w'' = w_3 - w_1$ and $\mu = 1 + e^{-(w'+w'')} - e^{-w'} - e^{-w''} \propto \mathbb{E}_t[\sigma_y \sigma_z]$. (Top) Dynamics when $\mu \ge 0$. (Bottom) Dynamics when $\mu < 0$.

$w_3 - w_1$ △ 0 $w_2 - w_1$	∅	—	/	\	△
+ + +	1	0	0	0	0
+ + −	$e^{-w'}$	$1 - e^{-w'}$	0	0	0
− + +	$e^{-w''}$	0	$1 - e^{-w''}$	0	0
− + −	$e^{-(w'+w'')}$	$e^{-w''}(1 - e^{-w'})$	$e^{-w'}(1 - e^{-w''})$	0	μ

$w_3 - w_1$ △ 0 $w_2 - w_1$	∅	—	/	\	△
+ + +	1	0	0	0	0
+ + −	$e^{-w'}$	$1 - e^{-w'} + \frac{1}{2}\mu e^{w''}$	0	$-\frac{1}{2}\mu e^{w''}$	0
− + +	$e^{-w''}$	0	$1 - e^{-w''} + \frac{1}{2}\mu e^{w'}$	$-\frac{1}{2}\mu e^{w'}$	0
− + −	$e^{-(w'+w'')}$	$e^{-w''}(1 - e^{-w'}) + \frac{1}{2}\mu$	$e^{-w'}(1 - e^{-w''}) + \frac{1}{2}\mu$	0	0

6.3 Swendsen-Wang dynamics induces reversible Markov chain We consider here a general family of "Swendsen-Wang" dynamics. Consider a graph $G(V, E)$ and a set B of *bonds*, each bond $b \in B$ being a subset of E. Each bond is associated to an energy function $U_b : \Sigma \to \mathbb{R}$. The cluster

dynamics is constructed in terms of per-bond operations, which depend on a function which takes a bond b and a configuration σ and returns a probability distribution $\mathbb{P}_b^\sigma$ on $\mathcal{P}(b)$. For $b \in B$ and $\omega_b \subseteq b$, we define the "compatibility" between σ and σ' as

$$P_b^{\sigma\to\sigma'}[\omega_b] = \begin{cases} \mathbb{P}_b^\sigma[\omega_b] & \text{if } \mathbb{P}_b^{\sigma'}[\omega_b] > 0 \\ 0 & \text{otherwise} \end{cases}.$$

Definition 6.1. *We say that the Swendsen-Wang dynamics satisfies the "cluster-balance" equation if*

$$P_b^{\sigma\to\sigma'}[\omega_b] = e^{-\Delta U_b} P_b^{\sigma'\to\sigma}[\omega_b]$$

for every $b \in B$, $\omega_b \subseteq b$ and $\sigma, \sigma' \in \Sigma$. Here $\Delta U_b = U_b(\sigma') - U_b(\sigma)$ is the difference of energy on the bond b.

Denote by Ω the set of all possible bond-clusterings, that is, $\Omega = \prod_{b\in B} \mathcal{P}(b)$. If $\omega \in \Omega$ is the result of the freezing mechanism of the Swendsen-Wang dynamics, then the resulting clustering κ on E is

$$\kappa = \bigcup_{b\in B} \omega_b.$$

Once κ is obtained, the final step of the Swendsen-Wang dynamics consists in flipping or keeping the spins in each cluster formed by κ independently with probability $\frac{1}{2}$.

One step of the Swendsen-Wang dynamics induces a Markov chain on the set of configurations Σ. The following result shows that this Markov chain has Gibbs for stationary distribution and is reversible.

Proposition 6.2. *Let $U : \Sigma \to \mathbb{R}$ be the total energy function defined as $U = \sum_{b\in B} U_b$. Assume that the cluster dynamics satisfy the cluster-balance Equation (6.1), and that, for all $b \in B$ and $\sigma \in \Sigma$, one has $\mathbb{P}_b^\sigma[\emptyset] > 0$. Then the Markov chain induced by the Swendsen–Wang dynamics is irreducible and reversible with respect to the Gibbs distribution associated to U, which is therefore its stationary distribution.*

Proof. Denote by T the Markov transition matrix induced by the Swendsen-Wang dynamics. Let $\sigma, \sigma' \in \Sigma$ be two spin configurations. From σ, there is a non-null probability that the cluster dynamics freezes nothing, and thus a positive probability of reaching σ', which implies that $T[\sigma, \sigma'] > 0$. The Markov chain is therefore irreducible. Moreover,

$$\begin{aligned} T[\sigma,\sigma'] &= \sum_{\kappa\subseteq E} \frac{1}{2^{\mathtt{Clust}(\kappa)}} \sum_{\substack{\omega\in\Omega \\ \kappa=\bigcup_{b\in B}\omega_b}} \prod_{b\in B} P_b^{\sigma\to\sigma'}[\omega_b] \\ &= \sum_{\kappa\subseteq E} \frac{1}{2^{\mathtt{Clust}(\kappa)}} \sum_{\substack{\omega\in\Omega \\ \kappa=\bigcup_{b\in B}\omega_b}} \prod_{b\in B} e^{-\Delta U_b} P_b^{\sigma'\to\sigma}[\omega_b] \\ &= e^{-\Delta U} \sum_{\kappa\subseteq E} \frac{1}{2^{\mathtt{Clust}(\kappa)}} \sum_{\substack{\omega\in\Omega \\ \kappa=\bigcup_{b\in B}\omega_b}} \prod_{b\in B} P_b^{\sigma'\to\sigma}[\omega_b] \\ &= e^{-\Delta U} T[\sigma',\sigma]. \end{aligned}$$

$\mathtt{Clust}(\kappa)$ denotes the number of connected components induced by κ (each isolated vertex counts for one component). Observe that if $\omega \in \Omega$ is incompatible either with σ or with σ', then $\prod_{b\in B} P_b^{\sigma\to\sigma'}[\omega_b] = 0$,

and hence such ω does not contribute to the decomposition. The proof finishes by noting that $e^{-\Delta U} = e^{U(\sigma)}e^{-U(\sigma')} = \dfrac{\mu(\sigma')}{\mu(\sigma)}$. ∎

6.4 Proof of Theorem 4.3 Let $\theta > \theta^{\max}$ be arbitrary close to $\theta^{\max}$. We have that $\eta(\theta) = 0$ which means that, with high probability, at least a proportion $1-\theta$ of the points in $V^{(n)}$ fall in clusters $C \in \mathcal{C}(\kappa^{(n)})$ formed by $\kappa^{(n)} \sim \mathbb{P}^{\kappa^{(n)}}$ of size negligible compared to $\left|V^{(n)}\right|$. Note that for each cluster $C \in \mathcal{C}(\kappa^{(n)})$, there are only two spin configurations on C due to edge constraints, say σ_C and $\overline{\sigma_C}$ with $\sigma_C = -\overline{\sigma_C}$. Note also that we can sample $\sigma^\star \sim \mu^{\Lambda^{(n)}}$ by tossing independently a $\mathcal{B}$ernoulli$(\frac{1}{2})$ for each cluster $C \in C(\kappa^{(n)})$ and choosing the configuration σ_C or $\overline{\sigma_C}$ according to the result of the Bernoulli. Denote $\sigma^\star_C$ this result. Let $\tau^{(n)}$ be the resulting configuration of the algorithm. Remark that the proportion X_C of common spins between $\sigma^\star_C$ and $\tau^{(n)}_C$

$$X_C := \frac{1}{\left|C\right|}\sigma^\star_C \cdot \tau^{(n)}_C = \frac{1}{\left|C\right|}\left|\left\{x \in C \mid \sigma^\star_C(x) = \tau^{(n)}_C(x)\right\}\right|$$

is a random variable taking its value in $[0,1]$ and with mean $\frac{1}{2}$. (Because $\frac{1}{2}\frac{1}{\left|C\right|}\sigma_C \cdot \tau^{(n)}_C + \frac{1}{2}\frac{1}{\left|C\right|}\overline{\sigma_C} \cdot \tau^{(n)}_C = \frac{1}{2}\frac{\left|C\right|}{\left|C\right|} = \frac{1}{2}$.) Finally,

$$\frac{1}{\left|V^{(n)}\right|}\sigma^\star \cdot \tau^{(n)} = \sum_{C \in \mathcal{C}(\kappa^{(n)})} \frac{\left|C\right|}{\left|V^{(n)}\right|} X_C.$$

We now split this sum into two parts: let $\mathcal{C}^1$ be the set of clusters of size negligible, we write:

$$\sum_{C \in \mathcal{C}} \frac{\left|C\right|}{\left|V\right|} X_C = \sum_{C \in \mathcal{C}^1} \frac{\left|C\right|}{\left|V\right|} X_C + \sum_{C \in \mathcal{C} \setminus \mathcal{C}^1} \frac{\left|C\right|}{\left|V\right|} X_C.$$

The second term can be bounded between 0 and $1 - \sum_{C \in \mathcal{C}^1} \dfrac{\left|C\right|}{\left|V\right|}$ so

$$\sum_{C \in \mathcal{C}^1} \frac{\left|C\right|}{\left|V\right|} X_C \leq \frac{1}{\left|V\right|}\sigma^\star \cdot \tau \leq \sum_{C \in \mathcal{C}^1} \frac{\left|C\right|}{\left|V\right|} X_C + 1 - \sum_{C \in \mathcal{C}^1} \frac{\left|C\right|}{\left|V\right|}$$

We recognize in $\sum_{C \in \mathcal{C}^1} \dfrac{\left|C\right|}{\left|V\right|} X_C$ a weighted sum (with weights $\to 0$ when $n \to \infty$) of independent random variables with mean $\frac{1}{2}$ and bounded variance. By the weighted version of the Law of Large Numbers (sometimes called the Kolmogorov's law), this quantity is equivalent to $\frac{1}{2}\sum_{C \in \mathcal{C}^1(\kappa)} \dfrac{\left|C\right|}{\left|V\right|} \geq \frac{1}{2} - \frac{1}{2}\theta$. Then asymptotically

$$\tfrac{1}{2} - \tfrac{1}{2}\theta \leq \frac{1}{\left|V^{(n)}\right|}\sigma^\star \cdot \tau^{(n)} \leq \tfrac{1}{2} + \tfrac{1}{2}\theta$$

The final overlap is given by

$$\sigma^\star \cap \tau = \max\left(\frac{1}{\left|V\right|}\sigma^\star \cdot \tau, 1 - \frac{1}{\left|V\right|}\sigma^\star \cdot \tau\right) \leq \tfrac{1}{2} + \tfrac{1}{2}\theta.$$

So $\varepsilon^{\max} \leq \theta$ and θ being chosen arbitrary close to $\theta^{\max}$, we conclude that

$$\varepsilon^{\max} \leq \theta^{\max}.$$

References

[1] V. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, *Fast unfolding of communities in large networks*, Journal of Statistical Mechanics: Theory and Experiment, 2008 (2008), p. P10008, https://doi.org/10.1088/1742-5468/2008/10/P10008.

[2] V. Cataudella, G. Franzese, M. Nicodemi, A. Scala, and A. Coniglio, *Critical clusters and efficient dynamics for frustrated spin models*, Physical review letters, 72 (1994), p. 1541.

[3] L. Chayes and H. K. Lei, *Random cluster models on the triangular lattice*, Journal of statistical physics, 122 (2006), pp. 647–670, https://doi.org/10.1007/s10955-005-8078-7.

[4] A. Clauset, M. Newman, and C. Moore, *Finding community structure in very large networks*, Phys. Rev. E, 70 (2004), p. 066111, https://doi.org/10.1103/PhysRevE.70.066111.

[5] P. Diaconis, *The Markov chain Monte Carlo revolution*, Bulletin of the American Mathematical Society, 46 (2009), pp. 179–205, https://doi.org/10.1090/S0273-0979-08-01238-X.

[6] R. G. Edwards and A. D. Sokal, *Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm*, Phys. Rev. D, 38 (1988), pp. 2009–2012, https://doi.org/10.1103/PhysRevD.38.2009.

[7] C. Fortuin and P. Kasteleyn, *On the random-cluster model: I. introduction and relation to other models*, Physica, 57 (1972), pp. 536–564, https://doi.org/10.1016/0031-8914(72)90045-6.

[8] S. Fortunato, *Community detection in graphs*, Physics Reports, 486 (2010), pp. 75–174, https://doi.org/10.1016/j.physrep.2009.11.002.

[9] J. Gaudio and C. Guan, *Sharp exact recovery threshold for two-community Euclidean random graphs*, 2025, https://doi.org/10.48550/arXiv:2501.14830. ArXiv.

[10] D. Kandel, R. Ben-Av, and E. Domany, *Cluster Monte Carlo dynamics for the fully frustrated Ising model*, Phys. Rev. B, 45 (1992), pp. 4700–4709, https://doi.org/10.1103/PhysRevB.45.4700.

[11] L. Münster and M. Weigel, *Spin glasses and percolation*, Frontiers in Physics, 12 (2024), https://doi.org/10.3389/fphy.2024.1448175.

[12] M. Penrose, *Random Geometric Graphs*, vol. 5, Oxford University Press, 2003, https://doi.org/10.1093/acprof:oso/9780198506263.001.0001.

[13] J. Reichardt and S. Bornholdt, *Statistical mechanics of community detection*, Phys. Rev. E, 74 (2006), p. 016110, https://doi.org/10.1103/PhysRevE.74.016110.

[14] A. Sankararaman and F. Baccelli, *Community detection on Euclidean random graphs*, Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), (2018), pp. 2181–2200, https://doi.org/10.1137/1.9781611975031.142.

[15] A. Sankararaman, H. Vikalo, and F. Baccelli, *ComHapDet: a spatial community detection algorithm for haplotype assembly*, BMC Genomics, (2020), https://doi.org/10.1186/s12864-020-06935-x.

[16] Scikit-Learn Library, *Spectral clustering*. https://scikit-learn.org/stable/modules/generated/sklearn.cluster.SpectralClustering.html. [Online; accessed 25-April-2025].

[17] R. H. Swendsen and J.-S. Wang, *Nonuniversal critical dynamics in Monte Carlo simulations*, Phys. Rev. Lett., 58 (1987), pp. 86–88, https://doi.org/10.1103/PhysRevLett.58.86.

[18] A. Traag, P. Van Dooren, and Y. Nesterov, *Narrow scope for resolution-limit-free community detection*, Phys. Rev. E, 84 (2011), p. 016114, https://doi.org/10.1103/PhysRevE.84.016114.

[19] V. Traag, L. Waltman, and N. van Eck, *From Louvain to Leiden: guaranteeing well-connected communities*, Scientific Reports, 9 (2019), https://doi.org/10.1038/s41598-019-41695-z.