Higher-order Monte Carlo cluster dynamics for community detection in Euclidean graphs

Louis HAUSEUX
Centre Inria d'Université Côte d'Azur
NEO *team &* AYANA *team*
Sophia Antipolis, France
louis.hauseux@inria.fr

Nahuel SOPRANO-LOTO
Centre Inria de Paris
DYOGENE/MATHNET *team*
Paris, France
nahuel.soprano-loto@inria.fr

Konstantin AVRACHENKOV
Centre Inria d'Université Côte d'Azur
NEO *team*
Sophia Antipolis, France
konstantin.avratchenkov@inria.fr

***Abstract*—We study GIBBS distributions with competing interactions and propose a higher-order extension of the SWENDSEN–WANG dynamics that incorporates triangular bonds. The new dynamics preserves the same stationary distribution, alleviates frustration, and yields markedly better sampling. When applied to a synthetic Euclidean-graph community-detection benchmark, our algorithm outperforms existing methods.**

***Index Terms*—Monte Carlo Markov Chain, Gibbs energy, spin configuration, Swendsen-Wang dynamics, high-order interaction, percolation.**

I. INTRODUCTION

Graphs provide a natural framework for modelling many real-world systems, which makes efficient graph analysis essential [1]. A central task in graph analysis is to cluster vertices with similar 'behaviour' into communities, a problem known as *community detection* [2].

Algorithms inspired by statistical physics offer intuitive, mathematically elegant, and theoretically analyzable solutions [3], [4]. The central challenges in these types of algorithms are to (i) define an appropriate (BOLTZMANN-)GIBBS model and (ii) efficiently sample the resulting GIBBS distribution.

We propose an original model combining ferromagnetic and anti-ferromagnetic interactions. For the sampling part, when the vertices possess a spatial embedding, a particularly appealing approach is the SWENDSEN-WANG dynamics [5], [6]. This algorithm builds random clusters [7], [8] and flips them, generating a MARKOV chain whose stationary law is the GIBBS measure of the model. This method accelerates sampling from high-probability configurations by enhancing long-range spin correlations. However, when the system contains both ferromagnetic and anti-ferromagnetic interactions, its performance deteriorates [9], and the problem worsens when the interactions convey contradictory information—a phenomenon known as *frustration*.

Inspired by SWENDSEN-WANG, our novel dynamics incorporates higher-order interactions (not just edges between spins but also triangles). Despite these additional interactions, the energy of each configuration remains unchanged, so the stationary distribution is still the original GIBBS measure. The higher-order interactions reduce frustration and significantly improve sampling efficiency.

Experiments demonstrate its superiority on the challenging Euclidean random graph model of SANKARARAMAN & BACCELLI [10], relevant to haplotype assembly in biology [11].

Paper organization: § II introduces background on statistical-physics methods; § III presents our higher-order SWENDSEN–WANG dynamics and proposes a full algorithm for community detection based on spin correlation estimation; § IV demonstrates its performance on synthetic data; § V concludes.

II. PRELIMINARIES

A. Community detection, spin configurations and GIBBS energy

Let $G(V, E)$ be a graph (undirected, unweighted) on a set V of n nodes, E being the set of edges. We assume the existence of K communities $V = V_0 \dot\cup \ldots \dot\cup V_{K-1}$. The problem of community detection is to recover the communities $V_0, \ldots, V_{K-1}$ given the input graph G. Throughout the paper, we will also assume that K is given, and that we are in the *assortative* case, the latter meaning that there is a higher edge density within a community than between two nodes from different communities.

The stochastic block model [12] is a simple generative model for such a graph. The communities $\overline{x} \in \{0, \ldots, K-1\}$ of each node $x \in V$ are drawn independently according to a distribution $\mathbf{p} = (p_0, \ldots, p_{K-1})$, and each hypothetical edge $\{x, y\}$ is drawn independently with probability $P_{\overline{x},\overline{y}}$, where P is a $K \times K$ matrix giving the edge density between communities. The model we look at in § IV is a *geometric* stochastic block model. Introducing spatial constraints when drawing edges often makes this latter model much harder both to analyze and to solve algorithmically.

We will focus on *clustering* rather than *classification*, *i.e.* look at the communities up to any permutation on the set of communities $\{0, \ldots, K-1\}$. 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:

L.H. is supported by the Université Côte d'Azur (UCA) DS4H (ANR, reference number ANR-17-EURE-0004) and 3IA Côte d'Azur for partial funding of his PhD thesis.

N.S.L. is supported by the SNS (Smart Networks and Services) JU and its members, funded by the European Union under Grant Agreement number 101139161.

DEFINITION 1. (Overlap.) *Let $\sigma^\star : V \to \{0,\dots,K-1\}$ be the Ground-Truth of the model, defined as the 'true' communities in the generative model at issue. For an estimator of the communities $\sigma : V \to \{0,\dots,K-1\}$, we define its* overlap *with $\sigma^\star$ as*

$$\max_{\pi} |\{x \in V \mid \sigma_x = \pi(\sigma^\star_x)\}|,$$

where the maximum is taken over all permutations π of the set of communities $\{0,\dots,K-1\}$.

In the terminology of statistical physics, an estimator of the communities, $\sigma : V \to \{0,\dots,K-1\}$, can be viewed as a *spin configuration*, and we proceed to define an energy, or Hamiltonian, associated with such configurations. Assume that we have two arbitrary subsets $F, \tilde{F} \subseteq E$, respectively called *ferromagnetic* and *anti-ferromagnetic* edges. These sets will be chosen appropriately as a function of the graph in § IV. The GIBBS energy associated with σ is defined as

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

This energy can be interpreted as follows: For each ferromagnetic edge $e \in F$, we add a penalty $\beta_e \geq 0$ if the configuration σ does not satisfy ferromagnetism, *i.e.* if the corresponding spins belong to different communities. Similarly, we add a penalty γ_e for each anti-ferromagnetic edge $e \in \tilde{F}$ unsatisfied, *i.e.* when the corresponding spins belong to the same community. The lower the energy $U(\sigma)$ of a configuration σ, the more probable that configuration is.

The GIBBS distribution is the probability measure over the set of configurations $\Sigma := \{0,\dots,K-1\}^n$, which assigns to a configuration $\sigma \in \Sigma$ the probability mass

$$d\mu(\sigma) = \frac{1}{Z} e^{-U(\sigma)}, \quad (2)$$

where the normalization term Z is called the *partition function*. As motivation, we mention that for the stochastic block model, the log-likelihood of a configuration σ coincides with the GIBBS energy in Eq. 1 (see the Appendix).

The statistical physics model we use has the triple advantage of being simple, intuitive, and general. Interestingly, the *modularity* of a graph [13], widely used for community detection in graphs [14], [15], can be written as a particular case of a GIBBS energy [3]. Note also that our model generalizes the Constant POTTS Model [4].

B. GIBBS *sampler and* SWENDSEN-WANG *dynamics*

The state space Σ grows exponentially with n; hence an algorithm capable of sampling efficiently from the GIBBS distribution is required.

Many Monte Carlo MARKOV Chain (MCMC) algorithms have been proposed to do so. Some of them have a local dynamics using *single-spin update*, *e.g.* GLAUBER dynamics. With GLAUBER dynamics updating spins one at a time, communities can indeed be well recovered *locally*. For nodes embedded in Euclidean space, however, it fails to find the correct *global* configuration when two well-labeled local spin clusters contradict each other.

The SWENDSEN-WANG dynamics [5] has a more global update mechanism, capable of updating entire clusters of spins in a single step. In each step, it randomly bonds neighboring spins if they agree in state, forming clusters. Then, all spins within each cluster are flipped together with some probability. This dynamics has proved to be very efficient on spin models with a geometric structure (in the original paper, the nodes were on the 2D grid).

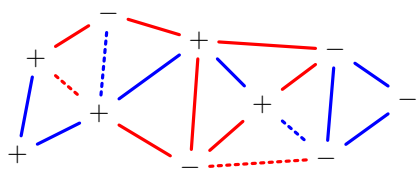

Fig. 1. Ferromagnetic F and anti-ferromagnetic edges $\tilde{F}$. Unsatisfied edges by the spin configuration σ are in dashed lines.

The EDWARDS & SOKAL formalism [6] makes it possible to incorporate into the SWENDSEN–WANG dynamics both ferromagnetic and anti-ferromagnetic 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. 1, 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 $p = 1 - e^{-\omega_e}$, where
$$\omega_e = \begin{cases} \beta_e & \text{if } e \in F \\ \gamma_e & \text{if } e \in \tilde{F} \end{cases}.$$
3) For each cluster C of the frozen graph resulting, flip all its spins with probability $\frac{1}{2}$, and keep them unchanged with probability $\frac{1}{2}$.

See Fig. 2 for an illustration. The SWENDSEN–WANG dynamics induces a MARKOV chain on the set of configurations $\Sigma := \{-,+\}^V$ whose stationary distribution is the GIBBS distribution (Eq. 2) [6].

C. KBD *algorithm (higher-order* GIBBS *sampling)*

The SWENDSEN-WANG dynamic is particularly effective in purely ferromagnetic systems. In this case, the correlation between two spins σ_x and σ_y is equal to the probability that the two nodes $x, y \in V$ are connected in the Random-Cluster model [7], [9]:

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

This means that the bond dynamics will percolate precisely when the inverse temperature β induces long-range spin correlations in the GIBBS measure. In short, the β that fits the GIBBS energy works for percolation.

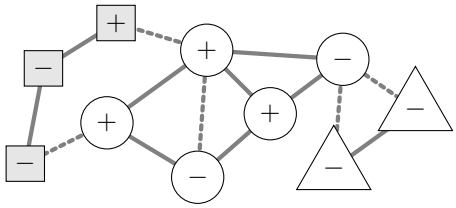

Fig. 2. Percolation on the satisfied edges of Fig. 1. Not retained edges are shown with dashed lines. Three clusters (squares, circles, triangles) are formed. The spins in each cluster are flipped: Compared to Fig. 1, the squares are inverted, whereas the other two clusters are left unchanged.

However, with antiferromagnetism the system percolates before spins align at long range, so a correct β (*e.g.* the one fitting with likelihood, see the Appendix) is much larger than the percolation inverse temperature and the random-cluster dynamics freezes. Moreover, the presence of opposite interactions can lead to contradictory information. Even the *ground state* (the configuration that minimizes energy) may fail to satisfy certain bonds. These contradictory interactions, known as *frustration*, increase even more the mixing time of the block-update chain.

One solution is to artificially change the system's temperature β, γ [9].

KANDEL, BEN-AV & DOMANY (KBD) [16] proposed a more elaborate solution for the fully frustrated ISING model on the 2D grid, consisting of using higher-order bonds (namely "plaquettes"). Their dynamics retains the same GIBBS energy for each configuration σ, yet freezes far fewer bonds.

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

III. PROPOSED METHOD

In our problem, the input data are a weighted graph $G(V, F \cup \tilde{F}, (\beta_e), (\gamma_e))$. Our original community detection algorithm consists of the following steps:

i) For the case $K = 2$ (two communities) considered here, we transfer the energy in Eq. 1 from edges to ferromagnetic triangles, which serve as new bonds. The new dynamics induces the same GIBBS measure for stationary distribution.

ii) Optionally we rescale all β_e, γ_e by a constant $c \in (0, 1)$ to tune cluster size.

iii) Starting from a random initial configuration σ^0, we then perform `Steps` iterations of the new 'SWENDSEN-WANG triangle' dynamics. Denoting by σ^s the configuration obtained at step $s \in \{1, \dots, \texttt{Steps}\}$, we can estimate the spin correlation for every pair x, y as follows:

$$\mathbb{E}_{\sigma \sim \text{GIBBS}}[\sigma_x \sigma_y] \approx \hat{S}_{x,y} = \frac{1}{\texttt{Steps}} \sum_{s=1}^{\texttt{Steps}} \sigma_x^s \sigma_y^s.$$

iv) The resulting spin correlation matrix is viewed as a complete graph, substantially easier to analyze for community detection. An algorithm such as LOUVAIN [14], [15] yields very good results in practice. For fair comparison with the Constant POTTS Model [4], we employed that algorithm in our final step.

The originality of our method lies in the first point i, which we describe in detail in the present section. In §III-D we provide an interpretation of the dynamics that partly explains its markedly better performance compared with the "classical" SWENDSEN-WANG dynamics. Point i consists of generalizing the ideas behind the KBD algorithm to the case of a more general graph than the 2D grid. For simplicity, we will consider the case $K = 2$, with spins taking values in $\{-, +\}$, but the ideas can be generalised to $K \geq 3$.

A. Triangle classification

Let us look at the triangles formed by edges of $F \cup \tilde{F}$. There are four possible configurations (see Tab. I), which can be divided into two groups: those that are *inherently contradictory* (△ and △), and those that may be contradictory or not depending on the spin configuration (△ and △), that we call *potentially contradictory*. Call T the set of potentially contradictory triangles. We partition $T = T_1 \dot{\cup} T_2$, where T_1 is the set of all pure ferromagnetic triangles (△), and T_2 is the set of triangles with one ferromagnetic and two anti-ferromagnetic edges (△).

The idea is to distribute the energy in Eq.1 over these triangles $t \in T$ and to only freeze them when the energy with configuration σ is minimal (*i.e.* all edges are satisfied by σ). We call such a σ-minimal-energy triangle *satisfied*. See Fig. 3.

TABLE I
POSSIBLE TRIANGLES AND THEIR RESPECTIVE ENERGY WHEN EACH EDGE (FERROMAGNETIC OR ANTI-FERROMAGNETIC) HAS A WEIGHT ω.

ENERGY	**Inherently contradictory**		**Potentially contradictory**	
Spins σ \ Triangles	△	△	△	△
+ + +	3ω	ω	2ω	**0**
+ + −	ω	ω	2ω	2ω
− + +	ω	3ω	**0**	2ω
− + −	ω	ω	2ω	2ω

B. Energy transfer

We now describe in detail the energy transfer mechanism (item i above), which redistributes edge weights onto triangular bonds. We define $\Omega_e = \begin{cases} \beta & \text{if } e \in F \\ \gamma & \text{if } e \in \tilde{F} \end{cases}$. For a triangle $t \in T$ and an edge $e \in t$, let $\omega_{e,t}$ denote the weight transferred from edge e to triangle t. To be in concordance with Tab. I, a first constraint is that each edge $e \in t$ of a triangle $t \in T$ must transfer the same weight to the triangle t

$$\forall t \in T, \forall e \in t, \omega_{e,t} =: \omega_t \text{ (same weight for each edge } e \in t).$$

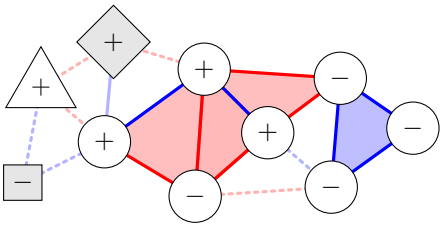

Fig. 3. Percolation on the satisfied triangles $t \in T_1 \cup T_2$ with the same configuration σ as in Fig. 1. One large cluster (circles) is formed, along with three single-site clusters. The spins of 'square' and 'diamond' clusters are flipped.

We shall also respect the linear constraints

$$\forall e, \sum_{t \in T, t \ni e} \omega_{e,t} \leq \Omega_e.$$

Finally, any leftover energy on an edge e is stored as

$$\omega_e := \Omega_e - \sum_{t \in T,\, t \ni e} \omega_{e,t} \geq 0.$$

Under these definitions, the energy in Eq.1 can be written as

$$U(\sigma) = \sum_{t \in T} 2\omega_t \mathbf{1}_{\sigma \not\sim t} + \sum_{e=\{x,y\} \in F} \omega_e \mathbf{1}_{\sigma_x \neq \sigma_y} + \sum_{e=\{x,y\} \in \tilde{F}} \omega_e \mathbf{1}_{\sigma_x = \sigma_y}, \quad (3)$$

where $\mathbf{1}_{\sigma \not\sim t}$ means that σ does not satisfy the restriction imposed by t.

C. Swendsen-Wang *triangle dynamics*

After transferring the energy (see §III-B), the proposed dynamics is similar to the classical Swendsen–Wang algorithm (§II-B), except that we now consider *bonds* $b \in B = T \cup F \cup \tilde{F}$, *i.e.* either triangles $t \in T$ or edges $e \in F \cup \tilde{F}$. In accordance with Eq. 3, the weight ω_b associated with any $b \in B$ is $\omega_b = \begin{cases} 2\omega_t & \text{if } b = t \in T \subseteq B, \\ \omega_e & \text{if } b = e \in F \cup \tilde{F} \subseteq B \end{cases}$. Since we now only check whether a bond is satisfied, the energy can be rewritten as $U(\sigma) = \sum_{b \in B} \omega_b \mathbf{1}_{\sigma \not\sim b}$.

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

1) Starting from a configuration $\sigma : V \to \{-, +\}$, keep only the bonds $b \in B$ satisfied.
2) Apply a percolation model to the retained bonds $b \in B$: "freeze" them with probability $p_b = 1 - e^{-\omega_b}$.
3) For each cluster $C \subseteq V$ that is formed, flip the spins with probability $\frac{1}{2}$, and keep them unchanged with probability $\frac{1}{2}$. In Fig. 3, the spins of 'square' and 'diamond' clusters are flipped.

D. Interpretation of the dynamic

Swendsen–Wang performs poorly in systems with both ferromagnetic and anti-ferromagnetic interactions, because the critical inverse temperature β_c (onset of long-range order) lies above the Random-Cluster model's percolation threshold β_{per}. This issue is further exacerbated by frustration.

The advantage of the triangle-based dynamics is that it can be viewed as follows: the bonds are the triangles $t \in T$, while the "vertices" are the edges $e \in F \cup \tilde{F}$. For a given configuration σ, we distinguish four types of edges $e = \{x, y\}$:

1) Truly-positive — if $e \in F$ and $\sigma_x = \sigma_y$.
2) Falsely-positive ··· if $e \in F$ and $\sigma_x \neq \sigma_y$.
3) Truly-negative — if $e \in \tilde{F}$ and $\sigma_x \neq \sigma_y$.
4) Falsely-negative ··· if $e \in \tilde{F}$ and $\sigma_x = \sigma_y$.

These edges can be grouped into two categories, the "True" (— and —) and the "False" (··· and ···). An edge configuration σ^{e} can therefore be viewed as a vector on $F \cup \tilde{F}$ taking values in $\{\texttt{True}, \texttt{False}\}$, and we use the notation $\sigma \to \sigma^{\text{e}}$ for the mapping from a spin configuration σ to its edge configuration σ^{e}. We restrict attention to the admissible configurations $\sigma^{\text{e}} \in \Sigma^{\text{e}} := \left\{ \sigma^{\text{e}} \;\middle|\; \exists \sigma,\ \sigma \to \sigma^{\text{e}} \right\}$ such that there exists σ with $\sigma \to \sigma^{\text{e}}$.

Note that, $|\Sigma^{\text{e}}| = \frac{1}{2^{\texttt{Clust}}} |\Sigma|$, where $\texttt{Clust}$ denotes the number of connected components of the graph $(V, F \cup \tilde{F})$.

The triangles $t \in T_1 \cup T_2$ are purely ferromagnetic bonds linking three $\texttt{True}$ spins. The associated Gibbs distribution is

$$d\mu(\sigma^{\text{e}}) = \frac{1}{Z^{\text{e}}} e^{-U(\sigma^{\text{e}})} \mathbf{1}_{\sigma^{\text{e}} \in \Sigma^{\text{e}}}.$$

One can interpret the triangle dynamics as a standard random cluster process on a hypergraph where edges of the original graph become sites (subject to a constraint of consistency). While this complicates the theoretical analysis (*e.g.* loss of monotonicity of the measures, see [8] § 11.5, p. 333), it intuitively explains why allowing triangle bonds reduces frustration: the dynamics now becomes purely ferromagnetic.

In practice, as will be demonstrated in the next section, the experiments with the new method show very good results.

IV. Experiments: the Geometric Stochastic Block Model

To evaluate our approach, we tested and compared it on a hard instance of the Sankararaman & Baccelli model [10], recently investigated by Gaudio & Guan [17] in the $\log n$ degree regime.

This case is challenging because it introduces substantial frustration that spreads through the geometric structure of the space and many standard community detection algorithms failed on it (*e.g.* spectral clustering [18] or single-spin-update Glauber dynamics). Most of them, such as Louvain [14], are simply not suited to this problem (a graph with negative weights). Even the algorithms proposed by Sankararaman & Baccelli [19] and Gaudio & Guan [17] for this specific model did not work under our chosen parameters; their algorithms are designed to operate asymptotically when the degree of the nodes tends to infinity.

To the best of our knowledge (and according to our experiments), only the LEIDEN [15] algorithm combined with the Constant POTTS Model [4] produced good results. We therefore compared our method with this algorithm.

We considered two versions of our method: (i) SWENDSEN-WANG edges, *i.e.* the Swendsen–Wang dynamics without any energy transfer to triangles; (ii) SWENDSEN-WANG triangles, the new dynamics described in §III.

A. *The* SANKARARAMAN & BACCELLI *model*

The tested model [10], [17] consists of placing N points uniformly on the 2D unit torus and constructing on them a geometric graph $G(V,\tau)$ that depends on a threshold $\tau > 0$. We then assign communities at random and apply on this geometric graph a stochastic block model (see §II-A), defined by a single parameter $p \in (\frac{1}{2}, 1)$: the 2×2 density matrix P is given by $P_{+,+} = P_{-,-} = p_{\text{in}} = p$ and $P_{+,-} = P_{-,+} = p_{\text{out}} = 1-p$. Thanks to the resulting sub-graph SBM $\subseteq G$, we can define the ferromagnetic edges $F \leftarrow$ edges(SBM) and the anti-ferromagnetic edges $\tilde{F} \leftarrow$ edges$(G) \setminus$ edges(SBM). They are weighted respectively by $\beta \leftarrow -c\log\left(\frac{p_{\text{out}}}{p_{\text{in}}}\right) = -c\log\left(\frac{1-p}{p}\right)$ and $\gamma \leftarrow -c\log\left(\frac{1-p_{\text{in}}}{1-p_{\text{out}}}\right) = \beta$. Setting $c \leftarrow 1$ fits the likelihood on a configuration σ with the GIBBS energy (see the Appendix, the geometric case is similar to the standard one).

We chose the following parameters: $N \leftarrow 1000$, $\tau = \sqrt{\frac{20}{1000\pi}}$ so that every node x has about 20 neighbors, $p \leftarrow \frac{3}{4}$ and $c \leftarrow \frac{1}{4}$.

This model is challenging because spatial location provides no information about the community. Further, the stochastic block model introduces substantial noise. While it is relatively easy to reconstruct the two communities locally, the propagation of false information through the graph makes it very difficult to determine whether two spatially distant nodes belong to the same community or not.

This benchmark also models haplotype-assembly: sequencing reads must be assigned to the two parental chromosomes, a task equivalent to a binary community detection problem in a noisy spatial graph [11].

The goal is to recover the communities given $F, \tilde{F}, \beta, \gamma$. A realization of such a graph is given in Fig. 4.

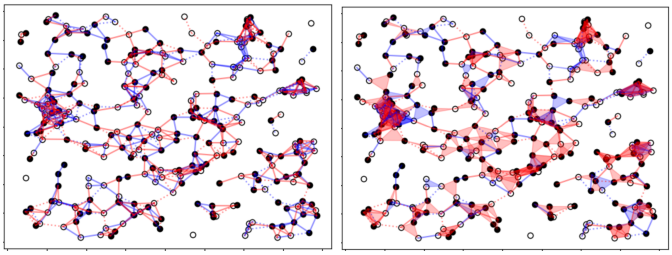

Fig. 4. One geometric graph with $N \leftarrow 300$ (we erased wrap-around edges due to the torus geometry). Spin colors correspond to the Ground-Truth. On the right, the same graph with the energy transferred onto the triangles.

B. *Our algorithms*

Our SWENDSEN-WANG triangles version of the algorithm is described in §III. The SWENDSEN-WANG edges version is the same but without energy transfer and not looking at the triangles.

In our experiments, for the SWENDSEN-WANG triangles version, more than 97% of the energy was transferred to the triangles; the remaining $\lesssim 3\%$ stayed on the edges.

Starting from a random configuration σ^0, we then ran `Steps` $\leftarrow 1000$ iterations of each dynamics.

After computing the spin correlation matrix, we then applied the LEIDEN algorithm [15] using the Constant POTTS Model [4] as the objective function.

C. LEIDEN *algorithm and Constant* POTTS *Model*

The LEIDEN [15] algorithm, a successor to the LOUVAIN [14] algorithm, is a graph-based community-detection algorithm. The randomness it introduces in its process allows it to perform well with cost functions other than modularity [13]. For example, it works well when the objective function is the Constant POTTS Model [4], which accepts a graph with negative weights. We ran this algorithm on the graph $F \cup \tilde{F}$ weighted by $+\beta$ and $-\gamma$, respectively.

As in the final step of our algorithm, the resolution parameter was set to 0. Because the algorithm returned more than two communities, we assigned the spin $+$ to every node in the largest community and the spin $-$ to all the other nodes.

D. *Results*

All the results we present are averaged over 100 graph realizations.

1) Comparison between the two dynamics SWENDSEN-WANG *triangles and* SWENDSEN-WANG *edges:* Fig. 5 compares the two cluster-dynamics at the same temperature β, γ. Because the triangle version freezes far fewer edges, its largest cluster stays below 50% during the first iterations (*vs* $\approx 90\%$ for the edge version). Consequently, the overlap (see Def. 1) with Ground-Truth rises much faster for the triangle version.

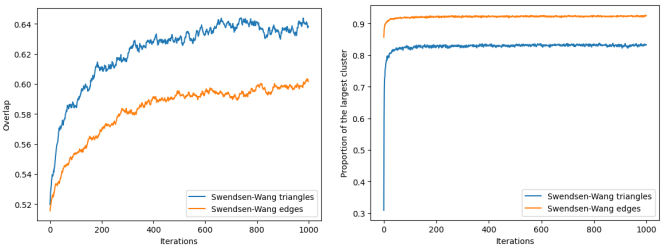

Fig. 5. (Left) Comparison of the averaged overlap (see Def. 1) with the Ground-Truth (fraction of correctly clustered nodes) and (Right) the averaged size of the largest cluster as a function of the number of iterations. In blue, SWENDSEN-WANG triangles dynamics; in orange, SWENDSEN-WANG edges. SWENDSEN-WANG triangles is initially (with an unfavorable random initial σ^0 configuration) much less frozen than SWENDSEN-WANG edges.

2) Final overlap with Ground-Truth: In Fig. 6, the accuracy distribution (*i.e.* overlap with Ground-Truth) is shown for the two versions of our algorithm as well as for the Constant Potts Model applied directly to the weighted graph $F \cup \tilde{F}$. Our Swendsen-Wang triangles approach clearly outperforms both the Swendsen-Wang edges variant and the Constant Potts Model baseline, being the only one to achieve almost perfect accuracy.

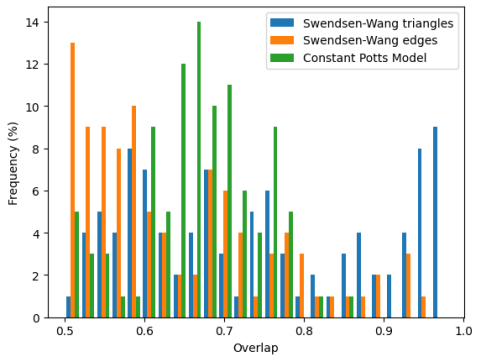

Fig. 6. Overlap with Ground-Truth distribution (over 100 realizations) resulting from the three algorithms: our two versions, Swendsen-Wang triangles and Swendsen-Wang edges, together with the Constant Potts Model applied to the graph $F \cup \tilde{F}$ weighted by $+\beta$ and $-\gamma$, respectively.

V. Conclusion

We have introduced a higher-order extension of the Swendsen–Wang cluster dynamics that redistributes pairwise Gibbs energy onto carefully selected triangular bonds. This "triangle-based dynamics" preserves the original Gibbs stationary distribution, alleviates frustration by freezing far fewer bonds, and therefore enhances long-range updates without altering the model temperature.

On the demanding 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. It achieves almost perfect overlap ($\geq$90%) with Ground-Truth communities one time out of four and reaches its first plateau markedly faster, demonstrating better mixing and sampling efficiency.

Beyond community detection, the proposed framework is generic: any pairwise Gibbs model on a sparse graph can in principle be augmented with higher-order bonds chosen to relieve local frustration. Future work will explore three main directions:

- Extending the construction to $K > 2$ communities via simplicial complexes of order > 2.
- Deriving theoretical properties for such dynamics.
- Investigating applications to large-scale biological data sets, such as long-read haplotype assembly and metagenomic binning.

These directions suggest that higher-order Monte Carlo cluster dynamics could become a versatile tool whenever frustration hampers traditional MCMC sampling.

Appendix

The SBM-likelihood of a configuration σ matches a Gibbs energy function for an appropriate choice of *inverse temperatures* β_e and γ_e. In fact, let $G(V, E)$ be a graph over n nodes V and K communities generated by the stochastic block model with community probability $\mathbf{p}$ and edge density matrix P. For simplicity, we assume the model is *homogeneous*, *i.e.* $P_{k,k} = p_{\text{in}} \in (0,1)$ and $P_{k,k'} = p_{\text{out}} \in (0, p_{\text{in}})$ if $k \neq k'$. $G(V, E)$ is generated by drawing independent community labels $\overline{x}_1, \dots, \overline{x}_n \sim \mathbf{p}$ and for each pair $1 \leq i < j \leq n$, independently, an edge $\{x, y\}$ is drawn in E with probability p_{in} if $\overline{x} = \overline{y}$ and with probability p_{out} if $\overline{x} \neq \overline{y}$. By independence, the likelihood $L(\sigma; G)$ of a spin configuration σ can be factored as follows

$$
\begin{aligned}
L(\sigma; G) &= \prod_{\{x,y\} \in E} p_{\text{in}}^{1-\mathbf{1}_{\sigma_x \neq \sigma_y}} p_{\text{out}}^{\mathbf{1}_{\sigma_x \neq \sigma_y}} \\
&\quad \prod_{\{x,y\} \notin E} (1 - p_{\text{out}})^{1-\mathbf{1}_{\sigma_x = \sigma_y}} (1 - p_{\text{in}})^{\mathbf{1}_{\sigma_x = \sigma_y}} \\
&\propto \prod_{\{x,y\} \in E} \left(\frac{p_{\text{out}}}{p_{\text{in}}} \right)^{\mathbf{1}_{\sigma_x \neq \sigma_y}} \prod_{\{x,y\} \notin E} \left(\frac{1 - p_{\text{in}}}{1 - p_{\text{out}}} \right)^{\mathbf{1}_{\sigma_x = \sigma_y}} \\
&\propto \exp \left[- \left(\beta \sum_{\{x,y\} \in F} \mathbf{1}_{\sigma_x \neq \sigma_y} + \gamma \sum_{\{x,y\} \in \tilde{F}} \mathbf{1}_{\sigma_x = \sigma_y} \right) \right].
\end{aligned}
$$

And $L(\sigma; G) \propto \frac{1}{Z} e^{-U(\sigma)}$ taking $F \leftarrow E$, $\tilde{F} \leftarrow E^c$, $\beta \leftarrow -\log\left(\frac{p_{\text{out}}}{p_{\text{in}}}\right)$ and $\gamma \leftarrow -\log\left(\frac{1-p_{\text{in}}}{1-p_{\text{out}}}\right)$. □

References

[1] K. Avrachenkov and M. Dreveton, *Statistical Analysis of Networks*. Now publishers, 2022.

[2] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 3, pp. 75–174, 2010.

[3] J. Reichardt and S. Bornholdt, "Statistical mechanics of community detection," *Phys. Rev. E*, vol. 74, p. 016110, Jul 2006.

[4] A. Traag, P. Van Dooren, and Y. Nesterov, "Narrow scope for resolution-limit-free community detection," *Phys. Rev. E*, vol. 84, p. 016114, Jul 2011.

[5] R. H. Swendsen and J.-S. Wang, "Nonuniversal critical dynamics in monte carlo simulations," *Phys. Rev. Lett.*, vol. 58, pp. 86–88, Jan 1987.

[6] R. G. Edwards and A. D. Sokal, "Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm," *Phys. Rev. D*, vol. 38, pp. 2009–2012, Sep. 1988.

[7] C. Fortuin and P. Kasteleyn, "On the random-cluster model: I. introduction and relation to other models," *Physica*, vol. 57, no. 4, pp. 536–564, 1972.

[8] G. Grimmett, *The random-cluster model*. Springer, 2006, vol. 333.

[9] L. Münster and M. Weigel, "Spin glasses and percolation," *Frontiers in Physics*, vol. 12, 2024.

[10] A. Sankararaman and F. Baccelli, "Community detection on Euclidean random graphs," *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 2181–2200, 2018.

[11] A. Sankararaman, H. Vikalo, and F. Baccelli, "ComHapDet: a spatial community detection algorithm for haplotype assembly," *BMC Genomics*, 2020.

[12] P. Holland, K. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps," *Social Networks*, vol. 5, no. 2, pp. 109–137, 1983.

[13] A. Clauset, M. Newman, and C. Moore, "Finding community structure in very large networks," *Phys. Rev. E*, vol. 70, p. 066111, Dec 2004.

[14] V. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, p. P10008, oct 2008.

[15] V. Traag, L. Waltman, and N. van Eck, "From Louvain to Leiden: guaranteeing well-connected communities," *Scientific Reports*, vol. 9, no. 5233, 2019.
[16] D. Kandel, R. Ben-Av, and E. Domany, "Cluster Monte Carlo dynamics for the fully frustrated Ising model," *Phys. Rev. B*, vol. 45, pp. 4700–4709, Mar 1992.
[17] J. Gaudio and C. Guan, "Sharp exact recovery threshold for two-community Euclidean random graphs," 2025, arXiv.
[18] Scikit-Learn Library, "Spectral clustering," https://scikit-learn.org/stable/modules/generated/sklearn.cluster.SpectralClustering.html, [Online; accessed 25-April-2025].
[19] A. Sankararaman and F. Baccelli, "Community detection on a spatial graph," 2017, GitHub project. [Online]. Available: https://github.com/abishek90/Community-Detection-on-a-Spatial-Graph/blob/master/simulation_spectral.py