

Phase Transition for Stochastic Block Model with more than $\sqrt{n}$ Communities (II)

Alexandra Carpentier*, Christophe Giraud[†] and Nicolas Verzelen[‡]

Abstract

A fundamental theoretical question in network analysis is to determine under which conditions community recovery is possible in polynomial time in the Stochastic Block Model (SBM). When the number K of communities remains smaller than $\sqrt{n}$ —where n denotes the number of nodes in the observed graph—, non-trivial community recovery is possible in polynomial time above, and only above, the Kesten–Stigum (KS) threshold, originally postulated using arguments from statistical physics.

When $K \geq \sqrt{n}$, Chin, Mossel, Sohn, and Wein [Chi+25] recently proved that, in the *sparse regime*, community recovery in polynomial time is achievable below the KS threshold by counting non-backtracking paths. This finding led them to postulate a new threshold for the many-communities regime $K \geq \sqrt{n}$. Subsequently, Carpentier, Giraud, and Verzelen [CGV25] established the failure of low-degree polynomials below this new threshold across *all density regimes*, and demonstrated successful recovery above the threshold in *certain moderately sparse* settings. While these results provide strong evidence that, in the many community setting, the computational barrier for community recovery lies at the threshold proposed in [Chi+25], the question of achieving recovery above this threshold still remains open in most density regimes.

The present work is a follow-up to [CGV25], in which we prove Conjecture 1.4 stated therein by:

- 1- Constructing a family of motifs satisfying specific structural properties; and
 - 2- Proving that community recovery is possible above the proposed threshold by counting such motifs.
- Together with [Chi+25] and [CGV25], our results provide a complete picture of the computational barrier for community recovery in the SBM with $K \geq \sqrt{n}$ communities. They also indicate that, in moderately sparse regimes, the optimal algorithms appear to be fundamentally different from spectral methods.

1 Introduction

Network analysis aims to understand the structure of random interactions among individuals or objects. A network over n individuals is typically represented as an undirected graph with n nodes, where edges encode observed binary interactions between pairs of individuals. Examples include social networks (friendships or follower relations), biological networks (gene–gene or protein–protein interactions), and information networks (such as email exchanges or citation graphs). A fundamental statistical model for such data is the *Stochastic Block Model* (SBM), introduced by Holland, Laskey, and Leinhardt [HLL83]. In the SBM, the n nodes are partitioned into $K < n$ latent groups or *communities*. Conditional on these (unobserved) community labels, edges are generated independently, with probabilities depending only on the communities of the connected nodes. In its simplest form, the within-community connection probability is p , while the between-community connection probability is q , with $p > q$. This model captures community structure in a mathematically tractable way and has been widely studied in statistics, computer science, and probability theory [Abb18].

The central inferential task in SBM analysis is to recover the latent community memberships from an observed network. In the sparse regime, where q scales as $1/n$, the seminal paper of Decelle, Krzakala, Moore, and Zdeborová [Dec+11] conjectured—based on the replica heuristic from statistical physics—that, when

*Institut für Mathematik – Universität Potsdam, Potsdam, Germany. Alexandra.Carpentier@uni-potsdam.de

[†]Laboratoire de Mathématiques d’Orsay, Université Paris-Saclay, CNRS, France. Christophe.Giraud@universite-paris-saclay.fr

[‡]INRAE, Institut Agro, MISTEA, Univ. Montpellier, France. Nicolas.Verzelen@inrae.fr

K is fixed and $n \rightarrow +\infty$, community recovery in polynomial-time is possible only above the Kesten–Stigum (KS) threshold

$$\frac{n\lambda^2}{K\lambda + K^2q} > 1. \quad (1)$$

Considerable effort has been devoted to confirming this conjecture. Polynomial-time algorithms provably achieving non-trivial recovery above the KS threshold have been developed in [Mas14; MNS18; BLM15; AS15; Chi+25]. Some hardness results below the KS threshold have been established within the Low-Degree Polynomial framework [LG24; SW25; Din+25], providing strong evidence for computational intractability in this regime. While most prior research has focused on small numbers of communities, recent attention has turned to the regime where the number of communities K grows with n . Early work by Chen and Xu [CX16] proposed polynomial-time algorithms achieving partial recovery near the KS threshold. Subsequent studies [SZ24; Chi+25] have shown that when $K = o(\sqrt{n})$, recovery above the KS threshold is possible in polynomial time, while low-degree polynomials fail below the KS threshold, thus confirming the predictions of [Dec+11] in this regime.

When $K \gg \sqrt{n}$, the computational barrier for recovery remained poorly understood until recently. Chin, Mossel, Sohn, and Wein [Chi+25] made a breakthrough by showing that, in the sparse setting (where q scales as $1/n$) with $K \gg \sqrt{n}$ communities, partial recovery is achievable *below* the KS threshold using a polynomial-time algorithm based on non-backtracking statistics. Their algorithm succeeds in the sparse regime as soon as

$$\lambda \gtrsim_{\log} \left(q + \frac{\lambda}{K} \right)^{1 - \log_n(K)}, \quad (2)$$

where $\gtrsim_{\log}$ hides polylogarithmic factors, and $\log_n$ denotes the logarithm in base n , so that $K = n^{\log_n(K)}$. The failure of low-degree polynomials below the threshold (2) was later established by Carpentier, Giraud, and Verzelen [CGV25], not only in the sparse setting (q scaling as $1/n$) considered in [Chi+25], but across all density regimes satisfying $q = o(1)$. Furthermore, [CGV25] shows that community recovery is possible above the threshold (2) using m -clique counting when $q \asymp n^{-2/(m+1)}$, for $m \in \{3, 4, 5, \dots\}$, and by counting self-avoiding paths of length $m - 1$ when $q \asymp n^{-(m-2)/(m-1)}$, for $m \in \{3, 4, 5, \dots\}$. For the remaining density regimes, Conjecture 1.4 of [CGV25] predicts that community recovery remains possible above the threshold (2) through the enumeration of specific motifs satisfying certain structural properties.

The present paper is a follow-up to [CGV25], dedicated to proving this conjecture. Our main contributions are :

1. To construct a new family of motifs fulfilling the structural properties required in Conjecture 1.4 of [CGV25];
2. To prove that an algorithm based on counting such motifs successfully recovers the communities above the threshold (2).

These results complement the contributions of [Chi+25] and [CGV25], and together establish that the computational barrier for community recovery in the SBM with $K \geq \sqrt{n}$ communities lies at (2). Moreover, they indicate that the optimal algorithms in this regime appear to be unrelated to spectral methods, echoing similar observations made for Gaussian mixture models [EGV24].

Organisation and notation. In Section 2, we describe the statistical setting, and we recall the Conjecture 1.4 of [CGV25] that we prove in this paper. In Section 3, we introduce the family of motifs used for successful recovery of the communities above the threshold (2), and we state our main results. The proofs are provided in Section 4 and 5.

We denote by $\log(\cdot)$ the natural logarithm, and by $\log_n(\cdot)$ the logarithm in base n , i.e. $K = n^{\log_n(K)}$. For better readability, we write $\mathbf{1}\{A\}$ for the indicator function $\mathbf{1}_A$ of the set A . We also use the short notation $[D]$ for the set $\{1, \dots, D\}$. In the discussion of the results, we use the symbol $a \lesssim_D b$ (respectively $a \lesssim_{\log} b$) to state that a is smaller than b up to a possible polynomial factor in D (resp. up to a poly-log factor), and we use $a \asymp b$ to state that $a \lesssim b \lesssim a$.

2 Setting and goal

2.1 Stochastic block model and motif counting

Stochastic Block Model. Let K be an integer in $[2, n]$, and $p, q \in (0, 1)$, with $p > q$. We assume that Y^* is sampled according to the following distribution:

1. $z_1, \dots, z_n \stackrel{i.i.d.}{\sim} \text{Uniform}\{1, \dots, K\}$;
2. Conditionally on $z_1, \dots, z_n$, the entries $(Y_{ij}^*)_{i < j}$ are sampled independently, with Y_{ij}^* distributed as a Bernoulli random variable with parameter $q + \lambda \mathbf{1}\{z_i = z_j\}$, where $\lambda = p - q > 0$.

To get simpler formulas, in the remaining of the paper, we will work with the quasi-centered adjacency matrix

$$Y_{ij} = (Y^*)_{ij} - q, \text{ for any } 1 \leq i < j \leq n. \quad (3)$$

We will also use the notation

$$\bar{q} = q(1 - q), \quad \text{and} \quad \bar{p} = \bar{q} + \lambda(1 - 2q). \quad (4)$$

Community recovery. Our goal is to recover the partition $\mathcal{C}^* = \{\mathcal{C}_1^*, \dots, \mathcal{C}_K^*\}$ of $[n]$ induced by the community assignment $z_1, \dots, z_n$, i.e.

$$\mathcal{C}_k^* := \{i \in [n] : z_i = k\}, \quad k = 1, \dots, K.$$

We define the membership matrix

$$[x_{ij}]_{1 \leq i < j \leq n} = [\mathbf{1}\{i, j \text{ are in the same community}\} - 1/K]_{1 \leq i < j \leq n}.$$

If, for each $1 \leq i < j \leq n$, we produce an estimator $\hat{x}_{ij}$ equal to x_{ij} with probability $1 - o(n^{-2})$, then the connected components of $\hat{x}$ coincide with $\mathcal{C}^*$ with probability $1 - o(1)$. Hence, we focus henceforth on the problem of estimating x_{ij} . Our algorithm is based on motif counting, introduced below.

Motif counting. Let $G = (V, E)$ be a graph, henceforth referred as motif, on a node set $V = \{v_1, \dots, v_D\}$. Fix $1 \leq i < j \leq n$ and denote Π_{ij} the set of injections $\pi : V \rightarrow \{1, \dots, n\}$ fulfilling $\pi(v_1) = i$ and $\pi(v_2) = j$. Then, we define the motif counting

$$P_G = \sum_{\pi \in \Pi_{ij}} P_{G, \pi}, \quad \text{with} \quad P_{G, \pi}(Y) = \prod_{(v, v') \in E} Y_{\pi(v), \pi(v')}.$$

We refer to P_G as a motif counting since, when applied to the vanilla adjacency matrix Y^* , the polynomial $P_G(Y^*)$ merely counts the occurrence in Y^* of the motif G “attached” at i and j .

2.2 Conjecture 1.4 from [CGV25]

Our goal is to prove Conjecture 1.4 of [CGV25] that predicts that community recovery is possible above the threshold (2) through the enumeration of specific motifs satisfying certain structural properties. Before recalling the details of the conjecture, we provide some heuristic arguments and sketchy computations motivating it.

To start with, we notice that the signal Condition (2) can be written $\lambda \geq_{\log} \left(\frac{\lambda}{K}\right)^{1 - \log_n(K)} \vee q^{1 - \log_n(K)}$ or equivalently

$$\lambda \geq_{\log} \left(\frac{K}{n}\right) \vee q^{1 - \log_n(K)}. \quad (5)$$

In the density regime $q \asymp n^{-1/r}$ with $r > 1$, under the condition $\lambda \geq_{\log} q^{1 - \log_n(K)}$, we have

$$\lambda^r \geq_{\log} q^r K \asymp \frac{K}{n}. \quad (6)$$

Combining (6) and (5), we then observe that when $q \asymp n^{-1/r}$ with $r > 1$ the signal Condition (2) reduces to

$$\lambda \geq_{\log} q^{1-\log_n(K)}. \quad (7)$$

Let us motivate the structural properties of G appearing in the conjecture. First of all, we can readily check that for non-connected motifs, the count $P_G(Y)$ has zero correlation with x_{ij} . Hence, we restrict our attention to connected motifs G . Introducing the notation

$$\mathbb{P}_{ij} := \mathbb{P}[\cdot | z_i = z_j], \quad \text{and} \quad \mathbb{P}_{ij\neq} := \mathbb{P}[\cdot | z_i \neq z_j], \quad (8)$$

a direct computation gives $\mathbb{E}_{ij\neq}[P_G] = 0$. So, for testing $z_i = z_j$ against $z_i \neq z_j$ with a small error, we seek for a connected motif G fulfilling

$$\mathbb{E}_{ij}[P_G]^2 \gg \text{var}_{ij}(P_G) \vee \text{var}_{ij\neq}(P_G). \quad (9)$$

The expectation can be simply evaluated (see Equations (28) and (30))

$$\mathbb{E}_{ij}[P_G] \asymp \left(\frac{n\lambda^{|E|/(|V|-2)}}{K} \right)^{|V|-2} = \left(\frac{n\lambda^r}{K} \right)^{|V|-2}, \quad (10)$$

with $r := |E|/(|V|-2)$. So, according to (6), we have $\mathbb{E}_{ij}[P_G] \gtrsim 1$. Given a labeling π of V , we write $\pi(G)$ for the corresponding labelled graph with nodes $\pi(V)$ and edges $\pi(E) = \{(\pi(u), \pi(v)) : (u, v) \in E\}$. We define $|E^\neq|$ as the number of edges in $\pi(G)$ between nodes of distinct communities. A direct computation gives

$$\mathbb{E}_{ij}[P_{G,\pi}^2] = \mathbb{E}_{ij} \left[\prod_{\substack{(i,j) \in \pi(E) \\ z_i = z_j}} \bar{p} \prod_{\substack{(i,j) \in \pi(E) \\ z_i \neq z_j}} \bar{q} \right] = \mathbb{E}_{ij} [\bar{p}^{|E|-|E^\neq|} \bar{q}^{|E^\neq|}] \asymp \lambda^{|E|} \mathbb{E}_{ij} \left[\left(\frac{q}{\lambda} \right)^{|E^\neq|} \right], \quad (11)$$

where $\bar{p}, \bar{q}$ are defined in (4), with $\bar{p} \sim \lambda$ since $\lambda \geq_{\log} K^{1/r} q$ according to (6). As explained in [CGV25], we expect that the leading terms in the variance decomposition

$$\text{var}_{ij}(P_G) = \sum_{\pi, \pi'} \text{cov}_{ij}(P_{\pi, G}, P_{\pi', G}),$$

are those for which $\pi(G) = \pi'(G)$. Writing $|\text{Aut}(G)| \leq |V|^{|V|}$ for the number of automorphisms of G , and recalling (11) and $|E| = r(|V|-2)$, we then expect to have

$$\text{var}_{ij}(P_G) \vee \text{var}_{ij\neq}(P_G) \stackrel{?}{\lesssim} \sum_{\pi \in \Pi_V} |\text{Aut}(G)| \text{var}_{ij}(P_{G,\pi}) \lesssim (n\lambda^r)^{|V|-2} \mathbb{E}_{ij} \left[(q/\lambda)^{|E^\neq|} \right], \quad (12)$$

where $\lesssim$ hides some factors depending on $|E|$ and $|V|$ only.

Let $\ell(G)$ denote the number of distinct communities in $\pi(G)$. Combining (12) with (6) and $\mathbb{P}_{ij}[\ell(G) = \ell] \asymp K^{-(|V|-1-\ell)}$, we then get

$$\text{var}_{ij}(P_G) \vee \text{var}_{ij\neq}(P_G) \lesssim \left(\frac{n\lambda^r}{K} \right)^{|V|-2} \sum_{\ell=1}^{|V|-1} K^{\ell-1} \mathbb{E}_{ij} \left[K^{-|E^\neq|/r} \mathbb{1}[\ell(G) = \ell] \right]. \quad (13)$$

Taking for granted that the upper-bound in (13) is valid up to constants or log factors, comparing (13) and (10), a test based on P_G will fulfill (9), as soon as the sum in (13) remains small. This will occur in particular if the motif G fulfills $|E^\neq| \geq r(\ell(G) - 1)$ for any partition of the nodes into $\ell(G)$ communities. This leads to the following conjecture of [CGV25].

Conjecture 2.1 (Conjecture 1.4 in [CGV25]). *Let $G = (V, E)$ be a connected graph, with $(v_1, v_2) \notin E$, and such that for any partition of V in ℓ groups, with v_1, v_2 in the same group, the number $|E^\neq|$ of edges in E between distincts groups fulfills*

$$|E^\neq| \geq r(\ell - 1) \quad (14)$$

with $r := |E|/(|V|-2)$. Then, if $q \asymp n^{-1/r}$ and $\lambda \geq_{\log} q^{1-\log_n(K)}$, we can recover the communities in the SBM with an algorithm based on P_G , counting of the occurrence of the motif G .

In the next section, for any rational $r > 1$,

- we construct a connected motif G fulfilling the key condition (14) ;
- we prove that when $q \asymp n^{-1/r}$ and (2) hold, we can recover the communities in the SBM with an algorithm based on P_G .

The family of motifs G is based on the blow-up of a cycle. Writing $r = \gamma + a$ with γ an integer larger than 1 and a a rational in $(0, 1)$, the motif is constructed by considering a γ -blow-up of a cycle of length κ , that we link evenly to i and j with $a\kappa\gamma$ edges —called fasteners in the following. The length κ of the cycle is chosen in such a way that $a\kappa\gamma$ is a large enough even integer. Let us describe with more details this construction.

3 Main Result

3.1 Construction of the blow-up graph with fasteners

Consider any rational number $a \in (0, 1)$, any positive integer γ . Fix $\kappa \geq 3 \vee (2a^{-1})$ an integer such that $a\kappa\gamma$ is an even integer. In the following, we construct the blow-up graph with fasteners $G_{\kappa, \gamma, a} = (V, E)$.

Let C_κ be a simple cycle on vertices $[\kappa]$. For each vertex $\omega \in [\kappa]$, define a *layer*

$$L_\omega := \{v_{\omega, t} : t = 1, \dots, \gamma\},$$

consisting of γ vertices that represent the “blow-up” of the cycle node ω . The collection of all such vertices forms the *cycle nodes* of the graph:

$$V_{\text{cyc}} := \bigcup_{\omega=1}^{\kappa} L_\omega, \quad |V_{\text{cyc}}| = \kappa\gamma.$$

Between every two consecutive layers L_ω and $L_{\omega+1}$ (with the convention $L_{\kappa+1} = L_1$), we insert a complete bipartite graph:

$$E_{\text{cyc}} := \bigcup_{\omega=1}^{\kappa} \{(v_{\omega, t}, v_{\omega+1, t'}) : 1 \leq t, t' \leq \gamma\},$$

so that every vertex in a layer L_ω is connected to all γ vertices in $L_{\omega-1}$ and all γ vertices in $L_{\omega+1}$. This produces a total of $|E_{\text{cyc}}| = \kappa\gamma^2$ edges and the graph induced by $G_{\text{cyc}} = (V_{\text{cyc}}, E_{\text{cyc}})$ is 2γ -regular. We refer to G_{cyc} as the cycle part of the graph G .

We then introduce the two additional *distinguished* vertices v_1 and v_2 , and define the set of nodes $V := V_{\text{cyc}} \cup \{v_1, v_2\}$. We connect v_1, v_2 to V_{cyc} in such a way that $|E| = r(|V| - 2) = (\gamma + a)(|V| - 2)$. Accordingly, we connect v_1 or v_2 to each node in V_{cyc} with frequency a . We denote by $V_{\text{fst}} \subseteq V_{\text{cyc}}$ the cyclic vertices to be connected to v_1 or v_2 , which we call the *fastener nodes*. Informally, the set V_{fst} whose cardinality is equal to $a\kappa\gamma$ is defined such that the nodes V_{fst} are as evenly spread as possible. More formally, for each layer $\omega = 1, \dots, \kappa$, we define the integers s_ω by

$$s_0 = 0, \quad s_\omega = \lfloor a\omega\gamma \rfloor - (s_0 + \dots + s_{\omega-1}). \quad (15)$$

Note, that $s_\omega \in \{\lfloor a\gamma \rfloor, \lceil a\gamma \rceil\}$. In layer L_ω , exactly s_ω vertices are selected into V_{fst} . By enumerating V_{fst} in the lexicographic order and alternatively assigning these nodes to two sets $V_{\text{fst},1}$ or to $V_{\text{fst},2}$, we partition V_{fst} into two subsets of size $a\kappa\gamma/2$ —which is an integer, as we have assumed $a\kappa\gamma$ to be an even integer.

Finally, we introduce the collection E_{fst} of *fastener edges* by

$$E_{\text{fst}} := \{(v_1, v) : v \in V_{\text{fst},1}\} \cup \{(v_2, v) : v \in V_{\text{fst},2}\}.$$

The complete edge set of the graph is then $E := E_{\text{cyc}} \cup E_{\text{fst}}$. This construction is illustrated in Figure 1.

Interestingly, we observe

$$|V| = \kappa\gamma + 2 ; |E| = \kappa\gamma^2 + a\kappa\gamma.$$

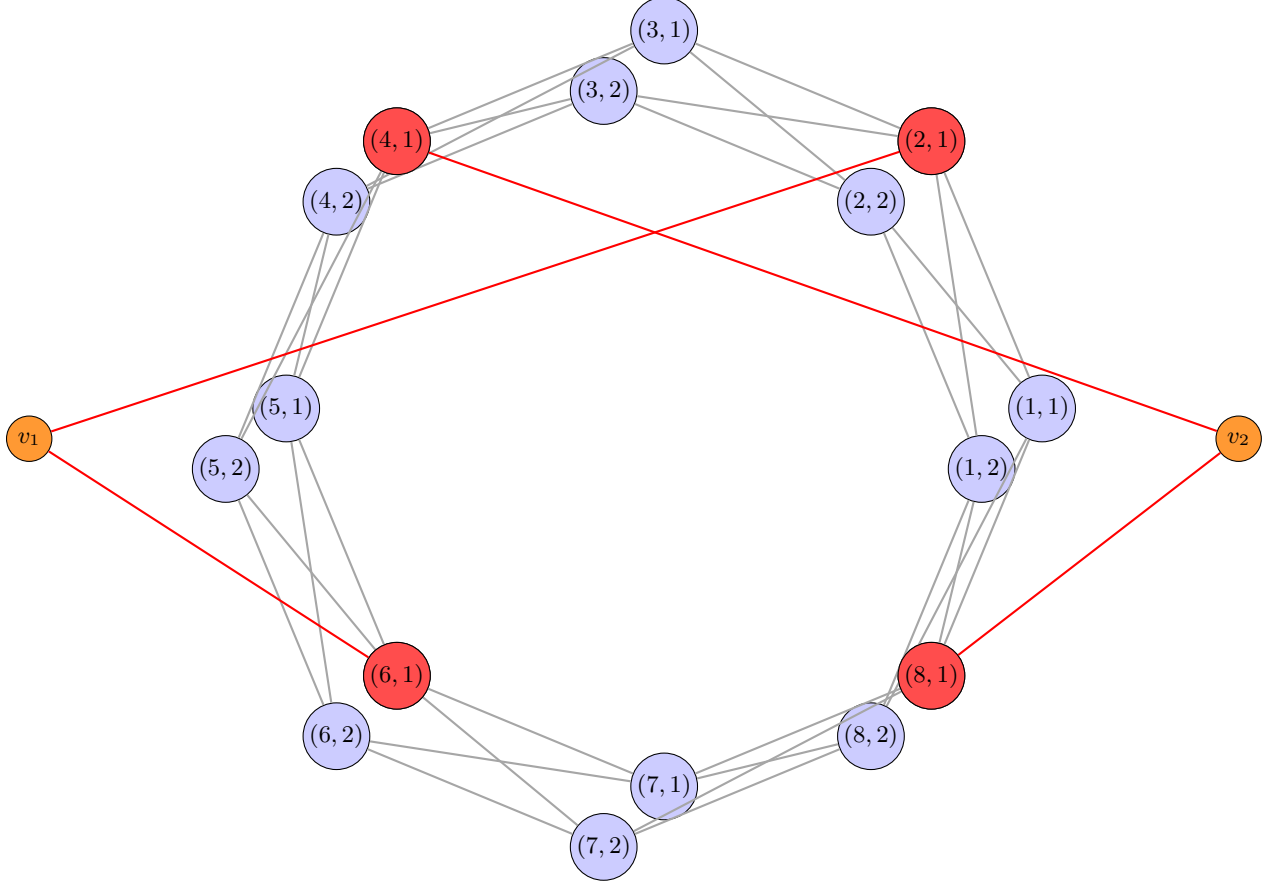


Figure 1: Illustration of the blow-up graph with fasteners $G_{\kappa,\gamma,a}$, in the case where $\gamma = 2, \kappa = 8, a = 0.25$. The distinguished nodes v_1, v_2 are in orange, the *fastener nodes* (in V_{fst}) are in red, the other cycle nodes are in blue. The *fastener edges* (in E_{fst}) are in red, while the other edges are in gray.

so that the graph $G_{\kappa,\gamma,a}$ satisfies

$$\frac{|E|}{|V| - 2} = \gamma + a .$$

The following proposition ensures that the motif $G_{\kappa,\gamma,a}$ fulfills the Condition (14) required in Conjecture 2.1. We recall that $\kappa \geq (3 \vee 2/a)$ is such $a\kappa\gamma$ is an even integer.

Proposition 3.1. *Fix any positive integer $I \leq |V_{\text{cyc}}|$ and consider any partition of V into $I+1$ communities such that both v_1 and v_2 are in the same community. Define $E^\neq \subset E$ as the set of edges between nodes of distinct communities. Then, we have*

$$|E^\neq| \geq I(\gamma + a) .$$

3.2 Counting Blow-up Motifs

Consider a blow-up graph with fasteners $G = G_{\kappa,\gamma,a}$. For $i < j$, we remind that $\Pi_{i,j}$ is the set of injections $\pi : V \rightarrow \{1 \dots, n\}$ such that $\pi(v_1) = i$ and $\pi(v_2) = j$, and we set

$$R_{ij} = \sum_{\pi \in \Pi_{i,j}} P_{G,\pi}(Y), \quad \text{with} \quad P_{G,\pi}(Y) = \prod_{(v,v') \in E} Y_{\pi(v),\pi(v')}, \quad (16)$$

where Y is the “centered” adjacency matrix (3).

In the following proposition, we control the mean and the variance of R_{ij} both when $z_i = z_j$ and when $z_i \neq z_j$. We recall that the conditional probabilities $\mathbb{P}_{ij}$ and $\mathbb{P}_{ij}$ are defined in (8).

Proposition 3.2. *Assume that $q \leq 1/4$, $q + 2\lambda \leq 1$, and $n \geq 2\kappa\gamma + 4$. We also assume that $\kappa\gamma$ is an even integer and that $\kappa \geq 3 \vee 2/a$. Let $i < j$ and let R_{ij} be defined by (16). We have*

$$\mathbb{E}_{ij} [R_{ij}] = \frac{(n-2)!}{(n-\kappa\gamma-2)!} \left(\frac{\lambda^{\gamma+a}}{K} \right)^{\kappa\gamma}, \quad \mathbb{E}_{ij} [R_{ij}] = 0. \quad (17)$$

In addition, if for some $\rho > 1$,

$$\left(\frac{\lambda^2}{2\bar{q}} \right)^{\gamma+a} \geq \frac{2K^2(\kappa\gamma)^5 \rho}{n}; \quad \lambda^{\gamma+a} \geq \frac{2K(\kappa\gamma)^5 \rho}{n} \quad (18)$$

then, we have

$$\text{var}_{ij}(R_{ij}) \bigvee \text{var}_{ij}(R_{ij}) \leq \frac{4}{\rho} \cdot \mathbb{E}_{ij}^2 [R_{ij}]. \quad (19)$$

As for the clique counting or self-avoiding path counting problems in [CGV25], relying on R_{ij} alone together with a Markov type bound is not sufficient. For this reason, we rely again on a Median-of-Means post-processing and we use the same notation as in the latter. In particular, we fix $\Lambda = 24 \log(n)$ and we assume for simplicity that $(n-2)/\Lambda$ is an integer. Recall the definition of $N = (n-2)/\Lambda + 2$ and let $J^{(1)}, \dots, J^{(\Lambda)}$ be a partition of $[n] \setminus \{i, j\}$ into Λ disjoint parts. For $\ell = 1, \dots, \Lambda$, we define $\Pi_{i,j}^{(\ell)}$ has the set of injections $\pi : V \rightarrow \{i, j\} \cup J^{(\ell)}$, such that $\pi(v_1) = i$ and $\pi(v_2) = j$. Then, as in (31) in [CGV25] for clique counts, we introduce the blow-up count $R_{ij}^{(\ell)} := \sum_{\pi \in \Pi_{i,j}^{(\ell)}} P_{G,\pi}(Y)$, and we define M_{ij} as a median of the set $\{R_{ij}^{(1)}, \dots, R_{ij}^{(\Lambda)}\}$. We estimate $x_{ij} = \mathbf{1}_{z_i=z_j} - \frac{1}{K}$ by

$$\hat{x}_{ij} = \mathbf{1} \left\{ M_{ij} > \frac{(N-2)!}{2(N-\kappa\gamma-2)!} \left(\frac{\lambda^{\gamma+a}}{K} \right)^{\kappa\gamma} \right\} - \frac{1}{K}, \quad \text{where } N = \frac{n-2}{24 \log(n)} + 2. \quad (20)$$

We can now state our main result.

Theorem 3.3. *There exists a numerical constant c such that the following holds. Assume that $q \leq 1/4$, $q + 2\lambda \leq 1$, that $N := 2 + (n-2)/(24 \log(n))$ is an integer, and $N \geq 2\kappa\gamma + 4$, and that $\kappa\gamma a$ is an even integer. Provided that*

$$\left(\frac{\lambda^2}{2\bar{q}} \right)^{\gamma+a} \geq c \frac{K^2(\kappa\gamma)^5 \log(n)}{n}; \quad \lambda^{\gamma+a} \geq c \frac{K(\kappa\gamma)^5 \log(n)}{n},$$

we have for $\hat{x}_{ij}$ defined by (20) that $\mathbb{P}(\hat{x}_{ij} = x_{ij}) \geq 1 - n^{-3}$.

Proof of Theorem 3.3. The proof follows exactly the same lines as that of Theorem 2.5 in [CGV25] to the difference that we build upon Proposition 3.2 instead of Proposition 2.4 in [CGV25]. We skip the details. $\square$

Corollary 3.4. *Let $\bar{q} = n^{-1/r}$ for some $r = \gamma + \alpha/\beta$ where γ , α , and β are positive integers with $\alpha < \beta$. Consider the blow-up graph with fasteners $G_{\kappa,\gamma,a}$ with $\kappa = 2\beta\gamma$ and $a = \alpha/\beta$. When $n \geq c_0 \beta \gamma^2 \log(n)$, $\lambda \leq 1 - 2q$, and*

$$\lambda \geq w'_r \log^{1/r}(n) \left(\frac{\lambda}{K} + \bar{q} \right)^{1 - \log_n(K)}, \quad (21)$$

with w'_r depending only on r , then the estimator $\hat{x}$ based on the number of blow-up motifs recovers the communities with probability at least $1 - 1/n$.

If $q = n^{-1/r}$, with r a rational number, recovering the communities above the threshold (2) is feasible with a polynomial of fixed degree, only depending on r . When $r > 1$ is not a rational number, we can still count blow-up motif for some rational number $\bar{r}$ close to r . Indeed, consider any $\epsilon < 1$. There exists an

integer $\bar{\beta} \leq 2/\lceil \epsilon r^2 \rceil \vee 1$ and a rational number $\bar{r} = \bar{\gamma} + \bar{\alpha}/\bar{\beta}$ such that $|\bar{r} - r| \leq \epsilon r^2/2$. Here, we have $\bar{\gamma} = \lfloor r \rfloor$. Choosing $\bar{\kappa} = 2\bar{\beta}\bar{\gamma}$, we deduce from Theorem 3.3, that the estimator $\hat{x}$ based on counting occurrence of blow-up motifs $G_{\bar{\kappa}, \bar{\gamma}, \bar{\alpha}/\bar{\beta}}$ recovers the communities with probability at least $1 - 1/n$ as long as

$$\lambda \geq w_r'' \epsilon^{-5} \log^{1/r}(n) \left(\frac{\lambda}{K} + \bar{q} \right)^{1 - \log_n(K)} n^\epsilon .$$

The corresponding polynomial has a degree of the order ϵr . In particular, if we take $\epsilon = \log(\log(n))/\log(n)$, we establish that, for any irrational r , it is possible to recover the communities above the threshold (2) with a polynomial of degree $O(\log(n)/\log(\log(n)))$. Whether this polynomial can be computed (or well-approximated) in polynomial-time remains an open question.

4 Proof of Proposition 3.1

In this proof, we need additional notation. Let $V_{\text{cyc},0}$ denote the set of cycle nodes having the same community as v_1 and v_2 . For $k = 1, \dots, I$ let $V_{\text{cyc},k}$ be the set of cycle nodes of community k . For short, we write $V_{\text{cyc},\neq 0} = \cup_{k=1}^I V_{\text{cyc},k}$.

We consider separately cycle edges and fastener edges. For $k = 0, \dots, I$, we write $E_{\text{cyc},k}^\uparrow \subset E_{\text{cyc}}$ for the collection of cycle edges that are incident to a node in $V_{\text{cyc},k}$ and a node from a different community — these edges will be called *boundary* edges. For $k = 0, \dots, I$, we define $V_{\text{fst},k} := V_{\text{cyc},k} \cap V_{\text{fst}}$, the subset of fastener nodes within community k . Note that $|V_{\text{fst},k}|$ also corresponds to the number of fastener edges within community k . For short, we also write $V_{\text{fst},\neq 0} = \cup_{k=1}^I V_{\text{fst},k}$.

First, we consider the specific and trivial case where $I = 1$ and $V_{\text{cyc},1} = V_{\text{cyc}}$. Then, we obviously have

$$|E^\neq| = |V_{\text{fst}}| = a\kappa\gamma \geq \gamma + 1 \geq (\gamma + a)I ,$$

since $\kappa \geq 2/a$. We assume henceforth $V_{\text{cyc},k} \neq V_{\text{cyc}}$ for any k .

The proof mainly divides into the two following lemmas. The first one states that the boundary edges of any community that does not contain v_1 or v_2 — namely the sets of edges $E_{\text{cyc},k}^\uparrow$ for $k \in \{1, \dots, I\}$ — is of size at least 2γ .

Lemma 4.1. *If $|V_{\text{cyc},k}| < \kappa\gamma$, i.e. $V_{\text{cyc},k} \neq V_{\text{cyc}}$, we have*

$$|E_{\text{cyc},k}^\uparrow| \geq 2\gamma .$$

The second lemma lower bounds the sum of twice the number of fasteners $V_{\text{fst},\neq 0}$ that do not belong to the same community as v_1 or v_2 , plus the number of boundary edges for the community of v_1, v_2 , namely $E_{\text{cyc},0}^\uparrow$. If the nodes $V_{\text{cyc},\neq 0}$ were sampled uniformly, then $|V_{\text{fst},\neq 0}|$ would be of the order of $a|V_{\text{cyc},\neq 0}|$. However, the number of fasteners in $V_{\text{fst},\neq 0}$ can be much smaller for unfavorable configurations such that fastener nodes are in the same community as v_1 or v_2 . Nevertheless, if this happens, then the number of boundary edges in $E_{\text{cyc},0}^\uparrow$ must be high accordingly to compensate this drop in the number of fasteners in $V_{\text{fst},\neq 0}$.

Lemma 4.2. *We have*

$$|E_{\text{cyc},0}^\uparrow| + 2|V_{\text{fst},\neq 0}| \geq 2a|V_{\text{cyc},\neq 0}| \geq 2aI . \quad (22)$$

Let us explain how Proposition 3.1 follows from Lemmas 3.1 and 4.1. First, we observe that an edge between two communities is either a fastener edge and is therefore incident to $V_{\text{fst},\neq 0}$ or is a cycle edge and therefore both arises in $E_{\text{cyc},k}^\uparrow$ and $E_{\text{cyc},k'}^\uparrow$ for $k \neq k'$ being the two communities to which the nodes of the edge belong. Hence, we have the following decomposition

$$2|E^\neq| \geq \sum_{k=0}^I |E_{\text{cyc},k}^\uparrow| + 2|V_{\text{fst},\neq 0}| ,$$

and Lemmas 4.1 and 4.2 ensure that

$$2|E^\neq| \geq 2\gamma I + 2aI ,$$

which concludes the proof of the Proposition 3.1.

Proof of Lemma 4.1. We consider three cases, (i) at least a layer does not intersect $V_{\text{cyc},k}$; (ii) each layer intersects $V_{\text{cyc},k}$ and $|V_{\text{cyc},k}| \leq (\kappa - 1)\gamma$; (iii) each layer intersects $V_{\text{cyc},k}$ and $|V_{\text{cyc},k}| > (\kappa - 1)\gamma$.

Case (i): at least a layer does not intersect $V_{\text{cyc},k}$. In this case there exist two layers i_1 and i_2 (not necessarily distinct) such that

$$L_{i_1} \cap V_{\text{cyc},k} \neq \emptyset; \quad L_{i_1-1} \cap V_{\text{cyc},k} = \emptyset; \quad L_{i_2} \cap V_{\text{cyc},k} \neq \emptyset; \quad L_{i_2+1} \cap V_{\text{cyc},k} = \emptyset,$$

with the convention that $L_0 = L_\kappa$ and $L_{\kappa+1} = L_1$. There are γ edges between L_{i_1-1} and one node of $V_{\text{cyc},k}$ in L_{i_1} . Similarly, there are γ edges between L_{i_2+1} and one node of $V_{\text{cyc},k}$ in L_{i_2} . Since $\kappa \geq 3$, these edges are distinct and we deduce $|E_{\text{cyc},k}^\uparrow| \geq 2\gamma$.

Case (ii): each layer intersects $V_{\text{cyc},k}$, with $|V_{\text{cyc},k}| \leq (\kappa - 1)\gamma$. Since in this case each layer intersects $V_{\text{cyc},k}$, it follows from the definition of the graph that each node v in $V_{\text{cyc}} \setminus V_{\text{cyc},k}$ is connected to $V_{\text{cyc},k}$ by at least two edges corresponding to nodes of $V_{\text{cyc},k}$ in the previous layer and in the following layer of that of v . As a consequence,

$$|E_{\text{cyc},k}^\uparrow| \geq 2|V_{\text{cyc}} \setminus V_{\text{cyc},k}| = 2[\kappa\gamma - |V_{\text{cyc},k}|] \geq 2\gamma,$$

as soon as $|V_{\text{cyc},k}| \leq (\kappa - 1)\gamma$.

Case (iii): each layer intersects $V_{\text{cyc},k}$, with $|V_{\text{cyc},k}| > (\kappa - 1)\gamma$. It remains to consider the case where $|V_{\text{cyc},k}| \in [(\kappa - 1)\gamma + 1, \kappa\gamma - 1]$ which is non empty only if $\gamma > 1$. Then, we consider the complementary set $V' = V_{\text{cyc}} \setminus V_{\text{cyc},k}$ which satisfies $|V'| \leq \gamma - 1 \leq \kappa(\gamma - 1)$. Since $V_{\text{cyc},k}$ and V' have a symmetric role for counting $|E_{\text{cyc},k}^\uparrow|$, we can apply the same arguments as in Case (i) and (ii), switching the role of $V_{\text{cyc},k}$ and V' . Accordingly, $|E_{\text{cyc},k}^\uparrow| \geq 2\gamma$, and the proof of Lemma 4.1 is complete. $\square$

Proof of Lemma 4.2. Write $\mathcal{L}_0 = \{\omega \in [\kappa] : L_\omega \cap V_{\text{cyc},0} \neq \emptyset\}$ the collection of layers that intersect $V_{\text{cyc},0}$. To start with, we consider three cases (a) $\mathcal{L}_0 = \emptyset$, (b) $\mathcal{L}_0 = [\kappa]$, and (c) $|\mathcal{L}_0| \in [2; \kappa - 1]$.

Case (a): $\mathcal{L}_0 = \emptyset$. In this case no fastener node belongs to the community of v_1, v_2 , therefore we have

$$|V_{\text{fst}, \neq 0}| = |V_{\text{fst}}| = a|V_{\text{cyc}}| \geq a|V_{\text{cyc}, \neq 0}|,$$

and (22) holds.

Case (b): $\mathcal{L}_0 = [\kappa]$. In this case we observe as in the proof of Lemma 4.1 (case (ii)), that any node v in $V_{\text{cyc}} \setminus V_{\text{cyc},0}$ is connected to two nodes in $V_{\text{cyc},0}$, one in the previous layer and one in the following layer. Thus, we get

$$|E_{\text{cyc},0}^\uparrow| \geq 2 \sum_{k=1}^I |V_{\text{cyc},k}| = 2|V_{\text{cyc}, \neq 0}| \geq 2a|V_{\text{cyc}, \neq 0}|,$$

and (22) holds.

Case (c): $|\mathcal{L}_0| \in [2; \kappa - 1]$. This is the intermediary case between cases (a) and (b). Consider $[\kappa]$ as the nodes of a cycle graph and $\mathcal{L}_0$ as a subset of nodes of this graph. Write $\#\text{CC}(\mathcal{L}_0) \geq 1$ for the number of connected components of the subgraph induced by $\mathcal{L}_0$.

Step 1: lower bound of $|E_{\text{cyc},0}^\uparrow|$. We partition $\mathcal{L}_0$ into $\mathcal{L}_{0,\text{in}}$, $\mathcal{L}_{0,\text{bd}}$, and $\mathcal{L}_{0,\text{sg}}$ where $\mathcal{L}_{0,\text{in}}$ are internal layers of $\mathcal{L}_0$ that is, they are not at the boundary of a connected component of $\mathcal{L}_0$, $\mathcal{L}_{0,\text{sg}}$ correspond to connected components that are singletons, and $\mathcal{L}_{0,\text{bd}}$ stands for all the remaining boundary layers.

Fix any $\omega \in \mathcal{L}_{0,\text{in}}$ and any node $v \in L_\omega \setminus V_{\text{cyc},0}$. This node v is connected by at least two edges to $V_{\text{cyc},0}$ (at least one node in the previous layer and one node in the following layer). Fix any $\omega \in \mathcal{L}_{0,\text{bd}}$. For any node $v \in L_\omega \setminus V_{\text{cyc},0}$, we also observe that v is connected by at least one edge to $V_{\text{cyc},0}$. Also, for any $v \in L_\omega \cap V_{\text{cyc},0}$, there are γ edges between v and nodes in a layer that does not belong to $\mathcal{L}_0$. Finally, we consider any $\omega \in \mathcal{L}_{0,\text{sg}}$ and any $v \in L_\omega \cap V_{\text{cyc},0}$. This node is connected to 2γ nodes outside the layers in $\mathcal{L}_0$.

Gathering these bounds, we get

$$\begin{aligned}
|E_{\text{cyc},0}^\uparrow| &\geq 2 \sum_{\omega \in \mathcal{L}_{0,\text{in}}} |L_\omega \setminus V_{\text{cyc},0}| + \sum_{\omega \in \mathcal{L}_{0,\text{bd}}} [|L_\omega \setminus V_{\text{cyc},0}| + \gamma] + 2 \sum_{\omega \in \mathcal{L}_{0,\text{sg}}} \gamma \\
&\geq 2 \sum_{\omega \in \mathcal{L}_0} |L_\omega \setminus V_{\text{cyc},0}| + 2\#\text{CC}(\mathcal{L}_0) ,
\end{aligned} \tag{23}$$

where we used that, for any $\omega \in \mathcal{L}_0$, $|L_\omega \setminus V_{\text{cyc},0}| \leq \gamma - 1$, and where we recall that $\#\text{CC}(\mathcal{L}_0)$ is the number of connected components of $\mathcal{L}_0$.

Step 2: lower bound of $|V_{\text{fst},\neq 0}|$. We now focus on the set $[\kappa] \setminus \mathcal{L}_0$ of layers that do not intersect $V_{\text{cyc},0}$. We start from

$$|V_{\text{fst},\neq 0}| \geq \sum_{\omega \in [\kappa] \setminus \mathcal{L}_0} |V_{\text{fst}} \cap L_\omega|.$$

Although $|V_{\text{fst}} \cap L_\omega| \in [\lfloor a\gamma \rfloor, \lceil a\gamma \rceil]$ for each ω , we have to be slightly more careful to control $|V_{\text{fst},\neq 0}|$. The set $[\kappa] \setminus \mathcal{L}_0$ decomposes into $\#\text{CC}(\mathcal{L}_0)$ connected components, since the set $\mathcal{L}_0$ decomposes in $\#\text{CC}(\mathcal{L}_0)$ connected components. We write $r_1, \dots, r_{\#\text{CC}(\mathcal{L}_0)}$ for their respective number of layers. Into the i -th connected components, there are at least $\lfloor ar_i \gamma \rfloor$ fasteners. From this and from $\sum_i r_i \geq \kappa - |\mathcal{L}_0|$, we deduce that

$$\begin{aligned}
|V_{\text{fst},\neq 0}| &\geq \sum_{i=1}^{\#\text{CC}(\mathcal{L}_0)} \lfloor ar_i \gamma \rfloor \geq a [\kappa - |\mathcal{L}_0|] \gamma - \#\text{CC}(\mathcal{L}_0) \\
&= a \sum_{\omega \in [\kappa] \setminus \mathcal{L}_0} |L_\omega \setminus V_{\text{cyc},0}| - \#\text{CC}(\mathcal{L}_0) .
\end{aligned} \tag{24}$$

Combining (23) and (24), we obtain

$$\begin{aligned}
|E_{\text{cyc},0}^\uparrow| + 2|V_{\text{fst},\neq 0}| &\geq 2a \left[\sum_{\omega \in [\kappa]} |L_\omega \setminus V_{\text{cyc},0}| \right] = 2a \left[\sum_{\omega \in [\kappa]} |L_\omega \cap V_{\text{cyc},\neq 0}| \right] \\
&\geq 2a|V_{\text{cyc},\neq 0}| ,
\end{aligned}$$

which concludes the proof of Lemma 4.2. $\square$

5 Proof of Proposition 3.2

This proposition is a straightforward consequence of the following result.

Proposition 5.1. *Assume that $q \leq 1/4$, $q + 2\lambda \leq 1$, $n \geq 2\kappa\gamma + 4$, and set $\bar{p} = \bar{q} + \lambda(1 - 2q)$. We also assume that $a\kappa\gamma$ is an even integer and $\kappa \geq 3 \vee 2/a$. Then, for any $1 \leq i < j \leq n$, we have*

$$\mathbb{E}_{ij} [R_{ij}] = \frac{(n-2)!}{(n-\kappa\gamma-2)!} \left(\frac{\lambda^{\gamma+a}}{K} \right)^{\kappa\gamma} , \tag{25}$$

$$\mathbb{E}_{ij} [R_{ij}] = 0, \tag{26}$$

$$\begin{aligned}
\text{var}_{ij}(R_{ij}) \vee \text{var}_{ij}(R_{ij}) &\leq \mathbb{E}_{12}^2[R_{12}](\kappa\gamma)^2 \left[\frac{2(\kappa\gamma)^3 K^2}{n} \left(\frac{\bar{q}}{\lambda^2} \vee \frac{\bar{q}}{\bar{p}} \right)^{\gamma+a} + \frac{2(\kappa\gamma)^3 K}{n} \left(\frac{\bar{p}}{\lambda^2} \vee 1 \right)^{\gamma+a} \right. \\
&\quad \left. + \left[\frac{2(\kappa\gamma)^3 K^2}{n} \left(\frac{\bar{q}}{\lambda^2} \vee \frac{\bar{q}}{\bar{p}} \right)^{\gamma+a} \right]^{\kappa\gamma} + \left[\frac{2(\kappa\gamma)^3 K}{n} \left(\frac{\bar{p}}{\lambda^2} \vee 1 \right)^{\gamma+a} \right]^{\kappa\gamma} \right] .
\end{aligned} \tag{27}$$

5.1 Notation and preliminaries

To control the expectation and the variance of R_{ij} , we need to introduce some graph notation as in Appendix A of [CGV25]. For a motif $G = (V, E)$ and a labeling $\pi \in \Pi_{ij}$, we define the labeled graph $\pi(G)$ as the graph with node set $\{\pi(v) : v \in V\}$ and edge set $\{(\pi(v), \pi(v')) : (v, v') \in E\}$. Given two labelings $\pi^{(1)}$ and $\pi^{(2)} \in \Pi_{ij}$, the labeled merged graph $G_\cup = (V_\cup, E_\cup)$ is defined as the union of $\pi^{(1)}(G^{(1)})$ and $\pi^{(2)}(G^{(2)})$, with the convention that two same edges are merged into a single edge. Similarly, we define the intersection graph $G_\cap = (V_\cap, E_\cap)$ and the symmetric difference graph $G_\Delta = (V_\Delta, E_\Delta)$ so that $E_\Delta = E_\cup \setminus E_\cap$. Here, $V_\cap$ (resp. V_Δ) is the set of nodes induced by the edges $E_\cap$ (resp. E_Δ) so that $G_\cap$ (resp. G_Δ) does not contain any isolated node.

The two following expressions are the starting point of our analysis. They easily derive from the definition (3) of Y :

$$\mathbb{E}[P_{G,\pi}] = \mathbb{E}\left[\prod_{(v,v') \in E} Y_{\pi(v)\pi(v')}\right] = \mathbb{E}\left[\prod_{(v,v') \in E} (\lambda \mathbf{1}_{z_{\pi(v)}=z_{\pi(v')}})\right], \quad (28)$$

$$\mathbb{E}[P_{G,\pi^{(1)}} P_{G,\pi^{(2)}}] = \mathbb{E}\left[\prod_{(i,j) \in E_\Delta} Y_{ij} \prod_{(i,j) \in E_\cap} Y_{ij}^2\right] = \mathbb{E}\left[\prod_{(i,j) \in E_\Delta} (\lambda \mathbf{1}_{z_i=z_j}) \prod_{(i,j) \in E_\cap} \bar{q} \left(\frac{\bar{p}}{\bar{q}}\right)^{\mathbf{1}_{z_i=z_j}}\right]. \quad (29)$$

Without loss of generality, we assume in the following that $(i, j) = (1, 2)$.

5.2 Results under the distribution $\mathbb{P}_{12}$

In what follows, we consider the graph $G := G_{\kappa,\gamma,a} = (V, E)$. We will study respectively the expectation and variance of the associated polynomial, both under $\mathbb{P}_{12}$ and under $\mathbb{P}_{\mathcal{Y}2}$.

Mean under $\mathbb{P}_{12}$. Since the graph G is connected, we have from Equation (28)

$$\mathbb{E}_{12}[P_{G,\pi}(Y)] = \lambda^{|E|} \mathbb{E}_{12}\left[\prod_{(i,j) \in \pi(E)} \mathbf{1}_{z_i=z_j}\right] = \frac{1}{K^{|V|-2}} \lambda^{|E|} = \frac{1}{K^{\kappa\gamma}} \lambda^{(\gamma+a)\kappa\gamma}, \quad (30)$$

since $z_1 = z_2$ a.s. under $\mathbb{P}_{12}$. The identity (25) follows.

Variance under $\mathbb{P}_{12}$. Let us turn to proving (27), which is the main part of the proof. Fix $\pi^{(1)}, \pi^{(2)} \in \Pi_{12}$. We start by controlling $\mathbb{E}_{12}[P_{G,\pi^{(1)}} P_{G,\pi^{(2)}}]$. Let $u \in \{0, \dots, |V_{\text{cyc}}|\}$ be the integer so that

$$|\text{range}(\pi^{(1)}) \cap \text{range}(\pi^{(2)})| = 2 + u.$$

Recall that $V_\cap, V_\Delta, E_\cap, E_\Delta$ stand for the common and symmetric-difference vertex- and edge-sets. Following (29), we have

$$\mathbb{E}_{12}[P_{G,\pi^{(1)}} P_{G,\pi^{(2)}}] = \lambda^{|E_\Delta|} \bar{q}^{|E_\cap|} \mathbb{E}_{12}\left[\prod_{(i,j) \in E_\Delta} \mathbf{1}_{z_i=z_j} \prod_{\substack{(i,j) \in E_\cap \\ z_i=z_j}} \frac{\bar{p}}{\bar{q}}\right].$$

Case $u = 0$. Then, we have $E_\cap = \emptyset$, $|V_\Delta| = |V_\cup| = 2\kappa\gamma + 2$ and $|E_\Delta| = 2\kappa\gamma(\gamma + a)$. Therefore

$$\begin{aligned} \mathbb{E}_{12}[P_{G,\pi^{(1)}} P_{G,\pi^{(2)}}] &= \lambda^{2\kappa\gamma(\gamma+a)} \mathbb{P}_{12}[z_i = z_1 \text{ for } i \in V_\Delta] \\ &= \lambda^{2\kappa\gamma(\gamma+a)} K^{-2\kappa\gamma} \\ &= \mathbb{E}_{12}[P_{G,\pi^{(1)}}] \mathbb{E}_{12}[P_{G,\pi^{(2)}}], \end{aligned} \quad (31)$$

by (30).

Case $1 \leq u \leq \kappa\gamma$. In this case $|V_\cup| = 2\kappa\gamma + 2 - u$ and there are $\kappa\gamma - u$ vertices that appear only in $\pi^{(2)}(V) \setminus \pi^{(1)}(V)$. All these vertices belong to G_Δ and each such vertex is connected (within G_Δ) to at least one

vertex in the intersection $\pi^{(1)}(V) \cap \pi^{(2)}(V)$ because G is connected. Let $E_{\Delta}^{(2)} \subset E_{\Delta}$ be the set of edges of G_{Δ} that arise from $\pi^{(2)}[G]$. In particular, $E_{\Delta}^{(2)}$ contains all the edges that have at least one endpoint in $\pi^{(2)}(V) \setminus \pi^{(1)}(V)$. Conditionally on the communities of the vertices of $\pi^{(1)}(V)$, the product of indicators on $E_{\Delta}^{(2)}$ enforces the community of the $\kappa\gamma - u$ vertices in $\pi^{(2)}(V) \setminus \pi^{(1)}(V)$. Hence, we have

$$\mathbb{E}_{12} \left[\prod_{(i,j) \in E_{\Delta}^{(2)}} \mathbf{1}_{z_i = z_j} \middle| z_{\pi^{(1)}(V)} \right] \leq \frac{1}{K^{\kappa\gamma - u}} \quad \text{a.s.}$$

Therefore (conditioning and removing $E_{\Delta}^{(2)}$), we get

$$\mathbb{E}_{12} [P_{G, \pi^{(1)}} P_{G, \pi^{(2)}}] \leq \lambda^{|E_{\Delta}|} \bar{q}^{|E_{\cap}|} \frac{1}{K^{\kappa\gamma - u}} \mathbb{E}_{12} \left[\prod_{(i,j) \in E_{\Delta} \setminus E_{\Delta}^{(2)}} \mathbf{1}_{z_i = z_j} \prod_{\substack{(i,j) \in E_{\cap} \\ z_i = z_j}} \frac{\bar{p}}{\bar{q}} \right].$$

Now, define $E_{\Delta}^{(1)} := E_{\Delta} \setminus E_{\Delta}^{(2)}$. Observe that $E_{\Delta}^{(1)}$ together with $E_{\cap}$ form a partition of the edges of the labeled copy of G coming from $\pi^{(1)}$. Given the community assignments $(z_i)_{i \in \pi^{(1)}(V)}$, denote $\ell(z)$ the number of distinct communities in $\pi^{(1)}(V)$, and $|E^{\neq}|$ the number of edges in the graph between distinct communities. We have

$$\begin{aligned} \mathbb{E}_{12} [P_{G, \pi^{(1)}} P_{G, \pi^{(2)}}] &\leq \lambda^{|E_{\Delta}|} \bar{p}^{|E_{\cap}|} \frac{1}{K^{\kappa\gamma - u}} \mathbb{E}_{12} \left[\left(\frac{\bar{q}}{\bar{p}} \right)^{|E^{\neq}|} \prod_{(i,j) \in E_{\Delta}^{(1)}} \mathbf{1}_{z_i = z_j} \right] \\ &\leq \lambda^{|E_{\Delta}|} \bar{p}^{|E_{\cap}|} \frac{1}{K^{\kappa\gamma - u}} \mathbb{E}_{12} \left[\left(\frac{\bar{q}}{\bar{p}} \right)^{(\ell(z)-1)(\gamma+a)} \prod_{(i,j) \in E_{\Delta}^{(1)}} \mathbf{1}_{z_i = z_j} \right], \end{aligned} \quad (32)$$

where, importantly, we apply Proposition 3.1 in the second line. Consider the graph $\bar{\mathcal{G}}_{\Delta}^{(1)} = (\pi^{(1)}(V), E_{\Delta}^{(1)})$. In Equation (32) above, the condition $\prod_{(i,j) \in E_{\Delta}^{(1)}} \mathbf{1}_{z_i = z_j} = 1$ enforces that the nodes of each of the connected components of $\bar{\mathcal{G}}_{\Delta}^{(1)}$ are in the same community. Let cc denote the number of connected components of $\bar{\mathcal{G}}_{\Delta}^{(1)}$ when we remove those containing $\pi^{(1)}(v_1) = 1$ or $\pi^{(1)}(v_2) = 2$. Then, conditionally to $\prod_{(i,j) \in E_{\Delta}^{(1)}} \mathbf{1}_{z_i = z_j} = 1$, we have $(\ell(z) - 1) \in [0, \text{cc}]$ and for any $x \in [0, \text{cc}]$, we have

$$\mathbb{P}_{12} \left[\ell(z) - 1 = x \middle| \prod_{(i,j) \in E_{\Delta}^{(1)}} \mathbf{1}_{z_i = z_j} = 1 \right] \leq \text{cc}^{\text{cc}} K^{-(\text{cc} - x)}. \quad (33)$$

Since the condition $\prod_{(i,j) \in E_{\Delta}^{(1)}} \mathbf{1}_{z_i = z_j} = 1$ enforces that the nodes of each of the connected components of $\bar{\mathcal{G}}_{\Delta}^{(1)}$ are in the same community, we get

$$\mathbb{P}_{12} \left[\prod_{(i,j) \in E_{\Delta}^{(1)}} \mathbf{1}_{z_i = z_j} = 1 \right] \leq \frac{1}{K^{\kappa\gamma + 1 - \text{cc} - 1}} = \frac{1}{K^{\kappa\gamma - \text{cc}}}$$

and, combining with (33)

$$\begin{aligned} \mathbb{E}_{12} \left[\mathbf{1}_{\ell(z)-1=x} \left(\frac{\bar{q}}{\bar{p}} \right)^{(\ell(z)-1)(\gamma+a)} \prod_{(i,j) \in E_{\Delta}^{(1)}} \mathbf{1}_{z_i = z_j} \right] &\leq \frac{1}{K^{\kappa\gamma - \text{cc}}} \cdot \text{cc}^{\text{cc}} K^{-(\text{cc} - x)} \left(\frac{\bar{q}}{\bar{p}} \right)^{x(\gamma+a)} \\ &= \frac{1}{K^{\kappa\gamma}} \cdot \text{cc}^{\text{cc}} K^x \left(\frac{\bar{q}}{\bar{p}} \right)^{x(\gamma+a)}. \end{aligned}$$

Since G is connected, each connected component of $\bar{\mathcal{G}}_{\Delta}^{(1)}$ must contain at least one node in $\pi^{(1)}(V) \cap \pi^{(2)}(V)$. As a consequence, we have $\text{cc} \leq u$. Combining Equation (32) with the last equation, and using $|E_{\Delta}| + 2|E_{\cap}| = 2|E|$, we get

$$\begin{aligned} \mathbb{E}_{12} [P_{G,\pi^{(1)}} P_{G,\pi^{(2)}}] &\leq \lambda^{|E_{\Delta}|} \bar{p}^{|E_{\cap}|} \frac{1}{K^{2\kappa\gamma-u}} u^u \sum_{x=0}^u \left(K \left(\frac{\bar{q}}{\bar{p}} \right)^{\gamma+a} \right)^x \\ &\leq \lambda^{2|E|} \left(\frac{\bar{p}}{\lambda^2} \right)^{|E_{\cap}|} \frac{1}{K^{2\kappa\gamma-u}} u^u \sum_{x=0}^u \left(K \left(\frac{\bar{q}}{\bar{p}} \right)^{\gamma+a} \right)^x. \end{aligned}$$

Lemma 5.2. *We have*

$$|E_{\cap}| \leq (\gamma + a)u.$$

Proof of Lemma 5.2. This result is a consequence of Proposition 3.1. Indeed, consider a community assignment such that the $2 + u$ nodes in $\pi^{(1)}(V) \cap \pi^{(2)}(V)$ belong to same community, whereas the $\kappa\gamma - u$ other nodes in $\pi^{(1)}(V) \setminus \pi^{(2)}(V)$ belong to distinct communities. Hence, for that assignment, we have $\kappa\gamma - u + 1$ communities and $|E^{\neq}| = |E| - |E_{\cap}|$. Then, Proposition 3.1 states that

$$|E| - |E_{\cap}| \geq (\gamma + a)[\kappa\gamma - u].$$

Since $|E| = (\gamma + a)\kappa\gamma$, Lemma 5.2 holds. $\square$

We deduce from the previous lemma that

$$\begin{aligned} \mathbb{E}_{12} [P_{G,\pi^{(1)}} P_{G,\pi^{(2)}}] &\leq \mathbb{E}_{12}^2 [P_{G,\pi}(Y)] \left[K \left(\frac{\bar{p}}{\lambda^2} \vee 1 \right)^{\gamma+a} \right]^u u^u \sum_{x=0}^u \left(K \left(\frac{\bar{q}}{\bar{p}} \right)^{\gamma+a} \right)^x \\ &\leq \mathbb{E}_{12}^2 [P_{G,\pi}(Y)] u^{u+1} \left[\left[K^2 \left(\frac{\bar{q}}{\lambda^2} \vee \frac{\bar{q}}{\bar{p}} \right)^{\gamma+a} \right]^u + \left[K \left(\frac{\bar{p}}{\lambda^2} \vee 1 \right)^{\gamma+a} \right]^u \right]. \end{aligned} \quad (34)$$

Combining the terms. Combining (31) and (34), we get

$$\begin{aligned} \text{var}_{12}(R_{12}) &= \sum_{u=0}^{\kappa\gamma} \sum_{\substack{\pi^{(1)}, \pi^{(2)} \in \Pi_{12} \\ |\text{range}(\pi^{(1)}) \cap \text{range}(\pi^{(2)})| = 2+u}} (\mathbb{E}_{12} [P_{G,\pi^{(1)}} P_{G,\pi^{(2)}}] - \mathbb{E}_{12} [P_{G,\pi^{(1)}}] \mathbb{E}_{12} [P_{G,\pi^{(2)}}]) \\ &\leq \sum_{u=1}^{\kappa\gamma} \binom{\kappa\gamma}{u}^2 \frac{(n-2)!u!}{(n-2\kappa\gamma+u-2)!} \mathbb{E}_{12} [P_{G,\pi^{(1)}} P_{G,\pi^{(2)}}] \\ &\leq \left(\frac{(n-2)!}{(n-\kappa\gamma-2)!} \right)^2 \sum_{u=1}^{\kappa\gamma} (\kappa\gamma)^{2u} \frac{1}{(n-\kappa\gamma-2)^u} \mathbb{E}_{12} [P_{G,\pi^{(1)}} P_{G,\pi^{(2)}}] \\ &\leq \left(\frac{(n-2)!}{(n-\kappa\gamma-2)!} \right)^2 \mathbb{E}_{12}^2 [P_{G,\pi}(Y)] \sum_{u=1}^{\kappa\gamma} \frac{u^{u+1} (\kappa\gamma)^{2u}}{(n-\kappa\gamma-2)^u} \left[\left[K^2 \left(\frac{\bar{q}}{\lambda^2} \vee \frac{\bar{q}}{\bar{p}} \right)^{\gamma+a} \right]^u + \left[K \left(\frac{\bar{p}}{\lambda^2} \vee 1 \right)^{\gamma+a} \right]^u \right] \\ &\leq \mathbb{E}_{12}^2 [R_{12}] (\kappa\gamma)^2 \left[\frac{2(\kappa\gamma)^3 K^2}{n} \left(\frac{\bar{q}}{\lambda^2} \vee \frac{\bar{q}}{\bar{p}} \right)^{\gamma+a} + \frac{2(\kappa\gamma)^3 K}{n} \left(\frac{\bar{p}}{\lambda^2} \vee 1 \right)^{\gamma+a} \right. \\ &\quad \left. + \left[\frac{2(\kappa\gamma)^3 K^2}{n} \left(\frac{\bar{q}}{\lambda^2} \vee \frac{\bar{q}}{\bar{p}} \right)^{\gamma+a} \right]^{\kappa\gamma} + \left[\frac{2(\kappa\gamma)^3 K}{n} \left(\frac{\bar{p}}{\lambda^2} \vee 1 \right)^{\gamma+a} \right]^{\kappa\gamma} \right]. \end{aligned}$$

This concludes the bound on $\text{var}_{12}(R_{12})$ in Equation (27).

5.3 Results under the distribution $\mathbb{P}_{\mathcal{Y}_2}$

Mean under $\mathbb{P}_{\mathcal{Y}_2}$. We have from Equation (28)

$$\mathbb{E}_{\mathcal{Y}_2} [P_{G,\pi}(Y)] = \lambda^{|E|} \mathbb{E}_{\mathcal{Y}_2} \left[\prod_{(i,j) \in \pi(E)} \mathbf{1}_{z_i=z_j} \right] = 0,$$

since $z_1 \neq z_2$ a.s. under $\mathbb{P}_{\mathcal{Y}_2}$. Equation (26) follows.

Variance under $\mathbb{P}_{\mathcal{Y}_2}$. We follow the same approach as for $\mathbb{P}_{\mathcal{I}_2}$. Let $\pi^{(1)}, \pi^{(2)} \in \Pi_{\mathcal{I}_2}$ such that $|\text{range}(\pi^{(1)}) \cap \text{range}(\pi^{(2)})| = 2 + u$, with $u \in \{0, \dots, \kappa\gamma\}$. Following (29), we again have

$$\mathbb{E}_{\mathcal{Y}_2} [P_{G,\pi^{(1)}} P_{G,\pi^{(2)}}] = \lambda^{|E_\Delta|} \bar{q}^{|E_\cap|} \mathbb{E}_{\mathcal{Y}_2} \left[\prod_{(i,j) \in E_\Delta} \mathbf{1}_{z_i=z_j} \prod_{\substack{(i,j) \in E_\cap \\ z_i=z_j}} \frac{\bar{p}}{\bar{q}} \right].$$

If $u = 0$, we have $E_\cap = \emptyset$ and since $z_1 \neq z_2$, we deduce that $\mathbb{E}_{\mathcal{Y}_2} [P_{G,\pi^{(1)}} P_{G,\pi^{(2)}}] = 0$. Then, for $u \in [1, \kappa\gamma]$, we argue as for $\mathbb{E}_{\mathcal{I}_2}$. In particular, the counterpart of (32) still holds up to the difference that under $\mathbb{E}_{\mathcal{Y}_2}$, the number of communities that are distinct of z_1 or of z_2 is $\ell(z) - 2$. Hence, we have

$$\mathbb{E}_{\mathcal{Y}_2} [P_{G,\pi^{(1)}} P_{G,\pi^{(2)}}] \leq \lambda^{|E_\Delta|} \bar{p}^{|E_\cap|} \frac{1}{K^{\kappa\gamma-u}} \mathbb{E}_{\mathcal{Y}_2} \left[\left(\frac{\bar{q}}{\bar{p}} \right)^{(\ell(z)-2)+(\gamma+u)} \prod_{(i,j) \in E_\Delta^{(1)}} \mathbf{1}_{z_i=z_j} \right].$$

Write again cc as the number of connected components of the graph $\bar{\mathcal{G}}_\Delta^{(1)} = (\pi^{(1)}(V), E_\Delta^{(1)})$ when we remove those containing $\pi^{(1)}(v_1) = 1$ or $\pi^{(1)}(v_2) = 2$. Then, conditionally to $\prod_{(i,j) \in E_\Delta^{(1)}} \mathbf{1}_{z_i=z_j} = 1$, we have $(\ell(z) - 2) \in [0, \text{cc}]$ and for any $x \in [0, \text{cc}]$

$$\mathbb{P}_{\mathcal{Y}_2} \left[\ell(z) - 2 = x \mid \prod_{(i,j) \in E_\Delta^{(1)}} \mathbf{1}_{z_i=z_j} = 1 \right] \leq \text{cc}^{\text{cc}} K^{-(\text{cc}-x)}.$$

Hence, we arrive as previously at

$$\mathbb{E}_{\mathcal{Y}_2} [P_{G,\pi^{(1)}} P_{G,\pi^{(2)}}] \leq \lambda^{2|E|} \left(\frac{\bar{p}}{\lambda^2} \right)^{|E_\cap|} \frac{1}{K^{2\kappa\gamma-u}} u^u \sum_{x=0}^u \left(K \left(\frac{\bar{q}}{\bar{p}} \right)^{\gamma+u} \right)^x,$$

and the counterpart of (34) holds:

$$\mathbb{E}_{\mathcal{Y}_2} [P_{G,\pi^{(1)}} P_{G,\pi^{(2)}}] \leq \mathbb{E}_{\mathcal{I}_2}^2 [P_{G,\pi}(Y)] u^{u+1} \left[\left[K^2 \left(\frac{\bar{q}}{\lambda^2} \vee \frac{\bar{q}}{\bar{p}} \right)^{\gamma+u} \right]^u + \left[K \left(\frac{\bar{p}}{\lambda^2} \vee 1 \right)^{\gamma+u} \right]^u \right].$$

Then, summing over all $\pi^{(1)}$ and $\pi^{(2)}$, we conclude as previously the bound on $\text{var}_{\mathcal{Y}_2}(R_{12})$ in Equation (27).

References

- [Abb18] Emmanuel Abbe. “Community Detection and Stochastic Block Models: Recent Developments”. In: *Journal of Machine Learning Research* 18.177 (2018), pp. 1–86.
- [AS15] Emmanuel Abbe and Colin Sandon. “Community Detection in General Stochastic Block Models: Fundamental Limits and Efficient Algorithms for Recovery”. In: *Proceedings of the 2015 IEEE 56th Annual Symposium on Foundations of Computer Science (FOCS)*. FOCS ’15. Washington, DC, USA: IEEE Computer Society, 2015, pp. 670–688.

- [BLM15] Charles Bordenave, Marc Lelarge, and Laurent Massoulié. “Non-backtracking spectrum of random graphs: community detection and non-regular ramanujan graphs”. In: *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*. IEEE. 2015, pp. 1347–1357.
- [CGV25] Alexandra Carpentier, Christophe Giraud, and Nicolas Verzelen. “Phase Transition for Stochastic Block Model with more than $\sqrt{n}$ Communities”. In: *arXiv preprint arXiv:2509.15822* (2025).
- [Chi+25] Byron Chin, Elchanan Mossel, Youngtak Sohn, and Alexander S. Wein. “Stochastic block models with many communities and the Kesten–Stigum bound - extended abstract”. In: *Proceedings of Thirty Eighth Conference on Learning Theory*. Ed. by Nika Haghtalab and Ankur Moitra. Vol. 291. Proceedings of Machine Learning Research. PMLR, 30 Jun–04 Jul 2025, pp. 1253–1258.
- [CX16] Yudong Chen and Jiaming Xu. “Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices”. In: *Journal of Machine Learning Research* 17.27 (2016), pp. 1–57.
- [Dec+11] Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. “Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications”. In: *Phys. Rev. E* 84 (6 Dec. 2011), p. 066106.
- [Din+25] Jingqiu Ding, Yiding Hua, Lucas Slot, and David Steurer. “Low degree conjecture implies sharp computational thresholds in stochastic block model”. In: *arXiv preprint arXiv:2502.15024* (2025).
- [EGV24] Bertrand Even, Christophe Giraud, and Nicolas Verzelen. “Computation-information gap in high-dimensional clustering”. In: *Proceedings of Thirty Seventh Conference on Learning Theory*. Ed. by Shipra Agrawal and Aaron Roth. Vol. 247. Proceedings of Machine Learning Research. PMLR, 30 Jun–03 Jul 2024, pp. 1646–1712.
- [HLL83] Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. “Stochastic blockmodels: First steps”. In: *Social networks* 5.2 (1983), pp. 109–137.
- [LG24] Yuetian Luo and Chao Gao. “Computational lower bounds for graphon estimation via low-degree polynomials”. In: *The Annals of Statistics* 52.5 (2024), pp. 2318–2348.
- [Mas14] Laurent Massoulié. “Community detection thresholds and the weak Ramanujan property”. In: *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*. 2014, pp. 694–703.
- [MNS18] Elchanan Mossel, Joe Neeman, and Allan Sly. “A proof of the block model threshold conjecture”. In: *Combinatorica* 38.3 (2018), pp. 665–708.
- [SW25] Youngtak Sohn and Alexander S. Wein. “Sharp Phase Transitions in Estimation with Low-Degree Polynomials”. In: *Proceedings of the 57th Annual ACM Symposium on Theory of Computing* (2025), pp. 891–902. arXiv: 2502.14407 [math.ST].
- [SZ24] Ludovic Stephan and Yizhe Zhu. “Community detection with the Bethe-Hessian”. In: *arXiv preprint arXiv:2411.02835* (2024).