arXiv:1311.4115v1 [math.PR] 17 Nov 2013

A Proof Of The Block Model Threshold Conjecture

Elchanan Mossel[*], Joe Neeman[†], and Allan Sly[‡]

November 19, 2013

Abstract

We study a random graph model named the "block model" in statistics and the "planted partition model" in theoretical computer science. In its simplest form, this is a random graph with two equal-sized clusters, with a between-class edge probability of q and a within-class edge probability of p.

A striking conjecture of Decelle, Krzkala, Moore and Zdeborová based on deep, non-rigorous ideas from statistical physics, gave a precise prediction for the algorithmic threshold of clustering in the sparse planted partition model. In particular, if $p = a/n$ and $q = b/n$, $s = (a-b)/2$ and $p = (a+b)/2$ then Decelle et al. conjectured that it is possible to efficiently cluster in a way correlated with the true partition if $s^2 > p$ and impossible if $s^2 < p$. By comparison, the best-known rigorous result is that of Coja-Oghlan, who showed that clustering is possible if $s^2 > Cp \ln p$ for some sufficiently large C.

In a previous work, we proved that indeed it is information theoretically impossible to to cluster if $s^2 < p$ and furthermore it is information theoretically impossible to even estimate the model parameters from the graph when $s^2 < p$. Here we complete the proof of the conjecture by providing an efficient algorithm for clustering in a way that is correlated with the true partition when $s^2 > p$. A different independent proof of the same result was recently obtained by Laurent Massoulie.

[*]U.C. Berkeley. Supported by NSF grant DMS-1106999, NSF grant CCF 1320105 and DOD ONR grant N000141110140

[†]U.T. Austin. Supported by NSF grant DMS-1106999 and DOD ONR grant N000141110140

[‡]U.C. Berkeley and the Australian National University. Supported by an Alfred Sloan Fellowship and NSF grant DMS-1208338.

1 Introduction

1.1 The Sparse Block Model

The block model is a random graph model defined in the following way

Definition 1.1 (The block model). *For $n \in \mathbb{N}$ and $p, q \in (0,1)$, let $\mathcal{G}(n,p,q)$ denote the model of random, $\pm$-labelled graphs on n vertices in which each vertex u is assigned (independently and uniformly at random) a label $\sigma_u \in \{\pm\}$, and then each possible edge (u,v) is included with probability p if $\sigma_u = \sigma_v$ and with probability q if $\sigma_u \neq \sigma_v$.*

If $p = q$, the planted partition model is just an Erdös-Renyi model, but if $p \gg q$ then a typical graph will have two well-defined clusters.

Interest in theoretical computer science in the average case analysis of the bisection problem led to intensive research of algorithm whose goal is to recover the partition [2–4, 6, 8, 13]. At the same time, the model is a classical statical model for communities [7] and the questions of identifiability of the parameters p and q and recovery of the clusters have been studied extensively, see e.g. [1, 15, 17]. We refer the readers to e.g. [14] for more background on the model.

1.2 The block models in sparse graphs

Until recently, all of the theoretical literature studying the block model focused on what we call the dense case, where the average degree is of order at least $\log n$ and the graph is connected. Indeed, it is clear that connectivity is required, if we wish to label all vertices accurately. However, the case of sparse graphs with constant average degree is well motivated from the perspective of real networks, see e.g. [11, 18].

Although sparse graphs are natural for modelling many large networks, the planted partition model seems to be most difficult to analyze in the sparse setting. Despite the large amount of work studying this model, the only results we know of that apply in the sparse case $p, q = O(\frac{1}{n})$ are those of Coja-Oghlan. Recently, Decelle et al. [5] made some fascinating conjectures for the cluster identification problem in the sparse planted partition model. In what follows, we will set $p = a/n$ and $q = b/n$ for some fixed a, b. Recall that $p = (a+b)/2$ and $s = (a-b)/2$.

Conjecture 1.2. *If $s^2 > p$ then the clustering problem in $\mathcal{G}(n, \frac{a}{n}, \frac{b}{n})$ is solvable as $n \to \infty$, in the sense that one can a.a.s. find a bisection which is positively correlated with the planted bisection.*

To put Coja-Oghlan's work into the context of this conjecture, he showed that if $s^2 > Cp \ln p$ for a large enough constant C, then the spectral method solves the clustering problem. Decelle et al.'s work is based on deep but non-rigorous ideas from statistical physics. In order to identify the best bisection, they use the sum-product algorithm (also known as belief propagation). Using the cavity method, they argue that the algorithm should work, a claim that is bolstered by compelling simulation results.

What makes Conjecture 1.2 even more interesting is the fact that if it is true, it represents a threshold for the solvability of the clustering problem. Indeed it was conjectured in [5] and proved in [14] that

Theorem 1.3. *If $s^2 < p$ then the clustering in $\mathcal{G}(n, \frac{a}{n}, \frac{b}{n})$ problem is not solvable as $n \to \infty$.*

It was further showed in [14] that $s^2 = p$ is the threshold for identifiability of the parameters a and b as conjectured by [5].

2 Our results

In our main result we prove Conjecture 1.2 by proving

Theorem 2.1. *If $s^2 > p$ then the clustering problem in $\mathcal{G}(n, \frac{a}{n}, \frac{b}{n})$ is solvable as $n \to \infty$, in the sense that one can a.a.s. find a bisection which is positively correlated with the planted bisection.*

We recently learned that Laurent Massoulie independently found a different proof of the conjecture [12].

2.1 Proof Strategy

It was conjectured in [5] that a popular algorithm, Belief Propagation initialized with i.i.d. uniform messages, detects communities all the way to the threshold. However, analysis of Belief propagation with random initial messages is a difficult task. In Krazakala et. al. [10] it was argued that a novel and very efficient spectral algorithm based on a non-backtracking matrix should also detect communities all the way to the threshold. Unfortunately we were unable to follow the path suggested in [10] and our algorithm for detection is not a spectral algorithm. Still, our analysis is based on the non-backtracking walk and we show that it can be implemented using matrix powering. The algorithm has very good theoretical running time $O(n \log^2 n)$ but the constant in the O needed for the proof that the algorithm is correct is very large, so the algorithm described is not nearly as efficient as the one in [10] (nor have we implemented it).

The basic intuition behind the proof is that we should be expecting a larger number of non-backtracking paths of a given length k between two vertices u and v if they have the same label, while a smaller number of non-backtracking paths is expected if the nodes u and v have different labels.

Instead of working with the number of non-backtracking walks, it is more convenient to work with a rank one correction, where an edge is represented by $1 - p/n$ and a non-edge by $-p/n$. With this alteration the expected contribution of each edge is 0. In order to be able to apply a statistical test based on non-backtracking paths for arbitrary pairs of vertices, the length of the paths k should be larger than the girth of the graph. On the other hand, longer backtracking walks are more likely to visit vertices and edges more than once. This makes the analysis more challenging. We now define the basic quantitates of interest.

Definition 2.2. *Let J_e denote the indicator of that the edge e is included in the graph and let $W_e = J_e - p/n$. For a non-backtracking path $\gamma = u_0, \dots, u_k$, let*

$$X_\gamma = X_{\gamma,p} = \prod_{i=1}^{k} W_{(u_{i-1}, u_i)}.$$

Let $\Gamma_{k,u,u'}$ denote the set of non-backtracking paths of length k from u to u' and let

$$N_{u,u',k} = N_{u,u',k} = \sum_{\gamma \in \Gamma_{k,u,u'}} X_\gamma.$$

Our analysis is based on counting non-backtracking paths. In order to do this, we would like to compute the expectation and variance of the number (or, more accurately, the weight) of paths. There is an obstacle, however, which is that on some very rare event there are many more paths than there should be. As a motivation, consider paths of length $\alpha \log n$ for a large constant α. The probability of an m-clique appearing around a particular vertex v is of the order $n^{-m^2/2} = e^{-\frac{m^2}{2} \log n}$, but on the event of its appearance, there are at least $(m-2)^k = e^{(m-2)\alpha \log n}$ non-backtracking paths of length k starting at v. If α is large enough, then we can choose $m \geq 4$ so that $\alpha \log(m-2) - m^2/2 - p$, and this event will contribute to the expected number of paths starting at v. In order to exclude such pathological cases, we will condition on an event that excludes cliques, along with some other problematic structures.

Definition 2.3. *We say that a graph is ℓ-tangle-free if every neighborhood of radius ℓ in the graph has at most one cycle. We write $\tilde{\mathbb{P}}$ and $\tilde{\mathbb{E}}$ for the probability measure conditioned on the graph being ℓ-tangle-free.*

Standard random graphs arguments imply that

Claim 2.4. *If $s, p = O(1)$ and $\ell = O(\log^{1/2} n)$ then the probability that the graph is ℓ-tangle free is $1 - n^{-1+o(1)}$.*

In our main technical result we obtain that:

Theorem 2.5. *If $s^2 > p$, then there exists an $0 < \alpha < \infty$ and $c_1 < \infty$ such that if $k = \alpha \log n$ then for all distinct u, u', v, v', it holds that*

$$\tilde{\mathbb{E}}[N_{u,u',k} \mid \sigma_u \sigma_{u'}] = \frac{1 + O(n^{-1/4})}{n} s^k \sigma_u \sigma_{u'},$$

$$\tilde{\mathbb{E}} N_{u,u',k}^2 \leq \frac{c_1 s^{2k}}{n^2},$$

$$\tilde{\mathbb{E}}[N_{u,u',k} N_{v,v',k} \mid \sigma_u \sigma_{u'}, \sigma_v \sigma_{v'}] = (1 + O(n^{-1/4})) \tilde{\mathbb{E}}[N_{u,u',k} \mid \sigma_u \sigma_{u'}] \tilde{\mathbb{E}}[N_{v,v',k} \mid \sigma_v \sigma_{v'}].$$

The first two statements say that $N_{u,u',k}$ is a random variable which is correlated with $\sigma_u \sigma'_u$ after the correct scaling. The later result says that for disjoint pairs of nodes, the variables $N_{u,u',k}$ and $N_{v,v',k}$ are asymptotically uncorrelated. The theorem (including stronger statements allowing to condition on large number of vertices) follows from Theorem 5.1, Theorem 6.1 and Theorem 7.1.

2.2 Almost linear time algorithm

Theorem 2.5 suggests a natural way to check if two vertices are in the same cluster by counting the number of weighted non-backtracking paths between them. It is the basis of the algorithm we develop to cluster the graph. We further show how it is possible to efficiently perform the algorithm using matrix powering.

Theorem 2.6. *There is an algorithm whose running time is $O(n \log^2 n)$ and such that if $s^2 > p$ then the algorithm returns a bisection with correlation $\epsilon(a, b) > 0$ with the planted bisection in $\mathcal{G}(n, \frac{a}{n}, \frac{b}{n})$ with probability $1 - o(1)$.*

With more care in the analysis the running time could be reduced to $O(n \log n)$ for a slightly modified algorithm.

3 The Algorithm and its Analysis

We will now describe the algorithm. We will first give a naïve description, and then we will show that this description may be implemented in almost linear time. We defer the proof of the algorithm's correctness until Section 8.

3.1 The Algorithm

Set $R = \log\log\log\log n$ to be a radius. Let $\delta' > 0$ be chosen so that $s^2(1-\delta')^2 = p(1-\delta')$ and let $\delta = \delta'/2$. Let k and c_1 be chosen as in Theorem 2.5 with the parameters $p' = p(1-\delta), s' = s(1-\delta)$ and $D = D(d', p')$ will be a large constant specified in the proof. The algorithm proceeds as follows:

- Pick at random $\lceil\sqrt{n}\rceil$ vertices V'' from the graph, leaving the graph $G' = (V', E')$
- Let w^* be a node in V'' whose number of neighbors in V' is closest to $\lceil\sqrt{\log\log n}\rceil$. Let Y be the set of its neighbours in V'.
- For each $v \in V'$ denote $S_v = B_R(v) \setminus B_{R-1}(v)$.
- For each $1 \le j \le \log n$, construct V_j by each time independently deleting $\lceil n\delta - \sqrt{n}\rceil$ vertices of $V' \setminus Y$ chosen independently.
- For each $v \in V'$ and $1 \le j \le \log n$ define

$$\tau_j(v) = \operatorname{sgn}\left(\sum_{u\in Y, u'\in S_v\cap V_j} N^{(j)}_{u,u',k} + D\frac{{s'}^{k+R+1}}{p'}|Y|\xi_{j,v}\right),$$

  where $\xi_{j,v}$ are i.i.d. random variables uniform on $[-1,1]$ and

$$N^{(j)}_{u,u',k} = \sum_{\substack{\gamma\in\Gamma_{k,u,u'}\\ \gamma\subset V_j}} X_{\gamma,p'}.$$

- For each $v \in V'$ let J_v be the first j such that $B_R(v) \cap V_j = \varnothing$ and $(S_v \cup Y) \subset V_j$, and 0 if no such j exists. Then set $\tau(v) = \tau_{J_v}(v)$ when $J_v \neq 0$. For v with $J_v = 0$ or for $v \in V \setminus V'$ choose $\tau(v)$ at independently at random uniformly from $\{1,-1\}$.

In order to prove Theorem 2.6, we will show that the output τ is correlated with the true partition with high probability. In the following section we describe how to evaluate the τ_j in time $O(p^R n\log n)$.

3.2 Efficient implementation of the algorithm

In order to implement the algorithm above we are looking for an efficient algorithm for calculating $N^{(k)}_{u,v}$ (moving the k into the superscript for ease of writing matrices).

Let V be the vertex set and A the adjacency matrix of the resulting graph. We recall the definition of $N^{(k)}_{u,v}$ and introduce some related matrices

Definition 3.1. *Let*

$$N^{(k)}_{u,v} = \sum_{\gamma} X_\gamma,$$

where the sum is over non-backtracking paths and with the convention that $N^{(0)} = I$. Let $\mathbb{1}$ be $n\times n$ matrix all of whose entries are 1 and

$$Q^{(k,\rho)}_{u,v} = \sum_{j=0}^{k/2} \rho^{2j} N^{(k-2j)}_{u,v}, \tag{1}$$

$$M = \begin{pmatrix} (1-p/n)A & -(1-p/n)^2(D-I) & -(p/n)(\mathbb{1}-A-I) & -(p/n)^2((n-1)I-D) \\ I & 0 & 0 & 0 \\ (1-p/n)A & -(1-p/n)^2 D & -(p/n)(\mathbb{1}-A-I) & -(p/n)^2((n-2)I-D) \\ 0 & 0 & I & 0 \end{pmatrix}, \quad (2)$$

$$\hat{M} = \begin{pmatrix} (1-p/n)A & -(1-p/n)^2 D & -(p/n)(\mathbb{1}-A-I) & -(p/n)^2((n-1)I-D) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad (3)$$

and set

$$\mathcal{Q}_k = \begin{pmatrix} Q^{(k,1-p/n)} & Q^{(k-1,1-p/n)} & Q^{(k,-p/n)} & Q^{(k-1,-p/n)} \end{pmatrix}^T \quad (4)$$

Lemma 3.2. *We have*

$$M\mathcal{Q}_k = \mathcal{Q}_{k+1}$$

and

$$\hat{M}\mathcal{Q}_k = \begin{pmatrix} N^{(k+1)} & 0 & 0 & 0 \end{pmatrix}^T.$$

Proposition 3.3. *For a graph on n vertices and m edges and for every vector z the matrix $N^{(k)}z$ can be computed in time $O((m+n)k)$.*

Proof. The proof follows from the fact that

$$N^{(k)}z = \hat{M}M^{k-1}\mathcal{Q}_0\begin{pmatrix} z & 0 & 0 & 0 \end{pmatrix}$$

and that each of the matrices $M, \hat{M}$ and $\mathcal{Q}_0$ are made of at most 16 blocks, each of which is a sum of a sparse matrix with $O(n+m)$ entries and a rank 1 matrix. ☐

We can now prove Theorem 2.6.

Proof. Note that the sums $\sum_{u\in Y, u'\in S_v\cap V_j} N_{u,u',k}$ are the only input needed from the graph at each iteration of the algorithm. The collection of all such sums can be read from the entries of $N^{(k)}z$, where N corresponds to the graph with the removed nodes and z is the indicator of the vertices in Y. Therefore, the running time of each iteration is $O((n+m)k)$. Since there are $\log n$ iterations and we have $m = O(n)$ and $k = O(\log n)$ we obtain that the running time of the algorithm is $O(\log^2 n)$. ☐

It remains to prove Lemma 3.2.

Proof. We will write $N^{(k)}_{u,w,v}$ to denote the above sum restricted to non-backtracking paths which move to w on their first step. Then we have the recursion

$$\begin{aligned} N^{(k)}_{u,v} &= \sum_{w\neq u} N^{(k)}_{u,w,v} \\ &= \sum_{w\neq u} W_{(u,w)}(N^{(k-1)}_{w,v} - N^{(k-1)}_{w,u,v}) \\ &= \sum_{w\neq u} W_{(u,w)}\left(N^{(k-1)}_{w,v} - W_{(u,w)}(N^{(k-2)}_{u,v} - N^{(k-2)}_{u,w,v})\right) \\ &= \sum_{w\neq u} W_{(u,w)}N^{(k-1)}_{w,v} - \sum_{w\neq u} W^2_{(u,w)}N^{(k-2)}_{u,v} + \sum_{w\neq u} W^2_{(u,w)}N^{(k-2)}_{u,w,v} \end{aligned}$$

By expanding the terms of the form $N^{(k-2)}_{u,w,v}$ repeatedly, we obtain by induction that

$$\begin{aligned}
N^{(k)}_{u,v} &= \sum_{w\neq u}\left[\Big(\sum_{j=0}^{\lfloor (k-1)/2\rfloor} W^{1+2j}_{(u,w)} N^{(k-1-2j)}_{w,v}\Big) - \Big(\sum_{j=0}^{\lfloor (k-2)/2\rfloor} W^{2+2j}_{(u,w)} N^{(k-2-2j)}_{u,v}\Big)\right] \\
&= \sum_{w\sim u}\left[\Big(\sum_{j=0}^{\lfloor (k-1)/2\rfloor} (1-p/n)^{1+2j} N^{(k-1-2j)}_{w,v}\Big) - \Big(\sum_{j=0}^{\lfloor (k-2)/2\rfloor} (1-p/n)^{2+2j} N^{(k-2-2j)}_{u,v}\Big)\right] \\
&+ \sum_{\substack{w\not\sim u\\ w\neq u}}\left[\Big(\sum_{j=0}^{\lfloor (k-1)/2\rfloor} (-p/n)^{1+2j} N^{(k-1-2j)}_{w,v}\Big) - \Big(\sum_{j=0}^{\lfloor (k-2)/2\rfloor} (-p/n)^{2+2j} N^{(k-2-2j)}_{u,v}\Big)\right]
\end{aligned}$$

The recursion above can be written using the matrix Q from (1) in the following way:

$$\begin{aligned}
N^{(k)} &= (1-p/n)AQ^{(k-1,1-p/n)} - (1-p/n)^2 DQ^{(k-2,1-p/n)} \\
&\quad - (p/n)(\mathbb{1} - A - I)Q^{(k-1,-p/n)} - (p/n)^2((n-1)I - D)Q^{(k-2,-p/n)}
\end{aligned}$$

Moreover,

$$\begin{aligned}
Q^{(k,1-p/n)} &= N^{(k)} + (1-p/n)^2 Q^{(k-2,1-p/n)} \\
&= (1-p/n)AQ^{(k-1,1-p/n)} - (1-p/n)^2(D-I)Q^{(k-2,1-p/n)} \\
&\quad - (p/n)(\mathbb{1} - A - I)Q^{(k-1,-p/n)} - (p/n)^2((n-1)I - D)Q^{(k-2,-p/n)},
\end{aligned}$$

and

$$\begin{aligned}
Q^{(k,-p/n)} &= N^{(k)} + (p/n)^2 Q^{(k-2,-p/n)} \\
&= (1-p/n)AQ^{(k-1,1-p/n)} - (1-p/n)^2 DQ^{(k-2,1-p/n)} \\
&\quad - (p/n)(\mathbb{1} - A - I)Q^{(k-1,-p/n)} - (p/n)^2((n-2)I - D)Q^{(k-2,-p/n)}.
\end{aligned}$$

Written in terms of the matrices $\mathcal{Q}$ from (4) and $M, \hat{M}$ defined in (2),(3) the recursions above can be written as

$$M\mathcal{Q}_k = \mathcal{Q}_{k+1}.$$

and

$$\hat{M}\mathcal{Q}_k = \begin{pmatrix} N^{(k+1)} & 0 & 0 & 0 \end{pmatrix}^T,$$

as claimed. □

4 Proof of the Correlation Theorem - Preliminaries

4.1 The conditional and unconditional measures

We begin by showing that probabilities of events that depend on small number of edges are close in the conditional and unconditional measures. We let Ξ denote the event that the graph is ℓ-tangle-free. Recall that $\ell = \Theta(\log^{1/2} n)$. Standard random graph arguments imply that if H is any graph with more edges than vertices, and $G \sim \mathcal{G}(n, q/n)$ for any q, then the probability that G contains a copy of H is $O(n^{-1})$. Of course, this also holds under $\mathbb{P}$, because $\mathbb{P}$ is stochastically dominated by $\mathcal{G}(n, \max\{a,b\}/n)$. Now, the complement of Ξ is contained in the event that G has a subgraph H with $O(\sqrt{\log n})$ vertices, such that $V(H) = E(H) + 1$. Since there at most $n^{o(1)}$ ways to choose such an H, it follows that $\mathbb{P}(\Xi) \geq 1 - n^{-1+o(1)}$.

Next, we note that the probability of a labelling is essentially unchanged between $\tilde{\mathbb{P}}$ and $\mathbb{P}$.

Lemma 4.1. *For any subset $A \subset V(G)$ and any labelling $\tau_A \in \{1,-1\}^A$,*

$$\tilde{\mathbb{P}}(\sigma_A = \tau_A) = (1 + O(n^{-1+o(1)}))2^{-|A|}.$$

Proof. For any labelling τ_A, the distribution of $\mathbb{P}$ given $\sigma_A = \tau_A$ is still stochastically dominated by $\mathcal{G}(n, \max\{a,b\}/n)$. Hence, $\mathbb{P}(\Xi \mid \sigma_A = \tau_A) \geq 1 - n^{-1+o(1)}$, and so

$$\tilde{\mathbb{P}}(\sigma_A = \tau_A) = \mathbb{P}(\sigma_A = \tau_A \mid \Xi) = \mathbb{P}(\Xi \mid \sigma_A = \tau_A)\frac{2^{-|A|}}{\mathbb{P}(\Xi)} = (1 + O(n^{-1+o(1)}))2^{-|A|}. \qquad \square$$

Next, we note that the probability of a single edge is essentially unchanged between $\tilde{\mathbb{P}}$ and $\mathbb{P}$.

Lemma 4.2. *Let $e = (v, v')$ and let S be a subset of edges of the complete graph on V s.t. $e \notin S$. Let $V(S)$ be the vertices adjacent to S. Then If $|S| = O(\log n)$ and $\mathbb{P}[\Xi \mid J_{S\cup\{e\}} = 1] > 0$ then*

$$\tilde{\mathbb{P}}[J_e = 1 \mid J_S = 1, \sigma_v, \sigma_{v'}, \sigma_{V(S)}] \geq (p + s\sigma_v\sigma_{v'})/n - O(n^{-2+o(1)}).$$

Proof. Set Ξ_e to be the event that the graph plus edge e is ℓ tangle free. Note that for every j_S it holds that

$$\mathbb{P}[\Xi \mid J_S = j_S] \geq \mathbb{P}[\Xi_e \mid J_S = j_S] \geq 1 - n^{-1+o(1)},$$

where the second inequality follows since adding in the edges $S \cup \{e\}$ will only create a local double loop if there is a path of length at most 2ℓ between two vertices in the edge set of $S \cup \{e\}$ that does not contain any edges from $S \cup \{e\}$ which has probability bounded above by $4(|S|+1)^2(a+b)^{2\ell}/n$. Now

$$\begin{aligned}
&\tilde{\mathbb{P}}[J_e = 1 \mid J_S = j_S] \\
&= \frac{\mathbb{P}[J_e = 1, J_S = j_S, \Xi]}{\mathbb{P}[J_e = 1, J_S = j_S, \Xi] + \mathbb{P}[J_e = 0, J_S = j_S, \Xi]} \\
&= \frac{\frac{p+s\sigma_v\sigma_{v'}}{n}\mathbb{P}[\Xi_e \mid J_S = j_S, \sigma_v, \sigma_{v'}, \sigma_{V(S)}]}{\frac{p+s\sigma_v\sigma_{v'}}{n}\mathbb{P}[\Xi_e \mid J_S = j_S, \sigma_v, \sigma_{v'}, \sigma_{V(S)}] + (1 - \frac{p+s\sigma_v\sigma_{v'}}{n})\mathbb{P}[\Xi \mid J_S = j_S, J_e = 0, \sigma_v, \sigma_{v'}, \sigma_{V(S)}]} \\
&= \frac{\frac{p+s\sigma_v\sigma_{v'}}{n}(1 - n^{-1+o(1)})}{\frac{p+s\sigma_v\sigma_{v'}}{n}(1 - n^{-1+o(1)}) + (1 - n^{-1+o(1)})} = \frac{p + s\sigma_v\sigma_{v'}}{n} - O(n^{-2+o(1)}).
\end{aligned}$$

$\square$

The claim above states that the conditional distribution of an edge is close in $\mathbb{P}$ and $\tilde{\mathbb{P}}$. Next we we wish to prove a similar claim for small collections of edges. For a set of edges S we let $V(S)$ denote its vertex set.

Lemma 4.3. *Let S, S', T be disjoint subsets of edges with $|S| + |S|' + |T| = O(\log n)$. Let $m : T \to \{0, 1, \ldots, C\log n\}$ be a sequence of moments. Assuming $\mathbb{P}[\Xi \mid J_{S\cup T} = 1] > 0$ we have that*

$$\left| \frac{\tilde{\mathbb{E}}[\prod_{e\in T}[J_e - p/n]^{m_e} \mid J_S = 1, J_{S'} = 0, \sigma_{V(S\cup S'\cup T)}]}{\prod_{(u,u')\in T, m_{(u,u')}=1}\frac{s\sigma_u\sigma_{u'}}{n}\prod_{(u,u')\in T, m_{(u,u')}>1}\frac{p+s\sigma_u\sigma_{u'}}{n}} - 1 \right| \leq |T|n^{-0.9}.$$

Proof. We prove this by induction on $|T|$. Let $(u, u') \in T$, then

Denote $\gamma = v = v_0, v_1, \ldots, v_{|\gamma|} = v'$ and let $T^* = T \setminus \{(u,u')\}$.

$$\begin{aligned}
&\tilde{\mathbb{E}}[(J_{(u,u')} - p/n)^{m_{(u,u')}} \prod_{e \in T^*} [J_e - p/n]^{m_e} \mid J_S = 1, J_{S'} = 0, \sigma_{V(S \cup S' \cup T)}] \\
&= (1-p/n)^{m_{(u,u')}} \tilde{\mathbb{P}}[J_{(u,u')} = 1 \mid J_S = 1, J_{S'} = 0, \sigma_{V(S \cup S' \cup T)}] \\
&\cdot \tilde{\mathbb{E}}[\prod_{e \in T^*} [J_e - p/n]^{m_e} \mid J_S = 1, J_{S'} = 0, J_{(u,u')} = 1, \sigma_{V(S \cup S' \cup T)}] \\
&+ (-p/n)^{m_{(u,u')}} \tilde{\mathbb{P}}[J_{(u,u')} = 1 \mid J_S = 1, J_{S'} = 0, \sigma_{V(S \cup S' \cup T)}] \\
&\cdot \tilde{\mathbb{E}}[\prod_{e \in T^*} [J_e - p/n]^{m_e} \mid J_S = 1, J_{S'} = 0, J_{(u,u')} = 0, \sigma_{V(S \cup S' \cup T)}]
\end{aligned}$$

The result follows by using the inductive hypothesis and Lemma 4.2.

□

4.2 Combinatorial path bounds

A crucial ingredient in the proof is obtaining bounds on the number of various types of paths in terms of how much they self intersect both by intersecting a previous vertex on the path and in terms of repeating an edge of the path.

Definition 4.4. *Given a path $p = (v_1, \ldots, v_k)$ in the graph, we say that an edge (v_i, v_{i+1})*

- *is* new *if for all $j \le i$, $v_j \ne v_{i+1}$*
- *is* old *if there is some $j < i$ such that $(v_i, v_{i+1}) = (v_j, v_{j+1})$.*
- *Otherwise, we say that (v_i, v_{i+1}) is* returning *(in this case $v_{i+1} = v_j$ for $j < i$ but (v_i, v_{i+1}) is not one of the previous edges).*

Let $k_n(p), k_o(p)$ and $k_r(p)$ to be the number of new, old, and returning edges respectively.

*We say that a path is ℓ-*tangle-free *if every neighborhood of radius ℓ has at most one cycle.*

For the rest of this subsection we fix α, let n be the size of the graph, and set $k = \alpha \log n$. Note that for every new edge in a path, the number of distinct vertices in the path increases by one, as does the number of distinct edges. For a returning edge, only the number of edges increases, while for an old edge, neither increases. Therefore we easily see that:

Claim 4.5. *The number of vertices visited by the path is $k_n + 1$ and the number of edges is $k_n + k_r$.*

Lemma 4.6. *If $k_r \ge 1$ then for sufficiently large n there are at most*

$$n^{k_n + k_r + \alpha \log(4ek_r) - k_r/2}$$

paths with a fixed starting and ending point, satisfying $k_n(p) = k_n$ and $k_r(p) = k_r$.

The point of Lemma 4.6 is that it implies that paths with large k_r are so rare that they do not contribute any weight. Indeed, recall that $k = \alpha \log n$ for some α to be determined. Choose k^* large enough (depending on α) so that

$$\alpha \log(2p) + \alpha \log(4ek^*) - k^*/2 < -4.$$

It then follows from Lemma 4.6 that if Γ is the collection of all paths of length k with $k_r(\gamma) \ge k^*$ then

$$\begin{aligned}\sum_{\gamma\in\Gamma}\left(\frac{2p}{n}\right)^{k_n(\gamma)+k_r(\gamma)} &\le \sum_{k_r\ge k^*}(2p)^k n^{\alpha\log(4ek_r)-k_r/2}\\ &= \sum_{k_r\ge k^*} n^{\alpha\log(2p)+\alpha\log(4ek_r)-k_r/2}\\ &= n^{-4+o(1)}.\end{aligned}$$

Note that for any path γ and any labelling σ,

$$|\tilde{\mathbb{E}}[X_\gamma \mid \sigma]| \le \left(\frac{2p}{n}\right)^{k_r(\gamma)+k_n(\gamma)}$$

(since $k_r(\gamma) + k_n(\gamma)$ is the number of edges in γ). Hence, we have:

Corollary 4.7. *Let $k^* = k^*(\alpha, p)$ be defined as above. Then*

$$\sum_{\gamma:k_r(\gamma)\ge k^*} |\tilde{\mathbb{E}}[X_\gamma \mid \sigma]| \le n^{-4+o(1)}.$$

Note that there is no non-backtracking restriction here, or in Lemma 4.6.

We now prove the lemma.

Proof. Suppose that for all i, we decide in advance whether (v_i, v_{i+1}) will be new, old, or returning. There are at most $\binom{k}{k_n\ k_o\ k_r}$ ways to make this choice. Fix an i and suppose that v_i has already been determined. If (v_i, v_{i+1}) is new then there are at most n choices for v_{i+1}. If (v_i, v_{i+1}) is returning then there are at most k_n choices for v_{i+1} because k_n vertices are visited by the path. Otherwise, $(v_i, v_{i+1}) = (v_j, v_{j+1})$ for some $j < i$. Note that the number of possible j is at most $2k_r + 2$. To see this note that whenever a new edge of the form (v_i, v_j) for $j < i$ is introduced, at least one returning edge must be introduced before the next time v_i is visited. Hence, the total number of choices is at most

$$\begin{aligned}\binom{k}{k_n\ k_o\ k_r} n^{k_n} k_n^{k_r}(2k_r+2)^{k_o} &\le \frac{k^{k_o+k_r}}{k_o!k_r!} n^{k_n} k^{k_r}(4k_r)^{k_o}\\ &= n^{k_n}\frac{k^{2k_r}}{k_r!}\frac{(4kk_r)^{k_o}}{k_o!}\\ &\le n^{k_n}\left(\frac{ek^2}{k_r}\right)^{k_r}\left(\frac{4ekk_r}{k_o}\right)^{k_o}\\ &= n^{k_n+k_r}\left(\frac{ek^2}{nk_r}\right)^{k_r}\left(\frac{4ekk_r}{k_o}\right)^{k_o}.\end{aligned}$$

Now, the quantity $(C/x)^x$ is increasing in x as long as $x \le C/e$. Applying this with $C = 4ekk_r$ and the values $k_0 \le k \le C/e$ we have

$$\left(\frac{4ekk_r}{k_o}\right)^{k_o} \le (4ek_r)^k = (4ek_r)^{\alpha\log n} = n^{\alpha\log(4ek_r)},$$

On the other hand, $ek^2/(nk_r) \le n^{-1/2}$ for sufficiently large n. Hence, the total number of paths is at most

$$n^{k_n+k_r} n^{-k_r/2} n^{\alpha\log(4ek_r)}. \qquad \square$$

We also require good bounds in the case where k_r is a small constant in which case the bound of Lemma 4.6 is larger (by a polynomial in n) than $n^{k_n+k_r-1}$. We can avoid this by asking the path to be tangle-free.

Lemma 4.8. *If $k_r \geq 1$ then there are at most*

$$k^{3k_r+4k_r k/\ell} n^{k_n-1}$$

ℓ-tangle-free paths with a fixed starting and ending point, satisfying $k_n(p) = k_n$ and $k_r(p) = k_r$.

Corollary 4.9. *Recall that $k = \alpha \log n$. If $\ell >> \log\log n$ then there are at most*

$$n^{k_n-1+o(1)}$$

ℓ-tangle-free paths with a fixed starting and ending point,

Proof. The proof is similar to the last one, except that we will count old edges more carefully. We say that a neighbor w of v is *short* if there is a cycle in p that contains (v, w) and has length at most 2ℓ; otherwise we say that w is a *long* neighbor. Because p is ℓ-tangle-free, each vertex $v \in p$ has at most 2 short neighbors.

For this proof, the *degree* $d(v) = d(v, p)$ of $v \in p$ is the number of $w \in p$ such that either (v, w) or $(w, v) \in p$. Let $D(p)$ be the set of vertices with degree at least 3. Note that because p is non-backtracking, an old edge (v_i, v_{i+1}) is already determined by the path until v_i unless $v_i \in D$.

Now fix $v \in D$. Let $m(v)$ be the number of times that v was visited. Let $m_{\text{in}}(v)$ be the number of times that v was visited from a long neighbor and $m_{\text{out}}(v)$ be the number of the times we move to a long neighbor. Note in the times where the path move from a short vertex to a different short vertex, the path is determined if the path moves to an old edge. At all other time there are at most $d(v)$ choice of following a long edge. Thus the total number of ways to chose old edges starting at v is bounded by

$$m(v)^{m_{\text{out}}(v)+m_{\text{in}}(v)} d(v)^{m_{\text{in}}(v)+m_{\text{out}}(v)} \leq m(v)^{2m_{\text{out}}(v)+2m_{\text{in}}(v)},$$

where the second inequality follows from the fact that $d(v) \leq m(v)$.

Repeating this for all $v \in D$, we see that the number of ways to choose all the old edges in p is at most

$$\prod_{v \in D} m(v)^{2m_{\text{out}}(v)+2m_{\text{in}}(v)} \tag{5}$$

Next, we invoke the tangle-free condition in order to bound $m_{\text{in}}(v)$ and $m_{\text{out}}(v)$. Indeed, after leaving v via a long neighbor, we must wait at least 2ℓ steps before visiting v again; hence $m_{\text{out}}(v) \leq k/(2\ell)$. A similar argument shows that $m_{\text{in}}(v) \leq k/(2\ell)$, and hence the number of ways to choose the old edges is at most

$$\left(\prod_{v \in D} m(v) \right)^{2k/\ell}.$$

By the AM-GM inequality,

$$\prod_{v \in D} m(v) \leq \left(\frac{1}{|D|} \sum_{v \in D} m(v) \right)^{|D|} \leq \left(\frac{1}{|D|} \sum_{v \in D} m(v) \right)^{2k_r},$$

where the second inequality follows because every time the walk returns to its old path, it creates at most 2 vertices of degree higher than 2 (one when the walk returns, and one when it leaves again).

Since $m(v) \le k$ for every v, the quantity above is bounded by k^{2k_r}. Plugging this back into (5), we see that the number of ways to choose all the old edges in p is at most $k^{4k_r k/\ell}$.

Now we will count the other kinds of edges. First of all, note that if we specify which edges are returning and we also specify the first new edge after each returning edge, then the old edges are also determined (because every edge after a returning edge but before the next new edge is either old or returning). Therefore, the number of ways to specify which edges are old, new, or returning is at most $\binom{k}{k_r}^2$. As in the previous proof, there are at most k^{k_r} ways to choose the returning edges. As for the new edges - the last new edge must hit the final vertex - so it has no choices. Therefore, there are n^{k_n-1} ways to choose the new edges and the total number of paths is at most

$$\binom{k}{k_r}^2 n^{k_n-1} k^{k_r} k^{4k_r k/\ell} \le k^{3k_r+2k_r k/\ell} n^{k_n-1}. \qquad \square$$

Next, we count paths with a few tangles. We say that a path has t ℓ-tangles if by deleting t edges (and no less than t), we can turn it into an ℓ-tangle-free path.

Lemma 4.10. *If $k_r \ge 1$ then there are at most*

$$k^{3k_r+4k_r(1+k/\ell+2t)} n^{k_n-1}$$

paths from a fixed starting point to a fixed endpoint with t ℓ-tangles, and which satisfy $k_n(p) = k_n$ and $k_r(p) = k_r$.

Proof. The only place we used the ℓ-tangle-free condition in the previous proof was in the counting of old edges, where we used it to bound $m_{\text{in}}(v)$ and $m_{\text{out}}(v)$. Now, if we are allowed t tangles then it is no longer immediately clear which neighbors of v should be the short neighbors. However, there are at most $\binom{d(v)}{2} \le m(v)^2$ ways to choose two short neighbors. Having chosen them, we can bound $m_{\text{out}}(v)$ (and similarly $m_{\text{in}}(v)$) by $k/(2\ell) + t$ because whenever we leave v, we either have to take a long path of length 2ℓ or we can take a shortcut, which adds a tangle. Applying this modified counting to the previous proof, we have at most

$$\left(\prod_{v\in D} m(v)\right)^{2+2k/\ell+4t} \le k^{4k_r(1+k/\ell+2t)}$$

ways to choose old edges. $\square$

Later, we will need to consider paths that are allowed to backtrack a certain number of times, and which are required to pass through certain vertices. For a set of vertices U, let $\Gamma^t_{k,\ell,r,u,u'}(U)$ be the collection of paths from u to u' that have length k, have t ℓ-tangles, have at most r backtracks, and pass through all the vertices contained in U.

Lemma 4.11.

$$|\Gamma^t_{k,\ell,r,u,u'}(U)| \le k^{3k_r+4k_r(1+k/\ell+2t)+2r+2|U|} n^{k_n-1-|U|}.$$

Proof. This is essentially the same as the proof of Lemma 4.10. There are two differences: the first is that we are allowed r backtracks (which may be distributed in the path in at most $\binom{k}{r} \le k^r$ ways, each of which induces at most k new choices for a total of k^{2r}). The second difference is that instead of having k_n vertices to choose, we have $k_n - |U| - 1$ (the -1 coming from the fact that we have fixed the start point and end point). Then there are at most $k^{|U|}(|U|)! \le k^{2|U|}$ ways to intersperse the vertices from U with the rest of the path. $\square$

Corollary 4.12. *In particular, if $k = \alpha \log n$, $\ell \gg \log\log n$ and $k_r \le k^*$, for some constant k^*, and $t \le t^*$ for some constant t^* then*

$$|\Gamma^t_{k,\ell,r,u,u'}(U)| \le n^{k_n-1-|U|+o(1)}.$$

5 The first moment

Theorem 5.1. *Let $\Gamma(u,v)$ be the set of non-backtracking paths of length k. Let A be any subset of vertices with $u,v \in A$, and $|A| \le n^{0.1}$. Then for any labelling σ_A on A,*

$$\tilde{\mathbb{E}}\left[\sum_{\gamma\in\Gamma(u,v)} X_\gamma \mid \sigma_A\right] = (1+O(n^{-0.8}))\frac{\sigma_u\sigma_v s^k}{n}.$$

5.1 Decomposition into segments

Definition 5.2. *Consider a path γ. We say that a collection of paths $\zeta^{(1)},\dots,\zeta^{(r)}$ is a SAW-decomposition of γ if*

- *each $\zeta^{(i)}$ is a self-avoiding path;*
- *the interior vertices of each $\zeta^{(i)}$ are not contained in any other $\zeta^{(j)}$, nor is any interior vertex of $\zeta^{(i)}$ equal to the starting or ending vertex of γ; and*
- *the $\zeta^{(i)}$ cover γ, in the sense that every edge traversed by γ is also traversed by some $\zeta^{(i)}$.*

Given a SAW decomposition as above, we let D denote the set of vertices that are the endpoint of some $\zeta^{(i)}$ and we let m_i denote the number of times that $\zeta^{(i)}$ was traversed in γ.

Note that since $\zeta^{(i)}$ and $\zeta^{(j)}$ share no interior vertices, every time that the path γ begins to traverse $\zeta^{(i)}$, it must finish traversing $\zeta^{(i)}$. Moreover, the fact that $\zeta^{(i)}$ and $\zeta^{(j)}$ share no interior vertices implies that they are edge-disjoint, and so for each fixed i, every edge in $\zeta^{(i)}$ is traversed the same number (i.e. m_i) of times.

There is a natural way to construct a SAW-decomposition of a path γ. Consider a path γ between u and v, and let D^- be the subset of γ's vertices that have degree 3 or more in γ. Let

$$D = D^- \cup \{u,v\} \cup \{w \in \gamma : \gamma \text{ backtracks at } w\}. \tag{6}$$

Then γ may be decomposed into a collection of self-avoiding walks between vertices in D. To be precise, suppose that γ is given by $u = u_0, u_1, \dots, u_k = v$. Let $j_1 > 0$ be minimal so that $u_{j_1} \in D$ and let $\gamma^{(1)}$ be the path $u_0,\dots,u_{j_1}$. Inductively, if $j_{i-1} < k$ then let $j_i > j_{i-1}$ be minimal so that $u_{j_i} \in D$ and set $\gamma^{(i)}$ to be the path $u_{j_{i-1}},\dots,u_{j_i}$. It follows from this definition that the interior nodes in each $\gamma^{(i)}$ are degree 2 in γ; hence, each $\gamma^{(i)}$ is self-avoiding, and any pair $\gamma^{(i)},\gamma^{(j)}$ are either identical, or their interior vertices are disjoint. Finally, let $\{\zeta^{(1)},\dots,\zeta^{(r)}\}$ be $\{\gamma^{(i)}\}$, but with duplicates removed.

Definition 5.3. *We call the preceding construction of $\zeta^{(1)},\dots,\zeta^{(r)}$ the* canonical SAW-decomposition of *γ.*

We remark that the same construction works for any set D that is larger than the one defined in (6).

Definition 5.4. *For a set of vertices A, if we run the preceding construction, but with*

$$D = A \cup D^- \cup \{u,v\} \cup \{w \in \gamma : \gamma \text{ backtracks at } w\}.$$

instead of as defined in (6)*, then we call the resulting SAW-decomposition the A-*canonical SAW-decomposition of γ.

Lemma 5.5. *If $\zeta^{(1)}, \dots, \zeta^{(r)}$ is the canonical SAW-decomposition of γ then $r \le 2k_r(\gamma) + b + 1$, where b is the number of backtracks in γ.*

If $\zeta^{(1)}, \dots, \zeta^{(r)}$ is the A-canonical SAW-decomposition of γ then $r \le 2k_r(\gamma) + b + 1 + |A|$.

Proof. Every returning edge in γ increases r by at most 2, since it can create a new SAW component, and it can split an existing component into 2 pieces. Every backtrack in γ increases r by at most 1, since it can create a new SAW component. This proves the first statement; to prove the second, note that each vertex $v \in A$ creates at most one new component, since if $v \in D^-$ then it has no effect, while if $v \notin D^-$ then it has degree at most 2 in γ and so splitting the path that goes through v introduces at most one new component. □

We present the following lemma for the expected weight of a SAW-decomposition. Note, however, that the $\zeta^{(1)}, \dots, \zeta^{(r)}$ in the lemma do not need to be a SAW-decomposition (since we will sometimes find it convenient to apply the lemma to a subset of a SAW-decomposition).

Lemma 5.6. *Suppose $\zeta^{(1)}, \dots, \zeta^{(r)}$ are edge-disjoint, self-avoiding paths, where $\zeta^{(i)}$ is a path of length z_i between u_i and v_i. Let $m_1, \dots, m_r$ be any positive integers. Suppose that the $\bigcup_i \zeta^{(i)}$ has no ℓ-tangles, and at most $n^{0.1}$ edges. Let $D = \{u_1, v_1, \dots, u_r, v_r\}$. Then for any labelling σ_D on vertices in D,*

$$\tilde{\mathbb{E}}\left[\prod_i \prod_{e \in \zeta^{(i)}} W_e^{m_i} \mid \sigma_D\right] = (1 + O(n^{-0.8})) \prod_{i : m_i = 1} \frac{\sigma_{u_i}\sigma_{v_i} s^{z_i}}{n^{z_i}} \prod_{i : m_i > 1} \frac{\sigma_{u_i}\sigma_{v_i} s^{z_i} + p^{z_i}}{n^{z_i}}.$$

Proof. The proof follows from Lemma 4.3, with S and S' being empty and T equal to $\bigcup_i \zeta^{(i)}$. Then for any labelling τ that is compatible with σ_D on D,

$$\tilde{\mathbb{E}}\left[\prod_i \prod_{e \in \zeta^{(i)}} W_e^{m_i} \mid \tau\right] = (1 + O(n^{-0.8})) \prod_{i : m_i = 1} \prod_{(u,v) \in \zeta^{(i)}} \frac{\tau_u \tau_v s}{n} \prod_{i : m_i > 1} \prod_{(u,v) \in \zeta^{(i)}} \frac{\tau_u \tau_v s + p}{n}. \tag{7}$$

Now we take the average over all assignments τ that agree with σ_D on D. For a single segment $\zeta^{(i)}$, we have

$$2^{-(z_i - 2)} \sum_\tau \prod_{(u,v) \in \zeta^{(i)}} \frac{\tau_u \tau_v s}{n} = \frac{\sigma_{u_i}\sigma_{v_i} s^{z_i}}{n},$$

where the sum ranges over all $2^{z_i - 2}$ labellings τ on $\zeta^{(i)}$ that agree with σ_{u_i} and σ_{v_i} at the endpoints u_i and v_i of $\zeta^{(i)}$. Similarly,

$$2^{-(z_i - 2)} \sum_\tau \prod_{(u,v) \in \zeta^{(i)}} \frac{\tau_u \tau_v s + p}{n} = \frac{\sigma_{u_i}\sigma_{v_i} s^{z_i} + p^{z_i}}{n}.$$

Going back to (7) and taking the average over all labelling τ that agree with σ_D on D, we see that the average factorizes into the products because each u that is not in D appears in exactly one $\zeta^{(i)}$. Hence (since the probability of a labelling of m vertices under $\tilde{\mathbb{P}}$ is $(1 + o(1))2^{-m}$ by Lemma 4.1),

$$\begin{aligned}\tilde{\mathbb{E}}\left[\prod_i \prod_{e \in \zeta^{(i)}} W_e^{m_i} \mid \sigma_D\right] &= (1 + O(n^{-0.8})) 2^{-m} \sum_\tau \tilde{\mathbb{E}}\left[\prod_i \prod_{e \in \zeta^{(i)}} W_e^{m_i} \mid \tau\right] \\ &= (1 + O(n^{-0.8})) \prod_{i : m_i = 1} \frac{\sigma_{u_i}\sigma_{v_i} s^{z_i}}{n^{z_i}} \prod_{i : m_i > 1} \frac{\sigma_{u_i}\sigma_{v_i} s^{z_i} + p^{z_i}}{n^{z_i}},\end{aligned}$$

where the sum is over τ that are compatible with σ_D, and $m = \sum_i (z_i - 2)$ is the number of vertices in $\bigcup_i \zeta^{(i)}$ that do not belong to D. □

5.2 The weight of tangle-free paths

The most immediate application of Lemma 5.6 is to bound the expected weight of tangle-free paths. Note that the path in the following lemma is not assumed to be non-backtracking (that will come into play once we count paths, however).

Lemma 5.7. *Let γ be an ℓ-tangle-free path of length k, and let $\zeta^{(1)}, \dots, \zeta^{(r)}$ be a SAW-decomposition of γ with endpoints D. Then for any labelling σ_D on D,*

$$\tilde{\mathbb{E}}[X_\gamma \mid \sigma_D] \le 2^{r+1} s^k n^{-(k_n+k_r)}.$$

Proof. First, recall from Definition 5.2 that z_i is the length of $\zeta^{(i)}$, and that m_i is the number of times that $\zeta^{(i)}$ is traversed by γ. We also write m_e for the number of times that an edge e is traversed, so that if $e \in \zeta^{(i)}$ then $m_e = m_i$.

Since γ is tangle-free, Lemma 5.6 implies that

$$\begin{aligned}\left|\tilde{\mathbb{E}}\left[\prod_i \prod_{e\in\zeta^{(i)}} W_e^{m_i} \mid \sigma_D\right]\right| &\le (1+o(1)) \prod_{i:m_i=1} \frac{s^{z_i}}{n^{z_i}} \prod_{i:m_i>1} \frac{s^{z_i}+p^{z_i}}{n^{z_i}} \\ &\le 2^{r+1} \prod_{i:m_i=1} \frac{s^{z_i}}{n^{z_i}} \prod_{i:m_i>1} \frac{p^{z_i}}{n^{z_i}} \\ &= 2^{r+1} \prod_{\substack{e\in\gamma \\ m_e=1}} \frac{s}{n} \prod_{\substack{e\in\gamma \\ m_e>1}} \frac{p}{n} \\ &\le 2^{r+1} n^{-k_n-k_r} s^k,\end{aligned}$$

where the last inequality follows because $1 < p < s^2$, and so $m_e > 1$ implies $p < s^2 \le s^{m_e}$, and because the total number of distinct edges present in γ is $k_n + k_r$. Finally, we apply Lemma 5.5 to bound r. ☐

Lemma 5.8. *Fix vertices u and v, and a collection of vertices U with $|U| = O(1)$ that does not contain u or v. Let $\Gamma_{U,k}^{NT}$ be the set of ℓ-tangle-free paths from u to v of length k that go through every vertex in U, which have at most $b = O(1)$ backtracks, and which satisfy $1 \le k_r \le k^*$. Let A be an arbitrary set of vertices with $|A| \le n^{0.9}$. If $k/\ell = o(\log n/\log\log n)$ then*

$$\sum_{\gamma\in\Gamma_{U,k}^{NT}} \tilde{\mathbb{E}}[X_\gamma \mid \sigma_A] \le s^k n^{-|U|-2+o(1)}$$

Proof. Let $\Gamma_{k_r,k_n,q} \subset \Gamma_{U,k}^{NT}$ be the subset of paths that intersect $|A|$ in at most q vertices, and that satisfy $k_r(\gamma) = k_r$, $k_n(\gamma) = k_n$. Now fix k_r, k_n and q, and consider $\gamma \in \Gamma_{k_r,k_n,q}$. We apply Lemma 5.7 to the $(A \cap V(\gamma))$-canonical SAW-decomposition of γ, and bound r using Lemma 5.5:

$$|\tilde{\mathbb{E}}[X_\gamma \mid \sigma_A]| \le k^{3k_r+b+2+q} n^{-k_n-k_r} s^k \le k^q n^{-k_n-k_r+o(1)} s^k. \tag{8}$$

Now consider Lemma 4.8. Out of all paths going through U and satisfying $k_r(\gamma) = k_r$, $k_n(\gamma) = k_n$, at most a $(k/|A|)^q$-fraction have an intersection with A of size q. Hence,

$$|\Gamma_{k_r,k_n,q}| \le k^{3k_r+4k_r k/\ell} n^{k_n-|U|-1} \left(\frac{k}{|A|}\right)^q \le n^{k_n-|U|-1-0.1q+o(1)}.$$

Applying this bound to (8), and summing over q, we obtain a geometric series with ratio $kn^{-0.1}$. This series is bounded by $(1 + o(1)$ times its first term, and so

$$\sum_{q\ge0} \sum_{\gamma\in\Gamma_{k_r,k_n,q}} |\tilde{\mathbb{E}}[X_\gamma \mid \sigma_A]| \le n^{-k_r-|U|-1+o(1)} s^k.$$

Summing over $k_r \geq 1$ and k_n (for which there are at most $k = n^{o(1)}$ choices each), we obtain the claimed bound. $\square$

5.3 The weight of tangled paths

Lemma 5.6 may also be exploited to bound the weight of tangled paths. In order to do this, we first take the canonical SAW-decomposition of a path γ and then we split it into short and long segments: say the short segments are those of length at most 4ℓ. Now, the union of long segments cannot contain any ℓ-tangles, because any ball of radius ℓ does not even contain a single cycle. We may then apply Lemma 5.6 to the long segments, and we may bound the contribution of short segments through other means.

Lemma 5.9. *Let γ be a path of length k with t ℓ-tangles, where $\ell = o(\log n)$. Let $\zeta^{(1)}, \dots, \zeta^{(r)}$ be a SAW-decomposition of γ with endpoints D. If $k_r = k_r(\gamma)$ is bounded by an absolute constant then*

$$\tilde{\mathbb{E}}[X_\gamma \mid \sigma_D] \leq 2^r s^k n^{-k_n - t + o(1)}.$$

The proof of Lemma 5.9 is the only part of our argument where is is crucial that we are working with the tangle-free measure. To explain the intuition behind the result, consider two extreme cases of paths with many tangles: in the first case, γ keeps returning to the same vertex, but along different paths. Then $t = k_r$, and so the bound of Lemma 5.9 is of the same order as the bound of Lemma 5.7. In the second case, there are only a few short cycles in a small neighborhood, but each of them is traversed many times. Under the tangle-free measure, we know in advance that some of the edges in these cycles are absent from the graph. But an edge that a priori cannot appear has an expected weight of $O(n^{-m_e})$ (as opposed to an edge which could appear, which has expected weight of $O(n^{-1})$). Now, if there are t tangles, then the set of edges that needs to be cut under the tangle-free measure satisfies $\sum m_e \geq t$; hence the penalty of n^{-t} in Lemma 5.9.

Proof. Let $A \subset \{\zeta^{(1)}, \dots, \zeta^{(r)}\}$ denote the collection of paths that have length at most 4ℓ; let B be the others. With some abuse of notation, we say that $e \in A$ if $e \in \zeta$ for some $\zeta \in A$ (and similarly for B).

$$\tilde{\mathbb{E}}\left[\prod_{e \in \gamma} W_e \mid \sigma_D\right] = \tilde{\mathbb{E}}\left[\prod_{e \in A} W_e \tilde{\mathbb{E}}\left[\prod_{e \in B} W_e \mid J_A, \sigma_D\right] \mid \sigma_D\right]. \tag{9}$$

Recall, by the discussion preceding Lemma 5.9, that $\bigcup_{\zeta \in B} \zeta$ contains no ℓ-tangles Hence, by Lemma 5.6,

$$\left|\tilde{\mathbb{E}}\left[\prod_{e \in B} W_e^{m_e} \mid J_A, \sigma_D\right]\right| \leq (1 + o(1)) \left(\frac{s}{n}\right)^{\#\{e \in B: m_e = 1\}} 2^r \left(\frac{p}{n}\right)^{\#\{e \in B: m_e > 1\}}.$$

Plugging this into (9),

$$\begin{aligned}
\left|\tilde{\mathbb{E}}\left[\prod_{e \in \gamma} W_e \mid \sigma_D\right]\right| &\leq 2^{r+1} \tilde{\mathbb{E}}\left|\prod_{e \in A} W_e \mid \sigma_D\right| \left(\frac{s}{n}\right)^{\#\{e \in B: m_e = 1\}} \left(\frac{p}{n}\right)^{\#\{e \in B: m_e > 1\}} \\
&\leq 2^{r+1} \tilde{\mathbb{E}}\left|\prod_{e \in A} W_e \mid \sigma_D\right| n^{-|B|} s^{\sum_{e \in B} m_e}, \qquad (10)
\end{aligned}$$

where the last inequality follows because $p < s^2$ and $s > 1$.

Next, we will bound $\tilde{\mathbb{E}}[\prod_{e\in A}|W_e| \mid \sigma_D]$. Take $F\subset A$ and let Ω_F be the event that none of the edges in F appear. Then

$$\begin{aligned}\mathbb{E}|1_{\Omega_F}\prod_{e\in A}W_e^{m_e}| &= \prod_{e\in F}|p/n|^{m_e}\cdot\prod_{e\in A\setminus F}\mathbb{E}|W_e|^{m_e}\\ &\le (p/n)^{\sum_{e\in F}m_e}\cdot(2a/n)^{|A\setminus F|}\\ &\le (2a/n)^{|A|+\sum_{e\in F}(m_e-1)},\end{aligned}\tag{11}$$

where we have used the very crude bounds $\mathbb{E}|W_e|^m \le 2a/n$ and $p\le a$.

Now, let $H=\{F\subset A : \sum_{e\in F}m_e \ge t \text{ and } |F|\le k_r\}$. We claim that $\Xi\subset\bigcup_{F\in H}\Omega_F$. Indeed, consider any graph $G\in\Xi$. Let F be a maximal subset of $E(\gamma)\setminus E(G)$ with the property that $\gamma\setminus F$ is connected. Then $\gamma\setminus F$ is tangle-free (if there is a tangle, then some edge in that tangle must be absent from G and removing that edge cannot disconnect $\gamma\setminus F$, thereby contradicting the maximality of F). Since γ has t tangles, we must have deleted t edges (counting multiplicity) to remove then; hence $\sum_{e\in F}m_e\ge t$. Since $\gamma\setminus F$ is connected, we must have $|F|\le k_r$. Hence, $F\in H$. Since $F\cap E(G)=\varnothing$, we have $G\in\Omega_F$. Recalling that G was an arbitrary element of Ξ, we must have $\Xi\subset\bigcup_{F\in H}\Omega_F$ as claimed. In particular, we have $\sum_{F\in H}1_{\Omega_F}\ge 1_\Xi$ and so, going back to (11),

$$\begin{aligned}\tilde{\mathbb{E}}|\prod_{e\in A}W_e^{m_e}| &= \frac{1}{\Pr(\Xi)}\mathbb{E}1_\Xi|\prod_{e\in A}W_e^{m_e}|\\ &\le \frac{1}{\Pr(\Xi)}\sum_{F\in H}\mathbb{E}1_{\Omega_F}|\prod_{e\in A}W_e^{m_e}|\\ &\le \frac{|H|}{\Pr(\Xi)}(2a/n)^{|A|+t-k_r},\end{aligned}$$

where we have used the fact that for every $F\in H$, $\sum_{e\in F}(m_e-1)=\sum_{e\in F}m_e-|F|\ge t-k_r$. Finally, $\Pr(\Xi)\ge\frac{1}{2}$, $|H|\le k^{k_r}$, and $|A|\le 4\ell r\le 16\ell k_r$ so

$$\tilde{\mathbb{E}}|\prod_{e\in A}W_e^{m_e}|\le 2k^{k_r}(2a/n)^{|A|+t-k_r}\le 2(2a)^{16\ell k_r+t}k^{k_r}n^{-|A|-t+k_r}.\tag{12}$$

Combining (12) with (10), we have

$$\left|\tilde{\mathbb{E}}\left[\prod_{e\in\gamma}W_e \mid \sigma_D\right]\right|\le 2^{r+2}k^{k_r}(2a)^{16\ell k_r+t}s^k n^{-|B|-|A|-t+k_r}.$$

Now, $|A|+|B|$ is the total number of edges traversed in γ, which is equal to k_n+k_r. Hence, the right hand side above is at most

$$2^{r+2}k^{k_r}(2a)^{16\ell k_r+t}s^k n^{-k_n-t},$$

which is smaller than the claimed bound. ☐

Lemma 5.10. *Fix vertices u and v, and a collection of vertices U with $|U|=O(1)$ that does not contain u or v. Let $\Gamma^T_{u,v,U,k}$ be the set of paths of length k from u to v with at least one ℓ-tangle, that pass through every vertex in U, and which have at most $b=O(1)$ backtracks. Let A be an arbitrary set of vertices with $|A|\le n^{0.9}$. If $k/\ell=o(\log n/\log\log n)$ then*

$$\sum_{\gamma\in\Gamma^{NT}_{U,k}}\tilde{\mathbb{E}}[X_\gamma\mid\sigma_A]\le s^k n^{-|U|-2+o(1)}$$

Proof. This proof is quite similar to the proof of Lemma 5.8, except that we apply Lemmas 5.9 and 4.10 instead of Lemmas 5.7 and 4.8.

Let $\Gamma_{k_n,q,t} \subset \Gamma^T_{U,k}$ be the subset of paths that intersect $|A|$ in at most q vertices, and that satisfy $k_n(\gamma) = k_n$, $t(\gamma) = t$. For fixed k_r, k_n and q and $\gamma \in \Gamma_{k_n,q,t}$, we apply Lemma 5.7 and Lemma 5.5 to obtain

$$|\tilde{\mathbb{E}}[X_\gamma \mid \sigma_A]| \le 2^{3k_r+b+2+q} n^{-k_n-t+o(1)} s^k \le 2^q n^{-k_n-t} s^k. \tag{13}$$

Now apply Lemma 4.10, noting that at most a $(k/|A|)^q$-fraction of the paths therein have an intersection with A of size q. Hence,

$$|\Gamma_{k_r,k_n,q}| \le k^{3k_r+4k_r(1+k/\ell+2t)} n^{k_n-t-|U|-1+o(1)} \left(\frac{k}{|A|}\right)^a \le n^{k_n-t-|U|-1-0.1a+o(1)}.$$

Applying this bound to (13), and summing over q, we obtain a geometric series with ratio $kn^{-0.1}$. This series is bounded by $(1+o(1)$ times its first term, and so

$$\sum_{q\ge 0} \sum_{\gamma \in \Gamma_{k_n,q,t}} |\tilde{\mathbb{E}}[X_\gamma \mid \sigma_A]| \le n^{-t-|U|-1+o(1)} s^k.$$

Summing over $t \ge 1$ and k_n (for which there are at most $k = n^{o(1)}$ choices each), we obtain the claimed bound. ☐

5.4 The first moment

Proof of Theorem 5.1. We further divide up Γ as follows:

- $\Gamma_0 \subset \Gamma$ is the subset of self-avoiding paths (i.e. paths with $k_r = 0$);
- $\Gamma_1^{NT} \subset \Gamma$ is the subset of paths with $1 \le k_r \le k^*$ and no ℓ-tangles;
- $\Gamma_1^T \subset \Gamma$ is the subset of paths with $1 \le k_r \le k^*$ and some ℓ-tangles; and
- $\Gamma_2 \subset \Gamma$ is the subset of paths with $k_r \ge k^*$.

Let Γ_0' consist of the paths in Γ' that are self-avoiding. Then for $\gamma \in \Gamma_0'$, its A-canonical SAW-decomposition consists of only a single segment, which is traversed only once. By Lemma 5.6,

$$\tilde{\mathbb{E}}[X_\gamma \mid \sigma_A] = (1+o(1))\tilde{\mathbb{E}}[X_\gamma \mid \sigma_u, \sigma_v] = (1+o(1))\frac{\sigma_u \sigma_v s^k}{n^k}.$$

Now, there are $(1 + O(n^{-0.8}))n^{k-1}$ paths in Γ_0' and so

$$\tilde{\mathbb{E}} \sum_{\gamma \in \Gamma_0'} [X_\gamma \mid \sigma_A] = (1 + O(n^{-0.8}))\frac{\sigma_u \sigma_v s^k}{n}. \tag{14}$$

This is the right hand side as claimed in the theorem; hence, we will need to show that the contributions of all other paths are negligible.

First, consider $\gamma \in \Gamma_1^{NT}$. By Lemma 5.8 with $U = \varnothing$ and $b = 0$, we see that the contribution of Γ_1^{NT} is at most $s^k n^{-2+o(1)}$. This is indeed of a smaller order than (14), and so we may ignore Γ_1^{NT}. The argument for $\gamma \in \Gamma_1^T$ is identical, except that we apply Lemma 5.10. For Γ_2, we simply appeal to Corollary 4.7, which implies that Γ_2's contribution is at most $n^{-4+o(1)}$, and therefore negligible. ☐

6 The second moment

Theorem 6.1. *Let $\Gamma_{u,v}(k)$ be the collection of non-backtracking paths between u and v of length k. If $k \geq \frac{\log n}{2\log s - \log p}$ and $k/\ell = o(\log n/\log\log n)$ and A is a set of vertices with $|A| \leq n^{0.9}$ then*

$$\tilde{\mathbb{E}}\left[\left(\sum_{\gamma\in\Gamma_{u,v}(k)} X_\gamma\right)^2 \mid \sigma_A\right] \leq (1+o(1))2n^{-2}s^{2k}\left(\frac{s^2}{s^2-p}\right)^2.$$

Suppose γ_1 and γ_2 are a pair of paths of length k from u to v. By reversing γ_2 and appending it to γ_1, we obtain a single path (γ, say) from u to itself which passes through v and backtracks at most once (at v). We consider the set of all γ that can be obtained in this way, and divide them into 4 classes:

- Γ_0 is the collection of such paths with $k_r = 0$. Such paths begin with a self-avoiding walk from u to v, after which they backtrack at v and walk back to u along exactly the same path. They have k edges, $k-1$ vertices, and every edge is visited twice.
- Γ_1 is the collection of such paths with $k_r = 1$. Ignoring multiplicity of edges, these paths look like a simple cycle with up to two "tails".
- Γ_2^{NT} is the collection of such paths with $2 \leq k_r \leq k^*$ and no tangles.
- Γ_2^T is the collection of such paths with $2 \leq k_r \leq k^*$ and at least one tangle.
- Γ_3 is the collection of such paths with $k^* \leq k_r$.

First, we consider Γ_0. The canonical SAW-decomposition of $\gamma \in \Gamma_0$ has exactly one component, and it is traversed twice. By Lemma 5.6, if A does not intersect with the interior of γ then $\tilde{\mathbb{E}}[X_\gamma \mid \sigma_A] = (1+o(1))(p/n)^k$. There are $(1+o(1))n^{k-1}$ such paths, giving a total weight of $(1+o(1))p^k/n$. As in the proofs of Lemmas 5.8 and 5.10, the contribution of γ whose interiors do intersect with A is negligible: each such intersection increases the right hand side of Lemma 5.6 by a factor of 2, while reducing the number of possible paths by a factor of $(k/|A|) \leq n^{-0.1+o(1)}$. Hence,

$$\sum_{\gamma\in\Gamma_0} \tilde{\mathbb{E}}[X_\gamma \mid \sigma_A] = (1+o(1))p^k/n. \tag{15}$$

Next, we consider Γ_1. Here we have two cases, depending on whether the returning edge occurs before or after the first visit to v. Let $\Gamma_1' \subset \Gamma_1$ be the collection of paths where the returning edge occurs on or before the kth step. In this case, the second half of the path must consist only of old edges (since if there is a new edge, then there would have to be a second returning edge). Since the first half of the path contains exactly one cycle, there are only two choices for the returning path: it may traverse the cycle in either direction. Note that because the two halves of the path have the same number of steps, the second half of the path must traverse the cycle the same number of times as the first half did. Hence, every edge in the cycle is traversed an even number of times, and every other edge is traversed twice.

Now, let $\Gamma_1'(\ell, c)$ be the set of cycles in Γ_1' that have ℓ edges in their cycle, each of which has $m_e = 2c$. Every $\gamma \in \Gamma_1'(\ell, c)$ has $k-(c-1)\ell$ distinct edges (each of which is traversed at least twice) and $k-(c-1)\ell$ distinct vertices (including u and v). By Lemma 5.6, if γ's interior does not intersect A then

$$\tilde{\mathbb{E}}[X_\gamma \mid \sigma_A] = (1+o(1))(p/n)^{k-(c-1)\ell}$$

To bound $|\Gamma_1'(\ell,c)|$, note that there are at most k choices for which edge is returning, and k choices for where it should return to. There are $n^{k-(m_c-1)\ell/2-2}$ ways to choose the vertices and 2 choices for which way to traverse the cycle on the second half of the path. As before, this bound on $|\Gamma_1'(\ell,c)|$ decreases exponentially if we require the paths to intersect with A; hence, we can ignore such intersecting paths. All together, we have

$$\sum_{\gamma\in\Gamma_1'(\ell,c)} \tilde{\mathbb{E}}[X_\gamma \mid \sigma_A] \le 2k^2(1+o(1))p^k n^{-2} = p^k n^{-2+o(1)}.$$

Taking the sum over all ℓ and c (there are at most k choices for each),

$$\sum_{\gamma\in\Gamma_1'} \tilde{\mathbb{E}}[X_\gamma \mid \sigma_A] \le k^2 p^k n^{-2+o(1)} \le p^k n^{-2+o(1)}. \tag{16}$$

Now consider $\Gamma_1'' = \Gamma_1 \setminus \Gamma_1'$. For path $\gamma\in\Gamma_1''$, the first k steps do not contain a returning edge; hence the first k steps make up a simple path. Let i be minimal so that the $(k+i+1)$th step of γ is new; let j be such that the $2k-j$th step of γ is returning. It follows that the first j edges of γ consist of a simple path where each edge is traversed twice. The same holds for edges $k-i+1$ through $k-1$. The rest of γ consists of a simple cycle of length $2k-2(i+j)$, each edge of which is traversed once. Let $\Gamma_1''(i,j)$ denote the set of such paths. By Lemma 5.6, if γ's interior does not intersect A then the expected weight of $\gamma\in\Gamma_1''(i,j)$ is

$$\tilde{\mathbb{E}}[X_\gamma \mid \sigma_A] = (1+o(1))(p/n)^{i+j}(s/n)^{2k-2(i+j)}.$$

Now, $|\Gamma_1''(i,j)| = (1+o(1))n^{2k-i-j-2}$ because $\gamma\in\Gamma_1''(i,j)$ has $2k-i-j$ distinct vertices (including u and v), and once those vertices and their order is fixed then γ is determined. As before, we may ignore paths whose interiors intersect A, and hence

$$\sum_{\gamma\in\Gamma_1''(i,j)} \tilde{\mathbb{E}}[X_\gamma \mid \sigma_A] = (1+o(1))n^{-2}s^{2k-2(i+j)}p^{i+j} = (1+o(1))n^{-2}s^{2k}\left(\frac{p}{s^2}\right)^{i+j}.$$

Summing over i and j, we have

$$\sum_{\gamma\in\Gamma_1''} \tilde{\mathbb{E}}[X_\gamma \mid \sigma_A] \le (1+o(1))n^{-2}s^{2k}\sum_{i,j=0}^{\infty}\left(\frac{p}{s^2}\right)^{i+j} = (1+o(1))n^{-2}s^{2k}\left(\frac{T}{T-1}\right)^2, \tag{17}$$

where $T = s^2/p$, which is larger than 1 if we are above the threshold.

To summarize,

Lemma 6.2.

$$\sum_{\gamma\in\Gamma_1} \tilde{\mathbb{E}}[X_\gamma \mid \sigma_A] \le (1+o(1))n^{-2}s^{2k}\left(\frac{s^2}{s^2-p}\right)^2.$$

Proof. We combine (16) with (17), noting that $k\to 0$ and $p<s^2$, which implies that (17) dominates. □

Proof of Theorem 6.1. By the discussion following the statement of the proposition,

$$\begin{aligned}\left(\sum_{\gamma\in\Gamma_{u,v}(k)} X_\gamma\right)^2 &= \sum_{\gamma_1,\gamma_2\in\Gamma_{u,v}(k)} X_{\gamma_1}X_{\gamma_2}\\ &= \sum_{\gamma\in\Gamma_0} X_\gamma + \sum_{\gamma\in\Gamma_1} X_\gamma + \sum_{\gamma\in\Gamma_2^{NT}} X_\gamma + \sum_{\gamma\in\Gamma_2^{T}} X_\gamma + \sum_{\gamma\in\Gamma_3} X_\gamma.\end{aligned}$$

Let us bound the expectations of each of these sums. First, (15) implies that

$$\tilde{\mathbb{E}} \sum_{\gamma \in \Gamma_0} X_\gamma \le (1+o(1))p^k/n \le (1+o(1))s^{2k}/n^2,$$

given the assumption on k in the proposition. Next, Lemma 6.2 gives

$$\tilde{\mathbb{E}} \sum_{\gamma \in \Gamma_1} X_\gamma \le (1+o(1))s^{2k}/n^2 \left(\frac{s^2}{s^2-p}\right)^2.$$

Combining the contributions from Γ_0 and Γ_1 gives the bound in the Proposition. It remains to show that all the other contributions are negligible. Setting $U = \{v\}$, Lemmas 5.8 and 5.10 imply that the contributions of Γ_2^T and Γ_2^{NT} are both $s^{2k}n^{-3+o(1)}$, while Corollary (4.7) implies that the contribution of Γ_3 is negligible also. ◻

7 Cross moments

Theorem 7.1. *Let u, u', v, v' be distinct vertices, and let $\Gamma_{u,v}(k)$ be the collection of non-backtracking paths between u and v of length k. Set $N_{u,v,k} = \sum_{\gamma \in \Gamma_{u,v}(k)} X_\gamma$. If $k/\ell = o(\log n/\log\log n)$ and A is a set of vertices containing u, u', v, v' with $|A| \le n^{0.1}$ then*

$$\tilde{\mathbb{E}}[N_{u,v,k}N_{u',v',k} \mid \sigma_A] = (1+O(n^{-1/4}))\mathbb{E}[N_{u,v,k} \mid \sigma_A]\mathbb{E}[N_{u',v',k} \mid \sigma_A]$$

Proof. Expanding the definition of $N_{u,v,k}$, we have

$$N_{u,v,k}N_{u',v',k} = \sum_{\gamma \in \Gamma_{u,v}(k)} \sum_{\gamma' \in \Gamma_{u',v'}(k)} X_\gamma X'_\gamma.$$

To prove the theorem, it suffices to consider $\mathbb{E}[X_\gamma X_{\gamma'} \mid \sigma_A]$ for each pair γ and γ'. For the main term, consider a pair γ, γ' of paths that are self-avoiding and vertex-disjoint, and whose interiors do not intersect A. Then Lemma 5.6 applies with $(\zeta^{(1)}, \zeta^{(2)}) = (\gamma, \gamma')$, and we have

$$\tilde{\mathbb{E}}[X_\gamma X_{\gamma'} \mid \sigma_A] = (1+O(n^{-0.8}))\frac{s^{2k}}{n^{2k}}.$$

There being at most n^{2k-2} such pairs (γ, γ'), we have

$$\sum_{\gamma,\gamma'} \tilde{\mathbb{E}}[X_\gamma X_{\gamma'} \mid \sigma_A] = (1+O(n^{-0.8}))\frac{s^{2k}}{n^2} = (1+O(n^{-1/4}))\mathbb{E}[N_{u,v,k} \mid \sigma_A]\mathbb{E}[N_{u',v',k} \mid \sigma_A],$$

where the second equality follows by Theorem 5.1, and the sum ranges over pairs γ, γ' that are self-avoiding and vertex-disjoint, and whose interiors do not intersect A. Since the right hand side above matches the right hand side of the theorem, it remains to show that all other pairs of paths have a negligible contribution. Let us first consider pairs γ, γ' that are self-avoiding and vertex-disjoint, but whose interiors do intersect A. If the intersection with A has size q, then Lemma 5.6 applied to the A-canonical SAW-decompositions of γ and γ' implies that

$$\tilde{\mathbb{E}}[X_\gamma X_{\gamma'} \mid \sigma_A] = (1+O(n^{-0.8}))2^q\frac{s^{2k}}{n^{2k}}.$$

On the other hand, there are at most $(2k/|A|)^q n^{2k-2} \le n^{2k-2-0.9q+o(1)}$ pairs γ, γ' that have an intersection with A of size q. Taking the geometric sum over $q \ge 1$, we see that the contribution of such pairs is at most

$$(1+O(n^{-0.8}))\frac{s^{2k}}{n^2}\sum_{q\ge 1} 2^q n^{-0.9q} \le O(n^{-0.7})\frac{s^{2k}}{n^2},$$

and so the contribution of these pairs is negligible.

We move on to pairs γ, γ' that do intersect; we want to show that they give a negligible contribution. It is possible to prove this directly, by considering many different cases for γ, γ', and their intersection. In order to leverage our earlier work, however, let us consider a modified path γ'', obtained by first traversing γ and then moving to u' and traversing γ'. Then γ'' has length $2k+1$. It is a path from u to v', and it traverses two predetermined vertices, v and u'. Moreover, $|X_\gamma X_{\gamma'}| \le n|X_{\gamma''}|$, because the two sides are equal, except that $X_{\gamma''}$ counts the edge (v, u') once more than $X_\gamma X_{\gamma'}$ does, and that edge has a weight of at least n^{-1}. Now, since γ and γ' intersect, it follows that $k_r(\gamma'') \ge 1$. Also, since γ and γ' are non-backtracking, it follows that γ'' has at most 2 backtracks (at v and u'). We will show that

$$\sum_{\gamma''} \mathbb{E}[X_{\gamma''}|\sigma_A] \le s^{2k} n^{-4+o(1)}, \tag{18}$$

where the sum ranges over γ'' from u to v' of length $2k+1$ that go through v and u' and have at most 2 backtracks. From this, it will follow that

$$\sum_{\gamma,\gamma'} \mathbb{E}[X_\gamma X_{\gamma'}|\sigma_A] \le s^{2k} n^{-3+o(1)},$$

where the sum ranges over non-backtracking paths γ and γ' of length k that go from u to v and u' to v' respectively, and that intersect at least once.

To prove (18), we distinguish three cases:

- Γ_1^{NT} denotes the set of γ'' with $1 \le k_r \le k^*$ and no ℓ-tangles;
- Γ_1^{T} denotes the set of γ'' with $1 \le k_r \le k^*$ and at least 1 ℓ-tangle; and
- Γ_2 denotes the set of γ'' with $k_r > k^*$.

By Lemma 5.8 with $U = \{v, u'\}$, the total contribution of Γ_1^{NT} is at most $s^{2k+1} n^{-4+o(1)}$. Similarly, Lemma 5.10 implies that the total contribution of Γ_1^T is at most $s^{2k+1} n^{-4+o(1)}$. Finally, Corollary 4.7 implies that Γ_2 contributes at most $n^{-4+o(1)}$. $\square$

8 Analysis of the algorithm

We recall some notation from the algorithm: G' is G, but with a random collection of $\lceil\sqrt{n}\rceil$ vertices removed. We choose some node $w^* \in G \smallsetminus G'$ that has many neighbors in G', and we set T to be those neighbors. For each $v \in V'$, S_v denotes the shell around v of radius $R = \log\log\log\log n$. V_j is the set of vertices that we consider in the jth iteration of the algorithm, and $N^{(j)}_{u,u',k}$ is the total weight of non-backtracking paths of length k from u to u' that only visit vertices in V_j.

The distribution of G' is simply a block model with fewer vertices, that is $G(n-\lceil\sqrt{n}\rceil, a/n, b/n)$. Let $G_j = (V_j, E_j)$ denote the graph obtained at iteration j. We need to argue that conditioned on a vertex neighborhood being removed, the distribution on the remaining graph is drawn (approximately) from the block model. The technical issue here is that the removed vertices are correlated and moreover we need to condition on some of their spins. In fact, this is the main reason for removing i.i.d. random vertices to obtain G_j instead of the more natural idea of removing neighborhoods. However, even if we remove i.i.d. vertices, we need to argue that the resulting graph looks like a block model due to the correlation and conditioning. Since the neighborhood of a single neighborhood does not contain too many vertices, this is possible to show.

For a vertex v in G_j, let T denote the set $S_v \cup Y$. We will be interested in the distribution of $(G_j, T, \sigma(T))$ and we would like to couple it with a configuration of $(G', T', \sigma'(T'))$ where G' is drawn from $G(n - \delta n + |T|, a/n, b/n)$ and T' is some fixed set of vertices of size $|T|$.

Lemma 8.1. *Fix a vertex $v \in V'$. Let P_1 denote the distribution of $(G_j, T, \sigma(T))$ conditioned on the graph structure of $B(v,R)$ and the events*

$$|B(v,R)| \le n^{0.1}, \quad B(v,R-1) \cap V_j = \varnothing, \quad S_v \subset V_j, \quad \sigma(T) = \sigma', \quad \sigma(B(v,R-1)) = \sigma''$$

for some σ', σ'' and where $T = S_v \cup Y$.

Let P_2 denote the distribution of $G(n - \delta n + |T|, a/b, b/n)$ conditioned on some set T' of size $|T|$ and $\sigma(v) = \sigma'(\psi(v))$, where ψ is a one to one map from T' to T. Then for large enough n, the measures P_1 and P_2 satisfy that

$$d_{TV}(P_1, P_2) \le n^{-0.3}.$$

Proof. The proof will couple the two configurations of spins and edges in the two graphs on $n(1-\delta) + |T|$ vertices. The coupling proceeds in the following way:

- $\sigma(T)$ and $\sigma(T')$ are coupled in the obvious way.
- Then we try to couple all other spins so they are completely identical.
- Finally, once the spins are identical, we will include exactly the same edges. This is possible since different edges are independent and the probabilities of including edges just depend on the end points.

The only non-trivial part of this proof is showing that we can perform the second step with high probability. Note that in $G(n - \delta n + |T|, a/b, b/n)$ all of the spins outside T' are i.i.d. Bernoulli 1/2. The conditional distribution outside T in G_j under the conditioning is also i.i.d. (since no edges are revealed). However, now the spins are biased, since we know they were not connected to the vertices of $B(v, r-1)$. Indeed, for each vertex v outside T we have that

$$\frac{P_1[\sigma_v = +]}{P_1[\sigma_v = -]} = \left(\frac{1 - \frac{a}{n}}{1 - \frac{b}{n}} \right)^{n_+ - n_-},$$

where $n_\pm$ is the number of $\pm$ in $\sigma(B(v, R-1))$. Thus

$$P_1[\sigma_v = +] = \frac{1}{2} + O(|B(v,R)|/n) = \frac{1}{2} + O(n^{-0.9}).$$

It is well known, see e.g. [16] that

$$d_{TV}(\mathrm{Bin}(n, 1/2), \mathrm{Bin}(n, 1/2 + x)) = O(x\sqrt{n})$$

which therefore implies that

$$d_{TV}(P_1, P_2) \le O(n^{-0.4}) \le n^{-0.3},$$

as needed. ∎

First we note that with high probability $|Y| = \lceil \sqrt{\log\log n} \rceil$ since the probability that there exists a vertex with that number of neighbours tends to one as the number of neighbours has a Binomial distribution. Similarly, with high probability the graph G' is ℓ-tangle-free. We will condition on these two facts from now on. We may assume without loss of generality that $\sigma(w^*) = +$.

We will denote $M_v = \sum_{v \in S_v} \sigma_v$ and $M^* = \sum_{v \in Y} \sigma_v$. Note that M^* is a sum of i.i.d. signs, each of which has probability $\frac{s}{p}$ (since we conditioned on $\sigma(w^*) = +$). Hence, with high probability

$$|M_* - \frac{s}{p}|Y|| \le |Y|^{3/4}. \tag{19}$$

In the statement of the lemma below we imagine that instead of running one iteration of the algorithm, we run infinitely many iterations so that $J_v \neq 0$ for all v. We will later show that w.h.p., in fact $J_v \le \log n$.

Lemma 8.2. *For a random vertex v, and any $\epsilon > 0$, it holds that*

$$\mathbb{P}\left[\left|\sum_{u \in Y, u' \in S_v} N^{(J_v)}_{u,u',k} - \frac{s'^k}{n} M_z M_*\right| > \frac{\epsilon s'^{k+R}}{n}|Y|\right] \to 0$$

as $n \to \infty$.

Proof. We will let $\mathbb{P}_2$ denote the measure defined by P_2 in Lemma 8.1 and write $\tilde{\mathbb{P}}_2$ for $\mathbb{P}_2$ conditioned on being ℓ-tangle-free. Then by Theorem 5.1 we have that.

$$\left|\tilde{\mathbb{E}}_2\left[\sum_{u \in Y, u' \in S_v} N^{(J_v)}_{u,u',k} - \frac{s'^k}{n} M_z M_* \mid \sigma_{S_v}, \sigma_Y\right]\right| = O(\frac{|S_v||Y|s'^k}{n^{5/4}}).$$

By Theorem 6.1 and Theorem 7.1 and using Cauchy Schwarz, we have that:

$$\tilde{\mathbb{E}}_2\left[\left(\sum_{u \in Y, u' \in S_v} N^{(J_v)}_{u,u',k} - \tilde{\mathbb{E}}_2 N^{(j)}_{u,u',k}\right)^2 \mid \sigma_{S_v}, \sigma_Y\right] = O(\frac{|S_v||Y|(|Y| + |S_v|)s'^{2k}}{n^2}).$$

and therefore

$$\mathbb{E}_2\left[\left|\sum_{u \in Y, u' \in S_v} N^{(J_v)}_{u,u',k} - \tilde{\mathbb{E}}_2 N^{(j)}_{u,u',k}\right| \mid \sigma_{S_v}, \sigma_Y\right] = O(\frac{|S_v|^{1/2}|Y|^{1/2}(|Y| + |S_v|)^{1/2} s'^k}{n})$$

Therefore by the triable inequality

$$\tilde{\mathbb{E}}_2\left[\left|\sum_{u \in Y, u' \in S_v} N^{(j)}_{u,u',k} - \frac{s'^k}{n} M_z M_*\right| \mid \sigma_{S_v}, \sigma_Y\right] \le O(\frac{s'^k}{n})\left(n^{-1/4}|Y||S_v| + |S_v|^{1/2}|Y|^{1/2}(|Y| + |S_v|)^{1/2}\right)$$

By Markov inequality, we therefore get that

$$\tilde{\mathbb{P}}_2\left[\left|\sum_{u \in Y, u' \in S_v} N^{(J_v)}_{u,u',k} - \frac{s'^k}{n} M_z M_*\right| > \frac{\epsilon s'^{k+R}}{n}|Y| \mid \sigma_{S_v}, \sigma_Y\right] \le O\left(s'^{-R}(n^{-1/4}|S_v| + |S_v||Y|^{-1/2} + |S_v|^{1/2})\right)$$

Since $\mathbb{E}_2\sqrt{|S_v|} \le \sqrt{p'^R} = o(s^R), \mathbb{E}_2[|S_v|] = p'^R$, and since $|Y| >> p'^R$, it follows that

$$\mathbb{E}_2[s'^{-R}\left(n^{-1/4}|S_v| + |S_v||Y|^{-1/2} + |S_v|^{-1/2}\right)] \to 0,$$

thus proving that

$$\tilde{\mathbb{P}}_2\left[\left|\sum_{u\in Y,u'\in S_v} N^{(J_v)}_{u,u',k} - \frac{s'^k}{n} M_z M_*\right| > \frac{\epsilon s'^{k+R}}{n}|Y|\right] \to 0$$

Since $\frac{d\tilde{\mathbb{P}}_2}{d\mathbb{P}_2} \leq 1 + o(1)$ (by Lemma 4.1) it follows that the same holds for $\mathbb{P}_2$ as needed. Applying Lemma 8.1 completes the result by proving that the same statement holds for $\mathbb{P}_1$. ☐

Definition 8.3. *Let T be a Galton-Watson branching process with Poisson(p) offspring distribution rooted at ρ and let $Z = Z(R)$ denote $|S_R(\rho)|$, the size of the Rth generation. Let η be given by the free measure Ising model on T with inverse temperature $\beta = \tanh^{-1}(s/p)$ constructed by first dividing the tree into clusters according to bond percolation with probability s/p and then assigning each cluster a random spin and assigning all vertices in that cluster that spin. We define a triple (η, η^+, η^-) to be the configuration generated this way where the cluster containing the root is given a random plus or minus spin. Let $\zeta = \eta_\rho$, let $\Psi = \sum_{v\in S_R(\rho)} \eta_v$ and define $\Psi^\pm$ similarly.*

Lemma 8.4. *There exist a $D > 0$ and $\epsilon > 0$ such that*

$$\mathbb{P}[\Psi^+ \geq \xi D s^R] \geq \frac{1}{2} + 2\epsilon. \tag{20}$$

Proof. Note that standard branching process theory implies that $\mathbb{E}[Z^2] = O(p^{2R})$. Furthermore, by symmetry, Ψ is symmetric about 0 and so if ξ is an independent uniform on $[0,1]$ then for any $D > 0$,

$$\mathbb{P}[\Psi \geq \xi D s^R] = \frac{1}{2}.$$

Moreover, analysis of multi-type branching processes going back to Kesten and Stigum [9] shows that

$$\mathbb{E}\Psi^2 = O(s^{2R})$$

provided $s^2 > p$. Also with this construction it is clear that $\Psi^+ - \Psi^-$ is simply twice the size of the percolation component of the origin intersected with level R. This is exactly given by a branching process with Poisson(s) offspring distribution so

$$\liminf \mathbb{P}[\Psi^+ - \Psi^- > s^R] > 0$$

and hence for large enough D and fixed $\epsilon > 0$,

$$\mathbb{P}[\Psi^+ \geq \xi D s^R] \geq \frac{1}{2} + 2\epsilon. \tag{21}$$

as needed. ☐

Lemma 8.5. *For $1 \leq i \leq \log n$ let (ζ_i, Z_i, Ψ_i) be iid copies of (ζ, Z, Ψ) above for $R = \log\log\log\log n$. For $v_1, \ldots, v_{\log n}$ be uniformly chosen vertices in V,*

$$d_{TV}\left(\{(\zeta_i, Z_i, \Psi_i)\}_{1\leq i\leq \log n}, \{(\sigma_{v_i}, |S_{v_i}|, M_{v_i})\}_{1\leq i\leq \log n}\right) \to 0$$

as $n \to \infty$.

Proof. We establish the result by coupling the two processes. By Markov's inequality with high probability $\sum_{i=1}^{\log n} Z_i \leq \log^2 n$.

Consider the following two processes. We reveal the branching process trees by sequentially revealing for each vertex how many children of each spin it has (Poisson(a) of the same spin and Poisson(b) of the opposite spin) down to level R in a breadth-first manner.

Similar we can reveal the neighbourhoods of the v_i and their spins in G by sequentially revealing the neighbours and spins of the vertices. The number of neighbours with the same spin is $\mathrm{Bin}(n - O(\log^2 n), (a+O(\frac{1}{n}\log^2 n))/n)$ and with different spins $\mathrm{Bin}(n-O(\log^2 n), (b+O(\frac{1}{n}\log^2 n))/n)$ where the $O(\log^2 n)$ quantities come from the change in the number of spins having already been revealed.

We couple these two processes with the usual coupling of Poisson and Binomial of these quantities, in each step we fail with probability $O((\log^2 n)/n)$ and so the coupling altogether fails with probability $o(1)$. $\square$

Denote the indicator variables

$$\mathcal{A}_{j,v} = \mathrm{sgn}(M_v + Ds^R \xi_{j,v}).$$

and

$$\mathcal{B}_{j,v} = \mathrm{sgn}(\sum_{u \in Y, u' \in S_v} N^{(j)}_{u,u',k} + D\frac{s'^{k+R+1}}{p'n}|Y|\xi_{j,v})$$

Lemma 8.6. *We have that*

$$\mathbb{P}\left[\frac{1}{|V'|}\sum_{v\in V'} \sigma_v \mathcal{A}_{j,v} \geq (3/2)\epsilon\right] \to 1$$

Proof. Let $v_1, \dots, v_{\log n}$ be a uniform sample without replacement from V'. Take the coupling in Lemma 8.5, and set

$$\mathcal{A}'_{j,v_i} = \mathrm{sgn}(\Psi_i + Ds^R \xi_{j,v_i}).$$

$$\mathbb{P}[\sum_{i=1}^{\log n} \sigma_{v_i} \mathcal{A}_{j,v_i} \geq 1.9\epsilon \log n] = \mathbb{P}[\sum_{i=1}^{\log n} \sigma_{v_i} \mathcal{A}'_{j,v_i} \geq 1.9\epsilon \log n] + o(1)$$

By equation (20)

$$\mathbb{P}[\sigma_{v_i}\mathrm{sgn}(\Psi_i + Ds^R\xi_i)] \geq 2\epsilon$$

so by a large deviation bound, since the $\mathcal{A}'_{j,v_i}$ are independent,

$$\mathbb{P}[\sum_{i=1}^{\log n} \sigma_{v_i} \mathcal{A}'_{j,v_i} \geq 1.9\epsilon \log n] \to 1.$$

The same now holds for $\mathcal{A}_{j,v_i}$. If we now partition V' to sets of size $\log n$ and use the fact that $\sigma_{v_i}\mathcal{A}'_{j,v_i}$ are ± 1, we obtain the claim of the lemma. $\square$

Lemma 8.7. *We have that for a random $v \in V'$, $\mathbb{P}[J_v = 0] \to 0$.*

Proof. Recall that with high probability we have that $|Y| = \sqrt{\log\log n}$. With high probability $|B_R(v)| \leq p^{2R}$. Condition on $|B_R(v)| \leq p^{2R}$. The probability that $B_{R-1} \cap V_j = \varnothing$ and $S_v \cup Y \subset V_j$ is bounded below by $e^{-c(p^{2R}+|Y|)} \geq (\log n)^{-1/2}$. Since these are independent events given $|B_R(v)|$ it follows that with probability tending to one $J_v \neq 0$. $\square$

We now show that the indicators $\mathcal{A}_{J_v,v}$ and $\mathcal{B}_{J_v,v}$ usually agree.

Lemma 8.8.
$$\tilde{\mathbb{E}}\left[\frac{1}{|V'|}\sum_{v\in V'}\mathcal{A}_{J_v,v}\mathcal{B}_{J_v,v}I(J_v\neq 0)\right]\to 1. \tag{22}$$

Proof. We would like to apply Lemma 8.2. Note the different settings between the current lemma and Lemma 8.2, as in the later $J^{(v)}$ is not bounded by $\log n$ and is never 0. However, by Lemma 8.7, the probability of $J_v \neq 0$ in the setting of the current lemma goes to 0 and therefore Lemma 8.2 implies that
$$\mathbb{P}\left[\{J_v=0\}\cup\left\{\left|\sum_{u\in Y,u'\in S_v}N^{(J_v)}_{u,u',k}-\frac{{s'}^k}{n}M_vM_*\right|>\frac{\epsilon {s'}^{k+R}}{n}|Y|\right\}\right]\to 0 \tag{23}$$

The event $\mathcal{A}_{J_v,v}\neq\mathcal{B}_{J_v,v}$ is equivalent to $\xi_{J_v,v}$ falling outside the interval with end-points $-M_v/(Ds'^R)$ and
$$-\frac{p'n\sum_{u\in Y,u'\in S_v}N^{(J_v)}_{u,u',k}}{|Y|Ds'^{k+R+1}}.$$

Since with high probability M_* is concentrated around $\frac{s'}{p}|Y|$ we have by (23) that the latter end point converges to
$$-\frac{p'n\frac{s'^k}{n}\frac{s'}{p'}|Y|M_v}{D\frac{s'^{k+R}}{n}\frac{s'}{p'}|Y|}=-\frac{M_v}{Ds'^R}.$$
Therefore the probability that $-\xi_{J_v,v}$ falls in the interval converges to 0 as needed. □

We can now complete the proof of Theorem 2.6.

Proof of Theorem 2.6.
Combining Lemmas 8.6, 8.8 and 8.7 we have that with high probability
$$\sum_{v\in V}\tau(v)\sigma(v)\geq\epsilon n.$$
yielding an algorithm recovering the a constant correlation with the true partition. To see that the running time is $O(n\log^2 n)$ we note that for each j once can compute all the path summations $\sum_{u\in Y,u'\in S_v}N^{(j)}_{u,u',k}$ in time $O(n\log n)$ using the construction in Section 3.2. □

8.1 Acknowledgments

The authors are grateful to Cris Moore and Lenka Zdeborová for stimulating and interesting discussions on many aspects of the block model.

References

[1] P.J. Bickel and A. Chen. A nonparametric view of network models and Newman-Girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50):21068–21073, 2009.

[2] R.B. Boppana. Eigenvalues and graph bisection: An average-case analysis. In *28th Annual Symposium on Foundations of Computer Science*, pages 280–285. IEEE, 1987.

[3] T.N. Bui, S. Chaudhuri, F.T. Leighton, and M. Sipser. Graph bisection algorithms with good average case behavior. *Combinatorica*, 7(2):171–191, 1987.

[4] A. Condon and R.M. Karp. Algorithms for graph partitioning on the planted partition model. *Random Structures and Algorithms*, 18(2):116–140, 2001.

[5] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physics Review E*, 84:066106, Dec 2011.

[6] M.E. Dyer and A.M. Frieze. The solution of some random NP-hard problems in polynomial expected time. *Journal of Algorithms*, 10(4):451–489, 1989.

[7] P.W. Holland, K.B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109 – 137, 1983.

[8] M. Jerrum and G.B. Sorkin. The Metropolis algorithm for graph bisection. *Discrete Applied Mathematics*, 82(1-3):155–175, 1998.

[9] H. Kesten and B. P. Stigum. Additional limit theorems for indecomposable multidimensional Galton-Watson processes. *Ann. Math. Statist.*, 37:1463–1481, 1966.

[10] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, Zdeborova L, and P. Zhang. Spectral redemption: clustering sparse networks. arXiv:1306.5550, 2013.

[11] J. Leskovec, K.J. Lang, A. Dasgupta, and M.W. Mahoney. Statistical properties of community structure in large social and information networks. In *Proceeding of the 17th international conference on World Wide Web*, pages 695–704. ACM, 2008.

[12] L. Massoulie. Community detection thresholds and the weak ramanujan property. arXiv:1311:3085, 2013.

[13] F. McSherry. Spectral partitioning of random graphs. In *42nd IEEE Symposium on Foundations of Computer Science*, pages 529–537. IEEE, 2001.

[14] E. Mossel, J. Neeman, and A. Sly. Stochastic block models and reconstruction. arXiv:1202.4124, 2013.

[15] K. Rohe, S. Chatterjee, and B. Yu. Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, 39(4):1878–1915, 2011.

[16] B. Roos. Binomial approximation to the poisson binomial distribution: The krawtchouk expansion. *Theory of Probability and Its Applications*, 45(2):258–272, 2001.

[17] T.A.B. Snijders and K. Nowicki. Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification*, 14(1):75–100, 1997.

[18] S.H. Strogatz. Exploring complex networks. *Nature*, 410(6825):268–276, 2001.