Electron. J. Probab. **23** (2018), no. 18, 1–28.
ISSN: 1083-6489 https://doi.org/10.1214/17-EJP128

Contiguity and non-reconstruction results for planted partition models: the dense case

Debapratim Banerjee*

Abstract

We consider the two block stochastic block model on n nodes with asymptotically equal cluster sizes. The connection probabilities within and between cluster are denoted by $p_n := \frac{a_n}{n}$ and $q_n := \frac{b_n}{n}$ respectively. Mossel et al. [27] considered the case when $a_n = a$ and $b_n = b$ are fixed. They proved the probability models of the stochastic block model and that of Erdős–Rényi graph with same average degree are mutually contiguous whenever $(a-b)^2 < 2(a+b)$ and are asymptotically singular whenever $(a-b)^2 > 2(a+b)$. Mossel et al. [27] also proved that when $(a-b)^2 < 2(a+b)$ no algorithm is able to find an estimate of the labeling of the nodes which is positively correlated with the true labeling. It is natural to ask what happens when a_n and b_n both grow to infinity. In this paper we consider the case when $a_n \to \infty$, $\frac{a_n}{n} \to p \in [0,1)$ and $(a_n - b_n)^2 = \Theta(a_n + b_n)$. Observe that in this case $\frac{b_n}{n} \to p$ also. We show that here the models are mutually contiguous if asymptotically $(a_n - b_n)^2 < 2(1-p)(a_n + b_n)$ and they are asymptotically singular if asymptotically $(a_n - b_n)^2 > 2(1-p)(a_n + b_n)$. Further we also prove it is impossible find an estimate of the labeling of the nodes which is positively correlated with the true labeling whenever $(a_n - b_n)^2 < 2(1-p)(a_n + b_n)$ asymptotically. The results of this paper justify the negative part of a conjecture made in Decelle et al. (2011) [17] for dense graphs.

Keywords: stochastic block model; planted partition model; threshold; phase transition; community detection; random network; linear statistics.
AMS MSC 2010: 05C80.
Submitted to EJP on November 22, 2016, final version accepted on November 27, 2017.

1 Introduction

In the last few years the stochastic block model has been one of the most active domains of modern research in statistics, computer science and many other related fields. In general a stochastic block model is a network with a hidden community structure where the nodes within the communities are expected to be connected in a different manner than the nodes between the communities. This model arises naturally in many problems of statistics, machine learning and data mining, but its applications further

*Dept. of Statistics, University of Pennsylvania. E-mail: dban@wharton.upenn.edu

extends to population genetics [30], where genetically similar sub-populations are used as the clusters, to image processing [32], [33], where the group of similar images acts as cluster, to the study of social networks, where groups of like-minded people act as clusters [29].

Recently a huge amount of effort has been dedicated to find out the clusters. Numerous different clustering algorithms have been proposed in literature. One might look at [22], [18], [13], [19], [10], [9], [16], [31], [25] for some references. One might also look at the review paper by Abbe [1] for a detailed study of the literature.

One of the easiest examples of the stochastic block model is the planted partition model where one have only two clusters of more or less equal size. Formally,

Definition 1.1. 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 in which each vertex u is assigned (independently and uniformly at random) a label $\sigma_u \in \{\pm 1\}$ and each edge between u and v are included independently with probability p if they have the same label and with probability q if they have different labels.

The case when p and q are sufficiently close to each other has got significant amount of interest in literature. Decelle et al. [17] made a fascinating conjecture in this regard.

Conjecture 1.1. Let $p = \frac{a}{n}$ and $q = \frac{b}{n}$ where a and b are fixed real numbers. Then the following are true.
i) If $(a-b)^2 > 2(a+b)$ then one can find almost surely a bisection of the vertices which is positively correlated with the original clusters.
ii) If $(a-b)^2 < 2(a+b)$ then the problem is not solveable.
iii) Further, there are no consistent estimators of a and b if $(a-b)^2 < 2(a+b)$ and there are consistent estimators of a and b whenever $(a-b)^2 > 2(a+b)$.

Coja-Oghlan [15] solved part $i)$ of the problem when $(a-b)^2 > C(a+b)$ for some large C and finally part $ii)$ and $iii)$ of Conjecture 1.1 was proved by Mossel et al. [27] and part $i)$ was solved by Mossel et al. [26] and Massoulié [24] independently.

Typically the problem is much more delicate when more than two communities are present in the sparse case. To keep things simple let us consider the general stochastic block model with k asymptotically equal sized blocks with connection probabilities within and between blocks are given by $\frac{a}{n}$ and $\frac{b}{n}$ respectively. It was conjectured in Mossel et al [27] that for k sufficiently large, there is a constant $c(k)$ such that whenever

$$c(k) < \frac{(a-b)^2}{a+(k-1)b} < k$$

the reconstruction problem is solvable in exponential time, it is not solvable if $\frac{(a-b)^2}{a+(k-1)b} < c(k)$ and solvable in polynomial time if $k < \frac{(a-b)^2}{a+(k-1)b}$. The upper bound is known as Kesten-Stigum threshold. Bordenave et al. [11] solved the reconstruction problem above a deterministic threshold by spectral analysis of non-backtraking matrix. One might look at Banks et al. [8] for the non solvability part. They proved that the probability models of stochastic block model and that of Erdős–Rényi graph with same average degree are mutually contiguous and the reconstruction problem is unsolvable if

$$d < \frac{2\log(k-1)}{k-1}\frac{1}{\lambda^2}.$$

Here $d = \frac{a+(k-1)b}{k}$ and $\lambda = \frac{a-b}{kd}$. Abbe et al. [2] provides an efficient algorithm for reconstruction above the Kesten-Stigum threshold. Abbe et al. [2] and Banks et al. [8]

also provide cases strictly below the Kesten-Stigum threshold where the problem is solvable in exponential time.

On the other hand, a different type of reconstruction problem was considered in Mossel et al. [28] for denser graphs. They considered two different notions of recovery. The first one is weak consistency where one is interested in finding a bisection $\hat{\sigma}$ such that σ and $\hat{\sigma}$ have correlation going to 1 with high probability. The second one is called strong consistency. Here one is interested in finding a bisection $\hat{\sigma}$ such that $\hat{\sigma}$ is either σ or $-\sigma$ with probability tending to 1. Mossel et al. [28] prove that weak consistency is possible if and only if $\frac{n(p_n-q_n)^2}{p_n+q_n} \to \infty$ and strong consistency is possible if and only if

$$\left(a_n + b_n - 2\sqrt{a_n b_n} - 1\right) \log n + \frac{1}{2} \log \log n \to \infty.$$

Here $a_n = \frac{np_n}{\log n}$ and $b_n = \frac{nq_n}{\log n}$ respectively. Abbe et al. [3] studied the same problem independently in the logarithmic sparsity regime. They prove that for $a = \frac{np_n}{\log n}$ and $b = \frac{nq_n}{\log n}$ fixed, $(a+b)-2\sqrt{ab} > 1$ is sufficient for strong consistency and that $(a+b)-2\sqrt{ab} \geq 1$ is necessary. We note that their results are implied by Mossel et al.[28].

However, according to the best of our knowledge questions similar to part $ii)$ and $iii)$ of Conjecture 1.1 have not yet been addressed in dense case (i.e. when a and b increase to infinity). This is the main focus of this paper.

Before stating our results we mention that the results in Mossel et al. [27] is more general than part $iii)$ of Conjecture 1.1. Let $\mathbb{P}_n$ and $\mathbb{P}'_n$ be the sequences of probability measures induced by $\mathcal{G}(n,p,q)$ and $\mathcal{G}(n, \frac{p+q}{2}, \frac{p+q}{2})$ respectively. Then [27] prove that whenever a and b are fixed numbers and $(a-b)^2 < 2(a+b)$, the measures $\mathbb{P}_n$ and $\mathbb{P}'_n$ are mutually contiguous i.e. for a sequence of events A_n, $\mathbb{P}_n(A_n) \to 0$ if and only if $\mathbb{P}'_n(A_n) \to 0$. Now part $iii)$ of Conjecture 1.1 directly follows from the contiguity. The proof in Mossel et al. [27] is based on calculating the limiting distribution of the short cycles and using a result on contiguity (Theorem 1 in Janson [21] and Theorem 4.1 in Wormald [35]). However, one should note that the result from [27] does not directly generalize to the denser case. Since, one requires the limiting distributions of short cycles to be independent Poisson in order to use Janson's result. In our proof instead of considering the short cycles we consider the "signed cycles" (to be defined later) which have asymptotic normal distributions. We also find a result analogous to Janson for the normal random variables in order to complete the proof.

On the other hand, the original proof of non-reconstruction from Mossel et al. [27] relies on the coupling of $\mathbb{P}_n$ and $\mathbb{P}'_n$ with probability measure induced by Galton Watson trees of suitable parameters. However, it is well known that when the graph is sufficiently dense i.e. $a_n \gg n^{o(1)}$ the coupling argument does not work. So our proof is based on fine analysis of conditional probabilities. Technically, this proof is closely related to the non-reconstruction proof in section 6.2 of Banks et al. [8] rather than the original proof given in Mossel et al. [27].

A natural question arises how far the arguments in this paper generalize to the multi-community case. Unfortunately, we do not have a definite answer for this problem. The fundamental difficulty here is the absence of locally tree like structure which is the essence of all the proofs in the sparse regime. However, we believe the similar thresholds are true even in dense case also. In fact, it was shown in Banerjee and Ma(2017) [7] that for the multi-community case the models are mutually singular much below the Kesten-Stigum threshold. We leave the problem for future research.

The paper is organized in the following manner. In Section 2 we build some preliminary notations and state our results. Section 3 is dedicated for building a result analogous to Theorem 1 in Janson [21]. In Section 4 we define signed cycles and find their asymptotic distributions. Section 5 is dedicated to complete the proofs of our

contiguity results. In Section 6 we prove the non-reconstruction result. Finally, the paper concludes with an Appendix containing a proof of a result from random matrix theory used in this paper.

2 Our results

Through out the paper a random graph will be denoted by G and $x_{i,j}$ will be used to denote the indicator random variable corresponding to an edge between the nodes i and j. Further $\mathbb{P}_n$ and $\mathbb{P}'_n$ will be used to denote the sequence of probability measures induced by $\mathcal{G}(n, p_n, q_n)$ and $\mathcal{G}(n, \frac{p_n+q_n}{2}, \frac{p_n+q_n}{2})$ respectively. For notational simplicity we denote $\frac{p_n+q_n}{2}$ by $\hat{p}_n$.

In this paper we shall consider the case when $(a_n - b_n)^2 = \Theta(a_n + b_n)$. We shall use the following notations through out the paper. We denote $c_n := \frac{(a_n-b_n)^2}{(a_n+b_n)}$, $d_n := \frac{p_n-q_n}{2}$ and $t_n = \frac{c_n}{2(1-\hat{p}_n)}$.

Further, for any two labeling of the nodes σ and τ, we define their overlap to be

$$\mathrm{ov}(\sigma, \tau) := \frac{1}{n}\left(\sum_{i=1}^{n} \sigma_i\tau_i - \frac{1}{n}\left(\sum_{i=1}^{n}\sigma_i\right)\left(\sum_{i=1}^{n}\tau_i\right)\right). \tag{2.1}$$

Now we define mutual contiguity of two sequences of measures as follows:

Definition 2.1. Let $\mathbb{P}_n$ and $\mathbb{Q}_n$ be two sequences of probability measures, such that for each n, $\mathbb{P}_n$ and $\mathbb{Q}_n$ both are defined on the same measurable space $(\Omega_n, \mathcal{F}_n)$. We then say that the sequences are mutually contiguous if for every sequence of measurable sets $A_n \subset \Omega_n$,

$$\mathbb{P}_n(A_n) \to 0 \Leftrightarrow \mathbb{Q}_n(A_n) \to 0.$$

Two sequences of probability measures $\mathbb{P}_n$ and $\mathbb{Q}_n$ are called asymptotically mutually singular if there exists a sequence of measurable sets A_n such that $\mathbb{P}_n(A_n) \to 1$ and $\mathbb{Q}_n(A_n^c) \to 1$ as $n \to \infty$.

We are now ready to state the main results of the paper.

Theorem 2.1. *i)If $a_n, b_n \to \infty$, $\frac{a_n}{n} \to p \in [0, 1)$ and $c_n \to c < 2(1-p)$, then the probability measures $\mathbb{P}_n$ and $\mathbb{P}'_n$ are mutually contiguous. So there does not exist an estimator (A_n, B_n) for (a_n, b_n) such that $|A_n - a_n| + |B_n - b_n| = o_p(a_n - b_n)$.*
ii)If $a_n, b_n \to \infty$, $\frac{a_n}{n} \to p \in [0, 1)$ and $c_n \to c > 2(1-p)$, then the probability measures $\mathbb{P}_n$ and $\mathbb{P}'_n$ are asymptotically mutually singular. Further there exists an estimator (A_n, B_n) for (a_n, b_n) such that $|A_n - a_n| + |B_n - b_n| = o_p(a_n - b_n)$.

Theorem 2.2. *If $a_n, b_n \to \infty$, $\frac{a_n}{n} \to p \in [0, 1)$ and $c_n \to c < 2(1-p)$, then there is no reconstruction algorithm which performs better than the random guessing i.e. for any estimate of the labeling $\{\hat{\sigma}_i\}_{i=1}^n$ we have*

$$\mathrm{ov}(\sigma, \hat{\sigma}) \xrightarrow{P} 0. \tag{2.2}$$

3 A result on contiguity

In this section we provide a very brief description of contiguity of probability measures. We suggest the reader to have a look at the discussion about contiguity of measures in Janson [21] for further details. In this section we state several propositions and apart from Proposition 3.4 and Proposition 3.3, all the proofs can be found in Janson [21].

Definition 2.1 of contiguity might appear a little abstract. However the following reformulation is perhaps more useful to understand the contiguity concept.

Proposition 3.1. *Two sequences of probability measures $\mathbb{P}_n$ and $\mathbb{Q}_n$ are mutually contiguous if and only if for every $\varepsilon > 0$ there exist $n(\varepsilon)$ and $K(\varepsilon)$ such that for all $n > n(\varepsilon)$ there exists a set $B_n \in \mathcal{F}_n$ with $\mathbb{P}_n(B_n^c), \mathbb{Q}_n(B_n^c) \leq \varepsilon$ such that*

$$K(\varepsilon)^{-1} \leq \frac{\mathbb{Q}_n(A_n)}{\mathbb{P}_n(A_n)} \leq K(\varepsilon). \ \forall A_n \subset B_n.$$

Although Proposition 3.1 gives an equivalent condition, verifying this condition is often difficult. However under the assumption of convergence of $\frac{d\mathbb{Q}_n}{d\mathbb{P}_n}$, one gets the following simplified result.

Proposition 3.2. *Suppose that $L_n = \frac{d\mathbb{Q}_n}{d\mathbb{P}_n}$, regarded as a random variable on $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$, converges in distribution to some random variable L as $n \to \infty$. Then $\mathbb{P}_n$ and $\mathbb{Q}_n$ are mutually contiguous if and only if $L > 0$ a.s. and $\mathrm{E}[L] = 1$.*

We now introduce the concept of Wasserstein's metric which will be used in the proof of Proposition 3.4.

Definition 3.1. Let F and G be two distribution functions with finite p th moment. Then the Wasserstein distance W_p between F and G is defined to be

$$W_p(F, G) = \left[\inf_{X \sim F, Y \sim G} \mathrm{E}\,|X - Y|^p \right]^{\frac{1}{p}}.$$

Here X and Y are random variables having distribution functions F and G respectively.

In particular, the following result will be useful in our proof:

Proposition 3.3. *Let F_n be a sequence of distribution functions and F be a distribution function. Then F_n converge to F in distribution and $\int x^2 dF_n(x) \to \int x^2 dF(x)$ if $W_2(F_n, F) \to 0$.*

The proof of Proposition 3.3 is well known. One might look at Mallows(1972)[23] for a reference.

With Proposition 3.2 in hand, we now state the most important result in this section. This result will be used to prove Theorem 2.1. Although, Proposition 3.4 is written in a complete different notation, one can check that it is analogous to Theorem 1 in Janson [21].

Proposition 3.4. *Let $\mathbb{P}_n$ and $\mathbb{Q}_n$ be two sequences of probability measures such that for each n, both of them are defined on $(\Omega_n, \mathcal{F}_n)$. Suppose that for each $i \geq 3$, $X_{n,i}$ are random variables defined on $(\Omega_n, \mathcal{F}_n)$. Then the probability measures $\mathbb{P}_n$ and $\mathbb{Q}_n$ are mutually contiguous if the following conditions hold:*

i) $\mathbb{P}_n \ll \mathbb{Q}_n$ *and* $\mathbb{Q}_n \ll \mathbb{P}_n$ *for each* n.

ii) For any fixed $k \geq 3$, one has $(X_{n,3}, \ldots, X_{n,k})\,|\mathbb{P}_n \xrightarrow{d} (Z_3, \ldots, Z_k)$ and $(X_{n,3}, \ldots, X_{n,k})|\mathbb{Q}_n \xrightarrow{d} (Z_3', \ldots, Z_k')$.

iii) $Z_i \sim N(0, 2i)$ and $Z_i' \sim N(t^{\frac{i}{2}}, 2i)$ are sequences of independent random variables. Here $|t| < 1$.

iv)

$$\mathrm{E}_{\mathbb{P}_n}\left[\left(\frac{d\mathbb{Q}_n}{d\mathbb{P}_n}\right)^2\right] \to \exp\left\{-\frac{t}{2} - \frac{t^2}{4}\right\} \frac{1}{\sqrt{1-t}}. \tag{3.1}$$

Further,

$$\frac{d\mathbb{Q}_n}{d\mathbb{P}_n}|\mathbb{P}_n \xrightarrow{d} \exp\left\{\sum_{i=3}^{\infty} \frac{2t^{\frac{i}{2}} Z_i - t^i}{4i}\right\}. \tag{3.2}$$

Proof. In this proof for simplicity we denote $\frac{d\mathbb{Q}_n}{d\mathbb{P}_n}$ by Y_n. We break the proof into two steps.

Step 1. In this step we prove the random variable in the right hand side of (3.2) is almost surely positive and has mean 1. Let us define

$$W := \exp\left\{\sum_{i=3}^{\infty} \frac{2t^{\frac{i}{2}} Z_i - t^i}{4i}\right\}$$

and

$$W^{(m)} := \exp\left\{\sum_{i=3}^{m} \frac{2t^{\frac{i}{2}} Z_i - t^i}{4i}\right\}.$$

As $Z_i \sim N(0, 2i)$,

$$\mathrm{E}\left[\exp\left\{\frac{2t^{\frac{i}{2}} Z_i - t^i}{4i}\right\}\right] = \exp\left\{\frac{4t^i \times 2i}{2 \times 16i^2} - \frac{t^i}{4i}\right\} = 1.$$

So $\{W^{(m)}\}_{m=3}^{\infty}$ is a martingale sequence and

$$\mathrm{E}\left[W^{(m)2}\right] = \prod_{i=3}^{m} \exp\left\{\frac{t^i}{2i}\right\} = \exp\left\{\sum_{i=3}^{m} \frac{t^i}{2i}\right\}.$$

Now

$$\sum_{i=3}^{\infty} \frac{t^i}{2i} = \frac{1}{2}\left(-\log(1-t) - t - \frac{t^2}{2}\right) \quad \forall \ |t| < 1.$$

So $W^{(m)}$ is a L^2 bounded martingale. Hence, W is a well defined random variable,

$$\mathrm{E}[W^2] = \exp\left\{-\frac{t}{2} - \frac{t^2}{4}\right\} \frac{1}{\sqrt{1-t}}$$

and $\mathrm{E}[W] = 1$.

Now observe that $Z_i \stackrel{d}{=} -Z_i$ for each i and whenever $|t| < 1$, the series $\sum_{i=3}^{\infty} \frac{t^i}{4i}$ converges. So

$$W^{-1} \stackrel{d}{=} \exp\left\{\sum_{i=3}^{\infty} \frac{2t^{\frac{i}{2}} Z_i + t^i}{4i}\right\}.$$

However, $E[W^{-1}] = \exp\left\{\sum_{i=3}^{\infty} \frac{t^i}{2i}\right\} < \infty$ implies $W > 0$ a.s.

Step 2. Now we come to the harder task of proving $Y_n \stackrel{d}{\to} W$. Since

$$\limsup_{n\to\infty} \mathrm{E}_{\mathbb{P}_n}\left[(Y_n)^2\right] < \infty$$

from condition $iv)$, the sequence Y_n is tight. Hence from Prokhorov's theorem there is a sub sequence $\{n_k\}_{k=1}^{\infty}$ such that Y_{n_k} converge in distribution to some random variable $W(\{n_k\})$. We shall prove that the distribution of $W(\{n_k\})$ does not depend on the sub sequence $\{n_k\}$. In particular, $W(\{n_k\}) \stackrel{d}{=} W$.

Since Y_{n_k} converges in distribution to $W(\{n_k\})$, for any further sub sequence $\{n_{k_l}\}$ of $\{n_k\}$, $Y_{n_{k_l}}$ also converges in distribution to $W(\{n_k\})$.

Given $\varepsilon > 0$ take m big enough such that

$$\exp\left\{\sum_{i=3}^{\infty} \frac{t^i}{2i}\right\} - \exp\left\{\sum_{i=3}^{m} \frac{t^i}{2i}\right\} < \varepsilon.$$

For this m, look at the joint distribution of $(Y_{n_k}, X_{n_k,3}, \ldots, X_{n_k,m})$. This sequence of $m-1$ dimensional random vectors with respect to $\mathbb{P}_{n_k}$ is also tight from condition $ii)$. So it has a further sub sequence such that

$$(Y_{n_{k_l}}, X_{n_{k_l},3}, \ldots, X_{n_{k_l},m})|\mathbb{P}_{n_{k_l}} \xrightarrow{d} (W(\{n_k\}), Z_3 \ldots, Z_m)\,.$$

Here we have used condition $ii)$ for the convergence of $(X_{n_{k_l},3}, \ldots, X_{n_{k_l},m})|\mathbb{P}_{n_{k_l}}$.

The most important part of this proof is to show, we can define the random variables $W^{(m)}$ and $W(\{n_k\})$ in such a way that there exist suitable σ algebras $\mathcal{F}_1 \subset \mathcal{F}_2$ such that $W^{(m)} \in \mathcal{F}_1$ and $W(\{n_k\}) \in \mathcal{F}_2$ and $\mathrm{E}\left[W(\{n_k\}) \mid \mathcal{F}_1\right] = W^{(m)}$.

From condition $iv)$ we have $\limsup_{n\to\infty} \mathrm{E}_{\mathbb{P}_n}\left[Y_n^2\right] < \infty$. As a consequence, the sequence the sequence $Y_{n_{k_l}}$ is uniformly integrable. This together with condition $i)$ will give us

$$1 = \mathrm{E}_{\mathbb{P}_{n_{k_l}}}\left[Y_{n_{k_l}}\right] \to \mathrm{E}[W(\{n_k\})] = 1. \tag{3.3}$$

Now take any positive bounded continuous function $f : \mathbb{R}^m \to \mathbb{R}$. By Fatou's lemma

$$\liminf \mathrm{E}_{\mathbb{P}_{n_{k_l}}}\left[f\left(X_{n_{k_l},3}, \ldots, X_{n_{k_l},m}\right) Y_{n_{k_l}}\right] \geq \mathrm{E}\left[f\left(Z_3, \ldots, Z_m\right) W(\{n_k\})\right]. \tag{3.4}$$

However for any constant ξ we have

$$\xi = \xi\, \mathrm{E}_{\mathbb{P}_{n_{k_l}}}\left[Y_{n_{k_l}}\right] \to \xi\, \mathrm{E}[W(\{n_k\})] = \xi$$

from (3.3).

So (3.4) holds for any bounded continuous function f. On the other hand replacing f by $-f$ we have

$$\lim \mathrm{E}_{\mathbb{P}_{n_{k_l}}}\left[f\left(X_{n_{k_l},3}, \ldots, X_{n_{k_l},m}\right) Y_{n_{k_l}}\right] = \mathrm{E}\left[f\left(Z_3, \ldots, Z_m\right) W(\{n_k\})\right]. \tag{3.5}$$

Now applying condition $ii)$ we have

$$\int f\left(X_{n_{k_l},3}, \ldots, X_{n_{k_l},m}\right) Y_{n_{k_l}} d\mathbb{P}_{n_{k_l}} = \int f\left(X_{n_{k_l},3}, \ldots, X_{n_{k_l},m}\right) d\mathbb{Q}_{n_{k_l}} \to \int f(Z'_3, \ldots, Z'_m) dQ. \tag{3.6}$$

Here Q is the measure induced by $(Z'_3, \ldots, Z'_m)$. In particular, one can take the measure Q such that $(Z_3, \ldots, Z_m)$ themselves are distributed as $(Z'_3, \ldots, Z'_m)$ under the measure Q. This is true due to the following observation.

$$\int f(Z'_3, \ldots, Z'_m) dQ = \mathrm{E}\left[f(Z_3, \ldots, Z_m) W^{(m)}\right]$$

for any bounded continuous function f. Since f is any bounded continuous function, we have

$$\int_A dQ = \mathrm{E}\left[\mathbb{I}_A W^{(m)}\right]$$

for any $A \in \sigma(Z_3, \ldots, Z_m)$. Here for any set A, $\mathbb{I}_A$ denotes the indicator function taking value one on A.

Now looking back into (3.5), we have for any $A \in \sigma(Z_3, \ldots, Z_m)$,

$$\mathrm{E}\left[\mathbb{I}_A W^{(m)}\right] = \mathrm{E}\left[\mathbb{I}_A W(\{n_k\})\right].$$

Since $W^{(m)}$ is $\sigma(Z_3, \ldots, Z_m)$ measurable, we have $W^{(m)} = \mathrm{E}\left[W(\{n_k\}) \mid \sigma(Z_3, \ldots, Z_m)\right]$

From Fatou's lemma

$$\mathrm{E}[W(\{n_k\})^2] \leq \liminf_{n\to\infty} \mathrm{E}_{\mathbb{P}_n}[Y_n^2] = \exp\left\{\sum_{i=3}^{\infty} \frac{t^i}{2i}\right\}.$$

As a consequence, we have

$$0 \leq \mathrm{E}\,|W(\{n_k\}) - W^{(m)}|^2 = \mathrm{E}[W(\{n_k\})^2] - \mathrm{E}[W^{(m)2}] < \varepsilon.$$

So $W_2(F^{W^{(m)}}, F^{W(\{n_k\})}) < \sqrt{\varepsilon}$. Here $F^{W^{(m)}}$ and $F^{W(\{n_k\})}$ denote the distribution functions corresponding to $W^{(m)}$ and $W(\{n_k\})$ respectively. As a consequence, $W_2(F^{W^{(m)}}, F^{W(\{n_k\})}) \to 0$ as $m \to \infty$. Hence by Proposition 3.3, $W^{(m)} \xrightarrow{d} W(\{n_k\})$.

On the other hand, we have already proved $W^{(m)}$ converge to W in L^2. So $W(\{n_k\}) \stackrel{d}{=} W$.

In Step 1 and Step 2 we verified all the conditions required to use Proposition 3.2. Now using Proposition 3.2 the proof of Proposition 3.4 is complete. □

Remark 3.1. One might observe that the second part in assumption $ii)$ of Proposition 3.4 is slightly weaker than (A2) in Theorem 1 of Janson [21]. For our purpose this is sufficient since we use the fact that $Y_n = \frac{d\mathbb{Q}_n}{d\mathbb{P}_n}$. However, in Theorem 1 of Janson [21] Y_n can be any random variable.

4 Signed cycles and their asymptotic distributions

We have discussed in the introduction that the proof of Mossel et al. [27] crucially used the fact that the asymptotic distribution of short cycles turn out to be Poisson. However, in the denser case one does not get a Poisson limit for the short cycles. So their proof does not work in the denser case. Here we consider instead the "signed cycles" defined as follows:

Definition 4.1. For a random graph G the signed cycle of length k is defined to be:

$$C_{n,k}(G) = \left(\frac{1}{\sqrt{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}}\right)^k \sum_{i_0,i_1,\ldots,i_{k-1}} (x_{i_0,i_1} - p_{n,\mathrm{av}}) \ldots (x_{i_{k-1}i_0} - p_{n,\mathrm{av}})$$

where $i_0, i_1, \ldots, i_{k-1}$ are all distinct and $p_{n,\mathrm{av}}$ is the average connection probability i.e. $p_{n,\mathrm{av}} = \frac{1}{n(n-1)} \sum_{i \neq j} \mathrm{E}[x_{i,j}]$. Observe that for $\mathcal{G}(n, p_n, q_n)$, $p_{n,\mathrm{av}}$ is equal to $\hat{p}_n$.

One should note that when $k = 3$ a similar kind of random variable was called "signed triangle" in Bubeck et al. [12]

It is intuitive that one might expect asymptotic normal distribution for $C_{n,k}$'s when $n \to \infty$ and $\hat{p}_n$ is sufficiently large. Our next result formalizes this intuition.

Proposition 4.1. *i)When $G \sim \mathbb{P}'_n$, $n(p_n + q_n) \to \infty$ and $3 \leq k_1 < \ldots < k_l = o(\log(\hat{p}_n n))$,*

$$\left(\frac{C_{n,k_1}(G)}{\sqrt{2k_1}}, \ldots, \frac{C_{n,k_l}(G)}{\sqrt{2k_l}}\right) \xrightarrow{d} N_l(0, I_l). \tag{4.1}$$

ii) When $G \sim \mathbb{P}_n$, $np_n \to \infty$, $c_n \to c \in (0, \infty)$ and $3 \leq k_1 < \ldots < k_l = o\big(\min(\log(\hat{p}_n n), \sqrt{\log(n)})\big)$,

$$\left(\frac{C_{n,k_1}(G) - \mu_1}{\sqrt{2k_1}}, \ldots, \frac{C_{n,k_l}(G) - \mu_l}{\sqrt{2k_l}}\right) \xrightarrow{d} N_l(0, I_l) \tag{4.2}$$

where $\mu_i = \left(\sqrt{\frac{c_n}{2(1-\hat{p}_n)}}\right)^{k_i}$ for $1 \leq i \leq m$.

The proof of Proposition 4.1 is inspired from the remarkable paper by Anderson and Zeitouni [4]. However, the model in this case is simpler which makes the proof less cumbersome. The fundamental idea is to prove that the signed cycles converge in distribution by using the method of moments and the limiting random variables satisfy the Wick's formula. At first we state the method of moments.

Lemma 4.1. *Let $(Y_{n,1},\dots,Y_{n,l})$ be a sequence of random vectors of l dimension. Then $(Y_{n,1},\dots,Y_{n,l}) \xrightarrow{d} (Z_1,\dots,Z_l)$ if the following conditions are satisfied:*

i)

$$\lim_{n\to\infty} \mathrm{E}[X_{n,1}\dots X_{n,m}] \tag{4.3}$$

exists for any fixed m and $X_{n,i} \in \{Y_{n,1},\dots,Y_{n,l}\}$ for $1 \le i \le m$.

ii) (Carleman's Condition)[14]

$$\sum_{h=1}^{\infty} \Big(\lim_{n\to\infty} \mathrm{E}[X_{n,i}^{2h}] \Big)^{-\frac{1}{2h}} = \infty \ \ \forall\, 1 \le i \le l.$$

Further,

$$\lim_{n\to\infty} \mathrm{E}[X_{n,1}\dots X_{n,m}] = \mathrm{E}[X_1\dots X_m].$$

Here $X_{n,i} \in \{Y_{n,1},\dots,Y_{n,l}\}$ for $1 \le i \le m$ and X_i is the in distribution limit of $X_{n,i}$. In particular, if $X_{n,i} = Y_{n,j}$ for some $j \in \{1,\dots,l\}$ then $X_i = Z_j$.

The method of moments is very well known and much useful in probability theory. We omit its proof.

Now we state the Wick's formula for Gaussian random variables which was first proved by Isserlis(1918)[20] and later on introduced by Wick[34] in the physics literature in 1950.

Lemma 4.2. *(Wick's formula)[34] Let $(Y_1,\dots,Y_l)$ be a multivariate mean 0 random vector of dimension l with covariance matrix Σ(possibly singular). Then $((Y_1,\dots,Y_l))$ is jointly Gaussian if and only if for any integer m and $X_i \in \{Y_1,\dots,Y_l\}$ for $1 \le i \le m$*

$$\mathrm{E}[X_1\dots X_m] = \begin{cases} \sum_\eta \prod_{i=1}^{\frac{m}{2}} \mathrm{E}[X_{\eta(i,1)}X_{\eta(i,2)}] & \text{for } m \text{ even} \\ 0 & \text{for } m \text{ odd.} \end{cases} \tag{4.4}$$

Here η is a partition of $\{1,\dots,m\}$ into $\frac{m}{2}$ blocks such that each block contains exactly 2 elements and $\eta(i,j)$ denotes the j th element of the i th block of η for $j = 1,2$.

The proof of the aforesaid Lemma is omitted. However, it is good to note that the random variables $Y_1,\dots,Y_l$ may also be the same. In particular, taking $Y_1 = \cdots = Y_l$, Lemma 4.2 also provides a description of the moments of Gaussian random variables. With Lemma 4.1 and 4.2 in hand, we now jump into the proof of Proposition 4.1.

Proof of Proposition 4.1:

At first we introduce some notations and some terminologies. We denote a word w to be an ordered sequence of integers (to be called letters) $(i_0,\dots,i_{k-1},i_k)$ such that $i_0 = i_k$ and all the numbers i_j for $0 \le j \le k-1$ are distinct. For a word $w = (i_0,\dots,i_{k-1},i_k)$, its length $l(w)$ is $k+1$. The graph induced by a word w is denoted by G_w and defined as follows. One treats the letters $(i_0,\dots,i_k)$ as nodes and puts an edge between the nodes $(i_j,i_{j+1})_{0\le j\le k-1}$. Note that for a word w of length $k+1$, $G_w = (V_w,E_w)$ is just a k cycle. For a word $w = (i_0,\dots,i_k)$ its mirror image is defined by $\tilde{w} = (i_0,i_{k-1},i_{k-2},\dots,i_1,i_0)$. Further for a cyclic permutation τ of the set $\{0,1,\dots,k-1\}$, we define $w^\tau := (i_{\tau(0)},\dots,i_{\tau(k-1)},i_{\tau(0)})$. Finally two words w and x are called paired if there is a cyclic permutation τ such that either $x^\tau = w$ or $\tilde{x}^\tau = w$. An ordered tuple of m words, $(w_1,\dots,w_m)$ will be called a sentence. For any sentence $a = (w_1,\dots w_m)$, $G_a = (V_a,E_a)$ is the graph with $V_a = \cup_{i=1}^m V_{w_i}$ and $E_a = \cup_{i=1}^m E_{w_i}$.

Proof of part i) We complete the proof of this part in two steps. In the first step the asymptotic variances of $(C_{n,k_1}(G),\dots,C_{n,k_l}(G))$ will be calculated and the second

step will be dedicated towards proving the asymptotic normality and independence of $(C_{n,k_1}(G),\dots,C_{n,k_l}(G))$.

Step 1: Observe that when $G\sim\mathbb{P}'_n$, the distribution of $C_{n,k_1}(G),\dots,C_{n,k_l}(G)$ is trivially independent of the labels σ_i and $\mathrm{E}[C_{n,k}(G)]=0$ for any k. Now we prove that $\mathrm{Var}(C_{n,k}(G))\sim 2k$ for any $k=o(\sqrt{n})$. Let for any word $w=(i_0,\dots,i_k)$, $X_w:=\prod_{j=0}^{k-1}\left(x_{i_j,i_{j+1}}-\hat{p}_n\right)$. Now observe that

$$\begin{aligned}\mathrm{Var}(C_{n,k})&=\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^k\mathrm{E}\left[(\sum_w X_w)^2\right]\\&=\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^k\mathrm{E}\left[\sum_{w,x}X_wX_x\right].\end{aligned}\tag{4.5}$$

Since both X_w and X_x are product of independent mean 0 random variables each coming exactly once, $\mathrm{E}[X_wX_x]\neq 0$ if and only if all the edges in G_w are repeated in G_x. Observe that since G_w and G_x are cycles of length k, this is satisfied if and only if w and x are paired. There are k many cyclic permutations τ of the set $\{0,\dots,k-1\}$ and for a given w and τ, there are only two possible choices of x such that w and x are paired. These choices are obtained when $x^\tau=w$ and $\tilde{x}^\tau=w$. As a consequence for any word w, exactly $2k$ words are paired with it. Now observe that when w and x are paired, X_wX_x is a product of k random variables each appearing exactly twice. As a consequence, $\mathrm{E}[X_wX_x]=(\hat{p}_n(1-\hat{p}_n))^k$. Also the total number of words is given by $n(n-1)\dots(n-k+1)$ for the choices of $i_0,\dots,i_{k-1}$. It is well known that

$$\frac{n(n-1)\dots(n-k+1)}{n^k}\to 1$$

whenever $k=o(\sqrt{n})$. So

$$\mathrm{Var}(C_{n,k})=2k\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^k n(n-1)\dots(n-k+1)\left(\hat{p}_n(1-\hat{p}_n)\right)^k\sim 2k\tag{4.6}$$

as long as $k=o(\sqrt{n})$. This completes **Step 1** of the proof.

Step 2: Now we claim that in order to complete **Step 2**, is enough to prove the following two limits.

$$\lim_{n\to\infty}\mathrm{E}\left[C_{n,k_1}(G)C_{n,k_2}(G)\right]\to 0\tag{4.7}$$

whenever $k_1\neq k_2$ and there exists random variables $Z_1,\dots,Z_l$ such that for any fixed m

$$\lim_{n\to\infty}\mathrm{E}[X_{n,1}\dots X_{n,m}]\to\begin{cases}\sum_\eta\prod_{i=1}^{\frac{m}{2}}\mathrm{E}[Z_{\eta(i,1)}Z_{\eta(i,2)}] & \text{for } m \text{ even}\\ 0 & \text{for } m \text{ odd.}\end{cases}\tag{4.8}$$

where $X_{n,i}\in\{\frac{C_{n,k_1}(G)}{\sqrt{2k_1}},\dots,\frac{C_{n,k_l}(G)}{\sqrt{2k_l}}\}$.

First observe that (4.8) will simultaneously imply part $i)$ and $ii)$ of Lemma 4.1. Implication of $i)$ is obvious. However, for $ii)$ one can take $X_{n,i}$'s to be all equal and from Wick's formula (Lemma 4.2) the limiting distribution of $X_{n,i}$'s are normal. It is well known that normal random variables satisfy Carleman's condition. On the other hand (4.8) also implies that the limit of $(\frac{C_{n,k_1}(G)}{\sqrt{2k_1}},\dots,\frac{C_{n,k_l}(G)}{\sqrt{2k_l}})$ is jointly normal. Hence applying (4.7), one gets the asymptotic independence.

We first prove (4.7). Observe that

$$\mathrm{E}\left[C_{n,k_1}(G)C_{n,k_2}(G)\right]=\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{k_1+k_2}{2}}\mathrm{E}\left[\sum_{w,x}X_wX_x\right].$$

However, here $l(w) = k_1+1$ and $l(x) = k_2+1$. So $\mathrm{E}\left[\sum_{w,x} X_w X_x\right] = 0$. As a consequence, (4.7) holds.

Now we prove (4.8). Let l_i be the length of any word corresponding to $X_{n,i}$. Observe that $l_i \in \{k_1+1, \ldots, k_l+1\}$ for any i. At first we expand the left hand side of (4.8).

$$\mathrm{E}[X_{n,1} \ldots X_{n,m}] = \left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{\sum_i (l_i-1)}{2}} \sum_{w_1,\ldots,w_m} E\left[X_{w_1} \ldots X_{w_m}\right]. \tag{4.9}$$

Here the graphs $G_{w_1}, \ldots, G_{w_m}$ are cycles of length $l_1 - 1, \ldots, l_m - 1$ respectively. So in order to have $E\left[X_{w_1} \ldots X_{w_m}\right] \neq 0$, we need each of the edges in $G_{w_1}, \ldots, G_{w_m}$ to be traversed more than once. The sentence $a := (w_1, \ldots, w_m)$ formed by such $(w_1, \ldots, w_m)$ will be called a weak CLT sentence. Given a weak CLT sentence a, we introduce a partition $\eta(a)$, of $\{1, \ldots, m\}$ in the following way. If i, j are in same block of the partition $\eta(a)$, then G_{w_i} G_{w_j} have at least one edge in common.

As a consequence, we can further write the left hand side of (4.9) in the following way.

$$\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{\sum_i (l_i-1)}{2}} \sum_{\eta} \sum_{w_1,\ldots,w_m \mid \eta = \eta(w_1,\ldots,w_m)} E\left[X_{w_1} \ldots X_{w_m}\right]. \tag{4.10}$$

Observe that each block in η should have at least 2 elements. Otherwise, in this case $E\left[X_{w_1} \ldots X_{w_m}\right] = 0$. As a consequence, the number of blocks in $\eta \leq [\frac{m}{2}]$.

Now we prove that if the number of blocks in $\eta < \frac{m}{2}$, then

$$\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{\sum_i (l_i-1)}{2}} \sum_{\eta} \sum_{w_1,\ldots,w_m \mid \eta = \eta(w_1,\ldots,w_m)} E\left[X_{w_1} \ldots X_{w_m}\right] \to 0.$$

If $\eta(w_1, \ldots, w_m)$ have strictly less than $\frac{m}{2}$ blocks, then a has strictly less than $\frac{m}{2}$ connected components. From Proposition 4.9 and Lemma 4.10 of Anderson and Zeitouni [4] it follows that in this case $\#V_a < \sum_{i=1}^m \frac{l_i-1}{2}$. However each connected component is formed by a union of several cycles so $V_a \leq E_a$. Now the following lemma gives a bound on the number of weak CLT sentences having strictly less than $\frac{m}{2}$ connected components.

Lemma 4.3. *Let $\mathcal{A}$ be the set of weak CLT sentences such that for each $a \in \mathcal{A}$, $\#V_a = t$. Then*

$$\#\mathcal{A} \leq 2^{\sum_i l_i} \left(C_1 \sum_i l_i\right)^{C_2 m} \left(\sum_i l_i\right)^{3(\sum_i l_i - 2t)} n^t. \tag{4.11}$$

The proof of Lemma 4.3 is rather technical and requires some amount of random matrix theory. So we defer its proof to the appendix. However, assuming Lemma 4.3, we have

$$\begin{aligned}
&\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{\sum_i (l_i-1)}{2}} \sum_{a\,:\,V_a < \sum_{i=1}^m \frac{l_i-1}{2}} E\left[X_{w_1} \ldots X_{w_m}\right] \\
&\leq \left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{\sum_i (l_i-1)}{2}} \sum_{t < \frac{\sum_i (l_i-1)}{2}} \sum_{e=t}^{\sum_i \frac{(l_i-1)}{2}} 2^{\sum_i l_i} \left(C_1 \sum_i l_i\right)^{C_2 m} \left(\sum_i l_i\right)^{3(\sum_i l_i - 2t)} n^t \hat{p}_n^e.
\end{aligned} \tag{4.12}$$

Now observe that $\sum_{e=t}^{\infty} \hat{p}_n^e \leq \frac{1}{1-\hat{p}_n} \hat{p}_n^t$. As we consider $p < 1$, we have for large enough n, $\frac{1}{1-\hat{p}_n} \leq D$ for some deterministic constant D. Plugging in this estimate in (4.12) we have

the first expression in (4.12) is lesser or equal to

$$D\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{\sum_i (l_i-1)}{2}} \sum_{t<\frac{\sum_i (l_i-1)}{2}} 2^{\sum_i l_i}\left(C_1\sum_i l_i\right)^{C_2 m}\left(\sum_i l_i\right)^{3m}\left(\sum_i l_i\right)^{3((\sum_i l_i-1)-2t)} n^t \hat{p}_n^t.$$

$$\leq D\left(\frac{2}{\sqrt{(1-\hat{p}_n)}}\right)^{\sum_i (l_i-1)} 2^m C_1^{C_2 m}\left(\sum_i l_i\right)^{(C_2+3)m} \underbrace{\sum_{t<\frac{\sum_i (l_i-1)}{2}} \left(\frac{(\sum_i l_i)^6}{n\hat{p}_n}\right)^{\sum_i \frac{l_i-1}{2}-t}}_{T_1((say))}. \tag{4.13}$$

Observe that T_1 is just a geometric series. When $k_l = o(\log(\hat{p}_n n))$ we have,

$$\left(\frac{(\sum_i l_i)^6}{n\hat{p}_n}\right) \leq \frac{(mk_l)^6}{n\hat{p}_n} \to 0.$$

Now, the lowest value of $\sum_{i=1}^m (l_i - 1) - 2t$ is 1. As the geometric series $\sum_{j=1}^\infty \kappa^j$, for $\kappa < 1$ is comparable to its first term, we can give the following final bound to (4.12),

$$C_3\left(\frac{2}{\sqrt{(1-\hat{p}_n)}}\right)^{\sum_i (l_i-1)} 2^m C_1^{C_2 m}\left(\sum_i l_i\right)^{(C_2+3)m} \frac{(\sum_i l_i)^3}{\sqrt{n\hat{p}_n}}. \tag{4.14}$$

Here C_3 is a universal constant. Observe that the dominant term in the numerator of (4.14) is

$$\left(\frac{2}{\sqrt{(1-\hat{p}_n)}}\right)^{\sum_i (l_i-1)} \leq \left(\frac{2}{\sqrt{(1-\hat{p}_n)}}\right)^{m(k_l-1)}.$$

However from our assumption $m(k_l-1)\log\left(\frac{2}{\sqrt{(1-\hat{p}_n)}}\right) - \frac{1}{2}\log(n\hat{p}_n) \to -\infty$. As a consequence, the first expression in (4.12) goes to 0.

Once this is proved all the other partitions left are pair partitions i.e. it has exactly $\frac{m}{2}$ many blocks. In particular, m is even. We now fix a partition η of this kind. Let for any $i \in \{1, \ldots, \frac{m}{2}\}$, $\eta(i,1) < \eta(i,2)$ be the elements in the i th block. Observe now that fixing a pair partition η and $(w_1, \ldots, w_m)$ such that $\eta(w_1, \ldots, w_m) = \eta$, the random variables $X_{w_{\eta(i_1,j)}}$ and $X_{w_{\eta(i_2,j)}}$ are independent when ever $i_1 \neq i_2$ for any $j \in \{1,2\}$. As a consequence, we now can rewrite (4.10) as follows:

$$\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{\sum_i (l_i-1)}{2}} \sum_{\eta} \sum_{w_1,\ldots,w_m \mid \eta=\eta(w_1,\ldots,w_m)} E\left[X_{w_1}\ldots X_{w_m}\right]$$

$$= o(1) + \left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{\sum_i (l_i-1)}{2}} \sum_{\eta \mid \eta \text{ pair parition}} \sum_{w_1,\ldots,w_m \mid \eta=\eta(w_1,\ldots,w_m)} \prod_{i=1}^{\frac{m}{2}} \mathrm{E}\left[X_{w_{\eta(i,1)}} X_{w_{\eta(i,2)}}\right] \tag{4.15}$$

Now observe that whenever $\prod_{i=1}^{\frac{m}{2}} \mathrm{E}\left[X_{w_{\eta(i,1)}} X_{w_{\eta(i,2)}}\right] \neq 0$, we have $w_{\eta(i,1)}$ and $w_{\eta(i,2)}$ are paired. In particular $l(w_{\eta(i,1)}) = l(w_{\eta(i,2)})$ and there are $(1+o(1))(2(l_{\eta(i,1)}-1))n^{l_{\eta(i,1)}-1}$ many such choices of $(w_{\eta(i,1)}, w_{\eta(i,2)})$ for every i. Here $l_{\eta(i,1)}$ is the common length of the words $w_{\eta(i,1)}$ and $w_{\eta(i,2)}$. On the other hand, in this case $\mathrm{E}\left[X_{w_{\eta(i,1)}} X_{w_{\eta(i,2)}}\right] =$

$(\hat{p}_n(1-\hat{p}_n))^{l_{\eta(i,1)}-1}$. Hence, we get the following final reduction to (4.15):

$$\begin{aligned}
&\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{\sum_i(l_i-1)}{2}} \sum_{\eta} \sum_{w_1,\ldots,w_m \mid \eta=\eta(w_1,\ldots,w_m)} E\left[X_{w_1}\ldots X_{w_m}\right] \\
&= o(1) + (1+o(1))\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{\sum_i(l_i-1)}{2}} \sum_{\eta \mid \eta \text{ pair parition}} \prod_{i=1}^{\frac{m}{2}} 2(l_{\eta(i,1)}-1) \\
&\quad \times \mathbb{I}_{l_{\eta(i,1)}=l_{\eta(i,2)}} n^{\frac{\sum_i(l_i-1)}{2}} \left(\hat{p}_n(1-\hat{p}_n)\right)^{\frac{\sum_i(l_i-1)}{2}} \\
&= o(1) + (1+o(1)) \sum_{\eta \mid \eta \text{ pair parition}} \prod_{i=1}^{\frac{m}{2}} 2(l_{\eta(i,1)}-1)\mathbb{I}_{l_{\eta(i,1)}=l_{\eta(i,2)}}.
\end{aligned} \tag{4.16}$$

This completes the proof. □

Proof of part ii) We now give a proof of part ii) of Proposition 4.1. Recall that $d_n = \frac{p_n - q_n}{2}$. We have

$$\begin{aligned}
C_{n,k}(G) &= \left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{k}{2}} \sum_{i_0,i_1,\ldots,i_{k-1}} (x_{i_0,i_1}-\hat{p}_n)\ldots(x_{i_{k-1}i_0}-\hat{p}_n) \\
&= \left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{k}{2}} \sum_{i_0,i_1,\ldots,i_{k-1}} (x_{i_0,i_1}-p_{i_0,i_1}+p_{i_0,i_1}-\hat{p}_n)\ldots(x_{i_{k-1}i_0}-p_{i_{k-1},i_k}+p_{i_{k-1},i_k}-\hat{p}_n) \\
&= \left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{k}{2}} \sum_{i_0,i_1,\ldots,i_{k-1}} (x_{i_0,i_1}-p_{i_0,i_1}+\sigma_{i_0}\sigma_{i_1}d_n)\ldots(x_{i_{k-1}i_0}-p_{i_{k-1},i_k}+\sigma_{i_{k-1}}\sigma_{i_k}d_n) \\
&= \left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{k}{2}} \sum_{i_0,i_1,\ldots,i_{k-1}} \left[(x_{i_0,i_1}-p_{i_0,i_1})\ldots(x_{i_{k-1}i_0}-p_{i_{k-1},i_k}) + d_n^k \prod_{j=0}^{k-1}\sigma_{i_j}\sigma_{i_{j+1}}\right] + V_{n,k}
\end{aligned} \tag{4.17}$$

where $p_{i,j} = p_n$ if $\sigma_i = \sigma_j$ and q_n otherwise. Here $V_{n,k}$ is obtained by taking the sum of all the remaining terms in the expansion of

$$\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{k}{2}} \sum_{i_0,i_1,\ldots,i_{k-1}} (x_{i_0,i_1}-p_{i_0,i_1}+\sigma_{i_0}\sigma_{i_1}d_n)\ldots(x_{i_{k-1}i_0}-p_{i_{k-1},i_k}+\sigma_{i_{k-1}}\sigma_{i_k}d_n)$$

apart from

$$\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{k}{2}} \sum_{i_0,i_1,\ldots,i_{k-1}} (x_{i_0,i_1}-p_{i_0,i_1})\ldots(x_{i_{k-1}i_0}-p_{i_{k-1},i_k})$$

and

$$\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{k}{2}} \sum_{i_0,i_1,\ldots,i_{k-1}} d_n^k \prod_{j=0}^{k-1}\sigma_{i_j}\sigma_{i_{j+1}}.$$

At first we prove that

$$\prod_{j=0}^{k-1}\sigma_{i_j}\sigma_{i_{j+1}} = 1 \tag{4.18}$$

irrespective of the values of σ_i's. The proof of this is straight forward since $i_0 = i_k$ we have

$$\prod_{j=0}^{k-1}\sigma_{i_j}\sigma_{i_{j+1}} = \prod_{j=0}^{k-1}\sigma_j^2 = 1.$$

As $d_n = \sqrt{\frac{c_n\hat{p}_n}{2n}}$, we have

$$\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{k}{2}} \sum_{i_0,i_1,\ldots,i_{k-1}} d_n^k = (1+o(1))\frac{d_n^k n^k}{(n\hat{p}_n(1-\hat{p}_n))^{\frac{k}{2}}} = (1+o(1))\left(\sqrt{\frac{c_n}{2(1-\hat{p}_n)}}\right)^k.$$

This explains the mean term. The proof of asymptotic normality and independence of

$$D_{n,k}(G) := \left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{k}{2}} \left[\sum_{i_0,i_1,\ldots,i_{k-1}} (x_{i_0,i_1} - p_{i_0,i_1})\ldots(x_{i_{k-1}i_0} - p_{i_{k-1},i_k})\right]$$

is exactly same as part i). We only note that here the variance is also $2k$. To see this, we have

$$d_n = \sqrt{\frac{c_n\hat{p}_n}{2n}}$$

and whenever, $k = o(\log(\hat{p}_n n))$ both

$$\lim_{n\to\infty}\left(\frac{(\hat{p}_n + d_n)(1-\hat{p}_n-d_n)}{\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{k}{2}} = 1 \tag{4.19}$$

and

$$\lim_{n\to\infty}\left(\frac{(\hat{p}_n - d_n)(1-\hat{p}_n+d_n)}{\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{k}{2}} = 1. \tag{4.20}$$

It is easy to see that $\mathrm{Var}\left(\frac{D_{n,k}(G)}{\sqrt{2k}}\right)$ lies between the left hand side of (4.19) and (4.20). As a consequence, $\mathrm{Var}\left(\frac{D_{n,k}(G)}{\sqrt{2k}}\right) \to 1$.

It is easy to observe that $\mathrm{E}[V_{n,k}]$ is always 0. Now our final task is to prove $\mathrm{Var}(V_{n,k}) \to 0$. This will prove that $V_{n,k} \xrightarrow{p} 0$ and the proof will be completed.

Let us fix a word w and let $\emptyset \subsetneq E_f \subsetneq E_w$ be any subset. Then

$$V_{n,k} = \sum_w V_{n,k,w}$$

where

$$V_{n,k,w} := \left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{k}{2}} \sum_{\emptyset\subsetneq E_f\subsetneq E_w} \prod_{e\in E_f} \sigma_e d_n \prod_{e\in E\setminus E_f}(x_e - p_e).$$

Here for any edge i,j, $x_e = x_{i,j}$, $p_e = p_{i,j}$ and $\sigma_e = \sigma_i\sigma_j$. Now

$$\mathrm{Var}(V_{n,k}) = \sum_{w,x} \mathrm{Cov}(V_{n,k,w}, V_{n,k,x}).$$

We now find an upper bound of $\mathrm{Cov}(V_{n,k,w}, V_{n,k,x})$.

At first fix any word w and the set $\emptyset \subsetneq E_f \subsetneq E_w$ and consider all the words x such that $E_w \cap E_x = E_w\setminus E_f$. As every edge in G_w and G_x appear exactly once,

$$\begin{aligned}
\mathrm{Cov}(V_{n,k,w}, V_{n,k,x}) &= \sum_{E_w\setminus E'\subset E_w\setminus E_f} \left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^k \prod_{e\in E'}(\pm d_n^2)\,\mathrm{E}\prod_{e\in E_w\setminus E'}(x_e-p_e)^2 \\
&= \sum_{E_w\setminus E'\subset E_w\setminus E_f} \left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^k (\pm d_n^{2\#E'})(1+o(1))\,(\hat{p}_n(1-\hat{p}_n))^{k-\#E'} \\
&\le \sum_{E_w\setminus E'\subset E_w\setminus E_f} (1+o(1))\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^k \left(\frac{c_n}{2}\right)^{\#E'}\left(\frac{\hat{p}_n}{n}\right)^{\#E'} \hat{p}_n^{k-\#E'} \\
&\le (C)^k \frac{1}{n^{k+\#E_f}}
\end{aligned} \tag{4.21}$$

where C is some known constant. The last inequality holds since $\#E' \geq \#E_f$ and $\#(E_w \backslash E' \subset E_w \backslash E_f) \leq 2^k$.

Observe that the graph corresponding to the edges $E_w \backslash E_f$ is a disjoint collection of straight lines. Let the number of such straight lines be ζ. Obviously $\zeta \leq \#(E_w \backslash E_f)$. The number of ways these ζ components can be placed in x is bounded by $k^\zeta \leq k^{\#(E_w \backslash E_f)}$ and all other nodes in x can be chosen freely. So there are at most $n^{k-\#V_{E_w \backslash E_f}} k^{\#(E_w \backslash E_f)}$ choices of such x. Here $V_{E_w \backslash E_f}$ is the set of vertices of the graph corresponding to $(E_w \backslash E_f)$. Observe that, whenever $k > \#E_f > 0$, $E_w \backslash E_f$ is a forest so

$$\#V_{E_w \backslash E_f} \geq \#(E_w \backslash E_f) + 1 \Leftrightarrow k - \#V_{E_w \backslash E_f} \leq \#E_f - 1.$$

As a consequence,

$$\sum_{x \mid E_w \cap E_x = E_w \backslash E_f} \mathrm{Cov}(V_{n,k,w}, V_{n,k,x}) \leq (C)^k \frac{1}{n^{k+\#E_f}} n^{\#E_f - 1} k^{\#(E_w \backslash E_f)} \leq (C)^k \frac{1}{n^{k+1}} k^k. \tag{4.22}$$

The right hand side of (4.22) does not depend on E_f and there are at most 2^k nonempty subsets E_f of E^w. So

$$\sum_x \mathrm{Cov}(V_{n,k,w}, V_{n,k,x}) \leq (2C)^k k^k \frac{1}{n^{k+1}}.$$

Finally there are at most n^k many w. So

$$\sum_w \sum_x \mathrm{Cov}(V_{n,k,w}, V_{n,k,x}) \leq (2C)^k k^k \frac{1}{n}. \tag{4.23}$$

Now we use the fact $k = o(\sqrt{\log(n)})$. In this case

$$k \log(2C) + k \log(k) \leq \sqrt{log(n)} \log(\sqrt{\log n}) = o(log(n)) \Leftrightarrow (2C)^k k^k = o(n).$$

This concludes the proof. □

5 Calculation of second moment and completion of the proof of Theorem 2.1

With Propositions 3.4 and 4.1 in hand the rest of the proof of Theorem 2.1 should be very straight forward. We at first prove that $\lim_{n\to\infty} \mathbb{E}\left(\frac{d\mathbb{P}_n}{d\mathbb{P}'_n}\right)^2$ is the right hand side of (3.1) with $t = \frac{c}{2(1-p)}$ whenever $\frac{a_n}{n} \to p \in [0,1)$.

Lemma 5.1. *Let $Y_n := \frac{d\mathbb{P}_n}{d\mathbb{P}'_n}$. Whenever $p_n \to p \in [0,1)$, we have*

$$\mathrm{E}_{\mathbb{P}'_n}[Y_n^2] \to \exp\left\{-\frac{t}{2} - \frac{t^2}{4}\right\} \frac{1}{\sqrt{1-t}}, \quad t = \frac{c}{2(1-p)} < 1.$$

Proof. The proof of Lemma 5.1 is similar to the proof of Lemma 5.4. in Mossel et al. [27]. The notations used in this proof are slightly different from that of Lemma 5.4 in Mossel et al. [27] for understanding case when p is not necessarily 0.

At first we introduce some notations. Given a labeled graph (G, σ) we define

$$W_{uv} = W_{uv}(G,\sigma) = \begin{cases} \frac{p_n}{\hat{p}_n} & \text{if } \sigma_u\sigma_v = 1 \text{ and } (u,v) \in E \\ \frac{q_n}{\hat{p}_n} & \text{if } \sigma_u\sigma_v = -1 \text{ and } (u,v) \in E \\ \frac{1-p_n}{1-\hat{p}_n} & \text{if } \sigma_u\sigma_v = 1 \text{ and } (u,v) \notin E \\ \frac{1-q_n}{1-\hat{p}_n} & \text{if } \sigma_u\sigma_v = -1 \text{ and } (u,v) \notin E \end{cases} \tag{5.1}$$

and define V_{uv} by the same formula, but with σ replaced by τ. Now

$$Y_n = \frac{1}{2^n} \sum_{\sigma \in \{1,-1\}^n} \prod_{(u,v)} W_{uv}$$

and

$$Y_n^2 = \frac{1}{2^{2n}} \sum_{\sigma,\tau} \prod_{(u,v)} W_{uv} V_{uv}.$$

Since $\{W_{uv}\}$ are independent given σ, it follows that

$$\mathrm{E}_{\mathbb{P}'_n}(Y_n^2) = \frac{1}{2^{2n}} \sum_{\sigma,\tau} \prod_{(u,v)} \mathrm{E}_{\mathbb{P}'_n}\left(W_{uv} V_{uv}\right).$$

Now we consider the following cases:

1. $\sigma_u \sigma_v = 1$ and $\tau_u \tau_v = 1$.
2. $\sigma_u \sigma_v = -1$ and $\tau_u \tau_v = -1$.
3. $\sigma_u \sigma_v = 1$ and $\tau_u \tau_v = -1$.
4. $\sigma_u \sigma_v = -1$ and $\tau_u \tau_v = 1$.

Let $t = \frac{c}{2(1-p)}$. We at first calculate $\mathrm{E}_{\mathbb{P}'_n}(W_{uv}V_{uv})$ for cases 1 and 3.
Case 1:

$$\begin{aligned}
\mathrm{E}_{\mathbb{P}'_n}(W_{uv}V_{uv}) &= \left(\frac{p_n}{\hat{p}_n}\right)^2 \hat{p}_n + \left(\frac{1-p_n}{1-\hat{p}_n}\right)^2 (1-\hat{p}_n). \\
&= \frac{p_n^2}{\hat{p}_n} + \frac{(1-p_n)^2}{1-\hat{p}_n} \\
&= \frac{(\hat{p}_n + d_n)^2}{\hat{p}_n} + \frac{(1-\hat{p}_n - d_n)^2}{1-\hat{p}_n} \\
&= 1 + d_n^2 (\frac{1}{\hat{p}_n} + \frac{1}{1-\hat{p}_n}) = 1 + \frac{d_n^2}{\hat{p}_n(1-\hat{p}_n)} = 1 + \frac{c_n}{2n(1-\hat{p}_n)} \\
&= 1 + \frac{t_n}{n}
\end{aligned} \tag{5.2}$$

where $d_n = \frac{p_n - q_n}{2}$ and $t_n = \frac{c_n}{2(1-\hat{p}_n)} = (1+o(1))t$ as before.
Case 3:

$$\begin{aligned}
\mathrm{E}_{\mathbb{P}'_n}(W_{uv}V_{uv}) &= \left(\frac{p_n}{\hat{p}_n} \cdot \frac{q_n}{\hat{p}_n}\right) \hat{p}_n + \left(\frac{1-p_n}{1-\hat{p}_n} \cdot \frac{1-q_n}{1-\hat{p}_n}\right) (1-\hat{p}_n). \\
&= \frac{p_n q_n}{\hat{p}_n} + \frac{(1-p_n)(1-q_n)}{1-\hat{p}_n} \\
&= \frac{(\hat{p}_n + d_n)(\hat{p}_n - d_n)}{\hat{p}_n} + \frac{(1-\hat{p}_n - d_n)(1-\hat{p}_n + d_n)}{1-\hat{p}_n} \\
&= 1 - d_n^2 (\frac{1}{\hat{p}_n} + \frac{1}{1-\hat{p}_n}) = 1 - \frac{d_n^2}{\hat{p}_n(1-\hat{p}_n)} = 1 - \frac{t_n}{n}
\end{aligned} \tag{5.3}$$

It is easy to observe that $\mathrm{E}_{\mathbb{P}'_n}(W_{uv}V_{uv}) = 1 + \frac{t_n}{n}$ and $1 - \frac{t_n}{n}$ for Case 2 and Case 4 respectively.

We now introduce another parameter $\rho = \rho(\sigma,\tau) = \frac{1}{n}\sum_i \sigma_i \tau_i$. Let $S_\pm$ be the number of $\{u,v\}$ such that $\sigma_u \sigma_v \tau_u \tau_v = \pm 1$ respectively. It is easy to observe that

$$\rho^2 = \frac{1}{n} + \frac{2}{n^2}(S_+ - S_-) \tag{5.4}$$

and

$$1-\frac{1}{n}=\frac{2}{n^2}(S_+ + S_-). \tag{5.5}$$

So

$$S_+ = (1+\rho^2)\frac{n^2}{4}-\frac{n}{2},\ \ S_- = (1-\rho^2)\frac{n^2}{4}. \tag{5.6}$$

Now

$$\begin{aligned}\mathrm{E}_{\mathbb{P}'_n}(Y_n^2) &= \frac{1}{2^{2n}}\sum_{\sigma,\tau}\left(1+\frac{t_n}{n}\right)^{S_+}\left(1-\frac{t_n}{n}\right)^{S_-}\\ &= \frac{1}{2^{2n}}\sum_{\sigma,\tau}\left(1+\frac{t_n}{n}\right)^{(1+\rho^2)\frac{n^2}{4}-\frac{n}{2}}\left(1-\frac{t_n}{n}\right)^{(1-\rho^2)\frac{n^2}{4}}.\end{aligned} \tag{5.7}$$

Observe that $t_n = (1+o(1))t$ is a bounded sequence. It is easy to check by taking logarithm and Taylor expansion that for any bounded sequence x_n,

$$\left(1+\frac{x_n}{n}\right)^{n^2} = (1+o(1))\exp\left\{nx_n-\frac{1}{2}x_n^2\right\}.$$

So we can write the right hand side of (5.7) as

$$\begin{aligned}&(1+o(1))\frac{1}{2^{2n}}\sum_{\sigma,\tau}e^{-\frac{t_n}{2}}\exp\left[\left(nt_n-\frac{t_n^2}{2}\right)\left(\frac{1+\rho^2}{4}\right)\right]\times\exp\left[\left(-nt_n-\frac{t_n^2}{2}\right)\left(\frac{1-\rho^2}{4}\right)\right]\\ =&(1+o(1))\frac{1}{2^{2n}}\sum_{\sigma,\tau}e^{-\frac{t_n}{2}-\frac{t_n^2}{4}}\exp\left[\frac{nt_n\rho^2}{2}\right]\\ =&(1+o(1))e^{-\frac{t_n}{2}-\frac{t_n^2}{4}}\frac{1}{2^{2n}}\sum_{\sigma,\tau}\exp\left[\frac{(1+o(1))tn\rho^2}{2}\right]\end{aligned} \tag{5.8}$$

From Lemma 5.5 in Mossel et al. [27]

$$\frac{1}{2^{2n}}\sum_{\sigma,\tau}\exp\left[\frac{(1+o(1))nt\rho^2}{2}\right]\to\frac{1}{\sqrt{1-t}}.$$

So the right hand side of (5.8) converges to

$$\exp\left\{-\frac{t}{2}-\frac{t^2}{4}\right\}\frac{1}{\sqrt{1-t}}$$

as required. □

Proof of Theorem 2.1:
Proof of part i) We take $X_{n,i} = C_{n,i}(G)$.

At first observe that when $p_n\to p\in[0,1)$ for any fixed i, $\mu_i := \left(\sqrt{\frac{c_n}{2(1-\hat{p}_n)}}\right)^i$ converges to $\left(\sqrt{\frac{c}{2(1-p)}}\right)^i$ as $n\to\infty$.

From Proposition 4.1 and Lemma 4.1 we see that $C_{n,i}(G)$'s satisfy all the required conditions for Proposition 3.4. Hence $\mathbb{P}_n$ and $\mathbb{P}'_n$ are mutually contiguous.

It is easy to see that the estimate $\hat{d}_n := \frac{1}{n-1}\sum_{i\neq j}x_{i,j}$ has mean $\frac{a_n+b_n}{2}$ and variance $O(\frac{a_n+b_n}{n})$. So

$$\hat{d}_n-\frac{a_n+b_n}{2} = o_p(\sqrt{a_n+b_n}) = o_p(a_n-b_n)$$

Suppose under $\mathbb{P}_n$ there exist estimators A_n of a_n and B_n of b_n such that

$$|A_n - a_n| + |B_n - b_n| = o_p(a_n - b_n).$$

Then $2(\hat{d}_n - B_n) - (a_n - b_n) = o_p(a_n - b_n)$ i.e.

$$\frac{2(\hat{d}_n - B_n)}{a_n - b_n} | \mathbb{P}_n \xrightarrow{P} 1.$$

However, from the fact that $\mathbb{P}_n$ and $\mathbb{P}'_n$ are mutually contiguous we also have

$$\frac{2(\hat{d}_n - B_n)}{a_n - b_n} | \mathbb{P}'_n \xrightarrow{P} 1$$

which is impossible.

Proof of part ii) It is easy to observe that $\mathbb{P}_n$ and $\mathbb{P}'_n$ are asymptotically singular as for any $k_n \to \infty$, $\frac{\mu_{k_n}}{\sqrt{2k_n}} \to \infty$. Now we construct estimators for a_n and b_n. Let us define

$$\hat{f}_{n,k_n} = \begin{cases} \left(\sqrt{2k_n} C_{n,k_n}(G)\right)^{\frac{1}{k_n}} & \text{if } C_{n,k_n}(G) > 0 \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to see that under $\mathbb{P}_n$ $\hat{f}_{n,k_n} \xrightarrow{P} \sqrt{\frac{c}{2(1-p)}}$ as $k_n \to \infty$. We have seen earlier that under $\mathbb{P}_n$

$$\begin{aligned} &\frac{\hat{d}_n - \frac{(a_n+b_n)}{2}}{\sqrt{a_n+b_n}} \xrightarrow{P} 0 \Rightarrow \frac{\hat{d}_n - \frac{(a_n+b_n)}{2}}{a_n+b_n} \xrightarrow{P} 0 \Rightarrow \sqrt{\frac{\hat{d}_n}{\frac{a_n+b_n}{2}}} \xrightarrow{P} 1. \\ \Rightarrow &\sqrt{\hat{d}_n} - \sqrt{\frac{a_n+b_n}{2}} = o_p(\sqrt{a_n+b_n}) = o_p(a_n - b_n) \end{aligned} \tag{5.9}$$

As $\hat{p}_n \to p$,

$$\sqrt{\frac{\hat{d}_n(1-\hat{p}_n)}{\frac{a_n+b_n}{2}(1-p)}} \xrightarrow{P} 1.$$

So

$$\sqrt{\hat{d}_n(1-\hat{p}_n)} - \sqrt{\frac{a_n+b_n}{2}(1-p)} = o_p(a_n - b_n) \; \forall \; p \in [0,1).$$

So $\sqrt{\hat{d}_n(1-\hat{p}_n)}\hat{f}_{n,k_n} - \frac{a_n-b_n}{2} = o_p(a_n - b_n)$ under $\mathbb{P}_n$. As a consequence, the estimators $\hat{A} = \hat{d}_n + \sqrt{\hat{d}_n(1-\hat{p}_n)}\hat{f}_{n,k_n}$ and $\hat{B} = \hat{d}_n - \sqrt{\hat{d}_n(1-\hat{p}_n)}\hat{f}_{n,k_n}$ have the required property. This concludes the proof. □

We end the discussion of this section by the following remark on the computation of the signed cycles.

Remark 5.1. In general the direct computation of the random variables $C_{n,k}(G)$'s take at least $O(n^k)$ amount of time. So it might appear that the statistics $C_{n,k}(G)$'s are not useful for any practical purpose. Fortunately, this is not the case. It was proved in Banerjee and Ma(2017) [7] that whenever k is odd, the difference between $C_{n,k}(G)$ and $\sum_{i=1}^n P_k(\lambda_i)$ converges in probability to 0 for any $k = o\left(\min(\log(\hat{p}_n n), \sqrt{\log(n)})\right)$. Here $\{\lambda_i\}_{1 \le i \le n}$ are the eigenvalues of the centered adjacency matrix of the graph and $P_k(\cdot)$ is the Chebyshev polynomial of degree k (look at (2.7)-(2.8) in Banerjee and Ma(2017) [7] for definition). The case when k is even is more complicated. In this case one can prove $C_{n,2k}(G) - \sum_{i=1}^n P_{2k}(\lambda_i) - E_{2k} \xrightarrow{P} 0$ where E_{2k} is an additional error term. One can prove

that $\mathrm{Var}[E_{2k}]$ converges to 0 and find the asymptotic value of $\mathrm{E}[E_{2k}]$ explicitly under additional growth conditions on $\hat{p}_n$. As a consequence, the signed cycles of growing orders can be computed by the spectral decomposition of the centered adjacency matrix of the graph. It is well known that this has $O(n^3 \log(n))$ time complexity. One might check Banerjee and Ma(2017) [7] for details.

6 Proof of non reconstructability

In this section we provide a proof of the non-reconstruction results stated in Theorem 2.2. Our proof technique relies on fine analysis of conditional probabilities. Technically, this proof is closely related to the non-reconstruction proof in section 6.2 of Banks et al. [8] rather than the original proof given in Mossel et al. [27]. At first we prove one Proposition and one Lemma which will be crucial for our proof.

Proposition 6.1. *Suppose $a_n, b_n \to \infty$, $\frac{a_n}{n} \to p \in [0,1)$, $c_n \to c$ and $c < 2(1-p)$. Then for any fixed r and any two configurations $(\sigma_1^{(1)}, \ldots, \sigma_r^{(1)})$, $(\sigma_1^{(2)}, \ldots, \sigma_r^{(2)})$*

$$\mathrm{TV}\left(\mathbb{P}_n(G|(\sigma_1^{(1)}, \ldots, \sigma_r^{(1)})), \mathbb{P}_n(G|(\sigma_1^{(2)}, \ldots, \sigma_r^{(2)}))\right) = o(1)$$

Here $\mathrm{TV}(\mu_1, \mu_2)$ is the total variation distance between two probability measures μ_1 and μ_2.

Proof. We know that

$$\begin{aligned}
&\mathrm{TV}\left(\mathbb{P}_n(G|\sigma_u^{(1)} \;\; u \in [r]), \mathbb{P}_n(G|\sigma_u^{(2)} \;\; u \in [r])\right) \\
&= \sum_G \left|(\mathbb{P}_n(G|\sigma_u^{(1)} \;\; u \in [r]) - \mathbb{P}_n(G|\sigma_u^{(2)} \;\; u \in [r])\right| \\
&= \sum_G \left|(\mathbb{P}_n(G|\sigma_u^{(1)} \;\; u \in [r]) - \mathbb{P}_n(G|\sigma_u^{(2)} \;\; u \in [r])\right| \frac{\sqrt{\mathbb{P}'_n(G)}}{\sqrt{\mathbb{P}'_n(G)}} \\
&\leq \left(\sum_G \mathbb{P}'_n(G)\right)^{\frac{1}{2}} \left(\sum_G \frac{\left(\mathbb{P}_n(G|\sigma_u^{(1)} \;\; u \in [r]) - \mathbb{P}_n(G|\sigma_u^{(2)} \;\; u \in [r]\right)^2}{\mathbb{P}'_n(G)}\right)^{\frac{1}{2}} \\
&= \left(\sum_G \frac{\left(\sum_{\tilde{\sigma}} \mathbb{P}_n(\tilde{\sigma}) \left(\mathbb{P}_n(G|\sigma^{(1)}, \tilde{\sigma}) - \mathbb{P}_n(G|\sigma^{(2)}, \tilde{\sigma})\right)\right)^2}{\mathbb{P}'_n(G)}\right)^{\frac{1}{2}}.
\end{aligned} \tag{6.1}$$

Here $\sigma^{(1)} := \left\{(\sigma_1^{(1)}, \ldots, \sigma_r^{(1)}\right\}$, $\sigma^{(2)} := \left\{(\sigma_1^{(2)}, \ldots, \sigma_r^{(2)})\right\}$ and $\tilde{\sigma}$ is any configuration on $\{r+1, \ldots, n\}$.

Now observe that

$$\begin{aligned}
&\left(\sum_{\tilde{\sigma}} \mathbb{P}_n(\tilde{\sigma}) \left(\mathbb{P}_n(G|\sigma^{(1)}, \tilde{\sigma}) - \mathbb{P}_n(G|\sigma^{(2)}, \tilde{\sigma}\right)\right)^2 \\
&= \sum_{\tilde{\sigma}, \tilde{\tau}} \mathbb{P}_n(\tilde{\sigma}) \mathbb{P}_n(\tilde{\tau}) \Big(\mathbb{P}_n(G|\sigma^{(1)}, \tilde{\sigma}) \mathbb{P}_n(G|\sigma^{(1)}, \tilde{\tau}) + \mathbb{P}_n(G|\sigma^{(2)}, \tilde{\sigma}) \mathbb{P}_n(G|\sigma^{(2)}, \tilde{\tau}) \\
&- \mathbb{P}_n(G|\sigma^{(1)}, \tilde{\sigma}) \mathbb{P}_n(G|\sigma^{(2)}, \tilde{\tau}) - \mathbb{P}_n(G|\sigma^{(2)}, \tilde{\sigma}) \mathbb{P}_n(G|\sigma^{(1)}, \tilde{\tau})\Big).
\end{aligned} \tag{6.2}$$

We shall prove that the value of

$$\sum_G \sum_{\tilde{\sigma}, \tilde{\tau}} \mathbb{P}_n(\tilde{\sigma}) \mathbb{P}_n(\tilde{\tau}) \frac{\mathbb{P}_n(G|\sigma^{(1)}, \tilde{\sigma}) \mathbb{P}_n(G|\sigma^{(2)}, \tilde{\tau})}{\mathbb{P}'_n(G)} \tag{6.3}$$

does not depend on $\sigma^{(1)}$ and $\sigma^{(2)}$ upto $o(1)$ terms. This will prove that the final expression in (6.1) goes to 0. As a consequence, the proof of Proposition 6.1 will be complete.

At first we recall the definition of $W_{uv}(G,\sigma)$ from (5.1). It is easy to observe that

$$\begin{aligned}
&\sum_{G}\sum_{\tilde{\sigma},\tilde{\tau}} \frac{\mathbb{P}_n(\tilde{\sigma})\mathbb{P}_n(\tilde{\tau})\left(\mathbb{P}_n(G|\sigma^{(1)},\tilde{\sigma})\mathbb{P}_n(G|\sigma^{(2)},\tilde{\tau})\right)}{\mathbb{P}'_n(G)} \\
&= \sum_{\tilde{\sigma},\tilde{\tau}} \frac{1}{2^{2(n-r)}} \sum_{G} \left(\prod_{uv} W(G,\sigma^{(1)},\tilde{\sigma})W(G,\sigma^{(2)},\tilde{\tau})\right)\mathbb{P}'_n(G) \qquad (6.4)\\
&= \frac{1}{2^{2(n-r)}} \sum_{\tilde{\sigma},\tilde{\tau}} \prod_{u,v} \mathrm{E}_{\mathbb{P}'_n}(W(G,\sigma^{(1)},\tilde{\sigma})W(G,\sigma^{(2)},\tilde{\tau})).
\end{aligned}$$

Observe that the sum in the final expression of (6.4) is taken over $(\tilde{\sigma},\tilde{\tau})$ so the configurations in $\sigma^{(1)}$ and $\sigma^{(2)}$ remain unchanged.

Now let us introduce the following parameters

$$\begin{aligned}
\rho^{\mathrm{fix}} &:= \frac{1}{r}\sum_{i=1}^{r}\sigma_i^{(1)}\sigma_i^{(2)} \\
S_{\pm}^{\mathrm{fix}} &:= \sum_{u,v\in[r]} I_{\{\sigma_u^{(1)}\sigma_v^{(1)}\sigma_u^{(2)}\sigma_v^{(2)}=\pm 1\}}
\end{aligned} \qquad (6.5)$$

where I_A denotes the indicator variable corresponding to set A. We similarly define

$$\begin{aligned}
\rho(\tilde{\sigma},\tilde{\tau}) &:= \frac{1}{n-r}\sum_{i=r+1}^{n}\tilde{\sigma}_i\tilde{\tau}_i \\
S_{\pm}(\tilde{\sigma},\tilde{\tau}) &:= \sum_{u,v\notin[r]} I_{\{\tilde{\sigma}_u\tilde{\sigma}_v\tilde{\tau}_u\tilde{\tau}_v=\pm 1\}}.
\end{aligned} \qquad (6.6)$$

Finally for each $u\in[r]$ define

$$S_{u,\pm}(\tilde{\sigma},\tilde{\tau}) = \#\{v\notin[r] \,:\, \tilde{\sigma_v}\tilde{\tau_v} = \pm\sigma_u^{(1)}\sigma_u^{(2)}\}. \qquad (6.7)$$

By using arguments similar to the proof of Lemma 5.1 one can show that the right hand side of the final expression of (6.4) further simplifies to

$$\begin{aligned}
&= \left(1+\frac{t_n}{n}\right)^{S_+^{\mathrm{fix}}}\left(1-\frac{t_n}{n}\right)^{S_-^{\mathrm{fix}}} \frac{1}{2^{2(n-r)}}\sum_{\tilde{\sigma},\tilde{\tau}}\left(1+\frac{t_n}{n}\right)^{S_+(\tilde{\sigma},\tilde{\tau})}\left(1-\frac{t_n}{n}\right)^{S_-(\tilde{\sigma},\tilde{\tau})}\times \\
&\qquad \prod_{u\in[r]}\left(1+\frac{t_n}{n}\right)^{S_{u,+}(\tilde{\sigma},\tilde{\tau})}\left(1-\frac{t_n}{n}\right)^{S_{u,-}(\tilde{\sigma},\tilde{\tau})} \\
&= \left(1+\frac{t_n}{n}\right)^{S_+^{\mathrm{fix}}}\left(1-\frac{t_n}{n}\right)^{S_-^{\mathrm{fix}}} \frac{1}{2^{2(n-r)}}\sum_{\tilde{\sigma},\tilde{\tau}}\left(1+\frac{t_n}{n}\right)^{\left(1+\rho(\tilde{\sigma},\tilde{\tau})^2\right)\frac{(n-r)^2}{4}-\frac{n-r}{2}}\times \\
&\qquad \left(1-\frac{t_n}{n}\right)^{\left(1-\rho(\tilde{\sigma},\tilde{\tau})^2\right)\frac{(n-r)^2}{4}} \prod_{u\in[r]}\left(1+\frac{t_n}{n}\right)^{n\frac{S_{u,+}(\tilde{\sigma},\tilde{\tau})}{n}}\left(1-\frac{t_n}{n}\right)^{n\frac{S_{u,-}(\tilde{\sigma},\tilde{\tau})}{n}}.
\end{aligned} \qquad (6.8)$$

It is easy to see that for any fixed $u\in[r]$ and $\sigma_u^{(1)}$, $\sigma_u^{(2)}$ when $\tilde{\sigma}$ and $\tilde{\tau}$ are chosen independently and uniformly over $\{\pm 1\}$ for each vertex $v\notin[r]$, both $\frac{S_{u,\pm}(\tilde{\sigma},\tilde{\tau})}{n} \overset{a.s.}{\to} \frac{1}{2}$. On the other hand $|S_{u,\pm}|\le n$. So both the quantities

$$\prod_{u\in[r]}\left(1+\frac{t_n}{n}\right)^{n\frac{S_{u,\pm}(\tilde{\sigma},\tilde{\tau})}{n}}$$

and

$$\prod_{u\in[r]}\left(1-\frac{t_n}{n}\right)^{n\frac{S_{u,-}(\tilde\sigma,\tilde\tau)}{n}}$$

are uniformly bounded over $\tilde\sigma$, $\tilde\tau$ and converge almost surely to $\exp\left(\frac{tr}{2}\right)$ and $\exp\left(-\frac{tr}{2}\right)$ under uniform independent assignment.

Now $S_+^{\rm fix}$ and $S_-^{\rm fix}$ are both bounded by r^2 also $t_n=(1+o(1))t$. So

$$\left(1+\frac{t_n}{n}\right)^{S_+^{\rm fix}}\left(1-\frac{t_n}{n}\right)^{S_-^{\rm fix}}=(1+o(1)).$$

On the other hand one can repeat the arguments in the proof of Lemma 5.1 to conclude that

$$\sum_{\tilde\sigma,\tilde\tau}\left(1+\frac{t_n}{n}\right)^{\left(1+\rho(\tilde\sigma,\tilde\tau)^2\right)\frac{(n-r)^2}{4}-\frac{n-r}{2}}\left(1-\frac{t_n}{n}\right)^{\left(1-\rho(\tilde\sigma,\tilde\tau)^2\right)\frac{(n-r)^2}{4}}\to\frac{1}{\sqrt{1-t}}\exp\left\{-\frac{t}{2}-\frac{t^2}{4}\right\}.$$

Combining all the arguments one gets the first expression in (6.8) converges to

$$\begin{aligned}&\frac{1}{\sqrt{1-t}}\exp\left\{-\frac{t}{2}-\frac{t^2}{4}\right\}\exp\left(\frac{tr}{2}\right)\exp\left(-\frac{tr}{2}\right)\\&=\frac{1}{\sqrt{1-t}}\exp\left\{-\frac{t}{2}-\frac{t^2}{4}\right\}.\end{aligned}$$

As a result

$$\sum_G\sum_{\tilde\sigma,\tilde\tau}\mathbb{P}_n(\tilde\sigma)\mathbb{P}_n(\tilde\tau)\frac{\mathbb{P}_n(G|\sigma^{(1)},\tilde\sigma)\mathbb{P}_n(G|\sigma^{(2)},\tilde\tau)}{\mathbb{P}'_n(G)}=(1+o(1))\frac{1}{\sqrt{1-t}}\exp\left\{-\frac{t}{2}-\frac{t^2}{4}\right\}$$

irrespective of the value of $\sigma^{(1)}$ and $\sigma^{(2)}$. So the final expression in (6.1) goes to 0. Hence the proof is complete. □

We now prove the following easy consequence of Proposition 6.1 which states that the posterior distribution of a single label is essentially unchanged if we know a bounded number of other labels.

Lemma 6.1. *Suppose S is a set of finite cardinality r, $u\notin S$ be a fixed node and π gives probability $\frac{1}{2}$ to both ±1. Then under the conditions of Proposition 6.1*

$$\mathrm{E}\left[\mathrm{TV}(\mathbb{P}_n(\sigma_u|G,\sigma_S),\pi)|\sigma_S\right]=o(1).$$

Proof. Observe that $\mathbb{P}_n(\sigma_u=i)=\pi(i)$ from the model assumption. So

$$\begin{aligned}&\mathrm{E}\left[\mathrm{TV}(\mathbb{P}_n(\sigma_u|G,\sigma_S),\pi)|\sigma_S\right]\\&=\sum_G\sum_{i=\pm1}\left|\mathbb{P}_n\left(\sigma_u=i|G,\sigma_S\right)-\mathbb{P}_n(\sigma_u=i)\right|\mathbb{P}_n(G|\sigma_S)\\&=\sum_{i=\pm1}\mathbb{P}_n(\sigma_u=i)\sum_G\left|\frac{\mathbb{P}_n\left(\sigma_u=i|G,\sigma_S\right)}{\mathbb{P}_n(\sigma_u=i)}-1\right|\mathbb{P}_n(G|\sigma_S)\\&=\sum_{i=\pm1}\mathbb{P}_n(\sigma_u=i)\sum_G\left|\frac{\mathbb{P}_n\left(\sigma_u=i\cap G\cap\sigma_S\right)\mathbb{P}_n(\sigma_S)}{\mathbb{P}_n(\sigma_u=i\cap\sigma_S)\mathbb{P}_n(G\cap\sigma_S)}-1\right|\mathbb{P}_n(G|\sigma_S)\\&=\sum_{i=\pm1}\mathbb{P}_n(\sigma_u=i)\sum_G\left|\frac{\mathbb{P}_n(G|\sigma_S,\sigma_u=i)}{\mathbb{P}_n(G|\sigma_S)}-1\right|\mathbb{P}_n(G|\sigma_S)\end{aligned}\tag{6.9}$$

Observe that

$$\mathbb{P}_n(G|\sigma_S) = \frac{1}{2}\left(\mathbb{P}_n(G|\sigma_S, \sigma_u = 1) + \mathbb{P}_n(G|\sigma_S, \sigma_u = -1)\right).$$

As a consequence, the final expression of the right hand side of (6.9) becomes

$$\frac{1}{2}\sum_{i=\pm 1} \mathbb{P}_n(\sigma_u = i)\mathrm{TV}\left(\mathbb{P}_n(G|\sigma_S, \sigma_u = i), \mathbb{P}_n(G|\sigma_S, \sigma_u = -i)\right).$$

So the proof is complete by applying Proposition 6.1. □

With Proposition 6.1 and Lemma 6.1 in hand, we now give a proof of Theorem 2.2.

Proof of Theorem 2.2:

Let $\hat{\sigma}$ be any estimate of the labeling of the nodes, σ be the true labeling and $f : \{1, 2\} \to \{\pm 1\}$ be the function such that $f(1) = 1$ and $f(2) = -1$.

It is elementary to check that

$$\frac{1}{2}\mathrm{ov}(\sigma, \hat{\sigma}) = \frac{1}{n}\left[N_{11} + N_{22} - \frac{1}{n}(N_{1\cdot}N_{\cdot 1}) - \frac{1}{n}(N_{2\cdot}N_{\cdot 2})\right]. \tag{6.10}$$

Here

$$\begin{aligned} N_{ij} &= \left|\sigma^{-1}\{f(i)\} \cap \hat{\sigma}^{-1}\{f(j)\}\right| \\ N_{i\cdot} &= \left|\sigma^{-1}\{f(i)\}\right| \\ N_{\cdot j} &= \left|\hat{\sigma}^{-1}\{f(j)\}\right|. \end{aligned} \tag{6.11}$$

So it is sufficient to prove that

$$\frac{1}{n^2}\,\mathrm{E}_{\mathbb{P}_n}\left[N_{ii} - \frac{1}{n}N_{i\cdot}N_{\cdot i}\right]^2 = \frac{1}{n^2}\,\mathrm{E}_{\mathbb{P}_n}\left[N_{ii}^2 - \frac{2}{n}N_{ii}N_{i\cdot}N_{\cdot i} + \frac{1}{n^2}N_{i\cdot}^2N_{\cdot i}^2\right] \to 0 \quad i \in \{1, 2\}.$$

Now

$$\begin{aligned} \mathrm{E}_{\mathbb{P}_n}\left[N_{ii}^2\right] &= \mathrm{E}_{\mathbb{P}_n}\left[\sum_{u,v} I_{\{\sigma_u = f(i)\}}I_{\{\sigma_v = f(i)\}}I_{\{\hat{\sigma}_u = f(i)\}}I_{\{\hat{\sigma}_v = f(i)\}}\right] \\ &= \mathrm{E}_{\mathbb{P}_n}\left[\mathrm{E}\left[\sum_{u,v} I_{\{\sigma_u = f(i)\}}I_{\{\sigma_v = f(i)\}}I_{\{\hat{\sigma}_u = f(i)\}}I_{\{\hat{\sigma}_v = f(i)\}}\right]|G\right] \\ &= \mathrm{E}_{\mathbb{P}_n}\left[\mathrm{E}\left[\sum_{u,v} I_{\{\sigma_u = f(i)\}}I_{\{\sigma_v = f(i)\}}\right] I_{\{\hat{\sigma}_u = f(i)\}}I_{\{\hat{\sigma}_v = f(i)\}}\,|G\right] \end{aligned} \tag{6.12}$$

The last step follows from the fact that $\hat{\sigma}$ is a function of G. Now

$$\begin{aligned} \mathrm{E}\left[I_{\{\sigma_u = f(i)\}}I_{\{\sigma_v = f(i)\}}|G\right] &= \mathrm{E}\left[I_{\{\sigma_u = f(i)\}}|G, \sigma_v = f(i)\right]\mathbb{P}_n\left(\sigma_v = f(i)|G\right) \\ &= (\pi(f(i)) + o(1))\mathbb{P}_n(G|\sigma_v = f(i))\frac{\mathbb{P}_n(\sigma_v = f(i))}{\mathbb{P}_n(G)} \\ &= (\pi^2(f(i)) + o(1))\frac{\mathbb{P}_n(G|\sigma_v = f(i))}{\mathbb{P}_n(G)} \end{aligned}$$

Here the second step follows from Lemma 6.1. Now,

$$\begin{aligned}
&\left|\mathrm{E}_{\mathbb{P}_n}\left[\mathrm{E}\sum_{u,v}\left(I_{\{\sigma_u=f(i)\}}I_{\{\sigma_v=f(i)\}}-\pi^2(f(i))\right)I_{\{\hat{\sigma}_u=f(i)\}}I_{\{\hat{\sigma}_v=f(i)\}}\,|G\right]\right| \\
&\leq \mathrm{E}_{\mathbb{P}_n}\left[\sum_{u,v}\left|\mathrm{E}\left[\left(I_{\{\sigma_u=f(i)\}}I_{\{\sigma_v=f(i)\}}-\pi^2(f(i))\right)I_{\{\hat{\sigma}_u=f(i)\}}I_{\{\hat{\sigma}_v=f(i)\}}\,|G\right]\right|\right] \\
&= \mathrm{E}_{\mathbb{P}_n}\left[\sum_{u,v}\left|\pi^2(f(i))I_{\{\hat{\sigma}_u=f(i)\}}I_{\{\hat{\sigma}_v=f(i)\}}\left(\frac{\mathbb{P}_n(G|\sigma_v=f(i))}{\mathbb{P}_n(G)}-1\right)+o(1)\right|\right] \\
&\leq \sum_{u,v}\sum_{G}|\mathbb{P}_n(G|\sigma_v=f(i))-\mathbb{P}_n(G)|+o(n^2) \\
&= o(n^2).
\end{aligned} \tag{6.13}$$

Here the last step follows from Proposition 6.1.

So we have

$$\mathrm{E}_{\mathbb{P}_n}\left[N_{ii}^2\right]=\sum_{u,v}\mathrm{E}_{\mathbb{P}_n}\left[\pi^2(f(i))I_{\{\hat{\sigma}_u=f(i)\}}I_{\{\hat{\sigma}_v=f(i)\}}\right]+o(n^2) \tag{6.14}$$

Similar calculations will prove that

$$\mathrm{E}_{\mathbb{P}_n}\left[N_{ii}N_{i\cdot}N_{\cdot i}\right]=n\sum_{u,v}\mathrm{E}_{\mathbb{P}_n}\left[\pi^2(f(i))I_{\{\hat{\sigma}_u=f(i)\}}I_{\{\hat{\sigma}_v=f(i)\}}\right]+o(n^3) \tag{6.15}$$

and

$$\mathrm{E}_{\mathbb{P}_n}\left[N_{i\cdot}^2N_{\cdot i}^2\right]=n^2\sum_{u,v}\mathrm{E}_{\mathbb{P}_n}\left[\pi^2(f(i))I_{\{\hat{\sigma}_u=f(i)\}}I_{\{\hat{\sigma}_v=f(i)\}}\right]+o(n^4). \tag{6.16}$$

Plugging in these estimates we have

$$\frac{1}{n^2}\,\mathrm{E}_{\mathbb{P}_n}\left[N_{ii}-\frac{1}{n}N_{i\cdot}N_{\cdot i}\right]^2=o(1).$$

This completes the proof. □

References

[1] E. Abbe. Community detection and stochastic block models: recent developments. *J. Mach. Learn. Res.*, To appear, 2017.

[2] E. Abbe and C. Sandon. Achieving the ks threshold in the general stochastic block model with linearized acyclic belief propagation. In *Advances in Neural Information Processing Systems 29*, pages 1334–1342, 2016.

[3] E. Abbe, A. S. Bandeira, and G. Hall. Exact recovery in the stochastic block model. *IEEE Transactions on Information Theory*, 62(1):471–487, 2016.

[4] G. W. Anderson and O. Zeitouni. A CLT for a band matrix model. *Probab. Theory Related Fields*, 134(2):283–338, 2006. MR-3447993

[5] G. W. Anderson, A. Guionnet, and O. Zeitouni. *An introduction to random matrices*. Cambridge University Press, Cambridge, 2010. MR-2222385

[6] D. Banerjee and A. Bose. Largest eigenvalue of large random block matrices: A combinatorial approach. *Random Matrices: Theory and Applications*, 6(02):1750008, 2017. MR-2760897

[7] D. Banerjee and Z. Ma. Optimal hypothesis testing for stochastic block models with growing degrees. *arXiv preprint arXiv:1705.05305*, 2017. MR-3656966

[8] J. Banks, C. Moore, J. Neeman, and P. Netrapalli. Information-theoretic thresholds for community detection in sparse networks. In *Conference on Learning Theory*, pages 383–416, 2016.

[9] 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.

[10] 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.

[11] C. Bordenave, M. Lelarge, and L. Massoulié. Non-backtracking spectrum of random graphs: community detection and non-regular Ramanujan graphs. *Ann. probab.*, To Appear.

[12] S. Bubeck, J. Ding, R. Eldan, and M. Z. Rácz. Testing for high-dimensional geometry in random graphs. *Random Structures & Algorithms*, 49(3):503–532, 2016. MR-3758726

[13] 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. MR-3545825

[14] T. Carleman. Les fonctions quasi analytiques(in French). Leçons professées au Collège de France. 1926. MR-0905164

[15] A. Coja-Oghlan. Graph partitioning via adaptive spectral techniques. *Combinatorics, Probability & Computing*, 19(2):227–284, 2010.

[16] A. Condon and R. M. Karp. Algorithms for graph partitioning on the planted partition model. *Random Structures & Algorithms*, 18(2):116–140, 2001. MR-2593622

[17] 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(6): 066106, Dec. 2011. MR-1809718

[18] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1): 1–38, 1977.

[19] M. E. Dyer and A. M. Frieze. The solution of some random np-hard problems in polynomial expected time. *J. Algorithms*, 10(4):451–489, Dec. 1989. MR-0501537

[20] L. Isserlis. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12(1/2):134–139, 1918. MR-1022107

[21] S. Janson. Random regular graphs: asymptotic distributions and contiguity. *Combin. Probab. Comput.*, 4(4):369–405, 1995.

[22] S. C. Johnson. Hierarchical clustering schemes. *Psychometrika*, 32(3):241–254, 1967. MR-1377557

[23] C. L. Mallows. A note on asymptotic joint normality. *Ann. Math. Statist.*, 43(2):508–515, 1972.

[24] L. Massoulié. Community detection thresholds and the weak ramanujan property. In *STOC 2014: 46th Annual Symposium on the Theory of Computing*, pages 1–10, New York, United States, June 2014. MR-0298812

[25] F. McSherry. Spectral partitioning of random graphs. In *42nd IEEE Symposium on Foundations of Computer Science*, pages 529–537, Oct 2001. MR-3238997

[26] E. Mossel, J. Neeman, and A. Sly. A Proof Of The Block Model Threshold Conjecture. *Combinatorica*, To Appear. MR-1948742

[27] E. Mossel, J. Neeman, and A. Sly. Reconstruction and estimation in the planted partition model. *Probab. Theory Related Fields*, 162(3–4):431–461, 2015.

[28] E. Mossel, J. Neeman, and A. Sly. Consistency thresholds for the planted bisection model. *Electron. J. Probab.*, 21:1–24, 2016. MR-3383334

[29] M. E. J. Newman, D. J. Watts, and S. H. Strogatz. Random graph models of social networks. *Proceedings of the National Academy of Sciences*, 99(suppl 1):2566–2572, 2002. MR-3485363

[30] J. K. Pritchard, M. Stephens, and P. Donnelly. Inference of population structure using multilocus genotype data. *Genetics*, 155(2):945–959, 2000.

[31] K. Rohe, S. Chatterjee, and B. Yu. Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.*, 39(4):1878–1915, 08 2011.

[32] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):888–905, Aug. 2000. MR-2893856

[33] M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis, and Machine Vision*. Thomson-Engineering, 2007.

[34] G. C. Wick. The evaluation of the collision matrix. *Phys. Rev.*, 80:268–272, Oct 1950.

[35] N. C. Wormald. Models of random regular graphs. In *Surveys in Combinatorics, 1999*, pages 239–298. Cambridge University Press, 1999. MR-0038281

7 Appendix

7.1 More general words and their equivalence classes

Here we only give a very brief description about the combinatorial aspects of random matrix theory required to prove Lemma 4.3. For more general information one should look at Chapter 1 of Anderson et al. [5] and Anderson and Zeiouni [4]. The definitions in this section have been taken from Anderson et al. [5] and Anderson and Zeitouni [4].

Definition 7.1. ($\mathcal{S}$ words) Given a set $\mathcal{S}$, an $\mathcal{S}$ letter s is simply an element of $\mathcal{S}$. An $\mathcal{S}$ word w is a finite sequence of letters $s_1 \dots s_n$, at least one letter long. An $\mathcal{S}$ word w is closed if its first and last letters are the same. Two $\mathcal{S}$ words w_1, w_2 are called equivalent, denoted $w_1 \sim w_2$, if there is a bijection on $\mathcal{S}$ that maps one into the other.

When $\mathcal{S} = \{1, \dots, N\}$ for some finite N, we use the term N word. Otherwise, if the set $\mathcal{S}$ is clear from the context, we refer to an $\mathcal{S}$ word simply as a word.

For any word $w = s_1 \dots s_k$, we use $l(w) = k$ to denote the length of w, define the weight $wt(w)$ as the number of distinct elements of the set $s_1, \dots, s_k$ and the support of w, denoted by $\mathrm{supp}(w)$, as the set of letters appearing in w. With any word w we may associate an undirected graph, with $wt(w)$ vertices and $l(w) - -1$ edges, as follows.

Definition 7.2. (Graph associated with a word) Given a word $w = s_1 \dots s_k$, we let $G_w = (V_w, E_w)$ be the graph with set of vertices $V_w = \mathrm{supp}(w)$ and (undirected) edges $E_w = \{\{s_i, s_i + 1\}, i = 1, \dots, k-1\}$.

The graph G_w is connected since the word w defines a path connecting all the vertices of G_w, which further starts and terminates at the same vertex if the word is closed. For $e \in E_w$, we use N_e^w to denote the number of times this path traverses the edge e (in any direction). We note that equivalent words generate the same graphs G_w (up to graph isomorphism) and the same passage-counts N_e^w.

Definition 7.3. (sentences and corresponding graphs) A sentence $a = [w_i]_{i=1}^n = [[\alpha_{i,j}]_{j=1}^{l(w_i)}]_{i=1}^n$ is an ordered collection of n words of length $(l(w_1), \dots, l(w_n))$ respectively. We define the graph $G_a = (V_a, E_a)$ to be the graph with

$$V_a = \mathrm{supp}(a), E_a = \{\{\alpha_{i,j}, \alpha_{i,j+1}\} | i = 1, \dots, n; j = 1, \dots, l(w_i) - 1\}\,.$$

Definition 7.4. (weak CLT sentences) A sentence $a = [w_i]_{i=1}^n$ is called a weak CLT sentence. If the following conditions are true:

1. All the words w_i's are closed.
2. Jointly the words w_i visit edge of G_a at least twice.
3. For each $i \in \{1, \dots, n\}$, there is another $j \neq i \in \{1, \dots, n\}$ such that G_{w_i} and G_{w_j} have at least one edge in common.

Note that these definitions are consistent with the ones given in Section 4. However, in Section 4, we defined these only for some specific cases required to solve the problem.

In order to prove Lemma 4.3, we require the following result from Anderson et al. [5].

Lemma 7.1. *(Lemma 2.1.23 in Anderson et al. [5]) Let $\mathcal{W}_{k,t}$ denote the equivalence classes corresponding to all closed words w of length $k+1$ with $\mathrm{wt}(w) = t$ such that each edge in G_w have been traversed at least twice. Then for $k > 2t-2$,*

$$\#\mathcal{W}_{k,t} \leq 2^k k^{3(k-2t+2)}.$$

Assuming Lemma 7.1 we now prove Lemma 4.3.

Proof of Lemma 4.3: Let $a = [w_i]_{i=1}^m$ be a weak CLT sentence such that G_a have $\mathcal{C}(a)$ many connected components. At first we introduce a partition $\eta(a)$ in the following way. We put i and j in same block of $\eta(a)$ if G_{w_i} and G_{w_j} share an edge. At first we fix such a partition η and consider all the sentences such that $\eta(a) = \eta$. Let $\mathcal{C}(\eta)$ be the number of blocks in η. It is easy to observe that for any a with $\eta(a) = \eta$, we have $\mathcal{C}(\eta) = \mathcal{C}(a)$. From now on we denote $\mathcal{C}(\eta)$ by $\mathcal{C}$ for convenience.

Let a be any weak CLT sentence such that $\eta(a) = \eta$. We now propose an algorithm to embed a into $\mathcal{C}$ ordered closed words $(W_1, \ldots, W_{\mathcal{C}})$ such that the equivalence class of each W_i belongs to $\mathcal{W}_{L_i,t_i}$ for some numbers L_i and t_i.

A similar type of argument can be found in Claim 3 of the proof of Theorem 2.2 in Banerjee and Bose(2017) [6].

An embedding algorithm: Let $B_1, \ldots, B_{\mathcal{C}}$ be the blocks of the partition η ordered in the following way. Let $m_i = \min\{j : j \in B_i\}$ and we order the blocks B_i such that $m_1 < m_2 \ldots < m_{\mathcal{C}}$. Given a partition η this ordering is unique. Let

$$B_i = \{i(1) < i(2) < \ldots < i(l(B_i))\}.$$

Here $l(B_i)$ denotes the number of elements in B_i.

For each B_i we embed the sentence $a_i = [w_{i(j)}]_{1 \leq j \leq l(B_i)}$ into W_i sequentially in the following manner.

1. Let $S_1 = \{i(1)\}$ and $\mathfrak{w}_1 = w_{i(1)}$.
2. For each $1 \leq c \leq l(B_i) - 1$ we perform the following.
   - Consider $\mathfrak{w}_c = (\alpha_{1,c}, \ldots, \alpha_{l(\mathfrak{w}_c),c})$ and $S_c \subset B_i$. Let $ne \in B_i \backslash S_c$ be the minimum index such that the following two conditions hold.
     (a) $G_{\mathfrak{w}_c}$ and $G_{w_{ne}}$ shares at least one edge $e = \{\alpha_{\kappa_1,c}, \alpha_{\kappa_1+1,c}\}$.
     (b) κ_1 is minimum among all such choices.
   - Let $w_{ne} = (\beta_{1,c}, \ldots, \beta_{l(w_{ne}),c})$ and $\{\beta_{\kappa_2,c}, \beta_{\kappa_2+1,c}\}$ be the first time e appears in w_{ne}. As $\{\beta_{\kappa_2,c}, \beta_{\kappa_2+1,c}\} = \{\alpha_{\kappa_1,c}, \alpha_{\kappa_1+1,c}\}$, $\alpha_{\kappa_1,c}$ is either equal to $\beta_{\kappa_2,c}$ or $\beta_{\kappa_2,c}$. Let $\kappa_3 \in \{\kappa_2, \kappa_2+1\}$ such that $\alpha_{\kappa_1,c} = \beta_{\kappa_3,c}$. If $\beta_{\kappa_2,c} = \beta_{\kappa_2+1,c}$, then we simply take $\kappa_3 = \kappa_2$.
   - We now generate $\mathfrak{w}_{c+1}$ in the following way

     $$\mathfrak{w}_{c+1} = (\alpha_{1,c}, \ldots, \alpha_{\kappa_1,c}, \beta_{\kappa_3+1,c}, \ldots, \beta_{l(w_{ne}),c}, \beta_{2,c}, \ldots, \beta_{\kappa_3,c}, \alpha_{\kappa_1+1,c}, \ldots, \alpha_{l(\mathfrak{w}_c),c}).$$

     Let $\tilde{a}_c := (\mathfrak{w}_c, w_{ne})$. It is easy to observe by induction that all $\mathfrak{w}_c$'s are closed words and so are all the w_{ne}'s. Also all the edges in the graph $G_{\tilde{a}_c}$ are preserved along with their passage counts in $G_{\mathfrak{w}_{c+1}}$.
   - Generate $S_{c+1} = S_c \cup \{ne\}$.
3. Return $W_i = \mathfrak{w}_{l(B_i)}$.

In the preceding algorithm we have actually defined a function f which maps any weak CLT sentence a into $\mathcal{C}$ ordered closed words $(W_1, \ldots, W_{\mathcal{C}})$ such that the equivalence class of each W_i belongs to $\mathcal{W}_{L_i,t_i}$ for some numbers L_i and t_i. Observe that given

two words $\mathfrak{w}_1$ and $\mathfrak{w}_2$, application of step 2 gives rise to a closed word $\mathfrak{w}_3$ where $l(\mathfrak{w}_3) = l(\mathfrak{w}_1) + l(\mathfrak{w}_2) - 1$. So

$$\begin{aligned} L_i = \sum_{j\in B_i} l(w_j) - (l(B_i)-1) < \sum_{j\in B_i} l(w_j). \\ \Rightarrow L_i + 1 \leq \sum_{j\in B_i} l(w_j) \\ \Rightarrow L_i + 1 - 2t_i \leq \sum_{j\in B_i} l(w_j) - 2t_i. \end{aligned} \tag{7.1}$$

Unfortunately f is not an injective map. So given $(W_1, \ldots, W_{\mathcal{C}})$ we find an upper bound to the cardinality of the following set

$$f^{-1}(W_1, \ldots, W_{\mathcal{C}}) := \{a | f(a) = (W_1, \ldots, W_{\mathcal{C}})\}$$

We have argued earlier $\mathcal{C}$ is the number of blocks in η. However, in general $(W_1, \ldots, W_{\mathcal{C}})$ does neither specify the partition η nor the order in which the words are concatenated with in each block B_i of η. So we fix a partition η with $\mathcal{C}$ many blocks and an order of concatenation $\mathcal{O}$. Observe that

$$\mathcal{O} = (\sigma_1(\eta), \ldots, \sigma_{\mathcal{C}}(\eta))$$

where for each i, $\sigma_i(\eta)$ is a permutation of the elements in B_i. Now we give an uniform upper bound to the cardinality of the following set

$$f^{-1}_{\eta,\mathcal{O}}(W_1, \ldots, W_{\mathcal{C}}) := \{a | \eta(a) = \eta \ ; \mathcal{O}(a) = \mathcal{O} \ \& f(a) = (W_1, \ldots, W_{\mathcal{C}})\}.$$

According to the algorithm any word W_i is formed by recursively applying step 2 to $(\mathfrak{w}_c, w_{ne})$ for $1 \leq c \leq l(B_i)$. Given a word $\mathfrak{w}_3 = (\alpha_1, \ldots, \alpha_{l(\mathfrak{w}_3)})$, we want to find out the number of two word sentences $(\mathfrak{w}_1, \mathfrak{w}_2)$ such that applying step 2 of the algorithm on $(\mathfrak{w}_1, \mathfrak{w}_2)$ gives $\mathfrak{w}_3$ as an output. This is equivalent to choose three positions $i_1 < i_2 < i_3$ from the set $\{1, \ldots, l(\mathfrak{w}_3)\}$ such that $\alpha_{i_1} = \alpha_{i_3}$. Once these three positions are chosen, $(\mathfrak{w}_1, \mathfrak{w}_2)$ can be constructed uniquely in the following manner

$$\begin{aligned} \mathfrak{w}_1 &= (\alpha_1, \ldots, \alpha_{i_1}, \alpha_{i_3+1}, \ldots, \alpha_{l(\mathfrak{w}_3)}) \\ \mathfrak{w}_2 &= (\alpha_{i_2}, \ldots, \alpha_{i_3}, \alpha_{i_1+1}, \ldots, \alpha_{i_2}). \end{aligned}$$

Total number of choices $i_1 < i_2 < i_3$ is bounded by $l(\mathfrak{w}_3)^3 \leq \left(\sum_{i=1}^m l(w_i)\right)^3$. For each block B_i, step 2 of the algorithm has been used $l(B_i)$ many times. So

$$f^{-1}_{\eta,\mathcal{O}}(W_1, \ldots, W_{\mathcal{C}}) \leq \left(\sum_{i=1}^m l(w_i)\right)^{3\sum_{i=1}^{\mathcal{C}} l(B_i)} = \left(\sum_{i=1}^m l(w_i)\right)^{3m}.$$

On the other hand, there are at most m^m many η's and for each η there are at most $\prod_{i=1}^{\mathcal{C}} l(B_i)! \leq m^m$ choices of $\mathcal{O}$. So

$$f^{-1}(W_1, \ldots, W_{\mathcal{C}}) \leq m^{2m} \left(\sum_{i=1}^m l(w_i)\right)^{3m} \leq \left(D_1 \sum_{i=1}^m l(w_i)\right)^{D_2 m} \tag{7.2}$$

for some known constants D_1 and D_2. Now we fix the sequence (L_i, t_i) and find an upper bound to the number of $(W_1, \ldots, W_{\mathcal{C}})$. From Lemma 7.1 we know the number of

choices of W_i is bounded by $2^{L_i-1}(L_i-1)^{L_i-2t_i+1}n^{t_i}$. So the total number of choices for $(W_1,\ldots,W_{\mathcal{C}})$ is bounded by

$$2^{\sum_{i=1}^{\mathcal{C}} L_i}\prod_{i=1}^{\mathcal{C}}(L_i-1)^{3(L_i-2t_i+1)}n^{t_i} \leq 2^{\sum_{i=1}^{m} l(w_i)}n^t\left(\sum_{i=1}^{m} l(w_i)\right)^{3(\sum_{i=1}^{m} l(w_i)-2t)}. \tag{7.3}$$

Now the number of choices (L_i,t_i) such that $\sum_{i=1}^{\mathcal{C}} L_i = \sum_{i=1}^{m} l(w_i) - \sum_{i=1}^{\mathcal{C}}(l(B_i)-1)$ and $\sum_{i=1}^{\mathcal{C}} t_i = t$ are bounded by

$$\binom{\sum_{i=1}^{m} l(w_i) - \sum_{i=1}^{\mathcal{C}}(l(B_i)-1)-1}{\mathcal{C}-1}\binom{t-1}{\mathcal{C}-1} \leq \binom{\sum_{i=1}^{m} l(w_i)-1}{\mathcal{C}-1}\binom{t-1}{\mathcal{C}-1} \leq \left(\sum_{i=1}^{m} l(w_i)\right)^{2m}. \tag{7.4}$$

Here the inequality follows since $\mathcal{C}\leq m$ and $t<\sum_{i=1}^{m}\frac{l(w_i)-1}{2}$. Finally we using the fact that $1\leq\mathcal{C}\leq m$ and combining (7.2), (7.3) and (7.4) we finally have

$$\begin{aligned}
&\#\mathcal{A} \leq \left(D_1\sum_{i=1}^{m} l(w_i)\right)^{D_2 m} \times 2^{\sum_{i=1}^{m} l(w_i)}n^t\left(\sum_{i=1}^{m} l(w_i)\right)^{3(\sum_{i=1}^{m} l(w_i)-2t)} \times \left(\sum_{i=1}^{m} l(w_i)\right)^{2m}\\
\Rightarrow &\#\mathcal{A} \leq 2^{\sum_i l(w_i)}\left(C_1\sum_i l(w_i)\right)^{C_2 m}\left(\sum_i l(w_i)\right)^{3(\sum_i l(w_i)-2t)} n^t
\end{aligned} \tag{7.5}$$

as required. □

Acknowledgments. The author thanks Elchanan Mossel and Zongming Ma for many useful discussions and their careful reading of the draft. He is grateful to Joe Neeman for useful discussions about non-reconstruction, Jian Ding for pointing out a small mistake in an earlier version of the draft and Cris Moore for pointing out an interesting reference. He thanks Adam Smith and Audra McMillan for their interest in this work and several useful discussions. Finally, he is grateful to both the anonymous referees for their careful reading of the draft, making several valuable comments and suggesting simplifications of many steps which have improved both the presentation and the technical aspects massively.