

ON IRREDUCIBLE CENTRAL LIMIT THEOREMS

FRANCESCO CARAVENNA, FRANCESCA COTTINI, AND GIOVANNI PECCATI

ABSTRACT. We consider sequences of homogeneous sums based on independent random variables and satisfying a central limit theorem (CLT). We address the following question:

“*In which cases is it not possible to reduce such an asymptotic result to the classical Lindeberg–Feller CLT through a restriction of the summation domain?*”

We provide several sufficient conditions for such irreducibility, expressed both in terms of (hyper)graphs Laplace eigenvalues, and of a certain notion of *combinatorial dimension*. Our analysis combines Cheeger-type inequalities with fourth moment theorems, showing that the irreducibility of a given CLT for homogeneous sums can be naturally encoded by the connectivity properties of the associated sequence of weighted hypergraphs. Several ad-hoc constructions are provided in the special case of quadratic forms.

CONTENTS

1. INTRODUCTION	2
1.1. Overview	2
1.2. Organization of the paper/tables	4
1.3. Acknowledgments	5
2. GENERAL SETTING AND MAIN RESULTS	5
2.1. Preliminaries	5
2.2. Irreducibility	6
2.3. Main results: spectral conditions	8
2.4. Main results: sparsity conditions	10
2.5. On necessary conditions for irreducibility	14
3. PRELIMINARIES, EXAMPLES AND APPLICATIONS	15
3.1. Universal fourth moment theorems	15
3.2. Reducible sequences verify a CLT	16
3.3. Examples of reducible sequences	17
3.4. Examples of irreducible sequences	19
4. IRREDUCIBILITY AND GRAPH SPECTRA: BACKGROUND AND PROOFS	23
4.1. Preliminaries: some connectivity estimates	23
4.2. Laplace spectra and irreducibility	27
4.3. Full irreducibility	28
5. IRREDUCIBILITY AND HYPERGRAPH SPECTRA	28
5.1. Hypergraphs and extended Cheeger’s inequalities	28
5.2. Irreducibility	31
6. IRREDUCIBILITY VIA COMBINATORIAL DIMENSIONS	34
6.1. Proof of Part (a) of Theorem 2.11	35

2020 *Mathematics Subject Classification*. 60F05; 60C05; 05C40.

Key words and phrases. Central Limit Theorem; Combinatorial Dimension; Cheeger Inequalities; Connectivity; Graphs; Homogeneous Sum; Hypergraphs; Irreducibility; Polynomial Chaos.

6.2. Proof of Part (b) of Theorem 2.11	36
7. IRREDUCIBILITY AND SPARSITY: AD HOC CONSTRUCTION FOR $d = 2$	38
7.1. Proof of Theorem 7.1-(1)	40
APPENDIX A. PROOFS OF SOME TECHNICAL RESULTS	43
A.1. Proof of Theorem 2.15	43
A.2. Proof of Theorem 4.8	44
A.3. Proof of Theorem 5.7	45
A.4. Proof of (6.7)	46
APPENDIX B. CARTESIAN PRODUCTS	47
REFERENCES	48

1. Introduction

1.1. Overview. This paper focuses on *homogeneous sums* — sometimes called *homogeneous polynomials* or *polynomial chaoses* — based on independent random variables, see (2.4) for a definition. These objects are quintessential examples of *degenerate U-statistics* [42, 50, 56, 63, 84], and play a pivotal role e.g. in the construction of *Wiener chaoses* over Gaussian or Poisson random measures [52, 69, 75, 78, 79], in the study of directed polymers [20–23], and in the analysis of Boolean (and more general) functions on discrete structures [28, 29, 45, 46, 59, 65, 77].

Over the past few decades, several central limit theorems (CLTs) have been established for sequences of homogeneous sums, ranging from the many ramifications of the fundamental contributions by P. de Jong [30–32, 37–39, 55] to the *universal CLTs* established in [70, 71] (see also [69, Chapter 11]), whose proofs combine the discrete Fourier techniques from [65, 82] with classical *fourth moment theorems* on Gaussian Wiener chaoses [69, 76].

Our goal in this paper is to investigate some previously unexplored aspects of these results. To this end, let us first recall the *universal de Jong CLT* established in [71, Theorem 1.10], formally stated in Theorem 3.1 below:

Let $\{Z_n(\mathbf{X}) : n \geq 1\}$ be a sequence of unit variance homogeneous sums, of fixed order $d \geq 2$ and based on a sequence of i.i.d. standard Gaussian random variables $\mathbf{X} = \{X_i\}$. Then, $Z_n(\mathbf{X})$ converges in distribution to a standard normal random variable $N \sim \mathcal{N}(0, 1)$ if and only if $\mathbb{E}[Z_n(\mathbf{X})^4] \rightarrow \mathbb{E}[N^4] = 3$. If $Z_n(\mathbf{X})$ converges in distribution to N , then this convergence continues to hold if \mathbf{X} is replaced by an arbitrary sequence of i.i.d. random variables $\{Y_i\}$ with mean zero and unit variance.[†]

The results from [71] have led to notable applications to directed polymers [21, 22], as well as to Salem–Zygmund CLTs for random polynomials [3], and to spectral fluctuations of non-Hermitian Gaussian random matrices [68]; see also [34, 55, 85].

[†]This statement is called a “de Jong CLT” since the characterization of a central limit theorem through a fourth-moment condition matches phenomena first identified in [30, 31]. We also recall that, since \mathbf{X} is a Gaussian family, the relation

$$Z_n(\mathbf{X}) \xrightarrow{d} N \iff \mathbb{E}[Z_n(\mathbf{X})^4] \rightarrow 3 \quad (1.1)$$

(where \xrightarrow{d} indicates convergence in distribution of random variables) follows from [76], together with the fact that each $Z_n(\mathbf{X})$ lies in the d th Wiener chaos associated with \mathbf{X} .

An intriguing observation made recently in [20] is that, in many cases relevant to the study of directed polymers and stochastic PDEs, the convergence in law of $Z_n(\mathbf{X})$ to a Gaussian distribution can be deduced directly from the classical Lindeberg-Feller CLT (see, e.g., [53, Theorem 4.12]) by restricting the summation domain of $Z_n(\mathbf{X})$ in such a way that *the homogeneous sum becomes a sum of independent random elements, up to a negligible remainder*. In this way, one can bypass the need to estimate higher-order moments or cumulants, such as those of order four. Throughout this paper, we will refer to such a simplified scenario as that of a *reducible CLT*.

The goal of our work is to understand to what extent this simplification applies, and in which cases it fails to hold. In particular, our results address the following question:

(Q1) *Assume that the homogeneous sums $\{Z_n(\mathbf{X})\}$ verify a CLT. Under which conditions is such a CLT irreducible, in the sense that it cannot be deduced from the classical Lindeberg–Feller theorem through a restriction of the summation domain?*

Motivated by Question **(Q1)**, we begin by introducing a rigorous notion of *(ir)reducibility* for homogeneous sums (see Definition 2.1). We then establish sufficient conditions for irreducibility and construct explicit families of sequences of homogeneous sums that satisfy *irreducible CLTs*—that is, they converge in distribution to a Gaussian limit, yet this convergence cannot be deduced by restricting the summation domain and applying the classical Lindeberg–Feller theorem.

The examples developed in this work are closely tied to the combinatorial structure of the summation sets E_n associated with the homogeneous sums $Z_n(\mathbf{X})$, with the property of irreducibility naturally arising from the connectivity features of E_n . Our main contributions, discussed in Section 2, rely in particular on tools from graph theory, including *Cheeger-type inequalities* [1, 2, 25, 60, 86] and their extensions to hypergraphs [7, 83]. We also present examples grounded in the concepts of *combinatorial dimension* and *fractional Cartesian products* [11–15, 17, 38, 71]. The inherent challenges involved in establishing *necessary conditions* for irreducibility are discussed in Subsection 2.5.

As illustrated below, our results complement and refine the recent characterizations of the asymptotic behaviour of homogeneous sums derived in [9, 49], and provide a novel perspective on the *fourth moment phenomenon* for chaotic random variables [69, 74, 76]. We observe that fourth moment theorems have recently played a crucial role in the derivation of CLTs for level sets of Gaussian waves [35, 64, 67, 73, 80], and have non-trivial counterparts in a noncommutative setting [33, 54]. See also I. Nourdin’s dedicated webpage for a comprehensive list of references around the fourth moment phenomenon.

Remark 1.1 (Spectral vs. domain reducibility for quadratic forms). The asymptotic normality of homogeneous sums of order 2 based on Gaussian variables can *always* be inferred from the Lindeberg–Feller theorem, via a spectral decomposition. Indeed, for a sequence $\{Z_n : n \geq 1\}$ in the *second Wiener chaos* of a separable Gaussian field [†], the following holds (see, e.g., [69, Proposition 2.7.11]):

(i) Each Z_n admits a spectral representation of the following form, for $N \in \mathbb{N} \cup \{+\infty\}$:

$$Z_n = \sum_{j=1}^N \gamma_{j,n} \xi_{j,n}, \quad (1.2)$$

[†]Recall that homogeneous sums of order 2 based on a i.i.d. standard Gaussian family \mathbf{X} are typical elements of the second Wiener chaos associated with \mathbf{X} (see [69, Sections 2.2 and 2.7.4]).

where $\{\xi_{j,n}\}$ are i.i.d. centered chi-square random variables with one degree of freedom and $\{\gamma_{j,n}\}$ are the eigenvalues of a suitable Hilbert–Schmidt operator.

- (ii) A direct application of the Lindeberg–Feller CLT yields the following equivalence: if $\mathbf{Var}(Z_n) \rightarrow 1$, then

$$Z_n \xrightarrow{d} N \sim \mathcal{N}(0, 1) \iff \max_{j \geq 1} |\gamma_{j,n}| \rightarrow 0. \quad (1.3)$$

The kind of reducibility studied in the present paper is of a stronger nature: we are interested in understanding whether a CLT for $\{Z_n\}$ can be deduced by the Lindeberg–Feller theorem via a *restriction of the summation domain*, rather than through implicit spectral conditions. As we shall demonstrate, the spectral criterion (1.3) does *not always* imply reducibility in our domain-based sense. For a direct application of criterion (1.3) to Gaussian quadratic forms with 0–1 coefficients, see Remark 3.2-(2). To the best of our knowledge, no analogue of the representation (1.2) exists for elements of Wiener chaoses of order $d \geq 3$. As a result, no systematic “spectral” reduction to the Lindeberg–Feller theorem is available for homogeneous sums of higher order; see also [49].

Remark 1.2 (Reducibility and U -statistics). The notions of reducibility and irreducibility introduced in Definition 2.1 below extend canonically from homogeneous sums to the broader class of degenerate U -statistics [31, 38, 40, 42, 56, 84]. This observation suggests a natural generalization of Question (Q1): one could ask which sequences of degenerate U -statistics satisfy a CLT that cannot be reduced to the Lindeberg–Feller framework. We view this as a distinct research direction, which we leave open for future investigation.

1.2. Organization of the paper/tables. In Section 2, we outline our general setting, introduce a rigorous definition of (ir)reducibility (Definition 2.1), and present some of our main results (Theorems 2.6, 2.11, and 2.13). We then discuss some of the challenges involved in identifying necessary conditions for irreducibility, see Subsection 2.5.

In Section 3, we present a wider discussion of our main results, illustrated with some explicit, non-trivial examples. The remaining sections are devoted to the proof of our results.

In Section 4, we prove our spectral criteria in the simplest setting of homogeneous sums of order $d = 2$, corresponding to *graphs*. After reviewing in Subsection 4.1 basic notions, normalized Laplacians, and related connectivity estimates (Cheeger’s inequalities), Subsection 4.2 contains the proof of Theorem 2.6, while Subsections 3.4 and 4.3 are devoted, respectively, to examples and to a strengthened notion of irreducibility.

Our spectral results are then extended in Section 5 to homogeneous sums of generic order $d \geq 2$, relying on an extension of Cheeger-type estimates to the setting of weighted hypergraphs (see Proposition 5.2 and Theorem 5.7).

Section 6 proves and illustrates Theorem 2.11, and Section 7 offers a detailed analysis of the ad-hoc construction for homogeneous sums of order 2 that appears in Theorem 2.13. Appendix A gathers additional proofs omitted in the main text, while Appendix B collects some preliminaries on Cartesian products of graphs.

For the reader’s convenience, all irreducible CLTs discussed in this paper are summarized in **Table A**, organized by the criterion used to establish irreducibility and the order of the corresponding homogeneous sums. Similarly, **Table B** displays a list of the reducible CLTs analyzed below, complete with their location.

From now on, every random object is assumed to be defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with \mathbb{E} denoting an expectation with respect to \mathbb{P} .

Criterion \ Order	$d = 2$	$d \geq 3$
<i>Spectral</i>	Theorem 2.6, Proposition 3.4 Examples (a), (b)-(i), (c), (d)-(ii), (d)-(iii) in Section 3.4	Theorem 5.7, Example 5.8, Example 5.9
<i>Combinatorial dimension</i>	Theorem 2.15	Theorem 2.11, Example 6.1, Proposition 6.2
<i>Ad-hoc construction</i>	Theorem 2.13, Section 7	—

Table A: Irreducible CLTs and where to find them.

Reducible CLT	Location
<i>Trivial</i>	Example 3.3-(a)
<i>Disjoint unions</i>	Example 3.3-(b)
<i>Variations of the Rook's graph</i>	Example 3.3-(b) and Section 3.4-(d)
<i>The (generalized) hypercube</i>	Example 3.3-(c) and Section 3.4-(b)

Table B: Reducible CLTs.

1.3. Acknowledgments. This research is supported by the Luxembourg National Research Fund (AFR/22/17170047/Bilateral-GRAALS). The first two authors acknowledge the support of INdAM/GNAMPA.

2. General setting and main results

2.1. Preliminaries. We consider a sequence of nonempty finite sets $\{V_n : n \geq 1\}$ with the property that

$$|V_n| \rightarrow \infty, \quad \text{as } n \rightarrow \infty. \quad (2.1)$$

We define $V := \bigcup_n V_n$ and let $\mathbf{X} = \{X_v : v \in V\}$ be a collection of i.i.d. standard normal random variables. For a fixed $d \geq 2$, we let $\{E_n : n \geq 1\}$ denote a sequence of sets such that

(1)

$$E_n \subset \underbrace{V_n \times \cdots \times V_n}_{d \text{ times}}, \quad n \geq 1, \quad \text{and} \quad |E_n| \rightarrow \infty;$$

(2) each E_n is *symmetric*: if $(v_1, \dots, v_d) \in E_n$, then $(v_{\sigma(1)}, \dots, v_{\sigma(d)}) \in E_n$ for all permutations σ of $[d] := \{1, \dots, d\}$;

(3) each E_n is *non-diagonal*: if $(v_1, \dots, v_d) \in E_n$, then $v_i \neq v_j$ for all $i \neq j$.

For every $n \geq 1$, we also consider a mapping

$$q_n : E_n \rightarrow \mathbb{R} : (v_1, \dots, v_d) \mapsto q_n(v_1, \dots, v_d),$$

that we assume to satisfy the following properties:

(a) q_n is symmetric: $q_n(v_1, \dots, v_d) = q_n(v_{\sigma(1)}, \dots, v_{\sigma(d)})$ for every permutation σ of $[d]$;

(b) as $n \rightarrow \infty$,

$$\sum_{(v_1, \dots, v_d) \in E_n} q_n(v_1, \dots, v_d)^2 := \|q_n\|^2 \rightarrow \infty. \quad (2.2)$$

Plainly, one can identify each q_n with a symmetric mapping on $V_n \times \cdots \times V_n$, with support equal to E_n . We will sometimes refer to the quantities $q_n(v_1, \dots, v_d)$ as the *coefficients* of the homogeneous sum, and to the squared coefficients

$$w_n(v_1, \dots, v_d) := q_n(v_1, \dots, v_d)^2 \quad (2.3)$$

as the *weights*.

As anticipated, in this work we will focus on sequences $\{Z_n : n \geq 1\}$ of *homogeneous sums* of order d , based on the Gaussian family \mathbf{X} and defined starting from the triplets (V_n, E_n, q_n) introduced above; for every $n \geq 1$, these are defined as follows:

$$Z_n = Z_n(\mathbf{X}) := \sum_{(v_1, \dots, v_d) \in E_n} q_n(v_1, \dots, v_d) \prod_{i=1}^d X_{v_i}. \quad (2.4)$$

One particularly interesting case considered below (studied, for instance, in [9] when $d = 2$) is that of constant coefficients, $q_n \equiv 1$. In this setting, the distribution of the random variable Z_n is entirely determined by the combinatorial structure of the set E_n .

Since the random variables X_v 's entering the definition (2.4) are independent, centered with unit variance, one has trivially that (using the notation (2.2))

$$\mathbb{E}[Z_n] = 0 \quad \text{and} \quad \mathbb{E}[Z_n^2] = d! \|q_n\|^2, \quad (2.5)$$

and we observe that, if $q_n^2 \equiv 1$ (constant weights), the second relation in (2.5) reduces to $\mathbb{E}[Z_n^2] = d! |E_n|$. We are particularly interested in those sequences of homogeneous sums satisfying a *central limit theorem* (CLT), that is, such that

$$\tilde{Z}_n = \tilde{Z}_n(\mathbf{X}) := \frac{Z_n(\mathbf{X})}{\sqrt{d! \|q_n\|^2}} \xrightarrow[n \rightarrow \infty]{d} N \sim \mathcal{N}(0, 1), \quad (2.6)$$

where \xrightarrow{d} indicates, as before, convergence in distribution.

It is by now a classical result (see the Introduction, as well as [69, Chapter 5 and Chapter 11]) that the convergence (2.6) can be established by a substantial simplification of the method of moments — via so-called *fourth moment theorems* — and has moreover a *universal nature* that is grounded in the polynomial chaos estimates from [65, 82]. For the reader's convenience, we discuss these results in Subsection 3.1 below (see, in particular, Theorem 3.1).

2.2. Irreducibility. Sometimes the CLT (2.6) can hold as a direct consequence of the classical *Lindeberg–Feller CLT for triangular arrays* [53, Theorem 4.12]. For instance, for $d \geq 1$ and $V = \mathbb{N}$, one easily verifies that the sequence

$$\tilde{Z}_n := \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} \prod_{\ell=1}^d X_{id+\ell} = \frac{(X_1 \cdots X_d) + (X_{d+1} \cdots X_{2d}) + \cdots + (X_{(n-1)d+1} \cdots X_{nd})}{\sqrt{n}} \quad (2.7)$$

with $\{X_i\}$ an i.i.d. standard Gaussian family, converges in distribution to a standard Gaussian law as a direct consequence of the usual CLT. As anticipated, the initial impetus for our work comes from reference [20], where the authors identified a class of highly non-trivial examples for which a *domain-wise* reduction of Theorem 3.1 to the Lindeberg–Feller setting remains possible. Their approach, motivated by applications to directed polymers, also extends to superpositions of multiple chaos orders, relies solely on second-moment computations, and provides natural criteria for proving Gaussian fluctuations in contexts such as singular stochastic PDEs [21].

As discussed in the Introduction, this paper aims to identify sequences of homogeneous sums whose Gaussian convergence *cannot* be deduced from a Lindeberg–Feller-type central limit theorem via a restriction of the summation domains; we call such sequences *irreducible*. Building on [20, Theorem 2.1], we now provide a rigorous definition of (ir)reducibility adapted to our setting.

Let the notation and assumptions of Section 2.1 prevail; for any subset $B \subset V_n$, we denote by $\sigma_n^2(B)$ the “contribution of B to the variance” defined by

$$\begin{aligned} \sigma_n^2(B) &:= d! \sum_{v_1, \dots, v_d \in B} \mathbf{1}_{E_n}(v_1, \dots, v_d) q_n(v_1, \dots, v_d)^2 \\ &= d! \sum_{\substack{(v_1, \dots, v_d) \in \\ E_n \cap (B \times \dots \times B)}} w_n(v_1, \dots, v_d), \end{aligned} \quad (2.8)$$

(where we have used the notation (2.3)) and observe that, when $w_n \equiv 1$, one has simply that

$$\sigma_n^2(B) = d! \left| E_n \cap \underbrace{(B \times \dots \times B)}_{d \text{ times}} \right|. \quad (2.9)$$

Definition 2.1 (Reducibility/Irreducibility). Fix $d \geq 2$ and consider the sequence of homogeneous sum $\{Z_n(\mathbf{X}) : n \geq 1\} = \{Z_n\}$ defined in (2.4). We say that $\{Z_n\}$ is reducible if for any $n \in \mathbb{N}$ there exist subsets (“boxes”) $B_1, \dots, B_{m_n} \subset V_n$ (where $B_i = B_i(n)$ may depend on n), such that $B_i \cap B_j = \emptyset$ for $i \neq j$ (that is, the boxes B_i are disjoint) and

- (i) $m_n \rightarrow \infty$;
- (ii) $\lim_{n \rightarrow \infty} \frac{1}{d! \|q_n\|^2} \sum_{i=1}^{m_n} \sigma_n^2(B_i) = 1$;
- (iii) $\lim_{n \rightarrow \infty} \frac{1}{d! \|q_n\|^2} \max_{i=1, \dots, m_n} \sigma_n^2(B_i) = 0$.

We say that $\{Z_n\}$ is irreducible if it is not possible to find disjoint sets $B_1, \dots, B_{m_n} \subset V_n$ such that (i), (ii) and (iii) hold. If $\{Z_n\}$ is irreducible and verifies (2.6), we say $\{Z_n\}$ verifies an irreducible CLT.

Remark 2.2. The fact that reducible sequences always verify a CLT was a key observation in [20]. For the sake of completeness, the argument is recalled in Subsection 3.2 below.

We refer to Section 3, see in particular Subsections 3.3 and 3.4, for a discussion of some explicit but non-trivial examples of reducible and irreducible sequences. Before proceeding, let us quickly compare the cases with constant vs. non-constant weights.

Remark 2.3. (1) (*Constant weights*) In the special case of *constant coefficients* $q_n \equiv 1$, or more generally constant weights $q_n^2 \equiv 1$, the conditions for irreducibility are simpler to state. Indeed, by virtue of (2.9), Properties (i), (ii) and (iii) of Definition 2.1 can be equivalently expressed in terms of $\{E_n\}$ as follows: as $n \rightarrow \infty$,

- (a) $m_n \rightarrow \infty$;
- (b) $\sum_{i=1}^{m_n} |E_n \cap (B_i \times \dots \times B_i)| \sim |E_n|$;
- (c) $\max_{i=1, \dots, m_n} |E_n \cap (B_i \times \dots \times B_i)| = o(|E_n|)$.

By extension, if a sequence $\{E_n\}$ verifies Properties (a)–(c), we say that $\{E_n\}$ is *reducible*.

- (2) (*From constant to non-constant weights*) Fix $d \geq 2$, set $q_n \equiv 1$, and consider a sequence $\{E_n\}$ verifying Properties (1)–(3), stated at the beginning of Section

2.1. Consider symmetric non-diagonal mappings $q_n : E_n \rightarrow \mathbb{R} : (v_1, \dots, v_d) \mapsto q_n(v_1, \dots, v_d)$ such that, for some $0 < \varepsilon < \eta < \infty$, one has that

$$\varepsilon \leq |q_n(v_1, \dots, v_d)| \leq \eta, \quad (v_1, \dots, v_d) \in E_n, \quad n \geq 1.$$

Then, it is easily seen that the sequence $\{E_n\}$ verifies Properties (a)—(c) of Point (1) if and only if the sequence $\{Z_n\}$ defined in (2.4) is irreducible in the sense of Definition 2.1. In view of this transfer principle, and despite a moderate loss of generality, we have chosen to center our analysis on the constant coefficient case throughout most of the paper—with the notable exception of Section 5.

In line with our initial objective, this paper aims to characterize those sequences $\{Z_n\}$ as in (2.4) that satisfy the CLT (2.6), but for which a domain-wise reduction is not possible. Our main contributions, presented in the forthcoming Subsections 2.3 and 2.4, revolve around two distinct sets of techniques:

- (I) *Spectral graph theory.* In the next Subsection 2.3, we will identify homogeneous sums associated with (hyper)graphs (possibly weighted) and use notable spectral estimates associated with their *Laplacian*—generally known as *Cheeger-type inequalities* [25]—in order to *characterize their irreducibility*. The proof of these results are presented in Sections 4 and 5. Our main references for spectral graph theory and Cheeger inequalities are the outstanding lecture notes by Luca Trevisan [87], as well as the classical texts [19, 26, 47]; see also [5, 51, 60–62]. The main tools allowing us to gauge the connectivity of hypergraphs are substantially inspired by the theory developed in [7, 83].
- (II) *Combinatorial dimensions and sparsity.* In Subsection 2.4, we will exploit the notion of *combinatorial dimension* (following the framework developed in [11, 13], see Definition 2.8) in order to derive a *general irreducibility* criterion for homogeneous sums whose support sets $\{E_n\}$ exhibit a *sparse structure*. The proof is presented in Section 6. These constructions build on ideas from [12, 14, 15, 17, 38, 71], and represent a departure from the spectral approaches. In the case $d = 2$, where no canonical notion of fractional product exists [14, 17], we study in Section 7 an explicit sparse construction yielding irreducible homogeneous sums (Theorem 2.13) that verify a CLT.

2.3. Main results: spectral conditions. We present our main results linking irreducibility to spectral properties. For ease of exposition, we first consider the simpler setting of *graphs*, then we discuss the case of *hypergraphs*.

2.3.1. **GRAPHS.** Let us consider the case of homogeneous sums of order $d = 2$ with $q_n \equiv 1$, in such a way that Z_n in (2.4) boils down to a quadratic form with 0-1 coefficients:

$$Z_n = \sum_{v_1, v_2 \in V_n} \mathbb{1}_{E_n}(v_1, v_2) X_{v_1} X_{v_2}, \quad (2.10)$$

where $\{E_n\}$ satisfy the requirements (1)–(3), as stated at the beginning of Section 2.1.

We associate with Z_n in (2.10) the finite undirected graph $G_n = (V_n, \mathcal{E}_n)$ with adjacency matrix $A_n(v, w) := \mathbb{1}_{E_n}(v, w)$, that is \mathcal{E}_n consists of unordered pairs $\{v, w\}$ with $(v, w) \in E_n$. Letting $\{v_1, \dots, v_{N_n}\}$ be an enumeration of V_n , we let D_n be the diagonal matrix with entries $d(v_1), \dots, d(v_{N_n})$, i.e. the degrees of each vertex. We also consider the *normalized*

Laplacian matrix $\mathcal{L}_n := I_n - D_n^{-1/2} A_n D_n^{-1/2}$ associated with G_n , whose eigenvalues are noted $0 = \mu_1^{(n)} \leq \dots \leq \mu_{N_n}^{(n)} \leq 2$.

Convention 2.4. Unless otherwise specified, every graph considered in the present and forthcoming sections is implicitly assumed to be undirected, simple (i.e. containing neither loops nor multiple edges), and with no isolated vertices.

Definition 2.5. Let $G_n = (V_n, \mathcal{E}_n)$, $n \geq 1$, be a sequence of graphs such that $|V_n|, |\mathcal{E}_n| \rightarrow \infty$, and write A_n for the adjacency matrix of G_n . Consider the sequence $\{Z_n\}$ defined as in (2.10) by identifying A_n with the indicator $\mathbb{1}_{E_n}$ of a symmetric and non-diagonal set $E_n \subset V_n \times V_n$. We refer to $\{Z_n\}$ as the *sequence of homogeneous sums generated by $\{G_n\}$* . We say that $\{G_n\}$ *generates an irreducible CLT*, if $\{Z_n\}$ verifies an irreducible CLT in the sense of Definition 2.1 — that is, if $\{Z_n\}$ is irreducible and

$$\tilde{Z}_n := \frac{Z_n}{\sqrt{2|E_n|}} \xrightarrow{d} N \sim \mathcal{N}(0, 1). \quad (2.11)$$

By extension, if the sequence $\{Z_n\}$ is reducible, we will say that $\{G_n\}$ *is reducible* or, more precisely, that $\{G_n\}$ *generates a reducible CLT*. Similar notions for hypergraphs and associated sequences of homogeneous sums will be introduced in Definition 5.5.

As discussed in Section 4, it is a classical fact that, for $k \geq 2$, the eigenvalue $\mu_k^{(n)}$ is closely related to the connectivity properties of the graph G_n generating Z_n , notably through (multiway) Cheeger-type inequalities [60, 61, 87]. Our main result in this setting, proved in Section 4, shows that these connectivity properties can be naturally related to the notions of reducibility and irreducibility introduced in Definition 2.1.

Theorem 2.6. *Let Z_n be as in (2.10), and let $G_n = (V_n, \mathcal{E}_n)$ be its associated graph. Suppose that there exists $k \geq 2$ such that, as $n \rightarrow \infty$,*

$$\liminf_n \mu_k^{(n)} > 0. \quad (2.12)$$

Then $\{Z_n\}$ is irreducible, in the sense of Definition 2.1.

A large collection of irreducible CLTs for quadratic forms of the type (2.10) — spanning in particular sequences $\{Z_n\}$ generated by *expanders* and by generic Cartesian products of regular graphs — is discussed in Subsection 3.4.

Remark 2.7 (Full irreducibility). Theorem 2.6 implies a stronger form of irreducibility, that we call *full irreducibility*, see Subsection 4.3.

2.3.2. HYPERGRAPHS. We next consider generic homogeneous sums of order $d \geq 2$, as defined in (2.4) (where the real coefficients $q_n(v_1, \dots, v_d)$ are arbitrary). In Section 5, we will show that the content of Theorem 2.6 can be extended to this general setting. To this end, we associate with each Z_n a finite *weighted hypergraph* $\mathcal{G}_n = (V_n, \mathcal{E}_n, w_n)$. Here, each *hyperedge* $e = \{v_1, \dots, v_d\} \in \mathcal{E}_n$ is a subset of the vertices V_n having cardinality d . Each hyperedge is also paired with the weight $w_n(v_1, \dots, v_d) := q_n(v_1, \dots, v_d)^2$ introduced above. In this context, it is therefore natural to introduce a general definition of the normalized Laplacian \mathcal{L}_n associated with \mathcal{G}_n and obtain an analogous version of Theorem 2.6 (see the discussion contained in Section 5, as well as Theorem 5.7 therein).

We refer to Section 5 for a precise statement of our results. As anticipated, we rely on an extension of Cheeger-type estimates for graphs, directly inspired by [7, 83].

2.4. Main results: sparsity conditions. We next present our main results which link irreducibility to a notion of sparsity, encoded by the concept of *combinatorial dimension*.

2.4.1. GENERAL RESULTS. The following definition — which was originally motivated by problems in harmonic analysis [10, 16] — is lifted from [11, 13, 15] and is meant to characterize those subsets of Cartesian products displaying a certain form of *sparsity*, resulting in the fact that they behave, in a precise sense, like Cartesian products of lower (and possibly non-integer) order.

Definition 2.8 (Combinatorial dimension). Let $\{V_n\}_{n \geq 1}$ be sets with $|V_n| \rightarrow \infty$. Fix an integer $d \geq 2$, as well as a real number $\alpha \in (0, d]$. We consider subsets

$$J_n \subseteq \underbrace{V_n \times \cdots \times V_n}_{d \text{ times}}, \quad n \geq 1,$$

(each J_n is not necessarily symmetric and has possibly diagonal components) and say that the sequence $\{J_n\}$ has *combinatorial dimension* equal to α if there exist finite constants $0 < c' \leq c$ such that the following two properties are verified for n sufficiently large:

- for all $A_1, \dots, A_d \subseteq V_n$,

$$|J_n \cap (A_1 \times \cdots \times A_d)| \leq c \max_{i=1, \dots, d} |A_i|^\alpha; \quad (2.13)$$

- moreover,

$$|J_n| \geq c' |V_n|^\alpha, \quad n \geq 1. \quad (2.14)$$

Remark 2.9. (1) For $d \geq 2$, the existence of sequences $\{J_n\}$ with arbitrary combinatorial dimension $\alpha \in [1, d]$ is established in [14, 17] through a random construction (see Subsection 2.4.3 for details).

(2) If $\{J_n\}$ has combinatorial dimension $\alpha > 0$, then as $n \rightarrow \infty$

$$|J_n| \asymp |V_n|^\alpha, \quad \text{i.e.} \quad c_1 |V_n|^\alpha \leq |J_n| \leq c_2 |V_n|^\alpha \quad (2.15)$$

for some $0 < c_1 \leq c_2 < \infty$. In particular, when $\alpha < d$, then $|J_n| = o(|V_n|^d)$; more generally, for every $\{A^{(n)}\}$ such that $A^{(n)} \subset V_n$ and $|A^{(n)}| \rightarrow \infty$, one has that

$$|J_n \cap (A^{(n)} \times \cdots \times A^{(n)})| \leq c |A^{(n)}|^\alpha = o(|A^{(n)}|^d).$$

Such an estimate aligns with our heuristic characterisation of non-trivial combinatorial dimensions as signatures of sparsity.

(3) The combinatorial dimension of a given sequence of sets $\{J_n\}$ is not always well-defined, even when the cardinality $|J_n|$ grows as some power of $|V_n|$. To see this, consider the case where $J_n = W_n^d$, with $W_n \subset V_n$ such that $|W_n| \asymp |V_n|^\beta$ for some $\beta < 1$. Then $|J_n| = |J_n \cap (W_n \times \cdots \times W_n)| = |W_n|^d \asymp |V_n|^{\beta d}$ and these relations are not compatible with both requirements (2.13) and (2.14) (which would entail respectively $d \leq \alpha$ and $\alpha \leq \beta d$).

Let us provide some basic examples.

Example 2.10. (a) (*Full Cartesian products*) The sequence $J_n = V_n \times \cdots \times V_n$, $n \geq 1$, has a trivial full combinatorial dimension equal to d , since conditions (2.13) and (2.14) are satisfied with $c = c' = 1$.

- (b) (*Sets with diagonal restrictions*) Fix $d \geq 2$ and consider a partition $\pi = \{b_1, \dots, b_{|\pi|}\}$ of $[d]$, where $|\pi| \in [d]$ is the number of blocks of π . For every $n \geq 1$, we define V_n^π to be the subset of the d -th Cartesian product of V_n composed of those vectors (v_1, \dots, v_d) such that $v_i = v_j$ if and only if i and j belong to the same block b of π . One easily sees that the sequence $\{V_n^\pi\}$ has combinatorial dimension $|\pi|$. Two straightforward cases correspond to the choices $\pi = \{[d]\}$ (fully diagonal set, in which case the combinatorial dimension is 1) and $\pi = \{\{1\}, \dots, \{d\}\}$ (fully non-diagonal set, in which case the combinatorial dimension is d). It is readily seen that, when $|\pi| < d$, the sequence $\{V_n^\pi\}$ cannot satisfy both Properties (2) (symmetry) and (3) (non-diagonal structure) from the beginning of Section 2.1. This prevents the use of such sets in constructing examples of irreducible CLTs.
- (c) (*Regular graphs*) For $d = 2$, we consider a sequence $\{E_n\}$ satisfying the properties (1)–(3) listed at the beginning of Section 2.1. We also fix $m \geq 1$, and assume that the mappings $(v, w) \mapsto \mathbb{1}_{E_n}(v, w)$, $v, w \in V_n$, correspond to the adjacency matrices of a sequence of m -regular graphs. Then, one has that $|E_n| = m|V_n|$, and a direct application of the *expander mixing lemma* (see e.g. [51, Lemma 2.5]) yields that, for all $n \geq 1$ and all $A_1, A_2 \subseteq V_n$,

$$|E_n \cap (A_1 \times A_2)| \leq O(m) \max\{|A_1|, |A_2|\},$$

implying that the sequence $\{E_n\}$ has combinatorial dimension 1.

The following result provides a sufficient condition for the irreducibility of a symmetric and non-diagonal sequence $\{E_n\}$, formulated in terms of its combinatorial dimension. The second part of the statement also establishes the existence of infinitely many non-trivial sequences $\{Z_n\}$ of homogeneous sums of order $d \geq 3$ that satisfy an irreducible CLT. (The case $d = 2$ will be studied in Theorem 2.15 through the use of a random construction.)

Theorem 2.11. *For $d \geq 2$, let $\{E_n\}$ satisfy Properties (1)–(3), as listed at the beginning of Section 2.1.*

- (a) *Assume that $\{E_n\}$ has combinatorial dimension*

$$1 < \alpha \leq d.$$

Then, $\{E_n\}$ is irreducible — recall Remark 2.3-(1).

- (b) *For every $d \geq 3$ and $b = 2, \dots, d-1$, there exists a sequence $\{E_n\}$ with combinatorial dimension*

$$\alpha = \frac{d}{b} \in (1, d),$$

such that the corresponding sequence $\{Z_n\}$ of homogeneous sums, defined as in (2.4) for $q_n \equiv 1$, satisfies an irreducible CLT.

We will prove Theorem 2.11 in Section 6. The construction of the irreducible sequences $\{E_n\}$ featured in the second part of the statement is detailed in Example 6.1: for this, we will take $V = \mathbb{N}^b$, $V_n = \{1, \dots, n\}^b$, and E_n with a structure close to that of a *fractional Cartesian product*, see [11–13, 15, 36, 71]. The exact computation of the combinatorial dimension of $\{E_n\}$ exploits a discrete Brascamp-Lieb-type inequality due to Fissler [43], while the asymptotic normality of the homogeneous sums follows from arguments already rehearsed in [71].

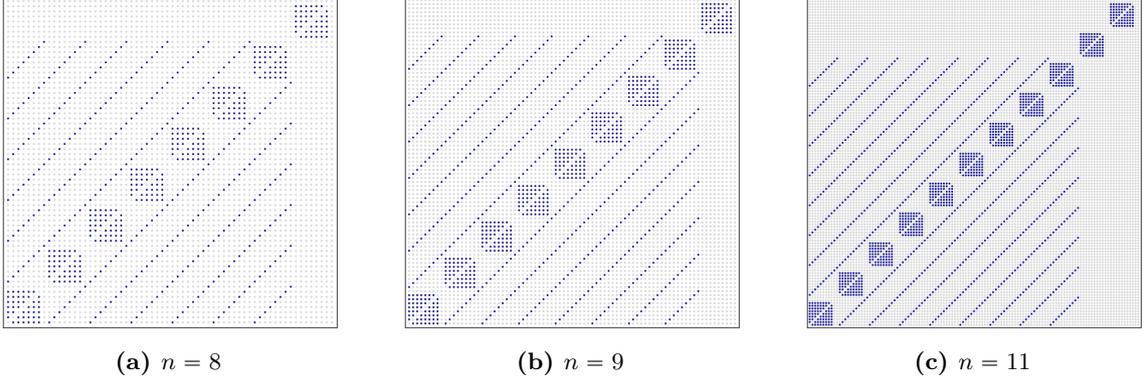


Figure 1. A visualization of the set E_n defined in (2.18) for $\beta = 0.9$ and different values of n , when $\mathbb{S}_v(a) = \{a\} \times \{1, \dots, \lfloor \beta n \rfloor\}$ and $\mathbb{S}_h(b) = \{1, \dots, \lfloor \beta n \rfloor\} \times \{b\}$. Each element of $V_n = \{(i, j) : i, j \in [n]\}$ has been identified with an element of $[n^2]$ by ordering V_n according to the lexicographic order. Pairs (v_1, v_2) not belonging to E_n are represented as grey dots. The fact that $\{E_n\}$ is irreducible implies that one cannot discard the diagonal bands framing the central blue tiles without drastically modifying the asymptotic behavior of $\{Z_n\}$.

2.4.2. AN AD-HOC CONSTRUCTION IN DIMENSION 2. It is a well-known fact (discussed e.g. in [14, 17]; see also Subsection 2.4.3 below) that there is no canonical notion of fractional Cartesian products in dimension $d = 2$. This explains why the second part of Theorem 2.11 does not address the two-dimensional case.

To compensate for this limitation (and to probe the sharpness of our results), we study a class of irreducible sequences of quadratic forms Z_n of the type (2.10), constructed to satisfy the following properties: (i) the combinatorial dimension of the corresponding sequence $\{E_n\}$ is not defined (see Remark 2.12); (ii) the spectra of the associated normalized Laplacians appear to be analytically intractable in their most general form.

Our explicit construction is realized as follows. For any $n \in \mathbb{N}$, we set

$$V_n = \{1, \dots, n\} \times \{1, \dots, n\} = [n]^2.$$

Fix $\beta \in (0, 1]$ and, for each $a, b \in \{1, \dots, n\}$, consider subsets $\mathbb{S}_v(a) \subseteq \{a\} \times \{1, \dots, n\}$ and $\mathbb{S}_h(b) \subseteq \{1, \dots, n\} \times \{b\}$ (the labels v, h stand for *vertical* and *horizontal*) with

$$|\mathbb{S}_v(a)| = |\mathbb{S}_h(b)| = \lfloor \beta n \rfloor. \quad (2.16)$$

We introduce the following equivalence relations on V_n : for any $v_1, v_2 \in V_n$,

$$\begin{aligned} v_1 \stackrel{h}{\sim} v_2 &\iff \text{for some } b \text{ one has } v_1, v_2 \in \mathbb{S}_h(b) \text{ with } v_1 \neq v_2, \\ v_1 \stackrel{v}{\sim} v_2 &\iff \text{for some } a \text{ one has } v_1, v_2 \in \mathbb{S}_v(a) \text{ with } v_1 \neq v_2. \end{aligned} \quad (2.17)$$

We define the sequence $\{Z_n\}$ according to (2.10), where E_n has the form

$$E_n = \{(v_1, v_2) \in V_n \times V_n : \text{either } v_1 \stackrel{h}{\sim} v_2 \text{ or } v_1 \stackrel{v}{\sim} v_2\}. \quad (2.18)$$

In words, $v_1 \stackrel{h}{\sim} v_2$ means that v_1, v_2 are on the same horizontal line within a set $\mathbb{S}_h(\cdot)$, and similarly for $v_1 \stackrel{v}{\sim} v_2$. We refer to Figure 1 for some graphical representations of E_n .

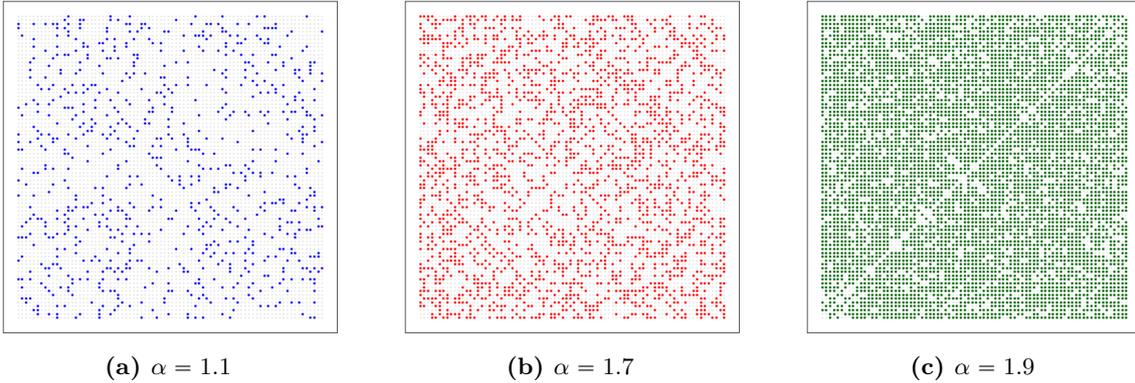


Figure 2. Three realisations of the (symmetric and non-diagonal) random point configuration $\tilde{\mathbf{E}}_n(\alpha; 2)$, as defined in Section 2.4.3, for $n = 80$ and different values of $\alpha \in (1, 2)$.

Remark 2.12 ($\{E_n\}$ has no combinatorial dimension). As discussed in Section 7, one has that $|E_n| = 2n \binom{\lfloor \beta n \rfloor}{2} \sim \beta^2 n^3 \asymp |V_n|^{3/2}$ as $n \rightarrow \infty$. As anticipated, for every $\beta \in (0, 1]$ the sequence $\{E_n\}$ has no definite combinatorial dimension, because for every n and $a \in [n]$

$$|E_n \cap (\mathbb{S}_v(a) \times \mathbb{S}_v(a))| = |\mathbb{S}_v(a)| \cdot (|\mathbb{S}_v(a)| - 1) \sim |\mathbb{S}_v(a)|^2$$

hence the two relations (2.13) and (2.14) cannot be fulfilled for the same α .

The following result proves that, for $\beta > \frac{1}{2}$, the sequence $\{Z_n\}$ is irreducible, in fact it verifies an irreducible CLT.

Theorem 2.13. *Let the above notation and conventions prevail, and fix*

$$\frac{1}{2} < \beta \leq 1.$$

Then, $\{Z_n\}$ verifies an irreducible CLT, in the sense of Definition 2.1.

We prove this result in Section 7, where we actually obtain a strengthened version of Theorem 2.13 (see, in particular, Theorem 7.1). We refer to Subsection 3.4 — see in particular Example (d) — for a discussion of the irreducibility of $\{Z_n\}$ in some special cases.

2.4.3. RANDOM CONSTRUCTIONS. We now present a random construction from [14, 17], yielding for every $d \geq 2$ the existence of a sequence $\{E_n\}$ with arbitrary combinatorial dimension $\alpha \in (1, d)$. To state this result, we fix d and α as above and, for every $n \geq 1$, we consider an array

$$H^{(n)} = \left\{ H_{j_1, \dots, j_d}^{(n)} : (j_1, \dots, j_d) \in [n]^d \right\} \quad (2.19)$$

of i.i.d. Bernoulli random variables with parameter $n^{\alpha-d}$. We define the random sets

$$\mathbf{E}_n(\alpha; d) := \left\{ (j_1, \dots, j_d) \in [n]^d : H_{j_1, \dots, j_d}^{(n)} = 1 \right\}, \quad n \geq 1, \quad (2.20)$$

and, for every pair of constants $0 < \varepsilon < \eta$, we define the event $C_n(\varepsilon, \eta; \alpha, d)$ to be the set of all $\omega \in \Omega$ such that $|\mathbf{E}_n(\alpha; d)(\omega)| \geq \varepsilon n^\alpha$ and $|\mathbf{E}_n(\alpha; d)(\omega) \cap (A_1 \times \dots \times A_d)| \leq \eta \max_i |A_i|^\alpha$, for all $A_1, \dots, A_d \subseteq [n]$.

Theorem 2.14 (Lemma 2 in [14]). *Let the above notation and assumptions prevail. Then, for every $d \geq 2$ and every $\alpha \in (1, d)$ there exist constants $0 < \varepsilon_0 < \eta_0$ (depending on d, α) such that*

$$\lim_{n \rightarrow \infty} \mathbb{P} \{C_n(\varepsilon_0, \eta_0; \alpha, d)\} = 1.$$

It is easily seen that, for $d = 2$, the conclusion of Theorem 2.14 continues to hold if one replaces the set $\mathbf{E}_n(\alpha; 2)$ with the symmetrized and non-diagonal set $\tilde{\mathbf{E}}_n(\alpha; 2)$, obtained by modifying (2.20) as follows: (a) define $H_{j_1, j_1}^{(n)} = 0$ for all $j_1 \in [n]$, and (b) force symmetry by setting $H_{j_1, j_2}^{(n)} = H_{j_2, j_1}^{(n)}$, for all $1 \leq j_1 < j_2 \leq n$. In this way, the array

$$\left\{ \mathbf{1}_{\tilde{\mathbf{E}}_n(\alpha; 2)}(i, j) \right\} = \left\{ H_{i, j}^{(n)} \right\}$$

coincides with the (random) adjacency matrix of the Erdős-Renyi random graph $G(n, p_n)$, with $p_n = n^{\alpha-2}$; see e.g. [18]. Figure 2 displays several realizations of the random set $\tilde{\mathbf{E}}_n(\alpha; 2)$, for $n = 80$ and different values of α .

The next statement complements Theorem 2.11, by showing the existence of irreducible CLTs for homogeneous sums of order 2, that are associated with sequences $\{E_n\}$ of arbitrary combinatorial dimension $\alpha \in (1, 2)$. The (simple) proof is presented in Section A.1, and uses the probabilistic method by exploiting the content of Theorem 2.14 in the case $d = 2$. More refined results (immaterial for our statement) can be obtained by following the strategy outlined in [9, Example 2.3], exploiting the classical estimates from [57].

Theorem 2.15. *Fix $d = 2$ and $\alpha \in (1, 2)$. Then, there exists a sequence $\{E_n\}$ such that: (i) $\{E_n\}$ satisfies Properties (1)–(3), as listed at the beginning of Section 2.1, (ii) $\{E_n\}$ has combinatorial dimension α ; (iii) the associated sequence of homogeneous sums $\{Z_n\}$, defined as in (2.10), satisfies the irreducible CLT (2.11).*

The next question is left open for further research, and could be in principle attacked by refining the study of the sets $\mathbf{E}(\alpha; d)$ introduced in the present section.

(Q2) *Fix $d \geq 3$ and $\alpha \in (1, d)$. Is there a sequence $\{Z_n\}$ as in (2.4) such that:*

- (a) $q_n \equiv 1$,
- (b) $\{E_n\}$ has combinatorial dimension α ,
- (c) $\{\tilde{Z}_n\}$ verifies the (irreducible) CLT (2.6)?

Note that Theorem 2.11-(b) provides a positive answer to Question **(Q2)** for all $d \geq 3$ and $\alpha = d/b$, $b = 2, \dots, d - 1$.

2.5. On necessary conditions for irreducibility. The main contribution of this paper is the identification of several sufficient conditions for irreducibility, stemming from both spectral (Theorem 2.6 and its extensions to hypergraphs) and combinatorial (Theorems 2.11 and 2.13) criteria. By contrast, we have not been able to establish corresponding necessary conditions. This limitation is briefly discussed in the following list.

- (i) The spectral condition (2.12) is *not* necessary for irreducibility, as shown, for instance, in Example (d) of Section 3.4: one can construct sequences of irreducible homogeneous sums whose Laplace eigenvalues vanish at every order. We conjecture that necessary spectral conditions for irreducibility might be obtained by extending

“hard” Cheeger-type inequalities (such as those presented in the forthcoming Theorem 4.4) to incorporate constraints on the relative sizes of the blocks S_1, \dots, S_k . Some progress in this direction can be found, for instance, in [4].

- (ii) Similarly, Example (d) in Section 3.4 and Example 2.10 will show, respectively, that (a) irreducible CLTs may also arise in cases where the combinatorial dimension is not well-defined, and (b) having a combinatorial dimension strictly greater than 1 is not necessary for irreducibility, since there exist sequences of regular graphs with fixed degree that generate both irreducible and reducible CLTs (note that sequences of regular graphs with a fixed degree are always associated with summation domains having a combinatorial dimension equal to one; see Example 2.10-(c)). At the time of writing, we are not aware of any additional combinatorial characterizations of discrete symmetric sets that could help close this gap.

Remark 2.16. An alternative approach to studying the CLTs considered in the present paper is based on the use of *dependency graphs* and Stein’s method; see, for instance, the classical reference [6]. In the context of 0-1 quadratic forms as in (2.10), this would entail addressing questions of reducibility and irreducibility by means of the so-called *line graph* associated with G_n , that is, the graph whose vertices correspond to the edges of G_n and where two edges are connected if and only if they are adjacent (see e.g. [19, Section 1.4.5]). While some preliminary computations have shown that this approach may lead to suboptimal results, we leave this direction open for future investigation.

3. Preliminaries, examples and applications

In this section, we first recall some (by now) classical conditions which ensure the validity of the Central Limit Theorem for homogeneous sums. We then provide examples and applications which illustrate our setting and our main results.

3.1. Universal fourth moment theorems. The following result, based on the material discussed in [69, Chapter 5 and Chapter 11], corresponds to the *universal de Jong theorem* evoked in the Introduction.

Theorem 3.1 (Theorem 1.10 in [71]). *Let the assumptions and notation of Section 2.1 prevail. Then, the following three properties are equivalent, as $n \rightarrow \infty$:*

- (i) *The CLT (2.6) holds;*
- (ii) $\mathbb{E}[\tilde{Z}_n(\mathbf{X})^4] \rightarrow 3$;
- (iii) *For any collection $\mathbf{Y} = \{Y_v : v \in V\}$ of i.i.d. random variables with zero mean and unit variance,*

$$\tilde{Z}_n(\mathbf{Y}) := \frac{Z_n(\mathbf{Y})}{\sqrt{d! \|q_n\|^2}} \xrightarrow{d} N \sim \mathcal{N}(0, 1),$$

where $Z_n(\mathbf{Y})$ is defined as in (2.4).

- (iv) *For any collection $\mathbf{U} = \{U_v : v \in V\}$ of independent random variables with zero mean, unit variance and such that $\sup_{v \in V} \mathbb{E}[|U_v|^p] < \infty$ for some $p > 2$,*

$$\tilde{Z}_n(\mathbf{U}) \xrightarrow{d} N \sim \mathcal{N}(0, 1),$$

where the notation is analogous to that introduced at Point (iii).

Remark 3.2. (1) Under the above notation and assumptions, define \tilde{q}_n to be the mapping on $V_n \times \cdots \times V_n$ such that $\tilde{q}_n(v_1, \dots, v_d) = q_n(v_1, \dots, v_d) / \sqrt{d! \|q_n\|^2}$ if $(v_1, \dots, v_d) \in E_n$, and $\tilde{q}_n(v_1, \dots, v_d) = 0$ otherwise. For every $r = 1, \dots, d$, define the *contraction*

$$\begin{aligned} \tilde{q}_n \star_r \tilde{q}_n &: \underbrace{V_n \times \cdots \times V_n}_{2d-2r \text{ times}} \rightarrow \mathbb{R} \\ &: (v_1, \dots, v_{2d-2r}) \mapsto \sum_{a_1, \dots, a_r \in V_n} \tilde{q}_n(a_1, \dots, a_r, v_1, \dots, v_r) \tilde{q}_n(a_1, \dots, a_r, v_{r+1}, \dots, v_{2d-2r}) \\ &:= \tilde{q}_n \star_r \tilde{q}_n(v_1, \dots, v_{2d-2r}). \end{aligned}$$

Then, [71, Proposition 1.6] implies that Condition (ii) in Theorem 3.1 is verified if and only if, for all $r = 1, \dots, d-1$,

$$\sum_{v_1, \dots, v_{2d-2r} \in V_n} \tilde{q}_n \star_r \tilde{q}_n(v_1, \dots, v_{2d-2r})^2 \longrightarrow 0. \quad (3.1)$$

(2) For $d = 2$, consider the case of constant coefficients $q_n \equiv 1$ and, for $n \geq 1$, denote by $\lambda_{n,1}, \dots, \lambda_{n,|V_n|}$ the eigenvalues of the matrix $\{A_n(v, u) : u, v \in V_n\}$ such that $A_n(u, v)$ equals one or zero according as $(u, v) \in E_n$ or not (that is, A_n is the *adjacency matrix* of the undirected graph with vertices V_n induced by the symmetric set E_n). In this case, a direct computation, based on criterion (1.3), shows that Condition (ii) of Theorem 3.1 holds if and only if, as $n \rightarrow \infty$,

$$\max_u |\lambda_{n,u}| = o(|E_n|^{1/2}). \quad (3.2)$$

Now denote by $\Delta(n)$ the *maximal degree* of the graph associated with the adjacency matrix A_n . It is easily seen that, since $\max_u |\lambda_{n,u}| \leq \Delta(n)$, condition (3.2) is implied by the stronger relation

$$\Delta(n) = o(|E_n|^{1/2}). \quad (3.3)$$

Also, if such a graph is d_n -regular, then relation (3.2) is equivalent to $d_n = o(|V_n|)$. See [9, Proposition 1.11] for a comprehensive statement applying to generic collections of i.i.d. random variables, and also [24, Section 3.1].

3.2. Reducible sequences verify a CLT. In order to provide some intuition, we now explain why *reducible sequences satisfy a Lindeberg-Feller-type CLT*, following [20]. Assume that the sequence $\{Z_n\}$ is reducible in the sense of Definition 2.1, and define

$$T_n := \sum_{i=1}^{m_n} \sum_{\substack{(v_1, \dots, v_d) \in \\ E_n \cap (B_i \times \cdots \times B_i)}} q_n(v_1, \dots, v_d) \prod_{\ell=1}^d X_{v_\ell} := \sum_{i=1}^{m_n} Z_{n,i}.$$

Then, for fixed n , one has that the random variables $\{Z_{n,i} : i = 1, \dots, m_n\}$ are stochastically independent and such that $\mathbf{Var}(Z_{n,i}) = \sigma_n^2(B_i)$. Moreover, Property (ii) yields the asymptotic relations (as $n \rightarrow \infty$)

$$\mathbf{Var}(T_n) = \sum_{i=1}^{m_n} \sigma_n^2(B_i) \sim d! \|q_n\|^2 \quad \text{and} \quad \mathbb{E}[(\tilde{Z}_n - \tilde{T}_n)^2] = \mathbb{E}[\tilde{Z}_n^2] - \mathbb{E}[\tilde{T}_n^2] \rightarrow 0, \quad (3.4)$$

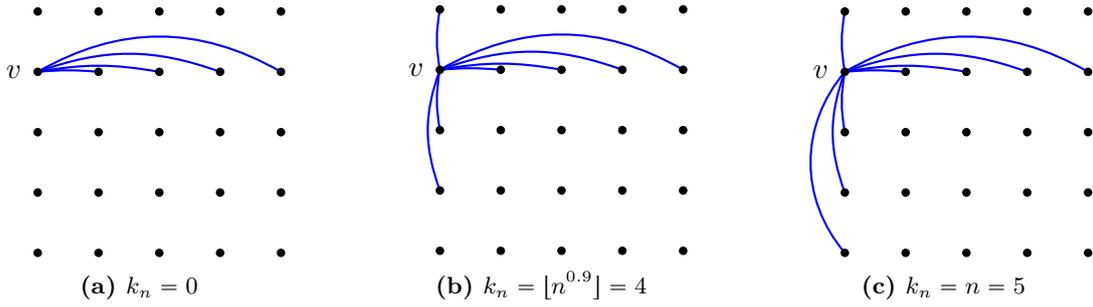


Figure 3. The edges adjacent to the vertex $v = (2, 1)$ in the graph underlying Example 3.3-(b), for $n = 5$ and different choices of k_n .

where

$$\tilde{T}_n := \frac{1}{\sqrt{d! \|q_n\|^2}} \sum_{i=1}^{m_n} Z_{n,i} := \sum_{i=1}^{m_n} \tilde{Z}_{n,i}.$$

Since every $\tilde{Z}_{n,i}$ is an element of the d th Wiener chaos associated with the Gaussian family \mathbf{X} , one can use standard hypercontractive estimates [69, Corollary 2.8.14] to deduce that, for all $p > 2$ and for some absolute constant $c_{p,d}$,

$$\sum_{i=1}^{m_n} \mathbb{E}[|\tilde{Z}_{n,i}|^p] \leq \frac{c_{p,d}}{(d! \|q_n\|^2)^{p/2}} \sum_{i=1}^{m_n} (\sigma_n^2(B_i))^{p/2} \leq c_{p,d} \max_{i=1, \dots, m_n} \sigma_n^2(B_i)^{p/2-1} = o(1), \quad n \rightarrow \infty, \quad (3.5)$$

where the last relation follows by combining Properties (ii) and (iii) of Definition 2.1. Applying Markov inequality as in [20, formulae (4.12)–(4.14)], one sees that (3.5) implies that $\tilde{T}_n, \tilde{Z}_n \xrightarrow{d} N \sim \mathcal{N}(0, 1)$ by the Lindeberg-Feller CLT, see e.g. [53, Theorem 4.12].

The next two sections present several examples of reducible and irreducible sequences, focusing mainly on the case $d = 2$, which corresponds to the results of Section 2.3.1 (except for Example 3.3 (a)). Throughout Sections 3.3–3.4, we consider sequences of simple graphs $\{G_n = (V_n, \mathcal{E}_n)\}$, which may vary from one example to another, and adopt the notation

$$0 = \mu_1^{(n)} \leq \dots \leq \mu_{|V_n|}^{(n)} \leq 2$$

for the eigenvalues of the normalized Laplacian associated with G_n . More background can be found in Section 4.

Further illustrations in the case $d \geq 3$ will require some non-trivial notions associated with hypergraphs and their adjacency structure, and will be discussed in the forthcoming Section 5, see in particular Examples 5.8 and 5.9.

3.3. Examples of reducible sequences. We start by illustrating some explicit examples of *reducible* sequences.

Example 3.3. (a) A trivial example of reducible sequence is obtained by considering the random variables $Z_n := \sum_{i=0}^{n-1} \prod_{\ell=1}^d X_{id+\ell}$, $n \geq 1$, appearing in (2.7). Indeed, in this case one has that $V_n = \{1, \dots, dn\} = [dn]$, and one can take $m_n = n$ and

$$B_i = B_i(n) = \{(i-1)d + 1, \dots, id\}, \quad i = 1, \dots, n.$$

- (b) We now build a non-trivial example associated with the sequence $V_n := \{(i, j) : i, j \in [n]\}$, $n \geq 1$. Let k_n be an integer such that $0 \leq k_n \leq n$ and define the set $E_n \subset V_n \times V_n$ as follows: given $v_1 = (a, b) \in V_n$, $v_2 = (i, j) \in V_n$, then $(v_1, v_2) \in E_n$ if and only if:

- either $a = i$ with $b \neq j$;
- or if $a \neq i$ with $b = j$ and, additionally, $i \leq k_n$.

Let us henceforth assume that $k_n = o(n)$. Then, it is an exercise to show that $|E_n| \sim n^3$. Moreover, since $k_n = o(n)$, one has that the family $\{E_n\}$ is reducible in the sense of Remark 2.3-(1), with $m_n = n$ and

$$B_i = B_i(n) = \{(i, \ell) : \ell = 1, \dots, n\}, \quad i = 1, \dots, n.$$

As already observed in Remark 3.2-(2), one can identify the function $(v_1, v_2) \mapsto \mathbb{1}_{E_n}(v_1, v_2)$ with the adjacency matrix A_n of the undirected graph G_n induced by E_n on the vertex set V_n . An illustration of such a graph for $n = 5$ and different choices of k_n can be found in Fig. 3. A graphical representation of the reducible set E_n for $n = 8, 9, 11$ and $k_n = \lfloor n^{0.9} \rfloor$ is provided in Fig. 4: in such a picture, the blue tiles around the diagonal correspond to the disjoint sets $E_n \cap (B_i \times B_i)$, $i = 1, \dots, n$, whereas the collection of all red dots corresponds to the set

$$E_n \setminus \bigcup_{i=1}^{m_n} (B_i \times B_i).$$

The sequence of graphs illustrated in Figure 3-(a) corresponds to the limiting case $k_n = 0$, yielding that G_n is the disjoint union of n copies of K_n , that is, $G_n = K_n \sqcup \dots \sqcup K_n$, $n \geq 1$. Now write $\{\mu_k^{(n)}\}$ for the Since each G_n has exactly n connected components, one has that $0 = \mu_1^{(n)} = \dots = \mu_n^{(n)}$, $n \geq 1$, so that $\mu_k^{(n)} \rightarrow 0$ for every fixed k . This behavior is consistent with Theorem 2.6.

- (c) For $n \geq 2$, the *hypercube* Q_n is the n -regular graph with vertex set V_n given by all strings of the type $x = (x_1, \dots, x_n) \in \{0, 1\}^n$, and such that $x, y \in V_n$ are connected by an edge ($x \sim y$) if and only if they differ exactly by one coordinate, that is, $\sum_{i=1}^n |x_i - y_i| = 1$. One has that $|V_n| = 2^n$. Also, if one defines

$$E_n := \{(x, y) \in V_n^2 : x \sim y\}, \quad n \geq 1, \tag{3.6}$$

one has that $|E_n| = n2^{n-1}$ (twice the number of edges in Q_n). We now show that the sequence $\{E_n\}$ is reducible in the sense of Remark 2.3-(1). To see this, for $h < n$, consider the set $V_h = \{0, 1\}^h$ and, for $z \in V_h$, define

$$B_n^{(z)} := \{x \in V_n : x_i = z_i \ \forall i = 1, \dots, h\},$$

in such a way that $|B_n^{(z)}| = 2^k$ with $k = n - h$. Choosing sequences $h_n, k_n \rightarrow \infty$ as $n \rightarrow \infty$ such that $h_n + k_n = n$, $\frac{k_n}{n} \rightarrow 1$ as $n \rightarrow \infty$ [†], one has that the family of partitions $\{B_n^{(z)} : z \in V_{h_n}\}$, $n \geq 1$, satisfies the three requirements (a)–(c) of Remark 2.3-(1). Indeed, (a) follows since $m_n := |V_{h_n}| = 2^{h_n} \rightarrow \infty$. Moreover, for any $z \in V_{h_n}$ one has that $|E_n \cap (B_n^{(z)} \times B_n^{(z)})| = |\{(x, y) \in E_n : x, y \in B_n^{(z)}\}| = k_n 2^{k_n}$,

[†]Choose for instance $h_n = \lfloor \log n \rfloor$ and $k_n = n - \lfloor \log n \rfloor$.

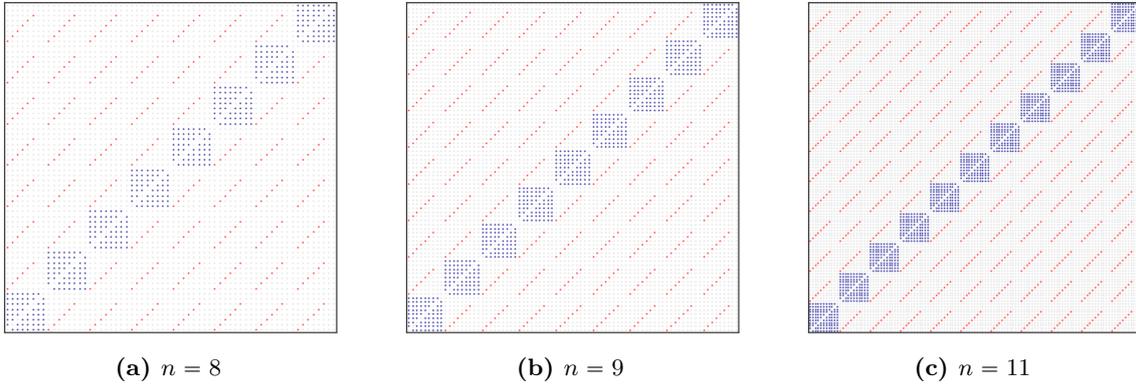


Figure 4. The set $E_n \subset V_n \times V_n$ (represented as the union of red dots and blue dots) from Example 3.3-(b), for $n = 8, 9, 11$ and $k_n = \lfloor n^{0.9} \rfloor$, where each element of $V_n = \{(i, j) : i, j \in [n]\}$ has been identified with an element of $[n^2]$ by ordering V_n according to the lexicographic order. Pairs (v_1, v_2) not belonging to E_n are represented as grey dots. In this case, the reducibility of $\{E_n\}$ implies that the red dots can be removed from E_n without affecting the asymptotic behavior of the homogeneous sum Z_n defined in (2.4) for $q_n \equiv 1$.

which implies that (b) and (c) are verified because

$$\frac{|E_n \cap (B_n^{(z)} \times B_n^{(z)})|}{|E_n|} = \frac{k_n 2^{k_n}}{n 2^n} = \frac{k_n}{n} \frac{1}{2^{h_n}} \xrightarrow{n \rightarrow \infty} 0,$$

uniformly in $z \in V_{h_n}$, and

$$\frac{\sum_{z \in V_{h_n}} |E_n \cap (B_n^{(z)} \times B_n^{(z)})|}{|E_n|} = \frac{2^{h_n} k_n 2^{k_n}}{n 2^n} = \frac{k_n}{n} \xrightarrow{n \rightarrow \infty} 1.$$

We observe that, classically, $Q_n = K_2^{\square n}$, that is, Q_n is the Cartesian product of n copies of the complete graph over two vertices.

3.4. Examples of irreducible sequences. In this section, we make use of the lexical conventions put forward in Definition 2.5 and present several examples of irreducible sequences of graphs that generate irreducible CLTs.

We start with a general fact showing that, under fairly general conditions, Cartesian products of regular graphs always generate irreducible CLTs. For definitions and properties of Cartesian products, we refer to Appendix B, as well as to [48, Sections 4, 5 and 33] or [19, Section 1.4].

Proposition 3.4. *Consider a sequence $G_n^0 = (V_n^0, \mathcal{E}_n^0)$, $n \geq 1$, of connected graphs, with normalized Laplace spectra given by*

$$0 = \mu_1^{(0,n)} < \mu_2^{(0,n)} \leq \dots \leq \mu_{|V_n^0|}^{(0,n)} \leq 2, \quad n \geq 1.$$

Assume that (i) G_n^0 is d_n -regular for some $d_n \geq 1$, (ii) $|V_n^0| \rightarrow \infty$, and (iii) $\liminf_{n \rightarrow \infty} \mu_2^{(0,n)} > 0$. Then, for any fixed integer $m \geq 2$, the sequence of product graphs

$$G_n := \left(G_n^0\right)^{\square m}, \quad n \geq 1,$$

generates an irreducible CLT, in the sense of Definition 2.5.

Proof. Writing $|V_n^0| := N_n$, one has that G_n is a md_n -regular connected graph with N_n^m vertices. Denote by $0 = \mu_1^{(n)} < \mu_2^{(n)} \leq \dots \leq \mu_{N_n^m}^{(n)} \leq 2$, $n \geq 1$, the normalized Laplace eigenvalues of G_n . Then, using the content of Appendix B (see also [19, Section 1.4.6]) and the regularity of G_n one has that $\mu_2^{(n)} = \mu_2^{(0,n)}/m$ (with multiplicity equal to m times the multiplicity of $\mu_2^{(0,n)}$ in G_n^0), and the irreducibility of $\{G_n\}$ immediately follows from Theorem 2.6. The fact that $\{G_n\}$ generates a CLT is a consequence of Remark 3.2-(2), and of the property that, trivially,

$$\frac{md_n}{N_n^m} \leq \frac{m}{N_n^{m-1}} \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

□

Remark 3.5. We stress that Proposition 3.4 holds without any assumption on the numerical sequence $\{d_n\}$. In particular, one may have $d_n \asymp |V_n|$, in which case (by virtue of Remark 3.2-(2)) the sequence $\{G_n^0\}$ does not generate a CLT (and is therefore irreducible, see Subsection 3.2). Heuristically, Proposition 3.4 shows that, given such a $\{G_n^0\}$, the m -fold Cartesian power operation

$$G_n^0 \mapsto (G_n^0)^{\square m}$$

creates enough sparsity to induce Gaussian fluctuations, yet without altering the irreducible structure. See the forthcoming Examples (b) and (c) for two illustrations of this phenomenon.

To analyze some of the examples described below, we will use the following facts:

- for $n \geq 2$, the complete graph K_n over n vertices has normalized Laplace spectrum given by

$$\left\{ 0, \frac{n}{n-1} \right\}, \quad \text{with respective multiplicities } 1 \text{ and } (n-1); \quad (3.7)$$

- for $n \geq 2$, the complete bipartite graph $K_{n,n}$ over $2n$ vertices has normalized Laplace spectrum given by

$$\{0, 1, 2\}, \quad \text{with respective multiplicities } 1, 2(n-1) \text{ and } 1. \quad (3.8)$$

See e.g. [19, Section 1.4]. For the rest of the section, and consistently with the notation adopted before, we write $\{X_i : i \geq 1\}$ to indicate a collection of i.i.d. $\mathcal{N}(0, 1)$ random variables.

Some explicit examples of irreducible sequences ($d = 2$):

- (a) (*Expanders*) Fix $m \geq 2$ and $\varepsilon > 0$, and consider a sequence of m -regular graphs $G_n = (V_n, \mathcal{E}_n)$, $n \geq 1$, such that $|\mathcal{E}_n|, |V_n| \rightarrow \infty$. According to the classical definition given e.g. in [51, Definition 2.2], the family $\{G_n\}$ is said to be a ε -*expander* if

$$\min_{0 < |S| \leq |V_n|/2} \frac{E(S, \bar{S})}{\text{Vol}(S)} := \tilde{\varphi}_2(G_n) \geq \frac{\varepsilon}{m}$$

for all n , where $E(S, \bar{S})$ indicates the number of edges of G_n connecting S with its complement \bar{S} , and $\text{Vol}(S)$ is the sum of the degrees of the vertices in S (see Section 4, in particular the discussion around formula (4.5), for more details). Using the Cheeger-type inequality stated below in (4.8), one sees that, in this case,

$\liminf_{n \rightarrow \infty} \mu_2^{(n)} \geq \varepsilon^2 / (2m^2) > 0$. Since m is fixed while $|V_n| \rightarrow \infty$, in particular $m/|V_n| \rightarrow 0$, one can now combine Remark 3.2-(2) with Theorem 2.6 to infer that $\{G_n\}$ generates an irreducible CLT, in the sense of Definition 2.5. An analogous conclusion holds in the case of a sequence of m_n -regular graphs verifying $\tilde{\varphi}_2(G_n) \geq \varepsilon_n/m_n$ for some ε_n such that $\liminf_{n \rightarrow \infty} \varepsilon_n/m_n > 0$, and such that $m_n = o(|V_n|)$. Canonical examples of expanders include *Ramanujan graphs* [51, 62, 66] and *random d -regular graphs* (which are expanders with high probability, see [27, 44, 81]).

- (b) (*Cartesian products of complete graphs*) This example significantly expands the content of Example (c) of Section 3.3. For integers $q, m \geq 2$, consider the m -fold Cartesian product $G(q, m) := K_q^{\square m}$. Then, combining (3.7) (for $n = q$) with the content of Appendix B (see also [48, Proposition 33.6] or [19, Section 1.4.6]), one sees that $G(q, m)$ is a $m(q-1)$ regular graph over q^m vertices with a normalized Laplace spectrum given by

$$\left\{ \frac{\ell q}{m(q-1)} : \ell = 0, 1, \dots, m \right\},$$

and respective multiplicities

$$\binom{m}{\ell} (q-1)^\ell, \quad \ell = 0, \dots, m.$$

We can immediately draw the following conclusions:

- (i) For fixed $m \geq 2$, the sequence of graphs $G_n := G(n, m) = K_n^{\square m}$, $n \geq 1$, generates an irreducible CLT as a direct consequence of Proposition 3.4 in the case $G_0^n = K_n$ (taking into account (3.7)). Note that $G(n, 2) = K_n^{\square 2}$ corresponds to the ‘‘Rook’s graph’’, studied e.g. in [58]. It is interesting to observe that the sequence $\{G(n, 2)\}$ can be identified with the sequence (V_n, \mathcal{E}_n) , $n \geq 1$, from Example 3.3-(b) *in the special case $k_n = n$* (as illustrated in Figure 3-(c)). This framework also corresponds to the example introduced in Subsection 2.4.2, with $\mathbb{S}_v(a) = \{a\} \times \{1, \dots, n\}$ and $\mathbb{S}_h(b) = \{1, \dots, n\} \times \{b\}$.
- (ii) For fixed $q \geq 2$, the sequence of graphs $G_n := G(q, n)$ is such that, for all $k \geq 2$, $\lim_{n \rightarrow \infty} \mu_k^{(n)} = 0$ and, in principle, one cannot use Theorem 2.6 to determine whether $\{G_n\}$ is reducible or not. The reducibility of $\{G_n\}$ can be established by using a direct argument. To see this, observe that G_n is isomorphic to the graph whose vertices are given by all vectors of the form $v = (v_1, \dots, v_n)$, where $v_i \in \{0, 1, \dots, q-1\}$ and $v \sim w$ if and only if $v_i \neq w_i$ for exactly one index $i = 1, \dots, n$; in particular, $G(2, n) \simeq Q_n$, the hypercube considered in Example 3.3-(c). The reducibility of $\{G_n\}$ can now be deduced by a straightforward variation of the arguments rehearsed in Example 3.3 (the details are left to the reader).

We observe that, if K_n is identified with the complete graph over $[n]$, then the sequence of homogeneous sums generated by $\{K_n\}$ (in the sense of Definition 2.5) is given for $n \in \mathbb{N}$ by

$$Z_n = \sum_{1 \leq i \neq j \leq n} X_i X_j = (X_1 + \dots + X_n)^2 - (X_1^2 + \dots + X_n^2).$$

It is easily seen that $\mathbf{Var}(Z_n) = 2n(n-1)$ and that, as $n \rightarrow \infty$, $\tilde{Z}_n := Z_n / \sqrt{2n(n-1)}$ converges in distribution towards $\frac{1}{\sqrt{2}}(N^2 - 1)$ with $N \sim \mathcal{N}(0, 1)$, that is a multiple

of a centered χ^2 random variable with one degree of freedom. See Remark 3.5, as well as [9, formula (1.4)]. Thus, as announced, the sequence of graphs K_n does not generate a CLT, while for any fixed $m \geq 2$ the cartesian products $K_n^{\square m}$ generate (an irreducible) CLT.

- (c) (*Cartesian products of bipartite graphs*) Fix $m \geq 2$. By virtue of (3.8), another direct application of Proposition 3.4 in the case $G_n^0 = K_{n,n}$, shows that the sequence $G_n = K_{n,n}^{\square m}$, $n \geq 1$, generates an irreducible CLT. We observe that G_n can be identified with the graph having vertex set $V_n = \{(v_1, \dots, v_m) : v_i \in [2n]\}$ and such that $v \sim w$ if and only if $v_i \neq w_i$ for exactly one index i , and either $v_i \in [n]$ and $w_i \in \{n+1, \dots, 2n\}$, or $w_i \in [n]$ and $v_i \in \{n+1, \dots, 2n\}$. See Figure 5 for an illustration of this graph in the case $n = 3$ and $m = 2$. If $K_{n,n}$ is identified with the bipartite graph over $[2n]$, then the sequence of homogeneous sums generated by $\{K_{n,n}\}$ can be written as

$$Z_n = 2 \left(\sum_{i=1}^n X_i \right) \times \left(\sum_{j=n+1}^{2n} X_j \right) \stackrel{\text{law}}{\equiv} 2n X_1 X_2, \quad n \geq 1.$$

See again Remark 3.5, as well as [9, Example 2.2]. Also in this case, the sequence of graphs $K_{n,n}$ does not generate a CLT, but for any fixed $m \geq 2$ the Cartesian products $K_{n,n}^{\square m}$ generate (an irreducible) CLT.

- (d) (*Cartesian products of graphs with isolated vertices*) Let $\{m_n\}$ be an integer-valued sequence such that $m_n \rightarrow \infty$ and consider the framework of Subsection 2.4.2 for $\mathbb{S}_v(a) = \{a\} \times \{1, \dots, m_n\}$ and $\mathbb{S}_h(b) = \{1, \dots, m_n\} \times \{b\}$. In this case, it is easily seen that the indicator $\mathbb{1}_{E_n}$ coincides with the adjacency matrix of the graph

$$U(n) := (K_{m_n} \sqcup \bar{K}_{n-m_n}) \square (K_{m_n} \sqcup \bar{K}_{n-m_n}),$$

where K_{m_n} is the complete graph over m_n vertices and \bar{K}_{n-m_n} is the trivial graph with $n - m_n$ isolated vertices. A direct inspection shows that the graph $U(n)$ has $(n - m_n)^2$ isolated vertices, $2(n - m_n)$ connected components isomorphic to K_{m_n} and one “giant” component isomorphic to $K_{m_n} \square K_{m_n}$ (some visualizations of E_n in the special setting of the present example are provided in Figure 1). Note that we already know that the sequence $\{K_{m_n}\}$ is associated with an irreducible CLT (see Point (i) in Example (b) of the present section).

We write G_n , $n \geq 1$, to denote the graph obtained from $U(n)$ after removing all isolated vertices. Using once again (3.7), one easily deduces that, for each n , the normalized Laplacian of G_n has spectrum $\{0, m_n/2(m_n - 1), m_n/(m_n - 1)\}$, with respective multiplicities $1 + 2(n - m_n)$, $2(m_n - 1)$ and $2(m_n - 1)^2 + 2(n - m_n)(m_n - 1)$. One can verify the following (details are left to the reader):

- (i) If $m_n = o(n)$ then $\{G_n\}$ generates a reducible CLT;

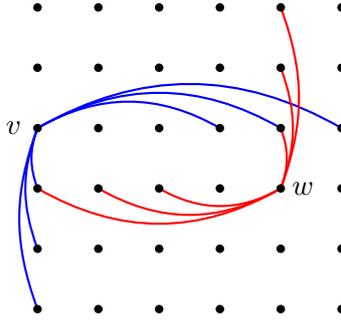


Figure 5. In blue: the edges adjacent to the vertex $v = (3, 1)$ in the graph $(K_{3,3})^{\square 2}$. In red: the edges adjacent to the vertex $w = (4, 5)$, in the same graph.

- (ii) If $m_n/n \rightarrow \beta \in (0, 1)$, then $\{G_n\}$ generates an irreducible CLT[†], and $\lim_{n \rightarrow \infty} \mu_k^{(n)} = 0$ for all k since the multiplicity of the zero Laplace eigenvalue is $1 + 2(n - m_n) \rightarrow \infty$;
- (iii) If $m_n/n \rightarrow 1$, then $\{G_n\}$ generates an irreducible CLT (by the same argument as for the case $\beta \in (0, 1)$), and there exists an integer $k \geq 2$ such that $\liminf_n \mu_k^{(n)} > 0$ if and only if $\limsup_n (n - m_n) < \infty$ (and, in this case, $\liminf_n \mu_k^{(n)} = 1/2$).

4. Irreducibility and graph spectra: background and proofs

In this section we prove Theorem 2.6, yielding a characterization of irreducible sequences $\{Z_n\}$ defined as in (2.10).

4.1. Preliminaries: some connectivity estimates.

4.1.1. GENERAL SETTING. Let $G = (V, \mathcal{E})$ be a finite (simple, undirected, loop-free) graph with $|V| = N \geq 2$ vertices, each one having a possibly different *degree* $d(v) := |\{e \in \mathcal{E} : v \in e\}|$, $v \in V$. As usual, we regard edges $e \in \mathcal{E}$ as unordered pairs $e = \{v, w\}$ and make extended use of the basic *degree sum formula* $\sum_{v \in V} d(v) = 2|\mathcal{E}|$. In resonance with Convention 2.4, for our purposes there is no loss of generality in assuming that the graph has no isolated vertices, that is, $d(v) > 0$ for all $v \in V$. For future use, we also fix an (arbitrary) enumeration $\{v_1, \dots, v_N\}$ of V .

We write \mathbb{R}^V to indicate the set of all functions $V \rightarrow \mathbb{R}$, whose generic element is written $\mathbf{x} = \{x_v : v \in V\}$. The space \mathbb{R}^V is canonically identified with the Euclidean space \mathbb{R}^N , while the Euclidean scalar product and norm on \mathbb{R}^V are denoted by $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$, respectively. Given $S \subseteq V$, we use the symbol $\mathbf{1}_S$ to indicate the element of \mathbb{R}^V such that x_v equals one

[†]Since the total number of edges in $K_{m_n} \square K_{m_n}$ scales as $\asymp \beta^3 n^3$, one has that the corresponding sequence of quadratic form $\{Z_n\}$ (defined as in (2.10)) is such that each Z_n can be written as the sum of two independent homogeneous sums of order 2,

$$Z_n = Z'_n + Z''_n,$$

in such a way that, as $n \rightarrow \infty$, $\{Z'_n\}$ is irreducible and such that

$$\mathbf{Var}(Z'_n) \sim \beta \mathbf{Var}(Z_n).$$

This implies the irreducibility of $\{Z_n\}$.

or zero according as $v \in S$ or $v \in \bar{S}$, where we used the notation

$$\bar{S} := V \setminus S. \quad (4.1)$$

Given $S, T \subseteq V$, we write

$$E(S, T) := |\{e = \{s, t\} \in \mathcal{E} : s \in S, t \in T\}|, \quad (4.2)$$

that is, $E(S, T)$ is the number of edges connecting S and T . We denote by $A := \{A(v, w) = \mathbb{1}_{\{v, w\} \in \mathcal{E}} : v, w \in V\}$ the $N \times N$ *adjacency matrix* of G and write

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$$

to indicate its eigenvalues, in such way that

$$\sum_{i=1}^N \lambda_i^2 = \text{Trace}(A^2) = \sum_{v \in V} A^2(v, v) = \sum_{v \in V} d(v) = 2|\mathcal{E}|.$$

We observe that, for every $S \subseteq V$, one has that

$$E(S, S) = \frac{1}{2} \langle \mathbf{1}_S, A \mathbf{1}_S \rangle, \quad E(S, \bar{S}) = \langle \mathbf{1}_S, A \mathbf{1}_{\bar{S}} \rangle,$$

where the factor $\frac{1}{2}$ is motivated by the fact that each edge $e = \{v, w\}$ with $v, w \in S$ is associated with two distinct ordered pairs (v, w) and (w, v) .

The *Laplacian matrix* of G is defined as

$$L := D - A, \quad \text{where} \quad D(v, w) := d(v) \mathbb{1}_{v=w}, \quad v, w \in V,$$

that is: D is the $N \times N$ diagonal *degree matrix*, with diagonal entries $d(v_1), \dots, d(v_N)$. We also introduce the *normalized Laplacian* of G as

$$\mathcal{L} := I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}},$$

whose eigenvalues (see [87, Section 3.2]) satisfy

$$0 = \mu_1 \leq \mu_2 \leq \dots \leq \mu_N \leq 2.$$

The second eigenvalue μ_2 is the so-called *spectral gap* of G . In the special case where G is a *d-regular graph*, i.e. $d(v) = d$ for all $v \in V$ and for some $d \geq 1$, the normalized Laplacian simplifies to $\mathcal{L} = I - \frac{1}{d} A$ and $\mu_i = 1 - \frac{\lambda_i}{d}$, $i = 1, \dots, N$.

For every nonzero vector $\mathbf{x} \in \mathbb{R}^V$, and writing $\mathbf{y} := D^{-\frac{1}{2}} \mathbf{x}$, we define the *Raleygh quotient* associated with \mathcal{L} and \mathbf{y} to be the ratio

$$\frac{\langle \mathbf{x}, \mathcal{L} \mathbf{x} \rangle}{\|\mathbf{x}\|^2} = \frac{\langle \mathbf{x}, D^{-\frac{1}{2}} L D^{-\frac{1}{2}} \mathbf{x} \rangle}{\|\mathbf{x}\|^2} = \frac{\langle \mathbf{y}, L \mathbf{y} \rangle}{\|D^{\frac{1}{2}} \mathbf{y}\|^2} = \frac{\sum_{u \sim v} (y_u - y_v)^2}{\sum_{v \in V} d(v) y_v^2} =: R_{\mathcal{L}}(\mathbf{y}).$$

Remark 4.1. In the sequel, we will use the following variational characterisation of the eigenvalues of \mathcal{L} (see e.g. [87, Section 1.3]): for all $k = 1, \dots, N$,

$$\mu_k = \min_{\substack{\mathcal{M} \subseteq \mathbb{R}^V \\ \dim(\mathcal{M})=k}} \max_{\mathbf{z} \in \mathcal{M} \setminus \{0\}} R_{\mathcal{L}}(\mathbf{z}), \quad (4.3)$$

where the minimum runs over all k -dimensional subspaces \mathcal{M} of \mathbb{R}^V . It follows that $\mu_1 = 0$ and the multiplicity of μ_1 is equal to the number of connected components of G .

4.1.2. CHEEGER'S INEQUALITIES. To introduce Cheeger's inequalities, we first define a quantity that measures how well a subset $S \subset V$ is connected to its complement \bar{S} , relative to the total number of edges incident to the vertices in S . Given $\emptyset \neq S \subseteq V$, the *edge expansion* of S is defined as

$$\varphi(S) := R_{\mathcal{L}}(\mathbf{1}_S) = \frac{E(S, \bar{S})}{\text{vol}(S)}, \quad \text{where we set } \text{vol}(S) := \sum_{v \in S} d(v).$$

and the second equality follows from the relation $\langle \mathbf{1}_S, (D - A)\mathbf{1}_S \rangle = \text{vol}(S) - 2E(S, S) = E(S, \bar{S})$ (see Figure 6 for illustrations). In particular, one has trivially that $\varphi(V) = 0$. The quantity $\text{vol}(S)$ is called the *volume of S* , and it equals $d|S|$ when G is d -regular.

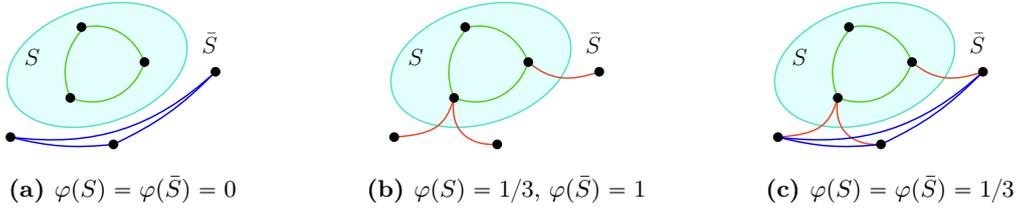


Figure 6. Computing edge expansions in various graphs: in each picture, the number of red edges equals the quantity $E(S, \bar{S})$ entering the definition of $\varphi(S)$.

For $k \geq 2$ we introduce the notation

$$\varphi_k(G) := \min_{\substack{S_1, \dots, S_k \subset V \\ \text{nonempty and disjoint}}} \max_{i=1, \dots, k} \varphi(S_i). \quad (4.4)$$

Remark 4.2. (1) (*A probabilistic interpretation of edge expansions*) Fix a nonempty $S \subseteq V$, and sample an edge adjacent to one of the vertices in S according to the following procedure: (i) sample a vertex v_0 in S with a probability proportional to its degree, and (ii) sample an edge e_0 adjacent to v_0 uniformly at random. Then, $\varphi(S)$ equals the probability that e_0 connects S and \bar{S} , that is, that $e_0 = \{a, b\}$ for some $a \in S$ and $b \in \bar{S}$. Note that, in the regular case, Step (i) above reduces to selecting one vertex uniformly at random within S .

(2) For every $k \geq 2$ one has that

$$\varphi_k(G) \leq \min_{\substack{S_1, \dots, S_k \\ \text{partition of } V \text{ with } S_i \text{ nonempty}}} \max_{i=1, \dots, k} \varphi(S_i),$$

so that, in particular,

$$\varphi_2(G) \leq \min_{\emptyset \subset S \subset V} \max \{ \varphi(S), \varphi(\bar{S}) \} = \min_{0 < |S| \leq |V|/2} \varphi(S) =: \tilde{\varphi}_2(G). \quad (4.5)$$

We now state the “easy direction” of *Cheeger's inequalities* of order k , and include a sketch of the proof — not only for completeness, but also because the result plays a crucial role in the arguments developed below. The proof relies on the min-max characterization of the eigenvalues of \mathcal{L} ; see Remark 4.1.

Proposition 4.3 (Cheeger's inequalities, easy direction). *Let $G = (V, \mathcal{E})$ be a finite graph respecting Convention 2.4 with $|V| = N$ and let μ_k be the k -th smallest eigenvalue of the normalized Laplacian \mathcal{L} , for $k = 2, \dots, N$. Then,*

$$\mu_k \leq 2 \varphi_k(G). \quad (4.6)$$

Sketch of the proof. Fix nonempty disjoint subsets $S_1, \dots, S_k \subset V$. A standard extension to irregular graphs of [87, Lemma 7.1] yields that, for all vectors $(\alpha_1, \dots, \alpha_k) \in \mathbb{R}^k$ not identically zero,

$$R_L \left(\sum_{i=1}^k \alpha_i \mathbf{1}_{S_i} \right) \leq 2 \frac{\sum_{i=1}^k \alpha_i^2 \text{vol}(S_i) \varphi(S_i)}{\sum_{i=1}^k \alpha_i^2 \text{vol}(S_i)} \leq 2 \max_{i=1, \dots, k} \varphi(S_i). \quad (4.7)$$

The conclusion follows from (4.3), first by computing $\max_{\mathbf{z} \in \mathcal{M} \setminus \{0\}} R_{\mathcal{L}}(\mathbf{z})$ when \mathcal{M} is the vector space generated by $\mathbf{1}_{S_1}, \dots, \mathbf{1}_{S_k}$, and then by taking the minimum over all nonempty disjoint subsets $S_1, \dots, S_k \subset V$. \square

The following statement (of which we omit the proof) is taken from [60, Theorem 3.8], and is a version of the ‘‘hard Cheeger inequality’’ for graph partitions; see also [61].

Theorem 4.4 (Cheeger’s inequality, hard direction). *Let $G = (V, \mathcal{E})$ be a finite graph, respecting Convention 2.4 and such that $|V| = N$, and let μ_k be the k -th smallest eigenvalue of its normalized Laplacian \mathcal{L} , for $k = 2, \dots, N$. Then, there exists an absolute constant $C > 0$ such that, for every $k = 2, \dots, N$, there exists a partition S_1, \dots, S_k of V with nonempty subsets verifying*

$$\max_{i=1, \dots, k} \varphi(S_i) \leq Ck^4 \sqrt{\mu_k}.$$

In particular, one has that $\varphi_k(G) \leq Ck^4 \sqrt{\mu_k}$.

We also remark that, in the case $k = 2$, one has the stronger estimate

$$\tilde{\varphi}_2(G) \leq \sqrt{2\mu_2}, \quad (4.8)$$

where $\tilde{\varphi}_2(G)$ is defined according to (4.5) — see e.g. [51, Theorem 2.4].

We conclude by stating an elementary bound, that will be exploited in the proof of Theorem 2.6.

Proposition 4.5. *Let $G = (V, \mathcal{E})$ be a finite graph as in Convention 2.4, with $|V| = N$. Fix an integer $1 \leq m \leq N$. Consider a collection B_1, \dots, B_m of nonempty disjoint subsets of V , and assume that*

$$\varphi(B_1) \leq \varphi(B_2) \leq \dots \leq \varphi(B_m). \quad (4.9)$$

Then, for all $k = 1, \dots, m$,

$$\mu_k \leq 2 \frac{\sum_{i=k}^m \text{vol}(B_i) \varphi(B_i)}{\sum_{i=k}^m \text{vol}(B_i)}. \quad (4.10)$$

Proof. Fix $k = 1, \dots, m$. Then (4.9) implies that

$$\frac{\sum_{i=k}^m \text{vol}(B_i) \varphi(B_i)}{\sum_{i=k}^m \text{vol}(B_i)} \geq \varphi(B_k).$$

The conclusion follows by observing that, again because of (4.9),

$$\varphi(B_k) = \min_{1 \leq j_1 < \dots < j_k \leq m} \max_{\ell=1, \dots, k} \varphi(B_{j_\ell}) \geq \varphi_k(G) \geq \frac{\mu_k}{2},$$

where we have used Proposition 4.3. \square

4.2. Laplace spectra and irreducibility. This subsection focuses on the proof of Theorem 2.6, which requires us to consider a sequence of quadratic sums $\{Z_n\}$ as in (2.10). We stress that such a framework corresponds to the case $d = 2$ and $q_n \equiv 1$ in our general setting (2.4), that the sequence $\{E_n\}$ is assumed to satisfy conditions (1)–(3) from the beginning of Section 2.1, and that the indicator $\mathbf{1}_{E_n}$ is identified with the adjacency matrix of an appropriate graph $G_n = (V_n, \mathcal{E}_n)$ such that $|E_n| = 2|\mathcal{E}_n|$. For $n \geq 1$, the normalized Laplace spectrum of G_n is written $0 = \mu_1^{(n)} \leq \mu_2^{(n)} \leq \dots \leq \mu_{N_n}^{(n)} \leq 2$.

The following elementary result (whose proof is left to the reader) uses the fact that, in the framework of the present section and using the notation (2.8) and (4.2), one has that

$$\sigma_n^2(B) = 2|E_n \cap (B \times B)| = 4E(B, B), \quad B \subseteq V.$$

Lemma 4.6 (Reducibility and graphs). *Let the assumptions and conventions of the present section hold. Then, the sequence $\{Z_n\}$ is reducible in the sense of Definition 2.1 (see also Remark 2.3-(a),(b),(c)) if and only if there exist partitions $\{B_1, \dots, B_{m_n}\} = \{B_1^{(n)}, \dots, B_{m_n}^{(n)}\}$ of the vertex sets V_n , such that, as $n \rightarrow \infty$,*

- (i') $m_n \rightarrow \infty$;
- (ii') $\sum_{i=1}^{m_n} E(B_i, B_i) \sim |\mathcal{E}_n|$;
- (iii') $\max_{i=1, \dots, m_n} E(B_i, B_i) = o(|\mathcal{E}_n|)$.

As indicated in the statement, the definition of the elements partition $\{B_1, \dots, B_{m_n}\}$ depends in principle on n ; this dependence has been omitted in (i'), (ii'), and (iii') to avoid overloading the notation.

We can now prove one of the main results of the paper.

Proof of Theorem 2.6. We reason by contradiction by assuming that (2.12) holds for some $k \geq 2$ and that there exists partitions $\{B_1, \dots, B_{m_n}\}$, $n \geq 1$, such that the three properties (i')–(iii') listed in Lemma 4.6 hold true. Without loss of generality we can always assume that the partition B_1, \dots, B_{m_n} satisfies the ordering relation (4.9) for all $n \geq 1$, so that (4.10) yields

$$\frac{\mu_k^{(n)}}{2} \leq \frac{\sum_{i=k}^{m_n} q_i^{(n)} \varphi(B_i)}{\sum_{i=k}^{m_n} q_i^{(n)}}, \quad \text{where } q_i^{(n)} := \frac{\text{vol}(B_i)}{\text{vol}(V_n)} = \frac{\text{vol}(B_i)}{2|\mathcal{E}_n|}. \quad (4.11)$$

To achieve the desired contradiction and conclude the proof, it is now sufficient to show that the right-hand side of the previous inequality converges to zero. Note that

$$\begin{aligned} \frac{1}{2} \sum_{i=1}^{m_n} E(B_i, \overline{B_i}) &= \frac{1}{2} \sum_{i=1}^{m_n} \text{vol}(B_i) - \sum_{i=1}^{m_n} E(B_i, B_i) \\ &= \frac{1}{2} \text{vol}(V_n) - \sum_{i=1}^{m_n} E(B_i, B_i) \\ &= |\mathcal{E}_n| - \sum_{i=1}^{m_n} E(B_i, B_i) = o(|\mathcal{E}_n|) \end{aligned} \quad (4.12)$$

by property (ii'). As a consequence,

$$\sum_{i=1}^{m_n} \frac{E(B_i, \overline{B_i})}{2|\mathcal{E}_n|} = \sum_{i=1}^{m_n} q_i^{(n)} \varphi(B_i) = o(1). \quad (4.13)$$

This implies that the numerator on the right-hand side of the inequality (4.11) vanishes, and it remains to show that the denominator is bounded away from zero. We conclude by observing that such a denominator converges indeed to 1, which is equivalent to the relation $\sum_{i=1}^{k-1} q_i^{(n)} \rightarrow 0$ since $\sum_{i=1}^{m_n} q_i^{(n)} = 1$. To see this, we simply write

$$q_i^{(n)} = \frac{\text{vol}(B_i)}{\text{vol}(V_n)} = \frac{\text{vol}(B_i)}{2|\mathcal{E}_n|} = \frac{E(B_i, B_i)}{|\mathcal{E}_n|} + \frac{E(B_i, \overline{B_i})}{2|\mathcal{E}_n|} = o(1), \quad (4.14)$$

uniformly for $i \in \{1, \dots, m_n\}$ (because of (iii')). \square

4.3. Full irreducibility. We conclude this section with a (slight) generalization of Theorem 2.6. Let $G_n = (V_n, \mathcal{E}_n)$, $n \geq 1$, be a sequence of graphs, as those studied in the previous section, and consider the associated sequence $\{Z_n\}$. We say that the sequence $\{G_n\}$ is *partially reducible* if the following conditions hold.

- There exist $\varrho \in (0, 1]$ and a subset $V'_n \subseteq V_n$, $n \geq 1$, such that (recall (4.1))

$$\text{vol}(V'_n) = (\varrho + o(1))\text{vol}(V_n), \quad E(V'_n, \overline{V'_n}) = o(|\mathcal{E}_n|), \quad (4.15)$$

that is, V'_n contains asymptotically a fraction $\varrho > 0$ of vertices and it is loosely connected to its complement.

- There exist partitions B_1, \dots, B_{m_n} of V'_n , $n \geq 1$, where $B_i = B_i^{(n)}$ for $i = 1, \dots, m_n$, such that
 - (i'') $m_n \rightarrow \infty$;
 - (ii'') $\sum_{i=1}^{m_n} E(B_i, B_i) = E(V'_n, V'_n) + o(|\mathcal{E}_n|)$;
 - (iii'') $\max_{i=1, \dots, m_n} E(B_i, B_i) = o(|\mathcal{E}_n|)$.

We say that a sequence of graphs $\{G_n\}$ is *fully irreducible* if it is not partially reducible.

Remark 4.7. The second condition in (4.15) is equivalent to $\varphi(V'_n) = o(1)$, thanks to the first condition which yields $|\mathcal{E}_n| = \frac{1}{2}d_n|V_n| \sim \frac{1}{2}\frac{1}{\varrho}d_n|V'_n|$. Also note that the second condition in (4.15) follows by the first one in the extreme case $\varrho = 1$, simply because $E(V'_n, \overline{V'_n}) \leq d_n|\overline{V'_n}| = d_n(|V_n| - |V'_n|) = d_n|V_n|(1 - \varrho + o(1)) = o(d_n|V_n|) = o(|\mathcal{E}_n|)$.

With little effort, we can deduce a slightly improved version of Theorem 2.6, whose proof is presented in Appendix A.

Theorem 4.8. *If condition (2.12) holds, then $\{G_n\}$ is fully irreducible.*

5. Irreducibility and hypergraph spectra

In this section we derive a generalization of Theorem 2.6, extending our results to the setting of homogeneous sums of arbitrary order $d \geq 2$, possibly involving non-trivial coefficients q_n . To this end, we broaden the spectral framework developed in Section 4 by moving to the setting of weighted hypergraphs, partially following the approach of [7, 83]. The main result of this section, Theorem 5.7, strictly contains Theorem 2.6 as a particular case. Nevertheless, we chose to treat the case $d = 2$ separately in order to highlight its specific features and maintain clarity.

5.1. Hypergraphs and extended Cheeger's inequalities. A *hypergraph* \mathcal{G} is a pair $\mathcal{G} = (V, \mathcal{E})$ consisting of a finite vertex set V and a collection \mathcal{E} of non-empty subsets of V , referred to as *hyperedges* (or simply *edges* when the context is clear). To accommodate the most general setting, we consider *weighted* hypergraphs, where a non-negative weight

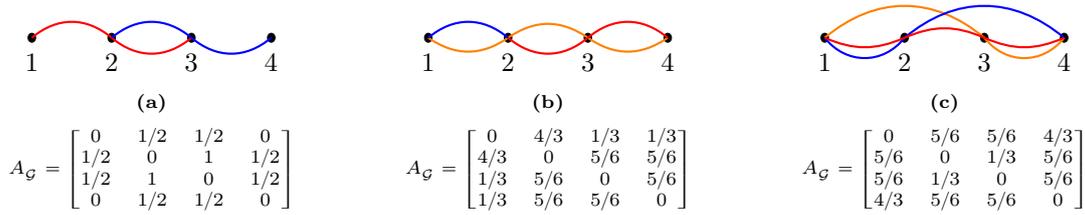


Figure 7. Three non-weighted hypergraphs ($w \equiv 1$) over the vertex set $V = [4]$, endowed with their adjacency matrices. Each hyperedge is represented as a bundle of 2-edges with the same colour, so that, noting $d = (d(1), d(2), d(3), d(4))$ the corresponding degree vector, one has the following configurations: **(a)** $\mathcal{E} = \{\{1, 2, 3\}, \{2, 3, 4\}\}$, $d = (1, 2, 2, 1)$; **(b)** $\mathcal{E} = \{\{1, 2\}, \{1, 2, 3, 4\}, \{2, 3, 4\}\}$, $d = (2, 3, 2, 2)$; **(c)** $\mathcal{E} = \{\{1, 2, 4\}, \{1, 2, 3, 4\}, \{1, 3, 4\}\}$, $d = (3, 2, 2, 3)$. The hypergraph in **(a)** is 3-uniform.

function $w = \{w(e) : e \in \mathcal{E}\}$ is assigned to the edges of \mathcal{G} . We exclude the presence of loops, and therefore assume that $|e| \geq 2$ for all $e \in E$. For $d \geq 2$, we say that \mathcal{G} is d -uniform if $|e| = d$ for every $e \in \mathcal{E}$.

It is evident that the notion of a hypergraph generalizes that of a standard graph, which corresponds to the special case of 2-uniform hypergraphs, that is, hypergraphs where each edge connects exactly two vertices. As in the case of graphs, two vertices $v, w \in V$ are said to be *adjacent*, written $v \sim w$, if there exists an edge $e \in \mathcal{E}$ such that $\{v, w\} \subseteq e$.

Consider a weighted hypergraph $\mathcal{G} = (V, \mathcal{E}, w)$ with $V = \{v_1, \dots, v_N\}$. Our goal is to define an extended version of the adjacency matrix associated with a graph, capable of capturing the connectivity structure of \mathcal{G} . To this end, we follow the approach of [7], which deals with the unweighted case where $w \equiv 1$. Writing $\mathcal{E}_{ij} := \{e \in \mathcal{E} : v_i, v_j \in e\}$, $i, j = 1, \dots, N$, the *adjacency matrix* $A_{\mathcal{G}}$ associated with \mathcal{G} is the $N \times N$ matrix that is zero on the diagonal, and otherwise defined as

$$(A_{\mathcal{G}})_{ij} := \sum_{e \in \mathcal{E}_{ij}} \frac{w(e)}{|e| - 1}, \quad i \neq j \in [N]. \quad (5.1)$$

Heuristic interpretation of (5.1): Fix $v_i \in V$ and assign mass 1 to every edge containing it. Assuming $w \equiv 1$, the adjacency matrix $A_{\mathcal{G}}$ quantifies the connectivity of v_i by distributing the unit mass of each edge $e \ni v_i$ uniformly among the other vertices $v_j \in e \setminus \{v_i\}$.

As a consequence of the choice of normalization in (5.1), if one sums the elements of the i th row (or column) of $A_{\mathcal{G}}$, one obtains exactly the total number of edges that include v_i as an element, a quantity that corresponds to the *degree* (or *weighted degree*, if $w \neq 1$) $d(v_i)$ of v_i . Such a quantity is defined as

$$d(v_i) = \sum_{j=1}^N (A_{\mathcal{G}})_{ij} = \sum_{\substack{j=1 \\ j \neq i}}^N \sum_{e \in \mathcal{E}_{ij}} \frac{w(e)}{|e| - 1} = \sum_{e \in \mathcal{E} : e \ni v_i} w(e), \quad v_i \in V. \quad (5.2)$$

See Figure 7 for some examples. Notice that, when $\mathcal{G} = G$ is a graph, the expression (5.1) coincides with the standard definition of the adjacency matrix associated with G .

As announced, we can now generalize the definitions introduced in Section 4. Denoting by D the diagonal matrix with degrees $d(v_1), \dots, d(v_N)$ as entries, we define the *Laplacian matrix* $L_{\mathcal{G}} := D - A_{\mathcal{G}}$ and the *normalized Laplacian matrix* associated with the hypergraph \mathcal{G} as $\mathcal{L}_{\mathcal{G}} := I - D^{-\frac{1}{2}} A_{\mathcal{G}} D^{-\frac{1}{2}}$. We observe that, for non-zero functions $\mathbf{g} = (g(v_1), \dots, g(v_N)) \in$

\mathbb{R}^V and $\mathbf{f} = (f(v_1), \dots, f(v_N)) := D^{-\frac{1}{2}}\mathbf{g}$, the *Raleygh quotient* associated with \mathcal{L}_G at \mathbf{f} is given by the ratio

$$\begin{aligned} \frac{\langle \mathbf{g}, \mathcal{L}_G \mathbf{g} \rangle}{\langle \mathbf{g}, \mathbf{g} \rangle} &= \frac{\langle \mathbf{g}, D^{-\frac{1}{2}} L_G D^{-\frac{1}{2}} \mathbf{g} \rangle}{\langle \mathbf{g}, \mathbf{g} \rangle} = \frac{\langle \mathbf{f}, L_G \mathbf{f} \rangle}{\langle D^{\frac{1}{2}} \mathbf{f}, D^{\frac{1}{2}} \mathbf{f} \rangle} \\ &= \frac{\sum_{v_i \sim v_j} \sum_{e \in \mathcal{E}_{ij}} \frac{w(e)}{|e|-1} (f(v_i) - f(v_j))^2}{\sum_{v_i \in V} d(v_i) f(v_i)^2} =: \mathcal{R}_G(\mathbf{f}). \end{aligned}$$

In particular, recalling the variational characterization of the eigenvalues $\mu_1 \leq \dots \leq \mu_N$ of \mathcal{L}_G (which is valid in our case since \mathcal{L}_G is symmetric; see Remark 4.1), one has that

$$\mu_k = \min_{\substack{\mathcal{M} \subseteq \mathbb{R}^V \\ \dim(\mathcal{M})=k}} \max_{\mathbf{f} \in \mathcal{M} \setminus \{0\}} R_{\mathcal{L}_G}(\mathbf{f}), \quad \forall k = 1, \dots, N, \quad (5.3)$$

from which one infers that $0 = \mu_1 \leq \mu_2 \leq \dots \leq \mu_N \leq 2$.

Given $\emptyset \neq S \subseteq V$, we define

$$\partial S := \{e \in \mathcal{E} : \exists v_i, v_j \in e \text{ with } v_i \in S, v_j \in \bar{S}\} \quad (5.4)$$

to be the set of edges with at least one vertex in S and one in its complement $\bar{S} := V \setminus S$. Moreover, we define the *volume* of S as

$$\text{vol}(S) = w(S) := \sum_{v \in S} d(v) = \sum_{v \in S} \sum_{e \in \mathcal{E}_v} w(e), \quad (5.5)$$

where $\mathcal{E}_v := \{e \in \mathcal{E} : v \in e\}$, while for $\mathcal{F} \subseteq \mathcal{E}$ we set

$$w(\mathcal{F}) := \sum_{e \in \mathcal{F}} w(e). \quad (5.6)$$

The *edge expansion* of a nonempty subset $S \subseteq V$ is given by

$$\varphi(S) := \frac{w(\partial S)}{\text{vol}(S)}, \quad (5.7)$$

and for $k \geq 2$, we set

$$\varphi_k(\mathcal{G}) := \min_{\substack{S_1, \dots, S_k \\ \text{nonempty and disjoint}}} \max_{i=1, \dots, k} \varphi(S_i).$$

Example 5.1. Consider the examples in Figure 7, and set $S = \{1, 2\}$, $\bar{S} = \{3, 4\}$. Then one has the following computations: **(a)** $\varphi(S) = \varphi(\bar{S}) = 2/3$; **(b)** $\varphi(S) = 2/5$, $\varphi(\bar{S}) = 1/2$; **(c)** $\varphi(S) = \varphi(\bar{S}) = 3/5$.

It is now possible to derive an analogue of the ‘‘easy direction’’ of the Cheeger inequality of order k (see Proposition 4.3) for hypergraphs. To this end, we take inspiration from [7, Theorems 3.3 and 4.1], where a bound is established for the hypergraph analogue of $\tilde{\varphi}(G)$ (see (4.5)) in the case $k = 2$.

Proposition 5.2. *Let $\mathcal{G} = (V, \mathcal{E}, w)$ be a weighted hypergraph as above. Then, for all $k = 2, \dots, N$,*

$$\mu_k \leq \frac{2(r(\mathcal{G}) - 1)^2}{cr(\mathcal{G}) - 1} \varphi_k(\mathcal{G}), \quad (5.8)$$

where $r(\mathcal{G})$ and $cr(\mathcal{G})$ are called, respectively, the rank and co-rank of \mathcal{G} and correspond to the maximum and the minimum of the cardinalities of the edges $e \in \mathcal{E}$.

Note that the inequality (5.8) reduces to (4.6) in the graph setting, where $r(\mathcal{G}) = cr(\mathcal{G}) = 2$. We outline the main steps of the proof, which extends to the case of hypergraphs the strategy used in the proof of Proposition 4.3.

Proof of Proposition 5.2. Let $S_1, \dots, S_k \subseteq V$ be non-empty disjoint subsets of V . For $\ell = 1, \dots, k$, let $\mathbf{f}_\ell := (\text{vol}(S_\ell))^{-\frac{1}{2}} \mathbf{1}_{S_\ell}$, where $\mathbf{1}_{S_\ell} := (\mathbb{1}_{S_\ell}(v_1), \dots, \mathbb{1}_{S_\ell}(v_N))$. For all not identically zero vectors $(\alpha_1, \dots, \alpha_k)$, we have

$$\begin{aligned} R_{\mathcal{G}} \left(\sum_{\ell=1}^k \alpha_\ell \mathbf{f}_\ell \right) &= \frac{\sum_{e \in \mathcal{E}} \sum_{\substack{v_i \sim v_j \\ v_i, v_j \in e}} \frac{w(e)}{|e|-1} \left(\sum_{\ell=1}^k \frac{\alpha_\ell}{\sqrt{\text{vol}(S_\ell)}} (\mathbb{1}_{S_\ell}(v_i) - \mathbb{1}_{S_\ell}(v_j)) \right)^2}{\sum_{v_i \in V} d(v_i) \left(\sum_{\ell=1}^k \frac{\alpha_\ell}{\sqrt{\text{vol}(S_\ell)}} \mathbb{1}_{S_\ell}(v_i) \right)^2} \\ &\leq \frac{2 \sum_{\ell=1}^k \frac{\alpha_\ell^2}{\text{vol}(S_\ell)} \sum_{e \in \mathcal{E}} \sum_{\substack{v_i \sim v_j \\ v_i, v_j \in e}} \frac{w(e)}{|e|-1} (\mathbb{1}_{S_\ell}(v_i) - \mathbb{1}_{S_\ell}(v_j))^2}{\sum_{\ell=1}^k \alpha_\ell^2}, \end{aligned}$$

where the last inequality is a consequence of the fact that S_1, \dots, S_k are pairwise disjoint, hence the sum over ℓ contains at most two non-zero terms. Observe indeed that for any $\ell = 1, \dots, k$, the term $(\mathbb{1}_{S_\ell}(v_i) - \mathbb{1}_{S_\ell}(v_j))$ does not vanish if and only if exactly one of the vertices v_i and v_j belongs to S_ℓ . As a consequence, we can rewrite

$$\sum_{v_i \sim v_j} \sum_{e \in \mathcal{E}_{ij}} (\mathbb{1}_{S_\ell}(v_i) - \mathbb{1}_{S_\ell}(v_j))^2 = \sum_{e \in \partial S_\ell} \sum_{\substack{v_i \sim v_j \\ v_i, v_j \in e}} (\mathbb{1}_{S_\ell}(v_i) - \mathbb{1}_{S_\ell}(v_j))^2 = \sum_{e \in \partial S_\ell} |e \cap S_\ell| |e \cap \bar{S}_\ell|,$$

and, for any $e \in \partial S_\ell$, we have $|e \cap S_\ell| |e \cap \bar{S}_\ell| \leq (|e| - 1)^2 \leq (r(\mathcal{G}) - 1)^2$. Using the bound $|e| - 1 \geq cr(\mathcal{G}) - 1$, we eventually obtain that

$$R_{\mathcal{G}} \left(\sum_{\ell=1}^k \alpha_\ell \mathbf{f}_\ell \right) \leq \frac{2(r(\mathcal{G}) - 1)^2}{cr(\mathcal{G}) - 1} \frac{\sum_{\ell=1}^k \alpha_\ell^2 \varphi(S_\ell)}{\sum_{\ell=1}^k \alpha_\ell^2} \leq \frac{2(r(\mathcal{G}) - 1)^2}{cr(\mathcal{G}) - 1} \max_{\ell=1, \dots, k} \varphi(S_\ell).$$

The conclusion now follows from (5.3), following the same route as in the graph setting. \square

As a consequence of Proposition 5.2, we state an estimate analogous to (4.10) for hypergraphs, that one can deduce from arguments similar to those in the proof of Proposition 4.5.

Proposition 5.3. *Let the above assumptions prevail, and fix integers $1 \leq m \leq N$. Consider a collection B_1, \dots, B_m of nonempty disjoint subsets of V , and assume that*

$$\varphi(B_1) \leq \varphi(B_2) \leq \dots \leq \varphi(B_m). \quad (5.9)$$

Then, for all $k = 1, \dots, m$,

$$\mu_k \leq \frac{2(r(\mathcal{G}) - 1)^2}{cr(\mathcal{G}) - 1} \frac{\sum_{i=k}^m \text{vol}(B_i) \varphi(B_i)}{\sum_{i=k}^m \text{vol}(B_i)}. \quad (5.10)$$

5.2. Irreducibility. The aim of this section is to state a generalization of Theorem 2.6 to the setting of homogeneous sums of a generic order $d \geq 2$. To this end, we consider a sequence $\{(V_n, E_n, q_n) : n \geq 1\}$ such that, for a fixed $d \geq 2$: (i) $\{V_n\}$ verifies (2.1), (ii) $\{E_n\}$ satisfies the subsequent requirements (1)–(3), and (iii) the symmetric coefficients $\{q_n\}$ verify (2.2). Following the convention (2.3), we continue to use the notation $w_n(\cdot) = q_n(\cdot)^2$.

For each $n \geq 1$, the triple (V_n, E_n, q_n) is canonically associated with the weighted hypergraph $\mathcal{G}_n = (V_n, \mathcal{E}_n, w_n)$ such that[†]

$$\mathcal{E}_n = \{\{v_1, \dots, v_d\} \subset V_n : (v_1, \dots, v_d) \in E_n\}, \quad \text{and} \quad (5.11)$$

$$w_n(\{v_1, \dots, v_d\}) = w_n(v_1, \dots, v_d) = q_n(v_1, \dots, v_d)^2. \quad (5.12)$$

Note that, by construction, each \mathcal{G}_n is d -uniform, and consequently $r(\mathcal{G}_n) = cr(\mathcal{G}_n) = d$ (see Proposition 5.2). Also, one has that $|E_n| = d!|\mathcal{E}_n|$. For $n \geq 1$, the Laplace spectrum of \mathcal{G}_n is written $0 = \mu_1^{(n)} \leq \mu_2^{(n)} \leq \dots \leq \mu_{N_n}^{(n)} \leq 2$.

As before, we denote by $\{Z_n\} = \{Z_n(\mathbf{X}) : n \geq 1\}$ the sequence of homogeneous sums defined in (2.4), and we observe that the second relation in (2.5) implies that

$$\mathbb{E}[Z_n^2] = d!^2 w_n(\mathcal{E}_n), \quad (5.13)$$

where we have used the notation (5.6) in the case $\mathcal{F} = \mathcal{E}_n$.

In order to generalize Lemma 4.6 in the context of hypergraphs, we define the set of edges entirely contained in a nonempty subset $S \subseteq V$ as follows:

$$\mathcal{E}_n(S, S) := \{e \in \mathcal{E}_n : e \subset S\}. \quad (5.14)$$

The following statement is the exact analogous of Lemma 4.6. The proof is left to the reader.

Lemma 5.4 (Reducibility and hypergraphs). *Let the above assumptions and conventions prevail. Then, the sequence $\{Z_n\}$ is reducible in the sense of Definition 2.1 if and only if there exist partitions $\{B_1, \dots, B_{m_n}\} = \{B_1^{(n)}, \dots, B_{m_n}^{(n)}\}$ of the vertex sets V_n , such that, as $n \rightarrow \infty$,*

- (i) $m_n \rightarrow \infty$;
- (ii) $\sum_{i=1}^{m_n} w_n(\mathcal{E}(B_i, B_i)) \sim w_n(\mathcal{E}_n)$;
- (iii) $\max_{i=1, \dots, m_n} w_n(\mathcal{E}(B_i, B_i)) = o(w_n(\mathcal{E}_n))$.

Definition 5.5. Let $\mathcal{G}_n = (V_n, \mathcal{E}_n, w_n)$, $n \geq 1$, be a sequence of weighted hypergraphs such that $|V_n|, |\mathcal{E}_n|, w_n(\mathcal{E}_n) \rightarrow \infty$. Consider a sequence $\{Z_n\}$ defined as in (2.4) such that (5.11)–(5.12) are verified. We say that $\{Z_n\}$ is a *sequence of homogeneous sums associated with $\{\mathcal{G}_n\}$* . As before, we say that $\{\mathcal{G}_n\}$ *generates an irreducible CLT*, if $\{Z_n\}$ verifies an irreducible CLT in the sense of Definition 2.1. If $\{Z_n\}$ is reducible, we will say that $\{\mathcal{G}_n\}$ *is reducible* or, more precisely, that $\{\mathcal{G}_n\}$ *generates a reducible CLT*.

Remark 5.6. A natural open question is whether, for $d \geq 3$, one can obtain a spectral characterization of Condition (ii) in Theorem 3.1, analogous to the necessary and sufficient condition (3.2) established in the case $d = 2$, but now based on the spectral analysis of hypergraphs developed in the previous section. We refer the reader to the breakthrough work by Herry, Malicet, and Poly [49] for significant progress in this direction.

We state the main result of the section, which establishes the sufficient condition for irreducibility.

[†]Due to the symmetry of E_n , the way in which the elements of a given d -subset $\{v_1, \dots, v_d\}$ are enumerated is immaterial.

Theorem 5.7. Fix $d \geq 2$ and let $\{Z_n\}$ be a sequence of homogeneous sum as in (2.4). Let $\mathcal{G}_n = (V_n, \mathcal{E}_n, w_n)$, $n \geq 1$, be the sequence of weighted hypergraphs associated to $\{Z_n\}$ via (5.11)–(5.12). If there exists $k \geq 2$ such that, as $n \rightarrow \infty$,

$$\liminf_n \mu_k^{(n)} > 0, \quad (5.15)$$

then $\{Z_n\}$ is irreducible in the sense of Definition 2.1.

The proof of Theorem 5.7 is provided in Appendix A: it consists of a technical generalization of the proof of Theorem 2.6.

Example 5.8 (Rook-like hypergraphs). For every $d \geq 3$, we will now build an example of a sequence of homogeneous sums $\{Z_n\}$, of order $d \geq 3$, and such that (i) they verify an irreducible CLT, (ii) irreducibility follows from Theorem 5.7 via the fact that the adjacency matrices (5.1) of the associated weighted hypergraphs (via (5.11)–(5.12)) coincide with those of the Rook's graph discussed in Example (b)-(i) of Section 3.4. To this end, for every $n > d$ large enough, define $\mathcal{G}_n = (V_n, \mathcal{E}_n, w_n)$ to be the hypergraph such that $V_n = [n]^d$, \mathcal{E}_n is given by all d -subset with the form

$$e = \{(a, b_1), (a, b_2), \dots, (a, b_d)\} \quad \text{or} \quad e = \{(b_1, a), (b_2, a), \dots, (b_d, a)\}$$

with $a \in [n]$ and b_1, \dots, b_d distinct, and we choose a constant weight (for convenience, we fix the weight so that it yields the same adjacency matrix of the Rook's graph)

$$w_n(e) \equiv (d-1) \binom{n-2}{d-2}^{-1}, \quad e \in \mathcal{E}_n.$$

It is easily seen that two vertices $v_1 = (a_1, b_1)$, $v_2 = (a_2, b_2)$ are adjacent in \mathcal{G}_n if and only if they are adjacent in the Rook's graph and, in this case, they are both contained in exactly $\binom{n-2}{d-2}$ edges. The choice of w_n then ensures that the adjacency matrix of \mathcal{G}_n , as given in (5.1), coincides with that of the Rook's graph. Writing $0 = \mu_1^{(n)} \leq \mu_2^{(n)} \leq \dots \leq \mu_{n^d}^{(n)}$ for the normalized Laplace eigenvalues of \mathcal{G}_n , one therefore has that $\mu_k^{(n)} \rightarrow \frac{1}{2}$ for all $k \geq 2$, and Theorem 5.7 implies that any associated sequence of homogeneous sums $\{Z_n\}$ (via (5.11)–(5.12)) is irreducible. We observe that

$$|\mathcal{E}_n| = 2n \binom{n}{d} \asymp n^{d+1}, \quad n \rightarrow \infty.$$

We now define $\{Z_n\}$ according to (2.4), with

$$q_n(v_1, \dots, v_d) = \begin{cases} \sqrt{d-1} \binom{n-2}{d-2}^{-1/2}, & \{v_1, \dots, v_d\} \in \mathcal{E}_n \\ 0, & \text{otherwise.} \end{cases}$$

In this way, $\mathbb{E}[Z_n^2] \asymp n$, and a direct computation shows that the numerical sequences defined in (3.1) scale as $O(n^{-1})$ for every $r = 1, \dots, d-1$, and thus $\{Z_n\}$ verifies an irreducible CLT.

Example 5.9 (3-uniform hypergraphs generating an irreducible CLT). We present the following example as an application of Theorem 5.7. Consider the vertex set

$$V_n = \{(a, b) : 1 \leq a \neq b \leq n\} \subseteq \mathbb{N} \times \mathbb{N}$$

in such a way that $N_n = |V_n| = n(n-1)$. For every n , we define the symmetric and non-diagonal set $\mathcal{E}_n \subseteq (V_n)^3$ consisting of all ordered triples (v_1, v_2, v_3) , where each $v_i = (a_i, b_i)$ is an element of V_n , satisfying the following conditions:

- For every pair $\ell \neq k$, the entries v_ℓ and v_k share exactly one coordinate: $|\{a_\ell, b_\ell\} \cap \{a_k, b_k\}| = 1$.
- The number of distinct labels among $\{a_1, b_1, a_2, b_2, a_3, b_3\}$ equals 3.

Note that E_n is the disjoint union of eight sets $E_n^{(1)}, \dots, E_n^{(8)}$, each one with a similar structure and of the same size as

$$E_n^{(1)} = \left\{ ((a, b), (b, c), (c, a)) : 1 \leq a \neq b \neq c \leq n \right\}.$$

Thus, $|E_n| = 8n(n-1)(n-2)$, and the hypergraph $\mathcal{G}_n = (V_n, \mathcal{E}_n)$, obtained from $\{E_n\}$ by setting $q_n \equiv 1$ and using (5.11), is unweighted and 3-uniform. For $n \geq 3$, we consider the homogeneous sum

$$Z_n := \sum_{v_1, v_2, v_3 \in V_n} \mathbb{1}_{E_n}(v_1, v_2, v_3) X_{v_1} X_{v_2} X_{v_3},$$

whose variance is commensurate to n^3 , as $n \rightarrow \infty$. To determine the irreducibility of $\{Z_n\}$, we observe that $\mathcal{G}_n = (V_n, \mathcal{E}_n)$ has adjacency matrix

$$(A_{\mathcal{G}_n})_{ij} = \sum_{e \in \mathcal{E}_{ij}} \frac{1}{|e| - 1} = \frac{1}{2} |\{e \in \mathcal{E}_n : v_i, v_j \in e\}|, \quad v_i, v_j \in V_n, \quad i \neq j.$$

In particular, when v_i and v_j are adjacent, say $v_i = (a, b)$ and $v_j = (b, c)$, they are contained in exactly two edges, which have (a, c) and (c, a) as their third vertices. This implies that $|\mathcal{E}_{ij}| = 2$, thus $(A_{\mathcal{G}_n})_{ij} = \mathbb{1}_{\{v_i \sim v_j\}}$ for every $v_i, v_j \in V_n$, where the adjacency relation \sim is given by

$$v_i \sim v_j \iff v_i = (a, b), v_j = (a', b') \text{ and either } \begin{cases} a = a', b \neq b', \text{ or} \\ b = b', a \neq a', \text{ or} \\ a = b', a' \neq b, \text{ or} \\ a' = b, a \neq b'. \end{cases}$$

One can verify that this equivalence relation induces a graph structure on the vertex set V_n , corresponding to the union of two graphs: the Rook's graph (see Example 3.3-(b)) with the diagonal removed, and a second graph—also isomorphic to the diagonal-free Rook's graph—obtained by permuting the two coordinates of each vertex. A direct computation yields that the degree of each vertex is $d_n = 4(n-2)$, and the spectrum of $A_{\mathcal{G}_n}$ is given by the integers $\{4(n-2), (2n-8), 0, -4\}$ with multiplicities $1, (n-1), n(n-1)/2$ and $-1 + (n-1)(n-2)/2$. As a consequence, the Laplace spectrum of \mathcal{G}_n is given by $\{0, n/(2(n-2)), 1, (n-1)/(n-2)\}$, with the same respective multiplicities. Since, for all fixed $k \geq 2$, one has that $\lim_{n \rightarrow \infty} \mu_k^{(n)} = \frac{1}{2}$, we deduce from Theorem 5.7 that $\{Z_n\}$ is irreducible. Finally, we remark that this example will be revisited in the subsequent section, where we will show that the hyperedge sets $\{\mathcal{E}_n\}$ can be realized as a special cases of a *fractional Cartesian product*, as defined in Example 6.1. As such, the asymptotic normality of the (normalized) sequence $\{Z_n\}$ will follow directly from Proposition 6.2.

6. Irreducibility via combinatorial dimensions

The aim of this section is to prove Theorem 2.11. The proof of Part (b) (see Section 6.2) is based on an explicit construction, described in full detail in Example 6.1 below.

The arguments used in this part are purely combinatorial and do not rely on the spectral analysis developed in the previous sections. We note, however, that one special instance of Example 6.1 was already addressed in Section 5 by means of hypergraph techniques (see Example 5.9).

From now on, we let the assumptions and notation in the statement of Theorem 2.11 prevail.

6.1. Proof of Part (a) of Theorem 2.11. Assume that $\{E_n\}$ has combinatorial dimension $1 < \alpha \leq d$. We have to show that it is not possible to find a sequence $\Pi_n := \{B_1, \dots, B_{m_n}\}$, $n \geq 1$, with $B_i = B_i(n)$, such that each Π_n is a partition of V_n , and, as $n \rightarrow \infty$,

$$\max_{i=1, \dots, m_n} \left| E_n \cap \underbrace{(B_i \times \dots \times B_i)}_{d \text{ times}} \right| = o(|E_n|), \quad (6.1)$$

$$\sum_{i=1}^{m_n} \left| E_n \cap \underbrace{(B_i \times \dots \times B_i)}_{d \text{ times}} \right| \asymp |E_n|, \quad (6.2)$$

(note that, if (6.1) and (6.2) are both verified, then, necessarily, $m_n \rightarrow \infty$).

We reason by contradiction, and assume that there exists a sequence of partitions $\{\Pi_n\}$ such that the two properties (6.1) and (6.2) are satisfied. Without loss of generality, we can assume that $|B_1| \geq |B_2| \geq \dots \geq |B_{m_n}|$, $n \geq 1$. For all $\beta > 0$, we set

$$K_n(\beta) := \max \{s \leq m_n : |B_s| \geq \beta |V_n|\},$$

with $\max \emptyset := 0$. Since each Π_n is a partition of V_n we have that

$$|V_n| = \sum_{s=1}^{m_n} |B_s| \geq \sum_{s=1}^{K_n(\beta)} |B_s^{(n)}| \geq K_n(\beta) \beta |V_n|,$$

where we have used the convention $\sum_{s=1}^0 := 0$, thus yielding the bound

$$K_n(\beta) \leq \frac{1}{\beta}, \quad \text{for all } \beta > 0. \quad (6.3)$$

In the sequel, we write $M(\beta) := \lceil \beta^{-1} \rceil + 1$, $\beta > 0$. We will also use the following estimate, valid for every $n \geq 1$ and every $t \leq m_n$:

$$\sum_{s=t}^{m_n} |B_s|^\alpha \leq |B_t|^{\alpha-1} \sum_{s=t}^{m_n} |B_s| \leq |B_t|^{\alpha-1} |V_n|, \quad (6.4)$$

where the first inequality uses the fact that $\alpha \in (1, d]$ and that the sequence $s \mapsto |B_s|$ is decreasing. Since (6.2) is in order and $\{E_n\}$ has combinatorial dimension α , using (2.15) for $J_n = E_n$ one infers that there exist a finite constant $\Gamma > 0$ and an integer $n_0 \geq 1$ such that, for all $n \geq n_0$,

$$\Gamma > \frac{|V_n|^\alpha}{\sum_{s=1}^{m_n} |E_n \cap (B_s \times \dots \times B_s)|} \geq \frac{|V_n|^\alpha}{c \sum_{s=1}^{m_n} |B_s|^\alpha} \geq \frac{|V_n|^{\alpha-1}}{c |B_1|^{\alpha-1}},$$

where the constant c is the one appearing on the right-hand side of (2.13) (choosing $J_n = E_n$), and the last estimate exploits (6.4) in the case $t = 1$. This implies that, for

$n \geq n_0$, $|B_1| \geq b|V_n|$, where $b := (c\Gamma)^{-\frac{1}{\alpha-1}}$. Now, assumption (6.1) combined once again with (2.15) (with $J_n = E_n$) yields that, as $n \rightarrow \infty$,

$$\frac{1}{|V_n|^\alpha} \sum_{s=1}^{M(b)-1} |E_n \cap (B_s \times \cdots \times B_s)| \rightarrow 0.$$

From this we infer that there exists an integer $n_1 \geq n_0$ such that, for all $n \geq n_1$

$$\Gamma > \frac{|V_n|^\alpha}{\sum_{s=M(b)}^{m_n} |E_n \cap (B_s \times \cdots \times B_s)|} \geq \frac{|V_n|^\alpha}{c \sum_{s=M(b)}^{m_n} |B_s|^\alpha} \geq \frac{|V_n|^{\alpha-1}}{c|B_{M(b)}|^{\alpha-1}},$$

where the last inequality uses (6.4) in the case $t = M(b)$. These relations imply that, for $n \geq n_1$, one necessarily has $|B_{M(b)}| \geq b|V_n|$, which is absurd, since it would yield

$$\frac{1}{b} < M(b) \leq K_n(b) \leq \frac{1}{b},$$

where we used (6.3). The proof of Part (a) is concluded.

6.2. Proof of Part (b) of Theorem 2.11. The following example shows that, at least for $d \geq 3$, one can easily build examples of sequences $\{E_n\}$ that have combinatorial dimensions strictly between 1 and d .

Example 6.1 (Fractional Cartesian products). The following construction is a variation of the definition of *fractional Cartesian products*, as discussed in [11, Chapters XII-XIII] and [13, 38, 71]. Fix $d \geq 3$ and $b = 1, \dots, d-1$. For every $n > b$, we define $V_n := \{\mathbf{v} = (v_1, \dots, v_b) \in [n]^b : v_s \neq v_t, s \neq t\}$. In what follows, generic elements $(\mathbf{v}_1, \dots, \mathbf{v}_d)$ of the Cartesian product $\underbrace{V_n \times \cdots \times V_n}_{d \text{ times}}$ will be written

$$(\mathbf{v}_1, \dots, \mathbf{v}_d) = (v_{(1,1)}, v_{(1,2)}, \dots, v_{(1,b)}; v_{(2,1)}, \dots, v_{(2,b)}; \dots; v_{(d,1)}, \dots, v_{(d,b)}), \quad (6.5)$$

where $(v_{(s,1)}, v_{(s,2)}, \dots, v_{(s,b)}) = \mathbf{v}_s \in V_n$, $s = 1, \dots, d$. We now fix a partition $S = \{S_1, \dots, S_d\}$ of the index set $I := \{(1,1), (1,2), \dots, (d,b)\}$ with the following properties:

- (a) $|S_i| = b$, $i = 1, \dots, d$;
- (b) for every $\ell = 1, \dots, d$, each S_i contains at most one index of the form (ℓ, s) , for $s = 1, \dots, b$.

We will also say that the partition S is *connected* if there do not exist partitions $\{B_1, B_2\}$ and $\{C_1, C_2\}$ of $[d]$ such that, for $i = 1, 2$

$$\{(\ell, s) : \ell \in B_i\} = \bigcup_{j \in C_i} S_j.$$

See Figure 8 for examples of such partitions.

For every $n > d$ we define the set $E_n^0 \subset V_n \times \cdots \times V_n$ as follows:

$$E_n^0 := \{(\mathbf{v}_1, \dots, \mathbf{v}_d) : v_{(s,u)} = v_{(t,v)} \text{ iff } (s,u), (t,v) \in S_i \text{ for some } i = 1, \dots, d\}, \quad (6.6)$$

where we used the notation (6.5). Finally, we define E_n to be the collection of those $(\mathbf{v}_1, \dots, \mathbf{v}_d)$ such that $(\mathbf{v}_{\sigma(1)}, \dots, \mathbf{v}_{\sigma(d)}) \in E_n^0$, for some permutation σ of $[d]$. In this way, each E_n automatically satisfies the properties (1) and (2) (symmetry), introduced at the beginning of Section 2.1. It is easily seen that a connected partition S with the properties (a) and (b) listed above always exists; moreover, standard combinatorial considerations yield that such an S can always be chosen so that each E_n also verifies (3) from the beginning

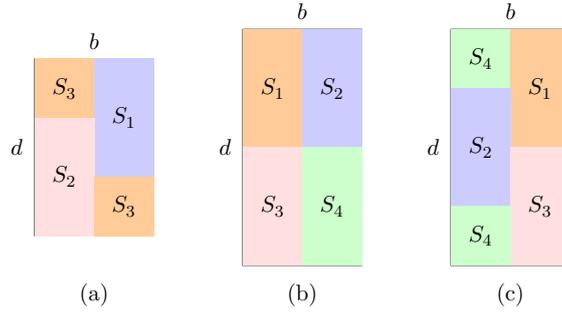


Figure 8. Some examples of partitions S of the index set I . Case (a): for $d = 3$ and $b = 2$ there are only four partitions $S = \{S_1, S_2, S_3\}$ verifying (a) and (b) and all are connected; in the example, we have $S_1 = \{(1, 2), (2, 2)\}$, $S_2 = \{(2, 1), (3, 1)\}$ and $S_3 = \{(3, 2), (1, 1)\}$. Case (b): for $d = 4$ and $b = 2$, an example of disconnected partition $S = \{S_1, S_2, S_3, S_4\}$ verifying (a) and (b) is given by $S_1 = \{(1, 1), (2, 1)\}$, $S_2 = \{(1, 2), (2, 2)\}$, $S_3 = \{(3, 1), (4, 1)\}$ and $S_4 = \{(3, 2), (4, 2)\}$ (for which $B_1 = C_1 = \{1, 2\}$ and $B_2 = C_2 = \{3, 4\}$). Case (c): for $d = 4$ and $b = 2$, an example of a connected partition is given by $S_1 = \{(1, 2), (2, 2)\}$, $S_2 = \{(2, 1), (3, 1)\}$, $S_3 = \{(3, 2), (4, 2)\}$ and $S_4 = \{(1, 1), (4, 1)\}$.

of Section 2.1, that is: E_n is non-diagonal. We claim that the sequence $\{E_n : n > d\}$ has combinatorial dimension $\alpha = \frac{d}{b}$, as per Definition 2.8. To see this, we first observe that, by definition, as $n \rightarrow \infty$, one has that $|V_n| \asymp n^b$, and

$$|E_n^0| \asymp |E_n| \asymp n^d \asymp |V_n|^{\frac{d}{b}},$$

proving that (2.14) is satisfied for $J_n = E_n$. On the other hand (e.g. by the triangle inequality), to prove (2.14) for $J_n = E_n$ it is sufficient to show that

$$|E_n^0 \cap (A_1 \times \cdots \times A_d)| \leq c \max_{i=1, \dots, d} |A_i|^{\frac{d}{b}}, \quad (6.7)$$

for some absolute constant c . It turns out that such a bound is a direct consequence of a classical estimate by Finner, stated in [43, Theorem 2.1] (see also [8, Proposition 2.8]). We refer to Appendix A.4 for a complete proof of (6.7).

The following statement shows that the sequence of homogeneous sums associated with the irreducible sequence $\{E_n\}$ verifies a CLT, thus concluding the proof of Part (b) of Theorem 2.11.

Proposition 6.2 (Fractional products yield irreducible CLTs). *For $d \geq 3$, let $\{E_n : n > k\}$ be the sequence of sets constructed in Example 6.1. Assume in addition that the underlying partition $S = \{S_1, \dots, S_k\}$ is connected, and verifies Properties (a) and (b) in Example 6.1. For $n > d$, we consider the sequence $\{Z_n\}$ defined as in (2.4) for $q_n \equiv 1$. Then, as $n \rightarrow \infty$, $\mathbb{E}[Z_n^2] = d!|E_n| \asymp n^d$, and the sequence $\tilde{Z}_n = \frac{Z_n}{\sqrt{d!|E_n|}}$ converges in distribution to a standard Gaussian random variable N .*

Proof. To show that \tilde{Z}_n converges in distribution to N , we can directly use [72, Proof of Proposition 6.6 and Proposition 6.8] to deduce that, for every thrice differentiable bounded function $h : \mathbb{R} \rightarrow \mathbb{R}$ with bounded derivatives,

$$|\mathbb{E}[h(\tilde{Z}_n)] - \mathbb{E}[h(N)]| \leq \frac{C}{|V_n|^{\frac{1}{2b}}}, \quad (6.8)$$

for some absolute constant C , so that the conclusion follows from the fact that $|V_n| \rightarrow \infty$. \square

Remark 6.3. The estimate (6.8) follows from [72, formula (6.72)], implying that the quantity $|\mathbb{E}[h(\tilde{Z}_n)] - \mathbb{E}[h(N)]|$ is bounded by a multiple of the sum

$$\frac{|E_n^\#|^{1/2}}{|E_n|} + \frac{\max_{\mathbf{i} \in V_n} |E_{n,\mathbf{i}}^*|^{1/4}}{|E_n|^{1/4}},$$

where $E_{n,\mathbf{i}}^* := \{(\mathbf{v}_1, \dots, \mathbf{v}_d) \in E_n : \mathbf{i} = \mathbf{v}_\ell \text{ for one } \ell \in [d]\}$, and $E_n^\# \subset E_n \times E_n$ is defined as the collection of all pairs $(F, G) := ((\mathbf{f}_1, \dots, \mathbf{f}_d), (\mathbf{g}_1, \dots, \mathbf{g}_d)) \in E_n \times E_n$ such that: (a) F and G have no elements in common, and (b) there exists $p \in [d-1]$, as well as distinct integers $\ell_1, \dots, \ell_p \in [d]$ such that replacing each \mathbf{f}_{ℓ_i} in F with the corresponding \mathbf{g}_{ℓ_i} in G (and vice versa) results in a (possibly different) element of $E_n \times E_n$. The claimed estimate (6.8) then follows from [72, formula (6.70)], implying that, for some absolute constant C , one has that $|E_n^\#| \leq C|V_n|^{2\alpha - \frac{1}{b}}$ and $\max_{\mathbf{i} \in V_n} |E_{n,\mathbf{i}}^*| \leq C|V_n|^{\alpha-1}$.

7. Irreducibility and sparsity: ad hoc construction for $d = 2$

This section provides the proof of Theorem 2.13, under the assumptions stated in Subsection 2.4.2. In particular, we prove a refined version of Theorem 2.13 (see Theorem 7.1 below). We not only show that any partition B_1, \dots, B_{m_n} of V_n satisfying (i) and (ii) cannot satisfy (iii) (this already implies irreducibility; see Definition 2.1 and Remark 2.3-(1)), but we also provide a quantitative bound contradicting (iii). In more detail, we provide a lower bound away from zero for the main contribution to the second moment coming from one of the sets in the partition (see inequality (7.5)), which is in conflict with (iii).

We briefly recall the notation of Section 2.4.2. For $n \in \mathbb{N}$, we set $V_n = [n]^2$, fix $\beta \in (0, 1]$ and for $a, b \in \{1, \dots, n\}$ consider the subsets $\mathbb{S}_v(a) \subseteq \{a\} \times \{1, \dots, n\}$ and $\mathbb{S}_h(b) \subseteq \{1, \dots, n\} \times \{b\}$ with $|\mathbb{S}_v(a)| = |\mathbb{S}_h(b)| = \lfloor \beta n \rfloor$.

We recall the equivalence relations $\overset{h}{\sim}$ and $\overset{v}{\sim}$ on V_n : for any $v_1, v_2 \in V_n$

$$v_1 \overset{h}{\sim} v_2 \iff \text{for some } b \text{ we have } v_1, v_2 \in \mathbb{S}_h(b) \text{ with } v_1 \neq v_2, \quad (7.1)$$

$$v_1 \overset{v}{\sim} v_2 \iff \text{for some } a \text{ we have } v_1, v_2 \in \mathbb{S}_v(a) \text{ with } v_1 \neq v_2, \quad (7.2)$$

and we write $v_1 \sim v_2$ if and only if $v_1 \overset{h}{\sim} v_2$ or $v_1 \overset{v}{\sim} v_2$. We set $E_n := \{(v_1, v_2) \in V_n \times V_n : v_1 \sim v_2\}$ and we consider the homogeneous sum

$$Z_n = \sum_{v_1, v_2 \in V_n} \mathbf{1}_{E_n}(v_1, v_2) X_{v_1} X_{v_2}, \quad (7.3)$$

where $\{X_v\}_{v \in \bigcup_n V_n}$ is a family of i.i.d. standard Gaussian random variables. We recall that Z_n is centered with second moment

$$\mathbb{E}[Z_n^2] = 2|E_n| = 2 \left\{ \sum_{b=1}^n |\mathbb{S}_h(b)| (|\mathbb{S}_h(b)| - 1) + \sum_{a=1}^n |\mathbb{S}_v(a)| (|\mathbb{S}_v(a)| - 1) \right\} \underset{n \rightarrow \infty}{\sim} 4\beta^2 n^3.$$

The following result establishes irreducibility for the sequence $\{Z_n\}$.

Theorem 7.1 (Irreducibility and quantitative bounds). *Fix any $\beta \in (\frac{1}{2}, 1]$. Then*

(1) Any partition $\{B_\ell = B_\ell(n, m_n)\}_{\ell=1, \dots, m_n}$, which satisfies as $n \rightarrow \infty$

$$\sum_{\ell=1}^{m_n} \sigma_n^2(B_\ell) = (4\beta^2 + o(1)) n^3 \quad (7.4)$$

must also satisfy

$$\max_{\ell=1, \dots, m_n} \sigma_n^2(B_\ell) \geq (2\beta^2(2\beta - 1) + o(1)) n^3. \quad (7.5)$$

(2) As a consequence, $\{Z_n\}$ satisfies an irreducible CLT in the sense of Definition 2.1.

Let us prove point (2), while the (longer) proof of the quantitative bound is presented separately in Subsection 7.1.

Proof of Theorem 7.1-(2). Assuming (1) in Theorem 7.1, the conditions (i), (ii) and (iii) in Definition 2.1 cannot hold together, thus $\{Z_n\}$ is irreducible. We only need to prove the convergence in distribution towards $N \sim \mathcal{N}(0, 1)$ by applying the Fourth Moment Theorem (see [71, Proposition 1.6] and [22, Theorem 4.2]). Recall $\tilde{Z}_n := Z_n / \sqrt{2|E_n|}$. Since $\lim_{n \rightarrow \infty} \mathbb{E}[\tilde{Z}_n^2] = 1$, we only need to prove that $\lim_{n \rightarrow \infty} \mathbb{E}[\tilde{Z}_n^4] = 3$. We have

$$\begin{aligned} \mathbb{E}[\tilde{Z}_n^4] &= \frac{1}{4|E_n|^2} \sum_{v_1, \dots, v_8 \in V_n} \mathbb{1}_{E_n}(v_1, v_2) \mathbb{1}_{E_n}(v_3, v_4) \mathbb{1}_{E_n}(v_5, v_6) \mathbb{1}_{E_n}(v_7, v_8) \mathbb{E}\left[\prod_{i=1}^8 X_{v_i}\right] \\ &= \frac{4}{|E_n|^2} \sum_{A, B, C, D \subseteq V_n} q_n(A) q_n(B) q_n(C) q_n(D) \mathbb{E}[X_A X_B X_C X_D], \end{aligned} \quad (7.6)$$

where for convenience we rearranged the sum over (*unordered*) subsets $A = \{v_1, v_2\}$, $B = \{v_3, v_4\}$, $C = \{v_5, v_6\}$, $D = \{v_7, v_8\}$ (recall that each vertex v_i is an *ordered* couple in the square $[n]^2$) and for each of these subsets, say $A \subseteq V_n$, we set $q_n(A) := \mathbb{1}_{E_n}(v_1, v_2)$ and $X_A := X_{v_1} X_{v_2}$.

Recall that the X_v 's are centered and independent and that all pairs $A, B, C, D \subseteq V_n$ contain two *distinct* vertices (the diagonal points of $V_n \times V_n$ are not in E_n). Therefore, the only non-zero contribution to $\mathbb{E}[X_A X_B X_C X_D] = \mathbb{E}[X_{v_1} X_{v_2} X_{v_3} X_{v_4} X_{v_5} X_{v_6} X_{v_7} X_{v_8}]$ is given by those terms where the individual X_v 's match either in pairs or in quadruples. We outline the possible cases below, recalling that $2|E_n| \sim 4\beta^2 n^3$ as $n \rightarrow \infty$.

(1) The X_v 's match in quadruples, i.e. $A = B = C = D$, and the contribution to (7.6) is negligible as $n \rightarrow \infty$:

$$\begin{aligned} 4|E_n|^{-2} \sum_{A \subseteq V_n} q_n(A)^4 \mathbb{E}[X_A^4] &= 4|E_n|^{-2} \sum_{\{v_1, v_2\} \subseteq V_n} \mathbb{1}_{E_n}(v_1, v_2) \mathbb{E}[X_{v_1}^4 X_{v_2}^4] \\ &= 18|E_n|^{-1} = O(n^{-3}). \end{aligned}$$

(2) The X_v 's match in pairs, however $A, B, C, D \subseteq V_n$ do not pair up two by two, for instance when $A = \{v_1, v_2\}$, $B = \{v_1, v_3\}$, $C = \{v_3, v_4\}$ and $D = \{v_2, v_4\}$. It is simple to see that their contribution to (7.6) is always either $O(n^{-1})$ or $O(n^{-2})$. To give a brief idea, let us consider the term when \sim is always \simeq , thus the indicator function of E_n imposes that $v_1 = (a_1, b_1)$, $v_2 = (a_1, b_2)$, $v_3 = (a_1, b_3)$ and $v_4 = (a_1, b_4)$. The corresponding term is then negligible for large n , indeed

$$4|E_n|^{-2} \sum_{v_1, \dots, v_4 \in V_n} \mathbb{1}_{E_n}(v_1, v_2) \mathbb{1}_{E_n}(v_1, v_3) \mathbb{1}_{E_n}(v_3, v_4) \mathbb{1}_{E_n}(v_2, v_4) \leq O(n^{-1}),$$

since the sum above can be bounded by the sum over the five degrees of freedom $a_1, b_1, b_2, b_3, b_4 \in \{1, \dots, n\}$. The other cases work similarly.

- (3) The X_v 's match in pairs and $A, B, C, D \subseteq V_n$ pair up two by two, i.e. either $A = B$ and $C = D$, or $A = C$ and $B = D$, or $A = D$ and $C = B$. If two distinct couples $A, B \subseteq V_n$ differ from each other but have a common element $v \in V_n$, the corresponding term can be treated similarly as in the previous case, thus giving a negligible contribution. Therefore, consider the case where all distinct couples $A, B \subseteq V_n$ are also disjoint: this is the only non-negligible contribution and gives exactly

$$3 \left(4|E_n|^{-2} \sum_{\substack{A, B \subseteq V_n \\ A \cap B = \emptyset}} q(A)^2 q(B)^2 \mathbb{E}[X_A^2 X_B^2] \right) = 3 \left(4|E_n|^{-2} \left(\sum_{\{v_1, v_2\} \subseteq V_n} \mathbb{1}_{E_n}(v_1, v_2) \right)^2 \right) = 3.$$

□

7.1. Proof of Theorem 7.1-(1). We divide the proof into three steps, which we first outline below.

7.1.1. STRATEGY OF THE PROOF. We assume that (7.4) holds. We fix $\eta \in (0, 1)$ small and we take n large enough so that (the factor $\frac{1}{2}$ is for later convenience)

$$\sum_{\ell=1}^{m_n} \sigma_n^2(B_\ell) \geq (4\beta^2 - \frac{1}{2}\eta) n^3. \quad (7.7)$$

We are going to show that there is $\bar{\ell} = \bar{\ell}_{n, \eta} \in \{1, \dots, m_n\}$ such that

$$\sigma_n^2(B_{\bar{\ell}}) \geq 2(\beta^2(2\beta - 1) - 11\sqrt{\eta} + O(\frac{1}{n})) n^3. \quad (7.8)$$

(The factor 11 multiplying $\sqrt{\eta}$ is immaterial, but it will be transparent to carry out explicit computations.) Since we can take $\eta > 0$ as small as we wish, this proves (7.5).

It remains to prove (7.8). Given a subset $A \subseteq V_n$, we denote its ‘‘rows’’ by

$$A(\cdot, b) := A \cap (\{1, \dots, n\} \times \{b\}) \quad \text{for } b \in \{1, \dots, n\}, \quad (7.9)$$

and similarly we denote its ‘‘columns’’ by

$$A(a, \cdot) := A \cap (\{a\} \times \{1, \dots, n\}) \quad \text{for } a \in \{1, \dots, n\}. \quad (7.10)$$

To help explanations, we refer to the labels $\ell \in \{1, \dots, m_n\}$ of the partition $\{B_\ell\}$ as *colors*. We are going to prove the following three steps.

- (1) *Almost each row and column has a ‘‘dominant’’ color, almost filling up $\mathbb{S}_h(b)$ or $\mathbb{S}_v(a)$:*

$\exists I, I' \subseteq \{1, \dots, n\}$ such that

$$|I| \geq (1 - \frac{1}{4}\sqrt{\eta}) n \quad \text{and} \quad \forall b \in I \quad \exists \ell_b: \quad |B_{\ell_b}(\cdot, b) \cap \mathbb{S}_h(b)| \geq (\beta - 2\sqrt{\eta}) n, \quad (7.11)$$

$$|I'| \geq (1 - \frac{1}{4}\sqrt{\eta}) n \quad \text{and} \quad \forall a \in I' \quad \exists \ell'_a: \quad |B_{\ell'_a}(a, \cdot) \cap \mathbb{S}_v(a)| \geq (\beta - 2\sqrt{\eta}) n. \quad (7.12)$$

- (2) *There is a color $\bar{\ell}$ which is dominant for a positive fraction of rows:*

$$\exists \bar{\ell}: \quad |\{b \in I: \ell_b = \bar{\ell}\}| \geq (\beta' - 7\sqrt{\eta}) n \quad \text{where } \beta' := 2\beta - 1 > 0. \quad (7.13)$$

- (3) *Color $\bar{\ell}$ fulfils relation (7.8).*

7.1.2. STEP 1. By symmetry, we only prove (7.11). We argue in three parts.

Part A. Recalling (2.8), for $B \subseteq V_n$ we define $\sigma_{n,h}^2(B)$ and $\sigma_{n,v}^2(B)$ by

$$\sigma_n^2(B) = 2 \underbrace{\sum_{v_1, v_2 \in B} \mathbb{1}_{v_1 \overset{h}{\sim} v_2}}_{\sigma_{n,h}^2(B)} + 2 \underbrace{\sum_{v_1, v_2 \in B} \mathbb{1}_{v_1 \overset{v}{\sim} v_2}}_{\sigma_{n,v}^2(B)}. \quad (7.14)$$

We prove in this part that

$$\sum_{\ell=1}^{m_n} \sigma_{n,h}^2(B_\ell) \geq (2\beta^2 - \frac{1}{2}\eta) n^3, \quad (7.15)$$

Note that by (7.1) and (7.2) we can write for $B \subseteq V_n$:

$$\sigma_{n,h}^2(B) = 2 \sum_{b=1}^n |B(\cdot, b) \cap \mathbb{S}_h(b)| (|B(\cdot, b) \cap \mathbb{S}_h(b)| - 1) \quad (7.16)$$

and

$$\sigma_{n,v}^2(B) = 2 \sum_{a=1}^n |B(a, \cdot) \cap \mathbb{S}_v(a)| (|B(a, \cdot) \cap \mathbb{S}_v(a)| - 1),$$

in particular

$$\sum_{\ell=1}^{m_n} \sigma_{n,h}^2(B_\ell) = \sum_{b=1}^n \Theta_{n,h}(b) \quad \text{where} \quad \Theta_{n,h}(b) := 2 \sum_{\ell=1}^{m_n} |B_\ell(\cdot, b) \cap \mathbb{S}_h(b)| (|B_\ell(\cdot, b) \cap \mathbb{S}_h(b)| - 1). \quad (7.17)$$

Since $(B_\ell(\cdot, b))_{\ell=1, \dots, m_n}$ is a partition of $\{1, \dots, n\} \times \{b\}$, we can bound

$$\Theta_{n,h}(b) = 2 \sum_{\ell=1}^{m_n} \sum_{v_1 \neq v_2 \in B_\ell(\cdot, b) \cap \mathbb{S}_h(b)} 1 \leq 2 \sum_{v_1 \neq v_2 \in \mathbb{S}_h(b)} 1 = 2[\beta n]([\beta n] - 1) \leq 2\beta^2 n^2, \quad (7.18)$$

which yields $\sum_{\ell=1}^{m_n} \sigma_{n,h}^2(B_\ell) \leq 2 \sum_{b=1}^n \beta^2 n^2 \leq 2\beta^2 n^3$. The same arguments apply to $\sigma_{n,v}^2$, hence

$$\sum_{\ell=1}^{m_n} \sigma_{n,v}^2(B_\ell) \leq 2\beta^2 n^3. \quad (7.19)$$

Finally, relation (7.15) follows by $\sigma_{n,h}^2(B_\ell) = \sigma_n^2(B_\ell) - \sigma_{n,v}^2(B_\ell)$ applying (7.7) and (7.19).

Part B. We now show that $\Theta_{n,h}(b)$ from (7.17) is close to its maximum $2\beta^2 n^2$ for most values of b : more precisely, setting

$$I := \left\{ b \in \{1, \dots, n\} : \Theta_{n,h}(b) > 2(\beta^2 - \sqrt{\eta}) n^2 \right\}, \quad (7.20)$$

we show that

$$|I| \geq (1 - \frac{1}{4}\sqrt{\eta}) n. \quad (7.21)$$

To this purpose, we can write

$$\sum_{\ell=1}^{m_n} \sigma_{n,h}^2(B_\ell) = \sum_{b \in I} \Theta_{n,h}(b) + \sum_{b \in I^c} \Theta_{n,h}(b) \leq |I| 2\beta^2 n^2 + (n - |I|) 2(\beta^2 - \sqrt{\eta}) n^2,$$

which can be rewritten as

$$\sqrt{\eta} 2n^2 |I| \geq \sum_{\ell=1}^{m_n} \sigma_{n,h}^2(B_\ell) - 2(\beta^2 - \sqrt{\eta}) n^3.$$

Plugging in the bound (7.15), we obtain precisely (7.21).

Part C. Given $b \in I$, see (7.20), let ℓ_b be a color for which $|B_\ell(\cdot, b) \cap \mathbb{S}_h(b)|$ is maximal:

$$\ell_b := \arg \max \{ |B_\ell(\cdot, b) \cap \mathbb{S}_h(b)| : 1 \leq \ell \leq m_n \}.$$

We show that

$$\forall b \in I: \quad |B_{\ell_b}(\cdot, b) \cap \mathbb{S}_h(b)| \geq (\beta - 2\sqrt{\eta}) n, \quad (7.22)$$

which completes the proof of (7.11).

To prove (7.22), we recall from (7.17) that, setting $k_\ell := |B_\ell(\cdot, b) \cap \mathbb{S}_h(b)|$, we have

$$\Theta_{n,h}(b) = 2 \sum_{\ell=1}^{m_n} k_\ell(k_\ell - 1) \quad \text{with} \quad k_\ell \geq 0 \quad \text{such that} \quad \sum_{\ell=1}^{m_n} k_\ell = |\mathbb{S}_h(b)| = \lfloor \beta n \rfloor. \quad (7.23)$$

The function $(k_1, \dots, k_m) \mapsto 2 \sum_{\ell=1}^m k_\ell(k_\ell - 1)$ with the above constraints is maximised for $m = 1$ and $k_1 = \lfloor \beta n \rfloor$, where it takes the value $2\lfloor \beta n \rfloor(\lfloor \beta n \rfloor - 1) \sim 2\beta^2 n^2$. Since $\Theta_{n,h}(b)$ is close to $2\beta^2 n^2$ for $b \in I$, see (7.20), this explains intuitively why $k_{\ell_b} = |B_{\ell_b}(\cdot, b) \cap \mathbb{S}_h(b)| = \max_{\ell=1, \dots, m_n} k_\ell$ should be close to βn , so that (7.22) holds. To make this precise, we simply bound

$$\begin{aligned} \Theta_{n,h}(b) &= 2k_{\ell_b}(k_{\ell_b} - 1) + 2 \sum_{\ell \neq \ell_b} k_\ell(k_\ell - 1) \\ &\leq 2k_{\ell_b}^2 + 2k_{\ell_b} \sum_{\ell \neq \ell_b} k_\ell = 2k_{\ell_b}^2 + 2k_{\ell_b} (\lfloor \beta n \rfloor - k_{\ell_b}) \leq 2k_{\ell_b} \beta n. \end{aligned} \quad (7.24)$$

Since $\Theta_{n,h}(b) \geq 2(\beta - \sqrt{\eta}) n^2$ for $b \in I$ and $\beta > \frac{1}{2}$, we see that k_{ℓ_b} fulfils (7.22). \square

7.1.3. STEP 2. We prove (7.13) by contradiction: we assume that

$$\forall \ell: \quad |\{b \in I: \ell_b = \ell\}| < (\beta' - 7\sqrt{\eta}) n \quad \text{with} \quad \beta' := 2\beta - 1, \quad (7.25)$$

and we deduce a contradiction with (7.12), namely

$$\exists a \in I' \quad \forall \ell: \quad |B_\ell(a, \cdot)| < (\beta - 2\sqrt{\eta}) n. \quad (7.26)$$

Recalling (7.11), let D be the union of all rows of boxes with dominant colors:

$$D := \bigcup_{b \in I} \{B_{\ell_b}(\cdot, b) \cap \mathbb{S}_h(b)\}.$$

By assumption (7.25), each color ℓ appears in less than $(\beta' - 7\sqrt{\eta}) n$ rows of D , therefore

$$\forall a \quad \forall \ell: \quad |B_\ell(a, \cdot) \cap D| < (\beta' - 7\sqrt{\eta}) n. \quad (7.27)$$

We claim that we can obtain the following bound for a suitable $a \in I'$:

$$\exists a \in I': \quad \forall \ell: \quad |B_\ell(a, \cdot) \cap D^c| \leq (1 - \beta + 5\sqrt{\eta}) n. \quad (7.28)$$

Summing (7.27) and (7.28) we obtain (7.26) (note that $\beta' + (1 - \beta) = \beta$).

It only remains to prove (7.28). We observe that by (7.11)

$$|D| \geq |I| (\beta - 2\sqrt{\eta}) n \geq (1 - \frac{1}{4}\sqrt{\eta}) (\beta - 2\sqrt{\eta}) n^2 \geq (\beta - 3\sqrt{\eta}) n^2. \quad (7.29)$$

We then define

$$J := \{a \in \{1, \dots, n\}: |D(a, \cdot)| > (\beta - 5\sqrt{\eta}) n\} \quad (7.30)$$

so that, by Lemma 7.2 below, we have

$$|J| \geq 2\sqrt{\eta}n. \quad (7.31)$$

Since $|I'| \geq (1 - \frac{1}{4}\sqrt{\eta})n > (1 - \sqrt{\eta})n$, see (7.12), we have $|I'| + |J| > n$ and therefore $I' \cap J \neq \emptyset$. We then select any $a \in I' \cap J$ and note that $|B_{\ell}(a, \cdot) \cap D^c| \leq |D(a, \cdot)^c| = n - |D(a, \cdot)|$, hence the bound (7.28) follows by the definition (7.30) of J . \square

It remains to prove the following elementary lemma (recall that $V_n = \{1, \dots, n\}^2$).

Lemma 7.2. *If $C \subseteq V_n$ satisfies $|C| \geq un^2$ for some $u \in (0, 1)$, then for any $u' < u$*

$$|\{a \in \{1, \dots, n\} : |C(a, \cdot)| > u'n\}| \geq \frac{u - u'}{1 - u'}n \geq (u - u')n.$$

Proof. Setting $J := \{a \in \{1, \dots, n\} : |C(a, \cdot)| > u'n\}$ we can bound

$$\begin{aligned} |C| &= \sum_{a \in J} |C(a, \cdot)| + \sum_{a \in J^c} |C(a, \cdot)| \leq |J|n + (n - |J|)u'n \\ &= (1 - u')n|J| + u'n^2, \end{aligned}$$

that is $|J| \geq \frac{|C| - u'n^2}{(1 - u')n}$. Plugging in $|C| \geq un^2$ completes the proof. \square

7.1.4. STEP 3. We finally prove (7.8). From (7.14) and (7.16) we can write

$$\sigma_n^2(B_{\bar{\ell}}) \geq \sigma_{n,h}^2(B_{\bar{\ell}}) = 2 \sum_{b=1}^n |B_{\bar{\ell}}(\cdot, b) \cap \mathbb{S}_h(b)| (|B_{\bar{\ell}}(\cdot, b) \cap \mathbb{S}_h(b)| - 1)$$

Restricting the sum to the set $B := \{b \in I : \ell_b = \bar{\ell}\}$, recalling (7.11) and (7.13), we obtain

$$\begin{aligned} \sigma_{n,h}^2(B_{\bar{\ell}}) &\geq 2(\beta' - 7\sqrt{\eta})n(\beta - 2\sqrt{\eta})n((\beta - 2\sqrt{\eta})n - 1) \\ &\geq 2(\beta' - 7\sqrt{\eta})(\beta - 2\sqrt{\eta} + O(\frac{1}{n}))^2 n^3 \geq 2(\beta'\beta^2 - 11\sqrt{\eta} + O(\frac{1}{n}))n^3, \end{aligned} \quad (7.32)$$

which completes the proof. \square

Remark 7.3. In (7.32) we only estimated the ‘‘horizontal’’ contribution to the variance $\sigma_{n,h}^2(B_{\bar{\ell}})$. Of course, by symmetry, a version of (7.13) holds for columns, for some color $\bar{\ell}'$, hence an estimate like (7.32) also holds for $\sigma_{n,v}^2(B_{\bar{\ell}'})$. Note that both $B_{\bar{\ell}}$ and $B_{\bar{\ell}'}$ have cardinality at least $(\beta' - 7\sqrt{\eta})(\beta - 2\sqrt{\eta})n^2$ by (7.11) and (7.13), hence they must overlap when $\beta\beta' = \beta(2\beta - 1) > \frac{1}{2}$, i.e. for $\beta > \frac{1+\sqrt{5}}{4} \simeq 0.81$: in this case $\bar{\ell} = \bar{\ell}'$, hence we can improve our final bound (7.8) by a factor 2:

$$\sigma_n^2(B_{\bar{\ell}}) = \sigma_{n,h}^2(B_{\bar{\ell}}) + \sigma_{n,v}^2(B_{\bar{\ell}}) \geq 4(\beta'\beta^2 - 11\sqrt{\eta} + O(\frac{1}{n}))n^3. \quad (7.33)$$

Appendix A. Proofs of some technical results

A.1. Proof of Theorem 2.15. Let $\Delta(\alpha, n)$ denote the maximal degree in the Erdős-Renyi random graph $G(n, p_n)$, where $p_n = n^{\alpha-2}$. By Theorem 2.14 (including the subsequent discussion) and Remark 3.2-(2), it is sufficient to show that, for α as in the statement, there exists $0 < \varepsilon < \alpha/2$ such that

$$\mathbb{P}[\Delta(\alpha, n) \geq n^{\alpha/2-\varepsilon}] \longrightarrow 0, \quad n \rightarrow \infty.$$

One has that

$$\mathbb{P}[\Delta(\alpha, n) \geq n^{\alpha/2-\varepsilon}] \leq n\mathbb{P}[B(n-1, p_n) \geq n^{\alpha/2-\varepsilon}] \leq n\mathbb{P}[B(n, p_n) \geq n^{\alpha/2-\varepsilon}],$$

where $B(k, p)$ denotes a binomial random variable with parameters (k, p) , and the conclusion follows from an application of the multiplicative Chernoff inequality: according e.g. to [41, Problem 1.6-(e)] we can bound $\mathbb{P}(B(k, p) \geq t kp) \leq 2^{-tkp}$ provided $t > 2e$, hence for $0 < \varepsilon < \min\{\alpha/2, 1 - \alpha/2\}$

$$\mathbb{P}[B(n, p_n) \geq n^{\alpha/2-\varepsilon}] \leq 2^{-(n^{1-\varepsilon-\alpha/2})n^{\alpha-1}} = 2^{-n^{\alpha/2-\varepsilon}} = o(n^{-1}), \quad n \rightarrow \infty,$$

thus yielding the desired conclusion.

A.2. Proof of Theorem 4.8. To prove Theorem 5.7 we follow very closely the proof of Theorem 2.6. By contradiction we assume that (2.12) holds for some $k \geq 2$ and that $\{G_n\}$ is partially reducible, i.e. there exists $\varrho \in (0, 1]$, $V'_n \subseteq V_n$ satisfying (4.15) and partitions B_1, \dots, B_{m_n} of V'_n such that the three properties (i''), (ii'') and (iii'') of partial reducibility hold. Ordering the sets as in (4.9), we can apply (4.10) which yields (4.11), that we copy for convenience:

$$\frac{\mu_k^{(n)}}{2} \leq \frac{\sum_{i=k}^{m_n} q_i^{(n)} \varphi(B_i)}{\sum_{i=k}^{m_n} q_i^{(n)}} \quad \text{where again} \quad q_i^{(n)} := \frac{\text{vol}(B_i)}{\text{vol}(V_n)} = \frac{\text{vol}(B_i)}{2|\mathcal{E}_n|}. \quad (\text{A.1})$$

This time the vector $(q_i^{(n)})_{1 \leq i \leq m_n}$ needs not be a probability, however we still have

$$\sum_{i=1}^{m_n} q_i^{(n)} = \frac{\text{vol}(V'_n)}{\text{vol}(V_n)} \rightarrow \varrho > 0. \quad (\text{A.2})$$

To complete the proof, it is then sufficient to show that

$$\sum_{i=1}^{m_n} q_i^{(n)} \varphi(B_i) = o(1) \quad \text{and} \quad \max_{1 \leq i \leq m_n} q_i^{(n)} = o(1), \quad (\text{A.3})$$

(so that $\sum_{i=k}^{m_n} q_i^{(n)} \rightarrow \varrho > 0$). Applying (4.13) and (4.14), it remains to show that

$$\sum_{i=1}^{m_n} E(B_i, \bar{B}_i) = o(|\mathcal{E}_n|). \quad (\text{A.4})$$

To this purpose, we need to modify (4.12) because B_1, \dots, B_{m_n} is only a partition of V'_n . To obtain a partition of the full set of vertices V_n , we define $B_{m_n+1} := \bar{V}'_n$. Arguing as in (4.12), we can now write

$$\frac{1}{2} \sum_{i=1}^{m_n+1} E(B_i, \bar{B}_i) = \frac{1}{2} \sum_{\substack{i,j=1 \\ i \neq j}}^{m_n+1} E(B_i, B_j) = |\mathcal{E}_n| - \sum_{i=1}^{m_n+1} E(B_i, B_i) \quad (\text{A.5})$$

and isolating the terms $i = m_n + 1$ in the first and last sums we obtain

$$\begin{aligned} \frac{1}{2} \sum_{i=1}^{m_n} E(B_i, \bar{B}_i) &= \left\{ |\mathcal{E}_n| - \frac{1}{2} E(V'_n, \bar{V}'_n) - E(\bar{V}'_n, \bar{V}'_n) \right\} - \sum_{i=1}^{m_n} E(B_i, B_i) \\ &= \left\{ |\mathcal{E}_n| - E(\bar{V}'_n, \bar{V}'_n) - o(|\mathcal{E}_n|) \right\} - E(V'_n, V'_n) - o(|\mathcal{E}_n|), \end{aligned}$$

where we applied (4.15) and the assumption of partial reducibility. We finally observe that, plainly, $|\mathcal{E}_n| = E(V'_n, V'_n) + E(\overline{V'_n}, \overline{V'_n}) + E(V_n, \overline{V'_n}) = E(V'_n, V'_n) + E(\overline{V'_n}, \overline{V'_n}) + o(|\mathcal{E}_n|)$ which completes the proof.

A.3. Proof of Theorem 5.7. We first introduce and recall the following notation. For $\ell \leq k$ and $S_1, \dots, S_\ell \subseteq V_n$ disjoint and nonempty subsets, we denote

$$\mathcal{E}(S_1, \dots, S_\ell) := \{ e \in \mathcal{E}_n : e \subset S_1 \cup \dots \cup S_\ell : \\ \exists v_1, \dots, v_\ell \in e \text{ with } v_1 \in S_1, \dots, v_\ell \in S_\ell \},$$

as the set of edges with elements in $S_1 \cup \dots \cup S_\ell$ and at least one of them in each S_i , for $i = 1, \dots, \ell$. In particular, we recall

$$\mathcal{E}_n(S, S) := \{ e \in \mathcal{E}_n : e \subset S \},$$

while

$$\mathcal{E}(S, \overline{S}) = \{ e \in \mathcal{E}_n : \exists v_1, v_2 \in e \text{ with } v_1 \in S, v_2 \in \overline{S} \} = \partial S. \quad (\text{A.6})$$

We now prove Theorem 5.7. The arguments follow the same guidelines as the proof of Theorem 2.6. However, the more complex structure of hypergraphs requires additional combinatorial details, which we briefly outline below.

We assume that (5.15) holds for some $k \geq 2$ and that there exists a sequence of partitions B_1, \dots, B_{m_n} , $n \geq 1$, such that the three properties (i), (ii), (iii) are verified. Moreover, without loss of generality, the relation (5.9) holds for $n \geq 1$. Recall that in this setting $r(\mathcal{G}) = cr(\mathcal{G}) = d$, hence (5.10) yields

$$\frac{\mu_k^{(n)}}{2(d-1)} \leq \frac{\sum_{i=k}^{m_n} q_i^{(n)} \varphi(B_i)}{\sum_{i=k}^{m_n} q_i^{(n)}}, \quad \text{where } q_i^{(n)} := \frac{\text{vol}(B_i)}{d w_n(\mathcal{E}_n)}. \quad (\text{A.7})$$

To conclude the proof, it suffices to show that the right-hand side of the previous inequality converges necessarily to zero.

Since the subsets B_i 's form a partition of V_n , it is possible to express the set ∂B_i , $i \in \{1, \dots, m_n\}$, in terms of a disjoint union:

$$\partial B_i = \mathcal{E}(B_i, \overline{B_i}) = \bigcup_{\alpha=1}^{d-1} \bigcup_{\substack{1 \leq i_1 < \dots < i_\alpha \leq m_n \\ i_1, \dots, i_\alpha \neq i}} \mathcal{E}(B_i, B_{i_1}, \dots, B_{i_\alpha}),$$

(recall (5.14) and (A.6)). Moreover, note that $\mathcal{E}_n \setminus \bigcup_{i=1}^{m_n} \mathcal{E}(B_i, B_i) = \bigcup_{i=1}^{m_n} \mathcal{E}(B_i, \overline{B_i})$, where the union in the right-hand side is not disjoint, yet with explicit cardinality

$$\left| \bigcup_{i=1}^{m_n} \mathcal{E}(B_i, \overline{B_i}) \right| = \sum_{\alpha=1}^{d-1} \frac{1}{\alpha+1} \sum_{\substack{1 \leq i_1 < \dots < i_\alpha \leq m_n \\ i_1, \dots, i_\alpha \neq i}} |\mathcal{E}(B_i, B_{i_1}, \dots, B_{i_\alpha})|,$$

where the factor $\frac{1}{\alpha+1}$ compensates for overcounting: indeed, each edge that intersects exactly $\alpha+1$ blocks is counted once for every choice of the distinguished B_i among them, and

hence appears $\alpha + 1$ times in the sum. As a consequence, by (ii) we have

$$\begin{aligned} \frac{1}{d} \sum_{i=1}^{m_n} w_n(\mathcal{E}(B_i, \overline{B}_i)) &= \frac{1}{d} \sum_{i=1}^{m_n} \sum_{\alpha=1}^{d-1} \sum_{\substack{1 \leq i_1 < \dots < i_\alpha \leq m_n \\ i_1, \dots, i_\alpha \neq i}} w_n(\mathcal{E}(B_i, B_{i_1}, \dots, B_{i_\alpha})) \\ &\leq \sum_{i=1}^{m_n} \sum_{\alpha=1}^{d-1} \frac{1}{\alpha+1} \sum_{\substack{1 \leq i_1 < \dots < i_\alpha \leq m_n \\ i_1, \dots, i_\alpha \neq i}} w_n(\mathcal{E}(B_i, B_{i_1}, \dots, B_{i_\alpha})) \quad (\text{A.8}) \\ &= w_n(\mathcal{E}_n) - \sum_{i=1}^{m_n} w_n(\mathcal{E}(B_i, B_i)) = o(w_n(\mathcal{E}_n)) \end{aligned}$$

and, then

$$\sum_{i=1}^{m_n} \frac{w_n(\mathcal{E}(B_i, \overline{B}_i))}{d w_n(\mathcal{E}_n)} = \sum_{i=1}^{m_n} \frac{w_n(\partial B_i)}{d w_n(\mathcal{E}_n)} = \sum_{i=1}^{m_n} q_i^{(n)} \varphi(B_i) = o(1), \quad (\text{A.9})$$

which already shows that the numerator in the bound (A.7) for $\mu_k^{(n)}$ vanishes. It remains to prove that the denominator is bounded away from zero. We still have

$$\sum_{i=1}^{m_n} q_i^{(n)} = \frac{\sum_{i=1}^{m_n} \text{vol}(B_i)}{d w_n(\mathcal{E}_n)} = \frac{\sum_{i=1}^{m_n} \sum_{v \in B_i} \sum_{e \in \mathcal{E}_v} w_n(e)}{d w_n(\mathcal{E}_n)} = \frac{\sum_{v \in V_n} \sum_{e \in \mathcal{E}_v} w_n(e)}{d w_n(\mathcal{E}_n)} = 1,$$

(see (5.5)). Therefore, we just need to prove that $\sum_{i=1}^{k-1} q_i^{(n)} \rightarrow 0$, which is implied by

$$q_i^{(n)} = \frac{\text{vol}(B_i)}{d w_n(\mathcal{E}_n)} = \frac{\sum_{v \in B_i} \sum_{e \in \mathcal{E}_v} w_n(e)}{d w_n(\mathcal{E}_n)} \leq \frac{w_n(\mathcal{E}(B_i, B_i))}{w_n(\mathcal{E}_n)} + \frac{(d-1) w_n(\mathcal{E}(B_i, \overline{B}_i))}{d w_n(\mathcal{E}_n)} = o(1), \quad (\text{A.10})$$

uniformly for $i \in \{1, \dots, m_n\}$, where we applied (iii) and (A.9).

A.4. Proof of (6.7). We adopt the notation and assumptions of Example 6.1; also, given a permutation ϱ of $[b]$ and $A \subseteq V_n$, we write A^ϱ to denote the class of all $(v_1, \dots, v_b) \in V_n$ such that $(v_{\varrho(1)}, \dots, v_{\varrho(b)}) \in A$. For every $s = 1, \dots, d$, write L_s to denote the set of those $\ell = 1, \dots, d$ such that $|S_\ell \cap \{(s, 1), \dots, (s, b)\}| = 1$ (note that the size of the previous intersection is either zero or one, by construction). We stress that $|L_s| = b$, for $s = 1, \dots, d$, and that each $\ell = 1, \dots, d$ is contained in exactly b distinct sets L_s . Without loss of generality, we always label the elements of L_s in such a way that, if $L_s = \{\ell_1, \dots, \ell_b\}$, then $\ell_1 < \ell_2 < \dots < \ell_b$. For $s = 1, \dots, d$, we denote by

$$\pi_s : [n]^d \rightarrow [n]^{L_s} : \mathbf{i} = (i_1, \dots, i_d) \mapsto \pi_s(\mathbf{i}) = (i_{\ell_1}, \dots, i_{\ell_b}),$$

where $\{\ell_1, \dots, \ell_b\} = L_s$. One can easily show that there exist permutations $\varrho_1, \dots, \varrho_d$ of $[b]$ such that

$$|E_n^0 \cap (A_1 \times \dots \times A_d)| = \int_{[n]^d} \prod_{s=1}^d \mathbf{1}_{A_s^{\varrho_s}}(\pi_s(\mathbf{i})) \nu_d(d\mathbf{i}) = \int_{[n]^d} \prod_{s=1}^d \mathbf{1}_{A_s^{\varrho_s}}(\pi_s(\mathbf{i}))^{1/b} \nu_d(d\mathbf{i}), \quad (\text{A.11})$$

where, for $t \geq 1$, ν_t stands for the counting measure on $[n]^t$. We can now directly apply [43, Theorem 2.1] and deduce that the right-hand side of (A.11) is bounded by

$$\prod_{s=1}^d \left(\int_{[n]^b} \mathbf{1}_{A_s^{\varrho_s}}(\mathbf{i}) \nu_b(d\mathbf{i}) \right)^{1/b}.$$

Since $[n]^b$ is a symmetric set, one has that

$$\int_{[n]^b} \mathbf{1}_{A_s^{e_s}}(\mathbf{i}) \nu_b(d\mathbf{i}) = |A_s|,$$

and the conclusion follows immediately. \square

Appendix B. Cartesian products

In this appendix, we recall some basic definitions and properties of Cartesian products of regular graphs. See e.g. [48, Chapters 4, 5 and 33] and [19, Section 1.4] for a full picture. Fix $d \geq 2$, and let $G = (V, \mathcal{E})$ be an undirected (loop-free) d -regular graph such that $|V| = N$. We denote by $\lambda_1 \geq \dots \geq \lambda_N$ the eigenvalues of the adjacency matrix of G , and by $0 = \mu_1 \leq \dots \leq \mu_N \leq 2$ the eigenvalues of the corresponding normalized Laplacian; see Section 4.1.1 for details. Fix $m \geq 2$: throughout the paper, we use the symbols

$$G^{\square m} = \underbrace{G \square \dots \square G}_{m \text{ times}}$$

to denote the m th Cartesian product of G . We recall that $G^{\square m}$ is the graph whose vertices are given by the set

$$V^m = \{v = (v_1, \dots, v_m) : v_i \in V\}$$

and such that $v = (v_1, \dots, v_m) \sim w = (w_1, \dots, w_m)$ if and only if there exists $j \in [m]$ such that $v_i = w_i$ for all $i \neq j$ and $\{v_j, w_j\} \in \mathcal{E}$ (that is, if and only if v_j, w_j are adjacent in G). For an arbitrary $(i_1, \dots, i_m) \in [N]^m$, we introduce the notation

$$\Lambda(i_1, \dots, i_m) := \lambda_{i_1} + \dots + \lambda_{i_m}, \quad \text{and} \quad M(i_1, \dots, i_m) := \mu_{i_1} + \dots + \mu_{i_m}. \quad (\text{B.1})$$

The following facts are used in several parts of the paper, and can be routinely checked.

- (1) $G^{\square m}$ is (dm) -regular and, consequently, the number of edges in $G^{\square m}$ is $2^{-1}dmN^m$.
- (2) The spectrum of the adjacency matrix of $G^{\square m}$ is given by the set

$$\{\Lambda(i_1, \dots, i_m) : (i_1, \dots, i_m) \in [N]^m\},$$

where we have used the notation introduced in the first part of (B.1).

- (3) The spectrum of the normalized Laplacian associated with $G^{\square m}$ is

$$\left\{ \frac{1}{m} M(i_1, \dots, i_m) : (i_1, \dots, i_m) \in [N]^m \right\},$$

where we have used the notation appearing in the second part of (B.1). To see this, recall that, since $G^{\square m}$ is (dm) -regular and its adjacency eigenvalues are the sums $\Lambda(i_1, \dots, i_m)$ defined above, the associated normalized Laplacian eigenvalues are

$$1 - \frac{\Lambda(i_1, \dots, i_m)}{dm} = \frac{1}{m} (\mu_{i_1} + \dots + \mu_{i_m}),$$

which yields the above formula. Note that the factor $1/m$ ensures that all eigenvalues of the normalized Laplacian of $G^{\square m}$ lie in $[0, 2]$, as they should.

References

- [1] N. Alon. Eigenvalues and expanders. *Combinatorica*, 6:83–96, 1986.
- [2] N. Alon and V. D. Milman. λ_1 , isoperimetric inequalities for graphs, and superconcentrators. *J. Comb. Theory, Ser. B*, 38(1):73–88, 1985.
- [3] J. Angst and G. Poly. Fluctuations in Salem-Zygmund almost sure central limit theorem. *Electron. J. Probab.*, 28:40, 2023. Id/No 44.
- [4] M. F. Anjos and J. Neto. Spectral bounds for graph partitioning with prescribed partition sizes. *Discrete Appl. Math.*, 269:200–210, 2019.
- [5] S. Arora, S. Rao, and U. Vazirani. Expander flows, geometric embeddings and graph partitioning. *J. ACM*, 56(2), Apr. 2009.
- [6] P. Baldi and Y. Rinott. On normal approximations of distributions in terms of dependency graphs. *Ann. Probab.*, 17(4):1646–1650, 1989.
- [7] A. Banerjee. On the spectrum of hypergraphs. *Linear Algebra and its Applications*, 614:82–110, 2021.
- [8] J. Bennett, A. Carbery, M. Christ, and T. Tao. The brascamp–lieb inequalities: Finiteness, structure and extremals. *Geom. Funct. Anal.*, 17:1343–1415, 2008.
- [9] B. Bhattacharya, S. Das, S. Mukherjee, and S. Mukherjee. Fluctuations of quadratic chaos. *Commun. Math. Phys.*, 405(10):51, 2024. Id/No 237.
- [10] R. Blei. Combinatorial dimension and certain norms in harmonic analysis. *Am. J. Math.*, 106:847–887, 1984.
- [11] R. Blei. *Analysis in integer and fractional dimensions*, volume 71 of *Camb. Stud. Adv. Math.* Cambridge: Cambridge University Press, 2001.
- [12] R. Blei. Measurements of interdependence. *Lith. Math. J.*, 51(2):141–154, 2011.
- [13] R. Blei and S. Janson. Rademacher chaos: tail estimates versus limit theorems. *Ark. Mat.*, 42(1):13–29, 2004.
- [14] R. Blei, Y. Peres, and J. Schmerl. Fractional products of sets. *Random Struct. Algorithms*, 6(1):113–119, 1995.
- [15] R. C. Blei. Fractional cartesian products of sets. *Ann. Inst. Fourier*, 29(2):79–105, 1979.
- [16] R. C. Blei. *Fractional dimensions and bounded fractional forms*, volume 331 of *Mem. Am. Math. Soc.* Providence, RI: American Mathematical Society (AMS), 1985.
- [17] R. C. Blei and T. W. Körner. Combinatorial dimension and random sets. *Israel J. Math.*, 47(1):65–74, 1984.
- [18] B. Bollobás. *Random graphs*. London-Orlando etc.: Academic Press (Harcourt Brace Jovanovich, Publishers). XVI, 447 p. hbk: £ 52.00; \$ 58.50; pbk: £ 26.00; \$ 29.95 (1985)., 1985.
- [19] A. E. Brouwer and W. H. Haemers. *Spectra of graphs*. Universitext. Springer, New York, 2012.
- [20] F. Caravenna and F. Cottini. Gaussian limits for subcritical chaos. *Electron. J. Probab.*, 27:1–35, 2022.
- [21] F. Caravenna, R. Sun, and N. Zygouras. Polynomial chaos and scaling limits of disordered systems. *J. Eur. Math. Soc. (JEMS)*, 19(1):1–65, 2017.
- [22] F. Caravenna, R. Sun, and N. Zygouras. Universality in marginally relevant disordered systems. *Ann. Appl. Probab.*, 27:3050–3112, 2017.
- [23] F. Caravenna, R. Sun, and N. Zygouras. The critical 2d stochastic heat flow. *Invent. Math.*, 233(1):325–460, 2023.
- [24] S. Chatterjee. A new method of normal approximation. *The Annals of Probability*, 36(4):1584 – 1610, 2008.
- [25] J. Cheeger. A lower bound for the smallest eigenvalue of the laplacian. In *Problems in Analysis*, pages 195–200. Princeton University Press, Princeton, 1971.
- [26] F. R. K. Chung. *Spectral graph theory*, volume 92 of *CBMS Regional Conference Series in Mathematics*. Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI, 1997.
- [27] N. Cook, L. Goldstein, and T. Johnson. Size biased couplings and the spectral gap for random regular graphs. *Ann. Probab.*, 46(1):72–125, 2018.
- [28] A. De, I. Diakonikolas, and R. A. Servedio. Deterministic approximate counting for juntas of degree-2 polynomial threshold functions. In *2014 IEEE 29th Conference on Computational Complexity (CCC)*, pages 229–240, 2014.
- [29] A. De and R. A. Servedio. Efficient deterministic approximate counting for low-degree polynomial threshold functions. In *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing, STOC '14*, page 832–841, New York, NY, USA, 2014. Association for Computing Machinery.

- [30] P. de Jong. A central limit theorem for generalized quadratic forms. *Probab. Theory Relat. Fields*, 75:261–277, 1987.
- [31] P. de Jong. A central limit theorem for generalized multilinear forms. *J. Multivariate Anal.*, 34:275–289, 1990.
- [32] P. de Jong. A central limit theorem with applications to random hypergraphs. *Random Struct. Algorithms*, 8(2):105–120, 1996.
- [33] A. Deya, S. Nouredine, and I. Nourdin. Fourth moment theorem and q -Brownian chaos. *Commun. Math. Phys.*, 321(1):113–134, 2013.
- [34] A. Deya and I. Nourdin. Invariance principles for homogeneous sums of free random variables. *Bernoulli*, 20(2):586–603, 2014.
- [35] G. Dierickx, I. Nourdin, G. Peccati, and M. Rossi. Small scale CLTs for the nodal length of monochromatic waves. *Commun. Math. Phys.*, 397(1):1–36, 2023.
- [36] Döbler, Kasprzak, and Peccati. Title missing. *Probab. Theory Relat. Fields*, 2022. Dettagli da completare.
- [37] C. Döbler. The Berry-Esseen bound in de Jong’s CLT. *Stat. Probab. Lett.*, 215:8, 2024. Id/No 110244.
- [38] C. Döbler, M. Kasprzak, and G. Peccati. The multivariate functional de Jong CLT. *Probab. Theory Relat. Fields*, 184(1-2):367–399, 2022.
- [39] C. Döbler and K. Krokowski. On the fourth moment condition for Rademacher chaos. *Ann. Inst. Henri Poincaré Probab. Stat.*, 55(1):61–97, 2019.
- [40] C. Döbler and G. Peccati. Quantitative de Jong theorems in any dimension. *Electron. J. Probab.*, 22:35, 2017. Id/No 2.
- [41] D. P. Dubhashi and A. Panconesi. *Concentration of measure for the analysis of randomized algorithms*. Cambridge: Cambridge University Press, 2009.
- [42] B. Efron and C. Stein. The jackknife estimate of variance. *Ann. Stat.*, 9:586–596, 1981.
- [43] H. Finner. A generalization of hölder’s inequality and some probability inequalities. *Ann. Probab.*, 20(4):1893–1901, 1992.
- [44] J. Friedman. On the second eigenvalue and random walks in random d -regular graphs. *Combinatorica*, 11(4):331–362, 1991.
- [45] C. Garban. Oded Schramm’s contributions to noise sensitivity. *Ann. Probab.*, 39(5):1702–1767, 2011.
- [46] C. Garban, G. Pete, and O. Schramm. The Fourier spectrum of critical percolation. *Acta Math.*, 205(1):19–104, 2010.
- [47] C. Godsil and G. Royle. *Algebraic graph theory*, volume 207 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2001.
- [48] R. Hammack, W. Imrich, and S. Klavžar. *Handbook of product graphs*. Discrete Math. Appl. (Boca Raton). Boca Raton, FL: CRC Press, 2nd ed. edition, 2011.
- [49] R. Herry, D. Malicet, and G. Poly. Limit distributions for polynomials with independent and identically distributed entries, 2024.
- [50] W. Hoeffding. A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.*, 19:293–325, 1948.
- [51] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc.*, 43(4):439–561, 2006.
- [52] S. Janson. *Gaussian Hilbert spaces*, volume 129 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1997.
- [53] O. Kallenberg. *Foundations of modern probability*. Probab. Appl. New York, NY: Springer, 1997.
- [54] T. Kemp, I. Nourdin, G. Peccati, and R. Speicher. Wigner chaos and the fourth moment. *Ann. Probab.*, 40(4):1577–1635, 2012.
- [55] Y. Koike. High-dimensional central limit theorems for homogeneous sums. *J. Theor. Probab.*, 36(1):1–45, 2023.
- [56] V. S. Korolyuk and Y. V. Borovskikh. *Theory of U -statistics. Updated and transl. from the Russian by P. V. Malyshev and D. V. Malyshev*, volume 273 of *Math. Appl., Dordr.* Dordrecht: Kluwer Academic Publishers, 1994.
- [57] M. Krivelevich and B. Sudakov. The largest eigenvalue of sparse random graphs. *Comb. Probab. Comput.*, 12(1):61–72, 2003.
- [58] R. Laskar and C. Wallis. Chessboard graphs, related designs, and domination parameters. *J. Stat. Plann. Inference*, 76(1-2):285–294, 1999.
- [59] G. Last, G. Peccati, and D. Yogeshwaran. Phase transitions and noise sensitivity on the Poisson space via stopping sets and decision trees. *Random Struct. Algorithms*, 63(2):457–511, 2023.

- [60] J. Lee, S. O. Gharan, and L. Trevisan. Multiway spectral partitioning and higher-order cheeger inequalities. *J. ACM*, 61(6), 2014.
- [61] A. Louis, P. Raghavendra, P. Tetali, and S. Vempala. Many sparse cuts via higher eigenvalues. In *Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing*, STOC '12, page 1131–1140, New York, NY, USA, 2012. Association for Computing Machinery.
- [62] A. Lubotzky. Expander graphs in pure and applied mathematics. *Bull. Amer. Math. Soc. (N.S.)*, 49(1):113–162, 2012.
- [63] A. Mandelbaum and M. S. Taqqu. Invariance principle for symmetric statistics. *Ann. Stat.*, 12:483–496, 1984.
- [64] D. Marinucci, M. Rossi, and A. Vidotto. Fluctuations of level curves for time-dependent spherical random fields. *Ann. Henri Lebesgue*, 7:583–620, 2024.
- [65] E. Mossel, R. O’Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. *Ann. Math. (2)*, 171(1):295–341, 2010.
- [66] M. R. Murty. Ramanujan graphs. *J. Ramanujan Math. Soc.*, 18(1):33–52, 2003.
- [67] M. Notarnicola. Matrix Hermite polynomials, random determinants and the geometry of Gaussian fields. *Ann. Henri Lebesgue*, 6:975–1030, 2023.
- [68] I. Nourdin and G. Peccati. Universal Gaussian fluctuations of non-Hermitian matrix ensembles: from weak convergence to almost sure CLTs. *ALEA, Lat. Am. J. Probab. Math. Stat.*, 7:341–375, 2010.
- [69] I. Nourdin and G. Peccati. *Normal Approximations with Malliavin Calculus: From Stein’s Method to Universality*. Cambridge University Press, 2012.
- [70] I. Nourdin, G. Peccati, G. Poly, and R. Simone. Classical and free fourth moment theorems: universality and thresholds. *J. Theor. Probab.*, 29(2):653–680, 2016.
- [71] I. Nourdin, G. Peccati, and G. Reinert. Invariance principles for homogeneous sums: universality of gaussian wiener chaos. *Ann. Probab.*, 38:1947–1985, 2010.
- [72] I. Nourdin, G. Peccati, and G. Reinert. Stein’s method and stochastic analysis of rademacher functionals. *Electron. J. Probab.*, 15:1703–1742, 2010.
- [73] I. Nourdin, G. Peccati, and M. Rossi. Nodal statistics of planar random waves. *Commun. Math. Phys.*, 369(1):99–151, 2019.
- [74] I. Nourdin, G. Peccati, and Y. Swan. Entropy and the fourth moment phenomenon. *J. Funct. Anal.*, 266(5):3170–3207, 2014.
- [75] D. Nualart. *The Malliavin calculus and related topics*. Probability and its Applications (New York). Springer-Verlag, Berlin, second edition, 2006.
- [76] D. Nualart and G. Peccati. Central limit theorems for sequences of multiple stochastic integrals. *Ann. Probab.*, 33(1):177–193, 2005.
- [77] R. O’Donnell. *Analysis of Boolean functions*. Cambridge: Cambridge University Press, 2014.
- [78] G. Peccati and M. Reitzner, editors. *Stochastic analysis for Poisson point processes. Malliavin calculus, Wiener-Itô chaos expansions and stochastic geometry*, volume 7 of *Bocconi Springer Ser.* Milano: Bocconi University Press; Cham: Springer, 2016.
- [79] G. Peccati and M. Taqqu. *Wiener chaos: Moments, cumulants and diagrams. A survey with computer implementation*, volume 1 of *Bocconi Springer Ser.* Milano: Bocconi University Press; Milano: Springer, 2011.
- [80] G. Peccati and A. Vidotto. Gaussian random measures generated by Berry’s nodal sets. *J. Stat. Phys.*, 178(4):996–1027, 2020.
- [81] D. Puder. Expansion of random graphs: new proofs, new results. *Invent. Math.*, 201(3):845–908, 2015.
- [82] V. I. Rotar’. Limit theorems for polylinear forms. *J. Multivariate Anal.*, 9:511–530, 1979.
- [83] S. Saha, K. Sharma, and S. Panda. On the laplacian spectrum of k-uniform hypergraphs. *Linear Algebra and its Applications*, 655:1–27, 2022.
- [84] R. J. Serfling. *Approximation theorems of mathematical statistics*. Wiley Ser. Probab. Math. Stat. John Wiley & Sons, Hoboken, NJ, 1980.
- [85] R. Simone. Universality of free homogeneous sums in every dimension. *ALEA, Lat. Am. J. Probab. Math. Stat.*, 12(1):213–244, 2015.
- [86] A. Sinclair and M. Jerrum. Approximate counting, uniform generation and rapidly mixing markov chains. *Inf. Comput.*, 82(1):93–133, 1989.
- [87] L. Trevisan. Lecture notes on graph partitioning, expanders and spectral methods. Lecture notes, 2016. Link: <https://lucatrevisan.github.io/books/expanders-2016.pdf>.

DIPARTIMENTO DI MATEMATICA E APPLICAZIONI, UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA, VIA
COZZI 55, 20125 MILANO, ITALY

Email address: `francesco.caravenna@unimib.it`

LPSM, SORBONNE UNIVERSITÉ, 4 PLACE JUSSIEU, 75005 PARIS, FRANCE

Email address: `francesca.cottini@sorbonne-universite.fr`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF LUXEMBOURG, MAISON DU NOMBRE 6, AVENUE DE LA
FONTE L-4364 ESCH-SUR-ALZETTE, LUXEMBOURG

Email address: `giovanni.peccati@uni.lu`