Exchangeability and Realizability: De Finetti Theorems on Graphs

T. S. Jayram and Jan Vondrák

IBM Almaden Research Center, San Jose, CA, USA
{jayram,jvondrak}@us.ibm.com

Abstract

A classic result in probability theory known as de Finetti's theorem states that exchangeable random variables are equivalent to a mixture of distributions where each distribution is determined by an i.i.d. sequence of random variables (an "i.i.d. mix"). Motivated by a recent application in [18] and more generally by the relationship of local vs. global correlation in randomized rounding, we study weaker notions of exchangeability that still imply the conclusion of de Finetti's theorem. We say that a bivariate distribution ρ is G-realizable for a graph G if there exists a joint distribution of random variables on the vertices such that the marginal distribution on each edge equals ρ.

We first characterize completely the G-realizable distributions for all symmetric/arc-transitive graphs G. Our main results are forms of de Finetti's theorem for general graphs, based on spectral properties. Let $\lambda_1(G) \geq \ldots \geq \lambda_n(G)$ denote the eigenvalues of the adjacency matrix of G.

1. We prove that if ρ is G_n-realizable for a sequence of graphs such that $\lim_{n\to\infty} \frac{\lambda_n(G_n)}{\lambda_1(G_n)} = 0$, then ρ is described by a probability matrix that is positive-semidefinite. For random variables on domains of size $|\mathcal{D}| \leq 4$, this implies that ρ must be an i.i.d. mix.
2. If ρ is G_n-realizable for a sequence of (n, d, λ)-graphs G_n (d-regular with all eigenvalues except for one bounded by λ in absolute value) such that $\lim_{n\to\infty} \frac{\lambda(G_n)}{d(G_n)} = 0$, then ρ is an i.i.d. mix.
3. If ρ is $\vec{G}_n$-realizable for a sequence of directed graphs such that each of them is an arbitrary orientation of an (n, d, λ)-graph G_n, and $\lim_{n\to\infty} \frac{\lambda(G_n)}{d(G_n)} = 0$, then ρ is an i.i.d. mix.

1998 ACM Subject Classification G.3 Probability and Statistics

Keywords and phrases exchangeability, de Finetti's Theorem, spectral graph theory, regularity lemma

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2014.762

1 Introduction

De Finetti's theorem [3] is a classic result in probability theory and statistics which states that exchangeable observations are equivalent to independent observations conditioned on a latent variable. Formally, a finite sequence of random variables $X_1, X_2, \ldots, X_n$ (or their joint distribution) is said to be *exchangeable* if their joint distribution is the same as that of the sequence $X_{\pi(1)}, X_{\pi(2)}, \ldots, X_{\pi(n)}$, for all permutations π on $[n]$. Note that every i.i.d. sequence satisfies this property and so does every convex combination of distributions determined by i.i.d. sequences (over the same domain). Following [21], we call such a distribution an *i.i.d. mix*. An infinite sequence of random variables is said to be exchangeable if every finite prefix is itself exchangeable. De Finetti's theorem states that every infinite exchangeable sequence is equivalent to an i.i.d. mix (of infinite sequences). Originally shown for Bernoulli random variables [4], it is now part of a broad theory of exchangeability; Kallenberg's book [8] is an excellent reference.

17th Int'l Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX'14) /
18th Int'l Workshop on Randomization and Computation (RANDOM'14).
Editors: Klaus Jansen, José Rolim, Nikhil Devanur, and Cristopher Moore; pp. 762–778
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Diaconis [5] was the first one to study exchangeability for finite sequences, and observe that de Finetti's theorem fails, e.g. for length two.[1] Diaconis and Freedman [6] considered a stronger form of exchangeability. Say that an exchangeable sequence $X_1, X_2, \ldots, X_k$ is n-*extendable*, for $n \geq k$, if it can be extended to an exchangeable sequence $X_1, X_2, \ldots, X_n$. Diaconis and Freedman [6] showed that every k-variate n-extendable distribution ρ is at variation distance at most $\frac{k(k-1)}{2n}$ from an i.i.d. mix, independent of the size of the domain of each X_i. For finite domains of size d, they showed a bound of $2dk/n$. This can be improved to $O(k\sqrt{d}/n)$ [19]. Here, we mostly deal with finite domains unless stated otherwise.

In recent work [18], a related question arose in the design of randomized rounding schemes for the Multiway Cut problem. A bivariate distribution ρ was defined in [18] to be *pairwise realizable*, if for each n there exist $X_1, X_2, \ldots, X_n$ such that the marginal distribution of (X_i, X_j) for every distinct $i, j \in [n]$ equals ρ. The authors showed that a distribution ρ is pairwise realizable if and only if ρ is an i.i.d. mix. Although this statement does not appear to have been formulated explicitly before, it is implicit in several works in the area of exchangeability (e.g. [21]). We observe here that it can be actually derived directly from the Diaconis-Freedman theorem (more generally for k-variate distributions and with the same quantitative bounds; see the last part of this section for details).

Since pairwise realizability is sufficient to derive the conclusion of the Diaconis-Freedman theorem (that such a bivariate distribution must be close to an i.i.d. mix), it is natural to ask whether the assumption could be weakened even further. In particular, is it necessary to assume that all pairs have the same distribution to conclude that this distribution is close to an i.i.d. mix? More precisely, we investigate the following concept.

G-realizable Distributions. We say that a bivariate distribution ρ is G-realizable for an undirected graph G if there exists a joint distribution of variables $\{X_v : v \in V(G)\}$ such that the distribution of (X_u, X_v) is ρ for each edge $\{u, v\} \in E(G)$. Similarly, ρ is $\vec{G}$-realizable for a directed graph $\vec{G}$ if the marginal distribution of (X_u, X_v) for each directed edge (u, v) is ρ. Note that the condition of being pairwise realizable is equivalent to being K_n-realizable for every n. The question we ask is, what distributions are G-realizable for a given graph G, and how does this depend on the structure of G? In particular, what kinds of graphs admit non-trivial G-realizable distributions and what kinds of graphs admit only i.i.d. mixes similar to de Finetti's theorem?

Motivation. Apart from pure mathematical curiosity, we are motivated by the question of *local vs. global correlation* in algorithm design [2, 1]. Generally speaking, mathematical relaxations of combinatorial optimization problems assign fractional values to elements or small subsets of the ground set, which can be interpreted as probabilities of certain local configurations. Hence a fractional solution can be viewed as a collection of local distributions. The question is whether these distributions can be realized globally, on the entire ground set. Usually this is possible only with some loss in the objective function, which leads to the design of approximation algorithms. In this paper, we study the basic question of characterizing the local distributions that can be realized globally for all pairs given by a certain graph.

[1] Take the distribution $(X, Y) \sim \rho$ on $\{0, 1\}^2$ given by $\Pr[X = 1, Y = 0] = \Pr[X = 0, Y = 1] = 1/2$. In particular, $\Pr[X = Y] = 0$. However, for any pair (U, V) of i.i.d. Bernoulli variables, $\Pr[U = V] \geq \frac{1}{2}$. Therefore, ρ cannot be a mix of i.i.d. distributions.

Our Results. First, we establish some basic properties of G-realizability. We show that it suffices to consider the *core* of G, that is, a minimal induced subgraph H of G such that H and G are homomorphically equivalent, whence being G-realizable is equivalent to being H-realizable. In particular, G-realizability for a bipartite graph G does not impose any restrictions at all. For arc-transitive directed graphs and for the analogous property of symmetric undirected graphs, we completely characterize the class of distributions that are G-realizable.

Our main results can be viewed as variants of the Diaconis-Freedman theorem (for $k = 2$), under weaker assumptions. For example, if ρ is K_n-realizable, the Diaconis-Freedman theorem shows that the distribution tends to an i.i.d. mix as $n \to \infty$. We show that even realizability on much sparser graphs leads to the same conclusion.

Realizability on Graphs with Spectral Properties. In the following, $\lambda_1 \geq \ldots \geq \lambda_n$ denote the eigenvalues of the adjacency matrix of G. For comparison with the Diaconis-Freedman theorem, recall that the eigenvalues of K_n are $\lambda_1 = n - 1$ and $\lambda_2 = \ldots = \lambda_n = -1$.

1. We prove that if ρ is G_n-realizable for a sequence of graphs such that $\lim_{n\to\infty} \frac{\lambda_n(G_n)}{\lambda_1(G_n)} = 0$, then ρ is described by a probability matrix that is positive-semidefinite. For random variables on domains of size $|\mathcal{D}| \leq 4$, this implies that ρ must be an i.i.d. mix.
2. If ρ is G_n-realizable for a sequence of (n, d, λ)-graphs G_n (d-regular with all eigenvalues except for one bounded by λ in absolute value) such that $\lim_{n\to\infty} \frac{\lambda(G_n)}{d(G_n)} = 0$, then ρ is an i.i.d. mix.
3. If ρ is $\vec{G}_n$-realizable for a sequence of directed graphs such that each of them is an arbitrary orientation of an (n, d, λ)-graph G_n, and $\lim_{n\to\infty} \frac{\lambda(G_n)}{d(G_n)} = 0$, then ρ is an i.i.d. mix.

Let us discuss some aspects of the above results. It is easy to see that being realizable on a bipartite graph G does not impose any restrictions at all and hence some condition forbidding bipartiteness is necessary to derive any non-trivial result. Result (1) applies to graphs that are "far from bipartite" in the sense that the normalized minimum eigenvalue is close to 0. (See [20] for an explicit connection.) More precisely (see Section 3.1), our result states that every G-realizable distribution is within a $|\frac{\lambda_n(G)}{\lambda_1(G)}|$-distance of a distribution whose probability matrix is positive semidefinite. For domains of size up to 4, this imples that ρ must be close to an i.i.d. mix (due to a connection between doubly nonnegative and completely positive matrices which we discuss in Section 3.2). It is an interesting question whether there is a distribution on a domain larger than 4 that is not an i.i.d. mix and realizable on a sequence of graphs such that $\lim_{n\to\infty} \frac{\lambda_n(G_n)}{\lambda_1(G_n)} = 0$.

Result (2) applies to all finite domains, by making a stronger assumption: Here, the distribution should be realizable on a family of *pseudorandom graphs*, defined in terms of normalized eigenvalues (see Section 3.3). The quantitative bound that we prove here is that a distribution on domain $\mathcal{D}$ realizable on an (n, d, λ)-graph must be $\frac{\lambda}{d}|\mathcal{D}|$-close to an i.i.d. mix. In contrast to the Diaconis-Freedman theorem, we have a dependence on $|\mathcal{D}|$ here which seems to be necessary. We show that for any (fixed) symmetric and triangle-free pseudorandom graph G_n, there is a canonical G_n-realizable distribution on a domain of size n which is at distance 1/2 from any i.i.d. mix. However, if a distribution on a fixed domain $\mathcal{D}$ is realizable on a family of pseudorandom graphs with $\frac{\lambda}{d} \to 0$, then it must be an i.i.d. mix.

Finally, (3) is our most technically involved result (which in the limit form subsumes result (2); however, the quantitative bounds here are much weaker). It shows that if the second normalized eigenvalue for a sequence $\{G_n\}$ tends to zero, then it is sufficient to assume just that ρ is $\vec{G}_n$-realizable for some *arbitrary orientation* of each G_n. In other words,

$(X_i, X_j) \sim \rho$ for some arbitrary direction of each edge $(i, j) \in E$, and this already implies that ρ is an i.i.d. mix. This is motivated by a version of de Finetti's theorem which holds under the weaker hypothesis that for every n, the marginal distributions induced by all *strictly increasing* sequences $(i_1, i_2, \ldots, i_n)$ are identical. (See [12] for a short proof using ergodic theory, or [21] for a finite version of this statement.) In particular, [21] proves that if $(X_i, X_j) \sim \rho$ for every $i < j$ for sufficiently many random variables, then ρ must be close to an i.i.d. mix. Our result implies that if $(X_i, X_j) \sim \rho$ for at least one (arbitrary) ordering of each pair $i, j \in [n]$, i.e. ρ is $\vec{T}$-realizable for some *tournament* $\vec{T}$, then ρ must be close to an i.i.d. mix. More generally, we prove this for orientations of pseudorandom graphs. Apart from properties of pseudorandom graphs, the proof of this result relies on the sparse regularity lemma, together with a counting argument adapted from [21].

1.1 Characterization of Realizable k-variate Distributions

We state the following (formal) strengthening of the Diaconis-Freedman theorem that illustrates the relationship between the notions of realizability and exchangeability, which are formally different but in some contexts closely related.

▶ **Proposition 1.** *Fix a k-variate distribution ρ. Suppose there are random variables $X_1, X_2, \ldots, X_n$ such that the marginal distribution of $(X_{i_1}, X_{i_2} \ldots, X_{i_k})$ equals ρ for every k-tuple $(i_1, i_2, \ldots, i_k)$ of distinct indices in $[n]$. Then ρ is $\frac{k(k-1)}{2n}$-close in variation distance to an i.i.d. mix. Therefore, if this property holds for all n, then ρ is an i.i.d. mix.*

Note that we assume only that every k distinct indices realizes the same distribution, rather than a full exchangeability of the sequence. Nonetheless, the result follows easily from the Diaconis-Freedman theorem as follows: Take a random permutation π of $[n]$ and consider the sequence $V_1, V_2, \ldots, V_n$, where $V_i = X_{\pi(i)}$ for $i \in [n]$. Then $V_1, V_2, \ldots, V_n$ is an exchangeable sequence that also realizes ρ on all k distinct indices. Therefore, ρ is $\frac{k(k-1)}{2n}$-close to an i.i.d. mix by the Diaconis-Freedman theorem. In the following, we give a short self-contained proof to illustrate the techniques that will be useful in the sequel. We will use the following standard fact from information theory.

▶ **Proposition 2** (Data-processing inequality for total variation distance)**.** *Let A, B, X be random variables such that X is independent of (A, B). Then for all functions h:*

$$d_{TV}(h(A, X), h(B, X)) \leq \max_x d_{TV}(h(A, x), h(B, x)) \leq d_{TV}(A, B)$$

Proof of Proposition 1. Pick uniformly a random k-tuple of distinct indices $(i_1, \ldots, i_k)$, and let $(Y_1, \ldots, Y_k) = (X_{i_1}, \ldots, X_{i_k})$. By the realizability property, $(Y_1, \ldots, Y_k) \sim \rho$. Define another distribution as follows: pick uniformly random and *independent* indices $j_1, \ldots, j_k \in [n]$ and define $(Z_1, \ldots, Z_k) = (X_{j_1}, \ldots, X_{j_k})$. Let $(Z_1, \ldots, Z_k) \sim \mu$.

We show that the two distributions are close to each other in variation distance. Note that the two distributions are obtained by choosing k indices by sampling with or without replacement, respectively, and then followed by applying the same (random) function $f(\ell) = X_\ell$ to each selected index ℓ. Since the sampling process was chosen independently of $X_1, X_2, \ldots, X_n$, the data-processing inequality implies that the variation distance between μ and ρ is at most that between sampling k times from $[n]$ with and without replacement. It is easy to see that this variation distance is upper-bounded by the probability that there are at least two equal indices among k independent samples from $[n]$; by the union bound, this probability is at most $\binom{k}{2}\frac{1}{n} = \frac{k(k-1)}{2n}$. ◀

2 Realizability on Graphs: Basic Properties

A distribution ρ over $\mathcal{D} \times \mathcal{D}$ is *$\vec{G}$-realizable* for a directed graph $\vec{G} = (V, A)$, if there is a joint distribution of random variables $(X_v : v \in V)$ such that $(X_i, X_j) \sim \rho$, for all arcs $(i,j) \in A$. Extend this definition to undirected graphs G by requiring that ρ is symmetric and that $(X_i, X_j) \sim \rho$, for all edges $\{i,j\} \in E$. We develop a few basic properties of G-realizability.

2.1 Realizability and Homomorphisms

A *homomorphism* from G to H is a mapping $f : V(G) \to V(H)$ such that $\{i,j\} \in E(G) \Rightarrow \{f(i), f(j)\} \in E(H)$ (for undirected graphs); and $(i,j) \in A(G) \Rightarrow (f(i), f(j)) \in A(H)$ (for directed graphs), for all i, j. We write $G \to H$ if such a homomorphism exists. We observe the following simple connection between realizability and homomorphisms.

▶ **Lemma 3.** *If ρ is H-realizable and $G \to H$ (directed or undirected), then ρ is G-realizable.*

Proof. Let $(Y_v : v \in V(H))$ be a collection of random variables such that for each edge $(i,j) \in E(H)$, (Y_i, Y_j) is distributed according to ρ. Let $f : G \to H$ be a homomorphism. Then we define a collection of random variables $(X_u : u \in V(G))$ such that $X_u = Y_{f(u)}$. For each edge $(i,j) \in E(G)$, we have $(f(i), f(j) \in E(H)$ and hence $(X_i, X_j) = (Y_{f(i)}, Y_{f(j)})$ is distributed according to ρ. This proves that ρ is G-realizable. ◀

It is well known (e.g., Proposition 3.5 in [17]) that for every graph G (directed or undirected), there is a unique minimal graph H such that $G \to H$ and $H \to G$. We call H the *core* of G and denote it as $C(G)$.

▶ **Corollary 4.** *A distribution ρ is G-realizable if and only if ρ is $C(G)$-realizable.*

In other words, it is only the core $C(G)$ that determines what distributions are G-realizable. For example, core of each bipartite graph is K_2 (a single edge), where every symmetric distribution is realizable. Containing a clique K_q is equivalent to the existence of a homomorphism $K_q \to G$. In particular, having a clique number $\omega(G) = q$ means that every G-realizable distribution satisfies the assumptions of Proposition 1 with $k = 2, n = q$, and is at distance at most $1/q$ from an i.i.d. mix. Conversely, having chromatic number $\chi(G) = k$ is equivalent to $G \to K_k$. This implies that any distribution realizable on K_k is also realizable on G. In particular, sampling twice without replacement from $[k]$ is realizable on G, which is at distance $1/k$ from an i.i.d. mix.

2.2 Symmetric (Edge/Arc-transitive) Graphs

Next, we consider the class of symmetric graphs, which possess a canonical realizable distribution — one corresponding to the adjacency matrix of the graph itself. We prove that in some sense all distributions realizable on symmetric graphs arise in this fashion.

▶ **Definition 5.** A directed graph $G = (V, E)$ is arc-transitive, if for each pair of directed edges $(u, v), (w, z) \in E$, there is an automorphism $\pi : V \to V$ such that $\pi(u) = w$ and $\pi(v) = z$. An undirected graph $G = (V, E)$ is symmetric, if for each pair of edges $\{u, v\}, \{w, z\} \in E$ and each pairing $u - w, v - z$ of their endpoints, there is an automorphism $\phi : V \to V$ such that $\pi(u) = w$ and $\pi(v) = z$.

We now characterize all G-realizable distributions for symmetric graphs. In short, the distributions are obtained by labeling the vertices of G arbitrarily with values in $\mathcal{D}$ and taking the labeling of a uniformly random edge. We need a preliminary fact.

▶ **Lemma 6.** *For an arc-transitive directed graph $\vec{G} = (V, A)$ and any arc $(u, v) \in A$, if π is a uniformly random automorphism of $\vec{G}$ then $(\pi(u), \pi(v))$ is a uniformly random arc in E.*

Proof. Let $\mathcal{A}$ denote the set of automorphisms of $\vec{G}$. Fix $(u, v) \in A$. For every arc $(w, z) \in A$ (possibly equal to (u, v)), define $\mathcal{A}_{wz} = \{\pi \in \mathcal{A} : \pi(u) = w, \pi(v) = z\}$. Fix an automorphism $\pi_1 \in \mathcal{A}_{wz}$ (which exists by arc-transitivity) and define $\phi : \mathcal{A} \to \mathcal{A}$ by $\phi(\pi) = \pi_1 \circ \pi$. We claim that ϕ is a bijection between $\mathcal{A}_{uv}$ and $\mathcal{A}_{wz}$: For each $\pi_0 \in \mathcal{A}_{uv}$, $(\phi(\pi_0))(u) = \pi_1(\pi_0(u)) = \pi_1(u) = w$ and $(\phi(\pi_0))(v) = \pi_1(\pi_0(v)) = z$. Similarly, $\phi^{-1}(\pi) = \pi_1^{-1} \circ \pi$ is the inverse of ϕ and maps $\mathcal{A}_{wz}$ to $\mathcal{A}_{uv}$. Therefore, $|\mathcal{A}_{uv}| = |\mathcal{A}_{wz}|$ and this holds for every $(w, z) \in A$. Thus a random automorphism is equally likely to map (u, v) to any other arc. ◀

▶ **Theorem 7.** *Let $G = (V, E)$ be a symmetric (undirected) or arc-transitive (directed) graph. Then ρ is a G-realizable distribution on $\mathcal{D} \times \mathcal{D}$ if and only if ρ is a convex combination of distributions ρ_f, where ρ_f for $f : V \to \mathcal{D}$ is defined as follows: (X, Y) is distributed according to ρ_f, if $(X, Y) = (f(u), f(v))$ where (u, v) is a uniformly random arc/edge of E (randomly ordered in the undirected case).*

Proof. First, if G is undirected, let us replace it by its bidirected version $\vec{G}$. Observe that a uniformly random arc (u, v) in $\vec{G}$ corresponds to a random ordering of a uniformly random edge in G, therefore our definition of ρ_f for directed/undirected graphs is consistent with this reduction to the directed case.

We prove that the distribution ρ_f is $\vec{G}$-realizable for every $f : V \to \mathcal{D}$. We define a random variable X_v for each vertex $v \in V$: $X_v = f(\pi(v))$ where π is a uniformly random automorphism of G. By Lemma 6, $(\pi(u), \pi(v))$ for any fixed edge $(u, v) \in E$ is uniformly distributed over all edges in E (uniformly over both orderings in the undirected case). Therefore, $(X_u, X_v) = (f(\pi(u)), f(\pi(v)))$ is distributed according to ρ_f.

Conversely, assume that ρ is $\vec{G}$-realizable and let $\{X_v : v \in V\}$ be a collection of random variables realizing ρ on each edge $(u, v) \in E$. Consider a pair of random variables (Y, Z) generated by $(Y, Z) = (X_u, X_v)$ where (u, v) is a uniformly random edge in E (randomly ordered in the undirected case). Since conditioned on (u, v), the distribution of (X_u, X_v) is ρ, the distribution of (Y, Z) is also ρ. We claim that the distribution of (Y, Z) is a convex combination of distributions ρ_f as in the statement above. To see this, consider some assignment of values $X_v = f(v)$ of nonzero probability. Conditioned on $X_v = f(v)\ \forall v \in V$, we have $(Y, Z) = (f(u), f(v))$ where (u, v) is a random edge — i.e., (Y, Z) is distributed according to ρ_f. Therefore, ρ is a convex combination of such distributions. ◀

In particular, the distribution ρ_{Id} on $\mathcal{D} = V$, obtained by taking $(X, Y) =$ *uniformly random edge of* G is G-realizable for any symmetric graph. We call ρ_{Id} the *canonical G-realizable distribution*. Using a classical theorem of Motzkin and Straus [16], we can show that the exact distance of ρ_{Id} from an i.i.d. mix is determined by the clique number of G.

▶ **Theorem 8.** *For an undirected symmetric graph G, the canonical G-realizable distribution ρ_{Id} is at variation distance exactly $1/\omega(G)$ from an i.i.d. mix, where $\omega(G)$ is the maximum size of a clique in G. This is the maximum distance from an i.i.d. mix among all G-realizable distributions.*

Proof. Let A be the adjacency matrix of G. The Motzkin-Straus theorem [16] states that

$$\max_{\|p\|_1, p \geq 0} p^T A p = 1 - \frac{1}{\omega(G)}.$$

Therefore, for any probability distribution p on V, we have $\sum_{a,b \in V} p_a p_b A_{ab} \leq 1 - \frac{1}{\omega(G)}$. By taking convex combinations, for any i.i.d. mix $\mu_{ab} = \sum \alpha_i p_a^i p_b^i$, we still have $\sum_{a,b \in V} \mu_{ab} A_{ab} \leq$

$1 - \frac{1}{\omega(G)}$. In other words, μ_{ab} has at most $1 - \frac{1}{\omega(G)}$ probability mass on the edges of G. However, the canonical G-realizable distribution ρ_{Id} is supported on the edges of G. Therefore, it is at distance at least $\frac{1}{\omega(G)}$ from μ. As we discussed above, every G-realizable distribution ρ must be at distance at most $\frac{1}{\omega(G)}$ from an i.i.d. mix, due to Proposition 1. Therefore, ρ_{Id} has distance exactly $\frac{1}{\omega(G)}$ from an i.i.d. mix. ◀

▶ **Example 9.** For a cycle C_n, the canonical C_n-realizable distribution ρ is defined as follows: $\rho(i, i+1 \bmod n) = \rho(i, i-1 \bmod n) = \frac{1}{2n}$; this is at variation distance 1/2 from an i.i.d. mix.

3 Realizability Based on Spectral Properties

In this section we consider undirected graphs. We showed in Theorem 7 that each fixed symmetric graph G possesses a rather rich collection of G-realizable distributions, similar to the structure of the graph itself. However, perhaps a more interesting question is: What fixed distributions are G_n-realizable for a family of graphs $\{G_n\}$ of growing size? Here we do not have many non-trivial examples, other than those where all the graphs G_n map homomorphically to a fixed symmetric graph H, and then we can realize all the H-realizable distributions. Similar to de Finetti's theorem, one can ask — what are the families of graphs that admit only i.i.d. mix distributions to be realized?

▶ **Example 10.** Consider any family of d-regular graphs $\{G_n\}$, for example d-regular expanders (even Ramanujan graphs). These graphs are $(d+1)$-colorable, therefore any K_{d+1}-realizable distribution is also realizable on each G_n. In particular, sampling from $[d+1]$ without replacement is realizable and at a fixed distance $\frac{1}{d+1}$ from an i.i.d. mix; i.e this family does not force a realizable distribution to be an i.i.d. mix.

On the other hand, even though this family may not contain any clique K_{d+1}, even any triangles, it seems to force realizable distributions to be quite close to an i.i.d. mix. This motivates us to investigate the relationship of realizability and spectral properties of graphs.

▶ **Definition 11.** The eigenvalues of a graph G are the eigenvalues of its adjacency matrix $A(G)$, $(A(G))_{ij} = 1$ if $\{i,j\} \in E(G)$ and 0 otherwise.

It is known that the eigenvalues are all real and contained in $[-\Delta, \Delta]$ where Δ is the maximum degree in G. We order the eigenvalues in a descending order and label them $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. By the trace formula, we have $\sum_{i=1}^{n} \lambda_i = 0$. The gap between λ_1 and λ_2 is related to expansion properties of G, while the minimum (most negative) eigenvalue λ_n is related to how close G is to a bipartite graph. For a bipartite graph, we have $\lambda_n = -\lambda_1$. Graphs where λ_n/λ_1 is close to zero are those where "MaxCutGain" is small: for any two disjoint sets $A, B \subset V$, the number of edges between A and B is not significantly larger than the number of edges inside A and B (see [20]). Thus these graphs can be considered "far from bipartite". In the following, we prove that this property imposes a natural condition on what distributions are G-realizable.

3.1 Graphs Without Large Negative Eigenvalues

We begin with a result that is proved using a standard spectral argument.

▶ **Lemma 12.** *Let ρ be a G-realizable distribution on $\mathcal{D} \times \mathcal{D}$ for a graph $G = (V, E)$, and let μ be the marginal distribution of X in $(X, Y) \sim \rho$. Then for any function $\phi : \mathcal{D} \to \mathbb{R}$,*

$$\mathop{\mathbb{E}}_{(X,Y)\sim\rho}[\phi(X)\phi(Y)] \geq \frac{\lambda_n(G)}{\lambda_1(G)} \mathop{\mathbb{E}}_{X\sim\mu}[\phi^2(X)].$$

Proof. We can assume that G has no isolated vertices (which add only zero eigenvalues to the spectrum). Let $(X_v : v \in V)$ be a collection of random variables realizing ρ on G. Since the distribution of (X_i, X_j) is ρ for each edge $(i,j) \in E$, we have

$$\mathop{\mathbb{E}}_{(X,Y)\sim\rho}[\phi(X)\phi(Y)] = \frac{1}{|E|}\sum_{\{i,j\}\in E}\mathop{\mathbb{E}}_{(X_v:v\in V)}[\phi(X_i)\phi(X_j)] = \frac{1}{|E|}\mathop{\mathbb{E}}_{(X_v:v\in V)}[\sum_{\{i,j\}\in E}\phi(X_i)\phi(X_j)].$$

Let A be the adjacency matrix of G. The sum inside the expectation is a quadratic form which can be lower-bounded using the minimum eigenvalue of A:

$$\sum_{\{i,j\}\in E}\phi(X_i)\phi(X_j) = \frac{1}{2}\sum_{i,j\in V}\phi(X_i)A_{ij}\phi(X_j) \geq \frac{1}{2}\lambda_n\sum_{i\in V}\phi^2(X_i).$$

Let us take the expectation over the distribution of $(X_v : v \in V)$. Using the fact that each $v \in V$ is in some edge and hence the marginal distribution of X_v is μ, we obtain

$$\mathop{\mathbb{E}}_{(X,Y)\sim\rho}[\phi(X)\phi(Y)] \geq \frac{\lambda_n}{2|E|}\mathop{\mathbb{E}}_{(X_v:v\in V)}[\sum_{i\in V}\phi^2(X_i)] = \frac{|V|\lambda_n}{2|E|}\mathop{\mathbb{E}}_{X\sim\mu}[\phi^2(X)] \leq \frac{\lambda_n}{\lambda_1}\mathop{\mathbb{E}}_{X\sim\mu}[\phi^2(X)],$$

since $2|E| = \mathbf{1}^T A\mathbf{1} \leq \lambda_1\|\mathbf{1}\|^2 = \lambda_1|V|$. This proves the lemma. ◀

Using the above lemma, we can perturb the distribution ρ and obtain a probability matrix which is positive semidefinite.

▶ **Theorem 13.** *If ρ is a G-realizable distribution, then ρ is at distance at most $|\frac{\lambda_n(G)}{\lambda_1(G)}|$ from a distribution ρ' whose probability matrix $\rho'_{ab} = \Pr_{(X,Y)\sim\rho'}[X = a, Y = b]$ is positive semidefinite.*

Proof. Let ρ be a G-realizable distribution on $\mathcal{D}\times\mathcal{D}$ and let μ be its marginal on the first variable, as in Theorem 12. Let $\delta_{ab} = 1$ if $a = b$ and 0 otherwise. We define

$$\rho'_{ab} = \frac{\rho_{ab} + |\frac{\lambda_n}{\lambda_1}|\mu_a\delta_{ab}}{1 + |\frac{\lambda_n}{\lambda_1}|}.$$

It can be checked easily that $\rho'_{ab} \geq 0$ and $\sum_{a,b\in\mathcal{D}}\rho'_{ab} = 1$, so this is a probability distribution. The total variantion distance between ρ and ρ' is

$$\begin{aligned}\frac{1}{2}\|\rho-\rho'\|_1 &= \frac{1}{2}\sum_{a,b\in\mathcal{D}}|\rho_{ab}-\rho'_{ab}| = \frac{1}{2(1+|\frac{\lambda_n}{\lambda_1}|)}\sum_{a,b\in\mathcal{D}}|(1+\left|\frac{\lambda_n}{\lambda_1}\right|)\rho_{ab} - (\rho_{ab}+\left|\frac{\lambda_n}{\lambda_1}\right|\mu_a\delta_{ab})| \\ &= \frac{|\frac{\lambda_n}{\lambda_1}|}{2(1+|\frac{\lambda_n}{\lambda_1}|)}\sum_{a,b\in\mathcal{D}}|\rho_{ab}-\mu_a\delta_{ab}| \leq \frac{|\frac{\lambda_n}{\lambda_1}|}{1+|\frac{\lambda_n}{\lambda_1}|}.\end{aligned}$$

We claim that ρ' is a positive-semidefinite matrix: For any $\phi : \mathcal{D} \to \mathbb{R}$, we have

$$\sum_{a,b\in\mathcal{D}}\rho'_{ab}\phi(a)\phi(b) = \frac{1}{1+|\frac{\lambda_n}{\lambda_1}|}\left(\sum_{a,b\in\mathcal{D}}\rho_{ab}\phi(a)\phi(b) + \left|\frac{\lambda_n}{\lambda_1}\right|\sum_{a\in\mathcal{D}}\mu_a\phi^2(a)\right).$$

By Theorem 12, $\sum_{a,b\in\mathcal{D}}\rho_{ab}\phi(a)\phi(b) \geq \frac{\lambda_n}{\lambda_1}\sum_{a\in\mathcal{D}}\mu_a\phi^2(a)$ (which is a negative number), so we obtain $\sum_{a,b\in\mathcal{D}}\rho'_{ab}\phi(a)\phi(b) \geq 0$. ◀

This has the following corollary which is one of the results claimed in the introduction.

▶ **Corollary 14.** *Let ρ be a distribution over $\mathcal{D} \times \mathcal{D}$ such that for an arbitrarily large n, ρ is G_n-realizable for a graph G_n such that $|V(G_n)| = n$ and*

$$\lim_{n\to\infty} \frac{\lambda_n(G_n)}{\lambda_1(G_n)} = 0,$$

then $\rho_{ab} = \Pr_{(X,Y)\sim\rho}[X = a, Y = b]$ is a positive-semidefinite matrix.

Therefore, distributions ρ realizable on a sequence of graphs with normalized minimum eigenvalue tending to 0 must be positive semidefinite. Since $\sum_{a,b\in\mathcal{D}} \rho_{ab} = 1$ and $\rho_{ab} \geq 0$, this seems close to the condition of being an i.i.d. mix: $\rho_{ab} = \sum_i q_i p_i(a) p_i(b)$ for distributions $\sum_i q_i = 1$ and $\sum_{a\in\mathcal{D}} p_i(a) = 1$. However, these two conditions have been extensively studied and they are not equivalent.

3.2 Completely Positive and Doubly Nonnegative Matrices

▶ **Definition 15.** A matrix $A \in \mathbb{R}^{n\times n}$ is *completely positive* if $\forall i,j; A_{ij} = \mathbf{v}_i \cdot \mathbf{v}_j$ for nonnegative vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^n_+$. A matrix $A \in \mathbb{R}^{n\times n}$ is *doubly nonnegative* if A is positive semidefinite and $\forall i,j; A_{ij} \geq 0$.

In our setting, completely positive matrices correspond to i.i.d. mixes (up to normalization), while doubly nonnegative matrices correspond to the distributions arising in Corollary 14. Clearly, every completely positive matrix is doubly nonnegative, but the opposite is not true. Nonetheless, the smallest known counterexamples are 5×5 matrices, and in fact for matrices up to 4×4 the two conditions are equivalent.

▶ **Theorem 16** ([15])**.** *A matrix in $\mathbb{R}^{4\times 4}$ is completely positive if and only if it is doubly nonnegative.*

Therefore, we obtain the following corollary for domains of size up to 4 (in particular, Boolean random variables).

▶ **Corollary 17.** *If ρ is a distribution on $\mathcal{D} \times \mathcal{D}$, $|\mathcal{D}| \leq 4$, and ρ is G_n-realizable for an arbitrarily large n on graphs G_n such that $|V(G_n)| = n$ and*

$$\lim_{n\to\infty} \frac{\lambda_n(G_n)}{\lambda_1(G_n)} = 0,$$

then ρ is an i.i.d. mix.

Proof. By Corollary 14, $\rho_{ab} \in \mathbb{R}^{\mathcal{D}\times\mathcal{D}}$ is a doubly nonnegative matrix, and hence for $|\mathcal{D}| \leq 4$ it is completely positive: $\rho_{ab} = \mathbf{v}_a \cdot \mathbf{v}_b$ for vectors $\mathbf{v}_a \geq 0$. For each coordinate i, let $q_i = \sum_{a\in\mathcal{D}} v_{ai}$ and define $p_i(a) = v_{ai}/q_i$ (we can assume $q_i > 0$, otherwise we remove that coordinate). We have $p_i(a) \geq 0$ and $\sum_{a\in\mathcal{D}} p_i(a) = \sum_{a\in\mathcal{D}} v_{ai}/q_i = 1$, so each $p_i(a)$ is a distribution on $\mathcal{D}$ and we have $\rho_{ab} = \sum_i q_i p_i(a) p_i(b)$. Since $\sum_{a,b\in\mathcal{D}} \rho_{ab} = 1$, it is easy to verify that $\sum q_i = 1$ as well. ◀

3.3 Realizability on Pseudorandom Graphs

Next, we show that if we assume that all eigenvalues except for one are close to 0, then G-realizable distributions must be close to an i.i.d. mix. We follow the exposition on pseudorandom graphs in [13].

▶ **Definition 18.** G is an (n, d, λ)-graph if G is a d-regular graph on n vertices such that all eigenvalues except for the largest one are bounded by λ in absolute value.

Note that the largest eigenvalue itself equals d. The definition above implies pseudorandom properties whenever λ/d is small. This is a rather strong definition of pseudorandomness; however, most of the known constructions of pseudorandom graphs fall in this category.

▶ **Definition 19.** For an undirected graph G and two sets of vertices $S, T \subseteq V(G)$ (not necessarily disjoint), we define $e(S,T) = |\{u \in S, v \in T : \{u,v\} \in E(G)\}|$ to denote the number of edges between S and T. For a directed graph $\vec{G}$, we define $\vec{e}(S,T) = |\{u \in S, v \in T : (u,v) \in E(G)\}|$ to denote the number of arcs from S to T.

Note that in the undirected case, each edge inside $S \cap T$ is counted twice in $e(S,T)$. In the directed case, each arc inside $S \cap T$ is counted once in $\vec{e}(S,T)$. (This is consistent with the view that an undirected graph can be viewed as a directed graph by replacing each undirected edge with the two arcs representing possible orientations of that edge.)

Next, we formulate the well-known *expander mixing lemma* (for undirected graphs).

▶ **Proposition 20** (Expander Mixing Lemma [13]). *Let G be an (n,d,λ)-graph. Then for any $S, T \subseteq V(G)$,*

$$\left| e(S,T) - \frac{d}{n}|S| \cdot |T| \right| \leq \frac{\lambda}{n}\sqrt{|S|(n-|S|)|T|(n-|T|)} \leq \lambda\sqrt{|S||T|}$$

In particular, this implies that (n,d,λ)-graphs for small values of $\frac{\lambda}{d}$ are good expanders (by taking $T = V \setminus S$).

▶ **Theorem 21.** *Let G be an (n,d,λ)-graph. If a distribution ρ is G-realizable, then ρ is $(\frac{\lambda}{d} \cdot \sqrt{|\mathcal{D}|})$-close in variation distance to an i.i.d. mix.*

Proof. Let ρ be a G-realizable distribution and let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ certify the G-realizability of ρ. Let $\vec{G}$ denote the directed graph obtained by replacing each undirected edge by two arcs in opposite directions. Choose an edge $\vec{e} = (i,j)$ uniformly from $E(\vec{G})$ which consists of nd arcs. By G-realizability, it follows that the distribution of (X_i, X_j) also equals ρ. As before, conditioned on $\mathbf{X} = \mathbf{x}$, for a uniformly chosen vertex $i \in V(G)$ the distribution of X_i is the empirical distribution $p_{\mathbf{x}}$. So the distribution μ of (X_i, X_j) when (i,j) is chosen uniformly from $[n]^2$ is an i.i.d. mix, a convex combination of products of such empirical distributions.

We now show that $d_{TV}(\mu, \rho)$ is small. By the data-processing inequality, it suffices to show that this holds when conditioned on $\mathbf{X} = \mathbf{x}$, for every $\mathbf{x}$. Let $\phi = (\phi_{ab})$ (respectively, ψ) be the probability distribution of (x_i, x_j) when when (i,j) is uniformly random in $[n]^2$ (respectively, (i,j) is uniformly random in $E(\vec{G})$). Let V_a denote the set of vertices labeled a, for each $a \in \mathcal{D}$. Note that the label sets partition V. Then $\phi_{ab} = |V_a| \cdot |V_b|/n^2$. On the other hand, it can be checked for both the cases $a = b$ and $a \neq b$ that $\psi_{ab} = e(V_a, V_b)/(nd)$.

Let $Q = \{(a,b) \in \mathcal{D} : \phi(a,b) < \psi(a,b)\}$ so that $d_{TV}(\phi, \psi) \leq \sum_{(a,b)\in Q}(\psi_{ab} - \phi_{ab})$. Fix $a \in \mathcal{D}$ and let $W_a = \bigcup_{b:(a,b)\in Q} V_b$. Observe that $\sum_{b:(a,b)\in Q} e(V_a, V_b) = e(V_a, W_a)$, since the V_a's are pairwise disjoint. Similarly $\sum_{b:(a,b)\in Q} |V_b| = |W_a|$. By the Expander Mixing Lemma:

$$\sum_{b:(a,b)\in Q} (\psi_{ab} - \phi_{ab}) = \frac{1}{nd}\Big(e(V_a, W_a) - \frac{d}{n}|V_a| \cdot |W_a|\Big) \leq \frac{\lambda}{nd}\sqrt{|V_a| \cdot |W_a|} \leq \frac{\lambda}{d}\sqrt{\frac{|V_a|}{n}} \tag{1}$$

By the Cauchy-Schwarz inequality, $\sum_{a\in\mathcal{D}} \sqrt{|V_a|/n} \leq \sqrt{|\mathcal{D}|} \cdot \sqrt{\sum_{a\in\mathcal{D}} |V_a|/n} = \sqrt{|\mathcal{D}|}$, because the label sets partition V. Using this bound, we conclude the proof using eq. (1) as follows:

$$d_{TV}(\phi, \psi) \leq \sum_{a\in\mathcal{D}} \sum_{b:(a,b)\in Q} (\psi_{ab} - \phi_{ab}) \leq \frac{\lambda}{d} \sum_{a\in\mathcal{D}} \sqrt{\frac{|V_a|}{n}} \leq \frac{\lambda}{d} \cdot \sqrt{|\mathcal{D}|}$$ ◀

▶ **Corollary 22.** *If for arbitrarily large n, ρ is G_n-realizable on an $(n, d(n), \lambda(n))$-graph G_n such that*

$$\lim_{n\to\infty} \frac{\lambda(n)}{d(n)} = 0$$

then ρ is an i.i.d. mix.

In other words, the only distributions realizable on a family of pseudorandom graphs with $\frac{\lambda}{d} \to 0$ are i.i.d. mixes. However, it is not true that a distribution realizable on a single pseudorandom graph with very good parameters must be close to an i.i.d. mix. In contrast to the Diaconis-Freeman theorem, some dependence on $|\mathcal{D}|$ seem to be necessary in Theorem 21.

▶ **Example 23.** Consider a *symmetric* (n, d, λ)-graph G. It is known that there are such graphs with arbitrarily small λ/d (see e.g. [11, 9]; these graphs are described as edge-transitive but in fact can be seen to be symmetric in the stronger sense of Theorem 5 as well). In addition, these graphs have high girth [14, 10]; for us it is sufficient that they are triangle-free. Consider the canonical distribution $(X, Y) \sim \rho$ where $(X, Y) = (u, v)$ is a (randomly ordered) uniformly random edge of G. This is G-realizable by Theorem 7. By Theorem 8, this distribution has variation distance $1/\omega(G) = 1/2$ from any i.i.d. mix.

4 Orientations of Pseudorandom Graphs

Since we know that realizability on pseudorandom graphs implies being an i.i.d. mix, one can ask whether it is really necessary to require the distribution ρ to be symmetric to start with. Perhaps being realizable on a suitable directed graph already implies being an i.i.d. mix and in particular being symmetric?

First, we have the following simple lemma which shows that in non-trivial cases the marginals of ρ must be symmetric.

▶ **Lemma 24.** *Let ρ be $\vec{G}$-realizable for a non-bipartite directed graph $\vec{G}$. Then $\forall a \in \mathcal{D}$:*

$$\Pr_{(X,Y)\sim\rho}[X = a] = \Pr_{(X,Y)\sim\rho}[Y = a].$$

Proof. We claim that there is a vertex v that is a head and also a tail of some edge: $\exists u, w \in V; (u, v) \in A(\vec{G})$ and $(v, w) \in A(\vec{G})$. If not, the orientation of $\vec{G}$ is such that vertices can be divided into head-only and tail-only; but this means that $\vec{G}$ is bipartite.

Thus, we have random variables X_u, X_v, X_w such that $(X_u, X_v) \sim \rho$ and also $(X_v, X_w) \sim \rho$. So, $\Pr_{(X,Y)\sim\rho}[X = a] = \Pr_{(X,Y)\sim\rho}[Y = a] = \Pr[X_v = a]$. ◀

However, this does not mean that ρ must be symmetric. An example is a directed cycle $\vec{C}_n$ with vertices identified with $\mathbb{Z}_n$, where we can have the following distribution: $X_i = Z + i \pmod n$, where Z is uniformly random in $\mathbb{Z}_n$. Then the distribution on each directed edge is given by $(X_i, X_{i+1}) = (j, j + 1)$ for each $j \in \mathbb{Z}_n$ with probability $1/n$. This distribution has symmetric marginals but it is not symmetric.

Nevertheless, if ρ is $\vec{G}_n$-realizable on a sufficiently dense directed graph $\vec{G}_n$, it seems to force ρ to be symmetric. One example of this is the transitive tournament $A = \{(i, j) : i, j \in [n], i < j\}$: If (X_i, X_j) has the same distribution for each $i < j$, then the distribution must be close to an i.i.d. mix. (This was proved implicitly by Trotter and Winkler [21].) Is being realizable on an arbitrary tournament sufficient to conclude that ρ must be close to an i.i.d. mix? In this section, we consider an even more general question: If ρ is realizable on some arbitrary orientation of a pseudorandom graph, does ρ have to be close to an i.i.d. mix? We prove that the answer is yes.

▶ **Theorem 25.** *Let ρ be a distribution such that for an arbitrarily large n, ρ is $\vec{G}_n$-realizable for some orientation $\vec{G}_n$ of an $(n, d(n), \lambda(n))$-graph and*

$$\lim_{n\to\infty} \frac{\lambda(n)}{d(n)} = 0.$$

Then ρ is an i.i.d. mix.

4.1 Directed Sparse Regularity Lemma

The main tool that we use here is a directed version of the *sparse regularity lemma*. The sparse regularity lemma of Kohayakawa and Rödl says roughly that any subgraph of a "well-behaved graph" of a certain density p, for example a pseudorandom graph, can be partitioned in such a way that most pairs of parts are regular with an error proportional to p. In addition, we need a directed version of this lemma. We follow the exposition in [7], Section 2.1.

▶ **Definition 26.** For a directed graph $\vec{G} = (V, A)$ and a parameter $p > 0$, we define the oriented p-density for a pair of sets $U, W \subseteq V$ as

$$d_p(U, W) = \frac{2\vec{e}(U, W)}{p|U||W|}.$$

▶ **Definition 27.** A directed graph $\vec{G} = (V, A)$ is (η, D, p)-bounded if, for any pair of disjoint sets $U, W \subseteq V$ with $|U|, |W| \geq \eta|V|$, we have $d_p(U, W) \leq D$.

▶ **Definition 28.** A pair of disjoint sets $U, W \subseteq V$ is (ϵ, p)-regular if for all $U' \subseteq U, |U'| \geq \epsilon|U|$ and $W' \subseteq W, |W'| \geq \epsilon|W|$, we have

$$|d_p(U', W') - d_p(U, W)| < \epsilon.$$

A partition $\mathcal{P} = \{V_0, V_1, \ldots, V_k\}$ of V is (ϵ, k, p)-regular if $|V_0| \leq \epsilon|V|$, $|V_1| = |V_2| = \ldots = |V_k|$ and for more than $(1-\epsilon)\binom{k}{2}$ pairs $\{i, j\} \subseteq [k], i \neq j$, we have that (V_i, V_j) and (V_j, V_i) are both (ϵ, p)-regular.

▶ **Lemma 29** (directed sparse regularity lemma, [7]). *For any real $\epsilon > 0$, $D > 1$ and integer $k_0 \geq 1$ there exists $\eta > 0$ and $K \geq k_0$ such that for every $0 < p \leq 1$, every (η, D, p)-bounded directed graph $\vec{G}$ admits an (ϵ, k, p)-regular partition for some $k_0 \leq k \leq K$.*

4.2 Application of the Sparse Regularity Lemma

Starting with an arbitrary orientation of a pseudorandom graph $\vec{G}$, we use the regularity lemma to identify a certain bipartite subgraph (A, B) of $\vec{G}$ where the orientation behaves also in a pseudorandom way, with a significant density β in one direction.

▶ **Lemma 30.** *For every $\epsilon \in (0, \frac{1}{2})$, there is $K \geq 2$ and $\gamma > 0$ such that given any orientation of an (n, d, λ)-graph $\vec{G}$ with $\lambda \leq \gamma d$, there are disjoint sets $A, B \subseteq V$, $|A| = |B| \geq \frac{n}{2K}$ and $\beta = \frac{\vec{e}(A,B)}{|A||B|} \geq \frac{d}{4n}$ such that*

$$\forall A' \subseteq A, B' \subseteq B; \left|\vec{e}(A', B') - \beta|A'||B'|\right| < 8\epsilon\beta|A||B|.$$

Proof. Fix $D = 4$ and $k_0 = 1$. Given $\epsilon \in (0,1)$, let $\eta > 0$ and $K \geq 1$ be the parameters given by Lemma 29. We set $\gamma = \min\{\eta, \frac{\epsilon}{4K}\}$. Then consider any orientation $\vec{G}$ of an (n,d,λ)-graph such that $\lambda \leq \gamma d$. We set $p = \frac{d}{n}$.

First we verify the condition of (η, D, p)-boundedness. Applying Proposition 20 to the undirected version of $\vec{G}$, we have

$$\left| e(S,T) - \frac{d}{n}|S||T| \right| \leq \lambda\sqrt{|S||T|} \leq \eta d\sqrt{|S||T|}$$

where $e(S,T) = \vec{e}(S,T) + \vec{e}(T,S)$ denotes the number of all edges between S and T (both directions). Our goal is to bound the normalized density $d_p(S,T) = \frac{2\vec{e}(S,T)}{p|S||T|}$. We have

$$\vec{e}(S,T) \leq \frac{d}{n}|S||T| + \eta d\sqrt{|S||T|} = p|S||T| + \eta pn\sqrt{|S||T|}.$$

For $|S|, |T| \geq \eta|V| = \eta n$, we obtain $\eta n \leq \sqrt{|S||T|}$, $\vec{e}(S,T) \leq 2p|S||T|$ and $d_p(S,T) = \frac{2\vec{e}(S,T)}{p|S||T|} \leq 4$. This verifies that $\vec{G}$ is (η, D, p)-bounded for $D = 4$. Consequently, Lemma 29 states that $\vec{G}$ admits an (ϵ, k, p)-regular partition $\mathcal{P} = \{V_0, V_1, \ldots, V_k\}$ where $k \leq K$.

Pick any (ϵ, p)-regular pair (V_i, V_j), $1 \leq i < j$. We have $|V_i| = |V_j| \geq \frac{1-\epsilon}{K}n \geq \frac{n}{2K}$. By Proposition 20 and the condition $\lambda \leq \gamma d \leq \frac{d}{4K}$, the number of undirected edges between V_i and V_j satisfies

$$\left| e(V_i,V_j) - \frac{d}{n}|V_i||V_j| \right| \leq \lambda\sqrt{|V_i||V_j|} \leq \frac{d}{4K}\sqrt{|V_i||V_j|} \leq \frac{d}{2n}|V_i||V_j|.$$

This implies that the number of undirected edges between V_i and V_j is $e(V_i,V_j) \geq \frac{d}{2n}|V_i||V_j|$. In at least one direction, assume from V_i to V_j, we get $\vec{e}(V_i,V_j) \geq \frac{d}{4n}|V_i||V_j|$. Then set $A = V_i$, $B = V_j$ and $\beta = \frac{\vec{e}(A,B)}{|A||B|}$. By the above we have $\beta \geq \frac{d}{4n}$. We claim that the conclusion of the lemma holds with these parameters.

We know that (A,B) is an (ϵ,p)-regular pair. This means that for any $A' \subseteq A$, $B' \subseteq B$ with $|A'| \geq \epsilon|A|, |B'| \geq \epsilon|B|$, we have

$$|d_p(A',B') - d_p(A,B)| < \epsilon.$$

Here, $d_p(A',B') = \frac{2\vec{e}(A',B')}{p|A'||B'|}$. So we can rewrite this bound as

$$\left| \frac{2\vec{e}(A',B')}{p|A'||B'|} - \frac{2\vec{e}(A,B)}{p|A||B|} \right| = \frac{2}{p}\left| \frac{\vec{e}(A',B')}{|A'||B'|} - \beta \right| < \epsilon.$$

Recalling that $p = \frac{d}{n} \leq 4\beta$,

$$\left| \vec{e}(A',B') - \beta|A'||B'| \right| < \frac{p}{2}\epsilon|A'||B'| \leq 2\beta\epsilon|A||B|.$$

We still have to handle the case where $|A'| < \epsilon|A|$ or $|B'| < \epsilon|B|$. Then applying Proposition 20 again (ignoring the regularity property of (A,B)), we obtain

$$\left| e(A',B') - \frac{d}{n}|A'||B'| \right| \leq \gamma d\sqrt{|A'||B'|}$$

and since $\gamma \leq \frac{\epsilon}{4K}$, $|A'||B'| \leq \epsilon|A||B|$ and $|A| = |B| \geq \frac{n}{2K}$, we get

$$\vec{e}(A',B') \leq \frac{d}{n}|A'||B'| + \gamma d\sqrt{|A'||B'|} \leq \frac{\epsilon d}{n}|A||B| + \frac{\epsilon d}{4K}\sqrt{|A||B|} \leq \frac{2\epsilon d}{n}|A||B|.$$

We have $\beta \geq \frac{d}{4n}$, thus we conclude that $\vec{e}(A',B') \leq 8\epsilon\beta|A||B|$. ◀

4.3 The Second Moment Argument

Once we have identified the regular pair (A, B), our goal is to prove that the realizability of ρ on this regular pair implies that ρ must be close to a symmetric distribution. We adapt an approach of Trotter and Winkler [21] for the case of a transitive tournament $\{(i, j) : i, j \in [n], i < j\}$. Roughly speaking, their argument is that if (X_i, X_j) has the same distribution for each $i < j$, then any particular value in $\mathcal{D}$ has a similar number of occurrences among $\{X_1, X_2, \ldots, X_{n/2}\}$ and in $\{X_{n/2+1}, \ldots, X_n\}$. Then, counting the pairs across the two blocks, the number of (a, b) pairs is similar to the number of (b, a) pairs for any $a, b \in \mathcal{D}$, which means that the distribution is close to symmetric. Technically, the proof involves a second moment computation and the Cauchy-Schwarz inequality. We show here that a similar argument still goes through in the setting of an (arbitrary) orientation of a pseudorandom graph.

The first part of the proof does not depend on the regularity of the directed pair (A, B). It uses only properties of the undirected pseudorandom graph G.

▶ **Lemma 31.** *Let $\xi \in (0, \frac{1}{2})$ and let ρ be a $\vec{G}$-realizable distribution for some orientation $\vec{G}$ of an (n, d, λ)-graph G (non-bipartite, without isolated vertices). Let $\{X_v : v \in V\}$ the random variables realizing ρ on $\vec{G}$. For every $a \in \mathcal{D}$ and $S \subseteq V$, define a random set*

$$S_a = \{v \in S : X_v = a\}.$$

Let $A, B \subseteq V$ be two disjoint sets such that $|A| = |B| \geq \frac{\lambda n}{\xi^2 d}$. Then for every $a, b \in \mathcal{D}$,

$$\Big| \mathbb{E}[|A_a||B_b| - |A_b||B_a|] \Big| \leq 4\xi\sqrt{\mu_a \mu_b}|A|^2$$

where $\mu_a = \Pr_{(X,Y)\sim\rho}[X = a]$.

Proof. We write the target quantity as follows:

$$\mathbb{E}[|A_a||B_b| - |A_b||B_a|] = \mathbb{E}[|A_a|(|B_b| - |A_b|) + (|A_a| - |B_a|)|A_b|].$$

By Cauchy-Schwarz,

$$\Big| \mathbb{E}[|A_a||B_b| - |A_b||B_a|] \Big| \leq \sqrt{\mathbb{E}[|A_a|^2]}\sqrt{\mathbb{E}[(|B_b| - |A_b|)^2]} + \sqrt{\mathbb{E}[|A_b|^2]}\sqrt{\mathbb{E}[(|A_a| - |B_a|)^2]}. \quad (2)$$

We estimate the second moments of $|A_a|$, $|A_b|$, etc. by relating $|A_a|^2$ to the number of edges inside A labeled (a, a), etc. This is possible, since G is pseudorandom and hence its edges are in some sense a good representation of all possible pairs.

Recall that A_a denotes the random subset of A whose random variables attain value a. Each vertex in G in contained in some directed edge; hence $\forall v \in V; \Pr[X_v = a] = \mu_a$ and $\mathbb{E}[|A_a|] = \mu_a|A|$ (see Lemma 24). Further, let us consider

$$e(S_a, T_a) = |\{u \in S, v \in T : \{u, v\} \in E(G), X_u = X_v = a\}|.$$

Since the distribution on each directed edge of $\vec{G}$ is ρ, the probability that $X_u = X_v = a$ for $\{u, v\} \in E(G)$ is ρ_{aa}. Note that the orientation does not matter here, because we are looking at the event of each endpoint having the same value. We obtain

$$\mathbb{E}[e(S_a, T_a)] = \rho_{aa} e(S, T).$$

On the other hand, $e(S_a, T_a)$ is equal to the number of edges between S_a and T_a. Since G is an (n, d, λ)-graph, Proposition 20 gives

$$\left| e(S_a, T_a) - \frac{d}{n}|S_a||T_a| \right| \leq \lambda\sqrt{|S_a||T_a|}.$$

We use this bound for the following choices of S and T:

- $e(A_a, A_a) \geq \frac{d}{n}|A_a|^2 - \lambda|A_a|$
- $e(B_a, B_a) \geq \frac{d}{n}|B_a|^2 - \lambda|B_a|$
- $e(A_a, B_a) \leq \frac{d}{n}|A_a||B_a| + \lambda\sqrt{|A_a||B_a|} \leq \frac{d}{n}|A_a||B_a| + \frac{1}{2}\lambda(|A_a| + |B_a|)$

using the arithmetic-geometric inequality in the last bullet. From here, we get

$$\begin{aligned} \frac{d}{n}\,\mathbb{E}[|A_a|^2] &\leq \mathbb{E}[e(A_a, A_a)] + \lambda|A_a|] \\ &= \rho_{aa}e(A, A) + \lambda\mu_a|A|. \end{aligned}$$

Using Proposition 20 again, we have $e(A, A) \leq \frac{d}{n}|A|^2 + \lambda|A|$. Therefore,

$$\mathbb{E}[|A_a|^2] \leq \rho_{aa}|A|^2 + \frac{n\lambda}{d}(\rho_{aa} + \mu_a)|A|. \tag{3}$$

A similar bound holds for $\mathbb{E}[|B_a|^2]$. Next, we estimate

$$\begin{aligned} \frac{d}{n}\,\mathbb{E}[|A_a||B_a|] &\geq \mathbb{E}[e(A_a, B_a) - \frac{1}{2}\lambda(|A_a| + |B_a|)] \\ &= \rho_{aa}e(A, B) - \frac{1}{2}\lambda\mu_a(|A| + |B|) \\ &\geq \rho_{aa}(\frac{d}{n}|A||B| - \lambda\sqrt{|A||B|}) - \frac{1}{2}\lambda\mu_a(|A| + |B|) \\ &\geq \rho_{aa}\frac{d}{n}|A||B| - \frac{1}{2}\lambda(\rho_{aa} + \mu_a)(|A| + |B|). \end{aligned}$$

From here,

$$\begin{aligned} \mathbb{E}[(|A_a| - |B_a|)^2] &= \mathbb{E}[|A_a|^2 - 2|A_a||B_a| + |B_a|^2] \\ &\leq \rho_{aa}|A|^2 - 2\rho_{aa}|A||B| + \rho_{aa}|B|^2 + \frac{2n\lambda}{d}(\rho_{aa} + \mu_a)(|A| + |B|). \end{aligned}$$

Since $|A| = |B|$, this simplifies to

$$\mathbb{E}[(|A_a| - |B_a|)^2] \leq \frac{4\lambda n}{d}(\rho_{aa} + \mu_a)|A|. \tag{4}$$

An analogous bound holds for $\mathbb{E}[(|A_b| - |B_b|)^2]$. Combining equations (2), (3) and (4), we conclude that

$$\begin{aligned} \Big|\mathbb{E}[|A_a||B_b| - |A_b||B_a|]\Big| &\leq \sqrt{\mathbb{E}[|A_a|^2]}\sqrt{\mathbb{E}[(|B_b| - |A_b|)^2]} + \sqrt{\mathbb{E}[|A_b|^2]}\sqrt{\mathbb{E}[(|A_a| - |B_a|)^2]} \\ &\leq \sqrt{\rho_{aa}|A|^2 + \frac{\lambda n}{d}(\rho_{aa} + \mu_a)|A|}\sqrt{\frac{4\lambda n}{d}(\rho_{bb} + \mu_b)|A|} \\ &\quad + \sqrt{\rho_{bb}|A|^2 + \frac{\lambda n}{d}(\rho_{bb} + \mu_b)|A|}\sqrt{\frac{4\lambda n}{d}(\rho_{aa} + \mu_a)|A|}. \end{aligned}$$

We assumed $|A| \geq \frac{\lambda n}{\xi^2 d}$, and also we have $\rho_{aa} \leq \mu_a$, $\rho_{bb} \leq \mu_b$, $\xi \leq \frac{1}{2}$, so we can simplify this bound:

$$\begin{aligned} \Big|\mathbb{E}[|A_a||B_b| - |A_b||B_a|]\Big| &\leq 4\sqrt{\mu_a\mu_b}\sqrt{|A|^2 + 2\xi^2|A|^2}\sqrt{2\xi^2|A|^2} \\ &\leq 4\xi\sqrt{\mu_a\mu_b}|A|^2. \end{aligned}$$

◀

Now we employ the properties of the regular pair (A, B).

▶ **Lemma 32.** *Let $\epsilon \in (0, \frac{1}{2})$, let ρ be a $\vec{G}$-realizable distribution for an orientation of an (n, d, λ)-graph G (non-bipartite, without isolated vertices) and let $A, B \subseteq V(G)$ be disjoint sets such that $\beta = \frac{\vec{e}(A,B)}{|A||B|} \geq \frac{d}{4n}$, $|A| = |B| \geq \frac{\lambda n}{\epsilon^2 d}$, and*

$$\forall A' \subseteq A, B' \subseteq B; \left|\vec{e}(A', B') - \beta|A'||B'|\right| < 8\epsilon\beta|A|^2.$$

Then

$$|\rho_{ab} - \rho_{ba}| < 20\epsilon.$$

Proof. Using the notation as above, let us define $S_a = \{v \in S : X_v = a\}$ and let us estimate $\vec{e}(A_a, B_b)$, the number of directed edges from A to B labeled (a, b). On the one hand, each directed edge gets labeled (a, b) with probability ρ_{ab}, so $\mathbb{E}[\vec{e}(A_a, B_b)] = \rho_{ab}\vec{e}(A, B)$. Similarly, $\mathbb{E}[\vec{e}(A_b, B_a)] = \rho_{ba}\vec{e}(A, B)$. On the other hand, by assumption we have

$$\left|\vec{e}(A_a, B_b) - \beta|A_a||B_b|\right| < 8\epsilon\beta|A|^2 \qquad \text{and} \qquad \left|\vec{e}(A_b, B_a) - \beta|A_b||B_a|\right| < 8\epsilon\beta|A|^2$$

for every particular choice of $A_a \subseteq A, A_b \subseteq A, B_a \subseteq B, B_b \subseteq B$. From here,

$$|\rho_{ab} - \rho_{ba}|\vec{e}(A, B) = \left|\mathbb{E}[\vec{e}(A_a, B_b) - \vec{e}(A_b, B_a)]\right| \leq \beta\left|\mathbb{E}[|A_a||B_b| - |A_b||B_a|]\right| + 16\epsilon\beta|A|^2$$

Since $|A| = |B| \geq \frac{\lambda n}{\epsilon^2 d}$, we can apply Lemma 31 with $\xi = \epsilon$, and we get

$$\left|\mathbb{E}[\big||A_a||B_b| - |A_b||B_a|]\right| \leq 4\epsilon\sqrt{\mu_a\mu_b}|A|^2.$$

Combining these two bounds, we get

$$|\rho_{ab} - \rho_{ba}|\vec{e}(A, B) \leq 4\beta\epsilon\sqrt{\mu_a\mu_b}|A|^2 + 16\beta\epsilon|A|^2 \leq 20\beta\epsilon|A|^2.$$

We also have $\vec{e}(A, B) = \beta|A||B| = \beta|A|^2$, so we conclude that $|\rho_{ab} - \rho_{ba}| \leq 20\epsilon$. ◀

Now we can finish the proof of Theorem 25.

Proof of Theorem 25. We assume that ρ is $\vec{G}_n$-realizable for arbitrarily large n, where $\vec{G}_n$ is an orientation of an $(n, d(n), \lambda(n))$-graph and

$$\lim_{n\to\infty} \frac{\lambda(n)}{d(n)} = 0.$$

Clearly, such graphs cannot have isolated vertices and cannot be bipartite. We prove that $|\rho_{ab} - \rho_{ba}| \leq 20\epsilon$ for every $\epsilon > 0$, which implies that ρ is symmetric.

For a given $\epsilon > 0$, let $K \geq 1$, $\gamma > 0$ be the constants given by Lemma 30. Then we choose n large enough so that $\frac{\lambda(n)}{d(n)} < \min\{\gamma, \frac{\epsilon^2}{2K}\}$, and ρ is $\vec{G}_n$-realizable on some orientation of an $(n, d(n), \lambda(n))$-graph. Lemma 30 gives a pair of disjoint sets $A, B \subseteq V$ such that $\beta = \frac{\vec{e}(A,B)}{|A||B|} \geq \frac{d}{4n}$ and

$$\forall A' \subseteq A, B' \subseteq B; \left|\vec{e}(A', B') - \beta|A'||B'|\right| < 8\epsilon\beta|A|^2.$$

Moreover, the parameters are chosen so that $|A| = |B| \geq \frac{n}{2K} \geq \frac{\lambda(n)n}{\epsilon^2 d(n)}$. Therefore, Lemma 32 applies and we conclude that $|\rho_{ab} - \rho_{ba}| \leq 20\epsilon$.

Since this holds for every $\epsilon > 0$, ρ is in fact a symmetric distribution ($\rho_{ab} = \rho_{ba}$) and we obtain that ρ is G_n-realizable for each G_n as an undirected graph. Therefore, Theorem 22 implies that ρ is an i.i.d. mix. ◀

References

1 Boaz Barak, Fernando G. S. L. Brandão, Aram W. Harrow, Jonathan A. Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In Howard J. Karloff and Toniann Pitassi, editors, *ACM STOC*, pages 307–326. ACM, 2012.

2 Boaz Barak, Prasad Raghavendra, and David Steurer. Rounding semidefinite programming hierarchies via global correlation. In Rafail Ostrovsky, editor, *FOCS*, pages 472–481. IEEE, 2011.

3 Bruno de Finetti. *Funzione Caratteristica di un Fenomeno Aleatorio*, pages 251–299. 6. Memorie. Academia Nazionale del Linceo, 1931.

4 Bruno de Finetti. La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institute Henry Poincaré*, 7(1):1–68, 1937.

5 Persi Diaconis. Finite forms of de Finetti's theorem on exchangeability. *Synthese*, 36(2):271–281, 1977.

6 Persi Diaconis and David Freedman. Finite exchangeable sequences. *The Annals of Probability*, pages 745–764, 1980.

7 Jair Donadelli and Yoshiharu Kohayakawa. A density result for random sparse oriented graphs and its relation to a conjecture of Woodall. *Electr. J. Comb.*, 9(1), 2002.

8 Olav Kallenberg. *Probabilistic symmetries and invariance principles*, volume 9. Springer, 2005.

9 Tali Kaufman and Alexander Lubotzky. Edge transitive Ramanujan graphs and symmetric LDPC good codes. In *Proceedings of the Forty-fourth Annual ACM Symposium on Theory of Computing*, pages 359–366, 2012.

10 Tali Kaufman and Alexander Lubotzky, 2014. personal communication.

11 Tali Kaufman and Avi Wigderson. Symmetric LDPC codes and local testing. In Andrew Chi-Chih Yao, editor, *ICS*, pages 406–421. Tsinghua University Press, 2010.

12 John F.C. Kingman. Uses of exchangeability. *The Annals of Probability*, pages 183–197, 1978.

13 Michael Krivelevich and Benny Sudakov. Pseudo-random graphs. *Bolyai Society Mathematical Studies*, 15:199–262, 2006.

14 Alexander Lubotzky and Roy Meshulam. A Moore bound for simplicial complexes. *Bull. London Math. Soc.*, 39:353–358, 2007.

15 John E. Maxfield and Henryk Minc. On the matrix equation X'X = A. *Proceedings of the Edinburgh Mathematical Society (Series 2)*, 13:125–129, 12 1962.

16 T.S. Motzkin and E.G. Straus. Maxima for graphs and a new proof of a theorem of Turán. *Canada J. Math.*, 17:533–540, 1965.

17 Jaroslav Nešetřil and Patrice Ossona de Mendez. *Sparsity - Graphs, Structures, and Algorithms.* Algorithms and Combinatorics 28. Springer, 2012.

18 Ankit Sharma and Jan Vondrák. Multiway cut, pairwise realizable distributions, and descending thresholds. In David B. Shmoys, editor, *STOC*, pages 724–733. ACM, 2014.

19 A. J. Stam. Distance between sampling with and without replacement. *Statistica Neerlandica*, 32(2):81–91, 1978.

20 Luca Trevisan. Max cut and the smallest eigenvalue. *SIAM J. Comput.*, 41(6):1769–1786, 2012.

21 William T. Trotter and Peter Winkler. Ramsey theory and sequences of random variables. *Combinatorics, Probability & Computing*, 7(2):221–238, 1998.