RESEARCH ARTICLE

Permutations with fixed pattern densities

Richard Kenyon[1] | **Daniel Král'[2,3]** | **Charles Radin[4]** | **Peter Winkler[5]**

[1]Department of Mathematics, Brown University, Providence, Rhode Island

[2]Faculty of Informatics, Masaryk University, Brno, Czech Republic

[3]Mathematics Institute, DIMAP and Department of Computer Science, University of Warwick, Coventry, UK

[4]Department of Mathematics, University of Texas, Austin, Texas

[5]Department of Mathematics, Dartmouth College, Hanover, New Hampshire

Correspondence

Richard Kenyon, Department of Mathematics, Brown University, Providence, RI 02912.
Email: rkenyon@math.brown.edu

Funding information

This research was supported by the National Science Foundation's Institute for Computational and Experimental Research in Mathematics, NSF, DMS-1208191, DMS-1509088 and DMS-0901475. Simons Foundation, 327929. European Research Council under the European Union's Seventh Framework Program, FP7/2007-2013. ERC, 259385.

Abstract

We study scaling limits of random permutations ("permutons") constrained by having fixed densities of a finite number of patterns. We show that the limit shapes are determined by maximizing entropy over permutons with those constraints. In particular, we compute (exactly or numerically) the limit shapes with fixed 12 density, with fixed 12 and 123 densities, with fixed 12 density and the sum of 123 and 213 densities, and with fixed 123 and 321 densities. In the last case we explore a particular phase transition. To obtain our results, we also provide a description of permutons using a dynamic construction.

KEYWORDS

graphons, permutation patterns, permutons

1 | INTRODUCTION

We study pattern densities in permutations. A *pattern* $\tau \in S_k$ in a permutation $\sigma \in S_n$ (with $k \leq n$) is a k-element subset of indices $1 \leq i_1 < \cdots < i_k \leq n$ whose image under σ has the same order as that under τ. For example the first three indices in the permutation 4312 have pattern 321. The *density* of $\tau \in S_k$ in $\sigma \in S_n$ is $\binom{n}{k}^{-1}$ times the number of such subsets of indices.

Pattern *avoidance* in permutations is a well-studied and rich area of combinatorics; see [21] for the history and the current state of the subject. Less studied is the problem of determining the range of possible densities of patterns, and the "typical shape" of permutations with constrained densities of (a fixed set of) patterns. We undertake such a study here. Specifically, we consider the densities of one or more patterns and consider the *feasible region*, or phase space $\mathbb{F}$, of possible values of densities

of the chosen patterns for permutations in S_n in the limit of large n. For densities in the interior of $\mathbb{F}$ we study the shape of a typical permutation with those densities, again in the large n limit. We note that the typical shape of pattern-avoiding permutations (which necessarily lie on the boundary of the feasible region $\mathbb{F}$) has also recently been investigated [1, 10, 15, 26–28].

To deal with these asymptotic questions we show that the size of our target sets of constrained permutations can be estimated by maximizing a certain function over limit objects called permutons. Furthermore when—as appears to be usually the case—the maximizing permuton is unique, properties of most permutations in the class can then be deduced from it. After setting up our general framework we work out several examples. To give further details we need some notation.

To a permutation $\pi \in S_n$ one can associate a probability measure γ_π on $[0, 1]^2$ as follows. Divide $[0, 1]^2$ into an $n \times n$ grid of squares of size $1/n \times 1/n$. Define the density of γ_π on the square in the ith row and jth column to be the constant n if $\pi(i) = j$ and 0 otherwise. In other words, γ_π is a geometric representation of the permutation matrix of π.

Define a *permuton* to be a probability measure γ on $[0, 1]^2$ with uniform marginals:

$$\gamma([a, b] \times [0, 1]) = b - a = \gamma([0, 1] \times [a, b]), \quad \text{for all } 0 \le a \le b \le 1. \tag{1}$$

Note that γ_π is a permuton for any permutation $\pi \in S_n$. Permutons were introduced in [16, 17] with a different but equivalent definition; the measure theoretic view of large permutations can be traced to [30] and was used in [13, 24] as an analytic representation of permutation limits equivalent to that used in [16, 17]; the term "permuton" first appeared, we believe, in [13].

Let Γ be the space of permutons. There is a natural topology on Γ, the weak topology on probability measures, which can equivalently be defined as the metric topology defined by the metric $d_\square$ given by $d_\square(\gamma_1, \gamma_2) = \max |\gamma_1(R) - \gamma_2(R)|$, where R ranges over aligned rectangles in $[0, 1]^2$. This topology is also the same as that given by the L^∞ metric on the cumulative distribution functions (CDFs) $G_i(x, y) = \gamma_i([0, x] \times [0, y])$. We say that a sequence of permutations π_n with $\pi_n \in S_n$ *converges* as $n \to \infty$ if the associated permutons converge in the above sense.

Extending the definition above, given a permuton γ the *pattern density of τ in γ*, denoted $\rho_\tau(\gamma)$, is by definition the probability that, when k points are selected independently from γ and their x-coordinates are ordered, the permutation induced by their y-coordinates is τ. For example, for γ with probability density $g(x, y)dx\, dy$, the density of pattern $12 \in S_2$ in γ is

$$\rho_{12}(\gamma) = 2 \int_{x_1 < x_2 \in [0,1]} \int_{y_1 < y_2 \in [0,1]} g(x_1, y_1)g(x_2, y_2)dx_1 dy_1 dx_2 dy_2. \tag{2}$$

It follows from results of [16, 17] that two permutons are equal if they have the same pattern densities (for all k).

The notion of pattern density for permutons generalizes the notion for permutations. Note however that the density of a pattern $\alpha \in S_k$ in a permutation $\tau \in S_n$ (defined to be the number of copies of α in τ, divided by $\binom{n}{k}$) will not generally be equal to the density of α in the permuton γ_τ; equality will only hold in the limit of large n.

1.1 | Results

Theorem 1 (restated from the somewhat different form in Trashorras [37]) is a large deviations theorem for permutons: it describes explicitly how many large permutations lie near a given permuton. The
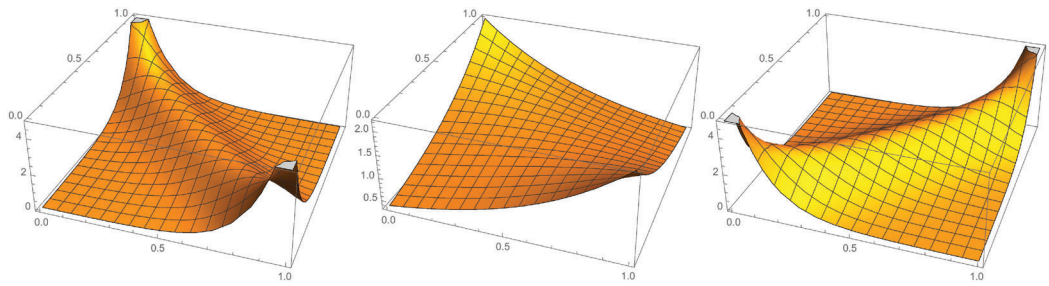
FIGURE 1 The permuton with fixed density ρ of pattern 12, shown for $\rho = 0.2, 0.4, 0.8$ [Colour figure can be viewed at wileyonlinelibrary.com]

statement is essentially that the number of permutations in S_n lying near a permuton γ is

$$n! e^{(H(\gamma)+o(1))n}, \tag{3}$$

where $H(\gamma)$ is the "permuton entropy" (defined below).

We use this large deviations theorem to prove Theorem 2, which describes both the number and (when uniqueness holds) limit shape of permutations in which a finite number of pattern densities have been fixed. The theorem is a variational principle: it shows that the number of such permutations is determined by the permuton entropy *maximized* over the set of permuton(s) having those fixed pattern densities.

Another construction we use replaces permutons by families of *insertion measures* $\{\mu_t\}_{t\in[0,1]}$, which is analogous to building a permutation by inductively inserting one element at a time into a growing list: for each $i \in [n]$ one inserts i into a random location in the permuted list of the first $i - 1$ elements. This construction is used to describe explicitly the entropy maximizing permutons with fixed densities of patterns of type $* * \cdots * i$ (here each $*$ represents an element not exceeding the length of the pattern, for example, $* * 2$ represents the union of the patterns 132 and 312). We prove that for this family of patterns the maximizing permutons are analytic, the entropy function as a function of the constraints is analytic and strictly concave, and the optimal permutons are unique and have analytic probability densities.

The most basic example to which we apply our results, the entropy-maximizing permuton for a fixed density ρ_{12} of 12 patterns, has probability density

$$g(x, y) = \frac{r(1 - e^{-r})}{(e^{r(1-x-y)/2} - e^{r(x-y-1)/2} - e^{r(y-x-1)/2} + e^{r(x+y-1)/2})^2},$$

where r is an explicit function of ρ_{12} (see Figure 1).

While maximizing permutons can be shown to satisfy certain explicit partial differential equations (PDEs) (see Section 8), they can also exhibit a very diverse set of behaviors. Even in one of the simplest cases, that of fixed density of the two patterns 12 and 123, the variety of shapes of permutons (and therefore of the approximating permutations) is remarkable (see Figure 7). In this case we prove that the feasible region of densities is the so-called "scalloped triangle" of Razborov [34, 35] which also describes the space of feasible densities for edges and triangles in the graphon model.

Another example which has been studied recently [11, 18, 19] is the case of the two patterns 123 and 321. In this case we describe a phase transition in the feasible region, where the maximizing permuton changes abruptly.

The variational principle can easily be extended to analyze other constraints that are continuous in the permuton topology. For constraints that are not continuous, for example the number of cycles of a fixed size, one can analyze an analogous "weak" characteristic, which is continuous, by applying the characteristic to patterns. For example, while the number of fixed points of a permuton is not well-defined, we can compute the expected number of fixed points for the permutation in S_n obtained by choosing n points independently from the permuton, and analyze this quantity in the large n limit. This computation will be discussed in a subsequent paper [22]; the result is that the expected weak number of fixed points is

$$\int_0^1 g(x, x)\, dx$$

when g has a continuous density. Similar expressions hold for cycles of other lengths.

1.2 | Analogies with graphons

For those who are familiar with variational principles for dense graphs [7, 9, 31, 32], we note the following differences between the graph case and the permutation case (see [25] for background on graph asymptotics):

1. Although permutons serve the same purpose for permutations that graphons serve for graphs, and (being defined on $[0, 1]^2$) are superficially similar, they are measures (not symmetric functions) and represent permutations in a different way. (One *can* associate a graphon with a limit of permutations, via comparability graphs of two-dimensional posets, but these have trivial entropy in the Chatterjee-Varadhan sense [9] and we do not consider them here.)
2. The classes of constrained (dense) graphs considered in [9] have size about e^{cn^2}, n being the number of vertices and the (nonnegative) constant c being the target of study. Classes of permutations in S_n are of course of size at most $n! \sim e^{n(\log n - 1)}$ but the constrained ones we consider here have size of order not $e^{cn \log n}$ for $c \in (0, 1)$, as one might at first expect, but instead $e^{n \log n - n + cn}$ where $c \in [-\infty, 0]$ is the quantity of interest.
3. The "entropy" function, that is, the function of the limit structure to be maximized, is bounded for graphons but unbounded for permutons. This complicates the analysis for permutations.
4. The limit structures that maximize the entropy function tend, in the graph case, to be combinatorial objects: step-graphons corresponding to what Radin, Ren and Sadun call "multipodal" graphs [33]. In contrast, maximizing permutons at interior points of feasible regions seem always to be smooth measures with analytic densities. Although they are more complicated than maximizing graphons, these limit objects are more suitable for classical variational analysis, for example, differential equations of the Euler-Lagrange type.

2 | VARIATIONAL PRINCIPLE

For convenience, we denote the unit square $[0, 1]^2$ by Q.

Let γ be a permuton with density g defined almost everywhere. We compute the *permutation entropy* $H(\gamma)$ of γ as follows:

$$H(\gamma) = \int_Q -g(x, y) \log g(x, y)\, dx\, dy, \tag{4}$$

where "$0 \log 0$" is taken as zero. Then H is finite whenever g is bounded (and sometimes when it is not). In particular for any $\sigma \in S_n$, we have $H(\gamma_\sigma) = -n(n \log n/n^2) = -\log n$ and therefore $H(\gamma_\sigma) \to -\infty$ for any sequence of increasingly large permutations even though $H(\lim \gamma_\sigma)$ may be finite. Note that H is zero on the uniform permuton (where $g(x, y) \equiv 1$) and negative (sometimes $-\infty$) on all other permutons, since the function $z \log z$ is concave downward. If γ has no density, we define $H(\gamma) = -\infty$.

We use the following large deviations principle, first stated in a somewhat different form by Trashorras (Theorem 1 in [37]); see also Theorem 4.1 in [29]. In Section 11 we give an alternative proof.

Theorem 1 ([37]). *Let Λ be a set of permutons, Λ_n the set of permutations $\pi \in S_n$ with $\gamma_\pi \in \Lambda$. Then:*

1. *If Λ is closed,*

$$\limsup_{n \to \infty} \frac{1}{n} \log \frac{|\Lambda_n|}{n!} \leq \sup_{\gamma \in \Lambda} H(\gamma); \tag{5}$$

2. *If Λ is open,*

$$\liminf_{n \to \infty} \frac{1}{n} \log \frac{|\Lambda_n|}{n!} \geq \sup_{\gamma \in \Lambda} H(\gamma). \tag{6}$$

To make a connection with our applications to large constrained permutations, fix some finite set $\mathcal{P} = \{\pi_1, \ldots, \pi_k\}$ of patterns. Let $\alpha = (\alpha_1, \ldots, \alpha_k)$ be a vector of desired pattern densities. We then define two sets of permutons:

$$\Lambda^{\alpha, \varepsilon} = \{\gamma \in \Gamma \mid |\rho_{\pi_j}(\gamma) - \alpha_j| < \varepsilon \text{ for each } 1 \leq j \leq k\} \tag{7}$$

and

$$\Lambda^\alpha = \{\gamma \in \Gamma \mid \rho_{\pi_j}(\gamma) = \alpha_j \text{ for each } 1 \leq j \leq k\}. \tag{8}$$

With that notation, and the understanding that $\Lambda_n^{\alpha, \varepsilon} = \Lambda^{\alpha, \varepsilon} \cap \gamma(S_n)$, where $\gamma(\alpha) = \gamma_\alpha$ as before, our first main result is:

Theorem 2.

$$\lim_{\varepsilon \downarrow 0} \lim_{n \to \infty} \frac{1}{n} \log \frac{|\Lambda_n^{\alpha, \varepsilon}|}{n!} = \max_{\gamma \in \Lambda^\alpha} H(\gamma).$$

The value $\max_{\gamma \in \Lambda^\alpha} H(\gamma)$ (which is guaranteed by the theorem to exist, but may be $-\infty$) will be called the *constrained entropy* and denoted by $s(\alpha)$. In Section 4 we will prove Theorem 2.

Theorem 2 puts us in a position to try to describe and enumerate permutations with some given pattern densities. It does not, of course, guarantee that there is just one $\gamma \in \Lambda^\alpha$ that maximizes $H(\gamma)$, nor that there is one with finite entropy. As we shall see it seems to be the case that interior points in feasible regions for pattern densities do have permutons with finite entropy, and *usually* just one optimizer. Points on the boundary of a feasible region (e.g., pattern-avoiding permutations) often have only singular permutons, and since the latter always have entropy $-\infty$, Theorem 2 will not be of direct use there.

3 | FEASIBLE REGIONS AND ENTROPY OPTIMIZERS

We collect here some general facts about feasible regions and entropy optimizers, making use of concavity of entropy and the "heat flow on permutons."

3.1 | Heat flow on permutons

The heat flow is a continuous flow on the space of permutons with the property that for any permuton $\mu = \mu_0$ and any positive time $t > 0$, μ_t has analytic density (and thus finite entropy).

The flow after time t is given by the action of the heat operator $e^{t\Delta}$ where Δ is the Laplacian on the square with reflecting boundary conditions. One can describe the flow concretely as follows. First, one can describe a permuton μ by its characteristic function $\hat{g}(u, v) = \mathbb{E}[e^{i(ux+vy)}]$. In fact since we are on the unit square we can use instead the discrete Fourier cosine series

$$\hat{g}(j, k) = \mathbb{E}[\cos(\pi j x) \cos(\pi k y)]$$

with $j, k \geq 0$. The uniform marginals condition is that $\hat{g}(j, k) = 0$ if $j = 0$ or $k = 0$ except for $\hat{g}(0, 0) = 1$.

The operator $e^{t\Delta}$ acts on the coefficients by multiplication by $e^{-\pi^2(j^2+k^2)t}$:

$$\hat{g}_t(j, k) = \hat{g}_0(j, k) e^{-\pi^2(j^2+k^2)t}.$$

Note that the heat flow preserves the marginals.

For any $t > 0$ the Fourier coefficients $\hat{g}_t(j, k)$ then decay exponentially quickly so that $\hat{g}_t(j, k)$ are the Fourier coefficients of a measure with analytic density.

3.2 | Elementary consequences for feasible regions

Let R be the feasible region for permutons with some finite set of n pattern densities τ. Let R_M be the subset of R consisting of points representable by a permuton with entropy at least $-M$, and R_* those representable by permutons with finite entropy.

Entropy is upper-semicontinuous on R (just as it is on the space Γ of all permutons, see Lemma 19). So R_M is closed.

Lemma 3. *R_* is dense in R.*

Proof. Any permuton γ may be perturbed (by the heat flow for small time, thus moving densities only a small amount) to achieve an analytic permuton, which will have finite entropy. ∎

Theorem 4. *Let C be a topological sphere in R_M, that is, a set homeomorphic to the unit sphere S^{n-1} in $\mathbb{R}^n$, where $n = \dim(R)$ is the number of patterns. If there is a continuous family of permutons in R_M parameterized by C then the interior of C is also contained in R_M.*

Proof. The hypothesis says that there is a continuous map $\phi : S^{n-1} \to \Gamma$ such that $\tau \circ \phi$ maps S^{n-1} to C. Since Γ is convex, ϕ extends to a map $\tilde{\phi}$ from the unit ball B_n to Γ. The composition $\tau \circ \tilde{\phi}$ provides a homotopy of C to a point and thus contains the interior of C. Since the entropy function is concave, the entropies of the points in the space of convex combinations are all at least $-M$. ∎

This statement applied to the case of single density yields

Corollary 5. *If R is the feasible region for a single density, then it is an interval on the interior of which entropy is finite and concave.*

Corollary 6. *Let C be a topological sphere in R_M. Suppose the maximizing permutons at points of C vary continuously on C. C contains in its interior no local minimum of the entropy.*

Proof. Concavity of entropy implies that the minimum of the entropy on the image of B_n occurs on the image of S^{n-1}. ∎

Theorem 7. *The maximizing permutons associated to two local maxima in R, at least one of which is a strict local maxima, are not connected by a continuous path of permutons.*

Proof. Connect the permutons representing the two local maxima by the path of their convex combinations. Concavity of entropy along this path implies the result. ∎

Note that for any pattern π there is a permuton which has zero density for that pattern (either the identity permuton or the "anti-identity" permuton). The *maximal π-density permuton(s)* are not known in general, although a lower bound on the maximal density is obtained from the permuton γ_π.

4 | PROOF OF THEOREM 2

Since we will be approximating H by certain Riemann sums, it is useful to define, for any permuton γ and any positive integer m, an approximating "step-permuton" γ^m as follows. Fixing m, denote by Q_{ij} the half-open square $((i-1)/m, i/m] \times ((j-1)/m, j/m]$; for each $1 \leq i, j \leq m$, we want γ^m to be uniform on Q_{ij} with $\gamma^m(Q_{ij}) = \gamma(Q_{ij})$. In terms of the density g^m of γ^m, we have $g^m(x, y) = m^2 \gamma(Q_{ij})$ for all $(x, y) \in Q_{ij}$.

To prove Theorem 2 we use the following result.

Proposition 8. *For any permuton γ, $\lim_{m \to \infty} H(\gamma^m) = H(\gamma)$, with $H(\gamma^m)$ diverging downward when $H(\gamma) = -\infty$.*

In what follows we will, in order to increase readability, write

$$\int_0^1 \int_0^1 - g(x, y) \log g(x, y) dx dy \tag{9}$$

as just $\int_Q -g \log g$. Also for the sake of readability, we will for this section only state results in terms of $g \log g$ rather than $-g \log g$; this avoids clutter caused by a multitude of absolute values and negations. Eventually, however, we will need to deal with an entropy function $H(\gamma) = \int_Q -g \log g$ that takes values in $[-\infty, 0]$.

Define

$$g_{ij} = m^2 \gamma(Q_{ij}). \tag{10}$$

We wish to show that the Riemann sum

$$\frac{1}{m^2} \sum_{0 \leq i, j \leq m} g_{ij} \log g_{ij}, \tag{11}$$

which we denote by $R_m(\gamma)$, approaches $\int_Q g \log g$ when γ is absolutely continuous with respect to Lebesgue measure, i.e., when the density g exists a.e., and otherwise diverges to ∞. We restate and prove this in slightly greater generality, not requiring that γ have uniform marginals or that it be normalized to have $\gamma(Q) = 1$. Note that that permutons arising later in this work often have unbounded density functions, thus will not be Riemann integrable; we need to show that nevertheless our particular Riemann sums $R_m(\gamma)$ do approximate the entropy integral.

Proposition 9. *Let γ be a finite measure on $Q = [0,1]^2$ and $R_m = R_m(\gamma)$. Then:*

1. *If γ is absolutely continuous with density g, and $g \log g$ is integrable, then $\lim_{m \to \infty} R_m = \int_Q g \log g$.*
2. *If γ is absolutely continuous with density g, and $g \log g$ is not integrable, then $\lim_{m \to \infty} R_m = \infty$.*
3. *If γ has a singular component, then $\lim_{m \to \infty} R_m = \infty$.*

Proof. We begin with the first case, where we need to show that for any $\varepsilon > 0$, there is an m_0 such that for $m \geq m_0$,

$$\int_Q g \log g - \frac{1}{m^2} \sum_{i,j=0}^{m} g_{ij} \log g_{ij} < \varepsilon . \tag{12}$$

Note that the function $u \log u$ is bounded below by $-1/e$ and since it is convex, the quantity on the left cannot be negative.

We need to truncate g. Let $A(t) = \{(x,y) \in Q : g(x,y) \geq t\}$ and choose $M > 1$ so that $\int_{A(M)} g \log g < \varepsilon/2$. Let $f = 0$ on $A(M)$ and $f = g$ otherwise, so that $\int_Q g \log g - \varepsilon/2 < \int_Q f \log f \leq \int_Q g \log g$. Then $\frac{1}{m^2} \sum_{i,j=0}^{m} f_{ij} \log f_{ij}$, which we will call R'_m, is at most R_m, so it suffices to show $\int f \log f - R'_m < \varepsilon/2$.

We next take a closer look at the function $u \log u$. For $0 \leq a \leq b \leq M$, define

$$d(a,b) = \max_{p \in [0,1]} \left(pa \log a + (1-p)b \log b - (pa + (1-p)b) \log(pa + (1-p)b) \right) \tag{13}$$

for $0 \leq a \leq b$. For fixed b, $d(a,b)$ increases as a decreases and reaches

$$d(0,b) = \max_{p \in [0,1]} ((1-p)b \log b - (1-p)b \log(1-p)b)$$
$$= -b \max_{p \in [0,1]} (1-p) \log(1-p) = b/e. \tag{14}$$

For fixed $a > 0$, the second derivative of $u \log u$—namely, $1/u$—is bounded above by $1/a$ on $[a,b]$, and we can apply the following calculus exercise:

Lemma 10. *Let w be a C^2 function on the real line with open domain containing the interval $[a,b]$, and suppose the second derivative w'' of w satisfies $0 \leq w''(x) \leq h$ for $x \in [a,b]$. Let*

$$v(x) = w(a) + \frac{x-a}{b-a} \left(w(b) - w(a) \right) - w(x) \tag{15}$$

be the vertical distance d between the graph of w and its $[a,b]$ secant. Then

$$\max_{x \in [a,b]} v(x) \leq h(b-a)^2. \tag{16}$$

Proof. Differentiating v at $x \in [a, b]$, we have

$$|v'(x)| = |\frac{w(b) - w(a)}{b - a} - w'(x)| = |w'(c) - w'(x)$$

$$= w''(s)|c - x| \leq h|c - x| \leq h(b - a), \qquad (17)$$

for some $c \in [a, b]$ and $s \in [a, b]$, via two applications of the Mean Value Theorem. But then, since $v(a) = 0$, we have $v(x) \leq h(b - a)(x - a) \leq h(b - a)^2$ as claimed. ∎

Applying Lemma 10 to $w(u) = u \log u$ with $x = pa + (1 - p)b$, we conclude that $d(a, b) \leq (b - a)^2 / a$ for $a > 0$.

To state the next lemma it will be convenient for us to define a function η on $[0,1]$ by $\eta(p) = -p \log p$ for $p \leq 1/e$ (including $\eta(0) = 0$) and $\eta(p) = 1/e$ for $1/e \leq p \leq 1$.

Lemma 11. *Let u be a measurable function on $Q = [0, 1]^2$ with $0 \leq u \leq M$ everywhere and $a \leq u \leq b$ on a set R of measure $1 - \delta$. If $c = \int_Q u$ and $D = \int_Q u \log u - c \log c$, then $D < d(a, b) + M\eta(\delta)$.*

Proof. Fix M, δ, a, and b, and suppose that the function u satisfies the conditions of the lemma's statement. If, in addition, $a \leq u \leq b$ everywhere in Q, then D is maximized (by convexity of $u \log u$) when u takes only the two values a and b; thus, in that case, D is bounded by $d(a, b)$.

On the other hand, let a and b be not fixed and suppose that u is constant on R, say with value $v = a = b$. Then, again by convexity of $u \log u$, to maximize D, v must be one of the extreme values 0 or M; and outside R, u can take only values 0 and M. Putting it another way, for M and δ fixed and u constant on R, D is maximized in one of the following two ways: (1) by taking $a = 0$ and setting $u = 0$ on a set of measure $1 - p \geq 1 - \delta$, and $u = M$ otherwise; or (2) by taking $a = M$ and setting $u = M$ on a set of measure $1 - p \geq 1 - \delta$, and $u = 0$ otherwise.

In Case (1),

$$D \leq pM \log M - pM \log pM = -Mp \log p. \qquad (18)$$

Since $-p \log p \leq -\delta \log \delta$ for $\delta \leq 1/e$, and since $-p \log p \leq 1/e$ always, we conclude that $D \leq M\eta(\delta)$.

In Case (2),

$$D \leq (1 - p)M \log M - (1 - p)M \log(1 - p)M = -M(1 - p) \log(1 - p). \qquad (19)$$

Since $-(1 - p) \log(1 - p) \leq -p \log p \leq \delta \log \delta$ for $\delta \leq 1/e$ and $-(1 - p) \log(1 - p) \leq 1/e$ always, we again conclude that $D \leq M\eta(\delta)$.

We now need to show that the sum $d(a, b) + M\eta(\delta)$ bounds D in general. Returning to the lemma's hypotheses concerning u, D is maximized when u takes on only the values a and b on the set R of measure $1 - \delta$. Let $c = \int_Q u$ and put $r = (|A|a + |B|b)/(|A| + |B|)$, where A is the subset of R on which $u = a$ and $B \subset R$ is the set on which $u = b$.

Define a new function v by $v = r$ on R and $v = u$ on $Q \setminus R$. Then

$$D = \int_R u \log u + \int_{Q \setminus R} u \log u - c \log c$$

$$= \left(\int_R u \log u - |R|r \log r \right) + |R|r \log r + \int_{Q \setminus R} v \log v - c \log c$$

$$\leq |R|d(a, b) + |R|r \log r + \int_{Q \setminus R} v \log v - c \log c$$

$$
\begin{aligned}
&= |R|d(a,b) + \int_Q v \log v - c \log c \\
&\leq |R|d(a,b) + M\eta(\delta),
\end{aligned}
\tag{20}
$$

where the last inequality follows from the cases (1) and (2) above applied to v. Finally since $|R| \leq 1$, $D \leq d(a,b) + M\eta(\delta)$ as claimed. ∎

Next we slice up f horizontally, choosing $0 = a_0 < a_1 < \cdots < a_k = M$ such that $\sum_{j=1}^k d(a_{j-1}, a_j) < \varepsilon/4$. This is easily done by choosing $a_1 < \varepsilon/8$ to get $d(0, a_1) < \varepsilon/5$, and taking advantage thereafter of $d(a, b)$ being quadratic in $b-a$, in particular, bounded by $(b-a)^2/a_1$ for $a \geq a_1$.

The sets $S_k = A(a_k) \setminus A(a_{k-1})$ must now be approximated by unions of grid cells. Any measurable $S \subset Q$ has Lebesgue measure equal to its outer measure $\inf\{\sum_{j=1}^\infty |C_j| \; : \; S \subseteq \cup C_j\}$ where each C_j is cell of the $m \times m$ grid on Q for some m. By picking m large enough, we can cover all but δ of the measure of S by a subset of cells of the $m \times m$ grid of total measure $|S| - \delta$. That is, we can find for any $\delta > 0$ an m_0 such that $m > m_0$ implies the existence of a set C of cells of the $m \times m$ grid such that the union U of the cells of C satisfies $|S \Delta U| < \delta$, where Δ is our symbol for symmetric difference. Applying this successively to our sets $S_1, S_2, \ldots, S_k$ yields m_k such that for $m > m_k$, there is a partition $C_1, \ldots, C_k$ of the cells of the $m \times m$ grid such that $|U_j \setminus S_j|/|S_j| < \varepsilon_j$ for each j with $|S_j| > 0$, where U_j is the union of the cells in C_j, and each $\varepsilon_j > 0$ is arbitrary. (If $|S_j| = 0$, we take $C_j = \emptyset$.) We choose each ε_j to satisfy $\eta(\varepsilon_j) < \varepsilon/4M$.

We now consider the contribution made to $\int f \log f - R'_m$, for each j, by the cells in C_j. Order the cells in C_j arbitrarily and let $U_j(i)$ be the ith. Let $p_i = |U_j(i) \setminus S_j|/|U_j(i)|$. Replacing u in Lemma 11 by f, Q by $U_j(i)$, and R by $U_j(i) \setminus S_j$, and noting that $a_{j-1} \leq f \leq a_j$ on all of $U_j(i)$ except on $U_j(i) \setminus S_j$ where $0 \leq f \leq M$, we conclude that

$$
m^2 \int_{U_j(i)} f \log f - f_i \log f_i < d(a_{j-1}, a_j) + M\eta(p_i),
\tag{21}
$$

where $f_i = m^2 \int_{U_j(i)} f$.

Summing over i and dividing through by m^2, we get

$$
\int_{U_j} f \log f - \frac{1}{m^2} \sum_i f_i \log f_i < |U_j|d(a,b) + \frac{M}{m^2} \sum_i \eta(p_i),
\tag{22}
$$

where the average of the p_i's is $p = |U_j \setminus S_j|/|U_j| < \varepsilon_j$. Since η is (weakly) concave, $\sum_i \eta(p_i)$ is maximized by setting all $p_i = p$ and we get

$$
\int_{U_j} f \log f - \frac{1}{m^2} \sum_i f_i \log f_i < |U_j|d(a_{j-1}, a_j) + |U_j|M\eta(\varepsilon_j).
\tag{23}
$$

Summing now over j with $|S_j| > 0$, we have

$$
\begin{aligned}
\int_Q f \log f - R'_m &= \int_Q f \log f - \frac{1}{m^2} \sum_j \sum_{i \in C_j} f_i \log f_i \\
&\leq \sum_j |U_j|d(a_{j-1}, a_j) + \sum_j |U_j|M\eta(\varepsilon_j) \\
&\leq \sum_j d(a_{j-1}, a_j) + \sum_j |U_j|\varepsilon/4 \\
&\leq \varepsilon/4 + \varepsilon/4 = \varepsilon/2
\end{aligned}
\tag{24}
$$

and we are done with the integrable case.

In the second case, where g is defined but $\int g \log g$ is infinite, we truncate g at $j = 2, 3, 4, \ldots$ to get integrable functions f_j with $\int_Q f_j \log f_j \to \infty$. We then apply the result of the first case to each f_j to get Riemann sums $R(j)_{m_j}$ that grow without bound; since each is below the corresponding Riemann sum $R_{m(j)}$ for g, we have $\lim_{m \to \infty} R_m \to \infty$ as claimed.

Finally, suppose γ has a singular component and let A be a set of Lebesgue measure zero for which $\gamma(A) = a > 0$. We first prove a cover lemma.

Lemma 12. *For any $\varepsilon > 0$ there is an m_2 such that $m > m_2$ implies that there are εm^2 squares of the $m \times m$ grid that cover at least half the γ-measure of A.*

Proof. Note first that if B is an open disk in Q of radius at most δ, then for $m > 1/(2\delta)$, then we can cover B with cells of an $m \times m$ grid of total area at most $64\delta^2$. The reason is that such a disk cannot contain more than $\lceil 2\delta/(1/m) \rceil^2 < (4\delta m)^2$ grid vertices, each of which can be a corner of at most four cells that intersect the disk. Thus, rather conservatively, the total area of the cells that intersect the disk is bounded by $(4/m^2) \cdot (4\delta m)^2 = 64\delta^2$. It follows that as long as a disk has radius at least $1/(2m)$, it costs at most a factor of $64/\pi$ to cover it with grid cells.

Now cover A with open disks of area summing to at most $\pi\varepsilon/64$. Let b_n be the γ-measures of the union of the disks of radii at least $1/2n$. Choose m_2 such that $b_{m_2} > a/2$ to get the desired result. ∎

Let M be given and use Lemma 12 to find m_2 such that for any $m \geq m_2$, there is a set $I \subset \{1, \ldots, m\}^2$ of size at most δm^2 such that $\gamma(\bigcup_I Q_{ij}) > a/2$, where $Q_{ij} = ((i-1)/m, i/m] \times ((j-1)/m, j/m]$ as before and δ is a small positive quantity depending on M and a, to be specified later. Then

$$
\begin{aligned}
R_m(\gamma) &= \sum_{ij} \frac{1}{m^2} g_{ij} \log g_{ij} \\
&\geq -1/e + \frac{1}{m^2} \delta m^2 \bar{g} \log \bar{g} = -1/e + \delta \bar{g} \log \bar{g},
\end{aligned}
\tag{25}
$$

where $\bar{g}$ is the mean value of g_{ij} over $(i,j) \in I$, the last inequality following from the convexity of $u \log u$. The $-1/e$ term is needed to account for possible negative values of $g \log g$.

But $\sum_I g_{ij} = m^2 \gamma(\bigcup_I Q_{ij}) > m^2 a/2$, so $\bar{g} > (m^2 a/2)/(\delta m^2) = a/(2\delta)$. Consequently

$$
R_m(\gamma) > -\frac{1}{e} + \delta \frac{a}{2\delta} \log \frac{a}{2\delta} = -\frac{1}{e} + \frac{a}{2} \log \frac{a}{2\delta} .
\tag{26}
$$

Taking

$$
\delta = \frac{a}{2} \exp\left(-2 \frac{\left(M + \frac{1}{e} \right)}{a} \right)
\tag{27}
$$

gives $R_m(\gamma) > M$ as required, and the proof of Proposition 8 is complete. ∎

We now prove Theorem 2.

Proof. The set $\Lambda^{\alpha,\varepsilon}$ of permutons under consideration consists of those for which certain pattern densities are close to values in the vector α. Note first that since the density function $\rho(\pi, \cdot)$ is continuous in the topology of Γ, Λ^α is closed and by compactness $H(\gamma)$ takes a maximum value on Λ^α.

Again by continuity of $\rho(\pi, \cdot)$, $\Lambda^{\alpha,\varepsilon}$ is an open set and we have from the second statement of Theorem 1 that for any ε,

$$\lim_{n \to \infty} \frac{1}{n} \log \frac{|\Lambda_n^{\alpha,\varepsilon}|}{n!} \geq \max_{\gamma \in \Lambda^{\alpha,\varepsilon}} H(\gamma) \geq \max_{\gamma \in \Lambda^\alpha} H(\gamma) \tag{28}$$

from which we deduce that

$$\lim_{\varepsilon \downarrow 0} \lim_{n \to \infty} \frac{1}{n} \log \frac{|\Lambda_n^{\alpha,\varepsilon}|}{n!} \geq \max_{\gamma \in \Lambda^\alpha} H(\gamma). \tag{29}$$

To get the reverse inequality, fix a $\gamma \in \Lambda^\alpha$ maximizing $H(\gamma)$. Let $\delta > 0$; since H is upper semi-continuous and Λ^α is closed, we can find an $\varepsilon' > 0$ such that no permuton γ' within distance ε' of Λ^α has $H(\gamma') > H(\gamma) + \delta$. But again since $\rho(\pi, \cdot)$ is continuous, for small enough ε, every $\gamma' \in \Lambda^{\alpha,\varepsilon}$ is indeed within distance ε' of Λ^α. Let Λ' be the (closed) set of permutons γ' satisfying $\rho(\pi_j, \gamma') \leq \varepsilon$; then, using the first statement of Theorem 1, we have thus

$$\lim_{n \to \infty} \frac{1}{n} \log \frac{|\Lambda_n'|}{n!} \leq H(\gamma) + \delta \tag{30}$$

and since such a statement holds for arbitrary $\delta > 0$, the result follows. ∎

5 | INSERTION MEASURES

A permuton γ can be described by a family of *insertion measures*. This description will be useful for constructing concrete examples, in particular for the so-called star models, which are discussed in Sections 6 and 7.

The insertion measures are a family of probability measures $\{\nu_x\}_{x \in [0,1]}$, with measure ν_x supported on $[0, x]$. This family is a continuum version of the process of building a random permutation on $[n]$ by, for each i, inserting i at a random location in the permutation formed from $\{1, \ldots, i-1\}$. Any permutation measure can be built this way. We describe here how any permuton can be built from a family of *independent* insertion measures, and conversely, how every permuton defines a unique family of independent insertion measures. An explicit example is given in the next section.

We first describe how to reconstruct the insertion measures from the permuton γ.

Let μ_x be the conditional law $\gamma|_{\{x\} \times [0,1]}$; by the Disintegration Theorem (see e.g., [8]), μ_x is defined for almost every x. Let $Y_x \in [0, 1]$ be the random variable with law μ_x. Let $Z_x \in [0, x]$ be the random variable (with law ν_x) giving the location of the insertion at time x, and let $F(x, \cdot)$ be its CDF. Then ν_x is related to μ_x in the following way, which we can take to be the definition of ν_x:

$$F(x, y) = \Pr(Z_x < y) = \Pr(Y_x < \tilde{y}) = G_x(x, \tilde{y}), \tag{31}$$

where $\tilde{y}$ is defined by $G(x, \tilde{y}) = y$. (Here and in what follows we use subscripts x and y on G to denote partial derivatives. Note that $G_x(x, y)$ is the CDF of μ_x, which exists for almost every x.)

More succinctly, we have

$$F(x, G(x, \tilde{y})) = G_x(x, \tilde{y}). \tag{32}$$

This equation defines F uniquely as a function of G at almost every x.

Conversely, given the insertion measures, Equation (32) is a differential equation for G. Concretely, after we insert x_0 at location $Y(x_0) = Z_{x_0}$, the image flows under future insertions according to the (deterministic) evolution

$$\frac{d}{dx} Y(x) = F(x, Y(x)), \quad Y(x_0) = Z_{x_0}. \tag{33}$$

If we let $\Psi_{[x,1]}$ denote the flow up until time 1, then the permuton is the push-forward under Ψ of ν_x:

$$\gamma_t = (\Psi_{[x,1]})_*(\nu_x). \tag{34}$$

See an example in the next section.

A more geometric way to see the correspondence between the insertion measures and γ is as follows. Project the graph of G in $\mathbb{R}^3$ onto the xz-plane; the image of the curves $G([0, 1] \times \{\tilde{y}\})$ are the flow lines of the vector field (33). The divergence of the flow lines at (x, y) is $f(x, y)$, the density associated with $F(x, y)$.

The permuton entropy can be computed from the entropy of the insertion measures as follows.

Lemma 13. *Assume the permuton γ has density $g(x, y)$ and the measures $\nu_x = \gamma|_{\{x\} \times [0,1]}$ have densities $f(x, y)$. Then*

$$H(\gamma) = \int_0^1 \int_0^x -f(x, y) \log(xf(x, y)) dy \, dx. \tag{35}$$

Proof. Differentiating (32) with respect to $\tilde{y}$ gives

$$f(x, G(x, \tilde{y})) G_y(x, \tilde{y}) = g(x, \tilde{y}). \tag{36}$$

Thus the RHS of (35) becomes

$$\int_0^1 \int_0^x -\frac{g(x, \tilde{y})}{G_y(x, \tilde{y})} \log \frac{xg(x, \tilde{y})}{G_y(x, \tilde{y})} \, dy \, dx. \tag{37}$$

Substituting $y = G(x, \tilde{y})$ with $dy = G_y(x, \tilde{y}) d\tilde{y}$ we have

$$\int_0^1 \int_0^1 -g(x, \tilde{y}) \log \frac{xg(x, \tilde{y})}{G_y(x, \tilde{y})} \, d\tilde{y} \, dx = H(\gamma) - \int_0^1 \int_0^1 g(x, \tilde{y}) \log x \, d\tilde{y} \, dx$$
$$+ \int_0^1 \int_0^1 g(x, \tilde{y}) \log G_y(x, \tilde{y}) \, d\tilde{y} \, dx. \tag{38}$$

Integrating over $\tilde{y}$ the first integral on the RHS gives

$$\int_0^1 -\log x \, dx = 1, \tag{39}$$

while the second integral is

$$\int_0^1 \int_0^1 \frac{\partial}{\partial x}(G_y \log G_y - G_y) \, dx \, d\tilde{y} = \int_0^1 (-1) d\tilde{y} = -1, \tag{40}$$

since $G(1, y) = y$ and $G(0, y) = 0$. So those two integrals cancel. ∎

6 | 12 PATTERNS

The number of occurrences $k(\pi)$ of the pattern 12 in a permutation of S_n has a simple generating function:

$$\sum_{\pi \in S_n} z^{k(\pi)} = \prod_{j=1}^{n}(1 + z + \cdots + z^j) = \sum_{i=0}^{\binom{n}{2}} C_i z^i. \tag{41}$$

One can see this by building up a permutation by insertions: when i is inserted into the list of $\{1, \dots, i-1\}$, the number of 12 patterns created is exactly one less than the position of i in that list.

For a fixed $z > 0$, we can interpret the LHS of (41) as defining a measure on permutations, giving a permutation π a probability proportional to $z^{k(\pi)}$. The product expression in the middle of (41) then indicates a way to sample from this distribution: independently for each j we pick one of the monomials $z^{y(j)}$ of $1 + z + \cdots + z^j$, with a probability proportional to its weight. The corresponding permutation is then obtained by the sequence of insertions $y(1), y(2), \dots, y(n)$. Note that these insertion probability measures are truncated geometric random variables with rate z. To sample a permutation with a fixed density $\rho \in [0, 1]$ of occurrences of pattern 12, or at least a density close to ρ, we should choose z in the above expression so that the monomial $C_{[\rho n^2/2]} z^{[\rho n^2/2]}$ is the maximal one.

When n is large we claim that the monomials $C_i z^i$ of (41) for i near $[\rho n^2/2]$ contain almost all the weight. To see this, note that under this sampling procedure the number of 12 patterns is a sum of n independent truncated geometric random variables, with the ith variable having variance at most $i^2/12$ (the variance is maximized when $z = 1$); the sum then has variance $O(n^3)$ and standard deviation $O(n^{3/2})$. So as $n \to \infty$ the resulting density is $\rho + O(n^{-1/2})$.

The value z can be determined as a function of ρ by Legendre duality (see below for an exact formula). Let r be defined by $e^{-r/n} = z$. In the limit of large n, setting $x = j/n$, the truncated geometric insertion densities converge in distribution to truncated exponentials, with densities

$$f(x, y) = \frac{r e^{-ry}}{1 - e^{-rx}} 1_{[0,x]}(y). \tag{42}$$

We can reconstruct the permuton from these insertion measures as discussed in the previous section; since as $n \to \infty$ the CDFs of the insertion measures converge, the limiting permuton is well-defined. The CDF of the insertion measure is

$$F(x, y) = \frac{1 - e^{-ry}}{1 - e^{-rx}}. \tag{43}$$

We need to solve the ODE (32), which in this case (to simplify notation we changed $\tilde{y}$ to y) is

$$\frac{1 - e^{-rG(x,y)}}{1 - e^{-rx}} = \frac{dG(x, y)}{dx}. \tag{44}$$

This can be rewritten as

$$\frac{dx}{1 - e^{-rx}} = \frac{dG}{1 - e^{-rG(x,y)}}. \tag{45}$$

Integrating both sides and solving for G gives the CDF

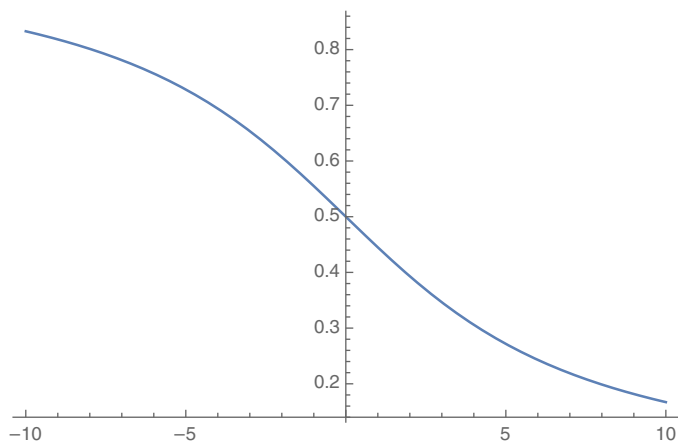$$G(x, y) = \frac{1}{r} \log\left(1 + \frac{(e^{rx} - 1)(e^{ry} - 1)}{e^r - 1}\right) \tag{46}$$

FIGURE 2 12 density as function of r [Colour figure can be viewed at wileyonlinelibrary.com]

which has density

$$g(x, y) = \frac{r(1 - e^{-r})}{(e^{r(1-x-y)/2} - e^{r(x-y-1)/2} - e^{r(y-x-1)/2} + e^{r(x+y-1)/2})^2}. \tag{47}$$

See Figure 1 for some examples for varying ρ.

The permuton entropy of this permuton is obtained from (35), and as a function of r it is (where $\text{Li}_2(x) = \sum_{j=1}^{\infty} \frac{x^j}{j^2}$ is the dilogarithm),

$$H(r) = -\frac{2\text{Li}_2(e^r)}{r} + \frac{\pi^2}{3r} - 2\log(1 - e^r) + \log(e^r - 1) - \log(r) + 2. \tag{48}$$

The density ρ of 12 patterns is the integral of the expectation of f:

$$\rho(r) = \frac{r(r - 2\log(1 - e^r) + 2) - 2\text{Li}_2(e^r)}{r^2} + \frac{\pi^2}{3r^2}; \tag{49}$$

see Figure 2 for ρ as a function of r.

Figure 3 depicts the entropy as a function of ρ.

7 | STAR MODELS

In this section, we study the density of patterns of the form $* * \cdots * k$, which we refer to as star models.

Equation (41) gives the generating function for occurrences of pattern 12. For a permutation π let $k_1 = k_1(\pi)$ be the number of 12 patterns. Let k_2 be the number of $* * 3$ patterns, that is, patterns of the form 123 or 213. A similar argument to that giving (41) shows that the joint generating function for k_1 and k_2 is

$$\sum_{k_1, k_2} C_{k_1, k_2} x^{k_1} y^{k_2} = \prod_{j=1}^{n} \left(\sum_{i=0}^{j} x^i y^{i(i-1)/2} \right). \tag{50}$$
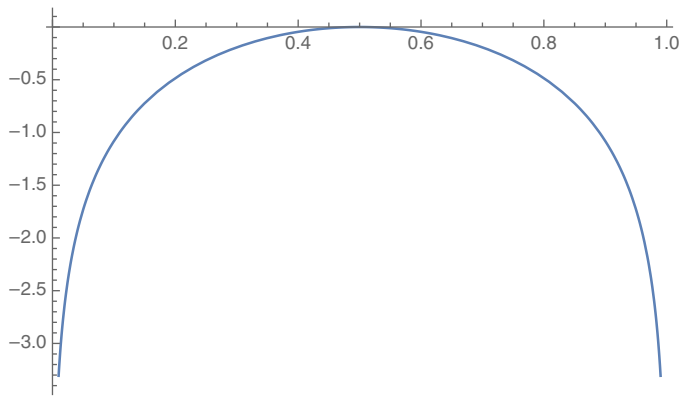
FIGURE 3 Entropy as function of 12 density [Colour figure can be viewed at wileyonlinelibrary.com]

More generally, letting k_3 be the number of patterns $**2$, that is, 132 or 312, and k_4 be the number of $**1$ patterns, that is, 231 or 321. The joint generating function for these four types of patterns is

$$\sum_{k_1,\ldots,k_4} C_{k_1,k_2,k_3,k_4} x^{k_1} y^{k_2} z^{k_3} w^{k_4} = \prod_{j=1}^{n} \left(\sum_{i=0}^{j} x^i y^{i(i-1)/2} z^{i(j-i)} w^{(j-i)(j-i-1)/2} \right). \tag{51}$$

One can similarly write down the joint generating function for all patterns of type $** \cdots * i$, with a string of some number k of stars followed by some i in $[k+1]$. (Note that with this notation, 12 patterns are $*2$ patterns.) These constitute a significant generalization of the Mallows model discussed in [36].

7.1 | The $*2/**3$ model

By way of illustration, let us consider the simplest case of $*2$ (that is, 12) and $**3$.

Theorem 14. *The feasible region for (ρ_{*2}, ρ_{**3}) is the region bounded below by the parameterized curve*

$$(2t - t^2, 3t^2 - 2t^3)_{t \in [0,1]} \tag{52}$$

and above by the parameterized curve

$$(1 - t^2, 1 - t^3)_{t \in [0,1]}. \tag{53}$$

The permutons on the boundaries are unique and supported on line segments of slopes ± 1, and are as indicated in Figure 4.

Proof. While this can be proved directly from the generating function (50), we give a simpler proof using the insertion density procedure. During the insertion process let $I_{12}(x)$ be the fractional number of 12 patterns in the partial permutation constructed up to time x. We let $Y_t \in [0, t]$ be the random variable giving the location of the insertion at time t. Then $I_{12}(x)$ is defined by $I_{12}(x) = \int_0^x Y_t \, dt$. On a small time interval $t \in [x, x + dx]$ the Y_t are independent and almost identically distributed, so by
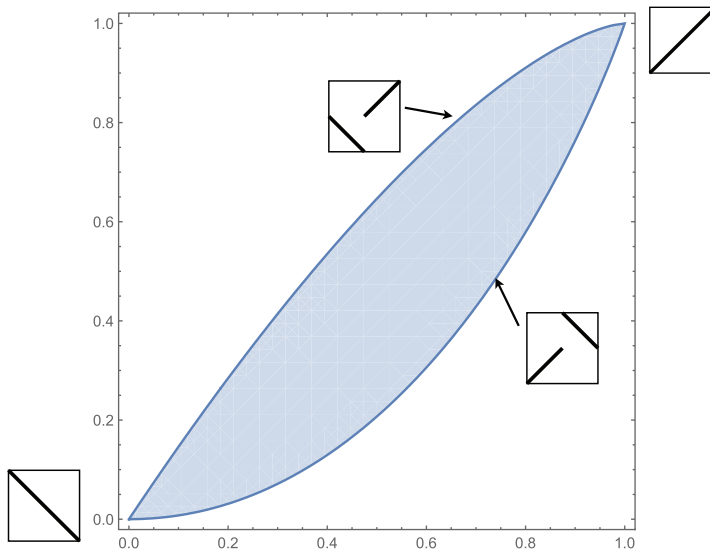
FIGURE 4 Feasible region for (ρ_{*2}, ρ_{**3}) [Colour figure can be viewed at wileyonlinelibrary.com]

the law of large numbers the number of insertions in a given subinterval is very close to dx times the Y_t-measure of that subinterval. This implies that

$$I_{12}(x + dx) - I_{12}(x) \approx \mathbb{E}[Y_t]dx,$$

and so

$$I_{12}(x) = \int_0^x \mathbb{E}[Y_t]\,dt.$$

Let $I_{**3}(x)$ be the fraction of $**3$ patterns created by time x. Note that when we insert i at position y_i in a permutation of $\{1, 2, \ldots, i - 1\}$, exactly $(y_i - 1)(y_i - 2)/2$ new $**3$ patterns are created (those ending in i). In the continuous limit, each infinitesimal insertion Y_t at time t thus creates $Y_t^2/2$ more $**3$ patterns, and as above we see that

$$I_{**3}(x + dx) - I_{**3}(x) \approx \mathbb{E}[Y_t^2/2]dx.$$

This implies that

$$I_{**3}(x) = \frac{1}{2}\int_0^x \mathbb{E}[Y_t]^2\,dt. \tag{54}$$

Let us fix $\rho_{12} = 2 \cdot I_{12}(1)$. To maximize $I_{**3}(x)$, we need to maximize

$$\int_0^1 (I'_{12}(t))^2\,dt \quad \text{subject to} \quad \int_0^1 I'_{12}(t)\,dt = \frac{\rho_{12}}{2} \tag{55}$$

with the additional constraint $0 \le I'_{12}(t) \le t$. This maximization is only achieved if for each t, $I'_{12}(t)$ is either zero or its maximal value t. In fact the maximum is achieved by inserting points at the beginning for as long as possible and then inserting points at the end, that is, $Y_t = 0$ up to $t = a$ and then $Y_t = t$
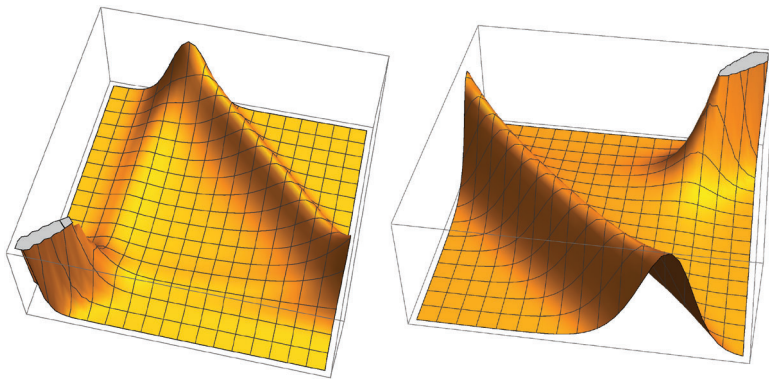
FIGURE 5 Permutons with $(\rho_{*2}, \rho_{**3}) = (0.5, 0.2)$, and $(0.5, 0.53)$ respectively [Colour figure can be viewed at wileyonlinelibrary.com]
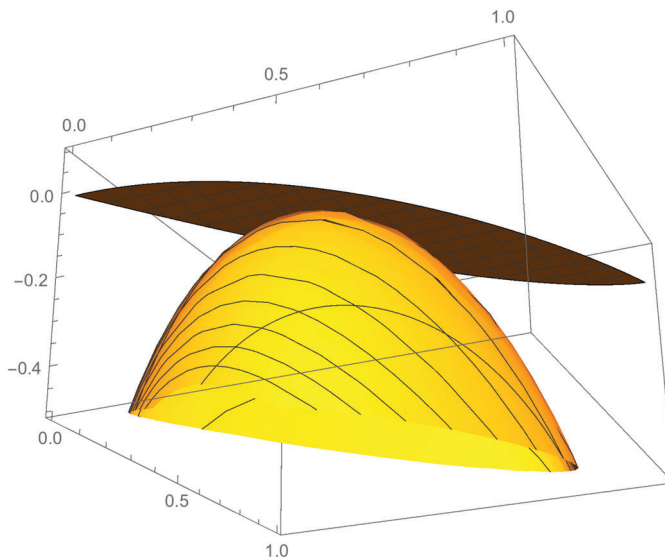

FIGURE 6 The entropy function on the parameter space for ρ_{12}, ρ_{**3} [Colour figure can be viewed at wileyonlinelibrary.com]

for $t \in [a, 1]$. The resulting permuton is then as shown in Figure 4, upper curve: on the square $[0, a]^2$ it is a descending diagonal and on the square $[a, 1]^2$ it is an ascending diagonal.

Likewise to minimize the above integral (55) we need to make the derivatives $I'_{12}(t)$ as equal as possible. Since $I'_{12}(t) \leq t$, this involves setting $I'_{12}(t) = t$ up to $t = a$ and then having it constant after that. The resulting permuton is then as shown in Figure 4: on the square $[0, a]^2$ it is an ascending diagonal and on the square $[a, 1]^2$ it is a descending diagonal.

A short calculation now yields the algebraic form of the boundary curves. ∎

Using the insertion density procedure outlined earlier, we see that the permuton as a function of x, y has an explicit analytic density (which cannot, however, be written in terms of elementary functions). The permutons for some values of (ρ_{*2}, ρ_{**3}) are shown in Figure 5.

The entropy $s(\rho_{*2}, \rho_{**3})$ is plotted in Figure 6. It is strictly concave (see Theorem 15) and achieves its maximal value, zero, precisely at the point $1/2, 1/3$, the uniform measure.

7.2 | Concavity and analyticity of entropy for star models

Theorem 15. *For a star model with a finite number of densities $\rho_1, \ldots, \rho_k$ of patterns $\tau_1 \ldots, \tau_k$ respectively, the feasible region is convex and the entropy $H(\rho_1, \ldots, \rho_k)$ is strictly concave and analytic on the feasible region. For each $\rho_1, \ldots, \rho_k$ in the interior of the feasible region there is a unique entropy-maximizing permuton with those densities, and this permuton has analytic probability density.*

One can construct examples where the feasible region is not strictly convex: for example, in the case of densities $*1$ and $*3$.

Proof. Let k_i be the length of the pattern τ_i.

The generating function for permutations of $[n]$ counting patterns τ_i is

$$Z_n(x_1, \ldots, x_k) = \sum_{\pi \in S_n} x_1^{n_1} \ldots x_k^{n_k}, \tag{56}$$

where $n_i = n_i(\pi)$ is the number of occurrences of pattern τ_i in π. The number of permutations with density ρ_i of pattern τ_i is the sum of the coefficients of the terms $x_1^{n_1} \ldots x_k^{n_k}$ with $n_i \approx \frac{n^{k_i}}{k_i!}\rho_i$; more precisely, we need $\frac{n_i}{n^{k_i}}k_i! \to \rho$. The entropy $H(\rho_1, \ldots, \rho_k)$ is the log of this sum, minus $\log n!$ (and normalized by dividing by n).

As discussed above, Z_n can be written as a product generalizing (51). Write $x_i = e^{a_i}$. Then the product expression for Z_n is

$$Z_n = \prod_{j=1}^{n} \sum_{i=0}^{j} e^{p(i,j)}, \tag{57}$$

where $p(i,j)$ is a polynomial in i and j with coefficients that are linear in the a_i. Using this product form, for appropriate values of x_i, the maximal terms in (56) will be those with $n_i \approx \frac{n^{k_i}}{k_i!}\rho_i$. This follows from the same argument as in the proof of the 12 case in Section 6: for fixed $x_i > 0$ the standard deviation of the number of occurrences of pattern τ_i will be $n^{k_i-1/2}$, so that $\frac{n_i}{n^{k_i}}k_i! = \rho_i + O(n^{-1/2})$.

For large n it is convenient to normalize the a_i by the appropriate power of n: write

$$x_i = e^{a_i} = \exp\left(\alpha_i / n^{k_i-1}\right). \tag{58}$$

Writing $i/n = t$ and $j/n = x$, the expression for $\log Z_n$ is then a Riemann sum, once normalized: In the limit $n \to \infty$ the "normalized free energy" F is

$$F := \lim_{n \to \infty} \frac{1}{n}(\log Z_n - \log n!) = \int_0^1 \left[\log \int_0^x e^{\tilde{p}(t,x)} \, dt\right] dx, \tag{59}$$

where $\tilde{p}(t,x) = p(nt, nx) + o(1)$ is a polynomial in t and x, independent of n, with coefficients which are linear functions of the α_i. Explicitly we have

$$\tilde{p}(t,x) = \sum_{i=1}^{k} \alpha_i \frac{t^{r_i}(x-t)^{s_i}}{r_i! s_i!}, \tag{60}$$

where $r_i + s_i = k_i - 1$ and, if $\tau_i = *\cdots *\ell_i$ then $s_i = k_i - \ell_i$.

We now show that F is concave as a function of the α_i, by computing its Hessian matrix. We have

$$\frac{\partial F}{\partial \alpha_i} = \int_0^1 \frac{\int_0^x \frac{t^{r_i}(x-t)^{s_i}}{r_i! s_i!} e^{\tilde{p}(t,x)} dt}{\int_0^x e^{\tilde{p}(t,x)} dt} dx = \int_0^1 \left\langle \frac{T^{r_i}(x-T)^{s_i}}{r_i! s_i!} \right\rangle dx, \tag{61}$$

where $T \in [0, x]$ is the random variable with (unnormalized) density $e^{\tilde{p}(t,x)}$, and $\langle \cdot \rangle$ is the expectation with respect to this probability measure.

Differentiating a second time we have

$$
\begin{aligned}
\frac{\partial^2 F}{\partial \alpha_j \partial \alpha_i} &= \int_0^1 \left\langle \frac{T^{r_i + r_j}(x - T)^{s_i + s_j}}{r_i! r_j! s_i! s_j!} \right\rangle - \left\langle \frac{T^{r_i}(x - T)^{s_i}}{r_i! s_i!} \right\rangle \left\langle \frac{T^{r_j}(x - T)^{s_j}}{r_j! s_j!} \right\rangle dx \\
&= \int_0^1 \text{Cov}\left[\frac{T^{r_i}(x - T)^{s_i}}{r_i! s_i!}, \frac{T^{r_j}(x - T)^{s_j}}{r_j! s_j!} \right] dx,
\end{aligned}
\tag{62}
$$

where Cov is the covariance.

The covariance matrix of a set of random variables with no linear dependencies is positive definite. Thus we see that the Hessian matrix is an integral of positive definite matrices and so is itself positive definite. This completes the proof of strict concavity of the free energy F.

Since Z_n is the (unnormalized) probability generating function, the vector of densities as a function of the $\{\alpha_i\}$ is obtained for each n by the gradient of the logarithm

$$
(\rho_1, \ldots, \rho_k) = \frac{1}{n} \nabla \log Z_n(\alpha_1, \ldots, \alpha_k).
\tag{63}
$$

In the limit we can replace $\frac{1}{n} \nabla \log Z_n$ by ∇F; by strict concavity of F its gradient is injective, and surjective onto the interior of the feasible region. In particular there is a unique choice of α_i's for every choice of densities in the interior of the feasible region. Note that the α_i's determine the insertion measures (these are the measures with unnormalized density $e^{\tilde{p}(t,x)}$), and thus the permuton itself, proving uniqueness of the entropy maximizer. Analyticity of the probability density is a consequence of analyticity of the associated differential equation (32).

By strict concavity of the free energy, we can relate the free energy to the entropy by the following standard argument. Referring back to the first paragraph of the proof, we have proven that, when $x_i = e^{\alpha_i}$, the generating function Z_n concentrates its mass on the terms $x_1^{n_1} \ldots x_k^{n_k}$ for which $n_i \approx \frac{n^{k_i}}{k_i!} \rho_i$ (where a_i and ρ_i are related by (63)), in the sense that a fraction $1 - o(1)$ of the total mass of Z_n is on these terms. The entropy is the log of the sum of the coefficients in front of these relevant terms. The entropy can thus be obtained from the free energy $\frac{1}{n} \log Z_n / n!$ by subtracting off $\frac{1}{n} \log(x_1^{n_1} \ldots x_k^{n_k})$. This shows that the entropy function H is the Legendre dual of F, that is,

$$
H(\rho_1, \ldots, \rho_k) = \max_{\{\alpha_i\}} \{ F(\alpha_1, \ldots, \alpha_k) - \sum \alpha_i \rho_i \}.
\tag{64}
$$

Analyticity of F implies that H is both analytic and strictly concave.

The "upper level sets" $\{\vec{\rho} : H(\vec{\rho}) \geq -M\}$ of H are convex by concavity of H. Their union is the interior of the feasible region, which, being an increasing union of convex sets, is convex. ∎

8 | PDES FOR PERMUTONS

For permutations with constraints on patterns of length 3 (or less) one can write explicit PDEs for the maximizers. It is possible that these may be used to show either analyticity or uniqueness, or both (although at the moment we have accomplished neither goal).

By Proposition 9 we may assume that our maximizing permutons have densities g, and moreover that $g \log g$ is integrable.

Let us first redo the case of 12-patterns, which we already worked out by another method in Section 6.

8.1 | Patterns 12

The density of patterns 12 is given in (2). Consider the problem of maximizing $H(\gamma)$ subject to the constraint $I_{12}(\gamma) = \rho$. This involves finding a solution to the Euler-Lagrange equation

$$dH + \alpha \, dI_{12} = 0 \tag{65}$$

for some constant α, for all variations $g \mapsto g + \varepsilon h$ fixing the marginals.

Given points $(a_1, b_1), (a_2, b_2) \in [0, 1]^2$ we can consider the change in H and I_{12} when we remove an infinitesimal mass δ from (a_1, b_1) and (a_2, b_2) (assuming the mass is positive there) and add it to locations (a_1, b_2) and (a_2, b_1). Note that two measures with the same marginals are connected by convolutions of such operations. The change in H to first order under such an operation is δ times (letting $S_0(p) := -p \log p$)

$$-S_0'(g(a_1, b_1)) - S_0'(g(a_2, b_2)) + S_0'(g(a_1, b_2)) + S_0'(g(a_2, b_1)) = \log \frac{g(a_1, b_1)g(a_2, b_2)}{g(a_1, b_2)g(a_2, b_1)}. \tag{66}$$

Under this operation the change in I_{12} to first order is δ times

$$\sum_{i,j=1}^{2} (-1)^{i+j} \left(\int_{a_i < x_2} \int_{b_j < y_2} g(x_2, y_2) dx_2 \, dy_2 + \int_{x_1 < a_i} \int_{y_1 < b_j} g(x_1, y_1) dx_1 \, dy_1 \right)$$

$$= \sum_{i,j=1}^{2} (-1)^{i+j} \left(G(a_i, b_j) + (1 - a_i - b_j + G(a_i, b_j)) \right). \tag{67}$$

Thus (65) becomes an explicit integral equation for g. If g is smooth, in the sense that g_{xy} exists, then we can differentiate (65) with respect to $a = a_1$ and $b = b_1$, getting a PDE

$$\frac{\partial}{\partial a} \frac{\partial}{\partial b} \log g(a, b) + 2\alpha g(a, b) = 0. \tag{68}$$

One can check that the formula (47) satisfies this PDE.

8.2 | Patterns 123

The density of patterns 123 is

$$I_{123}(\gamma) = 6 \int_{x_1 < x_2 < x_3, \, y_1 < y_2 < y_3} g(x_1, y_1)g(x_2, y_2)g(x_3, y_3)dx_1 \, dx_2 \ldots \, dy_3. \tag{69}$$

Under a similar perturbation as above the change in I_{123} to first order is δ times

$$dI_{123} = 6 \sum_{i,j=1}^{2} (-1)^{i+j} \left(\int_{a_i < x_2 < x_3, \, b_j < y_2 < y_3} g(x_2, y_2)g(x_3, y_3)dx_2 \, dx_3 \, dy_2 \, dy_3 \right.$$

$$+ \int_{x_1 < a_i < x_3, \; y_1 < b_j < y_3} g(x_1, y_1) g(x_3, y_3) dx_1 \, dx_3 \, dy_1 \, dy_3$$

$$+ \int_{x_1 < x_2 < a_i, \; y_1 < y_2 < b_j} g(x_1, y_1) g(x_2, y_2) dx_1 \, dx_2 \, dy_1 \, dy_2 \Bigg). \tag{70}$$

The middle integral here is a product

$$\int_{x_1 < a_i, \; y_1 < b_j} g(x_1, y_1) dx_1 \, dy_1 \int_{a_i < x_3, \; b_j < y_3} g(x_3, y_3) dx_3 \, dy_3$$

$$= G(a_i, b_j)(1 - a_i - b_j + G(a_i, b_j)). \tag{71}$$

Differentiating each of these three integrals with respect to both $a = a_1$ and $b = b_1$ (then only the $i = j = 1$ term survives) gives, for the first integral

$$g(a, b) \int_{a < x_3, \; b < y_3} g(x_3, y_3) dx_3 \, dy_3 = g(a, b)(1 - a - b + G(a, b)), \tag{72}$$

for the second integral

$$g(a, b)(1 - a - b + 2G(a, b)) + G_x(a, b)(-1 + G_y(a, b))$$

$$+ G_y(a, b)(-1 + G_x(a, b)), \tag{73}$$

and the third integral

$$g(a, b) \int_{x_1 < a, \; b < y_1} g(x_1, y_1) dx_1 \, dy_1 = g(a, b)G(a, b). \tag{74}$$

Summing, we get (changing a, b to x, y),

$$(dI_{123})_{xy} = 12G_{xy}(1 - x - y + 2G) + 12G_x G_y - 6G_x - 6G_y. \tag{75}$$

Thus under the assumption that g is smooth, the Euler-Lagrange equation is

$$(\log G_{xy})_{xy} + 6\alpha \big(2G_{xy}(1 - x - y + 2G) + 2G_x G_y - G_x - G_y\big) = 0. \tag{76}$$

This simplifies somewhat if we define $K(x, y) = 2G(x, y) - x - y + 1$. Then

$$(\log K_{xy})_{xy} + 3\alpha \big(2K_{xy}K + K_x K_y - 1\big) = 0. \tag{77}$$

In a similar manner we can find a PDE for the permuton with fixed densities of other patterns of length 3. In fact one can proceed similarly for longer patterns, getting systems of PDEs, but the complexity grows with the length.

9 | THE 12/123 MODEL

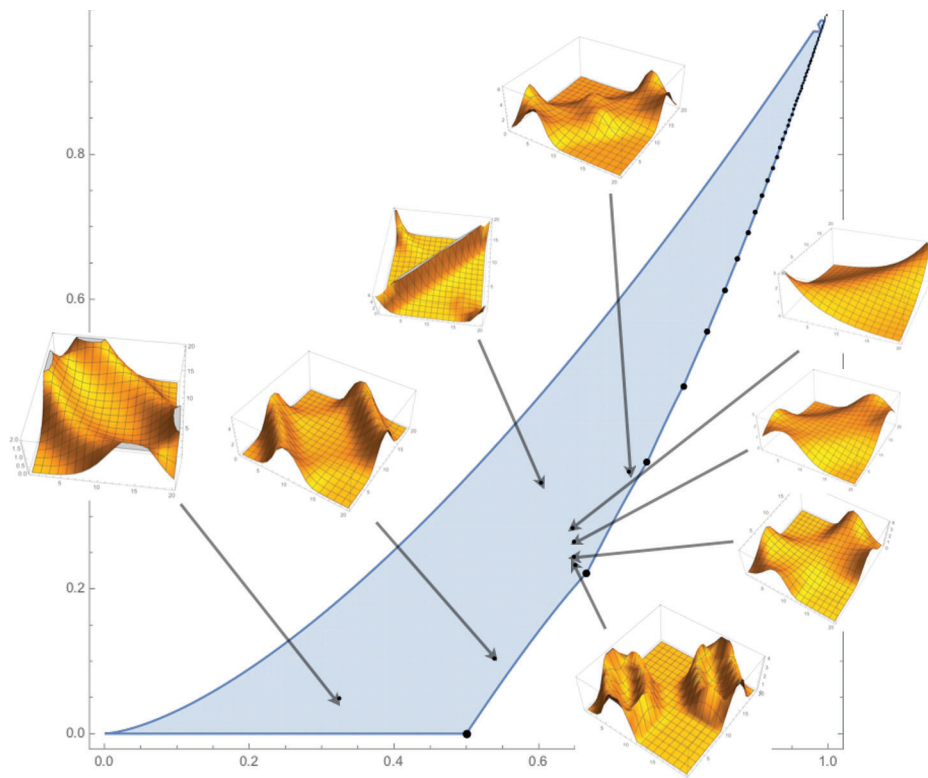When we fix the density of patterns 12 and 123, the feasible region has a complicated structure, see Figure 7.

FIGURE 7 The feasible region for ρ_{12} versus ρ_{123}, with corresponding permutons (computed numerically) at selected points [Colour figure can be viewed at wileyonlinelibrary.com]

Theorem 16. *The feasible region for ρ_{12} versus ρ_{123} is the same as the feasible region of edges and triangles in the graphon model.*

Proof. Let $\mathcal{R}$ denote the feasible region for pairs $\big(\rho_{12}(\gamma), \rho_{123}(\gamma)\big)$ consisting of the 12 density and 123 density of a permuton (equivalently, for the closure of the set of such pairs for finite permutations).

Each permutation $\pi \in S_n$ determines a (two-dimensional) poset P_π on $\{1, \ldots, n\}$ given by $i \prec j$ in P_π iff $i < j$ and $\pi_i < \pi_j$. The *comparability graph* $G(P)$ of a poset P links two points if they are comparable in P, that is, $x \sim y$ if $x \prec y$ or $y \prec x$. Then $i \sim j$ in $G(P_\pi)$ precisely when $\{i, j\}$ constitutes an incidence of the pattern 12, and $i \sim j \sim k \sim i$ when $\{i, j, k\}$ constitutes an incidence of the pattern 123. Thus the 12 density of π is equal to the edge density of $G(P_\pi)$, and the 123 density of π is the triangle density of $G(P_\pi)$—that is, the probability that three random vertices induce the complete graph K_3. This correspondence extends perfectly to limit objects, equating 12 and 123 densities of permutons to edge densities and triangle densities of graphons.

The feasible region for edge and triangle densities of graphs (now, for graphons) has been studied for many years and was finally determined by Razborov [34]; we call it the "scalloped triangle" $\mathcal{T}$. It follows from the above discussion that the feasibility region $\mathcal{R}$ we seek for permutons is a subset of $\mathcal{T}$, and it remains only to prove that $\mathcal{R}$ is all of $\mathcal{T}$. In fact we can realize $\mathcal{T}$ using only a rather simple two-parameter family of permutons.

Let reals a, b satisfy $0 < a \leq 1$ and $0 < b \leq a/2$, and set $k := \lfloor a/b \rfloor$. Let us denote by $\gamma_{a,b}$ the permuton consisting of the following diagonal line segments, all of equal density (see Figure 8):
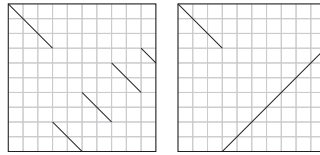
FIGURE 8 Support of the permutons $\gamma_{0.7,0.2}$ and $\gamma_{0.7,0}$

1. The segment $y = 1 - x$, for $0 \leq x \leq 1-a$;
2. The k segments $y = (2j-1)b - 1 + a - x$ for $1 - a + (j-1)b < x \leq 1 - a + jb$, for each $j = 1, 2, \ldots, k$;
3. The remaining, rightmost segment $y = 1 + kb - x$, for $1 - a + kb < x \leq 1$.

We interpret $\gamma_{a,0}$ as the permuton containing the segment $y = 1 - x$, for $0 \leq x \leq 1-a$, and the positive-slope diagonal from $(1-a, 0)$ to $(1, 1-a)$; finally, $\gamma_{0,0}$ is just the reverse diagonal from $(0, 1)$ to $(1, 0)$. These interpretations are consistent in the sense that $\rho_{12}(\gamma_{a,b})$ and $\rho_{123}(\gamma_{a,b})$ are continuous functions of a and b on the triangle $0 \leq a \leq 1, 0 \leq b \leq a/2$. (In fact, $\gamma_{a,b}$ is itself continuous in the topology of Γ, so all pattern densities are continuous.)

It remains only to check that the comparability graphons corresponding to these permutons match extremal graphs in [34] as follows:

- $\gamma_{a,0}$ maps to the upper left boundary of $\mathcal{T}$, with $\gamma_{0,0}$ going to the lower left corner while $\gamma_{1,0}$ goes to the top;
- $\gamma_{a,a/2}$ goes to the bottom line, with $\gamma_{1,1/2}$ going to the lower right corner;
- For $1/(k+2) \leq b \leq 1/(k+1)$, $\gamma_{1,b}$ goes to the kth lowest scallop, with $\gamma_{1,1/(k+1)}$ going to the bottom cusp of the scallop and $\gamma_{1,1/(k+2)}$ to the top.

It follows that $(a, b) \mapsto \big(\rho_{12}(\gamma_{a,b}), \rho_{123}(\gamma_{a,b})\big)$ maps the triangle $0 \leq a \leq 1, 0 \leq b \leq a/2$ onto all of $\mathcal{T}$, proving the theorem. ∎

It may be prudent to remark at this point that while the feasible region for 12 versus 123 density of permutons is the same as that for edge and triangle density of graphs, the *topography* of the corresponding entropy functions within this region is entirely different. In the graph case the entropy landscape is studied in [31–33]; one of its features is a ridge along the "Erdős-Rényi" curve (where triangle density is the 3/2 power of edge density). There is a sharp drop-off below this line, which represents the very high entropy graphs constructed by choosing edges independently with constant probability. The graphons that maximize entropy at each point of the feasible region all appear to be very combinatorial in nature: each has a partition of its vertices into finitely many classes, with constant edge density between any two classes and within any class, and is thus described by a finite list of real parameters.

The permuton topography features a different high curve, representing the permutons (discussed above) that maximize entropy for a fixed 12 density. Moreover, the permutons that maximize entropy at interior points of the region appear, as in other regions discussed above, always to be analytic.

We do not know explicitly the maximizing permutons (although they satisfy an explicit PDE, see Section 8) or the entropy function.

10 | 123/321 CASE

The feasible region for fixed densities ρ_{123} versus ρ_{321} is the same as the feasible region $\mathcal{B}$ for triangle density $x = d(K_3, G)$ versus anti-triangle density $y = d(\overline{K_3}, G)$ of graphons [19]. Let C be the line
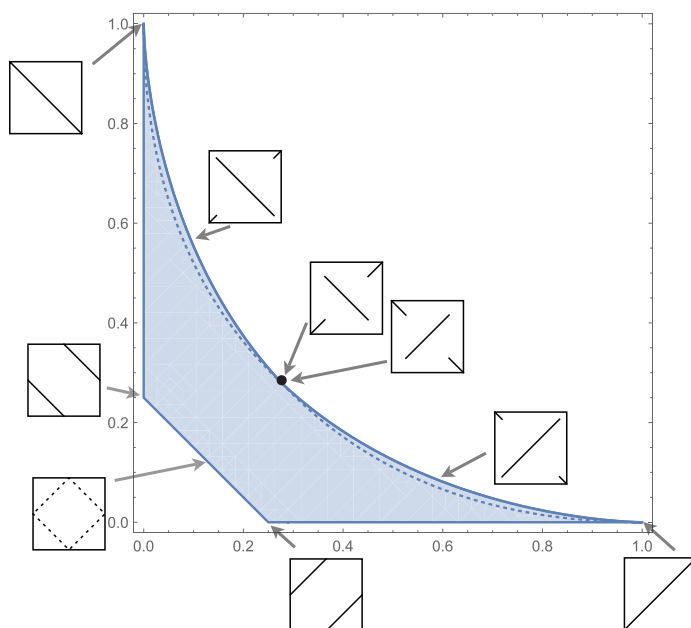
FIGURE 9 The feasible region for ρ_{123}, ρ_{321}. It is bounded above by the parameterized curves $(1 - 3t^2 + 2t^3, t^3)$ and $(t^3, 1 - 3t^2 + 2t^3)$ which intersect at $(x, y) = (0.278\ldots, 0.278, \ldots)$. The lower boundaries consist of the axes and the line $x + y = 1/4$ [Colour figure can be viewed at wileyonlinelibrary.com]

segment $x + y = \frac{1}{4}$ for $0 \leq x \leq \frac{1}{4}$, D the x-axis from $x = \frac{1}{4}$ to $x = 1$, and E the y-axis from $y = \frac{1}{4}$ to $y = 1$. Let F_1 be the curve given parametrically by $(x, y) = (t^3, (1 - t)^3 + 3t(1 - t)^2)$, for $0 \leq t \leq 1$, and F_2 its symmetric twin $(x, y) = ((1 - t)^3 + 3t(1 - t)^2, t^3)$. Then $\mathcal{B}$ is the union of the area bounded by C, D, E and F_1 and the area bounded by C, D, E and F_2.

The curves F_1 and F_2 cross at a concave "dimple" (r, r) where $r = s^3 = (1 - s)^3 + 3s(1 - s)^2)$, with $s \sim 0.653$ and $r \sim 0.278$ (see Figure 9).

To see that $\mathcal{B}$ is also the feasible region for 123 versus 321 density of permutons, an argument much like the one above for 12 versus 123 can be (and was, by [11]) given. Permutons realizing various boundary points are illustrated in Figure 9; they correspond to the extremal graphons described in [19]. The rest are filled in by parameterization and a topological argument (essentially Theorem 4 above, once we add an extra spike poking into R from the dimple to continuously interpolate between the two permutons at the dimple).

Of note for both graphons and permutons is the double solution at the dimple. These solutions are significantly different, as evidenced by the fact that their edge-densities (12 densities, for the permutons) differ. This multiplicity of solutions, if there are no permutons bridging the gap, suggests a phase transition in the entropy-optimal permuton in the interior of $\mathcal{B}$ in a neighborhood of the dimple. In fact, we can use a stability theorem from [18] to show that the phenomenon is real.

Before stating the next theorem, we need a definition: two n-vertex graphs are ε-*close* if one can be made isomorphic to the other by adding or deleting at most $\varepsilon \cdot \binom{n}{2}$ edges.

Theorem 17 (special case of Theorems 1.1 and 1.2 of [18]). *For any $\varepsilon > 0$ there is a $\delta > 0$ and an N such that for any n-vertex graph G with $n > N$ and $d(\overline{K_3}, G) \geq p$ and $|d(K_3, G) - M_p| < \delta$ is ε-close to a graph H on n vertices consisting of a clique and isolated vertices, or the complement of a graph consisting of a clique and isolated vertices. Here $M_p := \max\left((1 - p^{1/3})^3 + 3p^{1/3}(1 - p^{1/3})^2, (1 - q)^{1/3}\right)$*

where q is the unique real root of $q^3 + 3q^2(1 - q) = p$; that is, M_p is the largest possible value of $d(K_3, G)$ given $d(\overline{K_3}, G) = p$.

From Theorem 17 we derive the following lemma. Note that there are in fact many permutons representing the dimple (r, r) of the feasible region for 123 versus 321, but only two classes if we consider permutons with isomorphic comparability graphs to be equivalent. The class that came from the curve F_1 has 12 density $s^2 \sim 0.426$, the other $1 - s^2 \sim 0.574$. (Interestingly, the other end of the F_1 curve—represented uniquely by the identity permuton—had 12 density 1, while the F_2 class "began" at 12 density 0. Thus, the 12 densities crossed on the way in from the corners of $\mathcal{B}$.)

Lemma 18. *There is a neighborhood of the point (r, r) in the feasible region for patterns 123 and 321 within which no permuton has 12-density near $\frac{1}{2}$.*

Proof. Apply Theorem 17 with $\varepsilon = 0.07$ to get $\delta > 0$ with the property stated in the theorem. Let $\delta' = \min(\delta/2, (M_{r-\delta} - r)/2)$, which yields that $|M_p - r| \leq \delta/2$ for $p \in [r - \delta', r]$. So, if $|\rho_{123}(\gamma) - r| \leq \delta' \leq \delta/2$ and $p \in [r - \delta', r]$, then $|\rho_{123}(\gamma) - M_p| \leq \delta$ as required by the hypothesis of Theorem 17 (noting that $\rho_{123}(\gamma)$ is the triangle density of the comparability graph corresponding to γ). We conclude that any permuton γ such that $(\rho_{123}(\gamma), \rho_{321}(\gamma))$ lies in the rectangle $[r - \delta', r + \delta'] \times [r - \delta', r]$ has 1 2-density within 0.07 of either 0.426 or 0.574, thus outside the range $[0.496, 0.504]$.

The symmetric argument gives the same conclusion for $(\rho_{123}(\gamma), \rho_{321}(\gamma))$ in the rectangle $[r - \delta', r] \times [r - \delta', r + \delta']$. Since there are no permutons γ with both $(\rho_{123}(\gamma)$ and $\rho_{321}(\gamma)$ larger than r, the lemma follows. ∎

11 | PROOF OF THEOREM 1

For completeness, we now give a proof of Theorem 1. We begin with a simple lemma.

Lemma 19. *The function $H : \Gamma \to \mathbb{R}$ is upper semicontinuous.*

Proof. Let $\gamma_1, \gamma_2, \ldots$ be a sequence of permutons approaching the permuton γ (in the $d_\square$-topology); we need to show that $H(\gamma) \geq \limsup H(\gamma_n)$.

If $H(\gamma)$ is finite, fix $\varepsilon > 0$ and take m large enough so that $|H(\gamma^m) - H(\gamma)| < \varepsilon$; then since $H(\gamma_n^m) \geq H(\gamma_n)$ by concavity,

$$\limsup_n H(\gamma_n) \leq \limsup_n H(\gamma_n^m) = H(\gamma^m) < \varepsilon + H(\gamma) \tag{78}$$

and since this holds for any $\varepsilon > 0$, the claimed inequality follows.

If $H(\gamma) = -\infty$, fix $t < 0$ and take m so large that $H(\gamma^m) < t$. Then

$$\limsup_n H(\gamma_n) \leq \limsup_n H(\gamma_n^m) = H(\gamma^m) < t \tag{79}$$

for all t, so $\limsup_n H(\gamma_n^m) \to -\infty$ as desired. ∎

Let $B(\gamma, \varepsilon) = \{\gamma' | d_\square(\gamma, \gamma') \leq \varepsilon\}$ be the (closed) ball in Γ of radius $\varepsilon > 0$ centered at the permuton γ, and let $B_n(\gamma, \varepsilon)$ be the set of permutations $\pi \in S_n$ with $\gamma_\pi \in B(\gamma, \varepsilon)$.

Lemma 20. *For any permuton γ, $\lim_{\varepsilon \downarrow 0} \lim_{n \to \infty} \frac{1}{n} \log(|B_n(\gamma, \varepsilon)|/n!)$ exists and equals $H(\gamma)$.*

Proof. Suppose $H(\gamma)$ is finite. It suffices to produce two sets of permutations, $U \subset B_n(\gamma, \varepsilon)$ and $V \supset B_n(\gamma, \varepsilon)$, each of size

$$\exp\left(n \log n - n + n(H(\gamma) + o(\varepsilon^0)) + o(n)\right), \tag{80}$$

where by $o(\varepsilon^0)$ we mean a function of ε (depending on γ) which approaches 0 as $\varepsilon \to 0$. (The usual notation here would be $o(1)$; we use $o(\varepsilon^0)$ here and later to make it clear that the relevant variable is ε and not, e.g., n.)

To define U, fix $m > 5/\varepsilon$ so that $|H(\gamma^m) - H(\gamma)| < \varepsilon$ and let n be a multiple of m with $n > m^3/\varepsilon$. Choose integers $n_{i,j}$, $1 \le i, j \le m$, so that:

1. $\sum_{i=1}^{n} n_{i,j} = n/m$ for each j;
2. $\sum_{j=1}^{m} n_{i,j} = n/m$ for each i; and
3. $|n_{i,j} - n\gamma(Q_{ij})| < 1$ for every i, j.

The existence of such a rounding of the matrix $\{n\gamma(Q_{ij})\}_{i,j}$ is guaranteed by Baranyai's rounding lemma [2].

Let U be the set of permutations $\pi \in S_n$ with exactly $n_{i,j}$ points in the square Q_{ij}, that is, $|\{i : (i/n, \pi(i)/n) \in Q_{ij}\}| = n_{i,j}$, for every $1 \le i, j \le m$. We show first that U is indeed contained in $B_n(\gamma, \varepsilon)$. Let $R = [a, b] \times [c, d]$ be a rectangle in $[0, 1]^2$. R will contain all Q_{ij} for $i_0 < i < i_1$ and $j_0 < j < j_1$ for suitable i_0, i_1, j_0 and j_1, and by construction the γ_π-measure of the union of those rectangles will differ from its γ-measure by less than $m^2/n < \varepsilon/m$. The squares cut by R are contained in the union of two rows and two columns of width $1/m$, and hence, by the construction of π and the uniformity of the marginals of γ, cannot contribute more than $4/m < 4\varepsilon/5$ to the difference in measures. Thus, finally, $d_\square(\gamma_\pi, \gamma) < \varepsilon/m + 4\varepsilon/5 < \varepsilon$.

Now we must show that $|U|$ is close to the claimed size

$$\exp\left(n \log n - n - H(\gamma)n\right). \tag{81}$$

We construct $\pi \in U$ in two phases of m steps each. In step i of Phase I, we decide for each k, $(i-1)n/m < k \le in/m$, which of the m y-intervals $\pi(k)$ should lie in. There are

$$\binom{n/m}{n_{i,1}, n_{i,2}, \ldots, n_{i,m}} = \exp\left((n/m)h_i + o(n/m)\right) \tag{82}$$

ways to do this, where $h_i = -\sum_{j=1}^{m}(n_{i,j}/(n/m)) \log(n_{i,j}/(n/m))$ is the entropy of the probability distribution $n_{i,\cdot}/(n/m)$.

Thus, the number of ways to accomplish Phase I is

$$\begin{aligned}
\exp\left(o(n) + (n/m)\sum_i h_i\right) &= \exp\left(o(n) - \sum_{i,j} n_{i,j} \log(n_{i,j}/(n/m))\right) \\
&= \exp\left(o(n) - \sum_{i,j} n_{i,j}(\log(n_{i,j}/n) + \log m)\right) \\
&= \exp\left(o(n) - n \log m - \sum_{i,j} n_{i,j} \log \gamma(Q_{ij})\right) \\
&= \exp\left(o(n) - n \log m - n \sum_{i,j} \gamma(Q_{ij}) \log \gamma(Q_{ij})\right). \tag{83}
\end{aligned}$$

Recalling that the value taken by the density g^m of γ^m on the points of Q_{ij} is $m^2\gamma(Q_{ij})$, we have that

$$
\begin{aligned}
H(\gamma^m) &= \sum_{i,j} \frac{1}{m^2}\big(-m^2\gamma(Q_{ij})\log(m^2\gamma(Q_{ij}))\big) \\
&= -\sum_{i,j} \gamma(Q_{ij})(\log\gamma(Q_{ij}) + 2\log m) \\
&= -\sum_{i,j} \gamma(Q_{ij})(\log\gamma(Q_{ij}) + 2\log m) \\
&= -2\log m - \sum_{i,j} \gamma(Q_{ij})\log\gamma(Q_{ij}) .
\end{aligned}
\tag{84}
$$

Therefore we can rewrite the number of ways to do Phase I as

$$
\exp\big(n\log m + nH(\gamma^m) + o(n)\big).
\tag{85}
$$

In Phase II we choose a permutation $\pi_j \in S_{n/m}$ for each j, $1 \le j \le m$, and order the y-coordinates of the n/m points (taken left to right) in row j according to π_j. Together with Phase I this determines π uniquely, and the number of ways to accomplish Phase II is

$$
\begin{aligned}
(n/m)!^m &= \left(\exp\left(\frac{n}{m}\log\frac{n}{m} - \frac{n}{m} + o(n/m)\right)\right)^m \\
&= \exp\big(n\log n - n - n\log m + o(n)\big)
\end{aligned}
\tag{86}
$$

so that in total,

$$
\begin{aligned}
|U| &\ge \exp\big(n\log m + nH(\gamma^m) + o(n)\big)\exp\big(n\log n - n - n\log m + o(n)\big) \\
&= \exp\big(n\log n - n + nH(\gamma^m) + o(n)\big)
\end{aligned}
\tag{87}
$$

which, since $|H(\gamma) - H(\gamma^m)| < \varepsilon$, does the job.

We now proceed to the other bound, which involves similar calculations in a somewhat different context. To define the required set $V \supset B_n(\gamma, \varepsilon)$ of permutations we must allow a wide range for the number of points of π that fall in each square Q_{ij}— wide enough so that a violation causes Q_{ij} itself to witness $d_\square(\gamma_\pi, \gamma) > \varepsilon$, thus guaranteeing that if $\pi \notin V$ then $\pi \notin B_n(\gamma, \varepsilon)$.

To do this we take m large, $\varepsilon < 1/m^4$, and $n > 1/\varepsilon^2$. We define V to be the set of permutations $\pi \in S_n$ for which the number of points $(k/n, \pi(k)/n)$ falling in Q_{ij} lies in the range $[n(\gamma(Q_{ij}) - \sqrt{\varepsilon}), n(\gamma(Q_{ij}) + \sqrt{\varepsilon})]$. Then, as promised, if $\pi \notin V$ we have a rectangle $R = Q_{ij}$ with $|\gamma(R) - \gamma_\pi(R)| > \sqrt{\varepsilon}/m^2 > \varepsilon$.

It remains only to bound $|V|$. Here a preliminary phase is needed in which the exact count of points in each square Q_{ij} is determined; since the range for each $n_{i,j}$ is of size $2n\sqrt{\varepsilon}$, there are at most $\left(2n\sqrt{\varepsilon}\right)^{m^2} = \exp\big(m^2\log(2n\sqrt{\varepsilon})\big)$ ways to do this, a negligible factor since $m^2\log(n\sqrt{\varepsilon}) = o(n)$. For Phase I we must assume the $n_{i,j}$ are chosen to maximize each h_i but since the entropy function h is continuous, the penalty shrinks with ε. Counting as before, we deduce that here the number of ways to accomplish Phase I is bounded by

$$
\begin{aligned}
&\exp\big(n\log m + n(H(\gamma^m) + o(\varepsilon^0)) + o(n)\big) \\
&= \exp\big(n\log m + n(H(\gamma) + o(\varepsilon^0)) + o(n)\big).
\end{aligned}
\tag{88}
$$

The computation for Phase II is exactly as before and the conclusion is that

$$
\begin{aligned}
|V| \leq \exp & \left(n \log m - n + n(H(\gamma) + o(\varepsilon^0)) + o(n) \right) \\
& \times \exp \left(n \log n - n - n \log m + o(n) \right) \\
= \exp & \left(n \log n - n + n H(\gamma) + o(n) \right)
\end{aligned}
\tag{89}
$$

proving the lemma in the case where $H(\gamma) > -\infty$.

If $H(\gamma) = -\infty$, we need only the upper bound provided by the set V. Fix $t < 0$ with the idea of showing that $\frac{1}{n} \log \frac{|B_n(\gamma, \varepsilon_\gamma)|}{n!} < t$. Define V as above, first insuring that m is large enough so that $H(\gamma^m) < t-1$. Then the number of ways to accomplish Phase I is bounded by

$$
\exp \left(n \log m + n(H(\gamma^m) + o(\varepsilon^0)) + o(n) \right) < \exp \left(n \log m + n(t-1 + o(\varepsilon^0)) + o(n) \right)
\tag{90}
$$

and consequently $|V|$ is bounded above by

$$
\exp \left(n \log n - n + n(t-1) + o(n) \right) < \exp \left(n \log n - n + nt \right) .
\tag{91}
$$
∎

We are finally in a position to prove Theorem 1. If our set Λ of permutons is closed, then, since Γ is compact, so is Λ. Let $\delta > 0$ with the idea of showing that

$$
\lim_{n \to \infty} \frac{1}{n} \log \frac{|\Lambda_n|}{n!} \leq H(\mu) + \delta
\tag{92}
$$

for some $\mu \in \Lambda$. If not, for each $\gamma \in \Lambda$ we may, on account of Lemma 20, choose ε_γ and n_γ so that $\frac{1}{n} \log \frac{|B_n(\gamma, \varepsilon_\gamma)|}{n!} < H(\gamma) + \delta/2$ for all $n \geq n_\gamma$. Since a finite number of these balls cover Λ, we have too few permutations in Λ_n for large enough n, and a contradiction has been reached.

If Λ is open, we again let $\delta > 0$, this time with the idea of showing that

$$
\lim_{n \to \infty} \frac{1}{n} \log \frac{|\Lambda_n|}{n!} \geq H(\mu) - \delta .
\tag{93}
$$

To do this we find a permuton $\mu \in \Lambda$ with

$$
H(\mu) > \sup_{\gamma \in \Lambda} H(\gamma) - \delta/2 ,
\tag{94}
$$

and choose $\varepsilon > 0$ and n_0 so that $B_n(\mu, \varepsilon) \subset \Lambda$ and $\frac{1}{n} \log \left(\frac{|B_n(\mu, \varepsilon)|}{n!} \right) > H(\mu) - \delta/2$ for $n \geq n_0$.

This concludes the proof of Theorem 1.

REFERENCES

1. M. Atapour and N. Madras, *Large deviations and ratio limit theorems for pattern-avoiding permutations*, Combin. Probab. Comput. **23** (2014), 161–200.

2. Z. Baranyai, *On the factorization of the complete uniform hypergraph*. Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdös on his 60th birthday). Colloq. Math. Soc. János Bolyai, North-Holland, Amsterdam, Vol. 10, 1975, pp. 91–108.

3. D. Bevan, *Growth rates of permutation grid classes, tours on graphs, and the spectral radius*, arXiv:1302.2037v4 (2013).

4. D. Bevan, *Permutations avoiding 1324 and patterns in Lukasiewicz paths*, arXiv:1406.2890v2 (2015).

5. N. Bhatnagar and R. Peled, *Lengths of monotone subsequences in a mallows permutation*, arXiv:1306.3674v2 (2014).

6. G. Brightwell and N. Georgiou, *Continuum limits for classical sequential growth models*, Random Structures Algorithm. **36** (2010), 218–250.

7. S. Chatterjee and P. Diaconis, *Estimating and understanding exponential random graph models*, Ann. Statist. **41** (2013), 2428–2461.

8. J.T. Chang and D. Pollard, *Conditioning as disintegration*, Stat. Neerl. **51** (1997), 287–317.

9. S. Chatterjee and S.R.S. Varadhan, *The large deviation principle for the Erdös-Rényi random graph*, Eur. J. Comb. **32** (2011), 1000–1017.

10. T. Dokos and I. Pak, *The expected shape of random doubly alternating Baxter permutations*, arXiv:1401.0770 (2014).

11. S. Elizalde and M. Noy, *The regions obtained by pairs of densities of patterns of length 3*, preprint (2015).

12. G.B. Folland, *Real analysis*, 2nd ed., John Wiley & Sons Inc., New York, 1999.

13. R. Glebov, A. Grzesik, T. Klimošová, and D. Král', *Finitely forcible graphons and permutons*, J. Comb. Theory B. **110** (2015), 112–135.

14. J. Hladký, A. Máthé, V. Patel, and O. Pikhurko, *Poset limits can be totally ordered*, arXiv:1211.2473v2 (2013).

15. C. Hoffman, D. Rizzolo, and E. Slivken, *Pattern avoiding permutations and Brownian excursion*, arXiv:1406.5156v2 (2015).

16. C. Hoppen, Y. Kohayakawa, C.G. Moreira, and R.M. Sampaio, *Limits of permutation sequences through permutation regularity*, arxiv:1106.1663v1 (2011).

17. C. Hoppen, Y. Kohayakawa, C.G. Moreira, B. Ráth, and R.M. Sampaio, *Limits of permutation sequences*, J. Comb. Theory B. **103** (2013), 93–113.

18. H. Huang, N. Linial, H. Naves, Y. Peled, and B. Sudakov, *On the densities of cliques and independent sets in graphs*, arXiv:1211.4532v2 (2013).

19. H. Huang, N. Linial, H. Naves, Y. Peled, and B. Sudakov, *On the 3-local profiles of graphs*, arXiv:1211.3106v2 (2013).

20. S. Janson, *Poset limits and exchangeable random posets*, Combinatorica. **31** (2011), 529–563.

21. S. Kitaev. Patterns in permutations and words, Springer, Berlin (2011).

22. D. Král', R. Kenyon, C. Radin, and P. Winkler, in preparation.

23. R. Kenyon, C. Radin, K. Ren, and L. Sadun, *Multipodal structure and phase transitions in large constrained graphs*, arXiv:1405.0599v2 (2014).

24. D. Král' and O. Pikhurko, *Quasirandom permutations are characterized by 4-point densities*, Geometr. Funct. Anal. **23** (2013), 570–579.

25. L. Lovász, *Large networks and graph limits*, Amer. Math. Soc., Providence, RI, 2012.

26. N. Madras and H. Liu, *Random pattern-avoiding permutations*. Algorithmic Probability and Combinatorics, Contemp. Math., Amer. Math. Soc., Providence, RI, Vol. 520, 2010, pp. 173–194.

27. N. Madras and L. Pehlivan, *Structure of random 312-avoiding permutations*, arXiv:1401.6230v2 (2014).

28. S. Minor and I. Pak, *The shape of random pattern-avoiding permutations*, Adv. Appl. Math. **55** (2014), 86–130.

29. S. Mukherjee, *Estimation in exponential families on permutations*, arxiv: 1307.0978v3 (2015).

30. C.B. Presutti and W.R. Stromquist, *Packing rates of measures and a conjecture for the packing density of 2413*, Lond. Math. Soc. Lecture Notes. **376** (2010), 3–40.

31. C. Radin and L. Sadun, *Phase transitions in a complex network*, J. Phys. A: Math. Theor. **46** (2013), 305002.

32. C. Radin and L. Sadun, *Singularities in the entropy of asymptotically large simple graphs*, J. Stat. Phys. **158** (2015), 853–865.

33. C. Radin, K. Ren, and L. Sadun, *The asymptotics of large constrained graphs*, J. Phys. A: Math. Theor. **47** (2014), 175001.

34. A. Razborov, *On the minimal density of triangles in graphs*, Combin. Probab. Comput. **17** (2008), 603–618.

35. A. Razborov, *Flag algebras*, J. Symb. Logic. **72** (2007), 1239–1282.

36. S. Starr, *Thermodynamic limit for the Mallows model on S_n*, J. Math. Phys. **50** (2009), 095208.

37. J. Trashorras, *Large deviations for symmetrised empirical measures*, Theor. Probab. **21** (2008), 397–412.