

Vertex order in some large constrained random graphs

Hans Koch ^{*}

Abstract. In large random graphs with fixed edge density and triangle density, it has been observed numerically [9] that a typical graph is finite-podal, meaning that it has only finitely many distinct “types” of vertices. In particular, it seems to be a fundamental property of such graphs to have large groups of vertices that are all of the same type. In this paper we describe a mechanism that produces such behavior. By known results on graph limits, the problem reduces to the study of a constrained maximization problem for symmetric measurable functions (graphons) on the unit square. As a first step we prove that, for a wide range of parameter values, the constrained maximizers are in some sense monotone.

1. Introduction and main results

We consider a variational problem on the space \mathcal{G} of all symmetric measurable functions on $[0, 1]^2$ that take values in $[0, 1]$. Such functions arise in the asymptotic analysis of large simple graphs, and in this context, they are also called (labeled) graphons.

Consider first simple graphs with vertex set $I_n = \{1/n, 2/n, \dots, n/n\}$. For simplicity we identify such a graph with its incidence matrix $g : I_n \times I_n \rightarrow \{0, 1\}$. The edge density of g is the number of edges divided by $\binom{n}{2}$, and the triangle density is the number of triangles divided by $\binom{n}{3}$. To leading order in n , these two densities are given by

$$E(g) = \iint g(z) dz, \quad T(g) = \iiint g(z)g^{*2}(z) dz, \quad (1.1)$$

where $g^{*2}(x, y) = \int g(x, s)g(s, y) ds$. Here the single integral is over I_n and the double integrals are over $I_n \times I_n$, using normalized counting measure on these sets.

Graphons can be obtained as limits of such graphs, as $n \rightarrow \infty$. For precise statements and results see [5]. Roughly speaking, the values of a graphon represent limits of averages for graphs. The edge density $E(g)$ and triangle density $T(g)$ of a graphon $g \in \mathcal{G}$ are defined as in (1.1). But in this case, and from now on, simple integrals are over the unit interval $I = [0, 1]$ and double integrals are over the unit square $I \times I$, using Lebesgue measure. When talking about these densities, a pair of real values (ϵ, τ) will be called accessible if $\epsilon = E(g)$ and $\tau = T(g)$ for some graphon g . The set of accessible pairs is known explicitly [3,5]. It consists of the region $\epsilon^3 \leq \tau \leq \epsilon^{3/2}$ and part of the region $\epsilon(2\epsilon - 1) \leq \tau \leq \epsilon^3$.

Given an accessible pair (ϵ, τ) , denote by $Z_{\epsilon, \tau, n, \delta}$ the number of simple graphs with n vertices, whose edge densities belong to $[\epsilon - \delta, \epsilon + \delta]$, and whose triangle densities belong to $[\tau - \delta, \tau + \delta]$. Then the limit $\lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} n^{-2} \ln Z_{\epsilon, \tau, n, \delta}$ exists and agrees with the maximum value of the entropy

$$S(g) = \frac{1}{2} \int S(g(z)) dz, \quad S(g) = -g \log(g) - (1 - g) \log(1 - g), \quad (1.2)$$

taken over all graphons $g \in \mathcal{G}$ that satisfy $E(g) = \epsilon$ and $T(g) = \tau$. This was proved among other things in [8], using ideas and results from [4,7]. Roughly speaking, if n is

^{*} Department of Mathematics, University of Texas at Austin, Austin, TX 78712

large then a typical simple n -vertex graph g_n with edge density $E(g_n) \simeq \epsilon$ and triangle density $T(g_n) \simeq \tau$ is close to a graphon g that maximizes the entropy (1.2) subject to the constraints $E(g) = \epsilon$ and $T(g) = \tau$. The notion of closeness used here implies e.g. that if H is the density function for some other subgraph, then $H(g_n) \simeq H(g)$.

Generalizations of the above-mentioned result can be found in [10]. In this paper we only consider constraints on the densities of edges and triangles.

Definition 1.1. *We say that a graphon g_0 is a constrained entropy-maximizer if $\mathcal{S}(g) \leq \mathcal{S}(g_0)$ for every graphon g that satisfies $E(g) = E(g_0)$ and $T(g) = T(g_0)$.*

The simplest constrained entropy-maximizers are those that maximize \mathcal{S} subject to a single constraint $E(g) = \epsilon$. These graphons are constant almost everywhere on $I \times I$, so $T(g) = \epsilon^3$ in this case. In what follows, we will omit qualifiers like “almost everywhere” when it is clear what null sets should be ignored. Other known constrained maximizers include those whose density pairs (ϵ, τ) lie on the boundary of the accessible region [6]. Each of these graphons is finite-podal, in the sense that it has only finitely many distinct types of vertices (outside some set of measure zero).

Definition 1.2. *The type of a “vertex” $y \in I$ for a graphon g is the function $g(\cdot, y)$.*

Numerical results in [9] suggest that every graphon g that maximizes the entropy \mathcal{S} subject to the constraints $E(g) = \epsilon$ and $T(g) = \tau$ is in fact finite-podal. Furthermore, if $\tau > \epsilon^3$ then the constrained maximizer appears to be two-podal (two distinct types of vertices). This has been proved recently in [11], in the case where $\tau > \epsilon^3$ is sufficiently close to ϵ^3 , for any given $\epsilon \in (0, 1)$ different from $1/2$.

A basic property of finite-podal graphons is that they have large groups of vertices of the same type. As a possible step toward a more global result, we describe here a mechanism that forces some constrained entropy-maximizers to have this property. We restrict to a case where we can show that a constrained entropy-maximizer does not take the values 0 or 1. This case is described in the following lemma. Let $\mathcal{I} = I \times I$.

Definition 1.3. *We say that $g \in \mathcal{G}$ and $h \in \mathcal{G}$ are positively (negatively) correlated if the product $[g(z) - g(z_0)][h(z) - h(z_0)]$ is nonnegative (nonpositive) for almost all $z, z_0 \in \mathcal{I}$.*

Denote by \mathcal{G}° the set of all functions $g \in \mathcal{G}$ with the property that $\text{ess inf } g > 0$ and $\text{ess sup } g < 1$. Then we have the

Lemma 1.4. *Let $g \in \mathcal{G}$ be a constrained entropy-maximizer with positive entropy. Assume that g and g^{*2} are not negatively correlated. Then g belongs to \mathcal{G}° .*

Some properties of constrained entropy-maximizers that are not covered by this lemma will be described later in Theorem 2.1.

Let now g be a constrained entropy-maximizer that belongs to \mathcal{G}° . Then by the method of Lagrange multipliers, there exist real numbers α and β such that

$$S'(g) = \alpha + \beta g^{*2}. \quad (1.3)$$

Notice that S' is decreasing: $S'(g) = \log((1-g)/g)$. Assuming that g is not constant, we must have $\beta \neq 0$. Thus g and g^{*2} are either positively correlated ($\beta < 0$) or negatively

correlated ($\beta > 0$). A simple argument given in Section 2 shows that $\beta \leq 0$ implies $\tau \geq \epsilon^3$. Presumably the converse is true as well, but we have no proof for this.

Definition 1.5. We say that $g \in \mathcal{G}$ is ordered if $g(x_0, y_0) \geq g(x_1, y_1)$ for almost all $(x_0, y_0) \in \mathcal{I}$ and $(x_1, y_1) \in \mathcal{I}$ that satisfy $x_0 \leq x_1$ and $y_0 \leq y_1$. We say that g can be ordered if there exists an ordered $g^* \in \mathcal{G}$ and a measure-preserving map $v : I \rightarrow I$ such that $g^*(v(x), v(y)) = g(x, y)$ for almost all $(x, y) \in \mathcal{I}$.

Notice that, if $g = g^* \circ (v \times v)$ as described above, then $E(g) = E(g^*)$ and $T(g) = T(g^*)$ and $\mathcal{S}(g) = \mathcal{S}(g^*)$. Our main results are the following.

Theorem 1.6. Let $g \in \mathcal{G}$ be a constrained entropy-maximizer with positive entropy. Assume that g and g^{*2} are not negatively correlated. Then g can be ordered.

This ordering property can be used to prove the

Theorem 1.7. Under the same hypotheses as in Theorem 1.6, there exists a set $J \subset I$ of positive measure such that $g(\cdot, y_1) = g(\cdot, y_2)$ whenever $y_1, y_2 \in J$. In particular, g is constant on $J \times J$.

The remaining part of this paper is organized as follows. In Section 2 we give a generalization of Lemma 1.4 and introduce some notation. Section 3 is concerned mainly with the regularity of graphons that satisfy the variational equation (1.3). In Section 5 we prove Theorem 1.6, using the results from Section 4 concerning constrained local maxima of the triangle density $T(g)$. Section 6 is devoted to the proof of Theorem 1.7.

2. Additional observations

The following is a generalization of Lemma 1.4.

Theorem 2.1. Let $g \in \mathcal{G}$ be a constrained entropy-maximizer with positive entropy. Suppose that g does not belong to \mathcal{G}° . Then $g(z) = 0$ or $g(z) = 1$ on some set of positive measure. Furthermore, there exists $0 < \gamma < 1$ such that for almost every $z \in \mathcal{I}$,

- (a) if $g(z) = 0$ then $g^{*2}(z) > \gamma$,
- (b) if $0 < g(z) < 1$ then $g^{*2}(z) = \gamma$,
- (c) if $g(z) = 1$ then $g^{*2}(z) < \gamma$.

Notice that a graphon g with the properties (a), (b), and (c) is negatively correlated with g^{*2} . Our guess is that there are no constrained entropy-maximizers with these properties, except for densities (ϵ, τ) on the boundary of the accessible region.

Before giving a proof of Theorem 2.1 we introduce some notation and show that positive correlation implies $\tau \geq \epsilon^3$.

By analogy with incidence matrices, a graphon g will sometimes be regarded as the kernel of an integral operator G . More generally, to a function $h \in L^2(\mathcal{I})$ we associate the Hilbert-Schmidt operator $H : f \mapsto \int h(\cdot, y)f(y) dy$ on $L^2(I)$. The integral kernel of a product GH will be denoted by $g \star h$. Notice that the function g^{*2} defined after (1.1) is simply $g \star g$. Notice also that $T(g) = \text{tr}(G^3)$.

To come back to a claim made after (1.3), assume that $g \in \mathcal{G}$ and g^{*2} are positively correlated. The claim is that the densities $\epsilon = E(g)$ and $\tau = T(g)$ satisfy $\tau \geq \epsilon^3$. Indeed, using that $E(g) = \langle g, 1 \rangle \leq \langle g, g \rangle^{1/2}$ for the inner product in $L^2(\mathcal{I})$, we obtain

$$\begin{aligned} T(g) - E(g)^3 &\geq T(g) - \langle g, 1 \rangle \langle g, g \rangle \\ &= \frac{1}{2} \iiint [g(z) - g(z_0)] [g^{*2}(z) - g^{*2}(z_0)] dz dz_0 \geq 0. \end{aligned} \quad (2.1)$$

Finally, let us mention that (1.3) is the equation $D\mathcal{F}_{\alpha,\beta}(g) = 0$ for a critical point of the “free energy”

$$\mathcal{F}_{\alpha,\beta}(g) = \mathcal{S}(g) - \frac{1}{2}\alpha E(g) - \frac{1}{6}\beta T(g). \quad (2.2)$$

The critical points of interest are in general not maximizers of this free energy: it is not hard to show that all maximizers of $\mathcal{F}_{\alpha,\beta}$ for $\beta \leq 0$ are constant [7].

Definition 2.2. A set $B \subset \mathcal{I}$ is said to be symmetric if it is invariant under the reflection $(x, y) \mapsto (y, x)$. Two sets $B_0 \subset \mathcal{I}$ and $B_1 \subset \mathcal{I}$ will be called projection-disjoint if no vertical line $\{x\} \times I$ and no horizontal line $I \times \{y\}$ intersects both B_0 and B_1 .

Proof of Theorem 2.1. Let $g \in \mathcal{G}$ be a constrained entropy-maximizer. For $\varepsilon \geq 0$ define $\mathcal{I}_\varepsilon = \{z \in \mathcal{I} : \varepsilon < g(z) < 1 - \varepsilon\}$. Assuming $\mathcal{S}(g) > 0$, the set \mathcal{I}_ε has positive measure for $\varepsilon > 0$ sufficiently small. In what follows, we always assume that $0 < \varepsilon < 1/2$, and we identify functions on \mathcal{I}_ε with functions on \mathcal{I} that vanish outside \mathcal{I}_ε . For functions $h \in L^\infty(\mathcal{I}_\varepsilon)$ of norm $\|h\|_\infty < \varepsilon$, and for $F \in \{E, T, \mathcal{S}\}$, define $F_\varepsilon(h) = F(g + h) - F(g)$. Notice that F_ε is smooth near the origin, as a function from $L^\infty(\mathcal{I}_\varepsilon)$ to the real numbers. By assumption, $h = 0$ maximizes \mathcal{S}_ε subject to the constraints $E_\varepsilon(h) = T_\varepsilon(h) = 0$.

Suppose that g does not belong to \mathcal{G}° . In other words, $|\mathcal{I}_\varepsilon| < 1$ for all $\varepsilon > 0$.

Assume for contradiction that the derivatives $DE_\varepsilon(0)$ and $DT_\varepsilon(0)$ are linearly independent, for some $\varepsilon > 0$ and thus for all $\varepsilon > 0$ sufficiently small. Then there exists real numbers α and β (Lagrange multipliers) such that $S'(g) = \alpha + \beta g^{*2}$ almost everywhere on \mathcal{I}_ε . This holds for $\varepsilon > 0$ sufficiently small, with α and β independent of ε . If the set \mathcal{I}_0 has measure $|\mathcal{I}_0| = 1$ then $S'(g) = \alpha + \beta g^{*2}$ almost everywhere on \mathcal{I} . Given that $S'(r) \rightarrow +\infty$ as $r \searrow 0$ and $S'(r) \rightarrow -\infty$ as $r \nearrow 1$, this is possible only if \mathcal{I}_ε has measure 1 for some $\varepsilon > 0$. So we may assume that either $Z_0 = \{z \in \mathcal{I} : g(z) = 0\}$ or $Z_1 = \{z \in \mathcal{I} : g(z) = 1\}$ has positive measure.

Consider first the case $|Z_0| > 0$. Consider a small perturbation $g + sf$ of g , where f is the indicator function of Z_0 . Pick $\varepsilon > 0$ such that $DE_\varepsilon(0)$ and $DT_\varepsilon(0)$ are linearly independent. If $s > 0$ is sufficiently small, then there exists $h_s \in L^\infty(\mathcal{I}_\varepsilon)$ such that $E_\varepsilon(sf + h_s) = 0$ and $T_\varepsilon(sf + h_s) = 0$. This follows from the implicit function theorem. In fact, we get $\|h_s\|_\infty = \mathcal{O}(s)$ as $s \searrow 0$. Now notice that $\mathcal{S}(g + sf + h_s) = \mathcal{S}(g + sf) + \mathcal{S}_\varepsilon(h_s)$ and that

$$\mathcal{S}(g + sf) - \mathcal{S}(g) \sim s \ln(1/s), \quad (2.3)$$

where $u(s) \sim v(s)$ means that $u(s)/v(s)$ converges to a positive constant as $s \searrow 0$. Thus, $\mathcal{S}(g + sf + h_s) > 0$ for $s > 0$ sufficiently small, contradicting the assumption that g is

a constrained local entropy-maximizer. A similar contradiction is obtained in the case $|Z_1| > 0$. So the derivatives $DE_\varepsilon(0)$ and $DT_\varepsilon(0)$ must be linearly dependent for all $\varepsilon > 0$.

At this point we have established that the function g^{*2} is constant a.e. on \mathcal{I}_0 . If $|\mathcal{I}_0| = 1$ then g is constant a.e. as well and we are done. So we may assume that $|Z_0| > 0$ or $|Z_1| > 0$. Our goal is to use again a perturbation sf satisfying (2.3), and to compensate for the resulting change in T with another perturbation th .

Assume for contradiction that there is no positive $\gamma < 1$ for which (a) and (b) and (c) hold. Then g and g^{*2} are not negatively correlated. Thus, restricting to Lebesgue points for g , we can find symmetric sets $\mathcal{J}_0 \subset \mathcal{I}$ and $\mathcal{J}_1 \subset \mathcal{I}$ of positive measure such that $g(z_0) < g(z_1)$ and $g^{*2}(z_0) < g^{*2}(z_1)$ for all $z_0 \in \mathcal{J}_0$ and $z_1 \in \mathcal{J}_1$. Clearly $\mathcal{J}_0 \cap Z_1$ and $\mathcal{J}_1 \cap Z_0$ have zero measure. And since g^{*2} is constant in \mathcal{I}_0 , either $\mathcal{J}_0 \cap Z_0$ or $\mathcal{J}_1 \cap Z_1$ has positive measure. Consider first the case where $|\mathcal{J}_0 \cap Z_0| > 0$. If $\varepsilon > 0$ is sufficiently small then we can find a symmetric set $A_1 \subset \mathcal{J}_1$ of positive measure such that $g > \varepsilon$ on A_1 . In addition we choose symmetric sets $B \subset \mathcal{I}_0$ and $A_0 \subset \mathcal{J}_0 \cap Z_0$ of positive measure. Clearly the sets B , A_0 , and A_1 can be chosen in such a way that B and $A_0 \cup A_1$ are projection-disjoint. Now define $f = |A_1|f_0 - |A_0|f_1$, where f_0 and f_1 are the indicator function of A_0 and A_1 , respectively. Since $g^{*2}(z_0) < g^{*2}(z_1)$ for all $z_0 \in A_0$ and all $z_1 \in A_1$ we have

$$T(g + sf) = T(g) + 3bs + \mathcal{O}(s^2), \quad b = \langle g^{*2}, f \rangle < 0, \quad (2.4)$$

for $s > 0$ sufficiently small. Notice that sf has average zero and satisfies (2.3).

Our next goal is to undo the decrease in T with another perturbation th . To this end, we choose symmetric sets $B_0 \subset B$ and $B_1 \subset B$ that are projection-disjoint and have positive measure. Let h_0 and h_1 be the indicator function of B_0 and B_1 , respectively. Define $h = |B_1|h_0 - |B_0|h_1$ if this makes $D\mathcal{S}_0(g)h \geq 0$, or else define $h = |B_0|h_1 - |B_1|h_0$. Since B_0 and B_1 are projection-disjoint, we have $h_0 \star h_1 = h_1 \star h_0 = 0$. Furthermore, $\langle g^{*2}, h \rangle = 0$ since g^{*2} is constant on B . Thus, $T(g + th) - T(g) = 3at^2 + \mathcal{O}(t^3)$ with

$$a = \langle g, h^{*2} \rangle = |B_1|^2 \langle g, h_0^{*2} \rangle + |B_0|^2 \langle g, h_1^{*2} \rangle > 0. \quad (2.5)$$

Using that $t \mapsto \mathcal{S}(g + th)$ is analytic near $t = 0$, there exists $c > 0$ such that

$$\mathcal{S}(g + th) - \mathcal{S}(g) = tD\mathcal{S}_0(g)h + \frac{1}{2}t^2D^2\mathcal{S}_0(g)h^2 + \mathcal{O}(t^3) \geq -ct^2 \quad (2.6)$$

for $t > 0$ sufficiently small.

Notice that f and h both have average zero, so that $E(g + sf + th) = E(g)$. Furthermore, since $f \star h = h \star f = 0$ we have

$$T(g + sf + th) = T(g) + 3bs + 3at^2 + \mathcal{O}(s^2) + \mathcal{O}(t^3), \quad (2.7)$$

for $s, t > 0$ with $s + t$ small. Recall that $b < 0 < a$. So $T(g + th + sf) = T(g)$ along some continuous curve $s = \sigma(t) = |b|^{-1}at^2 + \mathcal{O}(t^3)$, for $t > 0$ sufficiently small. But $\mathcal{S}(g + sf + th) \geq \mathcal{S}(g + sf) - ct^2$ and $\mathcal{S}(g + sf)$ satisfies (2.3). Thus $\mathcal{S}(g + \sigma(t)f + th) > \mathcal{S}(g)$ for $t > 0$ sufficiently small, contradicting the assumption that g is a constrained local maximizer of \mathcal{S} . A similar contradiction is obtained in the case $|Z_1| > 0$. This concludes our proof of Theorem 2.1. QED

3. Regularity

In this section we describe regularity properties of the function $g^{\star 2}$ associated with a graphon $g \in \mathcal{G}$. Later we will restrict to functions g that satisfy the relation (1.3) for $\beta \neq 0$. Then g will have the same regularity as $g^{\star 2}$.

In what follows, g and h are fixed but arbitrary functions in $L^\infty(I)$. Let Y be the set of all points $y \in I$ for which $x \mapsto h(x, y)$ is measurable. By Fubini's theorem Y is measurable and $|Y| = 1$.

Proposition 3.1. *Let $f = g \star h$. Then there exist measurable sets $J_1 \subset J_2 \subset \dots \subset I$ with $|I \setminus J_m| \rightarrow 0$ as $m \rightarrow \infty$, such that the following holds for each m . The function $f(\cdot, y)$ is uniformly continuous when restricted to J_m , uniformly in y for $y \in Y$. Each point in J_m is a Lebesgue density point for J_m and a Lebesgue point for $f(\cdot, y)$ if $y \in Y$.*

Proof. We may assume that $\|g\|_\infty \leq 1$ and $\|h\|_\infty \leq 1$.

Let n be a fixed but arbitrary positive integer. By Lusin's theorem there exists a closed set $F \subset I$ such that the restriction of g to F is continuous, and such that $U = I \setminus F$ has measure $|U| \leq 4^{-n}$. Since F is compact, the restriction of g to F is uniformly continuous. Denote by ν_n the modulus of continuity of this restriction. Define $I_n = \{x \in I : |U_x| < 2^{-n}\}$ where $U_x = \{y \in Y : (x, y) \in U\}$. For every $x \in I \setminus I_n$ we have $|U_x| \geq 2^{-n}$, and thus $|I \setminus I_n| \leq 2^{-n}$. Consequently $|I_n| > 1 - 2^{-n}$. Let $F_x = \{y \in Y : (x, y) \in F\}$. Then for every $x, x_0 \in I_n$ and every $y \in Y$ we have

$$\begin{aligned} |f(x, y) - f(x_0, y)| &\leq \int |g(x, z) - g(x_0, z)| |h(z, y)| dz \\ &\leq |U_x \cup U_{x_0}| + \int_{F_x \cap F_{x_0}} |g(x, z) - g(x_0, z)| |h(z, y)| dz \\ &\leq 2^{1-n} + \nu_n(|x - x_0|). \end{aligned} \quad (3.1)$$

Denote by Z_m the set of points in $\bigcap_{n>m} I_n$ that are not Lebesgue density points for this set, and let $Z = \bigcup_m Z_m$. Setting $X = Y \setminus Z$ and $X_n = I_n \cap X$ define

$$J_m = \bigcap_{n>m} X_n, \quad \omega_m(r) = \inf_{n>m} [2^{1-n} + \nu_n(r)]. \quad (3.2)$$

Then every point in J_m is a Lebesgue density point for J_m . Furthermore $|J_m| \rightarrow 1$ as $m \rightarrow \infty$. By (3.1) we have

$$|f(x, y) - f(x_0, y)| \leq \omega_m(|x - x_0|), \quad x, x_0 \in J_m, \quad y \in Y. \quad (3.3)$$

Notice also that $\omega_m(r) \rightarrow 0$ as $r \rightarrow 0$. Thus $x \mapsto f(x, y)$ is uniformly continuous on J_m , uniformly in y for $y \in Y$.

Consider the function $f_0(x) = f(x, y)$ with $y \in Y$ fixed. Let $m \geq 1$ and $x_0 \in J_m$. Let $\gamma = f_0(x_0)$ and $\varepsilon > 0$. Since f_0 is continuous on J_m we have $f_0 \geq \gamma - \varepsilon$ on $Q \cap J_m$ for every sufficiently small interval Q centered at x_0 . Since x_0 is a Lebesgue density point for

J_m we have $|Q \cap J_m|/|Q| \rightarrow 1$ as $|Q| \rightarrow 0$. Thus the average of f_0 on Q is larger than $\gamma - 2\varepsilon$ if $|Q|$ is sufficiently small. This shows that x_0 is a Lebesgue point for f_0 . **QED**

Consider now the special case $h = g$. From (3.3) we obtain

$$|g^{*2}(x, y) - g^{*2}(x_0, y_0)| \leq \omega_m(|x - x_0|) + \omega_m(|y - y_0|), \quad x, y, x_0, y_0 \in J_m. \quad (3.4)$$

In other words, the restriction of g^{*2} to $J_m \times J_m$ is uniformly continuous, for each m .

With the sets X_n and J_m as described in the above proof, define

$$\tilde{I} = \liminf_n X_n = \bigcup_m J_m, \quad \tilde{\mathcal{I}} = \tilde{I} \times \tilde{I}. \quad (3.5)$$

Notice that $|I \setminus J_m| \leq 2^{-m}$, and that $I \setminus \tilde{I}$ has measure zero. Furthermore, every point in \tilde{I} is a Lebesgue density point for \tilde{I} .

Corollary 3.2. *Every point in $\tilde{\mathcal{I}}$ is a Lebesgue point for g^{*2} .*

Proof. Since every point in J_m is a Lebesgue density point for J_m , every point in $\mathcal{J}_m = J_m \times J_m$ is a Lebesgue density point for \mathcal{J}_m , and every point in $\tilde{\mathcal{I}}$ is a Lebesgue density point for $\tilde{\mathcal{I}}$.

Let $x_0, y_0 \in \tilde{I}$. Then (x_0, y_0) belongs to \mathcal{J}_m for sufficiently large m . Let $\gamma = g^{*2}(x_0, y_0)$ and $\varepsilon > 0$. Since g^{*2} is continuous on \mathcal{J}_m , we have $g^{*2} \geq \gamma - \varepsilon$ on $Q \cap \mathcal{J}_m$, for every sufficiently small square Q centered at (x_0, y_0) . Since (x_0, y_0) is a Lebesgue density point for \mathcal{J}_m we have $|Q \cap \mathcal{J}_m|/|Q| \rightarrow 1$ as $|Q| \rightarrow 0$. Thus the average of g^{*2} on Q is larger than $\gamma - 2\varepsilon$ if $|Q|$ is sufficiently small. Similarly, the average of g^{*2} on Q is smaller than $\gamma + 2\varepsilon$ if $|Q|$ is sufficiently small. This shows that (x_0, y_0) is a Lebesgue point for g^{*2} . **QED**

Corollary 3.3. *If g satisfies (1.3) then every point in $\tilde{\mathcal{I}}$ is a Lebesgue point for g .*

Assume now that $g \in \mathcal{G}^\circ$ is ordered. For $0 < x < 1$ and $y \in \tilde{I}$ define the limits $g(x\pm, y) = \lim_{\varepsilon \downarrow 0} g(x \pm \varepsilon, y)$ along points $x \pm \varepsilon$ in \tilde{I} . To simplify notation let $g(x, y) = g(0+, y)$ for $x \leq 0$ and $g(x, y) = g(1-, y)$ for $x \geq 1$. Define $\delta_1 g(x, y) = g(x+, y) - g(x-, y)$. Similarly define $\delta_1 g^{*2}(x, y) = g^{*2}(x+, y) - g^{*2}(x-, y)$. Notice that $\delta_1 g \leq 0$. By monotone convergence we have

$$\delta_1 g^{*2}(x, y) = \int \delta_1 g(x, z) g(z, y) dz, \quad x \in I, \quad y \in \tilde{I}. \quad (3.6)$$

Assume also that g satisfies (1.3) with $\beta \neq 0$. Using (3.6) we see that if $x \mapsto g(x, y)$ is continuous at x_0 for some $y \in \tilde{I}$, then $x \mapsto g(x, y)$ is continuous at x_0 for every $y \in \tilde{I}$. This proves the following

Proposition 3.4. *Assume that $g \in \mathcal{G}^\circ$ is ordered and satisfies (1.3). Then g can be modified on a set of measure zero in such a way that the following holds. There exists a countable set $Z \subset (0, 1)$ such that for every $y \in I \setminus Z$, the function $x \mapsto g(x, y)$ is continuous on I , except for jump discontinuities at points $x \in Z$.*

4. Local extrema

First we relate the constrained maximization problem for \mathcal{S} to a constrained maximization problem for T . This is useful for the following reason: if w is a measure-preserving isomorphism of \mathcal{I} , then $E(g \circ w) = E(g)$ and $\mathcal{S}(g \circ w) = \mathcal{S}(g)$. Lemma 3.2 below describes a situation where $T(g \circ w) > T(g)$.

Given any nonnegative real number $r < 1/2$, denote by \mathcal{G}^r the set of all functions $g \in \mathcal{G}$ satisfying $r < g < 1 - r$ almost everywhere on \mathcal{I} . If F is a real-valued function defined on a domain $\mathcal{D} \subset \mathcal{G}^r$, we say that $g_0 \in \mathcal{D}$ is a L^p -local maximizer of F on \mathcal{D} if there exists $\delta > 0$ such that $F(g_0) \geq F(g)$ whenever $g \in \mathcal{D}$ and $\|g - g_0\|_p < \delta$. Here $\|\cdot\|_p$ denotes the norm in $L^p(\mathcal{I})$. Define $\mathcal{G}_\epsilon^r = \{g \in \mathcal{G}^r : E(g) = \epsilon\}$ and

$$\Theta_{\epsilon,\tau}^r = \{g \in \mathcal{G}_\epsilon^r : T(g) = \tau\}, \quad \Sigma_{\epsilon,\sigma}^r = \{g \in \mathcal{G}_\epsilon^r : \mathcal{S}(g) = \sigma\}. \quad (4.1)$$

Proposition 4.1. *Let $0 < s < r < \epsilon < 1 - r$. Let g be a non-constant function in \mathcal{G}_ϵ^r . Set $\tau = T(g)$ and $\sigma = \mathcal{S}(g)$. Assume that g is a L^2 -local maximizer of \mathcal{S} on $\Theta_{\epsilon,\tau}^s$ with $\beta < 0$. Then g is a L^2 -local maximizer of T on $\Sigma_{\epsilon,\sigma}^r$.*

Proof. Under the given hypotheses we can find a function $h \in L^\infty(\mathcal{I})$ with $E(h) = 0$ such that $DS(g)h > 0$. By the variational equation (1.3) we also have $DT(g)h < 0$.

Assume for contradiction that g is not a L^2 -local maximizer of T on $\Sigma_{\epsilon,\sigma}^r$. Then there exist functions $g_1, g_2, \dots \in \Sigma_{\epsilon,\sigma}^r$ with $g_n \rightarrow g$ in the L^2 sense, such that $\tau_n = T(g_n)$ is larger than τ for all n . For $t \neq 0$ near zero we have

$$\mathcal{S}(g_n + th) - \sigma = t[DS(g_n)h + \mathcal{O}(t)], \quad (4.2)$$

with an $\mathcal{O}(t)$ bound that is uniform in n . So there exist n' and $t' > 0$ such that $\mathcal{S}(g_n + th) > \sigma$ whenever $n \geq n'$ and $0 < t \leq t'$. We also have

$$T(g_n + th) - \tau = (\tau_n - \tau) + t[DT(g_n)h + \mathcal{O}(t)], \quad (4.3)$$

with an $\mathcal{O}(t)$ bound that is uniform in n . By choosing $t' > 0$ sufficiently small we can make the term $[\dots]$ in this equation negative and bounded away from zero, for all $n \geq n'$ and $0 < t \leq t'$. Thus, if n is sufficiently large, then the right hand side of (4.3) is negative for $t = t'$. Given that both sides are positive for $t = 0$ it follows that for n sufficiently large, there exists $0 < t_n < t'$ such that $g_n + t_n h$ belongs to $\Theta_{\epsilon,\tau}^s$. At the same time $\mathcal{S}(g_n + t_n h) > \sigma$, as described above. This contradicts the assumption that g is a L^2 -local maximizer of \mathcal{S} on $\Theta_{\epsilon,\tau}^s$. **QED**

The next lemma plays a crucial role in our proof of Theorem 1.6. To make it more transparent, we first formulate an analogue for matrices. Since this matrix version will not be used we state it here without proof:

Let A be a symmetric square matrix with entries $A_{ij} \geq 0$. Pick two distinct rows of A , say A_k and A_n . To simplify the description assume that $A_{ij} = 0$ for $i, j \in \{k, n\}$. Replace A_k by the pointwise $\max(A_k, A_n)$ and A_n by the pointwise $\min(A_k, A_n)$. Similarly

for the columns $A_{.k}$ and $A_{.n}$. Unless the resulting matrix agrees with A , this operation increases $\text{tr}(A^3)$.

Let now x_0 and x_1 be distinct points in $(0, 1)$, and let δ_0 be half the smallest distance between any two distinct points in $\{0, x_0, x_1, 1\}$. Given any positive $\delta < \delta_0$ define $J_0 = [x_0 - \delta, x_0 + \delta]$ and $J_1 = [x_1 - \delta, x_1 + \delta]$ and $J = J_0 \cup J_1$. Then define $u : I \rightarrow I$ by setting $u(x) = x - x_0 + x_1$ if $x \in J_0$, and $u(x) = x - x_1 + x_0$ if $x \in J_1$, and $u(x) = x$ if $x \in J^c$. To simplify notation we write $u(x) = x'$. Let $g \in \mathcal{G}$ and define

$$f_\delta(x) = \iint_{P_x \times N_x} g(y, z) |g(x', y) - g(x, y)| |g(x', z) - g(x, z)| dydz, \quad (4.4)$$

for $x \in J$, where where

$$\begin{aligned} P_x &= \{y \in J^c : (x' - x)[g(x', y) - g(x, y)] > 0\}, \\ N_x &= \{z \in J^c : (x' - x)[g(x', z) - g(x, z)] < 0\}. \end{aligned} \quad (4.5)$$

Notice that $f_\delta(x)$ is a decreasing function of δ .

Lemma 4.2. *Assume that there exists a positive $\delta < \delta_0$ such that x_0 is a Lebesgue point for f_δ and $f_\delta(x_0) > 0$. Then $T(g \circ w) > T(g)$, where $w : \mathcal{I} \rightarrow \mathcal{I}$ is defined by*

$$w(x, y) = \begin{cases} (x', y) & \text{if } x \in J \text{ and } y \in P_x, \\ (x, y') & \text{if } y \in J \text{ and } x \in P_y, \\ (x, y) & \text{otherwise.} \end{cases} \quad (4.6)$$

Proof. Let $g_w = g \circ w$ and $h = g_w - g$. Denote by G , G_w , and H the integral operators with nuclei g , g_w , and h , respectively, as described in Section 2. Since h is supported in $(J \times J^c) \cup (J^c \times J)$, we have $\text{tr}(H^3) = 0$. Thus

$$\text{tr}(G_w^3) - \text{tr}(G^3) = 3\text{tr}(G^2H) + 3\text{tr}(GH^2). \quad (4.7)$$

For any measurable $F : J \rightarrow \mathbb{R}$ denote by \bar{F} the average value of F . Then $\text{tr}(G^2H) = 4\delta\bar{F}_1$, where

$$\begin{aligned} F_1(x) &= \int_{J^c} dy [g^{*2}(x, y)h(x, y) + g^{*2}(x', y)h(x', y)] \\ &= \int_{J^c} dy \int_{J^c} dz [g(x, z)g(z, y)h(x, y) + g(x', z)g(z, y)h(x', y)] + \mathcal{O}(\delta^2) \\ &= \int_{J^c} dy \int_{J^c} dz [g(x, z) - g(x', z)]g(z, y)h(x, y) + \mathcal{O}(\delta^2) \\ &= \int_{J^c} dy \int_{J^c} dz [g_w(x, z) - g(x', z)]g(z, y)h(x, y) \\ &\quad - \int_{J^c} dy \int_{J^c} dz h(x, z)g(z, y)h(x, y) + \mathcal{O}(\delta^2), \end{aligned} \quad (4.8)$$

with an $\mathcal{O}(\delta^2)$ bound that is independent of x . In the third equality we have used that $h(x, y) + h(x', y) = 0$. Similarly we have $\text{tr}(G^2 H) = 4\delta \overline{F_2}$, where

$$F_2(x) = \int_{J^c} dy \int_{J^c} dx g(z, y) h(z, x) h(x, y) + \mathcal{O}(\delta^2). \quad (4.9)$$

Thus, $F_1 + F_2 = F + \mathcal{O}(\delta^2)$, with

$$\begin{aligned} F(x) &= \int_{J^c} dy \int_{J^c} dz [g_w(x, z) - g(x', z)] g(z, y) h(x, y) \\ &= \int_{J^c} dy \int_{N_x} dz [g(x, z) - g(x', z)] g(z, y) h(x, y) \\ &= \int_{P_x} dy \int_{N_x} dz [g(x, z) - g(x', z)] g(z, y) [g(x, y') - g(x, y)] = f_\delta(x). \end{aligned} \quad (4.10)$$

Putting it all together,

$$\text{tr}(G_w^3) - \text{tr}(G^3) = 12\delta [\overline{f_\delta} + \mathcal{O}(\delta)]. \quad (4.11)$$

Under the given assumption, $\overline{f_\delta}$ has a positive limit as $\delta \rightarrow 0$. Thus, the right hand side of (4.11) is positive for sufficiently small $\delta > 0$. QED

5. Ordering

The main goal in this section is to give a proof of Theorem 1.6.

Proposition 5.1. *Let $g \in \mathcal{G}^\circ$ and define $Y(x_0, x_1) = \{y \in I : g(x_0, y) < g(x_1, y)\}$. Assume that either $Y(x_0, x_1)$ or $Y(x_1, x_0)$ has measure zero, for any two points $x_0, x_1 \in \tilde{I}$. Then g can be ordered.*

Proof. Consider $g \in \mathcal{G}^\circ$ satisfying the given assumption. Define

$$f(x) = \int g(x, y) dy. \quad (5.1)$$

It is straightforward to check that $Y(x_0, x_1)$ has measure zero if and only if $f(x_0) \geq f(x_1)$. Define $x_0 \preceq x_1$ to mean that either $f(x_0) > f(x_1)$, or else $f(x_0) = f(x_1)$ and $x_0 \leq x_1$. Then (I, \preceq) is totally ordered. Clearly f is decreasing from (I, \preceq) to (I, \leq) . Define

$$v(x) = |\{x' \in I : x' \preceq x\}|. \quad (5.2)$$

Furthermore, if A is a measurable set of real numbers of finite measure $a = |A|$, we define $RA = [0, a)$.

As was shown in [1], the function $v : I \rightarrow I$ is measure preserving, and $(Rf) \circ v = f$, where Rf denotes the decreasing rearrangement of f . Recall that the function f and Rf have the following level set decomposition

$$f(x) = \int_0^\infty \chi_{\{f>t\}}(x) dt, \quad (Rf)(x) = \int_0^\infty \chi_{R\{f>t\}}(x) dt. \quad (5.3)$$

Here $\{f > t\}$ denotes the set of all $x \in I$ such that $f(x) > t$.

Our goal is to obtain a rearrangement $\mathcal{R}g$ of the function g such that

$$(\mathcal{R}g)(v(x), v(y)) = g(x, y). \quad (5.4)$$

This identity also implies that $\mathcal{R}g$ is symmetric almost everywhere, since the range of v has full measure.

Consider first some arbitrary nonnegative function $F \in L^\infty(\mathcal{I})$. The corresponding restrictions $F(\cdot, y)$ are measurable on I , for almost every $y \in I$. Thus we can define a rearrangement R_1F of F by setting $(R_1F)(\cdot, y) = RF(\cdot, y)$. An analogous rearrangement R_2F is defined by setting $(R_2F)(x, \cdot) = RF(x, \cdot)$. Specifically we have

$$(R_jF)(x, y) = \int \chi_{R_j\{F>t\}}(x, y) dt, \quad j = 1, 2, \quad (5.5)$$

where

$$\begin{aligned} R_1A &= \{(x, y) \in \mathcal{I} : 0 \leq x < |A_{\cdot, y}|\}, & A_{\cdot, y} &= \{x \in I : (x, y) \in A\}, \\ R_2A &= \{(x, y) \in \mathcal{I} : 0 \leq y < |A_{x, \cdot}|\}, & A_{x, \cdot} &= \{y \in I : (x, y) \in A\}, \end{aligned}$$

for any set $A \subset \mathcal{I}$. By Fubini's theorem, if A is measurable then so are R_1A and R_2A . By the uniqueness of the level set decomposition, we have $\{R_jF > t\} = R_j\{F > t\}$ for all t . This shows e.g. that F and R_jF are equimeasurable.

Consider again the function g and define $\mathcal{R}g = R_2R_1g$. Clearly g and $\mathcal{R}g$ are equimeasurable. Let $g_{\cdot, y} = g(\cdot, y)$. For almost every $y \in I$ there exists a measure preserving function $v_y : I \rightarrow I$ such that $(Rg_{\cdot, y}) \circ v_y = g_{\cdot, y}$. In fact, using that $g_{\cdot, y}$ is decreasing almost everywhere on I , as a function from (I, \preceq) to (I, \leq) , we can take $v_y = v$ to be independent of y . The same applies to the functions $g_{x, \cdot} = g(x, \cdot)$. This shows that (5.4) holds for almost every $x, y \in I$. QED

Assume now that $g \in \mathcal{G}^\circ$ satisfies the equation (1.3) with $\beta \neq 0$. As shown in Section 3, there exists a set $\tilde{I} \subset I$ of measure 1 and an increasing sequence of measurable sets $J_m \nearrow \tilde{I}$ such that the restriction of g to each $J_m \times J_m$ is uniformly continuous. We may assume that every point of J_m is a Lebesgue density point for J_m .

Proposition 5.2. *Assume that $g \in \mathcal{G}^\circ$ satisfies (1.3) with $\beta \neq 0$. Then g is ordered if and only if for every $y \in \tilde{I}$ the function $x \mapsto g(x, y)$ is decreasing on \tilde{I} .*

Proof. The ‘‘if’’ part is obvious. To prove the ‘‘only if’’ part, assume that g is ordered. Let $y \in \tilde{I}$. Then there exists $M > 0$ such that $y \in J_m$ for all $m \geq M$. By continuity, g is

ordered on each of the sets $J_m \times J_m$. Thus $x \mapsto g(x, y)$ is decreasing on J_m for all $m \geq M$. This implies that $x \mapsto g(x, y)$ is decreasing on \tilde{I} . QED

Lemma 5.3. *Assume that $g \in \mathcal{G}^\circ$ satisfies (1.3) with $\beta < 0$. Assume that g cannot be ordered. Then for every $\varepsilon > 0$ there exists a measure-preserving bijection $w : \mathcal{I} \rightarrow \mathcal{I}$ such that $\|g \circ w - g\|_2 < \varepsilon$ and $T(g \circ w) > T(g)$.*

Proof. Under the given assumptions, Proposition 5.1 implies that there exists points $x_0, x_1 \in \tilde{I}$ and sets $Y, Z \subset I$ of positive measure such that $g(x_0, y) < g(x_1, y)$ for all $y \in Y$ and $g(x_0, z) > g(x_1, z)$ for all $z \in Z$. Since $J_m \nearrow \tilde{I}$ we have $x_0, x_1 \in J_m$ for large m . And since \tilde{I} has full measure, choosing m sufficiently large guarantees that $Y' = Y \cap J_m$ and $Z' = Z \cap J_m$ have positive measure.

We may assume that x_0 and x_1 belong to the open interval $(0, 1)$. Given any $\delta > 0$ define $J_0 = [x_0 - \delta, x_0 + \delta]$ and $J_1 = [x_1 - \delta, x_1 + \delta]$ and $J = J_0 \cup J_1$. We now choose $\delta > 0$ sufficiently small such that J_0 and J_1 are disjoint subsets of I , and such that $Y'' = Y' \cap J^c$ and $Z'' = Z' \cap J^c$ have positive measure, where $J^c = I \setminus J$.

Consider the sets P_x and N_x defined in (4.5), with $u : x \mapsto x'$ as described before (4.4). If $x = x_0$ then one of these sets includes Y'' and the other includes Z'' . So in this case E_x and N_x both have positive measure. Since g is uniformly continuous on $J_m \times J_m$, the same holds for every $x \in J_m$ that is sufficiently close (but not necessarily equal) to x_0 . And since x_0 is a Lebesgue density points for J_m , this applies to a set of points $x \in J_m$ that has positive measure. Thus x_0 is a Lebesgue point for f_δ and $f_\delta(x_0) > 0$. The claim now follows from Lemma 4.2. QED

Proof of Theorem 1.6. Let $g \in \mathcal{G}$ be a constrained entropy-maximizer with positive entropy $\sigma = \mathcal{S}(g)$. Assume that g and g^{*2} are not negatively correlated. Then g belongs to \mathcal{G}° by Lemma 1.4. Furthermore, as explained after Lemma 1.4, g satisfies the equation (1.3) with $\beta < 0$. We may assume that g is not constant.

Pick $r > 0$ such that $g \in \mathcal{G}^r$. Let $\epsilon = E(g)$. Then by Proposition 4.1, g is a L^2 -local maximizer of T on $\Sigma_{\epsilon, \sigma}^r$. And by Lemma 5.3 this implies that g can be ordered, as claimed. QED

6. Proof of Theorem 1.7

Let $g \in \mathcal{G}$ be a constrained entropy-maximizer with positive entropy. Assume that g and g^{*2} are not negatively correlated. Then g belongs to \mathcal{G}° by Lemma 1.4. Furthermore, as explained after Lemma 1.4, g satisfies the equation (1.3) with $\beta < 0$. By Theorem 1.6 we may assume that g is ordered.

By Proposition 3.4 we can also assume that for all $y \in I$ outside some countable set $Z \subset (0, 1)$, the function $x \mapsto g(x, y)$ is continuous on I , except for jump discontinuities at points $x \in Z$. In particular, we have $g(\cdot, y) \rightarrow g(\cdot, 0)$ pointwise, as $y \searrow 0$. It is not hard to see that the convergence is in fact uniform. Let $\tilde{I} = I \setminus Z$.

Using that $S'(g) = \alpha + \beta g^{*2}$ we have

$$S'(g(x, y)) - S'(g(x, \eta)) = \beta \int g(x, z) [g(z, y) - g(z, \eta)] dz. \quad (6.1)$$

Given $j \in \{1, 2\}$ let $y_j < \eta_j$ be two points in \tilde{I} . Define

$$\overline{S}_j'' = \int_0^1 S''(g_{j,s}) ds, \quad g_{j,s} = (1-s)g(\cdot, \eta_j) + sg(\cdot, y_j). \quad (6.2)$$

Then the equation (6.1) for $y = y_j$ and $\eta = \eta_j$ can be written as

$$\overline{S}_j''(x)f_j(x) = \beta \int g(x, z)f_j(z) dz, \quad f_j(x) = g(x, y_j) - g(x, \eta_j). \quad (6.3)$$

Equivalently, f_j satisfies the equation

$$f_j = W_j G f_j, \quad (W_j \phi)(x) = w_j(x)\phi(x), \quad w_j(x) = \frac{\beta}{\overline{S}_j''(x)}. \quad (6.4)$$

Since $y_j < \eta_j$ the function f_j is nonnegative. Furthermore, we see from (6.3) that either $f_j = 0$ on \tilde{I} or else $f_j > 0$ on \tilde{I} . Notice also that w_j is positive and bounded away from zero.

Consider W_j and G as linear operators on $L^2(I)$. Clearly W_j is bounded and G is compact. Both operators are self-adjoint, and W_j is positive. Thus, we can rewrite the equation $f_j = W_j G f_j$ more symmetrically as

$$h_j = A_j h_j, \quad h_j = W_j^{-1/2} f_j, \quad A_j = W_j^{1/2} G W_j^{1/2}. \quad (6.5)$$

Since A_j is a compact self-adjoint integral operator with positive kernel, and since h_j is a positive eigenfunction of A_j with eigenvalue 1, it is clear that $\lambda_j = 1$ is the largest eigenvalue of A_j and that it is simple. Using the corresponding Rayleigh quotient we see that

$$\lambda_j^{-1} = \inf_{f \neq 0} \frac{\langle f, \beta^{-1} \overline{S}_j'' f \rangle}{\langle f, G f \rangle}. \quad (6.6)$$

The infimum is attained if and only if f is a constant multiple of f_j .

Consider now two pairs (y_1, η_1) and (y_2, η_2) in $\tilde{I} \times \tilde{I}$ satisfying

$$y_1 \leq y_2 < \eta_2, \quad y_1 < \eta_1 \leq \eta_2. \quad (6.7)$$

Then

$$\begin{aligned} g_{1,s}(x) &= (1-s)g(x, \eta_1) + sg(x, y_1) \\ &\geq (1-s)g(x, \eta_2) + sg(x, y_2) = g_{2,s}(x). \end{aligned} \quad (6.8)$$

We will use that

$$\overline{S}_1'' - \overline{S}_2'' = \int [S''(g_{1,s}) - S''(g_{2,s})] ds = \overline{S}'''[g_{1,s} - g_{2,s}], \quad (6.9)$$

where

$$\overline{S}''' = \iint S'''(\bar{g}_{s,t}) ds dt, \quad \bar{g}_{s,t} = (1-t)g_{2,s} + tg_{1,s}. \quad (6.10)$$

Notice that

$$S'(g) = \log\left(\frac{1-g}{g}\right), \quad S''(g) = \frac{-1}{g(1-g)}, \quad S'''(g) = \frac{1}{g^2} - \frac{1}{(1-g)^2}. \quad (6.11)$$

Case 1. Consider first the case where $g(1, \eta) \geq 1/2$ for some positive $\eta \in \tilde{I}$. This occurs e.g. when $g(1, 0) > 1/2$. Let $J = (0, \eta] \cap \tilde{I}$. Then $g(x, y) \geq 1/2$ for all $x \in \tilde{I}$ and $y \in Y$. Let $y_1 = y_2 = 0$. If $f_1 = 0$ for all choices of $\eta_1 < 1$ in some subset $Y \subset J$ of positive measure, then g is constant on $\tilde{I} \times Y$ and we are done. So we may assume now that $f_1 \neq 0$. Similarly we may assume that $f_2 \neq 0$.

Consider now $\eta_1 < \eta_2$ in J . Since S''' is negative on the interval $(1/2, 1)$ we have $\overline{S_1''} \leq \overline{S_2''}$ and thus $w_1 \leq w_2$, implying that

$$\langle h, A_1 h \rangle \leq \langle h, A_2 h \rangle, \quad h \geq 0. \quad (6.12)$$

Assume for contradiction that there exists $\eta_1 < \eta_2$ in J such that $g(x, \eta_1) > g(x, \eta_2)$ for all x in some set $X \subset \tilde{I}$ of positive measure. Then the inequality (6.12) is strict unless h vanishes on X . Taking $h = h_1$ this leads to a contradiction. Thus we must have $g(\cdot, \eta_1) = g(\cdot, \eta_2)$ on \tilde{I} , for all $\eta_1 < \eta_2$ in J .

Case 2. Now consider the case where $g(0, \eta) \leq 1/2$ for some $\eta \in \tilde{I}$ smaller than 1. This occurs e.g. when $g(0, 1) < 1/2$. Let $J = [\eta, 1) \cap \tilde{I}$. Then $g(x, y) \leq 1/2$ for all $x \in \tilde{I}$ and $y \in J$. Let $\eta_1 = \eta_2 = 1$. By arguments analogous to those used in Case 1, we find that $g(\cdot, y_1) = g(\cdot, y_2)$ on \tilde{I} for all $y_1 < y_2$ in J .

Case 3. Next consider the case where $g(0, 1) = 1/2$ and $g(0, \eta) > 1/2$ for all $\eta \in \tilde{I}$. Equivalently we have $g(1, 0) = 1/2$ and $g(x, 0) > 1/2$ for all $x \in \tilde{I}$. Notice that, if $g(1, 0)$ were just a bit larger then we would be in Case 1. In order to estimate what happens near the point $(0, 1)$ it is convenient to use (6.6) in place of (6.5). Let $y_1 = y_2 = 0$. In what follows, η denotes a small positive number in \tilde{I} to be specified later. As a first step, we will prove that

$$\langle f_1, K f_1 \rangle \geq 0, \quad K = \overline{S_2''} - \overline{S_1''}, \quad (6.13)$$

for $\eta_1 < \eta_2$ in $J = (0, \eta] \cap \tilde{I}$. The inner product in this equation is the integral of $K f_1^2$ from 0 to 1. If we had $g(1, \eta) \geq 1/2$ as in Case 1, then $K \geq 0$ due to the monotonicity of S''' on $(1/2, 1)$. In the case at hand, we still have $K \geq 0$ on $[0, \xi]$ for any given $\xi \in (0, 1)$, provided that $\eta > 0$ has been chosen sufficiently small, say $\eta \leq \eta_0(\xi)$.

In what follows, our small positive parameter is $1 - \xi$, and we always assume that $0 < \eta \leq \eta_0(\xi)$. In order to prove (6.13), we will bound the values of the function $g_{1,s} - g_{2,s}$ on $[\xi, 1]$ in terms of the values of this function on $[0, \xi]$. Similarly for the function f_1 . This suffices, since the factor $\overline{S'''}$ in (6.9) can be made arbitrarily close to zero on $[\xi, 1]$ by choosing $\xi < 1$ sufficiently close to 1.

As in Case 1 we may assume that $f_1 \neq 0$ and $f_2 \neq 0$. Using positive upper and lower bounds on $\overline{S_j''}$ we have

$$0 \leq C^{-1}[g(x, \eta_1) - g(x, \eta_2)] \leq [g^{*2}(x, \eta_1) - g^{*2}(x, \eta_2)] \leq C[g(x, \eta_1) - g(x, \eta_2)], \quad (6.14)$$

for some $C > 0$ and all $x \in \tilde{I}$. The constant C only depends on the function g and the value of β . Notice that the function $g^{*2}(\cdot, \eta_1) - g^{*2}(\cdot, \eta_2)$ is decreasing. Thus, if $0 \leq x' < x \leq 1$ then

$$\begin{aligned} [g(x, \eta_1) - g(x, \eta_2)] &\leq C^2 [g(x', \eta_1) - g(x', \eta_2)], \\ [g_{1,s}(x) - g_{2,s}(x)] &\leq C^2 [g_{1,s}(x') - g_{2,s}(x')]. \end{aligned} \quad (6.15)$$

Similarly we can show that $f_1(x) \leq C^2 f_1(x')$. Applying these bounds for $0 \leq x' \leq \xi \leq x \leq 1$, with $\xi < 1$ sufficiently close to 1, it is now clear that (6.13) holds for $\eta_1 < \eta_2$ in $J = (0, \eta] \cap \tilde{I}$ and $\eta > 0$ sufficiently small. In this case we obtain

$$\frac{\langle f_1, \beta^{-1} \overline{S_2''} f_1 \rangle}{\langle f_1, G f_1 \rangle} \leq \frac{\langle f_1, \beta^{-1} \overline{S_1''} f_1 \rangle}{\langle f_1, G f_1 \rangle} = 1 = \inf_{f \neq 0} \frac{\langle f, \beta^{-1} \overline{S_2''} f \rangle}{\langle f, G f \rangle}. \quad (6.16)$$

Let $y_0 = \eta_1$ and $\eta_0 = \eta_2$ so that $f_0(x) = g(x, \eta_1) - g(x, \eta_2)$. As described after (6.4) we have either $f_0 = 0$ on \tilde{I} or $f_0 > 0$ on \tilde{I} . If we had $f_0 > 0$ on \tilde{I} then the inequality in (6.16) would be strict, which is a contradiction. Thus we must have $g(\cdot, \eta_1) = g(\cdot, \eta_2)$ on \tilde{I} , for all $\eta_1 < \eta_2$ in J .

Case 4. Finally consider the case where $g(1, 0) = 1/2$ and $g(1, \eta) < 1/2$ for all positive $\eta \in \tilde{I}$. Equivalently we have $g(0, 1) = 1/2$ and $g(x, 1) < 1/2$ for all $x \in \tilde{I}$. Notice that, if $g(0, 1)$ were just a bit smaller then we would be in Case 2. Let $\eta_1 = \eta_2 = 1$. By arguments analogous to those used in Case 3, we find that $g(\cdot, y_1) = g(\cdot, y_2)$ on \tilde{I} for all $y_1 < y_2$ in J .

In all four cases we find that $y \mapsto g(\cdot, y)$ is constant on J for some set $J \subset I$ of positive measure. Thus Theorem 1.7 is proved. **QED**

Acknowledgments. The author would like to thank C. Radin and L. Sadun for helpful discussions.

References

- [1] J.V. Ryff, *Measure preserving transformations and rearrangements*, J. Math. Anal. and Applics. **31** 449-458 (1970).
- [2] L. Lovász, B. Szegedy, *Limits of dense graph sequences*, J. Combin. Theory Ser. B **98**, 933-957 (2006).
- [3] A. Razborov, *On the minimal density of triangles in graphs*, Combin. Probab. Comput. **17**, 603-618 (2008).
- [4] S. Chatterjee, S.R.S. Varadhan, *The large deviation principle for the Erdős-Rényi random graph*, Eur. J. Comb. **32** 1000-1017 (2011).
- [5] L. Lovász, *Large networks and graph limits* AMS Colloquium Publications **60**, American Mathematical Society, Providence RI, (2012).
- [6] O. Pikhurko, A. Razborov, *Asymptotic structure of graphs with the minimum number of triangles*, arXiv:1204.2846.
- [7] S. Chatterjee, P. Diaconis, *Estimating and understanding exponential random graph models*, The Annals of Statistics **41**, 2428–2461 (2013).
- [8] C. Radin, L. Sadun, *Phase transitions in a complex network*, J. Phys. A: Math. Theor. **46**, 305002 (2013)

- [9] C. Radin, K. Ren, L. Sadun, *The asymptotics of large constrained graphs*, J. Phys. A: Math. Theor. **47**, 20pp (2014).
- [10] C. Radin, L. Sadun, *Singularities in the entropy of asymptotically large simple graphs*, J. Stat. Phys. **158**, 853-865 (2015).
- [11] R. Kenyon, C. Radin, K. Ren, L. Sadun, *Bipodal structure in oversaturated random graphs*, Preprint `mp_arc 15-100` (2015).