This talk is based on Balázs Szegedy's paper[1] "An information theoretic approach to Sidorenko's conjecture"; however, much of the presentation is in a different language (he does everything in terms of the Kullback–Leibler divergence, which is a function that is for certain purposes more amenable than entropy, though in my opinion is harder to understand intuitively). Another good resource for this topic is Gowers's blog post[2] "Entropy and Sidorenko's conjecture—after Szegedy", which is presented in the language of entropy, though it only seeks to prove very special cases of Szegedy's results.

1 Basics of information theory

Suppose X is a random variable on a finite space $\mathcal{X}$, with probability density $p(x) = \Pr(X = x)$ for $x \in \mathcal{X}$.

Definition 1. The (Shannon) entropy of X is defined by

$$\mathsf{H}(X) = \sum_{x \in \mathcal{X}} p(x) \log \frac{1}{p(x)} = \mathop{\mathbb{E}}_X \left[\log \frac{1}{p(X)} \right]$$

with the convention that $0 \log 0 = 0$.

The Shannon entropy is supposed to measure the amount of information conveyed by X, as measured by the number of bits it should take to store the outcome of X. By Jensen's inequality, we have that

$$\mathsf{H}(X) = \mathop{\mathbb{E}}_X \left[\log \frac{1}{p(X)} \right] \leq \log \left(\mathop{\mathbb{E}}_X \left[\frac{1}{p(X)} \right] \right) = \log |\mathcal{X}|$$

Moreover, equality is achieved by the uniform distribution on X.

For two (possibly dependent) random variables X, Y on spaces $\mathcal{X}, \mathcal{Y}$, we can define their joint entropy by the formula

$$\mathsf{H}(X, Y) = \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p(x, y) \log \frac{1}{p(x, y)}$$

where $p(x, y) = \Pr(X = x, Y = y)$. Equivalently, if we think of the pair (X, Y) as a random variable on the space $\mathcal{X} \times \mathcal{Y}$, then $\mathsf{H}(X, Y)$ is just the entropy of this single variable. We can similarly define the conditional entropy $\mathsf{H}(Y \mid X)$ to be the entropy of the random variable $(Y \mid X)$ on the space $\mathcal{X} \times \mathcal{Y}$, defined by

$$\Pr((Y \mid X) = (x, y)) = \Pr(Y = y \mid X = x)$$

Equivalently, we can define the conditional entropy by the formula

$$\mathsf{H}(Y \mid X) = \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p(x, y) \log \frac{p(x)}{p(x, y)}$$

Regardless of definition, it measures the amount of information gained by learning Y, given that we already know X. From the definitions and some simple computations, it follows that

$$\mathsf{H}(X, Y) = \mathsf{H}(X) + \mathsf{H}(Y \mid X)$$

which makes intuitive sense: the amount of information gained by learning X and Y is the same as the amount of information gained by first learning X and then learning Y, already knowing X. Inductively applying this fact, one can derive the *chain rule*:

$$\mathsf{H}(X_1, \ldots, X_n) = \mathsf{H}(X_1) + \mathsf{H}(X_2 \mid X_1) + \cdots + \mathsf{H}(X_n \mid X_1, \ldots, X_{n-1})$$

[1] `https://arxiv.org/pdf/1406.6738.pdf`
[2] `https://gowers.wordpress.com/2015/11/18/entropy-and-sidorenkos-conjecture-after-szegedy/`

If X and Y are independent, then

$$\mathsf{H}(X,Y)=\mathsf{H}(X)+\mathsf{H}(Y) \qquad \mathsf{H}(Y\mid X)=\mathsf{H}(Y)$$

At the other extreme, if the value of X determines the value of Y (i.e. if $Y=f(X)$ for some deterministic function f), then

$$\mathsf{H}(X,Y)=\mathsf{H}(X) \qquad \mathsf{H}(Y\mid X)=0$$

The important consequence of these results, for our purposes, is that we can always add "dummy" variables to our entropy: if X determines Y, then we can always add Y to any entropy involving X, and we can also add Y to the conditioning in any conditional entropy conditioning on X.

The final notion we will need is really one in probability theory. Suppose that $\mathcal{X}_1,\mathcal{X}_2,\mathcal{X}_3$ are three finite probability spaces, coming with random variables X_1,X_2,X_3 on them; suppose too that every point in $\mathcal{X}_i$ has positive probability under X_i, to avoid degeneracy. Suppose we have maps $\psi_i:\mathcal{X}_i\to\mathcal{X}_3$ for $i=1,2$ so that $\psi_i(X_i)=X_3$ for $i=1,2$. Let $\mathcal{X}_4$ be the fiber product of this configuration, i.e.

$$\mathcal{X}_4=\{(x_1,x_2)\in\mathcal{X}_1\times\mathcal{X}_2:\psi_1(x_1)=\psi_2(x_2)\}$$

Then a random variable X on $\mathcal{X}_4$ is called a *coupling* of X_1 and X_2 over X_3 if its marginal distributions on $\mathcal{X}_1$ and $\mathcal{X}_2$ equal X_1,X_2, respectively. Moreover, the most natural and most important coupling for our purposes will be the *conditionally independent* coupling X_4, defined by

$$\Pr(X_4=(x_1,x_2))=\frac{\Pr(X_1=x_1)\Pr(X_2=x_2)}{\Pr(X_3=\psi_1(x_1))}$$

Intuitively, $\mathcal{X}_4$ consists of all possible outcomes from $\mathcal{X}_1\times\mathcal{X}_2$ that agree on their induced outcome in $\mathcal{X}_3$, and a coupling is just a probability distribution on such outcomes. The conditionally independent coupling is the coupling that is "as independent as possible": if we observe a sample from X_3, then X_4 will be the independent distribution on all outcomes from $\mathcal{X}_1,\mathcal{X}_2$ that yield the outcome we observed in $\mathcal{X}_3$. The most important property that we will need of couplings is that the conditionally independent coupling maximizes entropy: i.e. for every coupling X, we have that

$$\mathsf{H}(X)\le\mathsf{H}(X_4)$$

This follows from the so-called submodularity of entropy, and is a relative version of the fact above, that the uniform distribution maximizes entropy. Heuristically, it just says that the conditionally independent coupling is the one that is most random: given the outcome of X_3, we impose no further dependencies on the outcome from $\mathcal{X}_1,\mathcal{X}_2$.

2 Sidorenko's Conjecture

Recall that we denote by $\mathsf{v}(G),\mathsf{e}(G)$ the number of vertices and edges, respectively, of a graph G. For two graphs H,G, we define

$$t(H,G)=\frac{|\hom(H,G)|}{\mathsf{v}(G)^{\mathsf{v}(H)}}$$

This is the fraction of maps $V(H)\to V(G)$ that map edges to edges, or equivalently the probability that a random map $V(H)\to V(G)$ will map edges to edges. Sidorenko's conjecture says that for all bipartite H and all G,

$$t(H,G)\ge t(K_2,G)^{\mathsf{e}(H)}$$

If G is a random graph, then $t(H,G)=t(K_2,G)^{\mathsf{e}(H)}+o(1)$ with high probability, and thus, if Sidorenko's conjecture is true, it is asymptotically tight.

Szegedy's "information theoretic" approach to Sidorenko's conjecture is based on the idea that if X is a random variable supported on $\hom(H, G)$, then

$$\log |\hom(H, G)| \geq \mathsf{H}(X)$$

and therefore

$$\log t(H, G) \geq \mathsf{H}(X) - \mathsf{v}(H) \log \mathsf{v}(G)$$

Similarly,

$$\log \left(t(K_2, G)^{\mathsf{e}(H)} \right) = \mathsf{e}(H)(\log(2\mathsf{e}(G)) - 2 \log \mathsf{v}(G))$$

Thus, to prove Sidorenko's conjecture, it suffices to find a random variable X supported on $\hom(H, G)$ so that

$$\begin{aligned} \mathsf{H}(X) &\geq \mathsf{e}(H)(\log(2\mathsf{e}(G)) - 2 \log \mathsf{v}(G)) + \mathsf{v}(H) \log \mathsf{v}(G) \\ &= \mathsf{e}(H) \log(2\mathsf{e}(G)) + (\mathsf{v}(H) - 2\mathsf{e}(H)) \log \mathsf{v}(G) \end{aligned}$$

To write this in a slightly nicer form, let V be a uniformly random vertex of G, and E a uniformly random oriented edge (i.e. we care about the order of the endpoints of E). Equivalently, V is uniform on $\hom(K_1, G)$ and E is uniform on $\hom(K_2, G)$. Then we have that

$$\mathsf{H}(V) = \log \mathsf{v}(G) \qquad\qquad \mathsf{H}(E) = \log(2\mathsf{e}(G))$$

Thus, to prove Sidorenko's conjecture for H, it suffices to find a random variable X on $\hom(H, G)$ such that

$$\mathsf{H}(X) \geq \mathsf{e}(H)\mathsf{H}(E) + (\mathsf{v}(H) - 2\mathsf{e}(H))\mathsf{H}(V)$$

We will actually also require the stronger (but very natural) condition that the marginal distribution on any edge of H is just the distribution of E. More formally, we define the following.

Definition 2. A *witness variable* on a (bipartite) graph H is an infinite family of random variables $X(G)$, one for each graph G, with the following properties:

1. $X(G)$ is a random variable on the space $\hom(H, G)$
2. For every edge $uv \in E(H)$, the marginal distribution of $X(G)$ on uv (formally, the induced distribution coming from the projection $\hom(H, G) \to \hom(uv, G)$) is uniform
3. $$\mathsf{H}(X(G)) \geq \mathsf{e}(H)\mathsf{H}(E(G)) + (\mathsf{v}(H) - 2\mathsf{e}(H))\mathsf{H}(V(G))$$

Generally, we'll suppress the G, and just talk about the variable X rather than $X(G)$. However, note that the definition of a witness variable requires that a variable on $\hom(H, G)$ be defined for *every* G. In particular, this implies that the space $\hom(H, G)$ is non-empty for every G, which implies that H must be bipartite.

The argument presented above proves the following result:

Theorem 3. *If H has a witness variable, then H satisfies Sidorenko's conjecture.*

Thus, the task at hand is to find witness variables on graphs. To build witness variables on new graphs, we will use inductive procedures, where we use various building operations to create new graphs from old graphs, and similarly combine the witness variables to get a witness variable on the new graph. To start, we need one graph with a witness variable.

Proposition 4. *K_2 has a witness variable, namely E, the uniform distribution on* $\hom(K_2, G)$.

Proof. E is certainly supported on $\hom(K_2, G)$ and has the correct marginal, by definition. So all we need to check is that for any graph G,

$$\mathsf{H}(E) \geq \mathsf{e}(K_2)\mathsf{H}(E) + (\mathsf{v}(K_2) - 2\mathsf{e}(K_2))\mathsf{H}(V)$$

But this certainly holds, since $\mathsf{v}(K_2) = 2\mathsf{e}(K_2)$, so the second term disappears, and we are just left with the equality $\mathsf{H}(E) = \mathsf{H}(E)$. $\square$

Our main technique for building up new graphs from old graphs will be gluing.

Definition 5. Given two graphs H_1, H_2 and vertex subsets $S_1 \subseteq V_1, S_2 \subseteq V_2$, and given a bijection $f : S_1 \leftrightarrow S_2$, we define the *glued graph* $H = H_1 \cup_f H_2$ in the natural way. We let

$$V(H) = (V(H_1) \sqcup V(H_2))/(s \sim f(s) : s \in S_1)$$

and let $E(H)$ be the image of $E(H_1) \sqcup E(H_2)$ under the gluing projection, except that we delete all parallel edges we get. We will denote by S the image of S_1 (and S_2) in H.

Lemma 6. *Suppose H_1, H_2 have witness variables X_{H_1}, X_{H_2}, and suppose that S_1, S_2 are independent sets in H_1, H_2, respectively. Suppose too that there is a bijection $f : S_1 \leftrightarrow S_2$ so that the marginal distributions $X_{H_1}|_S$ and $X_{H_2}|_S$ are identical (i.e. f induces a measure-preserving map $X_{H_1}|_{S_1} \to X_{H_2}|_{S_2}$). Let $H = H_1 \cup_f H_2$, and let X be the conditionally independent coupling of X_{H_1}, X_{H_2}. Then X is a witness variable for H.*

Proof. First, we need to check that H is indeed supported on $\hom(H, G)$. This follows from the fact that a homomorphism $H \to G$ is the same as two homomorphisms $H_1 \to G, H_2 \to G$ that agree on S (i.e. that are identified under f), and thus $\hom(H, G)$ is indeed the fiber product of $\hom(H_1, G), \hom(H_2, G)$ over $\hom(S, G)$.

Next, we will check the marginals. Since S_1, S_2 are independent sets, any edge in H must be contained in either H_1 or H_2 (and not both); without loss of generality, it's in H_1. But since X is a conditionally independent coupling, the distribution of this edge does not depend on the mapping of $H_2 \setminus S$, and thus the marginal of X is the same as the marginal of X_{H_1}, which we assumed was E.

Finally, we need to check the entropy inequality. Let X_S be the marginal of S, i.e. the induced random variable on $\hom(S, G)$. Since X determines and is determined by the mapping of the vertices, and since $H_1 \cup H_2 \cup S$ covers the vertices of H, we have that $\mathsf{H}(X) = \mathsf{H}(X_{H_1}, X_{H_2}, X_S)$. We now apply the chain rule:

$$\mathsf{H}(X) = \mathsf{H}(X_{H_1}, X_{H_2}, X_S) = \mathsf{H}(X_S) + \mathsf{H}(X_{H_1} \mid X_S) + \mathsf{H}(X_{H_2} \mid X_S, X_{H_1})$$

Now, recall that conditional on X_S, X_{H_2} is independent of X_{H_1}. So we may remove X_{H_1} from the final conditioning, and get

$$\mathsf{H}(X) = \mathsf{H}(X_S) + \mathsf{H}(X_{H_1} \mid X_S) + \mathsf{H}(X_{H_2} \mid X_S)$$

On the other hand, we may also write

$$\mathsf{H}(X_{H_1}) = \mathsf{H}(X_{H_1}, X_S) = \mathsf{H}(X_S) + \mathsf{H}(X_{H_1} \mid X_S)$$

and similarly for X_{H_2}. Plugging this in, we get that

$$\mathsf{H}(X) = \mathsf{H}(X_{H_1}) + \mathsf{H}(X_{H_2}) - \mathsf{H}(X_S)$$

It is worth noting that one of the most important properties of entropy is its submodularity, which means that we can always get an inequality of this form. However, in the conditionally independent setting, we actually get equality (which is good, because the inequality goes the wrong way for our purposes); this can be thought of as a form of inclusion-exclusion for conditionally independent random variables. By our assumption that X_{H_1}, X_{H_2} are witness variables, we get that

$$\begin{aligned}\mathsf{H}(X) &\geq [\mathsf{e}(H_1)\mathsf{H}(E) + (\mathsf{v}(H_1) - 2\mathsf{e}(H_1))\mathsf{H}(V)] + [\mathsf{e}(H_2)\mathsf{H}(E) + (\mathsf{v}(H_2) - 2\mathsf{e}(H_2))\mathsf{H}(V)] - \mathsf{H}(X_S) \\ &= [\mathsf{e}(H_1) + \mathsf{e}(H_2)]\mathsf{H}(E) + [\mathsf{v}(H_1) + \mathsf{v}(H_2) - 2\mathsf{e}(H_1) - 2\mathsf{e}(H_2)]\mathsf{H}(V) - \mathsf{H}(X_S)\end{aligned}$$

Observe that since S is independent, $\mathsf{e}(H_1)+\mathsf{e}(H_2) = \mathsf{e}(H)$ and $\mathsf{v}(H_1)+\mathsf{v}(H_2) = \mathsf{v}(H)+|S|$. Finally, observe that X_S is some (potentially very complicated) random variable on $V(G)^S$. Therefore, we have that

$$\mathsf{H}(X_S) \le \log |V(G)^S| = |S| \log \mathsf{v}(G) = |S|\mathsf{H}(V)$$

Putting this all together, we get that

$$\mathsf{H}(X) \ge \mathsf{e}(H)\mathsf{H}(E) + (\mathsf{v}(H) - 2\mathsf{e}(H))\mathsf{H}(V)$$

□

Therefore, if we can build up a graph H by starting with a single edge and always gluing along identically-distributed independent sets, we get a witness variable for H, and thus show that H satisfies Sidorenko's conjecture. Of course, checking that two independent sets in two different graphs have the same distribution might be hard; one convenient tool to do this in simple cases is the following lemma:

Lemma 7. *Suppose H has no isolated vertices. Let D be a random variable on $V(G)$ defined by*

$$\Pr(D = v) = \frac{\deg(v)}{2\mathsf{e}(G)}$$

Then if X is a witness variable, then the marginal of X on any vertex of H is precisely D.

Proof. Observe that D is the marginal of E on either of its vertices. Then since E is the marginal of X on every edge of H, D must be the marginal of X on any vertex of H. □

Corollary 8. *If H_1, H_2 have witness variables and H is formed by gluing them along a single non-isolated vertex (i.e. $|S| = 1$), then H has a witness variable.*

Proof. This follows immediately from Lemma 6. Indeed, $X_{H_1}|_S$ and $X_{H_2}|_S$ must both be distributed as D, since they both consist of a single vertex, and in particular they are identically distributed. □

Corollary 9. *Every tree has a witness variable, and thus satisfies Sidorenko's conjecture.*

Proof. We prove this by induction on the number of vertices; the base case $\mathsf{v}(H) = 2$ is the K_2 case proved above. Inductively, we may pick a leaf of H and write H as a gluing of a smaller tree H_1 and a single edge, glued along a single vertex. Then by the induction hypothesis and by Corollary 8, we are done. □

Corollary 10. *Let T be a tree, let $S \subseteq V(T)$ be an independent set, and let $f : S \to S$ be the identity map. Then $H = T \cup_f T$ has a witness variable, and thus satisfies Sidorenko's conjecture.*

In particular, by taking T to be a path, all even cycles satisfy Sidorenko's conjecture.

Proof. This again follows immediately from Lemma 6. We proved above that T has a witness variable X_T, and since we are gluing along the identity map, we automatically have that X_T induces the same distributed on $S_1 = S_2 = S$. Thus, we can take the conditionally independent coupling and get the desired witness variable on H. □

In Sidorenko's original paper, he proved using Cauchy–Schwarz and Hölder that trees and even cycles satisfy Sidorenko's conjecture, so these results in themselves are not too exciting. However, the next result uses the same tools to show that all so-called "tree-arrangeable" graphs satisfy Sidorenko's conjecture, which was the state of art before Szegedy's work.

Definition 11. Suppose we are given a bipartite graph $H = (A, B, E)$. We are allowed to extend H in two ways to get a new bipartite graph H':

- We may add a single vertex v to A and connect to a single vertex $b \in B$, or

- We may add a single vertex v to B and connect it a subset of $N(b)$, where b is some vertex in B and $N(b)$ is its neighborhood in A.

A graph is called *tree-arrangeable* if it may be built up from K_2 using a sequence of these allowed operations. The name "tree-arrangeable" comes from an equivalent description, which is a bit less intuitive in my opinion, and also less amenable to our purposes.

Tree-arrangeable graphs were introduced by Kim, Lee, and Lee, who proved that they satisfy Sidorenko's conjecture. Using the entropy tools we've already developed, we can provide a simple alternative proof.

Theorem 12. *Every tree-arrangeable graph has a witness variable, and thus satisfies Sidorenko's conjecture.*

Proof. We will prove this by induction on the number of vertices of $H = (A, B, E)$. In fact, the inductive statement will be stronger: we will prove that not only does H have a witness variable X_H, but that for every vertex $u \in B$, the marginal distribution of its neighborhood $N(u)$ is just the distribution of $\deg(u)$ independent neighbors of a sample of D.

The base case is $H = K_2$, for which both these statements are clear; the witness variable is just E, and the only neighborhoods are single vertices, for which we already know the marginals to be distributed as D.

For the inductive case, suppose this is true for all H with $n-1$ vertices. For a tree-arrangeable graph H' on n vertices, we may write it as an extension of some H on $n-1$ vertices according to one of the two operations defined above. First, suppose that H' is gotten from H by adding a new vertex v to A and connecting it to some $b \in B$; equivalently, H' is gotten by gluing H and K_2 along $S = \{b\}$. Then by Corollary 8, we know that the conditionally independent coupling of X_H with E will be a witness variable for H'. To check the condition about neighborhoods, observe that all neighborhoods in H' are the same as neighborhoods in H, except for $N(b)$. Its neighborhood in H' is just its neighborhood in H, plus the new vertex v. But since $X_{H'}$ is a conditionally independent coupling along b, we know that the distribution of v is independent from the distribution of the other neighbors of b. Since we assumed inductively that they were distributed as independent neighbors of D, the same is true in H'.

Now, assume instead that H' is gotten from H by adding a new vertex v to B and connecting it to some subset of $N_H(b)$ for some fixed $b \in B$. Equivalently, we can think of H' as a gluing of H with a star $K_{1,m}$ along m neighbors of b. Since $K_{1,m}$ can be built as a series of gluings of edges, we know by the same argument as above that in its witness variable, the marginal on the m leaves is just m independent neighbors of D. Moreover, by the inductive hypothesis, we know that the distribution of any m neighbors of b is also that of m independent neighbors of D. So by Lemma 6, this gluing will provide us a witness variable $X_{H'}$ for H'. To check the neighborhood condition, again observe that the only neighborhood of a vertex in B that changed is the neighborhood of v. But we automatically get that the distribution of $N(v)$ is that of m independent neighbors of D, since that was the marginal of this set in both $X_{K_{1,m}}$ and X_H. This completes the proof. $\square$

Observe that any bipartite graph which has one vertex b that is complete to the other side is tree-arrangeable. Indeed, we may build up such a graph by first adding a bunch of vertices to A, each connected to b, according to the first operation, and then adding vertices to B connected to whichever vertices of A we'd like, by the second operation. Thus, this result generalizes the theorem of Conlon, Fox, and Sudakov, which says that bipartite graphs with one complete vertex satisfy Sidorenko's conjecture.

Lemma 6 said that when we glue along identically-distributed independent sets, we can build new witness variables by a conditionally independent coupling. We can in fact do even better, and glue along forests. To do this, let H be a bipartite graph with no isolated vertices, and say that X is a *strong witness variable* for H if we have the inequality

$$\mathsf{H}(X) \geq \mathsf{e}(H)\mathsf{H}(E) + (\mathsf{v}(H) - 2\mathsf{e}(H))\mathsf{H}(D)$$

(the only difference being that $\mathsf{H}(V)$ was replaced by $\mathsf{H}(D)$). Note that since H has no isolated vertices, we have that $\mathsf{v}(H) \leq 2\mathsf{e}(H)$, and thus the second term is negative. Thus, since $\mathsf{H}(D) \leq \mathsf{H}(V)$, this inequality is indeed stronger than the inequality defining a witness variable. Note that for $H = K_2$, we still have equality if $X = E$, and thus K_2 has a strong witness variable. Therefore, for the purposes of inductively building up graphs from an edge and from gluings, the base case still works if we wish to maintain this stronger condition.

Lemma 13. *Suppose H_1, H_2 have strong witness variables X_{H_1}, X_{H_2}, and suppose that S_1, S_2 are vertex sets in H_1, H_2, respectively, with a bijection $f : S_1 \leftrightarrow S_2$ so that the marginal distributions $X_{H_1}|_S$ and $X_{H_2}|_S$ are identical. Suppose too that S_1, S_2 span forests (i.e. contain no cycles), and that f is an isomorphism of these forests. Let $H = H_1 \cup_f H_2$, and let X be the conditionally independent coupling of X_{H_1}, X_{H_2}. Then X is a strong witness variable for H.*

Proof. As in the proof of Lemma 6, we can use conditional independence and the chain rule to write

$$\begin{aligned}\mathsf{H}(X) &= \mathsf{H}(X_{H_1}, X_{H_2}, X_S)\\ &= \mathsf{H}(X_S) + \mathsf{H}(X_{H_1} \mid X_S) + \mathsf{H}(X_{H_2} \mid X_S, X_{H_1})\\ &= \mathsf{H}(X_S) + \mathsf{H}(X_{H_1} \mid X_S) + \mathsf{H}(X_{H_2} \mid X_S)\\ &= \mathsf{H}(X_S) + (\mathsf{H}(X_{H_1}) - \mathsf{H}(X_S)) + (\mathsf{H}(X_{H_2}) - \mathsf{H}(X_S))\\ &= \mathsf{H}(X_{H_1}) + \mathsf{H}(X_{H_2}) - \mathsf{H}(X_S)\end{aligned}$$

By assumption,

$$\begin{aligned}\mathsf{H}(X_{H_1}) &\geq \mathsf{e}(H_1)\mathsf{H}(E) + (\mathsf{v}(H_1) - 2\mathsf{e}(H_1))\mathsf{H}(D)\\ \mathsf{H}(X_{H_2}) &\geq \mathsf{e}(H_2)\mathsf{H}(E) + (\mathsf{v}(H_2) - 2\mathsf{e}(H_2))\mathsf{H}(D)\end{aligned}$$

Moreover, since H is gotten by gluing H_1 and H_2 along H,

$$\begin{aligned}\mathsf{v}(H) &= \mathsf{v}(H_1) + \mathsf{v}(H_2) - \mathsf{v}(S)\\ \mathsf{e}(H) &= \mathsf{e}(H_1) + \mathsf{e}(H_2) - \mathsf{e}(S)\end{aligned}$$

We get this formula for the edge count because we delete all parallel edges when gluing, and since f is an isomorphism, every edge in S comes from both S_1 and S_2, and thus one copy of it is deleted. So if we can prove that

$$\mathsf{H}(X_S) \leq \mathsf{e}(S)\mathsf{H}(E) + (\mathsf{v}(S) - 2\mathsf{e}(S))\mathsf{H}(D)$$

then we will get that

$$\begin{aligned}\mathsf{H}(X) &\geq (\mathsf{e}(H_1) + \mathsf{e}(H_2) - \mathsf{e}(S))\mathsf{H}(E) + [(\mathsf{v}(H_1) + \mathsf{v}(H_2) - \mathsf{v}(S)) - 2(\mathsf{e}(H_1) + \mathsf{e}(H_2) - \mathsf{e}(S))]\mathsf{H}(D)\\ &= \mathsf{e}(H)\mathsf{H}(E) + (\mathsf{v}(H) - 2\mathsf{e}(H))\mathsf{H}(D)\end{aligned}$$

as desired.

So it suffices to prove that $\mathsf{H}(X_S) \leq \mathsf{e}(S)\mathsf{H}(E) + (\mathsf{v}(S) - 2\mathsf{e}(S))\mathsf{H}(D)$. Note that since X is a conditionally independent coupling, we have that the marginals on all edges in S are E, and the marginals on all vertices in S are D. We will actually prove the following more general fact:

Proposition 14. *If F is a forest and Y is a random variable on $\hom(F, G)$ such that the marginal of Y on every edge of F is E, and the marginal on every vertex is D, then*

$$\mathsf{H}(Y) \leq \mathsf{e}(F)\mathsf{H}(E) + (\mathsf{v}(F) - 2\mathsf{e}(F))\mathsf{H}(D)$$

It's interesting to note that once we finish this proof, we will know that F has a strong witness variable, and for that variable the reverse inequality is also true. Thus, for forests, the entropy inequality for strong witness variables is actually an equality.

Proof. We prove this by induction on $\mathsf{v}(F)$. In the case $\mathsf{v}(F) = 2$, we have that F is either K_2, in which case we just need $\mathsf{H}(Y) \leq \mathsf{H}(E)$, which is true with equality, or else F is two independent vertices, in which case we need $\mathsf{H}(Y) \leq 2\mathsf{H}(V)$. This is also true, since Y is some variable on $V(G)^2$, so its entropy is bounded by the entropy of the uniform distribution, which is $2\mathsf{H}(V)$. So the base case is proved.

For the inductive case, since F is a forest, we can find some $v \in V(F)$ so that deleting v disonnects F. Therefore, we can write $V = V_1 \cup V_2$, where $V_1 \cap V_2 = \{v\}$, there is no edge between V_1 and V_2, and $|V_1|, |V_2| < |V(F)|$. Then Y is some coupling of $Y|_{V_1}$ and $Y|_{V_2}$, and thus its entropy is upper-bounded by the entropy of the conditionally independent coupling. Therefore, by the inclusion-exclusion property of the conditionally independent coupling, we get that

$$\mathsf{H}(Y) \leq \mathsf{H}(Y|_{V_1}) + \mathsf{H}(Y|_{V_2}) - \mathsf{H}(Y|_v)$$

Note that $Y|_{V_1}$ and $Y|_{V_2}$ are also distributions whose marginals on edges are E and on vertices are D, so by the inductive hypothesis,

$$\begin{aligned}\mathsf{H}(Y) &\leq (\mathsf{e}(V_1)\mathsf{H}(E) + (|V_1| - 2\mathsf{e}(V_1))\mathsf{H}(D)) + (\mathsf{e}(V_2)\mathsf{H}(E) + (|V_2| - 2\mathsf{e}(V_2))\mathsf{H}(D)) - \mathsf{H}(Y|_v) \\ &= (\mathsf{e}(V_1) + \mathsf{e}(V_2))\mathsf{H}(E) + (|V_1| + |V_2| - 1 - 2(\mathsf{e}(V_1) + \mathsf{e}(V_2)))\mathsf{H}(D) \\ &= \mathsf{e}(F)\mathsf{H}(E) + (\mathsf{v}(F) - 2\mathsf{e}(F))\mathsf{H}(D)\end{aligned}$$

□

This completes the proof of the proposition, and thus we are done. □

Corollary 15 (Stated somewhat informally)**.** *Consider the class of graphs that can be built up from a single edge by repeated gluings along isomorphic forests, where all the gluings are done so that the marginal distributions are equal along the glued forests. Then any graph in this class satisfies Sidorenko's conjecture.*

Proof. This follows immediately from Lemma 13; the only thing to be slightly careful about is that the notion of a strong witness variable is only stronger than the usual notion of a witness variable under the assumption that the graph has no isolated vertices. But this is no real issue; if H has isolated points, begin by deleting them and finding a strong witness variable for the remaining part of the graph. In particular, it is a witness variable. Now, adding back the isolated vertices (e.g. by gluing them along an empty independent set, using Lemma 6), we get a witness variable for H, so it satisfies Sidorenko's conjecture. □

Unfortunately, I haven't been able to think of any natural examples of graphs that can be constructed by gluings along forests, but cannot be constructed by gluings along independent sets. For unnatural examples, one can simply take two arbitrarily complicated graphs that have witness variables, and glue them along an edge; I think that in general, such examples cannot be constructed by only gluing along independent sets. I suspect that hypercubes can be built by gluings along forests, which would allow us to reprove a theorem of Hatami (using the techniques of weakly norming graphs) that shows that all hypercubes satisfy Sidorenko's conjecture; however, I have not yet found a way to construct such a sequence of gluings.

In Szegedy's paper, he proves Sidorenko's conjecture for an even larger class of graphs, which he calls *thick graphs*. However, even their definition is difficult to understand: they are graphs that arise in certain ways from certain complicated hypergraphs (called *reflection complexes*), and these hypergraphs can themselves be built up by various gluing operations akin to the ones above. The language of reflection complexes allows one to keep track of the gluing operations in a systematic fashion (in particular, it makes easy the important step of verifying that the sets along which we're gluing have identical marginals), so it does appear to be the correct framework to work in when one wants to derive the most general theorem possible from the entropy method. However, it is also a fairly complicated framework to define and to work in, and I don't understand it well enough to explain it. In theory, if a graph is thick, then it should be possible to prove that it satisfies Sidorenko's conjecture using only simple operations, as above, by manually keeping track of the information that the reflection complex stores automatically. However, I have been unable to actually do this for even fairly simple graphs (e.g. Szegedy reproves that hypercubes satisfy Sidorenko's conjecture using this technique, but I have been unable to understand the sequence of simple entropy steps that his proof should be encoding).