Freie Universität Berlin

Fachbereich Mathematik und Informatik

Master's thesis

On Sidorenko's conjecture

Submitted by:	Olaf Parczyk (4365433)
Submitted at:	18th november 2014
Thesis Advisor:	Prof. Dr. Tibor Szabó
Second Referee:	Prof. Dr. Martin Aigner

Zusammenfassung

Sei $t(H, G)$ die Wahrscheinlichkeit, dass eine gleichverteile geqählte Abbildung von einem bipartiten Graphen H in einen Graphen G ein Homomorphismus ist. Sidorenko's Vermutung [10] sagt aus, dass $t(H, G) \geq t(K_2, G)^{|E(H)|}$ für jeden Graphen G gilt. Insbesondere bedeutet dies, dass unter allen Graphen mit der selben Kantendichte der zufällige Graph die geringste Anzahl an Kopien von H besitzt. Die Vermutung wurde für verschiedene Klassen von Graphen bewiesen, einschließlich Bäume, gerade Kreise, Würfel und bipartite Graphen bei denen ein Knoten mit der kompletten anderen Seite verbunden ist.

Wir möchten hier zwei Arbeiten im Details untersuchen und deren Bestandteile präsentieren. Für die erste, „On the logarithmic calculus and Sidorenko's conjecture"von Li und Szegedy, [9] entwickeln wir eine Erweiterung der Methode. In der erst kürzlich erschienene Arbeit „Relative entropy and Sidorenko's conjecture"von Szegedy [12] wird eine viel stärkere Methode vorgestellt, die uns ermöglicht Graphen an Wäldern zusammenzukleben, wobei die Ungleichung erhalten bleibt. Wir erweitern auch diese Methode um die Vermutung für Würfel zu verifizieren und zum ersten Mal für die 1-subdivision des vollständigen Graphen K_m zu bewiesen. Außerdem untersuchen wir die Beschränkungen der Methode und beantworten teilweise eine Frage von Szegedy.

Abstract

Let $t(H, G)$ be the probability that a uniformly at random chosen map from a bipartite graph H to any graph G is a homomorphism. Sidorenko's conjecture [10] says that we have $t(H, G) \geq t(K_2, G)^{|E(H)|}$ for all G. In particular this says that among all graphs with the same edge density the random graph contains the minimum number of copies of H. The conjecture was proven for various classes of graphs, including trees, even cycles, cubes and bipartite graphs where one vertex is complete to the other side.

In this thesis we carefully examine two papers and give a detailed presentation of all ingredients. For the first one, 'On the logarithmic calculus and Sidorenko's conjecture' by Li and Szegedy [9], we present an extension of their method. The more recent paper 'Relative entropy and Sidorenko's conjecture' by Szegedy [12] provides a stronger method, which allows us to glue graphs on forests preserving Sidorenko's inequality. We also extend this method and apply it to reprove the conjecture for cubes and prove the conjecture for the 1-subdivision of the complete graph K_m. Moreover we analyse the limitations of the method and partially answer a question of Szegedy.

Contents

1 Introduction

1.1 Sidorenko's conjecture

Let H be a simple bipartite graph on $V(H) = \{1, \dots, n\}$ vertices with e edges and G any simple graph on N vertices with E edges. A map $a : V(H) \to V(G)$ is a **homomorphism** if all edges are mapped onto edges, i.e. for all $\{i, j\} \in E(H)$ we have $\{a(i), a(j)\} \in E(G)$. We denote the set of homomorphism from $V(H)$ to $V(G)$ by $Hom(H, G)$. For a uniformly at randomly chosen map the probability that it is a homomorphism is

$$t(H, G) = \frac{|Hom(H, G)|}{N^n}$$

which we call the **homomorphism density** of H in G. For a single edge we call K_2 $d := t(K_2, G) = \frac{2E}{N^2}$ the **edge density** of G. Note that usually the edge density is $E\binom{N}{2}^{-1}$, but asymptotically it is the same.

Sidorenko's conjecture can then be stated as:

Conjecture 1.1. For all bipartite graphs H and for all graphs G we have

$$t(H, G) \geq t(K_2, G)^e = d^e.$$

If for a graph H the statement is true for all G, then we say that H satisfies Sidorenko's conjecture, H has the Sidorenko property or shortly H is Sidorenko. Observe that for non-bipartite graphs H the conjecture is not true, because odd cycles can only be mapped to smaller odd cycles and therefore there are graphs, with any chromatic number, for which $Hom(H, G)$ is empty and $d > 0$.

In particular the conjecture says that asymptotically among all graphs with the same edge density d the random graph has the minimal number of copies of H. To make this clearer we fix H and take a closer look at $t(H, G)$. Let $sub(H, G)$ be the number of subgraphs in G isomporhic to H and $aut(H)$ the number of automorphism of H. Then certainly $aut(H) \cdot sub(H, G)$ counts all injective homomorphisms of H into G. In all non-injective homomorphisms at least two vertices are mapped to the same vertex and thus there are at most $\binom{n}{2}N^{n-1}$ of them. With N tending to infinity this gives us

$$t(H, G) = \frac{aut(G)sub(H, G)}{N^n} + O\left(\frac{1}{N}\right).$$

The conjecture then implies

$$sub(H, G) \geq \frac{d^e N^n - O(N^{n-1})}{aut(H)},$$

where the last term is asymptotically

$$sub(H, G(N, d)) = \frac{N(N-1)\dots(N-n+1)}{aut(H)}$$

the expected number of copies of H in the random graph on N vertices where every pair of distinct vertices is an edge with probability d (Erdős-Rényi random graph [5]).

We can also acquire a statement about the Turán number. For any graph H the Turán number $ex(N, H)$ is the maximum number of edges in a graph G on N vertices such that no subgraph is isomorphic to H. For a monograph on the topic consult [2]. If no subgraph is isomorphic to H then $sub(H, G) = 0$ and therefore we obtain from Sidorenko's conjecture $d^e \leq O\left(\frac{1}{N}\right)$. Using $d = \frac{2E}{N^2}$ we get $E = O(N^{\frac{2e-1}{e}})$ and since G was an arbitrary graph on N vertices

$$ex(N, H) = O(N^{2-\frac{1}{e}}).$$

This bound is to weak to give us anything new about any class of graphs, because an upper bound proven by Kövari, Sós and Turán is $O(N^{2-\frac{1}{s}})$ where s is the size of the smaller class in a bipartition of H.

1.2 First example

We do a first attempt to lower bound $t(H, G)$ by expressing it as the expected value of a random variable on the set of all maps with the uniform distribution. Let $w(i, j) = 1$ if and only if $\{i, j\} \in E(G)$ and 0 otherwise, then the expected value with respect to the uniform distribution is the edge density of G

$$\mathbb{E}_{[N]^2}(w) = \sum_{i,j \in [N]} \frac{1}{N^2} w(i, j) = \frac{2E}{N^2} = d.$$

Next we define the random variable $W : [N]^n \to \{0, 1\}$ by

$$W(a) = \prod_{\{i,j\} \in E(H)} w(a(i), a(j))$$

and since $W(a) = 1$ if and only if $a \in Hom(H, G)$ we get

$$\mathbb{E}_{[N]^n}(W) = \sum_{a \in [N]^n} \frac{1}{N^n} W(a) = \frac{Hom(H, G)}{N^n} = t(H, G).$$

Lemma 1.2. The star on n vertices with edges $E(H) = \{\{1, i\} : i = 2, \ldots, n\}$ as shown in Figure 1 satisfies Sidorenko's conjecture.

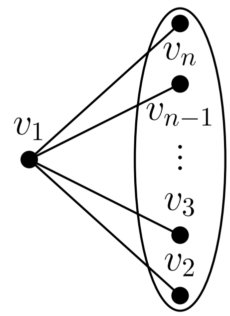

Figure 1: The star on n vertices

Proof.

$$\begin{aligned}
t(H,G) = \mathbb{E}_{[N]^n}(W) &= \sum_{a\in[N]^n} \frac{W(a)}{N^n} \\
&= \sum_{a\in[N]^n} \frac{1}{N^n} \prod_{i=2}^{n} w(a(1),a(i)) \\
&= \sum_{a(1)\in[N]} \frac{1}{N^n} \sum_{a(2),\ldots,a(n)\in[N]^{n-1}} \prod_{i=2}^{n} w(a(1),a(i)) \\
&= \sum_{a(1)\in N} \frac{1}{N} \prod_{i=2}^{n} \sum_{a(i)\in N} \frac{1}{N} w(a(1),a(i)) \\
&= \sum_{a(1)\in N} \frac{1}{N} \left(\sum_{a'\in N} \frac{1}{N} w(a(1),a') \right)^{n-1} \\
&\geq \left(\sum_{a(1)\in N a'\in N} \frac{1}{N^2} w(a(1),a') \right)^{n-1} \\
&= \left(\mathbb{E}_{[N]^2}(w)\right)^{n-1} = d^e = t(K_2,G)^e.
\end{aligned}$$

The inequality comes from an easy application of Jensen's inequality to the convex function $z = z^{n-1}$, which we will explain later in Section 1.4. We proved that the star satisfies Sidorenko's conjecture. □

1.3 More about the conjecture

Long before the conjecture was stated the inequality was proven for paths by Blackley and Roy [1] in 1965. In the 1990s Sidorenko stated the conjecture and proved it for trees, even cycles, complete bipartite graphs and bipartite graphs with one class of size at most three [10]. Hatami proved the conjecture for cubes

[7]. After that Conlon, Fox and Sudakov [4] proved that all bipartite graphs with one vertex complete to the other side are Sidorenko, that is there exists a bipartition $V(H) = A\dot{\cup}B$ of H into two partitesets and a vertex $a \in A$ such that for all $b \in B$ we have $\{a, b\} \in E(H)$.

In the paper from 2011 by Li and Szegedy [9] which we will analyse in Section 2 they give a recursive process for constructing new Sidorenko graphs out of old ones. The method gives us the possibility to glue certain graphs together on trees and apply a reflection operation to some subgraphs. This includes trees, even cycles and the bipartite graphs where one vertex is complete to the other side. At the end of the section we will give a slight extension which allows us to glue on tree like graphs containing some double-stars.

Two years later Kim, Lee and Lee [8] published two approaches to the conjecture. The first one is partly extending Li's and Szegedy's method to tree-arrangeable bipartite graphs, where a bipartite graph is called tree-arrangeable if the neighbours of one partiteset have a certain tree-like structure. We explain this at the end of Section 2. Secondly they show that the Cartesian product $T \times H$ of a tree T and a Sidorenko graph H is again Sidorenko, which we will state more precisely in Section 5.3. At the end we will partially answer a question of Szegedy in Section 5.5 about the relation of this paper to his new method from [12].

This method is presented in the very recent paper by Szegedy [12], where he extends the ideas of his previous one and gives us a method that allows us to glue graphs on independent sets and forests. It uses the relative entropy of probability distributions on graph homomorphisms and couplings. Roughly speaking the method gives us three classes of graphs $\mathfrak{G}_1 \subseteq \mathfrak{G}_2 \subseteq \mathfrak{G}$. Starting from a single edge each of them is obtained via iteratively gluing on independent sets ($\mathfrak{G}_1$), forests ($\mathfrak{G}_2$) or any graph ($\mathfrak{G}$) under certain additional conditions. The main theorem says that all graphs in $\mathfrak{G}_2$ satisfy Sidorenko's conjecture. We give a detailed discussion of the method, with basic examples and analyse all three sets in Section 4.

In Section 5 we give various new applications for the method. First we verify the statement by Sidorenko, that all bipartite graphs with one class of size at most 4 satisfy the conjecture. With a little bit more refined calculation we even extend the method beyond $\mathfrak{G}_2$ as suggested by Szegedy in [12]. In particular we reproduce Hatami's result and prove Sidorenko's conjecture for all cubes Q_d as Szegedy suggested in private communication. Furthermore we prove that the 1-subdivision of K_m is Sidorenko for all m which was, as far as we know, not proven before.

The smallest graph for which the conjecture is not know is a $K_{5,5}$ minus a 10-cycle or equivalently a Möbious ladder of length 5 as shown in Figure 2. We prove that this graph is not in $\mathfrak{G}$ and therefore this method cannot give us anything new in this direction.

Originally Sidorenko's conjecture was formulated in a more general setting. Let $g : [0, 1]^2 \to \mathbb{R}$ be a bounded non-negative measurable function then with

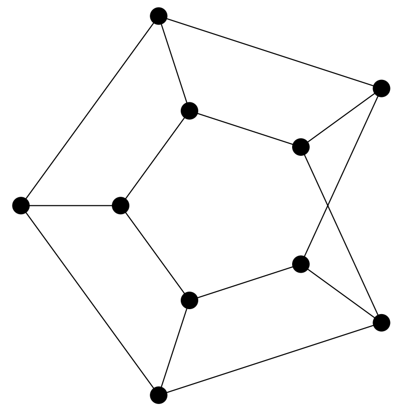

Figure 2: The $K_{5,5}$ minus C_{10} or equivalently möbius ladder of length 5.

the Lebesgue measure

$$t(H,g) := \mathbb{E}\Big(\prod_{\{i,j\}\in E(H)} g(x_i,x_j)\Big) = \int_{[0,1]^n} \prod_{\{i,j\}\in E(H)} g(x_i,x_j) \prod_{i=1}^{n} dx_i.$$

Now we can state the orginial version of Sidorenko's conjecture

Conjecture 1.3. For all bipartite graphs H and for all bounded non-negative measurable functions g we have

$$t(H,g) \geq t(K_2,g)^e.$$

Conjecture 1.3 implies Conjecture 1.1. For this take any graph G and define $g(x,y) = 1$ if and only if $\{\lceil x\cdot n\rceil, \lceil y\cdot n\rceil\} \in E(H)$ and zero otherwise. This can be seen as the adjacency matrix of G extended to $[0,1]^2$. Now every homomorphisms from H to G corresponds precisely to a cube of measure N^{-n} in $[0,1]^n$ and thus $t(H,g) = |Hom(H,G)|N^{-n} = t(H,G)$ for all H.

Another related concept is the study of quasi-random graphs and the forcing property. A sequence of graphs $\{G\}_{n\in\mathbb{N}}$ with edge density d is called quasi-random if for the 4-cycle C_4

$$t(C_4,G_n) = (1+o(1))d^e,$$

which is one of many equivalent properties shared by quasi random graphs [3]. Any graph H with which we can replace C_4 and get an equivalent condition for quasi-randomness is called forcing. In [4] it is stated as a conjecture that all bipartite graphs which contain a cycle are forcing. Equivalently a graph H which contains a cycle is forcing if the equality case in Conjecture 1.3 is only achieved by the constant function, i.e.

$$g \text{ constant } \Leftrightarrow t(H,g) = t(K_2,g)^e.$$

We do not give a proof for this here, because it would deviate to much.

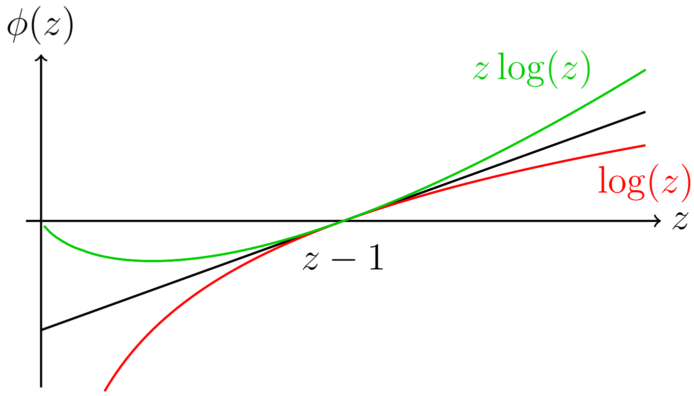

Figure 3: Plot of $z\log(z)$ and $\log(z)$ with tangent $z-1$ in 1.

1.4 Finite probability spaces and Jensen's inequality

All of the following chapters will use concavity and convexity of functions, especially Jensen's inequality. Let (X,μ) be a **finite probability space** consisting of a finite set X and a probability measure μ. A map $f : X \to \mathbb{R}$ with $\mathbb{E}_\mu(f) = 1$ is called a **density** on (x,μ). If not stated otherwise in this thesis all probability spaces are finite and therefore we do not need to worry about sigma-algebras.

For any interval C a real function $\phi : C \to \mathbb{R}$ is called convex (concave) if for all $c_1, c_2 \in C$ and for all $t \in [0,1]$ we have $\phi(tc_1+(1-t)c_2) \leq t\phi(c_1)+(1-t)\phi(c_2)$ (or with $\geq$ respectively). Examples are $z\log(z)$ for a convex and $\log(z)$ for a concave function as shown in Figure 3. Now we can state Jensen's inequality for probability spaces.

Lemma 1.4. For an interval $C \subseteq \mathbb{R}$ let $\phi : C \to \mathbb{R}$ be a convex (concave) function, $W : X \to C$ a random variable and f a density function on (X,μ), i.e. $\mathbb{E}_\mu(f) = 1$. Then

$$\begin{array}{ll} \text{convex } \phi & \text{concave } \phi \\ \mathbb{E}_\mu(\phi(W)) \geq \phi(\mathbb{E}_\mu(W)) & \mathbb{E}_\mu(\phi(W)) \leq \phi(\mathbb{E}_\mu(W)) \qquad (1) \\ \mathbb{E}_\mu(f\phi(W)) \geq \phi(\mathbb{E}_\mu(f\,W)) & \mathbb{E}_\mu(f\phi(W)) \leq \phi(\mathbb{E}_\mu(f\,W)). \qquad (2) \end{array}$$

In most of the applications the measure μ will be the uniform distribution and we just leave it.

Proof. We only proof the convex case. Let $z_0 = \mathbb{E}_\mu(W) \in C$, then for convex c there exist $a, b \in \mathbb{R}$ such that

$$\begin{aligned} az + b &\leq \phi(z)\,\forall z \in \mathbb{R} \\ az_o + b &= \phi(z_0). \end{aligned}$$

This implies $\phi(W(x)) \geq aW(x) + b$ for all $x \in X$. Using the linearity of expectation we get

$$\begin{aligned}\mathbb{E}_\mu(\phi(W)) &\geq \mathbb{E}_\mu(aW + b) \\ &= a\mathbb{E}_\mu(W) + b = az_0 + b \\ &= \phi(z_0) = \phi(\mathbb{E}_\mu(W)).\end{aligned}$$

The second part follows from the first one by converting μ to a new measure μ' on X via the density f. Alternatively for $z_0 = \mathbb{E}_\mu(f\,W)$ and the rest as above we get

$$\begin{aligned}\mathbb{E}_\mu(f\phi(W)) &\geq \mathbb{E}_\mu(f(aW + b)) \\ &= a\mathbb{E}(f\,W) + \mathbb{E}_\mu(f)b = az_0 + b \\ &= \phi(z_0) = \phi(\mathbb{E}_\mu(f\,W)).\end{aligned}$$

□

2 The logarithmic Calculus

This chapter builds on the first paper by Li and Szegedy [9] on the logarithmic calculus and is not essential for the understanding of the following chapters. This paper was formulated with the original version of the conjecture, but we state it in the more combinatorial setting. The logarithmic calculus can be considered as a symbolic way of proving inequalities between subgraph densities using conditional expectation and Jensen's inequality for $z = \log(z)$ and $z = z\log(z)$.

2.1 Basic examples

As a first example for the method we will now prove that the star on n vertices H is Sidorenko for a second time. For a n-variable function $h(x_1, \ldots, x_n)$ on $V(G)^n$ and a subset of the variables $S = \{x_1, \ldots, x_s\}$ we define the conditional expectation as the s-variable function $\mathbb{E}_S(h)$ on $V(G)^{n-s}$ defined by

$$\mathbb{E}_S(h)(x_1, \ldots, x_s) = \sum_{\{x_{s+1}, \ldots, x_n\} \in V(G)^{n-s}} N^{s-n} h(x_1, \ldots, x_n).$$

Note that $\mathbb{E}(h) = \mathbb{E}(\mathbb{E}_S(h))$. For ease of notation we identify the vertices of the graph and the variables. Recall the definition of $w(x_i, x_j)$ as being 1 if $\{x_i, x_j\}$ is an edge and 0 otherwise, $d = \mathbb{E}(w)$ and the product over all edges is W. Then the expected degree of a vertex is $d(x) = \mathbb{E}_x(w(x, y))$. We define a density function for the star

$$f_n = \prod_{i=2}^{n} w(x_1, x_i) d^{-1} d(x_1)^{2-n}$$

on $Hom(H, G)$ and extend it to $V(G)^{V(H)}$ by setting all new values to zero. Subsequently we will always do this and treat all non-defined values as zero.

Claim. We have $\mathbb{E}_{x_1}(f) = d(x_1)/d$ and thus $\mathbb{E}(f) = 1$, i.e. f defines indeed a density function.

Proof. For $n = 1$ the statement is clear. Assume the statement is true for $n-1$, then

$$\begin{aligned}\mathbb{E}_{x_1}(f_n) &= \mathbb{E}_{x_1}(\mathbb{E}_{\{x_1,\dots,x_{n-1}\}}(f_n)) \\ &= \mathbb{E}\left(f_{n-1}\mathbb{E}_{\{x_1,\dots,x_{n-1}\}}\left(\frac{w(x_1,x_n)}{d(x_1)}\right)\right) = \mathbb{E}(f_{n-1}) = \frac{d(x_1)}{d}.\end{aligned}$$

□

Throughout the following arguments it will always play a key-role which functions depend on which variables and which expected value we can pull out. We can now reprove Sidorenko's conjecture for the star.

Proof of Lemma 1.2. Taking the logarithm on both sides of the inequality of the conjecture we have to prove the following for all G

$$\log(t(H,W)) \geq e\log(d). \tag{3}$$

Then we calculate

$$\begin{aligned}\log(t(H,W)) &= \log(\mathbb{E}(\prod_{i=2}^{n} w(x_1,x_i))) \\ &= \log(\mathbb{E}(f d\, d(x_1)^{n-1})) \geq \mathbb{E}(f\log(d\, d(x_1)^{n-1})) \\ &= \mathbb{E}(f\log(d)) + (n-1)\mathbb{E}(f\log(d(x_1))) \\ &= \log(d)\mathbb{E}(f) + (n-1)\mathbb{E}(\mathbb{E}_{x_1}(f\log(d(x_1)))) \\ &= \log(d) + (n-1)\mathbb{E}(\log(d(x_1))\mathbb{E}_{x_1}(f)) \\ &= \log(d) + (n-1)d^{-1}\mathbb{E}(d(x_1)\log(d(x_1))) \\ &\geq \log(d) + (n-1)d^{-1}\mathbb{E}(d(x_1))\log(\mathbb{E}(d(x_1))) \\ &= \log(d) + (n-1)d^{-1}d\log(d) = n\log(d) = e\log(d),\end{aligned}$$

where we used Jensen's inequality Equation (2) with density f and the concavity of $z = \log(z)$ for the first estimate and Equation (1) with the convexity of $z = z\log(z)$ for the second. □

We can easily extend this proof to trees via the following density defined for a tree T on n vertices by

$$f_T = d^{-1}\prod_{i=1}^{n} d(x_i)^{1-\deg_T(x_i)} \prod_{\{x_i,x_j\}\in E(T)} w(x_i,x_j),$$

which can be seen as choosing one edge uniformly at random an then building the tree starting from this edge vertex by vertex.

Claim. The function f_T is indeed a density.

Proof. We prove by induction that $\mathbb{E}_{x_i}(f_T) = d(x_i)/d$. For $n = 1$ it is clear. Assume $n > 1$ and that the statement is true for $n - 1$. Consider a leaf $x_j \neq x_i$, let $S = \{x_1, \dots, x_n\} \setminus \{x_j\}$ and let T' be the tree obtained from T by deleting x_j. Then

$$\mathbb{E}_{x_i}(f_T) = \mathbb{E}_{x_i}(\mathbb{E}_S(f_T)) = \mathbb{E}_{x_i}(f_{T'}) = \frac{d(x_i)}{d}.$$

□

A more complex class of bipartite graphs are those where one vertex is complete to the other side, which were first considered by Conlon, Fox and Sudakov [4]. Let $V(H) = \{x, y_1, \dots, y_m, v_1, \dots, v_k\}$ be the vertices and x connected to $v_1, \dots, v_m$ and every y_t to a subset $S_t \subseteq \{v_1, \dots, v_m\}$. With $a_t = |S_t|$ we have for the number of edges $e = k + \sum_{t=1}^{m} a_t$.

Theorem 2.1. The graphs H defined above satisfies the Sidorenko conjecture.

Proof. We define a density f similar to the of the star by

$$f = \prod_{i=1}^{k} w(x, v_i) d^{-1} d(x)^{1-k}.$$

and

$$s_i = \mathbb{E}_{S_t}\Big(\prod_{v_j \in S_t} w(z, v_j)\Big).$$

Then we want to show Equation (3) for H and therefore we start with

$$\log(t(H, W)) = \log\Big(\mathbb{E}\big(f d(x)^{k-1} d \prod_{i=1}^{n} s_i\big)\Big)$$

because for every y_t only the neighbours S_t share any dependence we can move the expected value inside. Using the same calculations as in the last proof we immediately get

$$\log\Big(\mathbb{E}\big(f d(x)^{k-1} d \prod_{i=1}^{n} s_i\big)\Big) \geq k \log(d) + \sum_{t=1}^{m} \mathbb{E}(f \log(s_i)).$$

It remains to consider all terms of the last sum. Define a new density

$$f_t = s_t^{-1 d(x)^{a_t - k}} \prod_{i=1}^{k} w(x, v_i)$$

and $h_t = s_t d(x)^{1-a_t}$. Observe that then $f = d^{-1} f_t h_t$ and using $s_t = h_t d(x)^{a_t - 1}$ we get

$$\begin{aligned}
&\mathbb{E}(f_t \log(s_t)) = d^{-1}\mathbb{E}(f_t h_t \log(h_t)) + (a_t - 1)d^{-1}\mathbb{E}(\mathbb{E}_x(f_t h_t \log(d(x))))\\
\geq &d^{-1}\left(\mathbb{E}(f_t h_t \log(\mathbb{E}(f_t h_t))) + (a_t - 1)\mathbb{E}(d(x)\log(d(x)))\right)\\
= &a_t \log(d)
\end{aligned}$$

using Equation (2) with density f_t and the convexity of $z = z\log(z)$ for the first term and Equation (1) with $z = z\log(z)$ for the second. Together we get

$$\log(t(H, W)) \geq (k + \sum_{t=1}^{m} a_t)\log(d) = e\log(e).$$

□

A similar argument is possible for even cycles. The interesting observation is that in fact we obtain stronger inequalities

$$\mathbb{E}(d^{-1}d(x)^{k-1}\prod_{i=1}^{k} w(x_i, v_k)\log(s)) \geq |S|\log(d).$$

for any choice of neighbours $S \subseteq \{v_1, \ldots, v_n\}$ for a newly added vertex. This allows us us to glue various sub-stars onto the original star. This indicates that inequalities of this type can be used to produce new Sidorenko graphs by gluing.

2.2 Smoothness and gluing

Let $\mathcal{G}_m$ be the set of graphs in which m different vertices are labelled by the numbers $\{1, \ldots, m\}$. If H_1 and H_2 are in $\mathcal{G}_m$ then there product $H_1 H_2$ is defined as the graph obtained by identifying vertices with the same labels and reducing multiple edges. For a graph $H \in \mathcal{G}_m$ we define the **restricted subgraph density** in G as

$$t_S(H, G) = \mathbb{E}_S(\prod_{\{i,j\}\in E(H)} w(x_i, x_j)).$$

Definition 2.2. Let $H \in \mathcal{G}_m$ be a bipartite graph on m vertices such that the spanned subgraph on the labelled vertices S is a tree T. We say that H is smooth if

$$\mathbb{E}(f_T \log(t_S(H^*, G))) \geq |E(H^*)|\log(d)$$

where H^* is obtained from H by deleting the edges in T.

For an empty tree the statement is equivalent to Sidorenko's conjecture. An easy example for a smooth graph are two edges $\{x, y\}, \{y, z\}$, were only two neighbouring vertices x, y are labelled.

$$\begin{aligned}
&\mathbb{E}(w(x,y)d^{-1}\log(\mathbb{E}_{x,y}(w(y,z)))) \\
&\quad = d^{-1}\mathbb{E}(w(x,y)d(y)^{-1}d(y)\log(d(y))) \\
&\quad \geq d^{-1}\mathbb{E}(w(x,y)d(y)^{-1}d(y))\log(\mathbb{E}(w(x,y)d(y)^{-1}d(y))) \\
&\quad = d^{-1}d\log(d) = \log(d),
\end{aligned}$$

where we used Jensen's inequality with the convexity of $z = z\log(z)$ and the density $w(x,y)d(y)^{-1}$. The next two lemmas together show that smoothness is a strengthening of the Sidorenko property.

Lemma 2.3. Let $H \in \mathcal{G}_m$ be a smooth bipartite graph with tree T spanned on the labelled vertices and let T' be a non-empty sub-tree spanned on a subset of the labelled vertices. Then the graph H' obtained from H by unlabelling the vertices not in T' is again smooth.

Proof. It is enough to prove that unlabelling one leaf in H preserves smoothness. The general case is then an iteration of this step. Assume that x_m is connected to x_{m-1} and $S' = S \setminus \{x_m\}$. We have

$$\begin{aligned}
&\mathbb{E}(f_{T'}\log(t_{S'}(H'^*, W))) = \mathbb{E}(f_{T'}\log(\mathbb{E}_{x_m}(w(x_{m-1},x_m)t_S(H^*,W)))) \\
&\quad = \mathbb{E}(f_{T'}\log(\mathbb{E}_{x_m}(w(x_{m-1},x_m)d(x_{m-1})^{-1}d(x_{m-1})t_S(H^*,W)))) \\
&\quad = \mathbb{E}(f_{T'}\log(d(x_{m-1}))) + \mathbb{E}(f_{T'}\log(w(x_{m-1},x_m)d(x_{m-1})^{-1}t_S(H^*,W))) \\
&\quad \geq \mathbb{E}(\mathbb{E}_{x_{m-1}}(f_{T'}\log(d(x_{m-1})))) \\
&\qquad + \mathbb{E}(f_{T'}w(x_{m-1},x_m)d(x_{m-1})^{-1}\log(t_S(H^*,W))) \\
&\quad = d^{-1}\mathbb{E}(d(x_{m-1})\log(d(x_{m-1}))) + \mathbb{E}(f_T\log(t_S(H^*,W))) \\
&\quad \geq \log(d) + |E(H^*)|\log(d) = |E(H'^*)|\log(d),
\end{aligned}$$

where we used Jensen's inequality with density $w(x_{m-1},x_m)d(x_{m-1})^{-1}$ and the concavity of $z = \log(z)$ for the first estimate and the same arguments as in the basic examples for the second. □

Lemma 2.4. Assume that $\{x_1, x_2\} \in E(H)$ and $H \in \mathcal{G}_2$ is smooth. Then H is Sidorenko.

Proof. Using what we did for the previous graphs we get

$$\begin{aligned}
&\log(\mathbb{E}(t(H,W))) = \log(\mathbb{E}(dw(x_1,x_2)d^{-1}\prod_{\{x_i,x_j\}\in E(H^*)} w(x_i,x_j))) \\
&\quad \geq \log(d) + \mathbb{E}(w(x_1,x_2)d^{-1}\log(t(H^*,W))) \\
&\quad \geq \log(d) + |E(H^*)|\log(d) = |E(H)|\log(d).
\end{aligned}$$

Since this is Equation (3) for H this graph is Sidorenko. □

Preserving smoothness we can now unlabel vertices until only two are left and the resulting graph will be Sidorenko. Therefore these two lemmas imply that every smooth graph satisfies the Sidorenko conjecture. The next lemma extends the idea from the proof of Theorem 2.1. We want to glue graphs on there labelled vertices preserving smoothness and therefore the Sidorenko property.

Lemma 2.5. Let $H_1, H_2 \in \mathcal{G}_m$ be two smooth graphs such that the trees spanned on the labelled vertices S are identical with T in both graphs. Then $H = H_1 H_2 \in \mathcal{G}_m$ is again smooth.

Proof.

$$\begin{aligned}
&\mathbb{E}(f_T(t_S(H^*, W))) = \mathbb{E}(f_T \log(t_S(H_1^*, W) t_S(H_2^*, W))) \\
&\quad = \mathbb{E}(f_T \log(t_S(H_1^*, W))) + \mathbb{E}(f_T \log(t_S(H_2^*, W))) \\
&\quad = (|E(H_1)| - m + 1) \log(d) + (|E(H_2)| - m + 1) \log(d) \\
&\quad = (|E(H)| - m + 1) \log(d) = |E(H^*)| \log(d).
\end{aligned}$$

□

We can also extend the smooth part by adding new edges to the tree induced on the labelled vertices. Starting with the two edges from the beginning, where one was labelled, this gives us that all trees are smooth if at least on leaf is not labelled and thus another proof that trees are Sidorenko.

Lemma 2.6. Let $H \in G_m$ be a smooth graph and let $T' \in G_m$ be a tree such that H and T' induce the same tree T on the labelled vertices S. Then the graph $H' = HT'$ is again smooth and all vertices of T' are labelled.

Proof. It suffices to prove the statement for the case where a single edge is added. Let $\{y\} = V(T') \setminus V(T)$ be a leaf in T' and assume that y is connected to x_m. Then $S' = S \cup \{y\}$ and

$$\begin{aligned}
&\mathbb{E}(f_{T'} \log(t_{S'}(H'^*, W))) = \mathbb{E}(\mathbb{E}_S(f_{T'} \log(t_{S'}(H'^*, W)))) \\
&= \mathbb{E}(f_T \mathbb{E}_S(\log(t_{S'}(H'^*, W)) w(y, x_m) d(x_m)^{-1})).
\end{aligned}$$

We get rid of the last two terms by taking the expectation over y and then

$$\begin{aligned}
&= \mathbb{E}(f_T \log(t_{S'}(H'^*, W))) \\
&= \mathbb{E}(f_T \log(t_S(H^*, W))) \\
&\geq |E(H^*)| \log(d) = |E(H'^*)| \log(d),
\end{aligned}$$

because H' and H'^* have the same edges of which none contains y. □

Now we want to prove that we can reflect a tree on an independent set and obtain a smooth graph.

Lemma 2.7. Let $T \in \mathcal{G}_m$ be a tree such that the labelled points S are independent. Let H be the graph obtained from T^2 by labelling all vertices in one copy of T. Then H is smooth.

Proof. H is a graph on vertices $\{x_1, \ldots, x_{2n-m}\}$. Let T be a tree on vertices $N = \{x_1, \ldots, x_n\}$ and the labelled vertices $S = \{x_1, \ldots, x_m\}$. Observe that H^* has the same edges as T. Then it suffices to exclude the vertices in S for $t_S(T, W)$ and we have to prove the following statement

$$\mathbb{E}(f_T \log(t_S(T, W))) \geq |E(T)| \log(d).$$

Let $s = t_s(T, W)$ and $q = \prod_{i=1}^n d(x_i)^{\deg_H(x_i)-1}$, then

$$\begin{aligned}
&\mathbb{E}(f_t \log(s)) \\
&\quad = d^{-1}(\mathbb{E}(t_N(T, W)s^{-1}(sq^{-1}) \log(sq^{-1})) + \mathbb{E}(t_N(T, W)s^{-1}sq^{-1} \log(q))) \\
&\quad \geq \log(d) + \sum_{i=1}^n d^{-1}(\deg_H(x_i) - 1)\mathbb{E}(\mathbb{E}_{x_i} t_N(T, W)s^{-1}sq^{-1} \log(d(x_i))) \\
&\quad = \log(d) + \sum_{i=1}^n d^{-1}(\deg_H(x_i) - 1)\mathbb{E}(d(x_i) \log(d(x_i))) \\
&\quad \geq \log(d) + \sum_{i=1}^n (\deg_H(x_i) - 1) \log(d) = (n - 1) \log(d).
\end{aligned}$$

□

To apply this last lemma we can start with any tree and chose an independent set. Taking a path of length m such that the two endpoints are labelled and reflecting it we get that even cycles C_{2m} are smooth and therefore Sidorenko. Furthermore we get that the path of length m is smooth in C_{2m} and therefore we can glue two even cycles together on any path up to half the length of the smaller cycle using Lemma 2.7. Another application are the bipartite graphs were one vertex is complete to the other side. We obtain smooth graphs by reflecting sub-stars and get the hole graph by gluing everything together.

In the next section we will describe an extension of smoothness beyond trees. Later we will see a method that allow us to glue on forests and even more general graphs.

2.3 Extension of the method

In the paper of Li and Szegedy all the previous definitions and lemmas were stated for graphs $H \in \mathcal{G}_m$ which induce a tree T on the labelled vertices. We started with choosing one random edge and then building the tree. In principal this process is possible because $d(x)$ precisely gives us the proportion of the vertices we can take for the next neighbour of x. Especially there always is a neighbour for x because x was chosen as the neighbour of some earlier vertex.

We want to extend this idea to stars. If some vertex x has r neighbours S or more then we can chose a new neighbour for this r vertices. As in the proof of Theorem 2.1 this corresponds to $s = \mathbb{E}_S(\prod_{v \in S} w(x, v))$. We can include this into the density function of the base graph, which we then can glue with other graphs or apply the reflection operation.

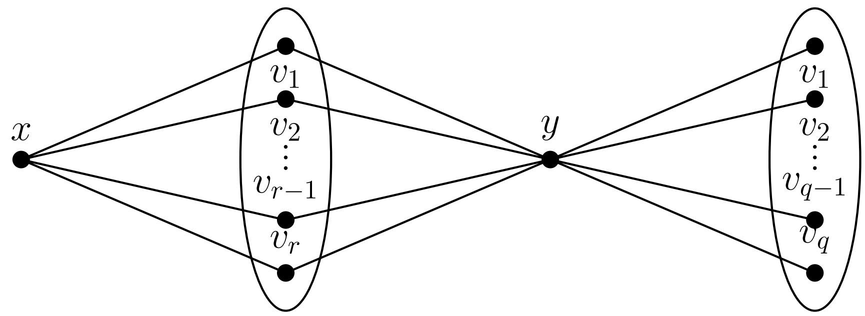

Figure 4: $T_{r,q}$

So we are able to slightly extend the definition of smoothness to graphs $H \in \mathcal{G}_m$ which induce a graph build in the following way: Start with one edge and then add as much neighbours as you want. While building this every time a vertex has at least k neighbours you can give these k vertices another neighbour. To avoid to much complexity in the expression and proofs we give one example for this extension.

Let $H \in \mathcal{G}_m$ be a bipartite graph on m vertices such that the spanned subgraph on the labelled vertices S is a graph $T_{r,q}$ consisting of a vertex x with r neighbours $V = \{v_1, \ldots, v_r\}$ which have another neighbour y which has q neighbours $\{w_1, \ldots, w_q\}$ as shown in Figure 4. We get the following density for $T_{r,q}$

$$f_{r,q} = \prod_{i=1}^{r} w(x, v_i) w(y, v_i) \prod_{i=1}^{q} w(y, w_i) d^{-1} d(x)^{1-r} s_r^{-1} d(y)^{1-q}.$$

where $s_r = \mathbb{E}_V(\prod_{i=1}^{r} w(x, v_i))$. We extend the definition of smoothness 2.2 by the following:

Definition 2.8. Let $H \in \mathcal{G}_m$ with $T_{r,q}$ and S as above. Then we say that in addition to Definition 2.2 H is smooth if

$$\mathbb{E}(f_R \log(t_S(H^*, G))) \geq |E(H^*)| \log(d)$$

where H^* is obtained from H by deleting the edges in R.

Lemmas 2.4 and 2.5 hold with the same proof as before. For Lemmas 2.3, 2.6 and 2.7 we have to take care of the role of y. As a demonstration we want to show first how unlabelling the vertex y works.

Lemma 2.9. For this let H be a graph which induces the graph $T_{r,0}$ on the labelled vertices S with the density

$$f_{r,0} = \prod_{i=1}^{r} w(x, v_i) w(y, v_i) d^{-1} d(x)^{1-r} s_r^{-1}$$

where y is labelled. After unlabelling we get the graph H' with a Star S_r induced on the labelled vertices $S' = S \setminus \{y\}$ and density

$$f = \prod_{i=1}^{r} w(x, v_i) d^{-1} d(x)^{1-r}.$$

Then H' is smooth.

Proof. We compute

$$\begin{aligned}
&\mathbb{E}(f \log(t_{S'}(H'^*, G))) = \mathbb{E}(f \log(\mathbb{E}_{S'}(\prod_{i=1}^{r} w(y, v_i) t_S(H^*, G))) \\
&= \mathbb{E}(f \log(\mathbb{E}_{S'}(\prod_{i=1}^{r} w(y, v_i) s_r^{-1} s_r \, t_S(H^*, G))) \\
&\geq \mathbb{E}(f \prod_{i=1}^{r} w(y, v_i) s_r^{-1} \mathbb{E}_{S'}(\log(s_r \, t_S(H^*, G))) \\
&= \mathbb{E}(f_{r,0} \log(s_r)) + \mathbb{E}(f_{r,0} \log(t_S(H^*, G))) \\
&\geq r \log(d) + \mathbb{E}(f_{r,0} \log(t_S(H^*, G))) \geq |E(H'^*)| \log(d),
\end{aligned}$$

where the first inequality comes from Jensen's inequality Equation 2 with density $\prod_{i=1}^{r} w(x, v_i) s^{-1}$ and the convexity of $z = \log(z)$ and the second is analogously to the proof of Theorem 2.1. □

Next we want to show how we can extend the labelled set by increasing r.

Lemma 2.10. Let $H \in \mathcal{G}_{r+q+2}$ be a smooth graph such that $T_{r,q}$ is the induced graph on the labelled vertices S and let $T_{r+1,q} \in \mathcal{G}_r + q + 2$ be such that $T_{r,q}$ is labelled. Then $H' = HT_{r+1,q}$ is again smooth and all vertices S' of $T_{r+1,q}$ are labelled.

Proof.

$$\begin{aligned}
&\mathbb{E}(f_{r+1,q} \log(t_{S'}(H'^*, G))) \\
&\quad = \mathbb{E}(f_{r+1,q} \mathbb{E}_{S'}(\log(t_{S'}(H'^*, G)))) \\
&\quad = \mathbb{E}\left(f_{r,q} \mathbb{E}_{S'}\left(\log(t_{S'}(H'^*, G)) \frac{s_r}{s_{r+1}} \frac{w(x, v_{r+1}) w(y, v_{r+1})}{d(x)}\right)\right),
\end{aligned}$$

where we get rid of $\frac{s_r w(q, v_{r+1})}{s_{r+1}}$ by taking the expectation over y and of $\frac{w(q, v_{r+1})}{d(x)}$ by taking it over v_{r+1}. Then

$$\begin{aligned}
&= \mathbb{E}(f_{r,q} \log(t_{S'}(H'^*, G))) \\
&= \mathbb{E}(f_{r,q} \log(t_S(H^*, G))) \\
&= |E(H^*)| \log(d) = |E(H'^*)| \log(d).
\end{aligned}$$

□

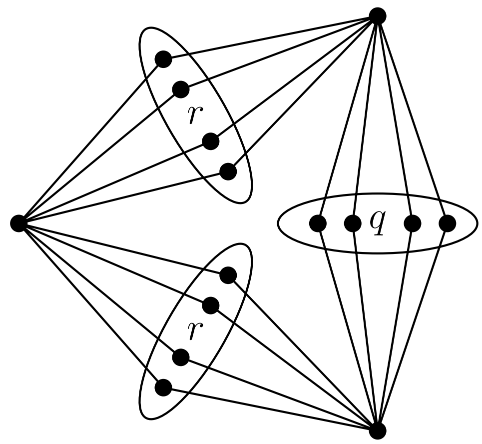

Figure 5: $T_{r,r,q}$

Lemma 2.7 extends in the natural way to $T_{r,q}$ and thus we can reflect it on the independent set consisting of x and the q neighbours of y giving us the smoothness of the graph $T_{r,r,q}$ shown in Figure 5 which was originally considered by Sidorenko in [10]. By Lemma 2.10 we can further increase the labelled part to get a graph with three vertices and distinct number of common neighbours. Unfortunately we are not able to produce all bipartite graphs with three vertices on one side with this method, because we can only add neighbours to the labelled part afterwords.

Note that our extension also works with more double-stars and thus gives us the possibility to prove Sidorenko's conjecture for graphs that were not considered up to this point like a sequence of double-stars, which we can also glue to a circle.

Another extension of the method by Li and Szegedy was given by Kim, Lee and Lee [8] to tree-arrangeable graphs. A bipartite graph H is called tree-arrangeable if there exists a bipartition $A\dot{\cup}B$ and a tree T on A such that for all vertices $u, v \in A$

$$N(u) \cap N(v) = \bigcap_{w \in P} N(w)$$

for any path P in T connecting u and v, where $N(u)$ is the set of neighbours of u in H. To proof that a graph is tree-arrangeable we need to find a tree T on an independent set A such that the vertices on a path have at least all the neighbours which both endpoints have. If we have only two vertices on one side then the graph is tree-arrangeable by just taking a single edge as T. They prove that all bipartite tree-arrangeable graphs are Sidorenko. We will prove this with the new method of Szegedy in Section 5.5.

Further examples for tree-arrangeable graphs are trees and bipartite graphs where one vertex is complete to the other side. For the latter one just take T

as the star centered at this vertex. This method really is an extension, because graphs with vertices $a_1, a_2 \in A$ such that $N(a) \subseteq N(a_i)$ for $i = 1$ or $i = 2$ are also tree-arrangeable, but this does not seem to follow from the method of Li and Szegedy. The tree that works for this example consists of an edge between $\{a_1, a_2\}$ and for every $a \neq a_1, a_2$ exactly one additional edge to a_1 if $N(a) \subseteq N(a_1)$ and to a_2 if not.

Cycles of length 6 and larger and the graphs in our extension, especially the graph in Figure 5, are in general not tree-arrangeable, because all pairwise neighbourhoods can be disjoint. Thus these two extensions go in different directions.

3 Entropy and graph homomorphisms

This chapter is devoted to prerequisites which we will need for the new method of Szegedy presented in Section 4.

3.1 Relative entropy

Entropy, a concept from information theory [6], is a measure for the uncertainty about the outcome of a random variable or the uniformity of a probability measure. The **entropy** of a measure μ on X is

$$H(\mu) = \mathbb{E}_\mu(-\log(\mu)) = -\sum_{x \in X} \log(\mu(x))\mu(x).$$

If $\mu(x) = 0$ then the corresponding summand is defined to be zero. This coincides with the limit $\lim_{x \to 0} x \log(x) = 0$. The entropy $H(\mu)$ is $\log(|X|)$ for the uniform distribution and 0 if one event has probability 1. The next lemma tells us that these are the extreme values. In principal entropy tells us how close a measure is to the uniform distribution.

Lemma 3.1. Properties of the entropy of μ:

(a) $H(\mu) \geq 0$

(b) Let $A = \{x \in X : \mu(x) \neq 0\} \subseteq X$ be the **support** of μ. Then

$$H(\mu) \leq \log(|A|).$$

Proof. (a) Using Jensen's inequality with Equation (1) for the convex function $z = -log(z)$ we get

$$H(\mu) = \mathbb{E}_\mu(-\log(\mu)) \geq -\log(\mathbb{E}_\mu(1)) = 0$$

(b) With Jensen's inequality Equation (1) for the concave function $z = \log(z)$ we get

$$H(\mu) = \mathbb{E}_\mu(\log(1/\mu)) \leq \log(\mathbb{E}_\mu(1/\mu)) = \log(|A|).$$

Remember that only elements in A appear with non-zero probability in the sum of the expected value.

□

Let μ and ν be two probability measures on X. The **cross entropy** of μ over ν is defined by

$$H(\mu, \nu) = \mathbb{E}_\mu(-\log(\nu)) = -\sum_{x \in X} \log(\nu(x))\mu(x).$$

If μ is absolutely continuous with respect to ν, i.e. $\nu(x) = 0$ implies $\mu(x) = 0$ we can define the **relative entropy** of μ with respect to ν as

$$D(\mu||\nu) = H(\mu, \nu) - H(\mu) = \mathbb{E}_\mu(\log(\mu/\nu)) = \sum_{x \in X} (\log \mu(x) - \log \nu(x))\mu(x),$$

where again every summand is defined to be zero, whenever $\mu(x)$ or $\nu(x)$ are zero.

Relative entropy is a measure for the loss of information if we approximate μ by ν. More precisely we will always set ν to be the uniform distribution on X, which implies that every measure on X is absolutely continuous with respect to ν. Then, in contrast to the normal entropy, the relative entropy is small when μ is close to the uniform distribution and large otherwise.

Lemma 3.2. Let $A \subseteq X$ be the support of μ, then

$$D(\mu||\nu) \geq -\log(\nu(A))$$

with equality if and only if $\mu(x) = \nu(x)/\nu(A)$ for all $x \in A$.

This implies

$$D(\mu||\nu) \geq 0 \tag{4}$$

for all μ and ν, because $-\log(\nu(A)) \geq 0$ for all $A \subseteq X$. In particular this says that the entropy of μ is always smaller than the cross entropy of μ with any ν.

Proof. Using Jensen's inequality Equation (1) for convexity of $z = -log(z)$ we get

$$\begin{aligned} D(\mu||\nu) &= \sum_{x \in A} (\log \mu(x) - \log \nu(x))\mu(x) = \sum_{x \in A} -\log\left(\frac{\nu(x)}{\mu(x)}\right)\mu(x) \\ &\geq -\log\left(\sum_{x \in A} \frac{\nu(x)}{\mu(x)}\mu(x)\right) = -\log(\nu(A)). \end{aligned}$$

The equality case in Jensen's inequality is achieved if and only if $\frac{\nu(x)}{\mu(x)} = c$ for all $x \in A$. Comparing both sides of the equation we get $c = \nu(A)$ as desired. □

3.2 Probability distributions of graph homomorphisms

Recall that $Hom(H,G) \subseteq V(G)^{V(H)}$ denotes the set of homomorphisms from H to G and $t(H,G) = \frac{|Hom(H,G)|}{N^n}$ denotes the probability that a random map $a : V(H) \to V(G)$ is a homomorphism.

Sidorenko's conjecture states that $t(H,G) \geq t(K_2,G)^e$, which is equivalent to

$$\log(t(H,G)) \geq |E(H)| \log(t(K_2,G)).$$

Let $\tau(H,G)$ be the uniform distribution on $Hom(H,G)$ and $\nu(H,G)$ the uniform distribution on $V(G)^{V(H)}$. We use the convention $D(\mu) := D(\mu||\nu(H,G))$ for any probability distribution μ on $V(G)^{V(H)}$. A probability distribution μ on $Hom(H,G)$ is extended to $V(G)^{V(H)}$ by setting all new values to zero.

Then the equality case of Lemma 3.2 implies $D(\tau(H,G)) = -log(t(H,G))$, because for $a \in Hom(H,G)$ we have

$$\tau(H,G)(a) = \frac{1}{|A|} = \frac{N^n}{N^n|A|} = \frac{\nu(H,G)(a)}{.}\nu(H,G)(A)$$

Let $D_e := D(\tau(K_2,G)) = -\log(t(K_2,G))$, then Sidorenko's conjecture for H is equivalent to the statement that

$$D(\tau(H,G)) \leq eD_e$$

holds for all G. By Lemma 3.2 every probability distribution μ on $Hom(H,G)$ satisfies

$$D(\tau(H,G)) = -\log(t(H,G)) = -\log(\nu(H,G)(Hom(H,G))) \leq D(\mu).$$

If μ also satisfies

$$D(\mu) \leq eD_e$$

then H satisfies Sidorenko's conjecture and we say that μ is a **witness** measure for H. Our goal is to construct a witness measure on $Hom(H,G)$ to prove Sidorenko's conjecture for H.

Another important probability distribution on $V(G)$ is the distribution κ where the probability of a vertex is proportional to it's degree. More precisely for $v \in V(G)$ we define $\kappa(v) = \frac{deg(v)}{2E}$. We will shortly write $D_v := D(\kappa)$. The role of κ is that it is the distribution of an end point of a uniformly chosen random edge.

3.3 Example

We look again at the proof of the star in the previous section using density and the random variable W. We now attempt to give a similar proof for the star going in the direction of the recent paper by Szegedy [12] using relative entropy and probability measures.

Proof of Lemma 1.2. Define a probability measure μ on $X = Hom(H, G)$. A homomorphism between H and G can be expressed by a vector of length $n = |V(H)|$ with entries in $V(G)$. For $(x_1, \dots, x_n) \in Hom(H, G)$ we define the measure

$$\mu((x_1, \dots, x_n)) = (2E)^{-1} \deg(x_1)^{2-n}$$

where $\deg(x_1)$ is always not zero because we started with a homomorphism. Summing over all $(x_1, \dots, x_n) \in Hom(H, G)$ immediately gives that this indeed defines a measure. Again this corresponds to a uniformly random chosen edge in G and $n-2$ further neighbours for one of the endpoints.

We compute

$$\begin{aligned} D(\mu) &= -\mathbb{E}_\mu(\log(\nu/\mu)) \\ &= -\mathbb{E}_\mu\left(\log\left(\frac{2E}{N^2} \cdot \left(\frac{\deg(x_1)}{N}\right)^{n-2}\right)\right) \\ &= -\mathbb{E}_\mu\left(\log\left(\frac{2E}{N^2}\right)\right) + (n-2)\mathbb{E}_\mu\left(\log\left(\frac{\deg(x_1)}{N}\right)\right). \end{aligned}$$

The second expected value only depends on x_1 so we can change it to the measure $\mu_1(x_1) = \frac{deg(x_1)}{2E}$ over $V(G)$ and get

$$\begin{aligned} &= -\log\left(\frac{2E}{N^2}\right) - (n-2)\mathbb{E}_{\mu_1}\left(\log\left(\frac{\deg(x_1)}{N}\right)\right) \\ &\leq -\log(d) - (n-1)\log\left(\frac{2E}{N^2}\right) = -e\log(t(K_2, G)) = eD_e. \end{aligned}$$

Thus μ is a witness measure and the star is Sidorenko. □

The same calculation work for any tree T on vertices $v_1, \dots, v_n$ with the measure

$$\mu_T((x_1, \dots, x_n)) = (2E)^{-1} \prod_{i=1}^{n} \deg_G(x_i)^{1-\deg_H(v_i)}.$$

As the density f_T this can be again seen as choosing one random edge and then building the tree starting from this edge. We found some witness measures for graphs. In the next chapter we describe a method to couple two probability measures preserving the Sidorenko property.

4 Coupling method

This chapter deals with the latest results by Szegedy [12] on Sidorenko's conjecture, which gives us the possibility to prove the conjecture for various new types of graphs. We want to iteratively construct witness measures using couplings. We first need some machinery that will allow us to combine probability distributions.

Let $\psi : X \to Y$ be a map between two probability spaces (X, μ) and (Y, ν). We can define a new measure μ' on Y by

$$\mu'(A) = \mu(\psi^{-1}(A)) \text{ for all } A \subseteq Y.$$

It is called the **marginal distribution** of (X, μ) on y with respect to ψ. If $\mu' = \nu$ then ψ is called **measure preserving**, i.e. the measure of the pre image of every subset is the same as the original measure.

4.1 Conditionally independent couplings

Let $\{(X_i, \mu_i)\}_{i=1}^3$ be three finite probability spaces and $\{\psi_i : X_i \to X_3\}_{i=1,2}$ two measure preserving maps, i.e. we have $\mu_3(A) = \mu_1(\psi_1^{-1}(A)) = \mu_2(\psi_2^{-1}(A))$ for all $A \subseteq X_3$. We call X_3 a **joint factor** of the probability spaces (X_1, μ_1) and (X_2, μ_2). Now we define X_4 as the set of $(x_1, x_2) \in X_1 \times X_2$ satisfying $\psi_1(x_1) = \psi_2(x_2)$ and $\mu_3(\psi_1(x_1)) \neq 0$ or all together

$$X_4 = \{(x_1, x_2) \in X_1 \times X_2 : \psi_1(x_1) = \psi_2(x_2), \mu_3(\psi_1(x_1)) \neq 0\}.$$

Note that $\mu_3(\psi_1(x_1)) = 0$ implies $\mu_1(x_1) = 0$. Let $\{\pi_i : X_4 \to X_i\}_{i=1,2}$ be the projections onto X_1,X_2 respectively. A measure μ on X_4 is called a **coupling** of (X_1, μ_1) and (X_2, μ_2) over the joint factor X_3 if π_1 and π_2 are measure preserving on (X_4, μ). This is equivalent to saying that the marginals of μ on X_1 and X_2 are μ_1 and μ_2 respectively. For example the marginal of μ on X_1 is

$$\mu|_{X_1}(x_1) = \sum_{x_2 \in X_2 : (x_1, x_2) \in X_4} \mu((x_1, x_2)) = \mu(\pi_1^{-1}(x_1)) = \mu_1(x_1).$$

The composition of two measure preserving maps is again measure preserving and thus are $\psi_1 \circ \pi_1$ and $\psi_2 \circ \pi_2$ for a coupling μ. Since $\psi_1(x_1) = \psi_2(x_2)$ holds for all $(x_1, x_2) \in X_4$ we have in fact a commutative diagram as shown in Figure 6 and in particular

$$\psi_1 \circ \pi_1 = \psi_2 \circ \pi_2 = \psi_1 \times \psi_2$$

For our purposes we want a coupling μ_4 where the probability is independent of the choice for $x_3 \in X_3$. Later this will give us the possibility to identify two graphs on certain sets leaving the respective rests independent. Let $x_3 \in X_3$ and $(x_1, x_2) \in X_4$ such that $\psi_1(x_1) = \psi_2(x_2) = x_3$. The events $A = \pi_1^{-1}(\{x_1\})$ and $B = \pi_2^{-1}(\{x_2\})$ are called **conditionally independent** over the event $Y = (\psi_1 \circ \pi_1)^{-1}(\{\psi_1(x_1)\})$ if $\mu(A \cap B|Y) = \mu(A|Y)\mu(B|Y)$. Assuming that μ_4 is a coupling we get $\mu_4(A) = \mu_1(x_1)$, $\mu_4(B) = \mu_1(x_2)$, $\mu_4(Y) = \mu_3(x_3)$ and therefore

$$\begin{aligned} \frac{\mu_4(A \cap B)}{\mu_4(Y)} &= \frac{\mu_4(A)}{\mu_4(Y)} \cdot \frac{\mu_4(B)}{\mu_4(Y)} \\ \Leftrightarrow \frac{\mu_4((x_1, x_2))}{\mu_3(\psi_1(x_1))} &= \frac{\mu_1(x_1)}{\mu_3(\psi_1(x_1))} \cdot \frac{\mu_2(x_2)}{\mu_3(\psi_1(x_1))}. \end{aligned}$$

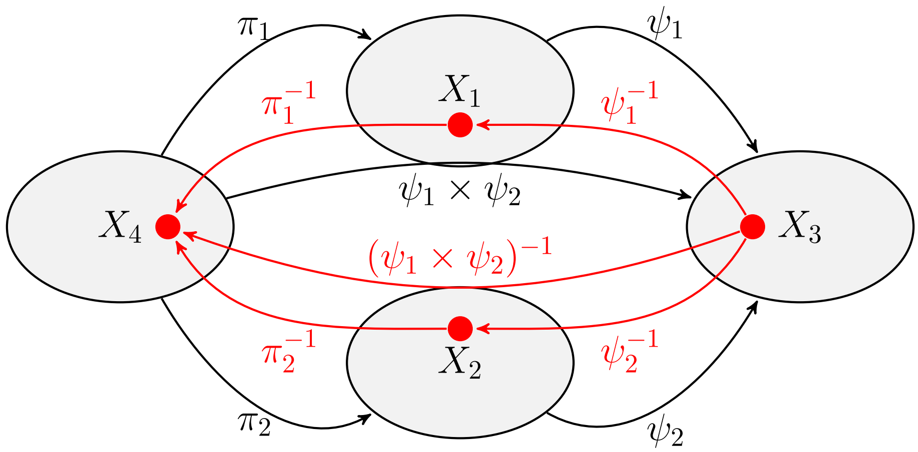

Figure 6: Commutative diagram for measure preserving maps. All the red subsets have the same measure in its respective probability space.

This motivates the following definition for the measure μ_4 on X_4

$$\mu_4((x_1, x_2)) := \frac{\mu_1(x_1)\mu_2(x_2)}{\mu_3(\psi_1(x_1))}. \tag{5}$$

Claim. Let us check that indeed μ_4 is a coupling, i.e. the projections π_i are measure preserving on (X_4, μ_4) for $i = 1, 2$.

Proof. W.l.o.g. $i = 1$ and let A be a subset of X_1.

$$\begin{aligned}
\mu_4(\pi_1^{-1}(A)) &= \sum_{\substack{(x_1,x_2)\in X_4 \\ x_1\in A}} \mu_4(x_1, x_2) = \sum_{\substack{(x_1,x_2)\in X_4 \\ x_1\in A}} \frac{\mu_1(x_1)\mu_2(x_2)}{\mu_3(\psi_1(x_1))} \\
&= \sum_{\substack{x_1\in A \\ \mu_3(\psi_1(x_1))\neq 0}} \frac{\mu_1(x_1)}{\mu_3(\psi_1(x_1))} \sum_{\substack{x_2\in X_2 \\ \psi_1(x_1)=\psi_2(x_2)}} \mu_2(x_2) \\
&= \sum_{\substack{x_1\in A \\ \mu_3(\psi_1(x_1))\neq 0}} \mu_1(x_1)\frac{\mu_2(\psi_2^{-1}(\psi_1(x_1)))}{\mu_3(\psi_1(x_1))} \\
&= \sum_{\substack{x_1\in A \\ \mu_3(\psi_1(x_1))\neq 0}} \mu_1(x_1) = \mu_1(A),
\end{aligned}$$

where the next to last equality comes from the fact that ψ_2 is measure preserving, i.e. the measure of the point $\psi_1(x_1)$ in X_3 is the same as the measure of its preimage in X_2. □

Taking $A = X_1$ this implies that indeed μ_4 is a probability measure. Together with the projections π_1 and π_2 we call μ_4 the (unique) **conditionally independent coupling** of X_1 and X_2 over the joint factor X_3. Observe that the pre-images $(\psi_1 \times \psi_2)^{-1}(x_3) = \psi_1^{-1}(x_3) \times \psi_2^{-1}(x_3)$ for $x_3 \in X_3$ give us a partition of X_4.

4.2 Relative entropy of couplings

In this part we want to establish statements for the relative entropy of conditionally independent couplings. We will repeatedly deal with similar terms. For the start consider the cross entropy of a conditionally independent coupling ν_4 and a coupling μ:

$$\begin{aligned} H(\mu, \nu_4) &= \sum_{(x_1,x_2)\in X_4} \log(\nu_4((x_1,x_2)))\mu((x_1,x_2)) \\ &= \sum_{(x_1,x_2)\in X_4} \log\left(\frac{\nu_1(x_1)\nu_2(x_2)}{\nu_3(\psi_1(x_1))}\right)\mu((x_1,x_2)). \end{aligned}$$

Then we can expand the logarithm to obtain three sums of the kind

$$\sum_{(x_1,x_2)\in X_4} \log(\nu_1(x_1))\mu((x_1,x_2)).$$

Since the part in the logarithm only depends on x_1 we can split the sum

$$= \sum_{\substack{x_1\in X_1 \\ \mu_3(\psi_1(x_1))\neq 0}} \log(\nu_1(x_1)) \sum_{\substack{x_2\in X_2 \\ \psi_1(x_1)=\psi_2(x_2)}} \mu((x_1,x_2)).$$

The last sum runs over all elements in X_2 which have the same value as x_1 and is thus equal to the measure of the pre image $\mu(\pi_2^{-1}(x_1))$. Finally using that π_2 is measure preserving on (X_4, μ) we get

$$= \sum_{\substack{x_1\in X_1 \\ \mu_3(\psi_1(x_1))\neq 0}} \log(\nu_1(x_1))\mu_1(x_1) = H(\mu_1, \nu_1).$$

With the same calculation we get for the second sum

$$\sum_{(x_1,x_2)\in X_4} \log(\nu_2(x_2))\mu((x_1,x_2)) = H(\mu_2, \nu_2)$$

and for the third we fix a summation by $x_3 \in X_3$ and use that $\psi_2 \circ \pi_2$ is measure preserving to get

$$\sum_{(x_1,x_2)\in X_4} \log(\nu_2(\psi_1(x_1)))\mu((x_1,x_2)) = H(\mu_3, \nu_3).$$

All together we get

$$H(\mu, \nu_4) = H(\mu_1, \nu_1) + H(\mu_2, \nu_2) - H(\mu_3, \nu_3).$$

This basically says that the cross entropy of a coupling μ over a conditionally independent coupling ν_4 does not depend on the choice of the coupling.

For the entropy of μ_4 this gives us

$$H(\mu_4) = H(\mu_1) + H(\mu_2) - H(\mu_3).$$

The main ingredient for the method of Szegedy is the following inclusion-exclusion type formula for conditionally independent couplings.

Lemma 4.1. Let μ_4 and ν_4 be conditionally independent couplings of X_1 and X_2 over the joint factor X_3, then

$$D(\mu_4||\nu_4) = D(\mu_1||\nu_1) + D(\mu_2||\nu_2) - D(\mu_3||\nu_3) \tag{6}$$

Proof.

$$\begin{aligned} D(\mu_4||\nu_4) &= H(\mu_4, \nu_4) - H(\mu_4) \\ &= H(\mu_1, \nu_1) + H(\mu_2, \nu_2) - H(\mu_3, \nu_3) - H(\mu_1) - H(\mu_2) + H(\mu_3) \\ &= D(\mu_1||\nu_1) + D(\mu_2||\nu_2) - D(\mu_3||\nu_3). \end{aligned}$$

□

The following lemma says that the relative entropy over couplings with respect to a conditionally independent couplings is minimized by the conditionally independent coupling.

Lemma 4.2. Let μ be any coupling of (X_1, μ_1) and (X_2, μ_2) over the joint factor X_3, then

$$D(\mu||\nu_4) \geq D(\mu_4||\nu_4) \tag{7}$$

Proof.

$$\begin{aligned} &D(\mu||\nu_4) - D(\mu_4||\nu_4) \\ &\quad = H(\mu, \nu_4) - H(\mu) - H(\mu_4, \nu_4) + H(\mu_4) \\ &\quad = H(\mu_4) - H(\mu) \\ &\quad = H(\mu_4, \mu_4) - H(\mu, \mu_4) + H(\mu, \mu_4) - H(\mu) \\ &\quad = H(\mu, \mu_4) - H(\mu) = D(\mu||\mu_4) \geq 0. \end{aligned}$$

This last relative entropy is a measure for the dependence between μ and μ_4. □

Note that we also get that the entropy over couplings is maximized by conditionally independent couplings. This coincides with the fact that entropy is maximized by the uniform distribution and the intuition that conditionally independent couplings are the most equally distributed couplings.

4.3 Coupling Sidorenko graphs

Now we are able to construct probability distributions on homomorphism sets which are iteratively obtained from the uniform distribution on edges using conditionally independent couplings.

Let μ be a probability distribution on $Hom(H, G)$ and $\beta : S \to V(H)$ be an injective map of some set S. Then the map $\mathrm{T}_\beta : Hom(H, G) \to V(G)^S$ defined by $\phi \mapsto \phi \circ \beta$ factors the probability space $(Hom(H, G), \mu)$. We get an induced probability distribution on $V(G)^S$ defined by $\mu|_\beta(\phi) = \mu(T_\beta^{-1}(\phi))$ for $\phi \in V(G)^S$. The probability space $(V(G)^S), \mu|_\beta)$ is called a **vertex factor** of μ.

Note that $T_\beta^{-1}(\phi)$ contains all possible ways how ϕ can be extended to a homomorphism in $Hom(H, G)$. If $S \subseteq V(H)$ then we denote by $\mu|_S$ the probability measure $\mu|_\beta$ where $\beta : S \to S$ is the identity. Then $\mu|_S$ is the marginal distribution on S with respect to T_β. For $\phi \in Hom(H, G)$ we then sometimes write $\phi|_S$ for $T_\beta(\phi)$. If S is empty then $\mu|_S$ is defined on a single point $V(G)^\emptyset$ and $D(\mu|_\emptyset) = (\log(1) - \log(1))1 = 0$.

Let H_1 and H_2 be two bipartite graphs and $(X_1, \mu_1) = (Hom(H_1, G), \mu_1)$ and $(X_1, \mu_1) = (Hom(H_2, G), \mu_1)$ two probability spaces. Our goal is to combine both probability distributions to a new probability distribution on the set of homomorphism of a graph glued together from H_1 and H_2. We want to define a coupling, thus we need a joint factor. Therefore we label the n gluing vertices in H_1 and H_2 by $\{\beta_i : [n] \to V(H_i)\}_{i=1,2}$ and assume that both induce the same vertex factor, i.e. the marginal distributions on the labelled vertices are the same. This is $\mu_3 := \mu_1|_{\beta_1} = \mu_2|_{\beta_2}$ with $X_3 = V(G)^{[n]}$ defines a vertex factor of μ_1 and μ_2. Then $\{\psi_i = \mathrm{T}_{\beta_i}\}_{i=1,2}$ are two measure preserving maps and X_3 is the joint factor of (X_1, μ_1) and (X_2, μ_2). We denote by $C(\mu_1, \mu_2, \beta_1, \beta_2)$ the conditionally independent coupling μ_4 of μ_1 and μ_2 over μ_3.

The assumption $\mu_1|_{\beta_1} = \mu_2|_{\beta_2}$ tells us that the measure of all possible ways to extend $\phi \in V(G)^{[n]}$ to a homomorphism in $Hom(H_1, G)$ or $Hom(H_2, G)$ is the same. We again have the commutative diagram illustrated in Figure 6.

Example 4.3. For an easy example let ν_1 and ν_2 be the uniform distribution on $Hom(H_1, G)$ and $Hom(H_2, G)$ and β_i be as above. Then for $\phi \in V(G)^{[n]}$ we get that ν_3 is uniformly distributed

$$\begin{aligned}\nu_3(\phi) = \nu_1|_{\beta_1}(v) &= \nu_1 \circ \mathrm{T}_{\beta_1}^{-1}(\phi) \\ &= \nu_1(\{\phi' \in V(G)^{V(H_1)} : \phi' \circ \beta = \phi\}) \\ &= \frac{|\{\phi \in V(G)^{V(H_1)} : \phi' \circ \beta = \phi\}|}{|V(G)|^{|V(H_1)|}} \\ &= \frac{|V(G)|^{V(H_1)-n}}{|V(G)|^{|V(H_1)|}} = V(G)^{-n}.\end{aligned}$$

Let ν_4 be the conditionally independent coupling of ν_1 and ν_2 over ν_3 then

$$\nu_4(\phi_1, \phi_2) = \frac{\nu_1(\phi_1)\nu_2(\phi_2)}{\nu_3(\psi_1(\phi_1))} = V(G)^{n-V(H_1)-V(H_2)}$$

is again uniformly distributed.

Since we need a probability distribution on the set of homomorphisms for every G we define the following concept: For a graph H we define a **probability model** f on the set of finite graphs G whose value $f(G)$ is a probability distribution on $Hom(H,G)$. We say that H is the **skeleton** of the probability model f.

Now let f_1 and f_2 be two probability models with skeletons H_1 and H_2 and $\{\beta_i : [n] \to V(H_i)\}_{i=1,2}$ two maps such that $f_1(G)|_{\beta_1} = f_2(G)|_{\beta_2}$ for all G. Then we call β_1 and β_2 a **joint vertex factor** of f_1 and f_2. The conditionally independent coupling $g = C(f_1, f_2, \beta_1, \beta_2)$ of f_1 and f_2 is the probability model g whose value on G is the conditionally independent coupling $g(G) = C(f_1(G), f_2(G), \beta_1, \beta_2)$ of $f_1(G)$ and $f_2(G)$ over $\beta_1([n])$ and with $f_1(G)|_{\beta_1} = f_2(G)|_{\beta_2}$. The skeleton of g is the graph obtained by identifying the vertices with the same labels in the disjoint union of H_1 and H_2. After identification we delete multiple edges in the same way as for smooths graphs in Section 2.

$C(f_1(G), f_2(G), \beta_1, \beta_2)$ is a probability measure over

$$X_4 = \left\{ (\phi_1, \phi_2) \in Hom(H_1, G) \times Hom(H_2, G) \;\middle|\; \begin{array}{c} \phi_1 \circ \beta_1 = \phi_2 \circ \beta_2 \\ \mu_3(\phi_1 \circ \beta_1) \neq 0 \end{array} \right\}$$

which is in fact isomorphic to

$$X_4 \cong \{\phi \in Hom(H, G) : \mu_3(\phi \circ \beta_1) \neq 0\}$$

by the isomorphism $\phi \mapsto (\phi_1, \phi_2)$ where $\phi_1 = \phi|_{V(H_1)}$ and $\phi_2 = \phi|_{V(H_2)}$. The property $\phi_1 \circ \beta_1 = \phi_2 \circ \beta_2$ is immediate because the images of β_1 and β_2 are the labelled points glued together. The requirement $\mu_3(\phi \circ \beta_1) \neq 0$ is no restriction because $\mu_3(\phi \circ \beta_1) = 0$ implies $\mu_1(\phi|_{V(H_1)}) = 0$.

Finally we constructed a probability distribution μ on $Hom(H, G)$ from two distributions on $Hom(H_1, G)$ and $Hom(H_2, G)$ using a conditionally independent coupling. Note that for $\phi \in V(G)^{[n]}$ we have that $\mathrm{T}_{\beta_1}^{-1}(\phi) \times \mathrm{T}_{\beta_2}^{-1}(\phi) \subseteq X_4$ is the set of possibilities which extends ϕ to a homomorphism from H to G.

Example 4.4. We take a closer look at some easy probability models. Let G be any graph. The uniform random edge model $G \mapsto \tau(e, G)$ gives us a probability distribution on $V(G)^2$. For two vertices $x, y \in V(G)$ the probability is $\tau(e, G)((x, y)) = \frac{1}{2E}$ when $\{x, y\} \in E(G)$ and zero otherwise, a uniformly random chosen edge.

Now we want to couple two edges $\{w, x\}$ and $\{y, z\}$ with probability model $\tau(e, G)$ on one vertex. We define $\beta_1(1) = w$ and $\beta_2(1) = y$. Check that the marginal distribution on a vertex is indeed κ

$$\tau(e, G)|_{\beta_i}(v) = \sum_{v \in e \in E(G)} \tau(e, G)(e) = \frac{deg(v)}{2E} = \kappa(v).$$

Then we get that $\tau(e,G)|_{\beta_1} = \kappa = \tau(e,G)|_{\beta_2}$ and therefore β_1 and β_2 define a joint vertex factor of twice $\tau(e,G)$. We define the conditionally independent coupling $f = (\tau(e,G), \tau(e,G), \beta_1, \beta_2)$.

$f(G)$ is a probability distribution on $V(G)^3$. For $x, y, z \in V(G)$ the probability is zero if $\{x,y\} \notin E(G)$ or $\{y,z\} \notin E(G)$ and otherwise we compute

$$\begin{aligned} f(G)((x,y,z)) = \mu_4((x,y,z)) = \frac{\mu_1((x,y))\mu_2((y,z))}{\mu_1|_{\beta_1}(y)} \\ = \frac{\tau(e,G)((x,y))\tau(e,G)((y,z))}{\kappa(y)} \\ = \frac{(2E)^{-1}(2E)^{-1}}{\deg(y)(2E)^{-1}} \\ = \frac{1}{\deg(y)2E}. \end{aligned}$$

The Skeleton of f is the star on 3 vertices. Observe that the probability distribution $f(G)$ is the same as the one we defined in Section 3.3 for proving that any tree is Sidorenko. In this way we can compute explicitly the probability model for small graphs.

Note that in the proof we used that the marginal distribution on a single vertex is κ. This can be generalized to the important property that marginals do not change under conditionally independent couplings. Let f_1 and f_2 be probability models with skeletons H_1 and H_2. Assume that $\{\beta_i : [n] \to V(H_i)\}_{i=1,2}$ defines a joint vertex factor. Let H be the skeleton of $g = C(f_1(G), f_2(G), \beta_1, \beta_2)$.

Lemma 4.5. For any subset $S_1 \subseteq V(H_1)$ or $S_2 \subseteq V(H_2)$ the marginal distributions of $g(G)$ are the same for all G as of $f_1(G)$, $f_2(G)$ repectively, i.e.

$$g(G)|_{S_1} = f_1(G)|_{S_1} \quad g(G)|_{S_2} = f_2(G)|_{S_2} \quad \forall G.$$

Proof. Let $S \subseteq V(H_1)$ ($S \subseteq V(H_2)$ goes analogously) and $\beta : S \to S$ the identity map. Then we get for the probability of one element $\phi \in V(G)^S$

$$\begin{aligned} g(G)|_S(\phi) = g(G)|_\beta(\phi) = g(G)(\mathrm{T}_\beta^{-1}(\phi)) = \mu_4(\mathrm{T}_\beta^{-1}(\phi)) \\ = \sum_{\substack{(\phi_1,\phi_2)\in X_4 \\ \mathrm{T}_\beta(\phi_1)=\mathrm{T}_\beta(\phi_2)=\phi}} \mu_4(\phi_1,\phi_2) \\ = \sum_{\substack{\phi_1\in X_1,\phi_2\in X_2,\mathrm{T}_\beta(\phi_1)=\mathrm{T}_\beta(\phi_2)=\phi \\ \mu_1|_{\beta_1}(\phi_1\circ\beta_1)\neq 0,\mathrm{T}_{\beta_1}(\phi_1)=\mathrm{T}_{\beta_2}(\phi_2)}} \frac{\mu_1(\phi_1)\mu_2(\phi_2)}{\mu_1|_{\beta_1}(\mathrm{T}_{\beta_1}(\phi_1))} \\ = \sum_{\substack{\phi_1\in X_1,\mathrm{T}_\beta(\phi_1)=\phi \\ \mu_1|_{\beta_1}(\phi_1\circ\beta_1)\neq 0}} \frac{\mu_1(\phi_1)}{\mu_1|_{\beta_1}(\mathrm{T}_{\beta_1}(\phi_1))} \sum_{\substack{\phi_2\in X_2,\mathrm{T}_\beta(\phi_2)=\phi \\ \mathrm{T}_{\beta_1}(\phi_1)=\mathrm{T}_{\beta_2}(\phi_2)}} \mu_2(\phi_2). \end{aligned}$$

Now we consider a fixed ϕ_1. We know that $S \subseteq V(H_1)$, this implies that $\beta(S) \cap V(H_2) \subseteq \beta_2([n])$ and thus $\mathrm{T}_\beta(\phi_1) = \phi$ and $\mathrm{T}_{\beta_2}(\phi_2) = \mathrm{T}_{\beta_1}(\phi_1)$ implies $\mathrm{T}_\beta(\phi_2) = \phi$. We sum over all those ϕ_2 an conclude

$$= \sum_{\substack{\phi_1 \in X_1, \mathrm{T}_\beta(\phi_1) = \phi \\ \mu_1|_{\beta_1}(\phi_1 \circ \beta_1) \neq 0}} \mu_1(\phi_1) \frac{\mu_2((\mathrm{T}_{\beta_2})^{-1}(\mathrm{T}_{\beta_1}(\phi_1)))}{\mu_1|_{\beta_1}(\mathrm{T}_{\beta_1}(\phi_1))}$$

$$= \sum_{\substack{\phi_1 \in X_1, \mathrm{T}_\beta(\phi_1) = \phi \\ \mu_1|_{\beta_1}(\phi_1 \circ \beta_1) \neq 0}} \mu_1(\phi_1) = \mu_1 \circ \mathrm{T}_\beta^{-1}(\phi).$$

□

Definition 4.6. We denote by $\mathfrak{U}$ the smallest set of probability models which contains the model $G \mapsto \tau(e, G)$ (uniform random edge) and is closed with respect to conditionally independent couplings over joint vertex factors. The set of skeletons of models in $\mathfrak{U}$ is denoted by $\mathfrak{G}$.

Notice that the fact that $f(G)$ is a probability distribution on $Hom(H, G)$ for all G implies that $Hom(H, G)$ is not empty. Taking G as a triangle and a single edge this in turn implies that every graph in $\mathfrak{G}$ has to be bipartite.

In particular it follows from the previous lemma that if a probability distribution on $Hom(H, G)$ is constructed according to a probability model in $\mathfrak{U}$ then its marginals on the edges of H are all identical to $\tau(e, G)$ and its marginals on vertices are identical to κ. This is the case because every probability model in $\mathfrak{U}$ is constructed via conditionally independent couplings from the model $G \mapsto \tau(e, G)$.

Definition 4.7. $\mathfrak{U}_1$ is the smallest set of probability models which contains the model $G \mapsto \tau(e, G)$ and is closed with respect to conditionally independent couplings over joint vertex factors under the restriction that the images $\beta_i([n])$ are independent sets. Correspondingly $\mathfrak{G}_1$ is the set of skeletons of the elements in $\mathfrak{U}_1$.

Note that $\mathfrak{U}_1 \subseteq \mathfrak{U}$ and therefore $\mathfrak{G}_1 \subseteq \mathfrak{G}$. With all this in our hands we can now easily prove Sidorenko's conjecture for $\mathfrak{G}_1$.

Proposition 4.8. Every element in $\mathfrak{U}_1$ is a family of witness measures. Consequently every graph in $\mathfrak{G}_1$ satisfies Sidorenko's conjecture.

Proof. $D(\tau(e, G)) \leq 1D_e$, thus $\tau(e, G)$ is a witness measure. Assume that f_1 and f_2 are probability models with Skeletons H_1 and H_2. Assume that the maps $\{\beta_i : [n] \to V(H_i)\}_{i=1,2}$ define a joint vertex factor such that the images of β_1 and β_2 are independent sets. Let H be the skeleton of $g = C(f_1, f_2, \beta_1, \beta_2)$.

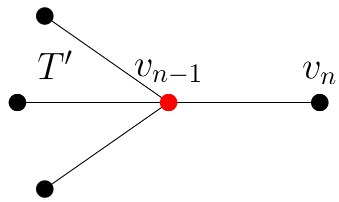

Figure 7: Recursively constructing the probability model for a tree T by gluing together T' and one edge on the red vertex.

Then with Lemma 4.1 and 3.2 and the induction hypothesis (IH)

$$\begin{aligned} D(g(G)) &\overset{(6)}{=} D(f_1(G)) + D(f_2(G)) - D(f_1(G)|_{\beta_1}) \\ &\overset{(4)}{\leq} D(f_1(G)) + D(f_2(G)) \\ &\overset{IH}{\leq} (|E(H_1)| + |E(H_2)|)D_e = |E(H)|D_e. \end{aligned}$$

□

Note that the $n = 0$ is allowed, because the emptyset is an independent set. In particular this tells us that the disjoint union of to Sidorenko graphs is again Sidorenko, which was already proven by Sidorenko himself [10].

With this proposition we now can prove Sidorenko's conjecture for some classes of graphs. In the following lemmas we show that trees, even cycles and bipartite graphs where one vertex is complete to the other side are in $\mathfrak{G}_1$ and therefore by Proposition 4.8 satisfy the Sidorenko conjecture.

Lemma 4.9. Any tree T is in $\mathfrak{G}_1$.

Proof. We prove by induction that any tree T on n vertices has a probability model in $\mathfrak{U}_1$ and thus is in $\mathfrak{G}_1$. For $n = 2$ we have a single edge. We know that the uniform random edge model $G \mapsto \tau(e, G)$ is in $\mathfrak{U}_1$ and therefore T is in $\mathfrak{G}_1$. Assume that all trees on $n - 1$ vertices have a probability model f in $\mathfrak{U}_1$.

Let T be a tree on n vertices $v_1, \ldots, v_n$ and assume that v_n is a leaf connected to v_{n-1}. By induction we know that the tree T' induced by the vertices $v_1, \ldots, v_{n-1}$ has a probability model f' in $\mathfrak{U}_1$. As stated before the uniform random edge model for the edge $\{v_{n-1}, v_n\}$ is in $\mathfrak{U}_1$. We define the labelling maps $\{\beta_i : [1] \to V(H_i\}_{i=1,2})$ by $\beta_i(1) = v_n$. By Lemma 4.5 we know that for every probability distribution in $\mathfrak{U}$ the marginals on vertices are identical to κ. Thus β_1 and β_2 define a joint vertex factor of f' and $\tau(e, G)$, because $f'(G)|_{\beta_1} = \kappa = \tau(e, G)|_{\beta_2}$. Since v_{n-1} is independent the conditional independent coupling $f = (f'(G), \tau(e, G), \beta_1, \beta_2)$ is in $\mathfrak{U}_1$. In Figure 7 we illustrate the gluing process, where the vertex in the image of β_1 and β_2 are coloured red. Wee see that the skeleton of f is the tree T and thus we are done. □

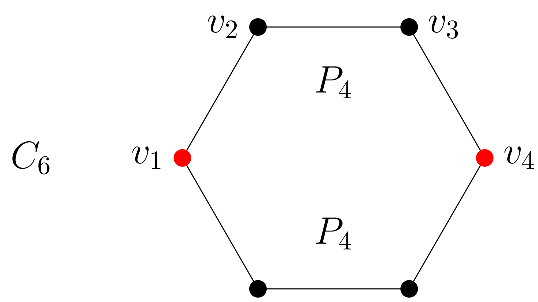

Figure 8: Constructing the probability model for an even cycle C_{2m} by gluing to paths of length m on its boundary vertices. Example for $m = 3$.

This explains how we always can glue any number of additional edge onto a graph in $\mathfrak{G}_1$. Sidorenko already proved in [10] that one can add additional edges to Sidorenko graphs.

Lemma 4.10. Even cycles C_{2m} are in $\mathfrak{G}_1$.

Proof. We construct a probability model for C_{2m}. The path P_{m+1} on $m+1$ vertices $v_1, \dots, v_{m+1}$ is a tree and therefore there is a probability model f in $\mathfrak{G}_1$ with skeleton P_m. We define $\{\beta_i : [2] \to V(H_i\}_{i=1,2})$ by $\beta_i(1) = v_{m+1}$ and $\beta_i(2) = v_1$. Then β_1 and β_2 define a joint vertex factor, because $f(G)|_{\beta_1} = f(G)|_{\beta_2}$. Observe that gluing a graph to itself on an independent set is always possible. Since v_1 and v_m are independent we get that the conditionally independent coupling of $f' = (f(G), f(G), \beta_1, \beta_2)$ has skeleton C_{2m} as shown in Figure 8. □

In the proof of the previous lemma we used that we can always glue a graph to itself on an independent set. We know from Lemma 4.5 that marginals do not change under coupling and therefore when constructing the probability model for a graph we can always glue previous graphs on independent sets. Applying this to stars gives us the possibility to prove that bipartite graph where one vertex is complete to the other side are Sidorenko.

Proof of Theorem 2.1. Recall that we have vertices $x, v_1, \dots, v_k, y_1, y_m$ where x is connected to $v_1, \dots, v_n$ and y_i is connected to $S_i \subseteq \{v_1, \dots, v_k\}$ and $a_t = |S_t|$. We recursively build the probability model $f = f_{k+1}$ for star on $k+1$ vertices $x, v_1, \dots, v_k$ as in the proof of Lemma 4.9. Now for every $t \in [m]$ we glue the stars on the vertex y_t with neighbours S_t onto the probability model f. For this define $\{\beta_i : [a_t] \to V(H_i\}_{i=1,2})$ by $\beta_i([a_t]) = S_t$, which gives us a joint vertex factor, because f was constructed from f_t. One step can be seen in Figure 9. We obtain that f is in $\mathfrak{U}_1$ and therefore the graph is Sidorenko. □

In general this gives us the possibility for any tree to reflect a sub-tree on an independent set and the result will be in $\mathfrak{G}_1$. Next we want to extend our gluing techniques.

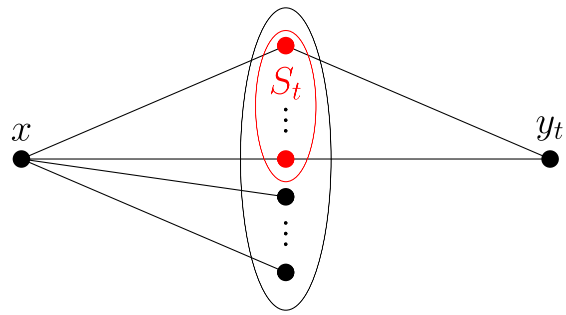

Figure 9: Reflecting a sub-star on an independent set to prove that bipartite graphs with one vertex xomplete to the other side are in $\mathfrak{G}_1$.

Definition 4.11. Let $\mathfrak{U}_2$ be the smallest set of probability models which contains the model $G \mapsto \tau(e, G)$ and is closed with respect to conditionally independent couplings over joint vertex factors which span forests. Correspondingly $\mathfrak{G}_2$ is the set of skeletons of the elements in $\mathfrak{U}_2$.

We have $\mathfrak{U}_1 \subseteq \mathfrak{U}_2 \subseteq \mathfrak{U}$ and thus $\mathfrak{G}_1 \subseteq \mathfrak{G}_2 \subseteq \mathfrak{G}$.

Theorem 4.12. Every element in $\mathfrak{U}_2$ is a family of witness measures. Consequently every graph in $\mathfrak{G}_2$ satisfies Sidorenko's conjecture.

To prove the theorem we need the following lemma.

Lemma 4.13. Let H be a forest and G any graph. Let μ be a probability measure on $Hom(H, G)$ such that the marginals on the edges of H are identical with $\tau(e, G)$ and the marginals on the vertices are identical with κ. Then

$$D(\mu) \geq D_e|E(H)| - D_v(2|E(H)| - |V(H)|).$$

Proof of Lemma 4.13. We go by induction on the number of vertices $|V(H)|$. For $|V(H)| = 0$ there is nothing to prove and for $|V(H)| = 1$ we get $D(\mu) = D_v$. If $|V(H)| = 2$ then either $|E(H)| = 1$ and then $D(\mu) = D_e$ or $|E(H)| = 0$ and $D(\mu) = D_v + D_v - D(\mu|_\emptyset) = 2D_v$ which is fine in both cases.

Now assume $V(H) \geq 3$ then since H is a forest there exists V_1 and V_2 such that $V(H) = V_1 \cup V_2$ and $V_1 \cap V_2$ is a single vertex v. We set $\mu_1 := \mu|_{V_1}$ and $\mu_2 := \mu|_{V_2}$, then μ is a coupling of μ_1 and μ_2 over $\mu_3 = \mu|_v$. Let μ_4 be the conditionally independent coupling of μ_1 and μ_2 over μ_3. With E_1 and E_2 as the edges contained in the forests induced by V_1 and V_2 respectively we get $|E(H)| = |E_1| + |E_2|$ and then using Lemma 4.2, 4.1 and 3.2 we get

$$\begin{aligned} D(\mu) &\overset{(7)}{\geq} D(\mu_4) \overset{(6)}{=} D(\mu_1) + D(\mu_2) - D(\mu|_v) \\ &\overset{(4)}{\geq} D_e|E_1| - D_v(2|E_1| - |V_1|) + D_e|E_2| - D_v(2|E_2| - |V_2|) - D_v \\ &= D_e|E(H)| - D_v(2|E(H)| - |V(H)|) \end{aligned}$$

□

Proof of Theorem 4.12. Let $f \in \mathfrak{U}_2$ be a probability model with skeleton H. We prove by induction that

$$D(f(G)) \leq D_e|E(H)| - D_v(2|E(H)| - |V(H)|). \tag{8}$$

Since we can assume that H has no isolated points we have $2|E(H)| \geq |V(H)|$ which implies $D(f(G)) \leq |E(H)|D_e$.

Clearly $\tau(e, G)$ with skeleton a single edge satisfies the inequality, because $D(\tau(e, G)) = D_e$. Let f_1 and f_2 be two probability models with skeletons H_1 and H_2. Assume that $\{\beta_i : [n] \to V(H_i)\}_{i=1,2}$ defines a joint vertex factor such that the images of β_1 and β_2 form an identical forests we call H_3. Let H be the skeleton of $g = C(f_1, f_2, \beta_1, \beta_2)$ then we get

$$\begin{aligned}
D(g(G)) &\overset{(6)}{=} D(f_1(G)) + D(f_2(G)) - D(f_1(G)|_{\beta_1}) \\
&\overset{IH}{\leq} D_e(|E(H_1)| + |E(H_1)| - |E(H_3)|) \\
&\quad - D_v(2|E(H_1)| + 2|E(H_2)| - 2|E(H_3)| - |V(H_1)| - |V(H_2)| + |V(H_3)|) \\
&= D_e|E(H)| - D_v(2|E(H)| - |V(H)|),
\end{aligned}$$

where the first two terms $D(f_i(G))$ are upper bounded by induction hypothesis (IH) and the third term $D(f_1(G)|_{\beta_1})$ is lower bounded by Lemma 4.13. □

In general it is a completely combinatorial and finite problem to decide membership in these classes. With the three dimensional cube and the 1-subdivision of K_4 we now give two precise examples.

Example 4.14. We want to prove that the 3-cube Q_3 is in $\mathfrak{G}_2$ and therefore Sidorenko. We start with two 4-cycles which are glued together on one edge. This graph is in $\mathfrak{G}_2$ because one vertex is complete to the other side and these graphs are in $\mathfrak{G}_1 \subseteq \mathfrak{G}_2$. We now take two copies of this graph H_1 and H_2 with probability models f_1 and f_2. Define the maps $\{\beta_i : [4] \to V(H_i)\}_{i=1,2}$ such that they map onto two opposite edges as indicated with red in Figure 10. This defines a joint vertex factor since we glue on the same graph and therefore $f_1(G)|_{\beta_1} = f_2(G)|_{\beta_2}$. Then by Theorem 4.12 the conditionally independent coupling $g = (f_1(G), f_2(G), \beta_1, \beta_2)$ lies in $\mathfrak{G}_2$ and has skeleton Q_3 as shown in Figure 10. Thus Q_3 is Sidorenko.

Example 4.15. Next we want to look at the 1-subdivision of K_4. This graph is obtained from the complete graph on 4 vertices by splitting every edge with a vertex. We start with two copies H_1 and H_2 of a 6-cycle with one additional edge which clearly is in $\mathfrak{G}_2$. We define $\{\beta_i : [4] \to V(H_i)\}_{i=1,2}$ such that the pendent vertex and two edges on the other side are labelled as indicated in Figure 11. We get a joint vertex factor and thus by Theorem 4.12 the conditionally independent coupling lies in $\mathfrak{G}_2$. The skeleton is the 1-subdivision of K_4 (compare Figure 11) and thus it is Sidorenko. In general the 1-subdivision of the complete graph K_m is not in $\mathfrak{G}_2$, but we will see later that it is Sidorenko.

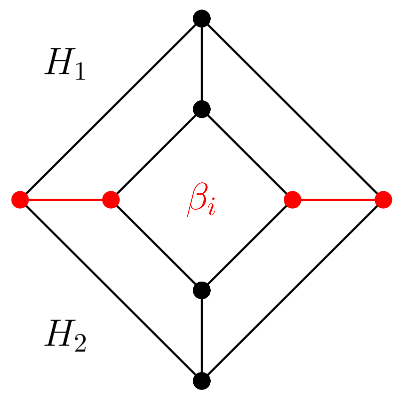

Figure 10: 3-cube is Sidorenko.

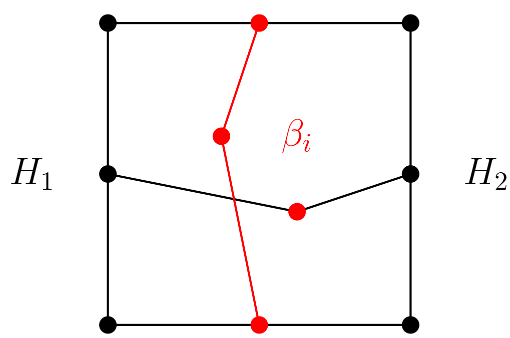

Figure 11: 1-subdivision of K_4 is Sidorenko.

We will keep on marking the induced graph of the image of β_1 and β_2 by red. But from now on we will omit the reference to the graph $D(f) := D(f(G))$ and leave out the details about the conditionally independent coupling, because the argument always is the same. Note that in fact Equation 8 is stronger then Sidorenko's conjecture, because if G is not regular, then κ is not uniformly distributed and therefore $D_v > 0$.

To achieve appropriate upper bounds for the relative entropy using the coupling method we need a lower bound for the relative entropy of the gluing part containing D_e times the number of edges. For forests we indeed get equality between the upper and lower bound. The question is what kind of bounds can we hope for if we do not glue on forests. Later we will see a recursive scenario where these terms get cancelled.

Example 4.16. Let us consider an easy example and try to find a lower bound for the relative entropy of the 6-cycle. To construct C_6 we follow the proof of Lemma 4.10 and start with two paths of length 3. Then these two graphs with probability models f_1 and f_2 are glued together on two vertices via $\{\beta_i\}_{i=1,2}$. We get the probability model f for C_6. By lemma 4.1 we get

$$\begin{aligned} D(C_6) &= D(f(G)) = D(f) \\ &\overset{(6)}{=} D(f_1) + D(f_2) - D(f_i|_{\beta_i}) \\ &= 6D_e - 4D_v - D(f_i|_{\beta_i}) \end{aligned}$$

and therefore we need an upper bound for $D(f_i|_{\beta_i})$. $f_i|_{\beta_i}$ is a distribution for two vertices which have an edge in their neighbourhood. Lemma 4.2 only gives us a lower bound of $2D_v$.

4.4 Boundaries of $\mathfrak{G}_1$ and $\mathfrak{G}_2$

In this part we want to develop a way to prove that a certain bipartite graph is in $\mathfrak{G}_2 \setminus \mathfrak{G}_1$. For a graph H not to be in $\mathfrak{G}_1$ we have to show that it is not possible to obtain the probability model for this graph from the uniform random edge model using couplings over independent sets. To prove this we consider any independent set A that disconnects H up to symmetry. Then we need to look at all possible partitions of the remaining vertices into two sets such that the vertices in A can be reached from both sets. Together with A these two sets induce the graphs H_1,H_2 and the maps $\{\beta_i\}_{i=1,2}$. Let f_1 and f_2 be two probability models in $\mathfrak{G}_1$ with skeletons H_1 and H_2 respectively. It is left to show that β_1 and β_2 do not define a joint vertex factor of f_1 and f_2, i.e. we need to find G such that $f_1(G)|_{\beta_1} \neq f_2(G)|_{\beta_2}$. How can we find such G? The following is true for all probability models.

Claim. A very important fact is that for all probability models $f \in \mathfrak{U}$ with skeleton H, all graphs G and any homomorphism $\phi \in Hom(H, G)$ the probability of $f(G)(\phi)$ is non-zero.

Proof. Let G be any graph. The probability distribution $\tau(K_2, G)$ corresponds to a uniformly random chosen edge and thus every homomorphism has non-zero probability. Let f_1 and f_2 be two probability models with the above property and assume that $\{\beta_i : [n] \to V(H_i)\}_{i=1,2}$ define a joint vertex factor. Let H be the skeleton of $g = (f_1(G), f_2(G), \beta_1, \beta_2)$ and $\phi \in Hom(H, G)$ any homomorphism. Then

$$g(G)(\phi) = \mu_4((\phi_1, \phi_2)) = \frac{f_1(G)(\phi_1) f_2(G)(\phi_2)}{f_1(G)|_{\beta_1}(T_{\beta_1}(\phi))}$$

which is non-zero by assumption, because ϕ_1 and ϕ_2 are homomorphisms. □

Intuitively it is clear that in any coupling vertices that did not have a common neighbour before cannot have one afterwords. Using the above claim we want to formulate this in the following lemma.

Lemma 4.17. Let H_1 and H_2 be graphs with probability models $f_1, f_2 \in \mathfrak{U}$ and let $\{\beta_i : [n] \to V(H_i)\}_{i=1,2}$ be such that $\beta_1([k])$ has a common neighbour in H_1 but $\beta_2([k])$ does not in H_2. Then β_1 and β_2 do not define a joint vertex factor and in particular $f_1(G)|_{\beta_1} \neq f_2(G)|_{\beta_2}$ for some G.

Proof. Let $G = H_2$ and $\phi \in Hom(H_2, G)$ the identity, then we get that $f_2(G)|_{\beta_2}(T_{\beta_2}(\phi)) \neq 0$ because $f_2(G)(\phi) \neq 0$. But in H_1 the vertices $\beta_1([k])$ do have a common neighbour and therefore it is not possible to extend $T_{\beta_1}(\phi)$ to a homomorphism in $Hom(H_1, G)$ and thus $f_1(G)|_{\beta_1}(T_{\beta_2}(\phi)) = 0$. □

The lemma tells us that if we find a set of vertices that have a common neighbour in H_1 but not in H_2 then there cannot be any coupling. As an example we will now proof that the 3-dimensional cube is not in $\mathfrak{G}_1$ and therefore in $\mathfrak{G}_2 \setminus \mathfrak{G}_1$.

Example 4.18. Let H be the three dimensional cube as in Figure 10. We have to consider all independent sets in $V(H)$ and check if there is any coupling possible. For every independent set A we have to check all possible partitions of the remaining vertices into two sets such that both reach all of A. Vertices in A will always be coloured red. All vertices that belong to one of the two sets will be coloured with green and together with A they induce H_1. H_2 is induced by all red and black vertices.

The largest possible independent set in H is of size 4 and is unique up to symmetry. The only possible partition for the rest is into two sets of two, because any three vertices have a common neighbour, which the fourth one cannot reach. As shown in Figure 12 on the left we can apply Lemma 4.17 to the vertices a, b and c, which do have a common neighbour in H_2 but not in H_1. Thus there cannot be a coupling using this independent set.

For independent sets of size three there is still only one possibility since any two from one side cover the hole other side. The rest is already a partition as shown in Figure 12 on the right and again vertices a, b and c do have a common neighbour in H_2 but not in H_1.

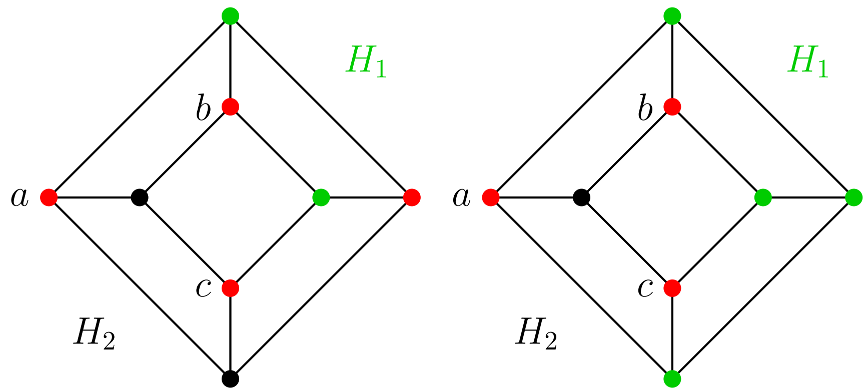

Figure 12: We cannot build the 3-cube by gluing on an independent set of size 4 or 3.

Since H is 3-regular there are two different independent sets of size two. Either we take two points from one side or two antipodal points. In both cases the graph stays connected and therefore there exists no partition. Obviously the same holds for independent sets of size one. We implicitly proved that the vertex connectivity of H is 3.

We checked all independent sets and in conclusion there are no two probability models f_1 and f_2 such that any coupling over an independent set gives us a probability model with skeleton H, i.e. $H \notin \mathfrak{G}_1$.

It is a little bite more complex, but with the same arguments one can prove that the 1-subdivision of K_4 is in $\mathfrak{G}_2 \setminus \mathfrak{G}_1$. To prove that a graph H is not in $\mathfrak{G}_2$ or even not in $\mathfrak{G}$ one has to check all possible forest in H or every subgraph for a partition that induces a coupling. We will see an example for this later.

5 Further applications of the method

5.1 Graph where one class is of size at most four

In his original paper [10] Sidorenko proves that all bipartite graphs with one class of size at most three are Sidorenko. He claims that the same calculations hold if we consider graphs where one class is of size at most 4. With the method of Szegedy we can now verify that this statement is also true. Let H be a bipartite graph with independent sets $A\dot{\cup}B = V(H)$, where $|A| \leq 4$. We assume the graph to be connected, because otherwise we consider its connected components and disjoint union of Sidorenko graphs are again Sidorenko.

If $|A| = 1$ then we have a star, which is Sidorenko. If $|A| = 2$ then either it is a tree or there are at least two common neighbours for the vertices in A which make these neighbours complete to the other side. In both cases we showed that H is Sidorenko. If for $A = \{x, y, y'\}$ these three vertices have a common neighbour we are again done. So assume they do not. We then start with two copies of the graph $T_{r,q}$ from Section 2.3 which are Sidorenko since all the $r \neq 0$ vertices are complete to the other side. Next we define $\{\beta_i : [q+1] \to \{x, w_1, \dots, w_q\}\}_{i=1,2}$ and glue both together along $\beta_i([q+1])$ as indicated in Figure 13. Then the resulting graph $T_{r,r,q}$ is Sidorenko. We can add $r' - r$ further neighbours to x and y', because the marginal distribution on this vertices did not change. Gluing edges onto x, y and y' gives us any possible graph H we need. Since H is connected we can assume w.l.o.g. $r \neq 0$ and $r' \neq 0$, whereas q could be zero.

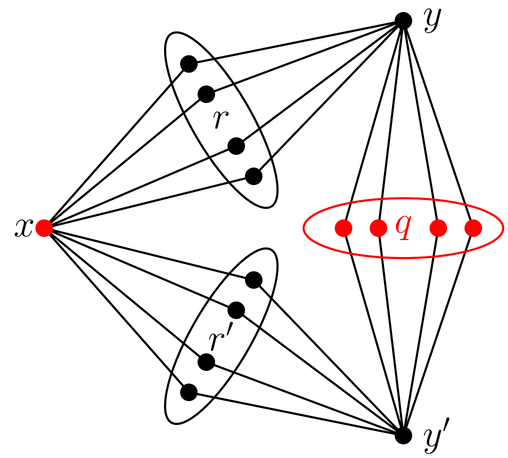

Figure 13: $T_{r,r',q}$

Now we need to consider $A = \{w, x, y, z\}$. Again, if there is a vertex connected to all of A we are done, because it is then complete to the whole other side which is A. Assume there is no such vertex. The examples from the last section (3-cube and 1-subdivision of K_4) now give us the building blocks for any graph of this kind. Basically we have to apply both constructions at once while replacing some vertices by more. As in the previous case we do not need to worry about vertices that are only connected to one vertex in A because we can add them afterwords. Therefore the graph looks like in Figure 14 and we have to take care of $\binom{4}{2} = 6$ sets for the pairwise neighbourhoods of two vertices in A and of $\binom{4}{3} = 4$ sets for threefold neighbourhoods. Note that some of these sets could be empty, but we know that the graph is connected. As shown in Figure 14, we want to glue some variants of the graphs H_1 and H_2 together on the red forest. The red vertex in the center is only there if s is odd, so we have s common neighbours for x and y in total.

We do a case distinction by the number $l = 0, \dots, 4$ which counts how many threefold neighbourhoods of A are non-empty.

- $l = 0$: This implies $t = r = r' = q = 0$. Since H is connected we can

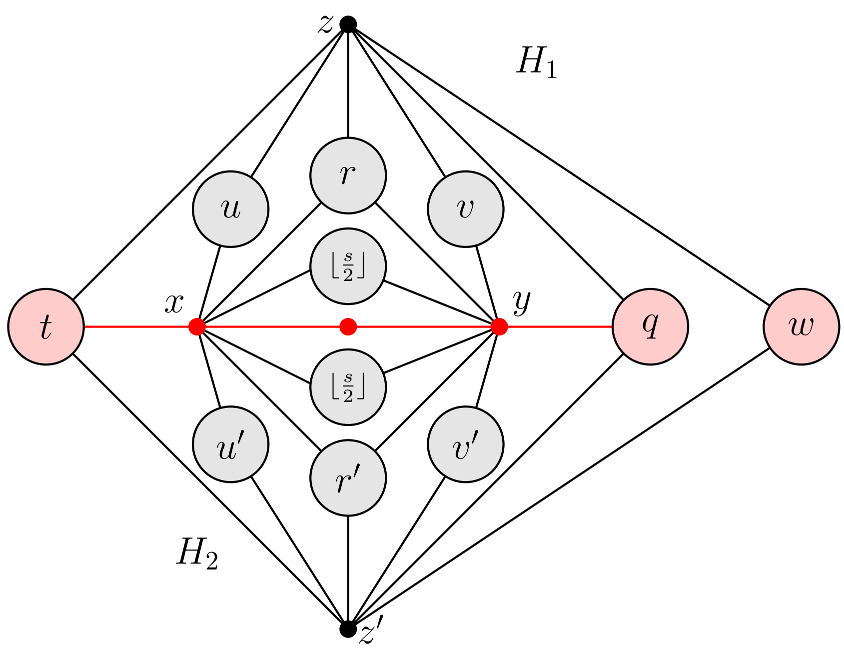

Figure 14: Constructing any bipartite graph with one class $A = \{x, y, z, z'\}$.

assume w.l.o.g. that x and y have $s > 0$ common neighbours and x and z have $u > 0$ common neighbours. Further we know that either $w > 0$ or one of u' and v' is non-zero. If $u' = v' = 0$ then we just take a double star H_2 on z and z' with w neighbours, $H_1 = T_{s,u,v}$ on x,y and z and glue both together on z. This graph is Sidorenko, because the induced probability distribution on a vertex is always κ and therefore we can always glue two Sidorenko graphs on a single vertex. The same works if $w = 0$ and one of v, v' and u' is zero.

Now for $w > 0$ we are left to check the cases where $u' > 0$ or $v' > 0$. If $v = 0$ and $v' = 0$ we can again glue a double star on x and y to the graph $T_{u,u',w}$ on x,z and z'. If $v = 0$ and $u' = 0$ then we need a blown up 8-cycle. We start with two edges with endpoints x and y and glue this on two edges with endpoints x and z'. Next we take another copy of this path of length four with the middle vertex z and glue both together on x and z'. The resulting 8-cycle, where every second vertex is one of x, y, z, z', allows us to add vertices until we have u,s,v' and w neighbours for the respective vertices.

Independently of w the only case left is $v > 0$, $u' > 0$ and $v' > 0$. Then let $H_1 = T_{1,1,\lceil \frac{s}{2} \rceil}$ on x, y, z and $H_2 = T_{1,1,\lceil \frac{s}{2} \rceil}$ on x, z, z' and add w neighbours to z and z' afterwords. We can glue H_1 and H_2 as shown in Figure 14 on x,y, one neighbour if s is odd and the set of size w. Afterwords we add

$u-1$,$v-1$,$u'-1$ and $v'-1$ neighbours to the respective vertices. This is possible because the marginal distribution on two of these vertices did not change during coupling and therefore these constructions are in $\mathfrak{G}_2$

- $l = 1$: Assume w.l.o.g. that x, z and z' do have a common neighbour, i.e. $t > 0$, but $r' = r = q = 0$. We have to distinguish with which vertices y has common neighbours.

  If y only has common neighbours with one other vertex in A, say w.l.o.g. x, then $s > 0$ and $v' = v = 0$. Take the graph H_1 induced by x,z, z' and their common neighbours, which is Sidorenko, since t vertices are complete to the other side. We can glue the double star H_2 on x and y with s common neighbours together with H_1 on x.

  Otherwise y has a common neighbour with at least two other vertices, say w.l.o.g. $v > 0$ and $v' > 0$ with $v > v'$. Then let $H_1 = T_{\lceil \frac{s}{2} \rceil, v, t}$ on x,y,z and $H_2 = T_{\lceil \frac{s}{2} \rceil, v, t}$ on x,y,z' and add w neighbours to z and z' afterwords. We can glue both graphs on x, y, one neighbour if s is odd and the sets of size t and w as shown with red in Figure 14. Then add u neighbours for x and z, u' neighbours for x and z' and the $v - v'$ remaining neighbours for y and z' to complete the graph to H. This is possible, because adding neighbours in both H_i was possible before gluing and thus it is afterwords. Since H_1 and H_2 are in $\mathfrak{G}_1$ and all other added neighbours were glued on independent sets H is in $\mathfrak{G}_2$

- $l \geq 2$: W.l.o.g. we can assume that x, y, z and x, y, z' have a common neighbour, i.e. $r > 0$ and $r' > 0$, where $r \geq r'$. Let H_1 be the graph on x,y,z with neighbour set of size r,t,$\lceil \frac{s}{2} \rceil$,$q$ and w and similarly let H_2 bet the same graph on x,y and z'. We glue both together on the forest spanned on the red vertices. Afterwords we add u,u',v,v' and $r - r'$ neighbours to their respective vertices to complete the graph to H. Again H is in $\mathfrak{G}_2$, because H_1 and H_2 are in $\mathfrak{G}_1$ and all other added neighbours were glued on independent sets.

All together we proved that any H is in $\mathfrak{G}_2$ if $|A| \leq 4$ and therefore satisfies the Sidorenko conjecture.

5.2 What about 5?

After the previous section the immediate question is: What can we say about graphs with at most 5 vertices on one side? Let us look at the building blocks. For the 1-subdivision of K_5 we easily prove that it is in $\mathfrak{G}$ (Figure 15) and later in Section 5.4 we will show that the 1-subdivision of K_m is Sidorenko for all m. The crucial fact will be that we glue on a smaller instance of the graph which is contained in both gluing parts and therefore we will be able to get rid of the negative term.

The complete bipartite graph $K_{5,5}$ minus a perfect matching (for 4 vertices this was the cube) is in $\mathfrak{G}$ as shown in Figure 16. This construction can be

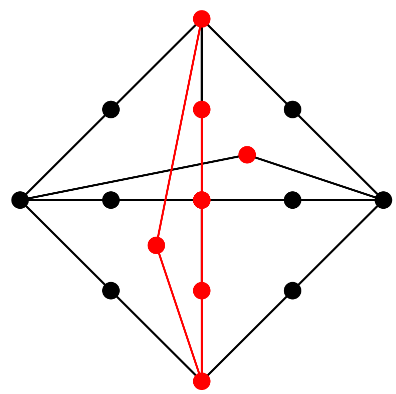

Figure 15: The 1-subdivision of K_5 is in $\mathfrak{G}$ glued together on a 6-cycle plus a single vertex.

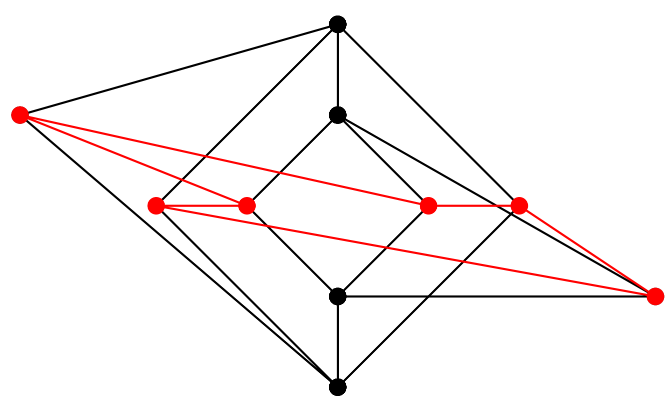

Figure 16: $K_{5,5}$ minus a perfect matching glued on a 6-cycle is in $\mathfrak{G}$.

generalized to the graph $K_{m,m}$ minus a perfect matching for all m in a straight forward way by gluing on the same graph for $m-1$. The problem with proving Sidorenko's conjecture for this graphs is that we glue on a graph which is not an induced subgraph of the parts we glue together, in the above case the 6-cycle. With some extra calculations one can show that for $m \neq 5, 8, 11, \dots$ this graph is Sidorenko, but we do not do it here.

It remains to look at 5 vertices where every three of them have a unique common neighbour. Recall the $K_{5,5}$ minus a ten-cycle is the smallest graph for which the conjecture is open. The graph described before consists of two copies of it glued together on the 5 vertices. We will now show that $K_{5,5} - C_{10}$ is not in $\mathfrak{G}$ as Szegedy [12] claimed and therefore the method in its current state fails for this graph.

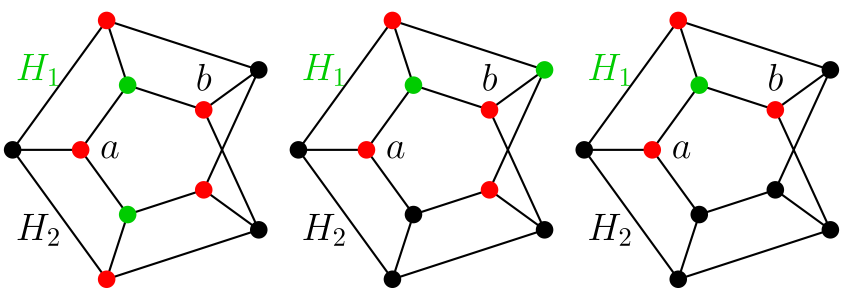

Figure 17: The $K_{5,5}$ minus a C_{10} with independent sets of size 5,4 and 3.

For the start let us prove that $H = K_{5,5} - C_{10}$ is not in $\mathfrak{G}_1$. To prove this we have to consider every independent set A which disconnects H, check all possible ways to partition the remaining vertices into two sets such that both reach A and prove that the induced graphs H_1 and H_2 cannot form a coupling.

The largest possible independent set A is of size 5. After the removal we are left with 5 independent vertices. We cannot take four vertices in one set, because the graph is 3-regular and therefore the fifth would not be able to reach all of A. We also cannot take only two vertices which have two common neighbours into one set, since they also cannot reach one vertex of A. But we can take any two neighbours into one set, which only have one common neighbour. The other three go into the second set. The result for H_1 and H_2 such that the intersection is A is illustrated in Figure 17 on the left. Again we have vertices a and b which have a common neighbour in H_1 but not in H_2. Lemma 4.17 then gives us that there is no coupling possible.

Since three vertices of one side cover the complete other side there is also only one independent set of size 4. After removal of this set we are left with three components, but we need to combine the two singletons to reach all of A because they only have degree three. Up to symmetry there is also only one possibility for a independent set of size three that disconnects H as in Figure 17 on the right. It immediately gives us the partition into two set. In both cases we get vertices a and b are as desired and we conclude with Lemma 4.17.

All other independent sets cannot disconnect H and therefore no partition is possible. Thus $K_{5,5} - C_{10} \notin \mathfrak{G}_1$.

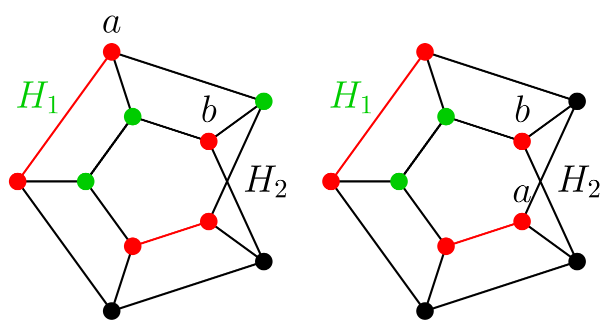

Figure 18: The $K_{5,5}$ minus a C_{10} with a tree that does not use any diagonal edges and disconnects the graph.

Now we want to address $\mathfrak{G}_2$. We need to check all subsets A of the vertices $V(H)$ that span a forest and disconnect H. All independent sets were already considered before and thus we only consider forests with at least one edge. Observe that H can be seen as a 10-cycle with 5 diagonals and thus there are two different kind of edges, the diagonals and the cycle edges.

First look at all forest not using any diagonal edges and at least one cycle edge. Then we can have at most 5 vertices, because the diagonal edges cover all vertices. But 4 vertices which do not induce a diagonal edge are not enough to disconnect the graph. Thus the only possibility is taking to cycle edges and one additional vertex as shown in Figure 18. We are left with three sets, which give us three possible partitions into H_1 and H_2 and we always have vertices a and b as desired.

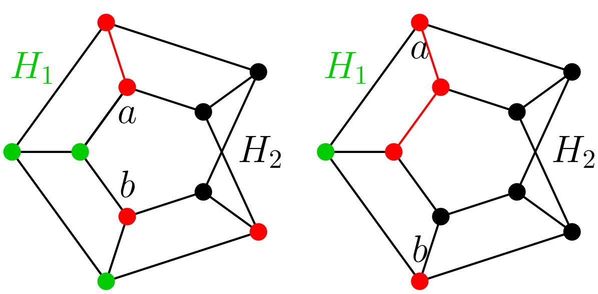

Figure 19: The $K_{5,5}$ minus a C_{10} with a tree that only edge is diagonal and disconnects the graph.

Now we consider forests with exactly one diagonal edge. Removing any diagonal edge leaves us two paths of length 4 with 4 diagonal edges, which we want to disconnect without using another diagonal edge. We can use at most

4 vertices, because the diagonal edges cover all vertices. It is not possible to disconnect the remaining graph with one vertex. If we want to disconnect it by two additional vertices we need to choose two which are on opposite sides on a 4-cycle. Up to symmetry there are two possibilities as shown in Figure 19.

In Figure 20 there are all remaining possibilities for a forest that disconnects the graph using exactly one diagonal edge and vertices a and b in every case. In the first row there are all possibilities two split the remaining graph into two sets. Either we choose an edge in the middle, which forces us to take another vertex as on the top left. We cannot take a fifth because then we have three components. Or we choose an independent vertex, which gives us two possibilities two choose another two edges which leave us with two components.

Now we have to look in where in these constructions we can add another vertex which does not create a second diagonal edge. In the next three rows there are all possible ways where we get three components with all choices for H_1 and H_2. The second row is the only possibility to add a vertex to the first example of Figure 19, the second row contains the only possibility to extend the first two examples of Figure 20 and the third row the only possibility to extend the third example without creating a second diagonal edge. We considered all possibilities starting from a diagonal edge and not having a second.

There cannot be a forest using three diagonal edges, because then we would have a 4-cycle. Therefore it is left to check that all forest with exactly two diagonal edges cannot induce a coupling as shown in Figure 21. We fix any two not neighbouring diagonal edges, which already gives us the first example on the top left. The next two possibilities are all possible choices for one additional vertex up to symmetry.

We can add another vertex to the example in the top middle in two different ways. Either we get the graph on the left of the second row, or the graph in the third row which has three possibilities for H_1 and H_2. The graph on the top right has also two possible additional vertices shown on the right of the second row. Adding any other vertex to one of this examples that does not create a cycle gives us the same example shown in the last row which also gives us a partition into three sets. In all of these examples we have vertices a and b which allow us to apply Lemma 4.17.

For $\mathfrak{G}$ it remains to check all subgraphs that are not forests. So there either has to an induced 4- or 6-cycle. If we remove a 6-cycle we are left with a path of length three. Avoiding a 4-cycle up to symmetry there is only one possibility to disconnect it as shown in Figure 22.

Removing a C_4 from H leaves us with a path of legnth 3 with 3 diagonal edges. We can now disconnect the graph using two more vertices by choosing the middle edge or one of the middle vertices and a non-neighbouring vertex. Both possibilities are treated in Figure 23 in the first row. If we want to use 7 vertices in total there are 3 left so disconnect the remaining graph. We cannot choose two vertices next to the C_4 and thus the only possible choice is shown on the top right. Since we need at least two vertices left the maximum number of vertices that we can remove is 8. For the two vertices left we can choose both from the same side which have one common neighbour, both from the same side

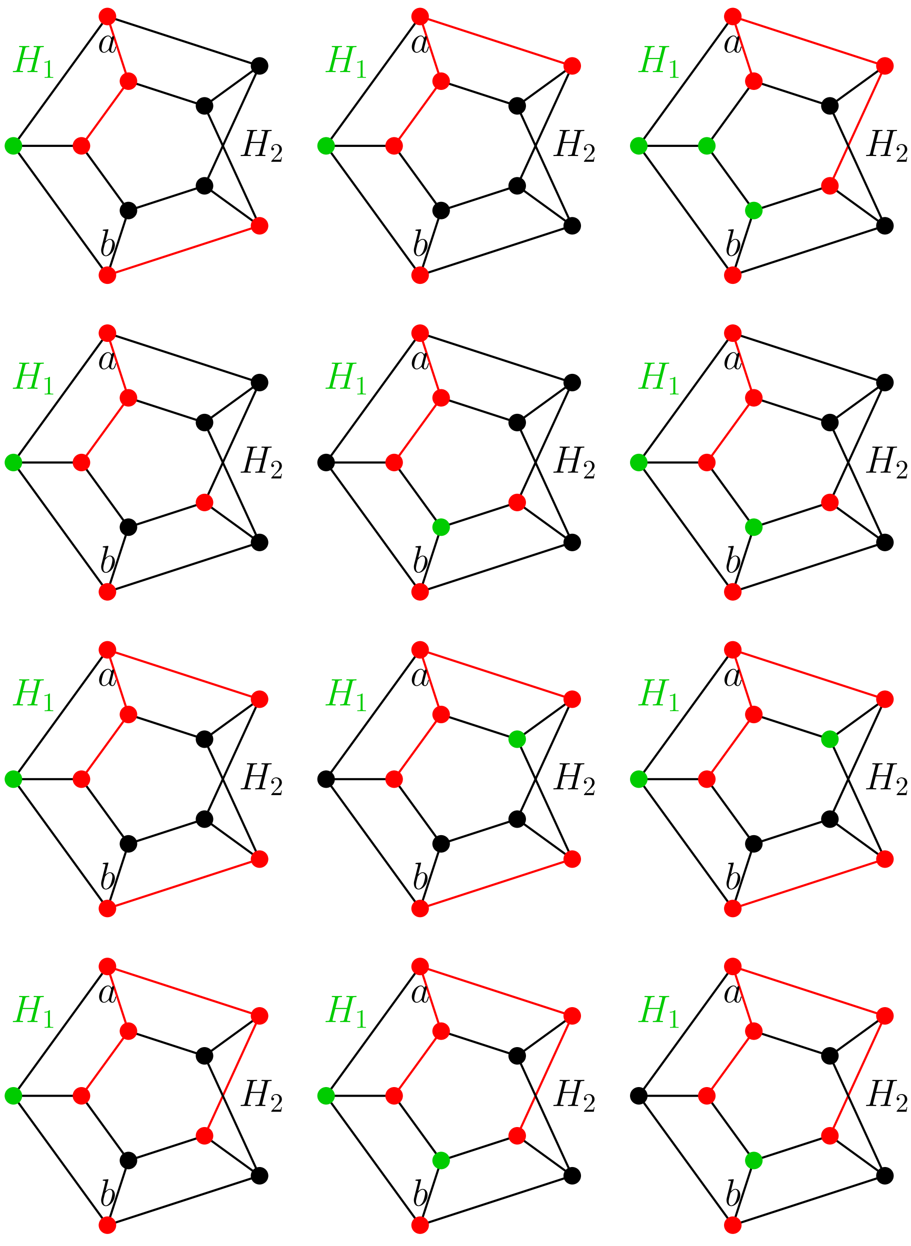

Figure 20: The $K_{5,5}$ minus a C_{10} with any choice for a forest using exactly one diagonal edge and disconnecting the graph.

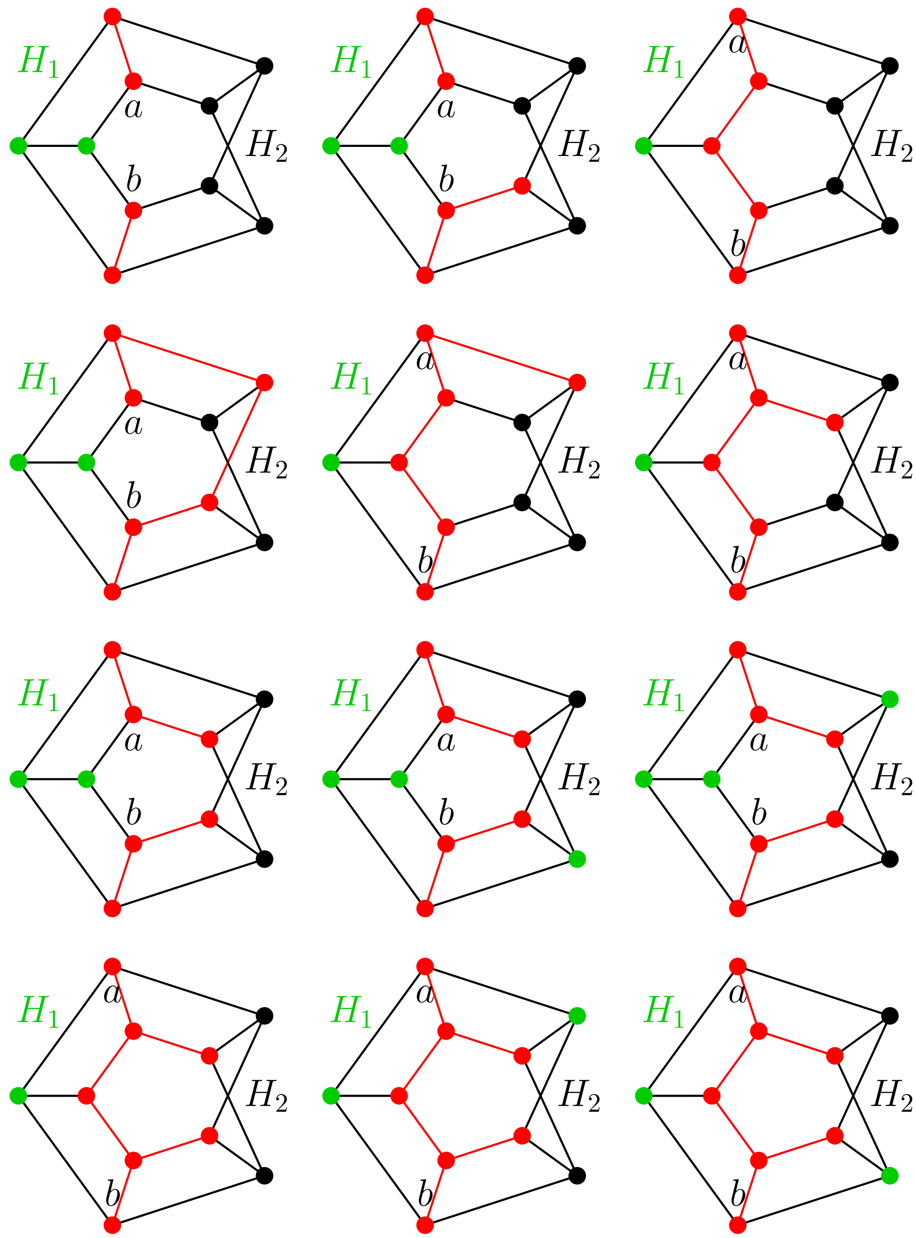

Figure 21: The $K_{5,5}$ minus a C_{10} with any choice for a forest using exactly two diagonal edges and disconnecting the graph.

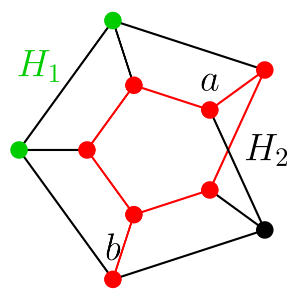

Figure 22: The $K_{5,5}$ minus a C_{10} with any choice for a subgraph that only cycle is of length 6 and is disconnecting the graph.

which have two common neighbours or both from different sides as shown in the bottom of Figure 23. In all cases we get two vertices a and b which have a neighbour in one graph but not the other and allow us to apply Lemma 4.17. Therefore no coupling on a graph is possible and all together we have proved that the $K_{5,5} - C_{10}$ is not in $\mathfrak{G}$.

5.3 All cubes are Sidorenko

As seen in the previous chapter the three dimensional cube is in $\mathfrak{G}_2$. We will now prove that all cubes are in $\mathfrak{G}$ and satisfy the Sidorenko conjecture. This result was first shown by Hatami in [7] who proved that norming graphs are Sidorenko and that all cubes are norming graphs.

We recursively define the **d-dimensional cube** Q_d. Q_0 is just a single vertex. Q_{d+1} is obtained from Q_d by taking to copies of Q_d and adding an edge for every vertex to its copy. Note that the following recursive and explicit formulas hold for the number of vertices and edges respectively

$$|V(Q_{d+1})| = 2|V(Q_d)| \qquad |V(Q_d)| = 2^d$$
$$|E(Q_{d+1})| = 2|E(Q_d)| + |V(Q_d)| \qquad |E(Q_d)| = d2^{d-1}$$

We want to recursively construct probability models for the cube Q_d. We cannot use the above constructions. Start with the probability model $f_1 = (G \mapsto \tau(e, G))$ with skeleton Q_1. Now we construct f_{d+1} from f_d. Take the conditionally independent coupling f of twice f_d where $(\beta_i : [2^{d-1}] \to V(Q_{d-1}))_{i=1,2}$ as shown in Figure 24. Then take the conditionally independent coupling f_{d+1} of twice this f where we define $(\delta_i : [2^d] \to V(Q_{d-1})\dot\cup V(Q_{d-1}))_{i=1,2}$ as in Figure 25. The skeleton of f_{d+1} is Q_{d+1}. This constructions shows that f_d is in $\mathfrak{G}$ for all d.

Theorem 5.1. f_d is a family of witness measures for all d and consequently all cubes are Sidorenko.

Proof. We want to obtain the same inequality (8) as in Theorem 2.1, but we cannot apply Proposition 4.8 since we do not glue on forests. Observe that

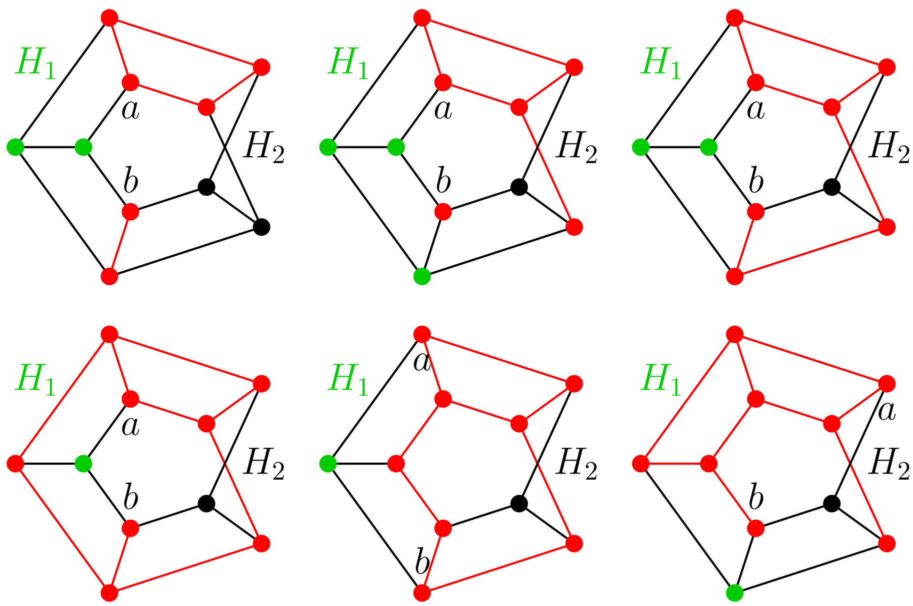

Figure 23: The $K_{5,5}$ minus a C_{10} with any other choice for a subgraph that is not a forest and disconnecting the graph.

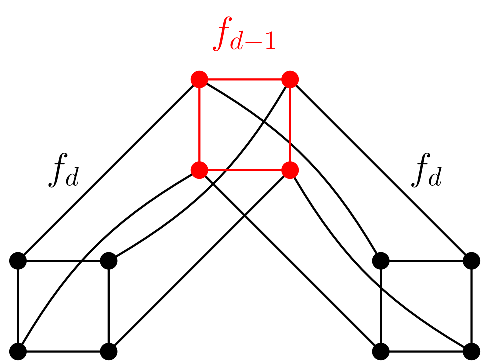

Figure 24: Gluing together twice the probability model f_d on the red vertices to get f. Example for $d = 3$.

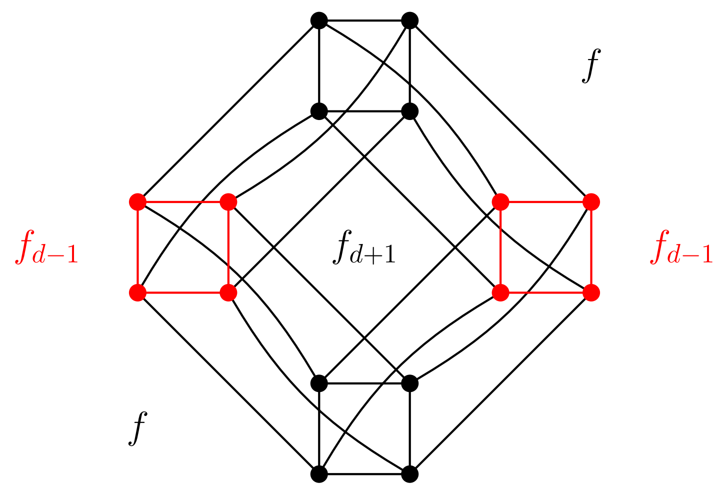

Figure 25: Gluing together twice the probability model f on the red vertices to get f_{d+1}. Example for $d = 3$.

we always glue on two disjoint cubes which gives us the possibility to obtain a recurrence relation.

From the definition of $\mathfrak{U}$ we know $D(f_0) = D_v$ and $D(f_1) = D_e$. Using what we obtained above and that $f|_{\beta_i} = f_{d-1}$ we achieve

$$\begin{aligned} D(f_{d+1}) &= D(f) + D(f) - D(f_{d+1}|_{\delta_i}) \\ &= 4D(f_d) - 2D(f|_{\beta_i}) - D(f_{d+1}|_{\delta_i}) \\ &= 4D(f_d) - 2D(f_{d-1}) - D(f_{d+1}|_{\delta_i}) \end{aligned}$$

The last term is the relative entropy of a coupling of two disjoint cubes of dimension $d-1$. To see this let $\mu_1 = f_{d+1}|_{\delta_i}|_{V(Q_{d-1})} = f_{d-1}$ and $\mu_2 = f_{d+1}|_{\delta_i}|_{V(Q_{d-1})} = f_{d-1}$ be the marginal distribution on both cubes, then $f_{d+1}|_{\delta_i}$ is a coupling of μ_1 and μ_2. Using Lemma 4.2 we can lower bound the relative entropy of this coupling by the relative entropy of the conditional independent coupling μ_4 of μ_1 and μ_2 over $f_{d+1}|_{\emptyset}$. Then

$$D(f_{d+1}|_{\delta_i}) \overset{(7)}{\geq} D(\mu_4) \overset{(6)}{=} 2D(f_{d-1}) - D(f_{d-1}|_{\emptyset}) = 2D(f_{d-1})$$

and therefore we get

$$D(f_{d+1}) \leq 4D(f_d) - 4D(f_{d-1}).$$

Repeated application for $i < d$ yields

$$\begin{aligned}
D(f_d) &\leq \cdots \leq i2^{i-1}D(f_{d-i}) - (i-1)2^i D(f_{d-1-i}) \\
&\leq i2^{i-1}(4D(f_{d-i-1}) - 4D(f_{d-i-2}))) - (i-1)2^i D(f_{d-1-i}) \\
&= (i2^{i-1}4 - (i-1)2^i)D(f_{d-i-1}) - i2^{i-1}4D(f_{d-i-2}) \\
&= (i+1)2^{(i+1)-1}D(f_{d-(i+1)}) - ((i+1)-1)2^{i+1}D(f_{d-1-(i+1)})
\end{aligned}$$

and so finally we get

$$D(f_d) \leq D_e d 2^{d-1} - D_v(d-1)2^d \leq |E(Q_d)|D_e,$$

which is precisely the same as Equation (8). Therefore Q_d is Sidorenko for all d. □

The proof for the d-dimensional cube can be generalized to any grid in any dimension using the recursion from above. One has to be careful and ensure that all the minus terms are cancelled. This result was also proven before by Kim, Lee and Lee [8] in their second extension to Cartesian products. For a tree T and bipartite graph H the Cartesian product $T \times H$ is a graph on $V(T) \times V(H)$ where (v_1, u_1) and (v_2, u_2) are adjacent if $u_1 = u_2$ and $\{v_1, v_2\} \in E(T)$ or $v_1 = v_2$ and $\{u_1, u_2\} \in E(H)$. They prove the following:

Theorem 5.2. If T is a tree and H is a bipartite graph having Sidorenko's property, then $T \times H$ also has Sidorenko's property.

We will discuss this with Szegedy's new method in Section 5.5. Taking paths of various length $P_1, P_2, \ldots, P_d$ which are Sidorenko gives us that the d-dimensional grid $P_1 \times P_2 \times \cdots \times P_n$ is again Sidorenko. In particular this also implies that all cubes $Q_d = K_2 \times K_2 \times \cdots \times K_2$ are Sidorenko.

We proved that Q_d is in $\mathfrak{G}$ for all d and earlier that $Q_3 \in \mathfrak{G}_2 \setminus \mathfrak{G}_1$. The question is what is the maximum d such that Q_d is in $\mathfrak{G}_2$. Or does this hold for all d? In Figure 26 we give the only possible construction for the 4-dimensional cube, using couplings over trees only. None of the methods we established suggest that this is not working, but we do not give a proof either. The red vertices induce a forest and give us the graphs H_1 and H_2 shown below. It remains to be shown that the induced probability distributions on the red vertices in H_1 and H_2 are the same, which in principal could be done by explicitly computing them in the sense of Example 4.4.

This construction really looks promising, which is in accordance with Szegedy, who mentioned in private communication that he thinks that cubes up to dimension 6 are in $\mathfrak{G}_2$.

5.4 1-subdivision of K_m is Sidorenko for all m

Similar to the cubes we can also deal with the 1-subdivision of K_m. First define the **1-subdivision** of K_m as the graph S_m which is obtained from the complete graph K_m by splitting every edge into two. A recursive construction becomes

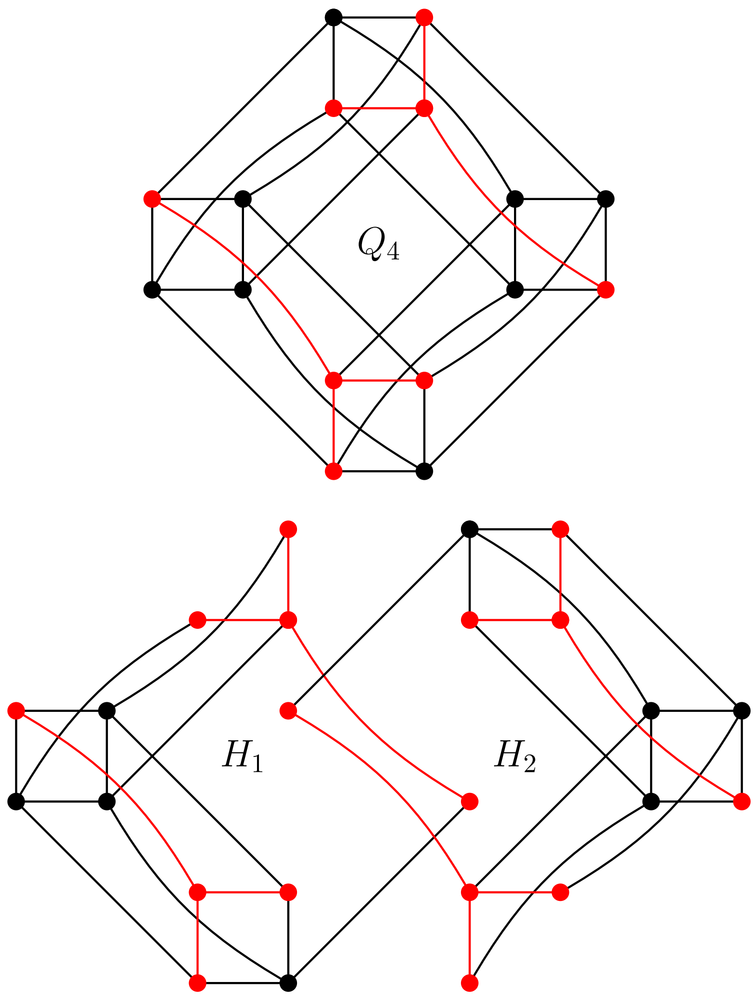

Figure 26: Possible construction to show that $Q_4 \in \mathfrak{G}_2$.

clear when we construct a probability model corresponding to S_m. The graph S_m has $m+\binom{m}{2}$ vertices (m for the vertices of K_m and one for every edge) and $2\binom{m}{2}$ (every edge is split in two) edges.

S_1 is just a single vertex, S_2 two adjoined edges and S_3 a 6-cycle. Starting from S_1 we can recursively construct a probability model for S_{m+1} by taking two copies of S_m, joining one edge to a corner and then gluing both together via $\{\beta_i\}_{i=1,2}$ on a S_{m-1} and a single vertex. We have already seen this construction for $m=4$ in Figure 11 and for $m=5$ in Figure 15. Let f_m be the probability model corresponding to S_m and f for S_m plus one adjoined edge. Again we want that the relative entropy of some terms cancel each other. Observe that we get $D(f)=D(f_m)+De-D_v$ by Lemma 4.1 and $D(f|_{\beta_i})\geq D(f_{m-1})+D_v$ by also using Lemma 4.2. So now we can do the following computation

$$\begin{aligned} D(f_{m+1}) &\overset{(6)}{=} D(f)+D(f)-D(f|_{\beta_i}) \\ &\leq 2D(f_m)-D(f_{m-1})+2D_e-3D_v. \end{aligned}$$

Repeated application for $i<m$ gives us

$$\begin{aligned} &D(f_m)\leq\dots \\ &\leq (i+1)D(f_{m-i})-iD(f_{m-i-1})+2\binom{i+1}{2}D_e-3\binom{i+1}{2}D_v \\ &\leq (i+1)[2D(f_{m-i-1})-D(f_{m-i-2})+2D_e-3D_v] \\ &-iD(f_{m-i-1})+2\binom{i+1}{2}D_e-\binom{i+1}{2}D_v \\ &=(i+2)D(f_{m-(i+1)})-(i+1)D(f_{m-(i+1)-1})+2\binom{i+2}{2}D_e-3\binom{i+2}{2}D_v. \end{aligned}$$

Finally after $m-3$ steps, using $D(f_1)=D_v$ and $D(f_2)=2D_e-D_v$ we get

$$\begin{aligned} &D(f_m) \\ &\leq\left(2\binom{m-1}{2}+2(m-1)\right)D_e-\left((m-1)+(m-2)+3\binom{m-1}{2}\right)D_v \\ &=2\binom{m}{2}D_e-\left(3\binom{m}{2}-m\right)D_v \\ &=|E(S_m)|D_e-(2|E(S_m)|-|V(S_m)|)D_v \end{aligned}$$

and therefore S_m is Sidorenko. This result was not proven before in any of the present resources.

5.5 A question of Szegedy

In this part we want to give a partial answer to the question of Szegedy [12] what the relationship between the results from Kim, Lee and Lee [8] and his new method is. First we will proof that their first approach to Sidorenko's conjecture is covered by Szegedys method, i.e. tree-arrangeable graphs are in $\mathfrak{G}_1$. Recall that a bipartite graph is tree-arrangeable if there exists a bipartition $A\dot{\cup}B$ and a tree T on A such that for all vertices $u, v \in A$

$$N(u) \cap N(v) = \bigcap_{w\in P} N(w)$$

for any path P in T connecting u and v, where $N(u)$ is the set of neighbours of u in H.

Theorem 5.3. All tree-arrangeable graphs are in $\mathfrak{G}_1$.

Proof. Let H be a tree-arrangeable graph with $V(H) = A\dot{\cup}B$, and T a tree on $A = \{a_1, \dots, a_n\}$ with the above property. We recursively build H starting from a vertex a_1 with neighbours $N(a_1)$ and following T. At step i we have the graph H_{i-1} which already contains all vertices $a_1, \dots, a_{i-1}$ and their neighbours. Assume that a_i is a neighbour of a_{i-1} in T and let T' be the graph induced by $a_1, \dots, a_i$ in T.

Now take a star on a_i with neighbours $N(a_i)$. We want to glue all vertices in $N(a_i)$ onto the respective vertices in H_{i-1} if they already exist. Let $N_i \subseteq N(a_i)$ be all the neighbours of a_i present in H_{i-1}. Thus we have to check that the marginal distributions on N_i are the same in the probability model for H_i and the star.

We know that for every vertex a_j for $j = 0, \dots, i-1$ and any path from a_i to a_j in T'

$$N(a_i) \cap N(a_j) = \bigcap_{w\in P} N(w).$$

Since a_i is a leaf in T' every path from a_i to another vertex has to use a_{i-1}. This implies that for all $j = 1, \dots, i-1$ we get $N_i \cap N(a_j) \subseteq N(a_i) \cap N(a_j) \subseteq N(a_{i-1})$. Using that $V(H_{i-1}) = \bigcup_{j=1}^{i-1}$ we get $N_i \subseteq N(a_{i-1})$. Since all vertices in N_i are neighbours of a_{i-1} the marginal distribution is the same as in the star centered at a_i. So we can glue the star to H_{i-1} on N_i giving us H_i.

After n steps we get the graph $H_n = H$ and since we only glued on independent sets $H \in \mathfrak{G}_1$. □

Next we want to investigate the second approach with Szegedy's methods. Taking any Sidorenko graph we cannot do anything with this method, because we need a probability model for H. So at least we need to assume that H is in $\mathfrak{G}$ with probability model f and $D(f) \leq eD_e$.

We will use a slightly different construction for $T \times H$. Let $\{T_v\}_{v\in V(H)}$ be vertex-disjoint copies of T. For every edge $\{v_1, v_2\} \in E(H)$ we place an edge between every vertex in T_{v_1} and T_{v_2}, so that they form a copy of $T \times K_2$ as shown in Figure 27.

We want to check that this construction gives us $T \times H$. Here a vertex $(v, u) \in V(T \times H)$ corresponds to the vertex v in $V(T_u)$. If $\{(v_1, u_1), (v_2, u_2)\}$ is an edge in $T \times H$ then there are two cases possible: Either $u_1 = u_2$ and

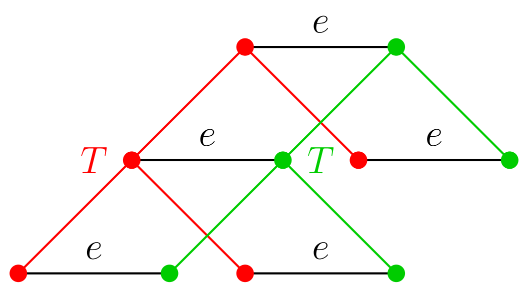

Figure 27: The Cartesian product of a tree T with a single edge e.

$\{v_1, v_2\} \in E(T)$, then there is an edge between v_1 and v_2 in T_{u_1}. Or $v_1 = v_2$ and $\{u_1, u_2\} \in E(H)$, then there is an edge between v_1 in T_{u_1} and v_1 in T_{u_2} because of the edge in H. The other way round every edge added in the construction above corresponds to an edge in $T \times H$ by similar arguments. Therefore the second construction gives us the same graph.

Let T be a tree and $H \in \mathfrak{G}$. What can we say about $T \times H$? We have a probability model $f \in \mathfrak{U}$ for H which is constructed from the uniform random edge model using conditionally independent couplings. We will try to reproduce this construction not starting from $\tau(e, G)$ but a probability model on $T \times K_2$ to get a probability model g for $T \times H$. We get a starting probability model g_0 for $T \times K_2$ in $\mathfrak{U}_2$ by just gluing 4-cycles together along edges as shown in Figure 27.

In the first step of the construction for f two edges $\tau(e, G)$ are glued together on a vertex. Therefore we take to copies of $T \times K_2$ with probability models g_0 and glue them along the tree T. This gives us a probability model in $\mathfrak{U}_2$.

In any other step of the construction for f two graphs H_1 and H_2 with probability models f_1 and f_2 in $\mathfrak{U}$ are glued together on a joint vertex factor $\{\beta_i\}_{i=1,2}$, i.e. $f_1|_{\beta_1} = f_2|_{\beta_2}$. The labelled vertices induce a graph H_0 and we get the probability model f' in $\mathfrak{U}$ with skeleton H'.

Therefore we take the two probability models g_1 and g_2 in $\mathfrak{U}$ with skeletons $T \times H_1$ and $T \times H_2$. We want to glue them together on a joint vertex factor $\{\eta_i\}_{i=1,2}$ to get the probability model g' in $\mathfrak{U}$ with skeleton $T \times H'$. We can construct $\{\eta_i : [n \cdot |V(T)|] \to V(T \times H_i)\}_{i=1,2}$ by extending $\{\beta_i : [n] \to V(H_i)\}_{i=1,2}$ from H_0 to $T \times H_0$ in the canonical way.

We need to prove that $g_1|_{\eta_1} = g_2|_{\eta_2}$, which seems plausible since we obtained g_1 and g_2 from g_0 in the same way as f_1 and f_2 from $\tau(e, G)$. Of course this is true if we glue a graph to itself or a former subgraph as its the case in all of our constructions, but in general it needs more precise investigations to formalize this statement.

Assume for the moment that we managed to prove that $g' \in \mathfrak{U}$ in all cases. Then we would face another problem: Assuming that $H' \in \mathfrak{G}$ is Sidorenko we have

$$D(f') = D(f_1) + D(f_2) - D(f_1|_{\beta_1}) \leq |E(H')|D_e$$

and we need

$$\begin{aligned} D(g') &= D(g_1) + D(g_2) - D(g_1|_{\eta_1}) \\ &\leq (|E(H')| \cdot |V(T)| + |V(H')| \cdot |E(T)|)D_e = |E(T \times H')|D_e. \end{aligned}$$

We are also not able to show this at the moment, because we need a lower bound for $D(g_1|_{\eta_1})$ where $T \times H'$ might contain cycles even if H' did not. To get a lower bound we would need to know more about the calculation of $D(f')$ and possible ways to get rid of the negative terms as we did for the cubes.

On the other hand this works if $f \in \mathfrak{U}_1$. Then we only glue on independent sets H_0 and therefore $T \times H_0$ would always be a forest and thus $g' \in \mathfrak{U}_2$. Summing this up we have proved that for any tree T and a graph H with probability model $f \in \mathfrak{U}_1$, which was constructed solely by reflection and gluing on former subgraphs, the Cartesian product $T \times H$ is in $\mathfrak{G}_2$ and therefore Sidorenko. Possible examples for H are trees, even cycles and bipartite graphs where one vertex is complete to the other side.

6 Concluding remarks

In the previous two sections we proved Sidorenko's conjecture for two classes of graphs in $\mathfrak{G}$, the d-dimensional cubes and the 1-subdivision of K_m. The constructions of the probability models used conditionally independent couplings over graphs that are not forests. The proofs did go through because we did glue on smaller instances of the graphs which appeared in both gluing parts. This way the negative terms were equalized by the positive terms. A proof like this is possible if we have a recursive construction for the graphs which gives us the right cancellations.

With some more detailed analysis of the relative entropy of the 6-cycle we should be able to prove that the $K_{5,5}$ minus a perfect matching is Sidorenko. An even beyond this there are much more classes of graphs in $\mathfrak{G}$ which allow a recursive construction and therefore the possibility to construct the probability measures in $\mathfrak{U}$.

To finish we list some more remarks and give ideas for future studies. Szegedy claims [12] that with his method one can prove the equivalent conjecture for some classes of 3-uniform hypergraphs, but note that Sidorenko proved that the conjecture fails in general [11].

We proved that the extension by Kim, Lee and Lee [8] to tree-arrangeable graphs is covered by Szegedys method. Even though the Cartesian product does not seem to go beyond $\mathfrak{G}$ we are so far not able to prove it with the method of Szegedy, because of the generality and complexity of the statement.

To really understand the impact of the method of Szegedy completely it would be necessary to find a complete combinatorial characterization of $\mathfrak{G}_1$, $\mathfrak{G}_2$

and $\mathfrak{G}$. For this we would need to extend the arguments given in Section 4.3 and 4.4 and explore what combinatorial properties make the difference between $Q_3 \in \mathfrak{G}_2 \setminus \mathfrak{G}_1$ and graphs that are in $\mathfrak{G}_1$.

The main tool of the method is the inclusion-exclusion formula (Lemma 4.1) for conditionally independent couplings. This immediately suggest an extension to a coupling of three or more probability distributions in the sense of the usual inclusion exclusion formula for multiple sets. This might be one approach to avoid the application of Lemma 4.17.

On the other hand we proved that $K_{5,5} - C_{10}$ is not in $\mathfrak{G}$. Since it is very hard to surpass $\mathfrak{G}$ with the coupling method and this seems to be the right tool for dealing with densities on graph homomorphisms this raises the provocative questions if Sidorenko's conjecture fails outside $\mathfrak{G}$. In contrast to this it would be really nice to prove the conjecture for any graph not in $\mathfrak{G}$.

Acknowledgements

I would like to express my gratitude to my thesis advisor Tibor Szabó who introduced me to the topic and guided me through the thesis. I also want to thank Balázs Szegedy for answering my questions about his papers and my family and friends who supported me.

References

[1] George R Blakley and Prabir Roy. A hölder type inequality for symmetric matrices with nonnegative entries. *Proceedings of the American Mathematical Society*, 16(6):1244–1245, 1965.

[2] Béla Bollobás. *Extremal graph theory*. Courier Dover Publications, 2004.

[3] Fan R. K. Chung, Ronald L. Graham, and Richard M. Wilson. Quasi-random graphs. *Combinatorica*, 9(4):345–362, 1989.

[4] David Conlon, Jacob Fox, and Benny Sudakov. An approximate version of sidorenko's conjecture. *Geometric and Functional Analysis*, 20(6):1354–1366, 2010.

[5] P Erdős and A Rényi. On random graphs i. *Publ. Math. Debrecen*, 6:290–297, 1959.

[6] Robert M Gray. *Entropy and information theory*, volume 1. Springer, 2011.

[7] Hamed Hatami. Graph norms and sidorenko's conjecture. *Israel Journal of Mathematics*, 175(1):125–150, 2010.

[8] Jeong Han Kim, Choongbum Lee, and Joonkyung Lee. Two approaches to sidorenko's conjecture. *arXiv preprint arXiv:1310.4383*, 2013.

[9] JL Li and Balazs Szegedy. On the logarithimic calculus and sidorenko's conjecture. *arXiv preprint arXiv:1107.1153*, 2011.

[10] Alexander F Sidorenko. Inequalities for functionals generated by bipartite graphs. *Diskretnaya Matematika*, 3(3):50–65, 1991.

[11] Alexander F Sidorenko. A correlation inequality for bipartite graphs. *Graphs and Combinatorics*, 9(2-4):201–204, 1993.

[12] Balazs Szegedy. Relative entropy and sidorenko's conjecture. *arXiv preprint arXiv:1406.6738*, 2014.

Eidesstattliche Erklärung zur Masterarbeit

Ich versichere, die Masterarbeit selbstständig und lediglich unter Benutzung der angegebenen Quellen und Hilfsmittel verfasst zu haben.
Ich erkläre weiterhin, dass die vorliegende Arbeit noch nicht im Rahmen eines anderen Prüfungsverfahrens eingereicht wurde.

Berlin, 18. November 2014