

CONSTRAINTS, MMSNP AND EXPANDER RELATIONAL STRUCTURES

GÁBOR KUN

ABSTRACT. We give a poly-time construction for a combinatorial classic known as Sparse Incomparability Lemma, studied by Erdős, Lovász, Nešetřil, Rödl and others: We show that every Constraint Satisfaction Problem is poly-time equivalent to its restriction to structures with large girth. This implies that the complexity classes CSP and Monotone Monadic Strict NP introduced by Feder and Vardi are computationally equivalent. The technical novelty of the paper is a concept of expander relations and a new type of product for relational structures: a generalization of the zig-zag product, the twisted product.

1. INTRODUCTION

The construction of graphs with large girth and chromatic number is a classic in probabilistic combinatorics. Many great mathematicians have contributed to this: Erdős [4] gave a probabilistic construction for graphs. Lovász [10] had a deterministic, but huge construction for hypergraphs. Nešetřil and Rödl gave a short probabilistic construction [16] for hypergraphs, see also Duffus, Rödl, Sands and Sauer[3]. Feder and Vardi showed [5] a more general statement known as Sparse Incomparability Lemma: They proved that for every CSP problem there is a randomized, poly-time algorithm that transforms every input structure of the CSP to an equivalent one of large girth.

Nešetřil and Matoušek gave a deterministic poly-time algorithm [15] to this in case of graphs. This was simplified in the recent work of Nešetřil and Siggers [17]. The main result of this paper is a deterministic, poly-time algorithm for the Sparse Incomparability Lemma for general CSP's.¹

Theorem 1. (*Algorithm*) *Let t, k be positive integers and τ a finite relational type. For every structure \mathbf{S} of type τ there exists a polynomial*

This research was supported by OTKA Grants no. T043671 and NK 67867, Subhash Khot's NSF Waterman Award CCF-1061938 and the MTA Rényi "Lendület" Groups and Graphs Research Group.

¹We will follow the terminology of relational structures, but our theorems hold in the special case of hypergraphs, too.

time constructible structure \mathbf{S}' of type τ with girth $> k$ such that for every structure \mathbf{T} of size $< t$ the equivalence $\mathbf{S} \in \text{CSP}(\mathbf{T}) \iff \mathbf{S}' \in \text{CSP}(\mathbf{T})$ holds. Moreover, $\mathbf{S}' \in \text{CSP}(\mathbf{S})$.

This theorem also answers a problem posed by Nešetřil, Kostochka and Smolíkova [7]. However, our work was primarily motivated by the paper of Feder and Vardi [5] on the dichotomy conjecture: They analyzed the complexity classes CSP and Monotone Monadic Strict NP (MMSNP). The class MMSNP contains the class CSP, and it has much bigger expressive power. Feder and Vardi proved that these classes are equivalent in a random sense. (For more on these classes and a simple proof see [9].) The only random part in their algorithm comes from their probabilistic proof of the Sparse Incomparability Lemma, so we can derandomize their result using Theorem 1.

Theorem 2. *Let τ be a finite relational type, $L \subseteq \text{Rel}(\tau)$ an MMSNP language. Then there is a finite set of relational structures $\mathcal{T} \subset \text{Rel}(\tau)$ such that*

- (1) *L has a polynomial time reduction to $\text{CSP}(\mathcal{T})$.*
- (2) *$\text{CSP}(\mathcal{T})$ has a polynomial time reduction to L .*

Note that the equivalence of the complexity classes CSP and MMSNP does not only mean that both of these classes contain an NP-complete problem. In particular, Theorem 2 shows that if dichotomy holds for CSP then it also holds for MMSNP.

Nešetřil and Matoušek used expander graphs to give a poly-time algorithm for the Sparse Incomparability Lemma in the case of graphs. Expander graphs are sparse but highly connected graphs. These play an important role in number theory, group theory and graph theory. Ajtai, Komlós and Szemerédi used expanders in their paper on parallel sorting [1]: this was the first time when expanders were used in computer science. "Optimal expander graphs", Ramanujan graphs were constructed by Margulis [14] and independently by Lubotzky, Phillips and Sarnak [11]. Simpler and simpler constructions were found in the last decade [2, 18]. Recently Lubotzky, Samuels and Vishne introduced a concept of Ramanujan complexes [12, 13].

On the other hand, for relational structures (hypergraphs) no similar construction or even definition is known. We introduce a concept of *expander relations*. We say that the r -ary relation R on S is an ε -expander relation if for every $S_1, \dots, S_r \subseteq S$ the number of relational tuples with the i th coordinate in S_i differs by less than $\varepsilon|R|$ from the expected value. We construct ε -expander relational structures with large girth and bounded degree in poly-time.

Theorem 3. (*Algorithm*) Let τ be a finite relational type, k a positive integer and $\varepsilon > 0$. Then for every $n > n_{\tau, \varepsilon, k}$ there exists a polynomial time constructible ε -expander \mathbf{S} of size n , type τ , maximal degree at most $M = M_{\tau, \varepsilon}$ and girth at least k .

In order to give this construction we define the *twisted product* of relational structures, a generalization of the so-called zig-zag product used by Reingold, Vadhan and Widgerson [18]. Alon, Schwartz and Shapira used a similar product called replacement product in their expander construction [2].

In Section 2 we give the basic definitions. Section 3 contains the novelties of this paper: the definition and properties of expander relational structures and the twisted product. In Section 4 we construct expander relational structures with large girth and bounded degree. In Section 5 we prove Theorem 1 and Theorem 2.

2. DEFINITIONS, NOTATIONS

We will work with finite relational structures throughout this paper: we denote these by boldface letters $\mathbf{A}, \mathbf{B}, \mathbf{C}, \dots$ and their base set by A, B, C, \dots , respectively. For an r -ary relational symbol R and relational structure \mathbf{A} with base set A let $R = R(\mathbf{A})$ denote the set of tuples of \mathbf{A} which are in relation R . Recall, that a homomorphism is a mapping which preserves all relations. Just to be explicit, for relational structures \mathbf{A}, \mathbf{B} of the same type τ a mapping $f : A \rightarrow B$ is a *homomorphism* $\mathbf{A} \rightarrow \mathbf{B}$ if for every r -ary relational symbol $R \in \tau$ and $(x_1, \dots, x_r) \in A^r$ the implication $(x_1, \dots, x_r) \in R(\mathbf{A}) \rightarrow (f(x_1), \dots, f(x_r)) \in R(\mathbf{B})$ holds. A *cycle* in a relational structure \mathbf{A} is either a minimal sequence of distinct points and distinct tuples $x_0, r_1, x_1, \dots, r_t, x_t = x_0$ where $t \geq 2$, each tuple r_i belongs to one of the relations $R(\mathbf{A})$ and each x_i is a coordinate of r_i and r_{i+1} , or, in the degenerate case, a relational tuple with at least one multiple coordinate. The *length* of the cycle is the integer t in the first case and 1 in the second case. The *girth* of a structure \mathbf{A} is the length of the shortest cycle in \mathbf{A} (if it exists; otherwise it is a forest and we define the girth to be infinity). The degree of an element x of \mathbf{S} is the number of relational tuples containing x (with multiplicity). Denote the maximal degree in \mathbf{S} by $\Delta(\mathbf{S})$. Given a function $f : S \rightarrow \mathbb{R}$ let $|f| = \sum_{x \in S} |f(x)|$ denote its first norm and $\max(f)$ its maximum, respectively.

For the relational structure \mathbf{A} set $CSP(\mathbf{A}) = \{\mathbf{B} : \mathbf{B} \text{ is homomorphic to } \mathbf{A}\}$. For a finite set \mathcal{T} of relational structures of the same type define

$CSP(\mathcal{T}) = \cup_{\mathbf{A} \in \mathcal{T}} CSP(\mathbf{A})$. The class CSP consists of languages of the form $CSP(\mathcal{T})$.

3. EXPANDER RELATIONS AND THE TWISTED PRODUCT

Definition 4. Given a finite relational structure \mathbf{A} , a relation $R \subseteq A^r$ and functions $f_1, \dots, f_r : A \rightarrow \mathbb{R}$ let us denote the sum

$$\sum_{(x_1, \dots, x_r) \in R(\mathbf{A})} \prod_{i=1}^r f_i(x_i)$$

by $R(f_1, \dots, f_r)$. For the subsets $S_1, \dots, S_r \subseteq A$ set $R(S_1, \dots, S_r) = R(\chi_{S_1}, \dots, \chi_{S_r})$. This equals the number of r -tuples $(x_1, \dots, x_r) \in R$ such that $x_1 \in S_1, \dots, x_r \in S_r$.

Definition 5. A nonempty r -ary relation $R \subseteq S^r$ is called an ε -expander relation if for every $S_1, \dots, S_r \subseteq S$ the inequality

$$\left| R(S_1, \dots, S_r) - |R| \frac{\prod_{i=1}^r |S_i|}{|S|^r} \right| \leq \varepsilon |R| \text{ holds.}$$

A relational structure \mathbf{S} is a (Δ, ε) -expander relational structure if every at least binary relation of \mathbf{S} is an ε -expander relation and $\Delta(\mathbf{S}) \leq \Delta$.

An expander graph is an expander relational structure: this is a trivial consequence of the Expander Mixing Lemma [6]. We postpone the study of such expanders until Section 4. Now we give several equivalent definitions for expander relations.

Lemma 6. For a finite r -ary relation $R \subseteq S^r$ the followings are equivalent.

- (1) For every $f_1, \dots, f_r : S \rightarrow [0; \infty)$,
 $\left| R(f_1, \dots, f_r) - |R| \frac{\prod_{i=1}^r |f_i|}{|S|^r} \right| \leq \varepsilon |R| \prod_{i=1}^r \max(f_i)$ holds.
- (2) For every $f_1, \dots, f_r : S \rightarrow [0; 1]$,
 $\left| R(f_1, \dots, f_r) - |R| \frac{\prod_{i=1}^r |f_i|}{|S|^r} \right| \leq \varepsilon |R|$ holds.
- (3) R is an ε -expander relation.

Proof. The implication (1) \rightarrow (2) is trivial. (3) is the special case of (2) when all the functions f_i are characteristic functions. We have to prove (3) \rightarrow (1):

$$\begin{aligned} & \left| R(f_1, \dots, f_r) - \prod_{i=1}^r |f_i| \frac{|R|}{|S|^r} \right| = \\ & \left| \int_{y_1=0}^{\max(f_1)} \dots \int_{y_r=0}^{\max(f_r)} \sum_{(e_1, \dots, e_r) \in R} \prod_{i=1}^r \chi_{\{y_i < f_i(e_i)\}} dy_1 \dots dy_r - \right. \\ & \left. \int_{y_1=0}^{\max(f_1)} \dots \int_{y_r=0}^{\max(f_r)} \frac{|R|}{|S|^r} \prod_{i=1}^r \left(\sum_{s \in S} \chi_{\{y_i < f_i(s)\}} \right) dy_1 \dots dy_r \right| \leq \\ & \left| \int_{y_1=0}^{\max(f_1)} \dots \int_{y_r=0}^{\max(f_r)} R(\{s : y_1 < f_1(s)\}, \dots, \{s : y_r < f_r(s)\}) - \right. \end{aligned}$$

$$\frac{|R|}{|S|^r} \prod_{i=1}^r |\{s : y_i < f_i(s)\}| dy_1 \dots dy_r \leq \int_{y_1=0}^{\max(f_1)} \dots \int_{y_r=0}^{\max(f_r)} \varepsilon |R| dy_1 \dots dy_r = \varepsilon |R| \prod_{i=1}^r \max(f_i). \quad \square$$

Definition 7. Let \mathbf{A} and \mathbf{B} be relational structures of type τ . We say that \mathbf{C} is a twisted product of \mathbf{A} and \mathbf{B} if the followings hold.

- (1) The base set of \mathbf{C} is the product set: $C = A \times B$.
- (2) The projection $\pi_B : A \times B \rightarrow B$ is a homomorphism $\mathbf{C} \rightarrow \mathbf{B}$.
- (3) For every r -ary relational symbol R of type τ , $1 \leq i \leq r$ and relational tuple $t = (t_1, \dots, t_r) \in R(\mathbf{B})$ there exists a bijection $\alpha_{t,i} : A \rightarrow C$ such that $\pi_B \circ \alpha_{t,i} = t_i$ and $(x_1, \dots, x_r) \in R(\mathbf{A}) \iff (\alpha_{t,1}(x_1), \dots, \alpha_{t,r}(x_r)) \in R(\mathbf{C})$.

If all the bijections in the definition are identical we get the direct product $\mathbf{A} \times \mathbf{B}$. In the case of simple, undirected graphs the last condition means that the preimage of every edge in \mathbf{B} is isomorphic to the direct product of \mathbf{A} and an edge. The celebrated zig-zag product [18] is a very special case (e.g. \mathbf{A} is a complete graph with loops). Two structures may have many different twisted products: we can choose many bijections freely. However, every twisted product of two expanders is an expander.

Lemma 8. *Consider an ε_A -expander \mathbf{A} and an ε_B -expander \mathbf{B} of type τ . If \mathbf{C} is the twisted product of \mathbf{A} and \mathbf{B} then \mathbf{C} is an $(\varepsilon_A + \varepsilon_B)$ -expander. And $\Delta(\mathbf{A})\Delta(\mathbf{B}) \geq \Delta(\mathbf{C})$ holds for the maximal degrees.*

Proof. Let R be an at least binary relation of type τ . We will prove that (2) of Lemma 6 holds for $R(\mathbf{C})$. Consider the functions $f_1, \dots, f_r : C \rightarrow [0; 1]$. Let $g_i : B \rightarrow \mathbb{R}$ denote the function $g_i(b) = \sum_{x \in \pi_B^{-1}(b)} f_i(x)$. Now $|g_i| = |f_i|$, and for every $b \in B$ the inequality $0 \leq g_i(b) \leq |A|$ holds. So the expander property of $R(\mathbf{B})$ implies that

$$\left| R(\mathbf{B})(g_1, \dots, g_r) - |R(\mathbf{B})| \frac{|g_1| \dots |g_r|}{|B|^r} \right| \leq \varepsilon_B |R(\mathbf{B})| |A|^r.$$

Given an r -tuple $\underline{b} = (b_1, \dots, b_r) \in R(\mathbf{B})$ consider the bijections $\alpha_{\underline{b},1}, \dots, \alpha_{\underline{b},r}$ determining the twisted product. Clearly $g_i(b_i) = |f_i|_{\pi_B^{-1}(b_i)} = |f_i \circ \alpha_{\underline{b},i}|$. We sum up all the error terms using $|A||B| = |C|$, $|R(\mathbf{A})||R(\mathbf{B})| = |R(\mathbf{C})|$ and the triangle inequality.

$$\begin{aligned} & \left| R(\mathbf{C})(f_1, \dots, f_r) - |R(\mathbf{C})| \frac{\prod_{i=1}^r |f_i|}{|C|^r} \right| = \\ & \left| R(\mathbf{C})(f_1, \dots, f_r) - |R(\mathbf{C})| \frac{\prod_{i=1}^r |g_i|}{|C|^r} \right| \leq \\ & \left| R(\mathbf{C})(f_1, \dots, f_r) - \frac{|R(\mathbf{A})|}{|A|^r} R(\mathbf{B})(g_1, \dots, g_r) \right| + \end{aligned}$$

$$\begin{aligned}
& \left| \frac{|R(\mathbf{A})|}{|A|^r} R(\mathbf{B})(g_1, \dots, g_r) - |R(\mathbf{C})| \frac{\prod_{i=1}^r |g_i|}{|C|^r} \right| = \\
& \left| \sum_{\underline{b} \in R(\mathbf{B})} \left(R(\mathbf{A})(f_1 \circ \alpha_{\underline{b},1}, \dots, f_r \circ \alpha_{\underline{b},r}) - \frac{|R(\mathbf{A})|}{|A|^r} \prod_{i=1}^r g_i(b_i) \right) \right| + \\
& \frac{|R(\mathbf{A})|}{|A|^r} \left| R(\mathbf{B})(g_1, \dots, g_r) - |R(\mathbf{B})| \frac{\prod_{i=1}^r |g_i|}{|B|^r} \right| \leq \\
& \sum_{\underline{b} \in R(\mathbf{B})} \left(\left| R(\mathbf{A})(f_1 \circ \alpha_{\underline{b},1}, \dots, f_r \circ \alpha_{\underline{b},r}) - \frac{|R(\mathbf{A})|}{|A|^r} \prod_{i=1}^r |f_i \circ \alpha_{\underline{b},i}| \right| \right) + \\
& \frac{|R(\mathbf{A})|}{|A|^r} \varepsilon_B |R(\mathbf{B})| |A|^r \leq \sum_{\underline{b} \in R(\mathbf{B})} \varepsilon_A |R(\mathbf{A})| + \varepsilon_B |R(\mathbf{C})| = (\varepsilon_A + \varepsilon_B) |R(\mathbf{C})|.
\end{aligned}$$

The statement about the maximal degrees follows immediately from the definition. \square

Now we have arrived at the crucial technical theorem of the paper: Two structures with small maximal degree have a twisted product with large girth if the first structure has large girth.

Theorem 9. (*Algorithm*) Consider the finite relational structures \mathbf{A} and \mathbf{B} of type τ . Suppose that the girth of \mathbf{A} is $\geq k$ and $|A|^{1/k} > \Delta(\mathbf{A})\Delta(\mathbf{B})$. Then there exists a twisted product \mathbf{C} of \mathbf{A} and \mathbf{B} with girth $\geq k$. The structure \mathbf{C} can be constructed in polynomial time (in $|\mathbf{A}|$ and $|\mathbf{B}|$).

Proof. We will define better and better twisted products of \mathbf{A} and \mathbf{B} . The number of cycles of minimal length will decrease in every step. We start with the direct product $\mathbf{C}_0 = \mathbf{A} \times \mathbf{B}$. Let \mathbf{C}_i denote the twisted product after Step i . of the algorithm. The bijections determining \mathbf{C}_i are denoted by $\alpha_{v,l}^i : A \rightarrow C$ (where v is an r -ary relational tuple of \mathbf{B} , $1 \leq l \leq r$ and $\pi_B \circ \alpha_{v,l}^i = v_l$).

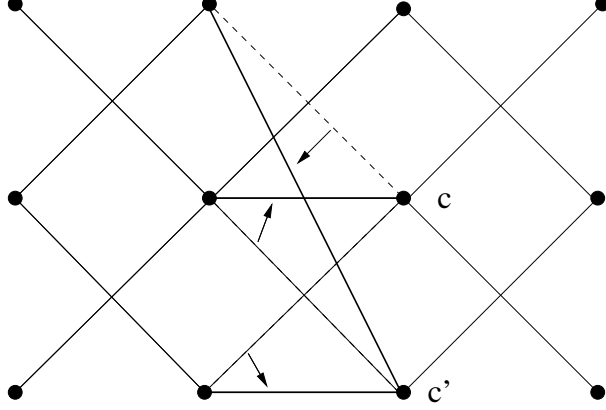
Now we describe Step $(i+1)$. Assume that the girth of \mathbf{C}_i is $j < k$. Let $t \in R(\mathbf{B})$ be an r -ary relational tuple, $1 \leq l \leq r$ and $c, c' \in \mathbf{C}_i$ such that their distance is $\geq k$ and $\pi_B(c) = \pi_B(c') = t_l$. We will specify other conditions on the choice of t, l, c and c' later.

Now we will change the bijection $\alpha_{t,l}^i$ but none of the other bijections defining \mathbf{C}_i . The relations of \mathbf{C}_{i+1} and \mathbf{C}_i will agree but the l^{th} coordinate of the tuples in $\pi_B^{-1}(t)$.

Set $\alpha_{t,l}^{i+1} = (aa') \circ \alpha_{t,l}^i$, where (aa') is the transposition of A flipping a and a' .

The figure illustrates this idea on the product of two undirected paths. The number of cycles of length four decreases. (We neglect the fact that undirected graphs have many degenerate cycles of length two

when considered as relational structures (digraphs). So we actually do not work with the shortest cycles in the figure.)



We call a cycle *short* if its length is j . We will prove that the number of short cycles is strictly less in \mathbf{C}_{i+1} than in \mathbf{C}_i . We correspond to short cycles in \mathbf{C}_{i+1} short cycles in \mathbf{C}_i . Let ξ denote the following bijection between the set of relational tuples of \mathbf{C}_{i+1} and \mathbf{C}_i . For a relational tuple u of \mathbf{C}_{i+1} set $\xi(u) = u$ if $\pi_B(u) \neq t$, else $\xi(u)_h = u_h$ if $h \neq l$ and $\xi(u)_l = (aa') \circ u_l$.

We will show that the image of a short cycle under ξ will be short. Call the tuple u *critical* if $\xi(u) \neq u$. This means that $\pi_B(u) = t$ and the l^{th} coordinate of u is c or c' . The l^{th} coordinate of a critical tuple is called *critical coordinate*. We call a pair of tuples (u_1, u_2) in \mathbf{C}_{i+1} a *cutting pair* if $\pi_B(u_1) \neq \pi_B(u_2)$ and the tuples intersect. Every cycle with length $\leq k$ has a cutting pair: otherwise its image under π_A would be a cycle, too.

Claim: Let t_1, \dots, t_m be a cycle in \mathbf{C}_{i+1} , where $m < \max\{k, 2j\}$ and (t_o, t_{o+1}) is a cutting pair. Assume that c (c') is a coordinate of both t_o and t_{o+1} . Then c (c') can not be a critical coordinate of t_o or t_{o+1} .

The Claim implies that the image of a short cycle under ξ is a (short) cycle: If the image of two intersecting tuples under ξ will not intersect then one should be a critical tuple and its critical coordinate should be in the intersection.

Proof. (of the Claim) We will prove by contradiction. We might suppose that $o = m$, say t_1 is critical, t_m is not and the critical coordinate c is in their intersection. If there is no other adjacent critical-noncritical

pair of tuples s.t. c' is the critical coordinate and it is in their intersection then the path $\xi(t_1), \dots, \xi(t_m)$ connects c and c' , hence $m \geq k$.

If there is such a pair, say (t_h, t_{h+1}) then we distinguish two cases:

If t_{h+1} is critical then $\xi(t_1), \dots, \xi(t_h)$ is a cycle, since $\xi(t_1)$ and $\xi(t_h)$ contains c' and the other tuples remained adjacent under ξ . Similarly, $\xi(t_{h+1}), \dots, \xi(t_m)$ is a cycle, hence $\min\{h, m - h\} \geq j$, so $m \geq 2j$, a contradiction.

In the other case, when t_h is critical the path $\xi(t_1), \dots, \xi(t_h)$ connects c' and c , hence $k \leq h < m$, a contradiction. \square

The main step of the algorithm does not increase the girth. The image of a short cycle under ξ will not be a cycle if it has a cutting pair such that exactly one of the two tuples is critical and the critical coordinate is in the intersection. The image of the other short cycles is still a cycle, and the cutting pairs are the same.

Let us do the main step of the algorithm for all possible triple $c \in \mathbf{C}, l, t$ (and arbitrary c') such that $\pi_B(c) = t_l$. This will hit every cycle of length $< \max\{k, 2j\}$, since the cutting pairs of a cycle do not change. If we iterate this $(\log_2(k) + 1)$ times then we will get the required girth.

The number of such triples is $O(|\mathbf{B}||\mathbf{A}|)$. We need to find c' in every step: this requires $O(|\mathbf{A}||\mathbf{B}|)$ time using Breadth First Search. We can exchange c and c' in the appropriate tuples in the same time. Altogether, the running time of the algorithm is $O(|\mathbf{A}||\mathbf{A}||\mathbf{B}|^2 \log(k))$. \square

4. CONSTRUCTION OF EXPANDERS WITH LARGE GIRTH

We prove Theorem 3 in this section. First we give a probabilistic existential proof in the spirit of [4, 5].

Lemma 10. *Let τ be a finite relational type, k a positive integer and $\varepsilon > 0$. Then there is a $\Delta > 0$ such that for every n large enough there exists a (Δ, ε) -expander of type τ on n vertices with girth $\geq k$.*

Proof. We consider a probability space on the set of relational structures with base set $\{1, \dots, 2n\}$. For every r -ary relational symbol $R \in \tau$ and r -tuple u let $\Pr(u \in R) = \frac{D}{n^{r-1}}$ independently, where the constant D will be chosen later. The expected number of cycles with length $\leq k$ is $O(c^k D^k)$, and the expected degree of a vertex is $O(cD)$, where c is a constant depending only on τ . Set Δ to be ten times the expected value of the degree of a vertex.

The Markov inequality implies that the number of elements covered by the cycles with length $\leq k$ is at most $n/4$ with probability $1 - o_n(1)$,

and the number of elements with degree at least Δ is at most $n/4$ with probability $\geq \frac{3}{5}$. Remove every element with large degree or covered by a short cycle (to get a structure on exactly n elements we may remove more), and consider the resulted structure \mathbf{A} with base set A . With probability $\frac{3}{5} - o_n(1)$ the girth of \mathbf{A} is $\geq k$ and the maximal degree of \mathbf{A} is bounded.

We have to prove the expander property. Consider the r -ary relation R and the subsets $S_1, \dots, S_r \subseteq \{1, \dots, 2n\}$. The probability that $|R(S_1, \dots, S_r) - Dn \frac{\prod_{i=1}^r |S_i|}{n^r}| < \frac{\varepsilon}{3} Dn$ is at most $2e^{-\frac{1}{36}\varepsilon^2 Dn}$ by the Chernoff bound. Since the number of the possible choices is 2^{2nr} this will hold for a D large enough with probability $1 - o_D(1)$ for every r , every r -ary relational symbol $R \in \tau$ and every $S_1, \dots, S_r \subseteq A$. In particular, $||R(\mathbf{A})| - Dn| < \frac{\varepsilon}{3} Dn$. Hence \mathbf{A} is an ε -expander. Altogether, with probability $\frac{3}{5} - o_n(1) - o_1(D)$ the structure \mathbf{A} is a (Δ, ε) -expander with girth $\geq k$. And this probability is positive if n and D are large enough. \square

Lemma 11. *Consider the d -regular undirected graph $G = (V, E)$ with second largest eigenvalue λ and the integer $k \geq 2$. Let \mathbf{S} be the relational structure with base set V and a single k -ary relation R_k :*

$$R_k = \{(a_1, \dots, a_k) : \forall i (a_i, a_{i+1}) \in E\}.$$

Then the relational structure \mathbf{S} is a $(kd^{k-1}, (k-1)\frac{|\lambda|}{d})$ -expander.

Proof. Note that $|R_i| = |V|d^{i-1}$ and the degree of every element is id^{i-1} . Set $\varepsilon = \frac{|\lambda|}{d}$. We prove by induction on k : First suppose that $k = 2$. We will use the expander mixing lemma [6]: for every $T, W \subseteq V$ the inequality $|E(T, W) - d \frac{|T||W|}{|V|}| \leq \lambda \sqrt{|T||W|}$ holds. This implies $|R_2(T, W) - \frac{|T||W|}{|V|^2}| \leq \varepsilon |R_2|$. Hence R_2 is an ε -expander relation.

Assume that we have proved the lemma for $(k-1)$. Consider the functions $x_1, \dots, x_k : S \rightarrow [0; 1]$. By Lemma 6 we need to show that $|R_k(x_1, \dots, x_k) - |R_k| \frac{\prod_{i=1}^k |x_i|}{|V|^k}| \leq \varepsilon(k-1)|R_k|$.

For $i = 1, \dots, k$ define the sequence of functions $y_i : S \rightarrow \mathbb{R}$ recursively. Let y_0 be the constant $\frac{1}{d}$ function and $y_{i+1}(a) = \sum_{(a,b) \in E} x_i(b)$.

Note that $|y_{i+1}| = R_2(y_i, x_{i+1})$. Clearly $0 \leq \max(y_i) \leq d^{i-1}$ and $R_i(x_1, \dots, x_i) = |y_i|$. Now we use the inductual hypothesis:

$$\begin{aligned} |R_k(x_1, \dots, x_k) - |R_k| \frac{\prod_{i=1}^k |x_i|}{|V|^k}| &= |R_2(y_{k-1}, x_k) - |R_k| \frac{\prod_{i=1}^k |x_i|}{|V|^k}| \leq \\ &= |R_2(y_{k-1}, x_k) - |R_2| \frac{|y_{k-1}||x_k|}{|V|^2}| + \left| |R_2| \frac{|y_{k-1}||x_k|}{|V|^2} - |R_k| \frac{\prod_{i=1}^k |x_i|}{|V|^k} \right| \leq \end{aligned}$$

$$\begin{aligned} & \varepsilon |R_2| \max(y_{k-1}) + \frac{d|x_k|}{|V|} \left| |y_{k-1}| - \frac{\prod_{i=1}^{k-1} |x_i|}{|V|^{k-1}} |R_{k-1}| \right| \leq \\ & \varepsilon |R_k| + \frac{d|x_k|}{|V|} (k-2) \varepsilon |R_{k-1}| \leq (k-1) \varepsilon |R_k| \end{aligned}$$

The structure \mathbf{S} is a $(kd^{k-1}, (k-1)\varepsilon)$ -expander. \square

Proof. (of Theorem 3) Assume that every relational symbol in R is at most r -ary. We know that for some d there exists a polynomial time construction of d -regular expander graphs with eigenvalue gap $|\frac{\lambda}{d}| < \frac{\varepsilon}{2r}$, see [14, 11].

On the other hand by Lemma 10 there exists an $\frac{\varepsilon}{2}$ -expander \mathbf{A} with girth at least k such that $|A|^{\frac{1}{k}} > rd^{r-1}\Delta(\mathbf{A})$ holds. If n is large enough then there exists such an \mathbf{A} of size $\log(n)$ by Theorem 10, and so we can find it in polynomial time.

We construct an expander graph G of size $\frac{n}{|A|}$ with the above properties. Lemma 11 shows how to construct an $\varepsilon \frac{r-1}{2r}$ -expander \mathbf{B} on the vertex set of G with maximal degree rd^{r-1} . The conditions of Lemma 9 hold for \mathbf{A} and \mathbf{B} , hence there exists a polynomial time constructible twisted product \mathbf{C} of \mathbf{A} and \mathbf{B} with girth at least k . Now \mathbf{C} is an ε -expander by Lemma 8 with maximal degree at most $M = rd^{r-1}\Delta(A)$ and girth at least k . \square

5. CSP vs MMSNP

Now we prove Theorem 2 showing that CSP and MMSNP are computationally equivalent. Feder and Vardi [5] proved the following (see [9] for a simple proof).

Theorem 12. *Let L be an MMSNP language. Then there is a finite set of relational structures \mathcal{T} and a positive integer k such that*

- (1) *L has a polynomial time reduction to $\text{CSP}(\mathcal{T})$.*
- (2) *$\text{CSP}(\mathcal{T})$ restricted to structures with girth at least k has a polynomial time reduction to L .*

Theorem 1 and Theorem ?? would imply Theorem 2. So we succeed to prove Theorem 1.

Lemma 13. *Consider the structures \mathbf{A}, \mathbf{B} and \mathbf{T} of type τ , where \mathbf{A} is an ε -expander. Suppose that every relational symbol in τ is at most r -ary and $\varepsilon|T|^r < 1$. Let \mathbf{C} be a twisted product of \mathbf{A} and \mathbf{B} . Then \mathbf{B} is homomorphic to \mathbf{T} iff \mathbf{C} is homomorphic to \mathbf{T} .*

Proof. By the definition of the twisted product there is a homomorphism $\pi_B : \mathbf{C} \rightarrow \mathbf{B}$. If \mathbf{B} is homomorphic to \mathbf{T} then so is \mathbf{C} . In order to prove the converse assume that there exists a homomorphism $\varphi : \mathbf{C} \rightarrow \mathbf{T}$. Let us define the mapping $\xi : B \rightarrow T$ in the following way.

For an element $b \in B$ let $\xi(b)$ be one of the elements of T such that $|\pi_B^{-1}(b) \cap \varphi^{-1}(\xi(b))| \geq \frac{|A|}{|T|}$. We will show that ξ is a homomorphism. Let R be an r -ary relational symbol in τ , $b = (b_1, \dots, b_r) \in R(\mathbf{B})$. We need to show that $(\xi(b_1), \dots, \xi(b_r)) \in R(\mathbf{T})$.

Set $S_i = \varphi^{-1}(\xi(b_i)) \cap \pi_B^{-1}(b_i)$. We succeed to show that there is a tuple $(c_1, \dots, c_r) \in R(\mathbf{C})$ with $c_i \in S_i$: In this case the tuple $(\xi(b_1), \dots, \xi(b_r)) = (\varphi(c_1), \dots, \varphi(c_r))$ would be in $R(\mathbf{T})$, since φ is a homomorphism.

Denote the bijections corresponding to b determining the twisted product \mathbf{C} by $\alpha_{b,i} : A \rightarrow \pi_B^{-1}(b_i)$. The tuple (c_1, \dots, c_r) (where $c_i \in \pi_B^{-1}(b_i)$ for every i) is in $R(\mathbf{C})$ iff $(\alpha_{b,1}^{-1}(c_1), \dots, \alpha_{b,r}^{-1}(c_r)) \in R(\mathbf{A})$.

We use the expander property of \mathbf{A} for the sets $\alpha_{b,i}^{-1}(S_i)$ for $1 \leq i \leq r$. Since $R(\mathbf{C})(S_1, \dots, S_r) = R(\mathbf{A})(\alpha_{b,1}^{-1}(S_1), \dots, \alpha_{b,r}^{-1}(S_r))$ we have

$$|R(\mathbf{C})(S_1, \dots, S_r) - |R(\mathbf{A})| \frac{\prod_{i=1}^r |S_i|}{|A|^r}| \leq \varepsilon |R(\mathbf{A})|.$$

On the other hand $|R(\mathbf{A})| \frac{\prod_{i=1}^r |S_i|}{|A|^r} > \varepsilon |R(\mathbf{A})|$ by the choice of the sets S_i and ε . Hence $R(\mathbf{C})(S_1, \dots, S_r) > 0$, there exists an appropriate tuple $(c_1, \dots, c_r) \in R(\mathbf{C})$. This completes the proof of the lemma. \square

Proof. (of Theorem 1) Let us choose r such that every relational symbol in τ is at most r -ary. Consider a $\frac{1}{r+1}$ -expander \mathbf{A} with girth $> k$ and bounded degree. Hence if $|A|$ is large enough then $|A|^{\frac{1}{k}} > \Delta(\mathbf{A})\Delta(\mathbf{S})$ holds. Such an expander \mathbf{A} can be constructed in polynomial time (of $|S|$) for fixed t and k . Now we can use Lemma 9 for $\mathbf{A} = \mathbf{A}$ and $\mathbf{B} = \mathbf{S}$ to construct a twisted product \mathbf{C} of girth at least k . Set $\mathbf{S}' = \mathbf{C}$. Lemma 13 implies Theorem 1. \square

REFERENCES

- [1] M. Ajtai, J. Komlós, E. Szemerédi, *Sorting in $c \log n$ parallel steps*, Combinatorica 3(1), (1983), 1–19.
- [2] N. Alon, Oded Schwartz, Asaf Shapira, *An elementary construction of constant-degree expanders*, 17th ACM-SIAM Symposium on Discrete Algorithms, (2007), 454–458.
- [3] D. Duffus, V. Rödl, B. Sands, N. Sauer, *Chromatic numbers and homomorphisms of large girth hypergraphs*, preprint, (2006).
- [4] P. Erdős, *Graph theory and probability*, Canad. J. Math., **11**, (1959), 34–38.
- [5] T. Feder, M. Y. Vardi: *The computational structure of monotone monadic SNP and constraint satisfaction: A study through Datalog and group theory*, SIAM J. Comput. 28, 1 (1999), 57–104.
- [6] S. Hoory, N. Linial, A. Wigderson, *Expander graphs and their applications*, <http://www.cs.huji.ac.il/~nati/>

- [7] A. Kostochka, J. Nešetřil, P. Smolíkova, *Colorings and homomorphisms of bounded degree and degenerate graphs*. Graph theory (Prague, 1998). Discrete Math. 233 (2001), no. 1-3, 257–276.
- [8] G. Kun: On the complexity of Constraint Satisfaction Problem, PhD thesis (in Hungarian), 2006.
- [9] G. Kun, J. Nešetřil: Forbidden lifts (NP and CSP for combinatorists), KAM-DIMATIA Series 2006-775 (to appear in European J. Comb.).
- [10] L. Lovász, *On chromatic number of finite set-systems*, Acta Math. Acad. Sci. Hungar. **19** 1968, 59–67.
- [11] A. Lubotzky, R. Phillips, P. Sarnak, *Ramanujan graphs*, Combinatorica **8**(3): 261 – 277, 1988.
- [12] A. Lubotzky, B. Samuels, V. Vishne, *Ramanujan complexes of type A_d* , Israel J. of Math., 2005, accepted.
- [13] A. Lubotzky, B. Samuels, V. Vishne, *Explicit constructions of Ramanujan complexes of type A_d* , Europ. J. of Combinatorics, 2005, submitted.
- [14] G. A. Margulis, *Explicit group-theoretical constructions of combinatorial schemes and their application to the design of expanders and concentrators*, J. Probl. Inf. Transm., **24**, No. 1, (1988), 39–46.
- [15] J. Matoušek, J. Nešetřil: Constructions of sparse graphs with given homomorphisms (to appear).
- [16] J. Nešetřil, V. Rödl: A short proof of the existence of highly chromatic hypergraphs without short cycle, J. Comb. Th. B **27** (1979), 225–227.
- [17] J. Nešetřil, M. H. Siggers: A new combinatorial approach to the Constraint Satisfaction Problem dichotomy conjecture, 32nd Symposium on the Mathematical Foundation of Computer Science, 2007, submitted.
- [18] O. Reingold, S. Vadhan, A. Wigderson, *Entropy, waves the zig-zag product, and new constant degree expanders*, Annals of mathematics, 155 (1), (2002), 157–187.

E-mail address: kungabor@cs.elte.hu

DEPARTMENT OF ALGEBRA AND NUMBER THEORY, EÖTVÖS UNIVERSITY,
BUDAPEST, PÁZMÁNY SÉTÁNY 1C, H-1117, HUNGARY