

# Secret sharing on large girth graphs

László Csirmaz, Péter Ligeti <sup>\*†</sup>

## Abstract

We investigate graph based secret sharing schemes and its information ratio, also called complexity, measuring the maximal amount of information the vertices has to store. It was conjectured that in large girth graphs, where the interaction between far away nodes is restricted to a single path, this ratio is bounded. This conjecture was supported by several result, most notably by a result of Csirmaz and Ligeti [7] saying that the complexity of graphs with girth at least six and no neighboring high degree vertices is strictly below 2. In this paper we refute the above conjecture. First, a family of  $d$ -regular graphs is defined iteratively such that the complexity of these graphs is the largest possible  $(d+1)/2$  allowed by Stinson's bound [13]. This part extends earlier results of van Dijk [10] and Blundo et al [3], and uses the so-called entropy method. Second, using combinatorial arguments, we show that this family contains graphs with arbitrary large girth. In particular, we obtain the following purely combinatorial result, which might be interesting on its own: there are  $d$ -regular graphs with arbitrary large girth such that any fractional edge-cover by stars (or by complete multipartite graphs) must cover some vertex  $(d+1)/2$  times.

## 1 Introduction

### 1.1 Motivation and notion

*Secret sharing* is a method for distributing some secret information between a set of participants by giving them partial knowledge of the secret in a way that only pre-described coalitions will be able to reconstruct the original secret from their respective parts. More precisely, let  $\mathcal{P}$  denote the set of *participants*. A family of subsets  $\mathcal{A} \subset 2^{\mathcal{P}}$  is called *access structure* if it is monotone increasing, i.e. if  $A \in \mathcal{A}$  and  $A \subseteq B$  then  $B \in \mathcal{A}$ . The elements of  $\mathcal{A}$  are called *qualified subsets*.

**Definition 1.** A perfect secret sharing  $\mathcal{S}$  realizing the access structure  $\mathcal{A}$  is a collection of random variables  $\xi_i$  for every  $i \in \mathcal{P}$  and  $\xi_s$  with a joint distribution such that

- (i) if  $A \in \mathcal{A}$ , then  $\{\xi_i : i \in A\}$  determines  $\xi_s$ ;
- (ii) if  $A \notin \mathcal{A}$ , then  $\{\xi_i : i \in A\}$  is independent of  $\xi_s$ .

In this paper we focus on the special case when every minimal element of  $\mathcal{A}$  has two elements; these are the so-called *graph-based schemes*. The participants are the vertices of a graph  $G = (V, E)$ , and a set of participants is qualified if there is an edge  $e \in E$  with endpoints in this set.

Traditionally the efficiency of a scheme is measured by the maximal amount of information some participant has to store compared to the size of the secret. This amount is called *information ratio*, or *complexity* of the scheme. The *complexity* of a graph  $G$  is the best lower bound on the complexity of schemes realizing  $G$ , and can be defined formally in following way:

<sup>\*</sup>Central European University, Eötvös Loránd University and Rényi Institute, Budapest

<sup>†</sup>e-mail: csirmaz@renyi.hu, turul@cs.elte.hu

**Definition 2.** The complexity (or information ratio) of the graph  $G = (V, E)$  is

$$c(G) = \inf_{\mathcal{S}} \max_{v \in V} \frac{H(\xi_v)}{H(\xi_s)},$$

where the infimum is taken over all perfect secret sharing schemes  $\mathcal{S}$  realizing the access structure defined by the graph  $G$ , and  $H(\xi)$  is the Shannon entropy of the random variable  $\xi$ .

One of the most interesting and challenging problems in this topic is to determine the information ratio exactly for particular graphs or families of graphs [1]. The main tool is to prove general or specific estimations for the information ratio and find constructions which realize these bounds.

## 1.2 Related works

For a comprehensible account on secret sharing, on its relevance, please consult the survey [1]. In terms of the number of vertices, the best lower bound for  $c(G)$  established so far is logarithmic in the number of vertices [3, 6, 10]; while it is known that  $c(G)$  is always  $\leq c \cdot n / \log n$  for some small constant  $c$  [8, 12].

The only known method for proving lower bounds is the entropy technique, we explain it in more detail in Section 2. On the other hand, every graph based secret sharing scheme yields an upper bound. One fundamental tool in constructing such schemes is Stinson's decomposition technique. While it is more general, we will use the following special case only.

**Theorem 1** (Stinson's Decomposition Theorem, [13]). *Suppose that the edges of the graph  $G$  can be fractionally covered by the edges of complete multipartite graphs such that every vertex is covered with weight at most  $k$ . Then  $c(G) \leq k$ .*  $\square$

Using Stinson's decomposition theorem for all star (spanned) subgraphs of a graph  $G$  with weight 0.5, it follows immediately that the complexity of  $G$  is at most  $(d + 1)/2$  where  $d$  is the maximal degree. This upper bound is known as the *Stinson's bound*.

To attack the problem whether Stinson's bound is tight, in [10] van Dijk defined a family of  $d$ -regular graphs and proved that Stinson's bound is asymptotically tight for that family. Later Blundo et al [3] showed that the complexity of this family actually matches Stinson's bound. In Section 4 we generalize their result by showing that Stinson's bound is tight for an even larger family of  $d$ -regular graphs.

The graphs in van Dijk's family have girth 6 and ratio  $(d + 1)/2$ . The complexity of other (sporadic) infinite graph families has been determined as well. Some examples are

- the edge graph of the  $d$ -dimensional hypercube which has complexity  $d/2$  and girth 4 [6];
- trees have complexity  $< 2$  and girth 0, for exact values consult [9];
- graphs with girth  $> 5$  and no adjacent vertices of degree at least three have complexity strictly below 2 [7].

These results supported the intuition that high complexity requires high connectivity and bounded girth, as the interaction between nodes diminishes exponentially as their distance grows. We show that this intuition was wrong by constructing graphs with arbitrary large complexity *and* arbitrary large girth at the same time. The result is achieved in two steps. First, a family  $\mathcal{G}_d$  of  $d$ -regular graphs is defined which extends the above mentioned van Dijk's graph family. Using the entropy method, explained in Section 2, we show that all graphs in  $\mathcal{G}_d$  have the maximal complexity allowed by Stinson's bound, namely  $(d + 1)/2$ . Second, in Section 5, using purely combinatorial arguments, we show that  $\mathcal{G}_d$  contains graphs with arbitrary large girth.

Any fractional cover of the edges by spanned stars automatically gives a secret sharing scheme with complexity equal to the maximal cover weight on the vertices. Consequently any fractional edge-cover by stars of a graph from  $\mathcal{G}_d$  covers some vertex at least  $(d+1)/2$  times, even the ones which have large girth. This is in sharp contrast to the easy fact that a tree has a star-cover where each vertex is covered at most twice.

### 1.3 Organization

The *entropy method* is a universal, but not complete, tool to establish lower bounds on the information ratio of arbitrary access structures, see [1, 2, 3, 5, 6, 10]. For the convenience of the interested reader, the method is explained in Section 2. Specific lemmas tailored for establishing lower bounds on graph-based access structures are stated and proved in Section 3. Section 4 contains the definition of the family  $\mathcal{G}_d$  of  $d$ -regular graphs along with the proof that they have complexity  $(d+1)/2$ . This graph family extends that of van Dijk from [10] and [3]. The existence of large girth graphs in  $\mathcal{G}_d$  is proved in Section 5. The construction and the results in this section are purely combinatorial, and may be interesting independently. Finally, Section 6 concludes the paper and lists some open problems.

## 2 The entropy method

We focus on the special case of graph based structures. The method can be summarized as follows. Consider any function  $f$  assigning non-negative real numbers to subsets of the vertices of  $G$  (the normalized entropy function) with the following properties:

1.  $f$  is monotone and submodular; moreover  $f(\emptyset) = 0$ ;
2. if  $A \subset B$ ,  $A$  is independent and  $B$  is not, then  $f(A) + 1 \leq f(B)$  (strict monotonicity)
3. if  $C$  is empty or independent,  $AC$  and  $BC$  are not independent (qualified), then  $f(AC) + f(AB) \geq f(C) + f(ABC) + 1$  (strict submodularity).

If for any such function  $f$  we have  $f(v) \geq \alpha$  for some vertex  $v$  of  $G$ , then the complexity of  $G$  is at least  $\alpha$ .

As in the formula above, throughout the paper we drop the  $\cup$  sign when denoting a union of subsets, and make no distinction between a vertex and the one-element subset containing that vertex. In particular,  $aAB$  denotes the subset  $\{a\} \cup A \cup B$ . Lower case letters  $a, b$ , etc will denote vertices of  $G$ , and capital letters  $A, B$ , etc denote subsets of vertices.

The correctness of the method follows from the observation that any secret sharing scheme is a collection of random variables, and  $f(A)$  can be chosen to be the total entropy of the shares given to the vertices in  $A$  divided by the entropy of the secret. Thus the relative size of participant  $v$ 's share is just  $f(v)$ . The above properties of  $f$  are the translations of basic Shannon inequalities and the fact that the distributed secret is independent from any independent set, and is determined by any non-independent subset of  $G$ . Consequently every secret sharing scheme assigns a share of relative size  $\alpha$  to some participant.

Following the entropy notation, we will use the following abbreviations:

$$\begin{aligned} I_f(A; B) &\stackrel{\text{def}}{=} f(A) + f(B) - f(AB); \\ I_f(A; B | C) &\stackrel{\text{def}}{=} f(AC) + f(BC) - f(C) - f(ABC). \end{aligned}$$

The submodularity property gives that both expressions are non-negative, moreover strict submodularity is equivalent to  $\mathbf{I}_f(A; B | C) \geq 1$  whenever  $C$  is either unqualified or empty, and  $AC$  and  $BC$  are qualified subsets.

In the formulas we frequently omit the function  $f$ , and any vertex, or subset of vertices stand for its  $f$  value as well. In particular, we use  $\mathbf{I}(A; B)$  and  $\mathbf{I}(A; B | C)$  instead of  $\mathbf{I}_f(A; B)$  and  $\mathbf{I}_f(A; B | C)$  whenever  $f$  is clear from the context.

### 3 Some general lemmas

We start by stating and proving some general lemmas which resemble to information theoretic arguments. Fix a function  $f$  satisfying properties 1–3 defined in Section 2.

**Lemma 2.** *If  $C$  is independent,  $AC$  and  $BC$  are not, then  $\mathbf{I}(A; B | C) \geq 1$ .*

*Proof.* This is just the restatement of the strong submodularity requirement.  $\square$

Let  $A$  and  $B$  be disjoint subsets of the vertices. We say that there is a *1-factor from  $B$  to  $A$*  if  $B = \{b_1, \dots, b_t\}$ , and there are  $t$  different elements  $a_1, \dots, a_t$  in  $A$  such that  $a_i b_j$  is an edge if and only if  $i = j$ .

**Lemma 3.** *Let  $A$  and  $B$  be disjoint subsets of the vertices,  $B$  is independent,  $A$  is not independent such that there is a 1-factor from  $B$  to  $A$ . Then  $\mathbf{I}(A; B) \geq |B|$ .*

*Proof.* The proof goes by induction on the number of elements in  $B$ . When  $B$  is empty, then  $|B| = 0$  and there is nothing to prove. Otherwise let  $ab$  be an edge in the 1-factor, and let the two sets be  $A = aA^*$  and  $B = bB^*$ . Then  $aB^*$  is independent,  $abB^*$  and  $aA^*B^*$  are not, thus we get  $\mathbf{I}(b; A^* | aB^*) \geq 1$  by Lemma 2. Then

$$\begin{aligned} \mathbf{I}(A; bB^*) - \mathbf{I}(A; B^*) &= (f(A) + f(bB^*) - f(bAB^*)) - (f(A) + f(B^*) - f(AB^*)) \\ &= f(bB^*) - f(B^*) + f(aA^*B^*) - f(abA^*B^*) \\ &= \mathbf{I}(b; A^* | aB^*) + \mathbf{I}(a; b | B^*) \geq 1 + 0. \end{aligned}$$

From here the induction hypothesis gives the claim.  $\square$

**Lemma 4.** *With the same assumptions as in Lemma 3,  $f(A) \geq |B| + 1$ .*

*Proof.* According to Lemma 3,  $\mathbf{I}(A; B) = f(A) + f(B) - f(AB) \geq |B|$ . Thus

$$f(A) \geq |B| + (f(AB) - f(B)) \geq |B| + 1$$

by strict monotonicity as  $B$  is independent and  $AB$  is not.  $\square$

**Lemma 5.** *Let  $A$  and  $B$  be disjoint subsets of the vertices such that neither  $A$  nor  $B$  is independent. Suppose  $B$  contains an independent subset  $B'$  and a 1-factor from  $B'$  to  $A$ . Then  $\mathbf{I}(A; B) \geq |B'| + 1$ .*

*Proof.* Let  $B = B'B''$  where  $B'$  is the independent set with the 1-factor. Then

$$\mathbf{I}(A; B'B'') = \mathbf{I}(A; B') + \mathbf{I}(A; B'' | B') \geq |B'| + 1$$

by Lemmas 3 and 2.  $\square$

## 4 The graph family $\mathcal{G}_d$

Let  $n_2, n_3, \dots$ , be integers, each one is at least 5, and  $n_2$  is even. We construct a sequence of bipartite graphs  $G_2, G_3, \dots$  as follows.  $G_2$  is the cycle with  $n_2$  nodes;  $A_2, B_2$  are the two independent sets consisting of the odd and even vertices, respectively.

Suppose  $G_d$  has been constructed; the equal size partition  $(A_d, B_d)$  of its vertices shows that  $G_d$  is bipartite;  $A_d$  and  $B_d$  are independent, and all edges of  $G_d$  go between  $A_d$  and  $B_d$ . (This property holds for  $G_d$  by induction.) Take  $n_{d+1}$  copies of  $G_d$  denoted as  $G_d^i$  with  $G_d^{n_{d+1}+1} = G_d^1$ . The vertex set of  $G_d^i$  is  $A_d^i \cup B_d^i$  where  $A_d^i$  and  $B_d^i$  are the (equal size) independent subsets of the vertices of  $G_d^i$ . To get  $G_{d+1}$  add an (arbitrary) 1-factor between  $B_d^i$  and  $A_d^{i+1}$  for all  $i = 1, 2, \dots, n_{d+1}$ , see

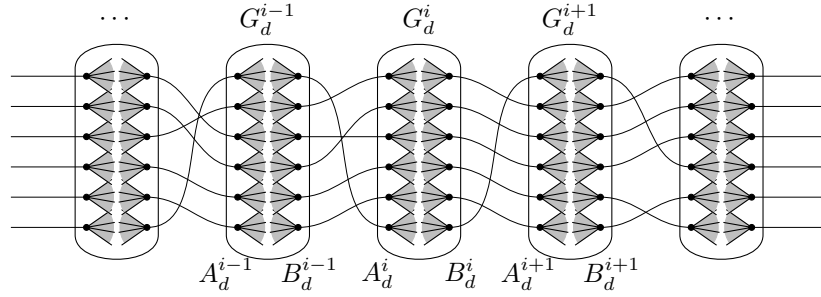


Figure 1: Structure of the graph  $G_{d+1}$

Figure 1. The equal size partition of  $G_{d+1}$  showing that  $G_{d+1}$  is bipartite is the union of vertices in  $A_d^i$  and the union of vertices in  $B_d^i$ , respectively. The graph family  $\mathcal{G}_d$  consists of all graphs  $G_d$  constructed this way.

**Claim 6.**  $G_d$  is a  $d$ -regular bipartite graph on  $n_2 n_3 \dots n_d$  vertices.

*Proof.* Immediate from the definition. □

According to Stinson's theorem [13], the information ratio of a  $d$ -regular graph is at most  $(d+1)/2$ . The next theorem claims that this bound is tight for graphs in  $\mathcal{G}_d$ . As van Dijk's graph family is contained properly in  $\mathcal{G}_d$ , this theorem extends the main result of [3].

**Theorem 7.** For any normalized entropy function  $f$  on  $G_d \in \mathcal{G}_d$  the following inequality holds:

$$\sum_{v \in G_d} f(v) \geq \frac{d+1}{2} |G_d|. \quad (1)$$

Consequently the information ratio of  $G_d$  is exactly  $(d+1)/2$ .

The crux of the proof is the following inequality which will be proved by induction along the construction of the graph  $G_d$ .

**Lemma 8.** For any normalized entropy function  $f$  on  $G_d$  the following inequality holds:

$$\sum_{v \in G_d} f(v) - f(G_d) \geq \frac{d}{2} |G_d| - 1. \quad (2)$$

Let us first see how to derive Theorem 7 from this lemma.

*Proof of Theorem 7 assuming Lemma 8.* The graph  $G_{d+1}$  consists of  $n_{d+1}$  copies of  $G_d \in \mathcal{G}_d$ ; each copy is connected by a 1-factor to the rest of  $G_{d+1}$ : half of the edges go to the previous copy, the other half of the edges go to the next copy; and the other endpoints of this 1-factor form an independent set of size  $|G_d|$ , see Figure 1. Thus, by Lemma 4,

$$f(G_d^i) \geq |G_d| + 1.$$

Applying Lemma 8 to each copy  $G_d^i$  separately, we get

$$\begin{aligned} \sum_{v \in G_{d+1}} f(v) &= \sum_{i=1}^{n_{d+1}} \sum_{v \in G_d^i} f(v) \geq \sum_{i=1}^{n_{d+1}} \left( f(G_d^i) + \frac{d}{2} |G_d| - 1 \right) \\ &\geq \sum_{i=1}^{n_{d+1}} \left( |G_d| + |G_d| \frac{d}{2} \right) = \frac{d+2}{2} |G_{d+1}|, \end{aligned}$$

which proves (1) when  $d$  is at least 3. For the case  $d = 2$  let  $a, b, c, d$  be four neighbor vertices on the cycle  $G_2$ . Now  $f(bc) \geq 3$  by Lemma 4, as  $ad$  is independent,  $bc$  is not, and there is a 1-factor from  $ad$  to  $bc$ . Consequently

$$\sum_{v \in G_2} f(v) \geq f(v_1 v_2) + f(v_3 v_4) + \dots \geq \frac{|G_2|}{2} \cdot 3.$$

Here we used  $f(b) + f(c) \geq f(bc)$ , which is a special case of submodularity.  $\square$

Now we turn to the proof of Lemma 8. The validity of the following decomposition of an expression similar to the left hand side of (2) follows immediately from the definitions.

**Fact 9.** Suppose  $n \geq 5$  and  $E_i$  are subsets of the vertices for  $i = 1, \dots, n$ . Then  $\sum_i f(E_i) - f(E_1 \dots E_n)$  can be written as the following sum of  $3 + (n - 4)$  entropy terms:

$$\begin{aligned} &\mathbf{I}(E_1; E_2) + \mathbf{I}(E_1 E_2; E_3) + \mathbf{I}(E_1 E_2 E_3; E_4 E_5 \dots E_n) + \\ &+ \mathbf{I}(E_4; E_5) + \mathbf{I}(E_4 E_5; E_6) + \dots + \mathbf{I}(E_4 E_5 \dots E_{n-1}; E_n). \end{aligned}$$

*Proof of Lemma 8.* First we prove (2) for the base case  $d = 2$ . The graph  $G_2$  is the cycle on  $n = |G_2| \geq 6$  vertices. Let us denote the vertices by  $v_1, \dots, v_n$ . The edges of  $G_2$  are  $v_i v_{i+1}$  and  $v_n v_1$ . For this graph inequality (2) rewrites to

$$\sum_{i=1}^n f(v_i) - f(G_2) \geq n - 1.$$

To prove it we use the decomposition in Fact 9 with  $E_i = \{v_i\}$  to rewrite the left hand side as the  $3 + (n - 4)$ -term sum

$$\begin{aligned} &\mathbf{I}(v_1; v_2) + \mathbf{I}(v_1 v_2; v_3) + \mathbf{I}(v_1 v_2 v_3; v_4 v_5 \dots v_n) + \\ &+ \mathbf{I}(v_4; v_5) + \mathbf{I}(v_4 v_5; v_6) + \dots + \mathbf{I}(v_4 v_5 \dots v_{n-1}; v_n). \end{aligned}$$

All the terms here are non-negative. Furthermore, by Lemma 3 we have

$$\begin{aligned} \mathbf{I}(v_1 v_2; v_3) &\geq 1, \\ \mathbf{I}(v_4 v_5; v_6) &\geq 1, \\ &\dots \\ \mathbf{I}(v_4 v_5 \dots v_{n-1}; v_n) &\geq 1; \end{aligned}$$

and by Lemma 5,  $I(v_1v_2v_3; v_4 \dots v_n) \geq 3$  witnessed by the independent set  $v_1v_3$ . These numbers add up to  $n - 1$  proving the base case.

Next suppose we know the inequality (2) for the graph  $G_d$ , and want to prove it for  $G_{d+1}$ . The graph  $G_{d+1}$  consists of  $n_{d+1} = n \geq 5$  copies of  $G_d$ . The vertex set of the  $i$ -th copy is  $V_i = A_i \cup B_i$ , where  $A_i$  and  $B_i$  are disjoint independent sets of equal size. The induction hypothesis tells us that

$$\sum_{v \in V_i} f(v) - f(V_i) \geq \frac{d}{2} |G_d| - 1$$

for every  $i$ . Since  $|G_{d+1}| = n \cdot |G_d|$ , we are done if we prove that

$$\begin{aligned} \sum_{i=1}^n f(V_i) - f(G_{d+1}) &\geq \left( \frac{d+1}{2} |G_{d+1}| - 1 \right) - n \left( \frac{d}{2} |G_d| - 1 \right) \\ &= \frac{n}{2} |G_d| + n - 1. \end{aligned} \tag{3}$$

Apply the decomposition using  $E_i = V_i$  to get

$$\begin{aligned} &I(V_1; V_2) + I(V_1V_2; V_3) + I(V_1V_2V_3; V_4 \dots V_n) + \\ &+ I(V_4; V_5) + I(V_4V_5; V_6) + \dots + I(V_4 \dots V_{n-1}; V_n) \end{aligned}$$

as an equivalent form of the left hand side here. By Lemma 5 each of these quantities, except for the third one, have value at least  $1 + |G_d|/2$ , as  $A_i \subset V_i$  is an independent set of size  $|G_d|/2$  connected to the other part by a 1-factor. As  $V_1V_2V_3$  has an independent set of size  $|G_d|$  (the vertex set  $A_1B_3$ ) connected by a 1-factor to  $V_4 \dots V_n$ , we have

$$I(V_1V_2V_3; V_4 \dots V_n) \geq |G_d| + 1.$$

The sum of these values is  $n|G_d|/2 + n - 1$ , as was required.  $\square$

## 5 Graphs in $\mathcal{G}_d$ with large girth

Graphs in  $\mathcal{G}_2$  are even length cycles, thus the girth is equal to the number of vertices – which can be arbitrary large. The first challenge is to find large girth graphs in  $\mathcal{G}_3$  where one can choose the 1-factors between the neighboring independent sets arbitrarily, see Figure 1. However, it is not clear how to control the interaction of those choices in order to avoid introducing short cycles. A natural approach would be choosing these 1-factors randomly. With too many 1-factors the graph will have constant girth with overwhelming probability, but maybe Lovasz Local Lemma [11] can be used to show that the girth is  $> g$  with non-zero (while exponentially small) probability. Unfortunately this approach failed as well the attempts to use algebraic construction.

Our method which finds large girth graphs in  $\mathcal{G}_{d+1}$  is based on the following idea. We choose all 1-factors between the different copies of the  $G_d$  graph identically. Then map  $G_{d+1}$  to  $G_d$  so that the image of the vertex  $v^i$  in the  $i$ -th copy of  $G_d$  is just  $v$ . Then the image of  $G_{d+1}$  is  $G_d^*$  which has the same edges as  $G_d$  plus a (maximal) 1-factor between the two independent sets  $A_d$  and  $B_d$  of  $G_d$ . If we know that  $G_d^*$  has no short cycles, then we can conclude that neither does  $G_{d+1}$ . Lemma 10 shows the existence of such a  $G_2^*$  graph on every large enough vertex number. To ease the description first we introduce the notion of  $\pi$ -graphs.

Let  $\pi$  be a permutation on  $1, \dots, n$ . A  $\pi$ -graph is a 3-regular bipartite graph on vertices  $a_i, b_i$  for  $1 \leq i \leq n$  such that  $a_i$  is connected to  $b_i$ , to  $b_{i+1}$  and to  $b_{\pi(i)}$ , where  $b_{n+1} = b_1$ .

A  $\pi$ -graph is bipartite, the two independent vertex classes are  $\{a_i\}$  and  $\{b_i\}$ . The edges  $a_i - b_{\pi(i)}$  form a (maximal) 1-factor.

**Lemma 10.** *There is a function  $N(g)$  such that for  $g > 3$  and  $n > N(g)$  there is a permutation  $\pi$  and a  $\pi$ -graph on  $2n$  vertices which has girth  $> g$ .*

*Proof.* The distance of two vertices is the length of the shortest path between them. Thus  $G$  has girth  $> g$  if for any edge  $uv$  of  $G$ , after deleting that edge the distance of  $u$  and  $v$  is at least  $g$ .

We construct the claimed  $\pi$ -graph by adding a 1-factor to a large cycle on  $2n$  vertices. The initial graph has vertices  $a_i, b_i$  for  $i = 1, \dots, n$  and edges  $a_i - b_i$  and  $a_i - b_{i+1}$ , where indices are understood modulo  $n$ . A new edge can be added between the 2-degree vertices  $a_i$  and  $b_j$  if their distance is at least  $g$ . For a 2-degree vertex  $v$  the number of vertices with distance  $< g$  from  $v$  is at most

$$1 + 2 + 4 + \dots + 2^{g-1} < 2^g.$$

Consequently one can add the next edge in the 1-factor in a greedy way until the number of (unmatched) 2-degree vertices among the  $a_i$  vertices goes below  $2^g$ .

At this point the graph has exactly  $(2^g - 1) + (2^g - 1)$  2-degree vertices, and girth  $> g$ . Next find (circular) intervals  $I_s$  of the indices  $1, \dots, n$  for  $1 \leq s < 2^g$  with the following properties:

1.  $I_s$  contains  $2^g + 1$  indices;
2. if  $i \in I_s$ , then both  $a_i$  and  $b_i$  are at distance  $\geq g$  from any 2-degree vertex;
3. if  $i \in I_s$  and  $j \in I_t$ ,  $s \neq t$ , then  $a_i$  and  $b_j$  has distance  $\geq g$ .

After picking  $s$  intervals, the number of indices which cannot be the midpoint of the next interval  $I_{s+1}$  is at most

$$2(2^g + s \cdot (2^g + 1))2^g(2^g + 1).$$

Consequently, if  $n > 2^{4g+3}$  then one can find such intervals. From each interval  $I_s$  pick two indices  $u_s < \ell_s$  such that the distance between the vertices  $b_{u_s}$  and  $a_{\ell_s}$  is at least  $g$ . As the interval has length  $2^g + 1$ , such a pair always exists. By construction, all distances between the vertices in the set

$$\{a_{\ell_s}, b_{u_s} : 1 \leq s < 2^g\}$$

are at least  $g$ , moreover any of them is at distance  $\geq g$  from any 2-degree vertex.

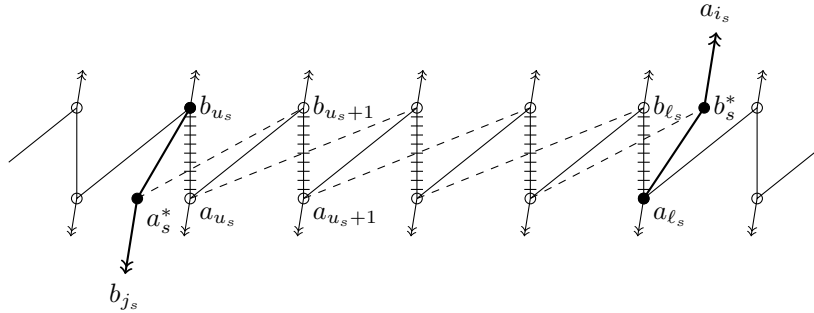


Figure 2: Swapping edges between  $a_s^*$  and  $b_s^*$  in  $G$

Let the 2-degree vertices be  $a_{i_s}$  and  $b_{j_s}$  where  $s$  runs from 1 to  $2^g - 1$ . Add  $2(2^g - 1)$  new vertices and some new edges to the graph. The new vertices are  $a_s^*$  and  $b_s^*$ , and the new edges are  $b_{j_s} - a_s^*$ ,  $a_s^* - b_{u_s}$ , and  $a_{i_s} - b_s^*$ ,  $b_s^* - a_{\ell_s}$ , see Figure 2. In this new graph  $G$  the new vertices  $a_s^*$  and  $b_s^*$  have degree 2; vertices  $b_{u_s}$  and  $a_{\ell_s}$  have degree 4, all other vertices have degree 3.  $G$  still has girth  $> g$ . Indeed, any cycle without the new vertices has length  $> g$ . As new vertices have degree 2, if a cycle contains  $a_s^*$ , then it also contains  $b_{j_s}$  and  $b_{u_s}$ ; similarly for  $b_s^*$ . A short cycle cannot contain a single new vertex  $a_s^*$  (or  $b_s^*$ ) only, as  $b_{j_s}$  and  $b_{u_s}$  are far from each other. So the



short cycle contains (at least) two new vertices connected by a short path, and this short path can only be  $a_s^* - b_{j_s} - \dots - a_{i_t} - b_t^*$ , or  $a_s^* - b_{j_s} - \dots - b_{j_t} - a_t^*$ , or  $b_s^* - a_{i_s} - \dots - a_{i_t} - b_t^*$ . Any two of the endpoints of these paths are far from each other, thus one cannot merge them into a short cycle.

Now  $G$  is bipartite, has girth  $> g$ , but it is not 3-regular. To make it 3-regular, we make the following changes. Delete the edges between  $a_i$  and  $b_i$  for  $u_s \leq i \leq \ell_s$  (after this  $a_{\ell_s}$  and  $b_{u_s}$  will have degree 3), and add the edges  $a_s^* - b_{u_s+1}$ ,  $a_{\ell_s-1} - b_s^*$ , and the edges  $a_{i-1} - b_{i+1}$  for  $u_s < i < \ell_s$  as depicted on Figure 2. Denote this new graph by  $G^*$ .

If  $G^*$  has a cycle, then the newly added (dashed) edges in the cycle can be replaced by three edges of  $G$ . This way the cycle cannot collapse, and after deleting edges traversed back and forth, the cycle in  $G$  has at least as many edges as it had in  $G^*$ , but not more than three times as much. Consequently, the girth of  $G^*$  is bigger than  $g/3$ .

Summing up, for any  $n > 2^{4g+3}$  we have constructed a  $\pi$ -graph on  $2n + 2(2^g - 1)$  vertices which has girth bigger than  $g/3$ . Thus we have proved the lemma with the function  $N(g) = 2^{12g+4}$ .  $\square$

Before continuing, let us show how Lemma 10 can be used to find a large girth graph in the family  $\mathcal{G}_3$ . Fix a  $\pi$ -graph  $G$  on  $2n$  vertices which has girth  $> g$  as given by Lemma 10. Let the vertices of  $G$  be labeled as  $a_i, b_i$  such that  $b_1, a_1, b_2, a_2, \dots, b_n, a_n$  is a cycle, and the additional 1-factor is given by the edges  $a_i - b_{\pi(i)}$ .

Let  $m \geq 5$ . The vertices of  $H$  are  $a_{i,j}$  and  $b_{i,j}$  where  $1 \leq i \leq n$  and  $1 \leq j \leq m$ , the indices are understood modulo  $n$  and  $m$ , respectively.  $a_{i,j}$  is connected to  $b_{i,j}$ ,  $b_{i+1,j}$  and  $b_{\pi(i),j+1}$ . For fixed  $j$ , the vertices  $a_{i,j}$  and  $b_{i,j}$  form a cycle on  $2n$  vertices – thus an instance of  $\mathcal{G}_2$ ; and the edges  $a_{i,j} - b_{\pi(i),j+1}$  form a 1-factor between the independent sets  $A_j$  and  $B_{j+1}$ . Consequently  $H$  is a  $\mathcal{G}_3$ -graph. The map  $\varphi(a_{i,j}) = a_i$ ,  $\varphi(b_{i,j}) = b_j$  is a graph homomorphism from  $H$  to  $G$ , which maps any cycle in  $H$  into a cycle in  $G$ . (As  $\varphi$  is onto, and both  $H$  and  $G$  are 3-regular, in the image no edge is traversed immediately backward.) As  $G$  has girth  $> g$ ,  $H$  has girth  $> g$  as well, as required.

To formalize the above construction, we introduce the following notation. Let  $G$  be a graph on even number of vertices,  $A$  and  $B$  be the equal size partition of the vertices, and  $\pi : A \rightarrow B$  be a one-to-one mapping. For any  $m > 1$  the graph  $\mathcal{H}(m, G, \pi)$  consists of  $m$  disjoint copies of  $G$  labeled as  $G^1, G^2, \dots, G^m$  with  $G^{m+1} = G^1$ . There is an additional 1-factor between  $A_j$  of  $G^j$  and  $B_{j+1}$  of  $G^{j+1}$  determined by the mapping  $\pi$ : the vertex  $a_j \in A_j$  is connected to  $\pi(a)_{j+1} \in B_{j+1}$ . The above graph  $H$  is just  $\mathcal{H}(m, C_{2n}, \pi)$ , where  $\pi$  is the map from Lemma 10.

For handling larger degree  $d$  the following stronger claim will be proved by induction of  $d$ .

**Lemma 11.** *For each  $g > 3$  there exists a  $d$ -regular graph  $G_d \in \mathcal{G}_d$  with independent vertex sets  $A_d, B_d$ , and a one-to-one map  $\pi_d : A_d \rightarrow B_d$  such that for all  $m > 1$  the graph  $\mathcal{H}(m, G_d, \pi_d)$  has girth  $> g$ .*

*Proof.* For  $d = 2$  the graph  $G_2$  is the cycle on  $N_2 = 2n \approx 2 \cdot 2^{12g+4}$  vertices, and  $\pi_2$  is the map given by Lemma 10. It satisfies the claim of the lemma as discussed above.

Suppose we have the graph  $G_d$  on even number of vertices and the map  $\pi_d$  as in the Lemma, and want to find  $G_{d+1}$  and  $\pi_{d+1}$ . Let  $t = 3|G_d|$ , and apply Lemma 10 to the girth  $gt$  to get an  $N_{d+1} \geq 5$  which is a multiple of  $|G_d|$ , and the map  $\pi$  on  $1, \dots, N_{d+1}$ , which is a one-to-one map between the odd and even indices. Let  $m_d = N_{d+1}/|G_d|$ . This is  $\geq 5$ , thus  $G_{d+1} = \mathcal{H}(m_d, G_d, \pi_d)$  is in  $\mathcal{G}_{d+1}$ , and has  $N_{d+1}$  vertices. These vertices can be labeled as  $v_i$  for  $1 \leq i \leq N_{d+1}$  such that

- a) the two independent sets of  $G_{d+1}$  are the odd-indexed vertices and the even-indexed vertices;
- b) if  $v_i - v_j$  is an edge in  $G_{d+1}$ , then the circular distance between  $i$  and  $j$  (that is, the smaller of  $|i - j|$  and  $N_{d+1} - |i - j|$ ) is at most  $t$ .

Indeed, as  $t = 3|G_d|$ , enumerate the vertices in the first copy of  $G_d$  first, then the vertices in the second copy, and so on, making sure that the independent sets get the even and odd indices, respectively. Finally let  $\pi_{d+1}$  be the map between the odd and even numbers between 1 and  $N_{d+1}$  as induced by the map  $\pi$  of Lemma 10.

We claim that  $G_{d+1}$  and  $\pi_{d+1}$  satisfies the conditions of the lemma. The only non-trivial case is that  $H = \mathcal{H}(m, G_{d+1}, \pi_{d+1})$  has girth  $> g$  for every  $m > 1$ . Consider a (short) cycle in  $H$ . If this cycle has vertices from the same copy of  $G_{d+1}$  only, then it has length  $> g$  as  $G_{d+1}$  has girth  $> g$  by induction.

Consequently the cycle must have vertices in different copies of  $G_{d+1}$ . Let  $v_{i_1}^j$  and  $v_{i_2}^j$  be neighbor points in the cycle in the  $j$ -th copy, where  $1 \leq i_1, i_2 \leq N_{d+1}$ . By construction, the circular distance of  $i_1$  and  $i_2$  is at most  $t$ , thus the  $v_{i_1}^j - v_{i_2}^j$  edge can be replaced by at most  $t$  edges now in the cycle  $v_1, v_2, \dots, v_{N_{d+1}-1}, v_{N_{d+1}}$ . But it means that the graph returned by Lemma 10 has a cycle of length  $\leq gt$ , which is a contradiction.  $\square$

The main theorem of this section is a simple corollary of Lemma 11.

**Theorem 12.** *There are arbitrary large girth graphs in  $\mathcal{G}_d$ .*

*Proof.* As  $H = \mathcal{H}(m, G_d, \pi_d)$  has girth  $> g$ , and  $G_d$  is a spanned subgraph of  $H$ ,  $G_d$  must have girth  $> g$  as well.  $\square$

## 6 Conclusion

Using the entropy method, it was shown that the general upper bound  $(d+1)/2$  on the complexity of graph based secret sharing schemes, known as Stinson's bound, is tight for a large class of inductively defined  $d$ -regular bipartite graphs. The class  $\mathcal{G}_d$  contains the graphs defined by van Dijk [10] and proved to be tight by Blundo et al [3].

In Section 5 it was proved that each  $\mathcal{G}_d$  contains graphs of arbitrary large girth using combinatorial arguments. This result refutes the widely believed conjecture that large girth graphs have bounded complexity – due to the exponentially diminishing interaction between the shares assigned to the vertices.

The construction of the graph family  $\mathcal{G}_d$  in Section 4 is not the most general one. When connecting instances of  $G_d$  by 1-factors as indicated on Figure 1, the instances  $G_d^i$  need not be isomorphic. The construction uses that they are of equal size, and the proof uses only that all of them are from  $\mathcal{G}_d$ . Stinson's bound is tight for graphs in this extended family of bipartite  $d$ -regular graphs.

While graphs in  $\mathcal{G}_d$  can have arbitrary large girth, they are highly connected: between any two vertices there are  $d$  edge-disjoint paths. High connectivity, however, is compatible with low complexity: the complete graph  $K_n$  has complexity 1.

The proof of Theorem 12 stating that  $\mathcal{G}_d$  contains graphs of girth  $> g$ , can be used to estimate the size of this graph.  $N_2 \approx g$  as  $\mathcal{G}_2$  contains cycles;  $N_3 \approx 12 \cdot 2^{12g+4}$  by Lemma 10; and the inductive construction of Lemma 11 implies  $N_{d+1} \approx 12 \cdot 2^{36g} N_d$ , a really huge number. A  $d$ -regular graph with girth  $> g$  must contain at least  $d^g$  vertices, and there are such graphs with essentially that many vertices. It seems plausible that the family  $\mathcal{G}_d$  contains a  $> g$  girth graph with about that much vertices. We leave to prove (or refute) this claim as an open problem.

## Acknowledgment

This research was partially supported by the Lendület program of the Hungarian Academy of Sciences. The authors thank the members of the Crypto Group of the Rényi Institute, and especially Gábor Tardos, the numerous insightful discussions on the topic of this paper.

## References

1. A. Beimel, *Secret-sharing schemes: a survey*, in: Y. M. Chee et al (eds), Coding and Cryptology, IWCC 2011 LNCS vol **6639**, Springer
2. C. Blundo, A. De Santis, R. De Simone, U. Vaccaro, *Graph decomposition and secret sharing schemes*, J. Crypt. **8** (1995) pp. 39–64.
3. C. Blundo, A. De Santis, R. De Simone, U. Vaccaro, *Tight bounds on the information rate of secret sharing schemes*, Des., Codes and Crypt. **11** (2) (1997) pp. 107–110.
4. E. F. Brickell, D. R. Stinson, *Some improved bounds on the information rate of perfect secret sharing schemes*, J. Crypt. **5** (1992) pp. 153–166.
5. L. Csirmaz, *The size of a share must be large*, J. Crypt. **10** (4) (1997) pp. 223–231.
6. L. Csirmaz, *Secret sharing schemes on graphs*, Studia Math. Hung. **44** (2007), 297–306.
7. L. Csirmaz, P. Ligeti, *On an infinite family of graphs with information ratio  $2 - 1/k$* , Computing **85** (1) (2009) pp. 127–136.
8. L. Csirmaz, P. Ligeti, G. Tardos, *Erdős-Pyber theorem for hypergraphs and secret sharing*, Graphs and Combinatorics **31** (5) (2015) pp. 1335–1346
9. L. Csirmaz, G. Tardos, *Optimal Information Rate of Secret Sharing Schemes on Trees*, IEEE Trans. on Inf. Theory **59** (4) (2013) pp. 2527–2530.
10. M. van Dijk, *On the information rate of perfect secret sharing schemes*, Des., Codes and Crypt. **6** (2) (1995) pp. 143–169.
11. P. Erdos, L. Lovasz, *Problems and results on 3-chromatic hypergraphs and some related questions*, in: *Infinite and Finite Sets*, volume 11 of Colloq. Math. Soc. J. Bolyai (1975) pp. 609–627,
12. P. Erdos, L. Pyber, *Covering a graph by complete bipartite graphs*, Disc. Math. **170** (1-3) (1997) pp. 249–251.
13. D. R. Stinson, *Decomposition construction for secret sharing schemes*, IEEE Trans. on Inf. Theory, **40** (1) (1994) pp. 118–125.