



Atkins, R., Rombach, P., & Skerman, F. (2016). Guessing numbers of odd cycles.

Early version, also known as pre-print

Link to publication record in Explore Bristol Research PDF-document

# University of Bristol - Explore Bristol Research General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: http://www.bristol.ac.uk/pure/about/ebr-terms.html

# **Guessing Numbers of Odd Cycles**

Ross Atkins, Puck Rombach and Fiona Skerman

Abstract. For a given number of colours, s, the guessing number of a graph is the base s logarithm of the size of the largest family of colourings of the vertex set of the graph such that the colour of each vertex can be determined from the colours of the vertices in its neighbourhood. An upper bound for the guessing number of the *n*-vertex cycle graph  $C_n$  is n/2. It is known that the guessing number equals n/2 whenever n is even or s is a perfect square [7]. We show that, for any given integer  $s \ge 2$ , if a is the largest factor of s less than or equal to  $\sqrt{s}$ , for sufficiently large odd n, the guessing number of  $C_n$  with s colours is  $(n-1)/2 + \log_s(a)$ . This answers a question posed by Christofides and Markström in 2011 [7]. We also present an explicit protocol which achieves this bound for every n.

Linking this to index coding with side information, we deduce that the information defect of  $C_n$  with s colours is  $(n+1)/2 - \log_s(a)$  for sufficiently large odd n. Our results are a generalisation of the s = 2 case which was proven in [3].

Keywords: guessing number, cycle graph, information defect, index codes, unicast, entropy

# 1. Introduction

Computing the guessing number (Definition 1.2) of a graph G, can be equivalent to determining whether the multiple unicast coding problem [9] is solvable on a network related to G. The guessing number of a graph, G, is also studied for its relation to the information defect of G and index coding with side information [1, 11]. Exact guessing numbers are known only for a small number specific classes of graphs, such as perfect graphs, or small cases of non-perfect graphs [2, 5, 6, 15]. In particular, the guessing number of odd cycles, which is the focus of this paper, was not known, except for small cases [7, 3]. Here we compute the guessing number of the cycle graph,  $C_n$ , by analysing optimal protocols for the "guessing game".

The guessing game was introduced by Riis in 2007 [14]. It is a cooperative *n*-player information game played on a graph with *n* vertices with *s* colours. The guessing game on the complete graph  $K_n$ with s = 2 colours is played as follows. Each of the *n* players are given a hat that is red or blue uniformly and independently at random. Each player can see everyone else's hat, but not their own. The players collaboratively aim to maximise the probability that all players guess the colour of their hats correctly. Much of the popularity of this puzzle is owed to the striking difference between the success probability achieved by uncoordinated random guessing and an optimal protocol, which are  $1/2^n$  and 1/2 respectively.

The general guessing game considered here differs from many other variants of multiplayer information games (for example: the "hat guessing game" [4], "Ebert's game" [10] and the "hats-on-a-line game" [12]) in the following critical ways:

- The colours are assigned to each player independently and uniformly.
- Every player must guess (no passing or remaining silent).

- Each player does not necessarily see every other player's colours; two players can see each other if and only if they are joined by an edge in a given graph G.
- The players guess simultaneously so no communication is possible once the colours are assigned.
- The guessing game is won only if *all* the players guess correctly. An incorrect guess by any single player would mean that the whole team of *n* players collectively lose the guessing game (unlike [4], for example which seeks to optimise the number of players who guess correctly).

It is known that the greatest probability of winning the guessing game can be achieved by a deterministic protocol [5]. Let G = (V, E) be a graph where  $V = \{v_1, v_2, \ldots, v_n\}$  is the set of vertices and  $E \subseteq {V \choose 2}$  is the edge set. We restrict our attention to undirected graphs, but the problem is generalizes to directed graphs in an obvious way.

**Definition 1.1 (Protocol, colouring).** For any positive integer s, we let  $\mathbb{Z}_s$ , the group of all residues modulo s, denote the *colour set*. A *colouring* of G with s colours is an n-tuple  $c = (c_1, c_2, \ldots, c_n)$  such that  $c_i \in \mathbb{Z}_s$ . The set of all colourings of G with s colours is denoted  $\mathbb{Z}_s^n$ . A *protocol* on G with s colours is an n-tuple  $\mathcal{P} = (f_1, f_2, f_3, \ldots, f_n)$  where for each i, the [deterministic] function  $f_i : \mathbb{Z}_s^n \to \mathbb{Z}_s$  is such that  $f_i(c)$  is dependent only on  $c_j$  for all j such that  $v_i v_j \in E$ , *i.e.* for any i and any two colourings  $c = (c_1, c_2, \ldots, c_n)$  and  $c' = (c'_1, c'_2, \ldots, c'_n)$ , if  $c'_j = c_j$  for all j such that  $v_i v_j \in E$  then  $f_i(c) = f_i(c')$ . The fixed set of  $\mathcal{P}$ , Fix( $\mathcal{P}$ ), is the set of all invariant colourings:

$$\operatorname{Fix}(\mathcal{P}) = \left\{ c \in \mathbb{Z}_s^n \mid c_i = f_i(c) \; \forall i \right\}.$$

**Definition 1.2 (Fixed number, fixed set).** The *fixed number* of  $\mathcal{P}$  is the size of its fixed set;  $\operatorname{fix}(\mathcal{P}) = |\operatorname{Fix}(\mathcal{P})|$ . A protocol  $\mathcal{P}$  is called *non-trivial* if  $\operatorname{Fix}(\mathcal{P}) \neq \emptyset$ . A protocol is called *optimal* if it has maximal fixed number.

**Definition 1.3 (Guessing number).** The guessing number of G with s colours is defined as

$$\operatorname{gn}(G, s) = \log_s \max_{\mathcal{P}} [\operatorname{fix}(\mathcal{P})]$$

We assign the *n*-tuple of colours  $c \in \mathbb{Z}_s^n$  uniformly at random to the set of players, who are each identified with a vertex of G. The guesses of the players are given by  $\mathcal{P}(c)$ , so the players win if and only if  $c = \mathcal{P}(c)$ . Hence, the probability that an optimal protocol  $\mathcal{P}$  wins is

$$\mathbb{P}(c = \mathcal{P}(c)) = \frac{\operatorname{fix}(\mathcal{P})}{|\mathbb{Z}_s^n|} = s^{\operatorname{gn}(G,s)-n}$$

Christofides and Markström [7] showed that, for a perfect graph G and any s,  $gn(G, s) = n - \alpha$  where  $\alpha$  is the size of the largest independent set in G. For example, the complete graph  $K_n$  is a perfect graph with  $\alpha = 1$ , so an optimal protocol on  $K_n$ , wins with probability 1/s. The 3-cycle and the even-cycle  $C_{2k}$  (for any positive integer k) are both perfect graphs with  $\alpha(C_3) = 1$  and  $\alpha(C_{2k}) = k$  so

$$\operatorname{gn}(C_3, s) = 2$$
 and  $\operatorname{gn}(C_{2k}, s) = k \quad \forall k.$  (1.1)

Henceforth, we shall consider only the cycle graphs  $C_n$  for odd  $n \ge 5$ . In [7], it is shown that

$$\operatorname{gn}(C_5, 2) = 5$$

and the analysis in [3] shows that

$$gn(C_n, 2) = \frac{n-1}{2}$$
, for odd  $n \ge 7$ .

For general s, Christofides and Markström define protocols called "the clique strategy" and "the fractional-clique strategy" [7]. The fractional clique strategy is only defined when the number of colours s is a perfect power, and it is shown to be optimal on the odd cycle whenever s is a perfect square, *i.e.* 

$$\operatorname{gn}(C_n, m^2) = \frac{n}{2} \qquad \forall n, m.$$
(1.2)

In Definition 3.2, a protocol  $\mathcal{P}_{fcp}$  is defined on odd cycles for any number of colours *s*. The protocol  $\mathcal{P}_{fcp}$  is equivalent to the clique-strategy when *s* is prime, and to the fractional-clique-strategy when *s* is a perfect square. The protocol  $\mathcal{P}_{fcp}$  is called the *fractional-clique-partition protocol* to emphasise that it is very closely related to Christofides and and Markström's fractional-clique strategy. Our main result in Theorem 5.6 states that, for any given *s*, this fractional-clique-partition protocol is optimal on any large enough odd cycle.

The rest of this paper is organised as follows. In Section 2, we summarise a few of the known results on guessing numbers, and introduce the concepts of entropy and mutual information, which we will use heavily in our proofs. In Section 3, we define the fractional-clique-partition protocol, which is a refinement of the protocol introduced in [7] and we prove that for odd n, as the number of colours grows, this protocol achieves a fix( $\mathcal{P}$ ) lies between  $s^{n/2}$  and  $s^{n/2}(1 - \mathcal{O}(n/\sqrt{s}))$  (Theorem 3.5). In Section 4, we lay the technical groundwork which is needed for Section 5. Then, in Section 5, we focus on the case of large n compared to s, and we prove that the fractional-clique-partition protocol is in fact optimal on large enough odd cycles (Theorem 5.6). In Section 6, we link this to index coding with side information and compute the size of an optimal index code for  $C_n$  with s colours when n is odd and sufficiently large.

# 2. Backround Material and Notation

Many of our proofs will use the concept of the entropy of a random variable. Entropy is defined in Definition 2.2 and we list three crucial properties in Proposition 2.3. In this paper we take most logarithms base s, including inside the definitions of entropy. In the rest of this section, we present a few known results on the guessing number, define some useful random variables on the cycle graph and a notion of entropy, all of which will be used extensively in our proofs. When possible, we are consistent with the definitions and notations given in [5, 6, 7, 2, 13, 14]. We start with a small, useful result that shows, intuitively, that we are allowed to "forget" some colours.

**Proposition 2.1.** Let G be a graph, let s and s' be positive integers with  $s' \leq s$ , and let  $\mathcal{P}$  be any protocol on G with s colours. There exists a protocol  $\mathcal{P}'$  on G with s' colours such that

$$\left\{ c \in Fix(\mathcal{P}) \mid 0 \le c_i < s' \ \forall i \right\} \subseteq Fix(\mathcal{P}')$$

*Proof.* If  $\mathcal{P} = (f_1, f_2, \dots, f_n)$  then define  $\mathcal{P}' = (f'_1, f'_2, \dots, f'_n)$  in the following way.

- If  $0 \le c_j < s'$  for all j such that  $v_i v_j \in E$ , and  $0 \le f_i(c) < s'$  then  $f'_i(c) = f_i(c)$ .
- If  $s' \leq c_j < s$  for any j such that  $v_i v_j \in E$ , or  $s' \leq f_i(c) < s$  then  $f'_i(c) = 0$ .

For any colouring  $c \in Fix(\mathcal{P})$ , if  $0 \le c_i < s'$  for all i, then  $\mathcal{P}'(c) = \mathcal{P}(c) = c$  so  $c \in Fix(\mathcal{P}')$ .

**Definition 2.2 (Entropy, mutual information).** Let  $A_1, \ldots, A_k$  be random variables which take values in a finite set  $\mathcal{A}$ . The *entropy* of  $A_1, \ldots, A_k$  is denoted  $H(A_1, \ldots, A_k)$  and is given by:

$$H(A_1,\ldots,A_k) = -\sum_{a_1,\ldots,a_k \in \mathcal{A}^k} \mathbb{P}(A_1 = a_1,\ldots,A_k = a_k) \log_s \mathbb{P}(A_1 = a_1,\ldots,A_k = a_k).$$

The mutual information of  $A_1$  and  $A_2$  is denoted  $I(A_1; A_2)$  and is given by:

$$I(A_1; A_2) = H(A_1) + H(A_2) - H(A_1, A_2)$$

Let B be another random variable taking values in  $\mathcal{A}$ . The conditional mutual information of  $I(A_1; A_2|B)$  is given by

$$I(A_1; A_2|B) = H(A_1, B) + H(A_2, B) - H(A_1, A_2, B) - H(B).$$
(2.1)

**Proposition 2.3.** Let  $A_1$ ,  $A_2$  be random variables which take values in a finite sets A.

- 1.  $H(A_1) \leq \log |\mathcal{A}|$  with equality if and only if  $A_1$  is uniformly distributed.
- 2.  $I(A_1; A_2) \ge 0$  with equality if and only if  $A_1$  and  $A_2$  are independent.

 $\square$ 

3.  $I(A_1; A_2|B) \ge 0$  with equality if and only if  $A_1$  and  $A_2$  are independent conditional on B.

For a proof of the results in Proposition 2.3 we refer the reader to [8].

- **Definition 2.4.** For a non-empty set S, we use the notation  $A \in_u S$  to mean A is a random variable distributed uniformly over all elements in S.
- **Definition 2.5 (Notation for C<sub>n</sub>).** The cycle graph,  $C_n$ , has *n* vertices  $V = \{v_1, v_2, \ldots, v_n\}$ . The edge set of  $C_n$  is

 $E = \{v_i v_{i+1} \mid i = 1, 2, 3, \dots, n\}$ 

(indices are always taken modulo n). In a slight abuse of notation, for any protocol  $\mathcal{P} = (f_1, f_2, f_3, \ldots, f_n)$  on  $C_n$  with s colours, we say  $f_i : \mathbb{Z}_s^2 \to \mathbb{Z}_s$  where

$$f_i(c) = f_i(c_{i-1}, c_{i+1}).$$

Recall that a protocol  $\mathcal{P}$  is non-trivial if  $\operatorname{Fix}(\mathcal{P}) \neq \emptyset$ . For a given non-trivial protocol  $\mathcal{P}$  on  $C_n$ , define  $X = (X_1, X_2, \ldots, X_n)$  to be a colouring chosen uniformly at random from  $\operatorname{Fix}(\mathcal{P})$ . *i.e.* 

$$X \in_u \operatorname{Fix}(\mathcal{P}).$$

Note that the random colouring  $X = (X_1, X_2, \ldots, X_n)$  is only defined for non-trivial protocols  $\mathcal{P}$ . To simplify notation we will sometimes denote the entropy of a tuple of  $X_i$ s by

$$h(i_1, i_2, i_3, \ldots) = H(X_{i_1}, X_{i_2}, X_{i_3}, \ldots)$$

Since  $X_i$  is determined by  $(X_{i-1}, X_{i+1})$  we must have  $H(X_{i-1}, X_i, X_{i+1}) = H(X_{i-1}, X_{i+1})$ so h(i-1, i, i+1) = h(i-1, i+1). In general we can freely remove the argument *i* from  $h(\ldots, i-1, i, i+1, \ldots)$  as long as we don't remove the arguments i-1 and i+1.

 $h(\dots, i-1, i, i+1, \dots) = h(\dots, i-1, i+1, \dots)$ (2.2)

To simplify notation even further, for integers j < k, let  $H_j^k$  denote the quantity

$$H_j^k = h(j, j+1, j+2, \dots, k-1) + h(j+1, j+2, j+3, \dots, k).$$

**Proposition 2.6.** For any three integers i, j, k such that  $1 \le i < j$  and  $j + 1 < k \le n$ .

$$H_i^k \le H_i^j + H_{i+1}^k$$

*Proof.* We add up the following inequalities:

$$\begin{aligned} h(i, i+1, \dots, k-1) &= h(i, \dots, j-1, j+1, \dots, k-1) \\ &\leq h(i, \dots, j-1) + h(j+1, \dots, k-1), \\ \text{and} \quad h(i+1, i+2, \dots, k) &= h(i+1, \dots, j, j+2, \dots, k) \\ &\leq h(i+1, \dots, j) + h(j+2, \dots, k). \end{aligned}$$

**Lemma 2.7.** If  $\mathcal{P}$  is a non-trivial protocol on  $C_n$  with  $s \ge 2$  colours and  $X \in_u fix(\mathcal{P})$ , then, for all i,  $\log_s fix(\mathcal{P}) = H(X), h(i) \le 1.$ 

*Proof.* The entropy of any random variable over a finite domain is maximised when the variable is uniformly distributed. Therefore,  $h(i) = H(X_i) \leq H(U)$  where U is a random variable uniformly distributed over  $\mathbb{Z}_s$ . Hence,

$$h(i) \le H(U) = -\sum_{s \to 0} \frac{1}{s} \log_s \frac{1}{s} = 1.$$

The variable X is uniformly distributed over  $Fix(\mathcal{P})$ . Therefore,

$$H(X) = -\sum \frac{1}{\operatorname{fix}(\mathcal{P})} \log_s \frac{1}{\operatorname{fix}(\mathcal{P})} = \log_s \operatorname{fix}(\mathcal{P}).$$

**Lemma 2.8.** If  $\mathcal{P}$  is a non-trivial protocol on  $C_n$  with  $s \geq 2$  colours and  $X \in_u fix(\mathcal{P})$ , then

$$H_j^k \le \sum_{i=j}^k H(X_i)$$

for any  $j \leq 1$  and  $j + 3 \leq k \leq n$ .

and

*Proof.* We prove this by induction on (k - j). Recall that  $h(i_1, i_2, i_3, \ldots) = H(X_{i_1}, X_{i_2}, X_{i_3}, \ldots)$ .

• Base case: k = j + 3. Since  $X_{j+1} = f_{j+1}(X_j, X_{j+2})$  and  $X_{j+2} = f_{j+2}(X_{j+1}, X_{j+3})$  we have

$$\begin{split} h(j,j+1,j+2) &= h(j,j+2) \leq h(j) + h(j+2) \\ h(j+1,j+2,j+3) &= h(j+1,j+3) \leq h(j+1) + h(j+3), \end{split}$$

respectively. Adding these together yields:

$$H_j^{j+3} = h(j, j+1, j+2) + h(j+1, j+2, j+3) \le h(j) + h(j+1) + h(j+2) + h(j+3).$$

• Inductive step:  $k \ge j + 4$ . Since  $X_{k-1} = f_{k-1}(X_{k-2}, X_k)$  we have

$$h(j+1, j+2, \dots, k) = h(j+1, j+2, \dots, k-2, k)$$
  
$$\leq h(j+1, j+2, \dots, k-2) + h(k)$$

By Proposition 2.3,  $I(X_j; X_{k-1}|X_{j+1}, X_{j+2}, \ldots, X_{k-2}) \ge 0$ . Adding these together yields

$$H_i^k \le H_i^{k-1} + h(k)$$

This completes the proof.

**Lemma 2.9.** Let  $\mathcal{P}$  be a non-trivial protocol on  $C_n$  with  $s \geq 2$  colours and let  $X \in_u fix(\mathcal{P})$ . Suppose  $1 = d(1), d(2), d(3), \ldots, d(k) = n$  is a sequence of positive integers with  $k \geq 2$ . If  $d(i + 1) \geq d(i) + 2$  for all i, then

$$2\log_s fix(\mathcal{P}) = H_{d(1)}^{d(2)} + H_{d(2)+1}^{d(3)} + \dots + H_{d(k-1)+1}^{d(k)}.$$

*Proof.* We proceed by induction on k.

• Base case: k = 2. Since  $X_1 = f_1(X_n, X_2)$  and  $X_n = f_n(X_{n-1}, X_1)$ , we have

$$H(X) = h(2,3,4,\ldots,n)$$
 and  $H(X) = h(1,2,3,\ldots,n-1),$ 

respectively. Adding these together gives  $H_1^n = 2H(X) = 2\text{fix}(\mathcal{P})$ .

• Inductive step. By Proposition 2.6, for any  $d(k-1) + 2 \le d(k) \le n-2$ , we have

$$H^n_{d(k-1)+1} = H^{d(k)}_{d(k-1)+1} + H^n_{d(k)+1}.$$

#### 3. The Fractional-Clique-Partition Protocol

In this section, we define the fractional-clique-partition protocol,  $\mathcal{P}_{fcp}$ , on odd cycles  $C_n$  with  $s \geq 2$  colours. Theorem 3.4 appears in [7] and serves as a good upper bound for any  $n \geq 4$  and all numbers of colours.

**Definition 3.1 (Factorization bijection).** It is easy to see that for any factorization ab = s, there exists a bijection between  $\mathbb{Z}_s$  and  $\mathbb{Z}_a \times \mathbb{Z}_b$ . Let  $\phi(z) \times \psi(z)$  be such a bijection. For ease of notation, a and b are assumed to be given in context. Let  $\pi$  be the inverse of this bijection, so that  $\pi(\phi(z), \psi(z)) = z$  for all  $z \in \mathbb{Z}_s$ .



FIGURE 1. The protocol  $\mathcal{P}_{fcp}$  on  $C_7$  with s = ab colours, where a < b. Each vertex  $v_i$  is subdivided into two nodes representing the first and second components  $(\phi(c_i)$  and  $\psi(c_i)$ , respectively). The red edges (••••) represent pairs of first-components that are copying each other. The blue edges (••••) represent pairs of second-components that are copying each other. The black edge (••••) joins a first-component  $(\phi(c_n))$  and a second-component  $(\psi(c_1))$  which are copying each other as much as possible. For a colouring  $c \in \operatorname{Fix}(\mathcal{P}_{fcp})$  on  $C_7$ , there are a different choices for each red edge, b different choices for each blue edge and a different choices for the black edge. Therefore, fix $(\mathcal{P}_{fcp}) = a^4b^3 = as^3$  for n = 7.

**Definition 3.2 (Fractional-clique-partition protocol).** Let  $n \geq 3$  be an odd integer, let s be a positive integer, let a be the greatest factor of s less than or equal to  $\sqrt{s}$  and let b = s/a. For any colouring  $c = (c_1, c_2, \ldots, c_n) \in \mathbb{Z}_s^n$ , let  $\phi(c_i)$  and  $\psi(c_i)$  be referred to as the first and second coordinates respectively of vertex  $v_i$ . The fractional-clique-partition protocol is the protocol  $\mathcal{P}_{fcp} = (f_1, f_2, \ldots, f_n)$  on  $C_n$  defined by:

$$f_{i}(c_{i-1}, c_{i+1}) = \pi \big( \phi(c_{i-1}), \psi(c_{i+1}) \big) \qquad \text{for } i = 2, 4, 6, \dots, n-1$$
  

$$f_{i}(c_{i-1}, c_{i+1}) = \pi \big( \phi(c_{i+1}), \psi(c_{i-1}) \big) \qquad \text{for } i = 3, 5, 7, \dots, n-2$$
  

$$f_{1}(c_{n}, c_{2}) = \pi \big( \phi(c_{2}), \phi(c_{n}) \big) \qquad \text{and}$$
  

$$f_{n}(c_{n-1}, c_{1}) = \pi \big( \psi(c_{1}) (\text{mod } a), \psi(c_{n-1}) \big).$$

Informally, vertices  $v_{2k-1}$  and  $v_{2k}$  are copying each others first coordinate and vertices  $v_{2k}$  and  $v_{2k+1}$  are copying each others second coordinate (for k = 1, 2, 3, ..., (n-1)/2). Additionally, the second coordinate of vertex  $v_1$  and the first coordinate of vertex  $v_n$  copy each other as much as possible - whenever the second coordinate of vertex  $v_1$  is less than a. An example of  $\mathcal{P}_{fcp}$  on  $C_7$  is illustrated in Figure 1.

**Proposition 3.3.** For a given integer  $s \ge 2$  and odd integer  $n \ge 3$ , if a is the greatest factor of s less than or equal to  $\sqrt{s}$ , then we have  $fix(\mathcal{P}_{fcp}) = as^{(n-1)/2}$ .

*Proof.* Let n = 2k + 1. We count the number of colourings of  $C_n$  for which the protocol  $\mathcal{P}_{fcp}$  guesses correctly. For any colouring  $c \in \text{Fix}(\mathcal{P}_{fcp})$ , there are k pairs of vertices copying each other's first coordinates and there are a different choices for  $\phi$  for each pair. Similarly, for each of the k pairs of vertices copying each other's second coordinates, there are b different choices for  $\psi$ . This yields  $a^k b^k$ 

possibilities. Additionally, the first coordinate of vertex  $v_n$  must equal the second coordinate of vertex  $v_1$ , for which there are *a* possible colours. Multiplying these together yields

$$\operatorname{fix}(\mathcal{P}_{fcp}) = a^{k+1}b^k = as^{(n-1)/2}.$$

**Theorem 3.4.** [7] For any integer  $n \ge 4$ , we have  $gn(C_n, s) \le \frac{n}{2}$ , with equality only if for any optimal protocol,  $\mathcal{P}$  the following is satisfied. If  $X \in_u Fix(\mathcal{P})$  then  $H(X_i) = 1$  for all i.

*Proof.* Let  $\mathcal{P}$  be an optimal protocol on  $C_n$  with s colours. By Lemmas 2.7, 2.8 and 2.9, we have

$$gn(C_n, s) = \log_s fix(\mathcal{P}) = H(X) = \frac{1}{2}H_1^n \le \frac{1}{2}\sum_{i=1}^n h(i) \le \frac{n}{2}$$

If  $gn(C_n, s) = n/2$ , then we must have equality throughout, which means that h(i) = 1 for all i.  $\Box$ 

Theorem 3.4 appears in [7]. This same paper also shows that the limit of  $gn(C_n, s) \to n/2$  as  $s \to \infty$ . We give a bound on the rate convergence to this limit in Theorem 3.5.

**Theorem 3.5.** If n is odd and  $s = m^2 - t$  for integers m and  $t \ge 0$  then there exists a protocol  $\mathcal{P}$  on  $C_n$  with s colours such that

$$fix(\mathcal{P}) \ge s^{n/2}\left(1-\frac{tn}{s}\right).$$

*Proof.* Consider the protocol  $\mathcal{P}' = \mathcal{P}_{fcp}$  on  $C_n$  with  $s' = m^2$  colours and let  $X' \in_u \operatorname{Fix}(\mathcal{P}')$ . By Theorem 3.4, we must have  $H(X'_i) = 1$  and therefore  $X'_i$  is uniformly distributed over  $\mathbb{Z}_{s'}$  for all i. By the union bound,

$$\mathbb{P}\left(X'_i < s \; \forall \, i\right) \ge 1 - \sum_{i=1}^n \mathbb{P}(X'_i \ge s) = 1 - \sum_{i=1}^n \frac{t}{m^2} = 1 - \frac{tn}{m^2}.$$

Now, let  $\mathcal{P}$  be a protocol on  $C_n$  with s colours such that  $c \in \operatorname{Fix}(\mathcal{P})$  for all colourings  $c \in \operatorname{Fix}(\mathcal{P}')$  such that  $c_i < s$  for all i (such a protocol must exist by Proposition 2.1). For this protocol,

$$\begin{aligned} \operatorname{fix}(\mathcal{P}) &\geq \operatorname{fix}(\mathcal{P}') \mathbb{P}\left(X'_i < s \;\forall\; i\right) \\ &\geq \operatorname{fix}(\mathcal{P}') \left(1 - \frac{tn}{m^2}\right) \\ &= (s+t)^{n/2} \left(1 - tn(s+t)^{-1}\right) \\ &\geq s^{n/2} \left(1 - \frac{tn}{s}\right). \end{aligned}$$

**Corollary 3.6.** If  $n \ge 4$  then  $gn(C_n, s) = \frac{n}{2} - \mathcal{O}\left(\frac{n}{\sqrt{s \log_e s}}\right)$  as  $s \to \infty$ .

*Proof.* For even n we have  $gn(C_n, s) = \frac{n}{2}$ . For odd n, let m be the smallest positive integer such that  $m^2 \ge s$ . This gives  $t = m^2 - s = \mathcal{O}(\sqrt{s})$ . If  $\mathcal{P}$  is the protocol constructed in Theorem 3.5, then

$$\operatorname{gn}(C_n, s) \ge \log_s \operatorname{fix}(\mathcal{P}) \ge \frac{n}{2} + \log_s \left(1 - \frac{tn}{s}\right) = \frac{n}{2} - \mathcal{O}\left(\frac{n}{\sqrt{s}\log_e s}\right).$$

7



FIGURE 2. The structure of Sections 4 and 5. An arrow  $A \to B$  indicates that A is used in the proof of B.

#### 4. Entropy Results

The bounds in Theorem 3.5 are only useful when n is small relative to s. In contrast, the purpose of the results in this section is to establish Lemma 4.9, which in turn will be used to prove Theorem 5.6 which only applies when n is large relative to s. To help orientate the reader through this section (and the next), Figure 2 shows which results are used to prove other results.

**Definition 4.1 (Flat function, semi-perfect function).** For any  $z \in \mathbb{Z}_s$  and for any function  $f : \mathbb{Z}_s^2 \to \mathbb{Z}_s$  let  $f^{-1}(z) = \{(x, y) \mid f(x, y) = z\}$ . The function f is called *flat* if and only if  $|f^{-1}(z)| = s$  for all z. Let  $U = (U_1, U_2) \in_u \mathbb{Z}_s^2$ . A *semi-perfect* function, f, is any flat function such that the  $U_1$  and  $U_2$  are conditionally independent given f(U) (Definition 2.2), *i.e.* 

$$I(U_1; U_2 \mid f(U)) = 0.$$

**Definition 4.2 ((k, \epsilon)-uniform).** For any positive integer k and any  $\epsilon > 0$ , a random variable Y

is called 
$$(k, \epsilon)$$
-uniform if Y takes values in a finite set  $\mathcal{Y}$  with  $|\mathcal{Y}| = k$  and, for any  $y \in \mathcal{Y}_{+}$ 

$$\left|\mathbb{P}(Y=y) - \frac{1}{k}\right| \le \epsilon.$$

**Proposition 4.3.** For any integer  $k \ge 2$ , any integer  $s \ge 2$  and any  $\epsilon > 0$ , there exists  $\delta > 0$  such that, for any random variable Y which takes k distinct values, if H(Y) is the entropy of Y (base s), then

 $H(Y) \ge \log_s k - \delta \implies Y \text{ is } (k, \epsilon) \text{-uniform.}$ 

*Proof.* For each k, it suffices to show this for all small enough  $\epsilon$ . Assume  $7k\epsilon < 1$ . We prove the contrapositive:

• Suppose that  $\mathbb{P}(Y = y) \geq \frac{1}{k} + \epsilon$  for at least one value y. Entropy is greatest when Y is as uniformly distributed as possible. Therefore,

$$\begin{split} H(Y) &= -\sum_{i} \mathbb{P}(Y=i) \log_{s} \mathbb{P}(Y=i) \\ &\leq -\left(\frac{1}{k} + \epsilon\right) \log_{s} \left(\frac{1}{k} + \epsilon\right) - \left(k - 1\right) \left(\frac{1}{k} - \frac{\epsilon}{k - 1}\right) \log_{s} \left(\frac{1}{k} - \frac{\epsilon}{k - 1}\right) \\ &= \log_{s} k - \left(\frac{1}{k} + \epsilon\right) \log_{s} (1 + k\epsilon) - \left(\frac{k - 1}{k} - \epsilon\right) \log_{s} \left(1 - \frac{k\epsilon}{k - 1}\right). \end{split}$$

Since  $0 < k\epsilon < \frac{1}{7}$ , we can use the identity,  $-\log_s(1-\gamma) \le (\gamma + \frac{5}{9}\gamma^2)\log_e s$  (valid for  $|\gamma| \le 1/7$ ), to simplify this expression.

$$\begin{split} H(Y) &\leq \log_s k - \frac{k\epsilon^2}{9} \left( 4 - 5k\epsilon + \frac{4}{k-1} + \frac{5k\epsilon}{(k-1)^2} \right) \log_e s \\ &\leq \log_s k - \frac{k\epsilon^2}{3} \log_e s. \end{split}$$

• Now suppose  $\mathbb{P}(Y = y) \leq \frac{1}{k} - \epsilon$  for at least one value y. The entropy would be greatest when Y is as uniformly distributed as possible. Therefore,

$$\begin{split} H(Y) &= -\sum_{i} \mathbb{P}(Y=i) \log_{s} \mathbb{P}(Y=i) \\ &\leq -\left(\frac{1}{k} - \epsilon\right) \log_{s} \left(\frac{1}{k} - \epsilon\right) - \left(k - 1\right) \left(\frac{1}{k} + \frac{\epsilon}{k-1}\right) \log_{s} \left(\frac{1}{k} - \frac{\epsilon}{k-1}\right) \\ &= \log_{s} k - \left(\frac{1}{k} - \epsilon\right) \log_{s} (1 - k\epsilon) - \left(\frac{k-1}{k} + \epsilon\right) \log_{s} \left(1 + \frac{k\epsilon}{k-1}\right). \end{split}$$

We can use the identity,  $-\log_s(1-\gamma) \leq (\gamma + \frac{5}{9}\gamma^2)\log_e s$ , again to simplify this expression.

$$\begin{split} H(Y) &\leq \log_s k - \frac{k\epsilon^2}{9} \left( 4 - 5k\epsilon + \frac{4}{k-1} - \frac{5k\epsilon}{(k-1)^2} \right) \log_e s \\ &\leq \log_s k - \frac{k\epsilon^2}{3} \log_e s. \end{split}$$

In either case,  $H(Y) < \log_s k - \delta$  for any  $\delta < \frac{k\epsilon^2}{3} \log_e s$ .

**Lemma 4.4.** For any integer  $s \ge 2$ , there exists positive constant  $\epsilon = \epsilon(s)$  that satisfies the following property. For any non semi-perfect function  $f : \mathbb{Z}_s^2 \to \mathbb{Z}_s$  and for any three  $(s, \epsilon)$ -uniform random variables  $Y_1, Y_2, Y_3$  over  $\mathbb{Z}_s$  satisfying  $Y_2 = f(Y_1, Y_3)$ , if  $(Y_1, Y_3)$  is  $(s^2, \epsilon)$ -uniform, then

$$I(Y_1; Y_3 | Y_2) \ge \frac{1}{2} \min \left\{ I(U_1; U_2 | f(U)) \mid f \text{ is a flat but not semi-perfect} \right\} = \delta_1,$$
  
=  $(U_1, U_2) \in_{\mathcal{U}} \mathbb{Z}^2.$ 

where  $U = (U_1, U_2) \in_u \mathbb{Z}_s^2$ .

*Proof.* The value  $\delta_1 = \delta_1(s) = \frac{1}{2} \min \left\{ I(U_1; U_2 | f(U)) \mid f \text{ is a flat but not semi-perfect} \right\}$  is well-defined for any  $s \ge 2$ , because there are only a finite number of functions  $f : \mathbb{Z}_s^2 \to \mathbb{Z}_s$ , and at least one of them is flat and not semi-perfect hence we can take the minimum of these. For example, the function  $f(x, y) = x + y \pmod{s}$  is flat but not semi-perfect. First, let  $\epsilon < \frac{1}{s^2(s+2)}$ , so that

$$\frac{1}{s^2} - (s-1)\epsilon > \frac{1}{s^2} - (s+1)\epsilon > \epsilon$$

We show that f is flat by contradiction. Since  $(Y_1, Y_3)$  is  $(s^2, \epsilon)$ -uniform:

• If  $|f^{-1}(z)| \ge s + 1$  then

$$\mathbb{P}(Y_2 = z) = \mathbb{P}((Y_1, Y_3) \in f^{-1}(z)) \ge (s+1)\left(\frac{1}{s^2} - \epsilon\right) = \frac{1}{s} + \left(\frac{1}{s^2} - (s+1)\epsilon\right) > \frac{1}{s} + \epsilon.$$

• If 
$$|f^{-1}(z)| \le s - 1$$
 then  
 $\mathbb{P}(Y_2 = z) = \mathbb{P}((Y_1, Y_3) \in f^{-1}(z)) \le (s - 1) \left(\frac{1}{s^2} + \epsilon\right) = \frac{1}{s} - \left(\frac{1}{s^2} - (s - 1)\epsilon\right) < \frac{1}{s} - \epsilon.$ 

$$\square$$

Both cases contradict the assumption that  $Y_2$  is  $(s, \epsilon)$ -uniform. Therefore f is a flat function and so

$$I(U_1; U_2 | f(U)) \ge 2\delta_1.$$

Moreover, since  $(Y_1, Y_3)$  is  $(s^2, \epsilon)$ -uniform, then U and  $(Y_1, Y_3)$  differ in distribution by less than  $\epsilon$ . Since mutual information is continuous, we can choose  $\epsilon$  small enough so that

$$|I(U_1; U_2|f(U)) - I(Y_1; Y_3|Y_2)| \le \delta_1$$

Then, by the triangle inequality,  $I(Y_1; Y_3 | Y_2) \ge \delta_1$ .

**Definition 4.5.** From now on, for any integer  $s \ge 2$ , let  $\epsilon = \epsilon(s) > 0$  be chosen small enough so that  $\epsilon \le \frac{1}{s^2(2s+1)}$  and  $\epsilon$  satisfies Lemma 4.4. Then let  $\delta_2 = \delta_2(s) > 0$  be chosen small enough to satisfy Proposition 4.3 for both k = s and  $k = s^2$  for this value  $\epsilon$ . Then, with  $\delta_1$  as defined in Lemma 4.4, let  $\delta = \min(\delta_1, \delta_2)$ .

**Lemma 4.6.** Let  $n \ge 5$  be an integer and let  $\mathcal{P}$  be any non-trivial protocol on  $C_n$  with  $s \ge 2$  colours. The random variables  $X_1, X_2, X_3, X_4, X_5$  (Definition 2.5) satisfy:

$$H_1^5 \le 3 + h(2,4) - I(X_2; X_4 | X_3).$$

*Proof.* By Lemma 2.7, it suffices to show  $H_1^5 \leq h(1) + h(3) + h(5) + h(2,4) - I(X_2; X_4|X_3)$ . By Shannon's Inequality (Proposition 2.3) we have:

$$h(2,3,4) + h(3) = h(2,3) + h(3,4) - I(X_2; X_4 | X_3),$$

$$(4.1)$$

$$h(1,2,3,4) + h(2,3) \le h(1,2,3) + h(2,3,4), \tag{4.2}$$

and  $h(2,3,4,5) + h(3,4) \le h(2,3,4) + h(3,4,5).$  (4.3)

Also, since  $X_i = f_i(X_{i-1}, X_{i+1})$  for i = 2, 3, 4 respectively we have:

$$h(1,2,3) = h(1,3) \le h(1) + h(3), \tag{4.4}$$

$$h(2,3,4) = h(2,4), \tag{4.5}$$

and 
$$h(3,4,5) = h(3,5) \le h(3) + h(5).$$
 (4.6)

The required result is the sum of equations (4.1), (4.2), (4.3), (4.4), (4.5) and (4.6).

**Lemma 4.7.** Let  $n \ge 5$  be an integer and let  $\mathcal{P} = (f_1, f_2, \ldots, f_n)$  be a non-trivial protocal on  $C_n$  with  $s \ge 2$  colours and let  $X \in_u fix(\mathcal{P})$ . For any j, if  $f_{j+2}$  is not semi-perfect or  $(X_{j+1}, X_{j+3})$  is not  $(s^2, \epsilon)$ -uniform then  $H_j^{j+4} \le 5 - \delta$ , for  $\delta$  as in Definition 4.5.

*Proof.* Without loss of generality let j = 1. There are 3 cases.

• If, for any  $i \in \{1, 2, 3, 4, 5\}$ , the variable  $X_i$  is not  $(s, \epsilon)$ -uniform, then  $h(i) \leq 1 - \delta_2$  (Proposition 4.3). In this case, by Lemma 2.8,

$$H_1^5 \le \sum_{i=1}^5 h(i) \le 5 - \delta_2.$$

• If  $(X_2, X_4)$  is not  $(s^2, \epsilon)$ -uniform, then  $h(2, 4) \le 2-\delta_2$  (Proposition 4.3). Therefore, by Lemma 4.6, we have

$$H_1^5 \le 3 + h(2,4) - I(X_2; X_4 | X_3) \le 5 - \delta_2.$$

• Otherwise,  $X_2, X_3, X_4$  are each  $(s, \epsilon)$ -uniform and  $(X_2, X_4)$  is  $(s^2, \epsilon)$ -uniform and  $f_3$  is not semiperfect. In this case, by Lemma 4.4, we have  $I(X_{j+1}; X_{j+3}|X_{j+2}) \geq \delta_1$ . By Lemma 4.6, we have

$$H_1^5 \le 3 + h(2,4) - I(X_2; X_4 | X_3) \le 5 - \delta_1$$

In all cases, we have  $H_1^5 \leq 5 - \delta$  because  $\delta = \min{\{\delta_1, \delta_2\}}$ .

_		
_	_	

**Lemma 4.8.** Let  $n \geq 7$  be an integer and let  $\mathcal{P} = (f_1, f_2, \dots, f_n)$  a non-trivial protocol on  $C_n$  with  $s \geq 2$  colours and let  $X \in_u fix(\mathcal{P})$ . For any j, if any of  $f_{j+2}$ ,  $f_{j+3}$  or  $f_{j+4}$  are not semi-perfect, or any of  $(X_{j+1}, X_{j+3})$ ,  $(X_{j+2}, X_{j+4})$  or  $(X_{j+3}, X_{j+5})$  are not  $(s^2, \epsilon)$ -uniform, then  $H_j^{j+6} \leq 7 - \delta$ .

*Proof.* Without loss of generality let j = 1. We treat each case individually, and use Lemma 4.7.

• If  $f_3$  is not semi-perfect or  $(X_2, X_4)$  is not  $(s^2, \epsilon)$ -uniform then

$$\begin{aligned} H_1^7 &= h(1,2,3,4,5,6) + h(2,3,4,5,6,7) \\ &= h(1,2,3,4,6) + h(2,3,4,5,7) \\ &\leq H_1^5 + h(6) + h(7) \\ &\leq (5-\delta) + 1 + 1. \end{aligned}$$

• If  $f_4$  is not semi-perfect or  $(X_3, X_5)$  is not  $(s^2, \epsilon)$ -uniform then

$$\begin{aligned} H_1^7 &= h(1,2,3,4,5,6) + h(2,3,4,5,6,7) \\ &= h(1,3,4,5,6) + h(2,3,4,5,7) \\ &\leq h(1) + H_2^6 + h(7) \\ &\leq 1 + (5-\delta) + 1. \end{aligned}$$

• If  $f_5$  is not semi-perfect or  $(X_4, X_6)$  is not  $(s^2, \epsilon)$ -uniform then

$$\begin{split} H_1^7 &= h(1,2,3,4,5,6) + h(2,3,4,5,6,7) \\ &= h(1,3,4,5,6) + h(2,4,5,6,7) \\ &\leq h(1) + h(2) + H_3^7 \\ &\leq 1 + 1 + (5 - \delta). \end{split}$$

г		
		L

**Lemma 4.9.** Let  $n \ge 7(\delta^{-1} + 2)$ . Suppose  $\mathcal{P} = (f_1, f_2, \ldots, f_n)$  is a non-trivial protocol on  $C_n$  with  $s \ge 2$  colours and let  $X \in_u fix(\mathcal{P})$  such that, for each j, either

- at least one of f<sub>j-1</sub>, f<sub>j</sub>, f<sub>j+1</sub> is not semi-perfect, or
  at least one of (X<sub>j-2</sub>, X<sub>j</sub>), (X<sub>j-1</sub>, X<sub>j+1</sub>), (X<sub>j</sub>, X<sub>j+2</sub>) is not (s<sup>2</sup>, ε)-uniform,

then  $fix(\mathcal{P}) < s^{(n-1)/2}$ 

*Proof.* Let m be an odd integer such that  $m > \delta^{-1}$  and  $7m \le n$ . By Lemma 2.9 and Lemma 4.8, we have

$$2H(X) \le \sum_{j=0}^{m-1} H_{7j+1}^{7j+7} + \sum_{i=7m}^{n-1} h(i)$$
$$\le m(7-\delta) + (n-7m)$$
$$= n - m\delta.$$

Since  $m > \delta^{-1}$ , this means that  $H(X) < \frac{n-1}{2}$ . Therefore fix $(\mathcal{P}) < s^{(n-1)/2}$  by Lemma 2.7. 

# 5. Guessing numbers of large odd cycles

In this section, we prove our main result in Theorem 5.6, which states that, for any given s, this fractional-clique-partition protocol is optimal on any large enough odd cycle.

**Definition 5.1 (Perfect function).** For any function  $f : \mathbb{Z}_s^2 \to \mathbb{Z}$ , let L(f, z) and R(f, z) denote the subsets

$$L(f, z) = \{x \mid f(x, y) = z \text{ for some } y\}$$
  
and 
$$R(f, z) = \{y \mid f(x, y) = z \text{ for some } x\}.$$

The function f is called a *perfect* function if it is semi-perfect and the cardinalities |L(f, z)|and |R(f, z)| do not depend on z, *i.e.* if |L(f, z)| = |L(f, z')| and |R(f, z)| = |R(z')| for all  $z, z' \in \mathbb{Z}_s$ .

**Proposition 5.2.** If f is a semi-perfect function then for all  $z \in \mathbb{Z}_s$  then

$$f^{-1}(z) = L(f, z) \times R(f, z).$$

Moreover |L(f,z)||R(f,z)| = s.

*Proof.* Let  $U = (U_1, U_2) \in_u \mathbb{Z}_s^2$  and for a given z, let L = L(f, z) and R = R(f, z). Since f is semiperfect, we have  $I(U_1, U_2 \mid f(U)) = 0$ . Therefore,  $U_1$  and  $U_2$  are conditionally independent given f(U). For any  $x \in L$  and any  $y \in R$ , we must have

$$\mathbb{P}(U_1 = x \land U_2 = y | f(U) = z) = \mathbb{P}(U_1 = x | f(U) = z) \mathbb{P}(U_2 = y | f(U) = z) > 0,$$
  
and  $f^{-1}(z) = L \times R.$ 

Furthermore, since  $U_1$  and  $U_2$  are independently uniformly distributed over  $\mathbb{Z}_s$  and U is uniformly distributed over  $\mathbb{Z}_s^2$ , we have

$$\frac{1}{s} = \mathbb{P}(f(U) = z) = \mathbb{P}(U_1 \in L \land U_2 \in R) = \mathbb{P}(U_1 \in L) \times \mathbb{P}(U_2 \in R) = \frac{|L|}{s} \times \frac{|R|}{s}.$$
  
e,  $|L||R| = s.$ 

Therefore, |L||R| = s.

**Lemma 5.3.** Let  $s \ge 2$  be an integer, let  $0 < \epsilon \le \frac{1}{s^2(2s+1)}$  be a constant. Let  $\mathcal{P} = (f_1, f_2, \ldots, f_n)$  be any non-trivial protocol on  $C_n$  with s colours and let  $X \in_u Fix(\mathcal{P})$ . If  $f_1$  and  $f_2$  are semi-perfect functions and  $(X_0, X_2)$  and  $(X_1, X_3)$  are  $(s^2, \epsilon)$ -uniform, then, for any  $c_1, c_2 \in \mathbb{Z}_s$ , we have

$$|\{c_0|f_1(c_0,c_2)=c_1\}|=|\{c_3|f_2(c_1,c_3)=c_2\}|.$$

*Proof.* We proceed by contradiction. Let  $S_0 = \{c_0 | f_1(c_0, c_2) = c_1\}$  and  $S_3 = \{c_3 | f_2(c_1, c_3) = c_2\}$ . Without loss of generality assume  $|S_0| < |S_3|$  so since  $|S_0| < s$  we must have  $|S_3| > (1 + \frac{1}{s}) |S_0|$ . Now since  $(X_0, X_2)$  is  $(s^2, \epsilon)$ -uniform,

$$\mathbb{P}(X_1 = c_1 \land X_2 = c_2) = \sum_{x \in S_0} \mathbb{P}((X_0, X_2) = (x, c_2)) \le |S_0| \left(\frac{1}{s^2} + \epsilon\right).$$

Similarly, since  $(X_1, X_3)$  is  $(s^2, \epsilon)$ -uniform,

$$\mathbb{P}(X_1 = c_1 \land X_2 = c_2) = \sum_{x \in S_3} \mathbb{P}((X_1, X_3) = (c_1, x)) \ge |S_3| \left(\frac{1}{s^2} - \epsilon\right).$$

However, since  $\epsilon \leq \frac{1}{s^2(2s+1)}$ , this implies

$$1 + \frac{1}{s} < \frac{|S_3|}{|S_0|} \le \frac{s^{-2} + \epsilon}{s^{-2} - \epsilon} \le \frac{\frac{1}{s^2} + \frac{1}{s^2(2s+1)}}{\frac{1}{s^2} - \frac{1}{s^2(2s+1)}} = 1 + \frac{1}{s},$$

which is a contradiction.

**Lemma 5.4.** Let  $\mathcal{P} = (f_1, f_2, \dots, f_n)$  be a non-trivial protocol on  $C_n$  with  $s \geq 2$  colours, let  $X \in_u$ Fix( $\mathcal{P}$ ) and let j be any index (indices taken modulo n). If  $f_{j-1}$ ,  $f_j$  and  $f_{j+1}$  are semi-perfect functions and  $(X_{j-2}, X_j)$ ,  $(X_{j-1}, X_{j+1})$  and  $(X_j, X_{j+2})$  are  $(s^2, \epsilon)$ -uniform, then  $f_j$  is a perfect function.

*Proof.* We proceed by contradiction. Without loss of generality, assume j = 0 and fix  $c_0, c'_0 \in \mathbb{Z}_s$  arbitrarily. Now choose  $c_{-1}, c_1 \in \mathbb{Z}_s$  such that  $f_0(c_{-1}, c_1) = c_0$  and choose  $c'_{-1}, c'_1 \in \mathbb{Z}_s$  such that  $f_0(c'_{-1}, c'_1) = c'_0$ . Also let  $c''_0 = f_0(c'_{-1}, c_1)$ . Now by Lemma 5.3,

$$|L(f_0, c_0)| = |\{x|f_0(x, c_1) = c_0\}| = |\{x|f_1(c_0, x) = c_1\}| = |R(f_1, c_1)|$$
  
and 
$$|L(f_0, c_0'')| = |\{x|f_0(x, c_1) = c_0''\}| = |\{x|f_1(c_0'', x) = c_1\}| = |R(f_1, c_1)|$$

Similarly

an

$$|R(f_0, c_0'')| = |\{x|f_0(c_{-1}', x) = c_0''\}| = |\{x|f_{-1}(x, c_0'') = c_{-1}'\}| = |L(f_{-1}, c_{-1}')|$$
  
d 
$$|R(f_0, c_0')| = |\{x|f_0(c_{-1}', x) = c_0'\}| = |\{x|f_{-1}(x, c_0') = c_{-1}'\}| = |L(f_{-1}, c_{-1}')|.$$

Recall that  $|L(f_0, z)| \cdot |R(f_0, z)| = s$  for all  $z \in \mathbb{Z}_s$  (Proposition 5.2). Therefore,  $|R(f_0, c'_0)| = |R(f_0, c''_0)|$  if and only if  $|L(f_0, c'_0)| = |L(f_0, c''_0)|$ . Hence,

$$|L(f_0, c_0)| = |L(f_0, c_0'')| = |L(f_0, c_0')|.$$

Similarly,  $|R(f_0, c_0)| = |R(f_0, c'_0)|$  (for arbitrary  $c_0, c'_0 \in \mathbb{Z}_s$ ) and therefore  $f_0$  is a perfect function.  $\Box$ 

**Lemma 5.5.** Let  $\mathcal{P} = (f_1, f_2, \ldots, f_n)$  be a non-trivial protocol on  $C_n$  with  $s \ge 2$  colours, such that  $f_j$  is a perfect function for some j. Then  $fix(\mathcal{P}) \le as^{(n-1)/2}$ , where a is the greatest factor of s less than or equal to  $\sqrt{s}$ .

*Proof.* Without loss of generality, assume j = 2. Since  $f_2$  is perfect, let  $l = |L(f_2, z)|$  and  $r = |R(f_2, z)|$ . Without loss of generality, assume  $l \leq r$  and therefore  $l \leq a$ . Then  $X_2$  takes at most s different values and  $X_1$ , conditioned on  $X_2 = z$  for any  $z \in \mathbb{Z}_s$ , takes at most l different values. Therefore, the pair  $(X_1, X_2)$  takes at most ls different values in  $\mathbb{Z}_s^2$  and  $h(1, 2) \leq \log_s(ls)$ . We have

$$\begin{aligned} H(X) &= h(1,2,3,\ldots,n) \\ &= h(1,2,4,6,\ldots,n-3,n-1) \\ &\leq h(1,2) + \sum_{i=1}^{(n-3)/2} h(2i+2) \\ &\leq \log_s(ls) + \frac{n-3}{2}. \end{aligned}$$

Therefore fix $(\mathcal{P}) = s^{H(X)} \le ls^{(n-1)/2} \le as^{(n-1)/2}$ .

**Theorem 5.6.** For any integer  $s \ge 2$ , let a be the greatest factor of s less than or equal to  $\sqrt{s}$ . There exists some  $N \in \mathbb{N}$  such that

$$gn(C_n, s) = \begin{cases} \frac{n}{2}, & \text{for even } n, \\ \frac{n-1}{2} + \log_s a, & \text{for odd } n > N, \end{cases}$$

and  $\mathcal{P}_{fcp}$  is an optimal protocol on  $C_n$  with s colours for any odd  $n \geq N$ .

*Proof.* Let  $\epsilon$  and  $\delta$  be the values given in Definition 4.5, let  $N = 7(\delta^{-1} + 2)$  and let  $\mathcal{P} = (f_1, f_2, \dots, f_n)$  be any non-trivial protocol on  $C_n$  with s colours. We have two cases:

Case one. For all j, either:

- at least one of the functions  $f_{j-1}$ ,  $f_j$  and  $f_{j+1}$  is not semi-perfect or
- at least one of  $(X_{j-2}, X_j)$ ,  $(X_{j-1}, X_{j+1})$ ,  $(X_j, X_{j+2})$  is not  $(s^2, \epsilon)$ -uniform.

Case two. There exists some j such that:

- the functions  $f_{j-1}$ ,  $f_j$  and  $f_{j+1}$  are all semi-perfect and
- $(X_{j-2}, X_j), (X_{j-1}, X_{j+1})$  and  $(X_j, X_{j+2})$  are all  $(s^2, \epsilon)$ -uniform.

For case one, we can conclude that  $\operatorname{fix}(\mathcal{P}) \leq s^{(n-1)/2} \leq \operatorname{fix}(\mathcal{P}_{fcp})$  by Lemma 4.9. In case two,  $f_j$  must be a perfect function (Lemma 5.4) and then  $\operatorname{fix}(\mathcal{P}) \leq as^{(n-1)/2} = \operatorname{fix}(\mathcal{P}_{fcp})$  (Lemma 5.5). In either case,  $\operatorname{fix}(\mathcal{P}_{fcp}) \geq \operatorname{fix}(\mathcal{P})$ . Hence  $\mathcal{P}_{fcp}$  is optimal.

# 6. An application to index coding with side information

In the problem of index coding with side information on a graph G, a sender aims communicate n messages  $c_1, c_2, \ldots, c_n$  (where  $c_i \in \mathbb{Z}_s$ ) to n receivers  $v_1, v_2, \ldots, v_n$  (the vertices of G). Each receiver,  $c_i$ , knows  $c_j$  in advance, for each j such that  $v_i v_j$  is an edge in G. The sender is required to broadcast a message to all receivers (the same message to all receivers) so that each receiver,  $v_i$ , can recover  $c_i$ . If m is the smallest integer such that the sender can achieve this by broadcasting one of only m different messages, then the *information defect* [13] of G with s colours is defined to be

$$\beta(G, s) = \log_s(m).$$

The relationship between the guessing number and information defect of a graph is well known. Explicitly, let  $\mathfrak{C}_s(G)$  be the *confusion graph* [1, 3] (also known as the "code graph" [7]), defined to have vertex set  $\mathbb{Z}_s^n$ , in which two vertices  $c, c' \in \mathbb{Z}_s^n$  are adjacent if and only if for some  $i \in [n], c_i \neq c'_i$ but for each j such that  $ij \in E(G)$  we have  $c_j = c'_j$ . Intuitively  $c, c' \in \mathbb{Z}_s^n$  are 'confusable' (joined by an edge in the confusion graph) if there is no protocol  $\mathcal{P}$ , for the guessing game on G, such that both  $c, c' \in \operatorname{Fix}(\mathcal{P})$  (*i.e.* c and c' cannot both be encoded with the same message from the sender.). If  $\chi(\mathfrak{C}_s(G))$  is chromatic number of the confusion graph of G and  $\alpha(\mathfrak{C}_s(G))$  is the size of the largest independent set in the confusion graph of G, then

$$\beta(G,s) = \log_s \chi(\mathfrak{C}_s(G))$$
 and  $\operatorname{gn}(G,s) = \log_s \alpha(\mathfrak{C}_s(G)).$ 

For any graph H, we have the identity  $\chi(H)\alpha(H) \ge |H|$  and so we have the identity [13]

$$\beta(G,s) + \operatorname{gn}(G,s) \ge \log_s |\mathfrak{C}_s(G)| = n.$$

We use this identity and the fact that the fractional-clique protocol  $\mathbb{P}_{fcp}$  is optimal (Theorem 5.6) to prove Theorem 6.1. This theorem in general is a new result, although the case s = 2 was proven combinatorially in [3]. Theorem 6.1 shows that the size of an optimal index code,  $\beta(G, s)$ , depends on the factorisation structure of the size of the alphabet, s, used for the input.

**Theorem 6.1.** For a given s, let b be the smallest factor of s which is at least  $\sqrt{s}$ . There exists some N such that for all odd n > N,

$$\beta(C_n, s) = \frac{n-1}{2} + \log_s b.$$

*Proof.* Write a = s/b. First by Theorem 5.6,  $gn(C_n, s) = (n-1)/2 + \log_s a$  for all large enough odd n. Therefore,

$$\beta(C_n, s) \ge n - \operatorname{gn}(C_n, s) = \frac{n-1}{2} + \log_s b.$$

To show that we in fact get equality, we define a set of  $bs^{(n-1)/2}$  possible messages with which the sender can solve the index coding with side information problem on  $C_n$ . Let  $\phi$  and  $\psi$  be defined as in Definition 3.1. This means that  $\phi \times \psi$  is a bijection from  $\mathbb{Z}_a \times \mathbb{Z}_b$  to  $\mathbb{Z}_s$ . Now for any colouring  $c = (c_1, c_2, \ldots, c_n) \in \mathbb{Z}_s^n$  let the sender broadcast the following values:

- For  $i = 1, 2, 3, \ldots, \frac{n-1}{2}$ , the sender broadcasts the residue  $\phi(c_{2i-1}) + \phi(c_{2i})$  modulo a and the residue  $\psi(c_{2i}) + \psi(c_{2i+1})$  modulo b.
- Additionally, the sender broadcasts the residue  $\psi(c_1) + \phi(c_n)$  modulo b.

The sender broadcasts  $\frac{n-1}{2}$  residues modulo a and  $\frac{n+1}{2}$  residues modulo b, and so the total number of possible messages that the sender might send is

$$m = a^{(n-1)/2} b^{(n+1)/2} = b s^{(n-1)/2}$$

Furthermore, each receiver,  $v_i$ , knows  $c_{i-1}$  and  $c_{i+1}$ , and so can recover both  $c_i$  because she can recover both  $\phi(c_i)$  and  $\psi(c_i)$ .

# 7. Acknowledgements

Puck Rombach is supported by AFOSR MURI Grant No. FA9550-10-1-0569 and ARO MURI Grant No. W911NF-11-1-0332.

### References

- Noga Alon, Avinatan Hasidim, Eyal Lubetzky, Uri Stav, and Amit Weinstein. Broadcasting with side information. arXiv preprint arXiv:0806.3246, 2008.
- [2] Rahil Baber, Demetres Christofides, Anh N. Dang, Søren Riis, and Emil R. Vaughan. Graph guessing games and non-shannon information inequalities. arXiv:1410.8349, 2014.
- [3] Zic Bar-Yossef, Yitzhak Birk, T.S. Jayram, and Tomer Kol. Index coding with side information. Information Theory, IEEE Transactions on, 57(3):1479–1494, 2011.
- [4] Steve Butler, Mohammad T. Hajiaghayi, Robert D. Kleinberg, and Tom Leighton. Hat guessing games. SIAM Review, 51(2):399–413, 2009.
- [5] Peter J. Cameron, Anh N. Dang, and Søren Riis. Guessing games on triangle-free graphs. arXiv preprint arXiv:1410.2405, 2014.
- [6] Gerard Jennhwa Chang, Keqin Feng, Liang-Hao Huang, and Mei Lu. The linear guessing number of undirected graphs. *Linear Algebra and its Applications*, 449:119–131, 2014.
- [7] Demetres Christofides and Klas Markström. The guessing number of undirected graphs. *The Electronic Journal of Combinatorics*, 18(1):P192, 2011.
- [8] Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. Wiley-Interscience, Hoboken, NJ, 2006.
- [9] Randall Dougherty and Kenneth Zeger. Nonreversibility and equivalent constructions of multiple unicast networks. *IEEE Trans. Inf. Theory*, 52:1287–1291, 2006.
- [10] Todd Ebert. Applications of recursive operators to randomness and complexity. PhD thesis, University of California Santa Barbara, 1998.
- [11] Maximilien Gadouleau and Søren Riis. Graph-theoretical constructions for graph entropy and network coding based communications. *Information Theory, IEEE Transactions on*, 57(10):6703–6717, 2011.
- [12] Maura B. Paterson and Douglas R. Stinson. Yet another hat game. The Electronic Journal of Combinatorics 17, 2010.
- [13] Søren Riis. Graph entropy, network coding and guessing games. arXiv preprint arXiv:0711.4175, 2007.
- [14] Søren Riis. Information flows, graphs and their guessing numbers. Electronic Journal of Combinatorics. 14, pages 44–61, 2007.
- [15] Taoyang Wu, Peter Cameron, and Søren Riis. On the guessing number of shift graphs. Journal of Discrete Algorithms, 7(2):220–226, 2009.

Ross Atkins University of Oxford Department of Statistics 1 South Parks Road Oxford OX1 3TG United Kingdom e-mail: ross.atkins@univ.ox.ac.uk

Puck Rombach University of California, Los Angeles Department of Mathematics 520 Portola Plaza CA 90095-1555 United States e-mail: rombach@math.ucla.edu Fiona Skerman Heilbronn Institute University of Bristol Howard House, Queen's Ave Bristol BS8 1SD United Kingdom e-mail: f.skerman@bristol.ac.uk