Shannon capacity and related problems in Information Theory and Ramsey Theory

Eyal Lubetzky

Based on Joint work
with Noga Alon and Uri Stav

Outline of talk

Shannon Capacity of a graph:

- Introduction
- Some bounds and properties ↔ Ramsey Numbers
- The capacity of a disjoint union. ↔ Ramsey Constructions

Application: Index Coding:

- Introduction
- Constructing non-linear index-codes beating the linear optimum ↔ Ramsey Constructions
- Open problems

Shannon capacity - Introduction

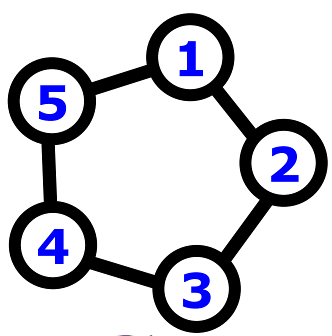
- Transmission over a noisy channel $\mathcal{C}$:
  - Input alphabet: $V = \{1, \ldots, n\}$
  - Output alphabet: $U = \{1, \ldots, m\}$
  - $\mathcal{C} : V \to P(U)$ maps each input letter to a set of possible output letters.

Goal ([Shannon '56]):

What is the maximal rate of zero-error transmission over a given noisy channel $\mathcal{C}$?

Single letter transmission over $\mathcal{C}$

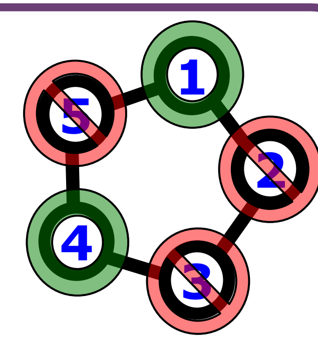
- Define the characteristic graph of a channel $\mathcal{C}$:
  - $G=(V,E)$ where $ij \in E \iff \mathcal{C}(i) \cap \mathcal{C}(j) \neq \emptyset$.
- The set $S \subset V$ guarantees zero error $\iff$ S is an independent set of G.
- OPT $= \alpha(G)$ for a single use of $\mathcal{C}$.



C_5

i	$\mathcal{C}(i)$
1	1,2
2	2,3
3	3,4
4	4,5
5	5,1

S :



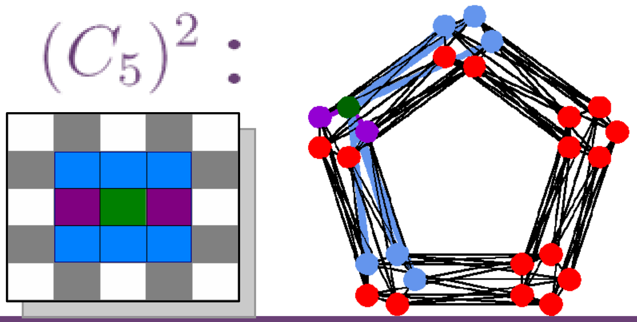
OPT = 2

Strong graph powers - definition

Q: Can we benefit from sending longer words over $\mathcal{C}$?

- Define G^k, the k^{th} strong graph power of G:
  - $V(G^k) = V(G)^k$
  - $(u_1, \ldots, u_k) \neq (v_1, \ldots, v_k)$ are adjacent $\iff$ for all i, either $u_i = v_i$ or $u_i v_i \in E(G)$.
- When G is the characteristic graph of $\mathcal{C}$, $uv \in E(G^k) \iff u$ and v are confusable in $\mathcal{C}$.

$(C_5)^2$:

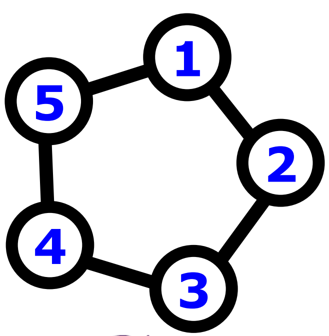


Strong graph powers - application

Q: Can we benefit from sending longer words over $\mathcal{C}$?

A: YES

- OPT= $\alpha(G^k)$ for sending k-letter words via $\mathcal{C}$.
- Block-coding shows $\alpha(G^k) \geq (\alpha(G))^k$.
- A strict inequality $\alpha(G^k) > (\alpha(G))^k$ is possible!

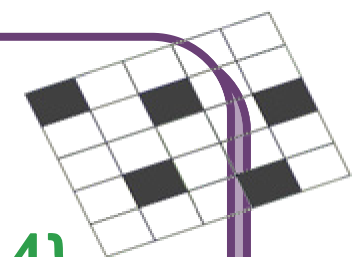


C_5

$\alpha(\mathrm{C}_5) = 2$ {1 , 4}

$\alpha((\mathrm{C}_5)^2) \geq 4$ {11, 14, 41, 44}

$\alpha((\mathrm{C}_5)^2) \geq 5$ {11, 23, 35, 42, 54}



Shannon capacity - definition

- The Shannon Capacity of G is defined to be:

$$c(G) = \lim_{k\to\infty} \left(\alpha(G^k)\right)^{1/k} = \sup_k \left(\alpha(G^k)\right)^{1/k}$$

$$\alpha(G^{k+l}) \geq \alpha(G^k)\alpha(G^l) \Longrightarrow \exists \lim = \sup$$

- $c(G)$ is the effective alphabet-size of $\mathcal{C}$ when sending zero-error transmission.
- E.g., if $c(G) = 7$, then for $k \gg 1$ we can send $\sim 7^k$ k-letter words via $\mathcal{C}$ without danger of confusion.

Shannon capacity: some bounds

- [Shannon '56]: $\alpha(G) \le c(G) \le \chi_f(\overline{G})$.
  Smallest graph unsettled by this was C_5.

  (Motivated [Berge '60] to study perfect graphs;
  WPGT proved by [Lovász '72], SPGT by [CRST '02].)
- [Haemers '78, '79]: algebraic upper bounds.
- [Lovász '79]: $c(G) \le \vartheta(G)$ (the Lovász ϑ func.), giving $c(C_5) = \sqrt{5}$.
- $c(G)$ remains unknown even for simple and small graphs, e.g. C_7 .

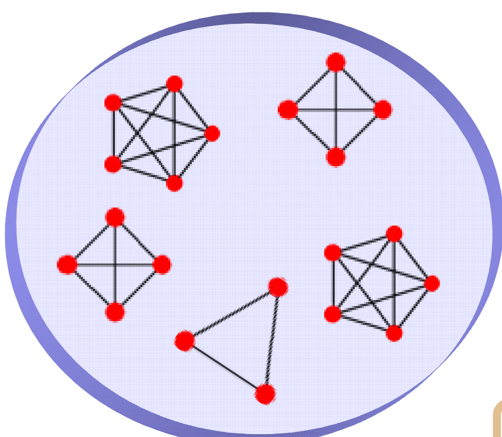
Shannon capacity: original bounds

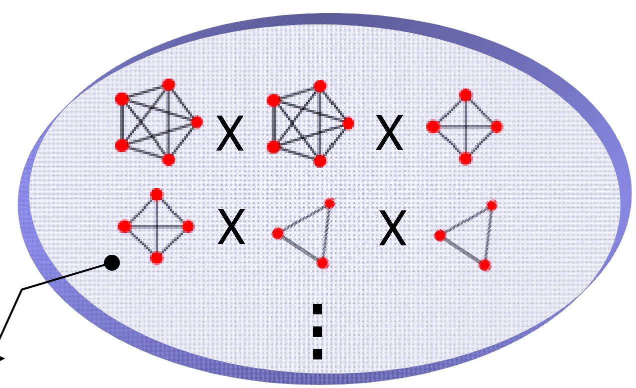
- [Shannon '56]: $\alpha(G) \le c(G) \le \chi_f(\overline{G})$.

By definition.

Similar to proving that $c(G) \le \chi(\overline{G})$:

If r cliques cover the vertices of G, then G^k can be covered by r^k cliques.





Cartesian product of cliques = clique

Shannon capacity: algebraic bound

- $A = (a_{ij}) \in M_n(\mathbb{F})$ represents G over $\mathbb{F}$ iff:
  - Diagonal entries are non-zero: $a_{ii} \neq 0 \quad \forall i \in [n]$
  - Off diagonal entries $a_{ij} = 0$ whenever $(i,j) \notin E$

$n \times n$ matrices over $\mathbb{F}$

- [Haemers '78, '79]:

If A represents G over $\mathbb{F}$, then $c(G) \leq \text{rank}_{\mathbb{F}}(A)$.

Proof:

- I independent set of $G \implies A[I:I] = \begin{pmatrix} \neq 0 & & 0 \\ & \ddots & \\ 0 & & \neq 0 \end{pmatrix}$ full rank

  $\implies \alpha(G) \leq \text{rank}_{\mathbb{F}}(A)$
- Higher powers: by definition, $A^{\otimes k}$ represents G^k:

  $\implies \alpha(G^k) \leq \text{rank}_{\mathbb{F}}(A^{\otimes k}) = (\text{rank}_{\mathbb{F}}(A))^k$ ■

Where is $c(G)$ attained?

- Shannon's χ bound gives examples of graphs where $c(G) = \alpha(G)$: 1-letter words are optimal.
- Lovász's ϑ function gives examples of graphs where $c(G) = \sqrt{\alpha(G^2)}$: 2-letter words are optimal.
- No known G with other finite optimal word-length.

Q: Can we approximate $c(G)$ by $\alpha(G), \ldots, \alpha(G^k)$ for some large finite k?

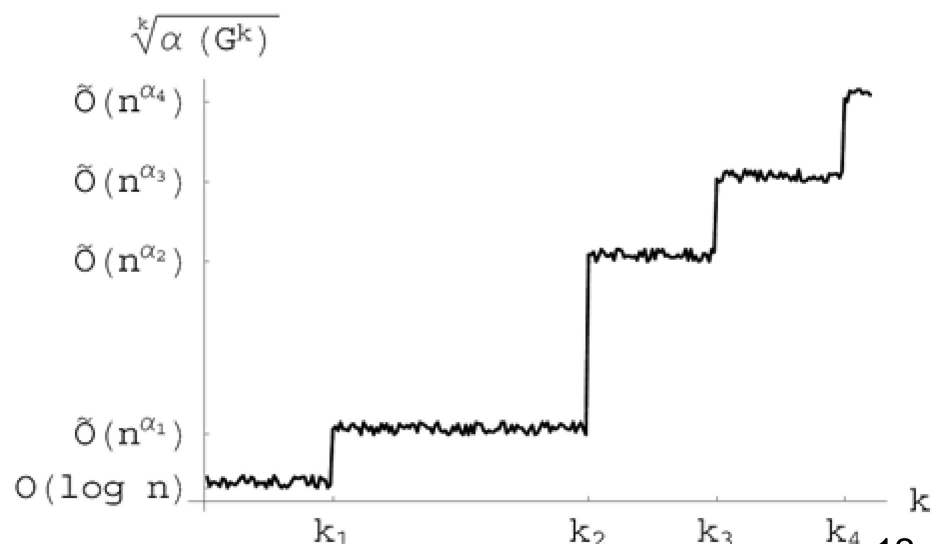
A: No, not even after we witness any finite number of improvements…

Rate increases between powers

- [Alon+L '06]: There can be any finite number of rate increases at any arbitrary locations:

> For every fixed $k_1 < k_2 < \ldots < k_s$ and $\varepsilon > 0$ there is a graph G so that for all j, $\max_{t<k_j} \alpha(G^t)^{\frac{1}{t}} \leq \left(\alpha(G^{k_j})^{\frac{1}{k_j}}\right)^{\varepsilon}$.

- Nevertheless, we can deduce some bound on $\alpha(G^{k+1})$ given $\alpha(G^k)$, using Ramsey Theory.



Shannon capacity and Ramsey No.

The Ramsey number $r(k,k)$ is the minimal integer r so that every 2-edge-coloring of the complete graph K_r has a monochromatic K_k.

- Suppose $\alpha(G)=5$. Then $\alpha(G^2)<165$ (!).
- [Erdős+McEliece+Taylor ‘71]: A tight bound of:

$$\text{If } \alpha(G)=k, \text{ then } \alpha(G^2)\le r(k+1,k+1)-1.$$

- Proof: color the edges of an independent set of G^2 according to the disconnected coordinate.

Ramsey Numbers

Sum of channels

- 2 senders combine separate channels, $\mathcal{C}_1$ and $\mathcal{C}_2$:
  - Each letter can be sent from either of the 2 channels.
  - Letters from $\mathcal{C}_1$ are never confused with those from $\mathcal{C}_2$.
- Characteristic graph is $G_1 + G_2$.

Disjoint union of individual char. graphs

- [Shannon ‘56]: $c(G+H) \geq c(G) + c(H)$,
  and conjectured that (=) always holds.

<u>Q:</u>

How can adding a separate channel $\mathcal{C}_2$ increase the capacity by more than $c(G_2)$?

$\mathcal{C}_1$: אבג...

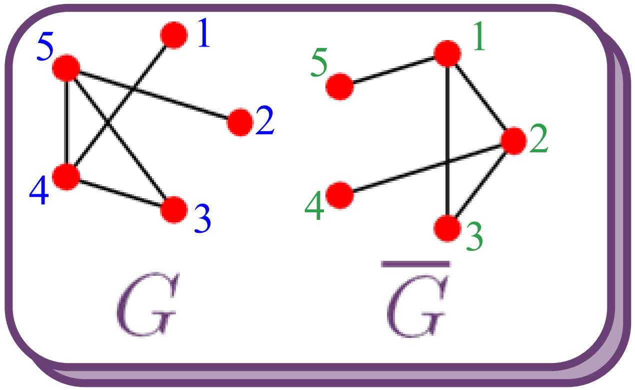
$\mathcal{C}_2$: $abc\ldots$

The Shannon capacity of a union

- [Alon '98] disproved Shannon's conjecture:

$$\exists\, G,H :\ c(G) \le k,\ c(H) \le k,\ c(G+H) \ge k^{\Omega\left(\frac{\log k}{\log\log k}\right)}$$

- Proof outline:
  - Suppose for some $G = ([n], E)$: $\begin{cases} c(G) \le n^{o(1)}, \\ c(\overline{G}) \le n^{o(1)}. \end{cases}$ $\alpha(G) \ll n$ $\omega(G) \ll n$
  - in the 2nd power
  - Ind. set $\{(i_G, i_{\overline{G}}) : i \in [n]\}$ implies $c(G+\overline{G}) \ge \sqrt{n}$.
- Such a G is a Ramsey graph!
- Proof applies an algebraic bound to a variant of the Ramsey construction by [Frankl+Wilson '81].



Multiple channels & privileged users

- [Alon+L '08]: The following stronger result holds:

For any fixed t and family $\mathcal{F} \subset 2^{[t]}$, $\exists\, G_1, \ldots, G_t$ so that: $\forall\, I \subset [t]$, $c(\sum_{i \in I} G_i)$ is "large" if I contains some $F \in \mathcal{F}$, and is "small" otherwise.

- E.g., $\mathcal{F} = \{F \subset [t] : |F| = k\}$ ensures that:
  - Any k senders combined have a high capacity.
  - Any group of $k-1$ senders has a low capacity.

Ramsey Theory revisited

- By-product: explicit construction for a Ramsey graph with respect to "rainbow" sub-graphs:

For any (large) n and $t \leq \sqrt{\frac{2\log n}{(\log\log n)^3}}$ there is an explicit t-edge-coloring of K_n, so that every induced subgraph on $\exp(O(\sqrt{\log n \log\log n}))$ vertices contains all t colors.

$n^{o(1)}$

Ramsey Constructions

Outline of talk - revisited

Shannon Capacity of a graph:

- Introduction
- Some bounds and properties ↔ Ramsey Numbers
- The capacity of a disjoint union. ↔ Ramsey Constructions

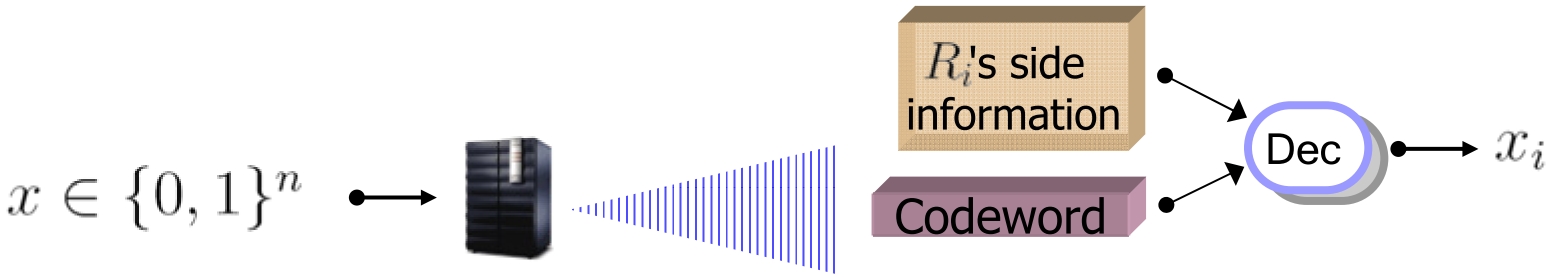
Application: Index Coding:

- Introduction
- Constructing non-linear index-codes beating the linear optimum ↔ Ramsey Constructions

- Open problems

Index Coding - Problem Definition

- [Birk+Kol '98],[Bar-Yossef+Birk+Jayram+Kol '06]:
  - Server broadcasts data to n receivers, $R_1, \ldots, R_n$.
  - Input data: $x \in \{0,1\}^n$.
  - Each R_i is interested in x_i, and knows some subset of the remaining bits.
  - Goal: design a code of minimal word length, so that: for every input word x, every R_i will be able to recover the bit x_i (using his side-information).

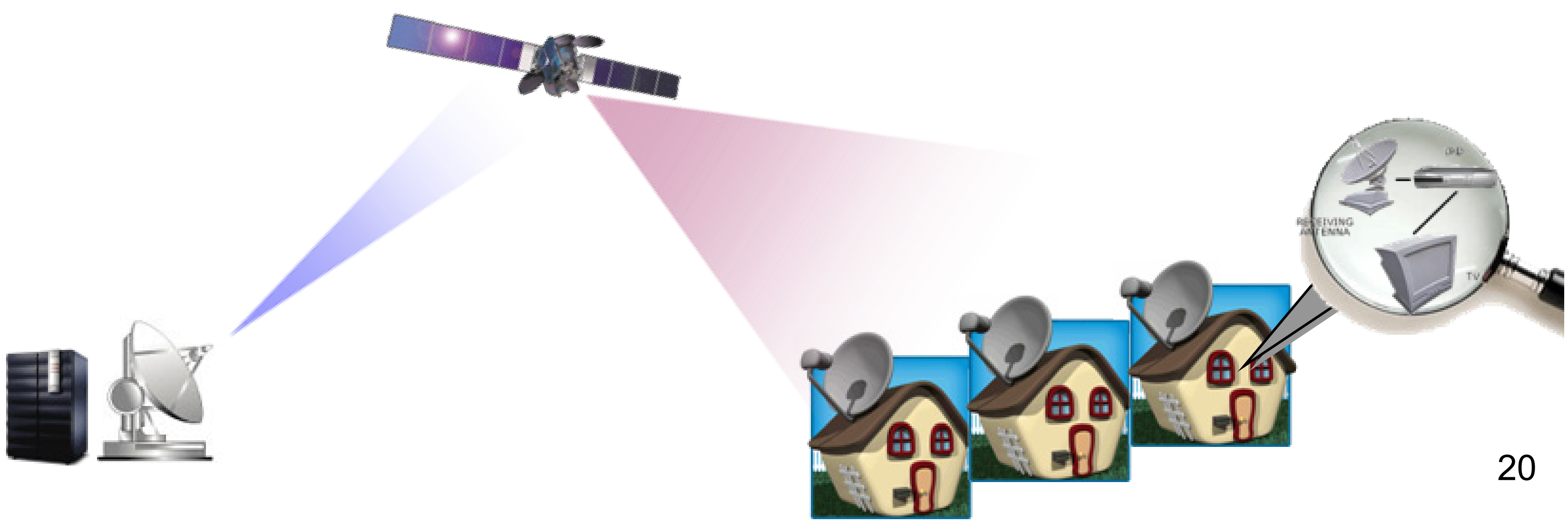


Motivation: Informed Source Coding

- Content broadcast to cashing clients:

Limited individual storage | Slow backward channel

- Clients inform server on known & required blocks.
- Goal: broadcast a short stream, allowing each client to recover its wanted data.



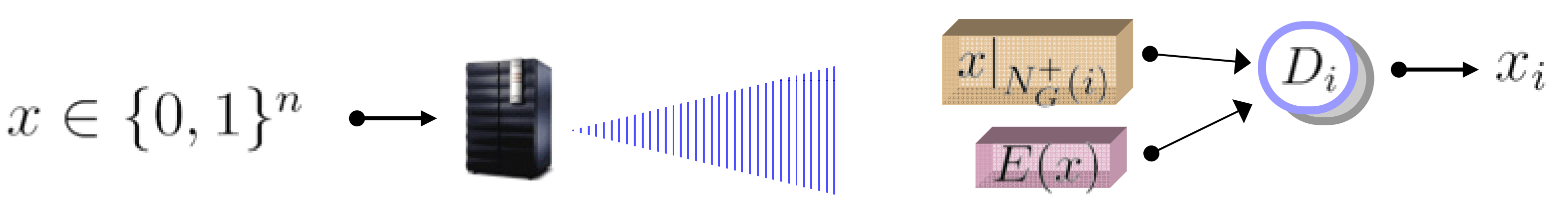
Index coding in terms of graphs

- Define the (directed) side-information graph:
  - Vertex set: $V = \{1, \ldots, n\}$.
  - (i, j) is an edge iff R_i knows the value of x_j .
- An *index code* of length ℓ for G is:
  - An encoding function: $E : \{0,1\}^n \to \{0,1\}^\ell$,
  - Decoding functions: $D_1, \ldots, D_n$,

so that $\forall i \in [n]$,$\forall x \in \{0,1\}^n$: $D_i(E(x), x|_{N^+_G(i)}) = x_i$.

Out-neighbors of R_i in G.

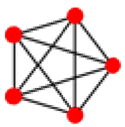
$\ell(G)$ = minimum length of an index code for G.



Index coding Examples

<u>Note:</u> For any graph G, $1 \leq \ell(G) \leq n$.

- Suppose every R_i knows all the bits except x_i :

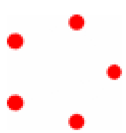


  - Side-information graph is the complete graph K_n.
  - A linear index code of length 1:

    $$E(x) = \bigoplus_{i=1}^{n} x_i \quad , \quad D_i(E(x), x|_{\{j:j\neq i\}}) = E(x) \oplus (\bigoplus_{j\neq i} x_j) = x_i \quad .$$

  - $\Longrightarrow \ell(G) = 1$.

- Similarly, if no R_i knows any of the bits:



  - Side-information graph is the edgeless graph.
  - Counting argument: code must contain 2^n distinct words, hence $\ell(G) = n$.

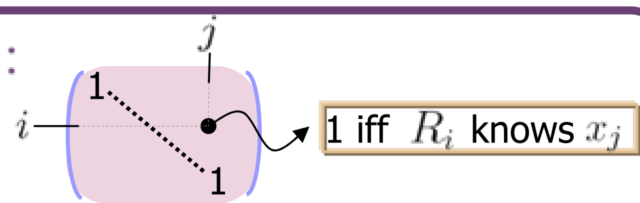
A linear index coding scheme

- Set: A_G : the adjacency matrix of G,
  $\{u_1, \ldots, u_r\}$: basis for $\mathrm{rows}(A_G + I)$ over $GF(2)$.
- Encoding: given $x \in \{0,1\}^n$, send $(u_1 \cdot x, \ldots, u_r \cdot x)$.
- Decoding: $$((A_G + I)x)_i = x_i + \sum_{j \in N_G^+(i)} x_j$$
  $\Longrightarrow R_i$ can reconstruct x_i.
- Altogether: $\ell(G) \leq \mathrm{rank}_2(A_G + I)$

Allows recovering $(A_G + I)x$

R_i knows these bits by definition.

$A_G + I$:



Optimal linear index codes

<u>Note:</u> For any spanning sub-graph $H \subset G$, $\ell(G) \leq \ell(H)$.

- $\Longrightarrow \ell(G) \leq \min_{H \subset G} \text{rank}_2(A_H + I) \;=: \text{minrk}_2(G)$
- [BBJK '06] showed:
  - $\text{minrk}_2(G)$ is the size of the optimal linear index code.
  - In many cases $\ell(G) = \text{minrk}_2(G)$.

    e.g., perfect graphs, acyclic graphs, holes, anti-holes,…
- The main conjecture of [BBJK '06]:

<u>Conj:</u> Linear index coding is **always** optimal, i.e., $\ell(G) = \text{minrk}_2(G)$ for any G.

Beating the linear optimum

- [L+Stav]: the conjecture of [BBJK '06] is false in, essentially, the strongest possible way:

For any $\varepsilon > 0$ and (large) n, $\exists\, G$ on n vertices so that:

1. Any linear index code for G requires $n^{1-\varepsilon}$ bits.
2. There exists a non-linear index code for G using n^{ε} bits.

Moreover, G is undirected and can be constructed explicitly.

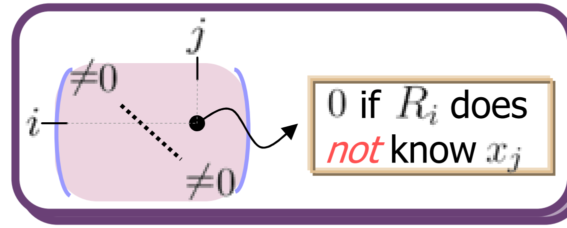
$\mathrm{minrk}_2(G) \geq n^{1-\varepsilon}$
(hardly improves trivial protocol of sending the entire word x)

$\ell(G) \leq n^{\varepsilon}$

Index coding - proof sketch

WANTED G such that $\mathrm{minrk}_2(G)$ is "large", and $\ell(G)$ is "small".

- Need $\ell(G)$ to be small regardless of $\mathrm{minrk}_2(G)$ …
- Use higher order fields:
  - Take $A = (a_{ij})$ representing G over $\mathbb{F}$:



  - Encode Ax using $\lceil \mathrm{rank}_{\mathbb{F}}(A) \log_2 |\mathbb{F}| \rceil$ bits.
  - Decoding: $a_{ii}^{-1}(Ax)_i = x_i + a_{ii}^{-1} \sum_{j \in N_G^+(i)} a_{ij} x_j$
- Generalizing $\mathrm{minrk}_2(G) \longrightarrow \mathrm{minrk}_{\mathbb{F}}(G)$, we have $\ell(G) \leq \lceil \mathrm{minrk}_{\mathbb{F}}(G) \log_2 |\mathbb{F}| \rceil$.

Index coding - proof sketch

WANTED: G such that $\mathrm{minrk}_2(G)$ is "large", and $\mathrm{minrk}_p(G)$ is "small".

$p \neq 2$

- Difficult to provide lower bounds on $\mathrm{minrk}_2(G)$…
- Use the Shannon capacity:
  - $\mathrm{minrk}_2(G) \geq c(G)$
  - $c(G \times \overline{G}) \geq n$
  - $\mathrm{minrk}_2(G \times \overline{G}) \leq \mathrm{minrk}_2(G)\,\mathrm{minrk}_2(\overline{G})$
- It follows that for every G on n vertices,

$\mathrm{minrk}_2(G)\,\mathrm{minrk}_2(\overline{G}) \geq n$.

$\mathrm{rank}_2(A_H + I)$ must be "large" for $\forall H \subset G$

$\{(i_G, i_{\overline{G}}) : i \in [n]\}$ is an independent set

Showing that $\mathrm{minrk}_2(\overline{G}) \leq n^{\varepsilon}$ will imply that $\mathrm{minrk}_2(G) \geq n^{1-\varepsilon}$

Index coding - proof sketch

WANTED G such that $\mathrm{minrk}_2(\overline{G})$ is "small", and $\mathrm{minrk}_p(G)$ is "small".

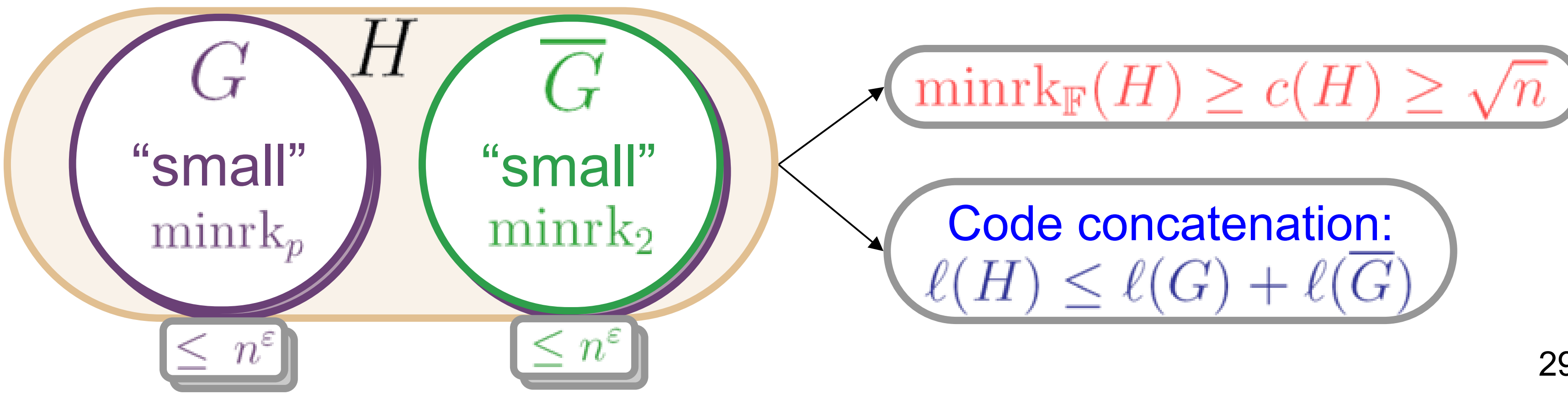
$p \neq 2$

- Such a G is a Ramsey graph.
- The construction of [Alon '98]: for some large primes $p \neq q$.
  - "small" $\mathrm{minrk}_p(G)$
  - "small" $\mathrm{minrk}_q(\overline{G})$
- Use Lucas' Theorem to extend this construction to any distinct primes.
- Choosing $q = 2$ completes the proof. ■

Beating linear codes over any field

- We constructed graphs where $\ell(G) \ll \mathrm{minrk}_2(G)$ using linear codes over higher-order fields.
- Q: Can $\ell(G)$ beat any linear index coding scheme, i.e., $\forall \mathbb{F}\ \ell(G) \ll \mathrm{minrk}_{\mathbb{F}}(G)$?


  A: YES (a corollary of the previous Thm).
- Take $H = G + \overline{G}$ for the previous G :



Multiple round index coding

- $t \geq 1$ rounds (each with its own input & setting):

  R_i is interested in the i th bit of each word.

  $G_1, \ldots, G_t$: side information graphs

- $\ell(G_1, \ldots, G_t)$: minimal length of such an index code.
- Multiple usage can improve the average rate!
- Example:
  - 1st usage: R_i knows the bits $\{x_j : j > i\}$
  - 2nd usage: R_i knows the bits $\{y_j : j < i\}$
  - In this case, $\ell(G_1) = \ell(G_2) = n$,

    yet $\ell(G_1, G_2) = n + 1$, largest possible gap!

G_1, G_2 are transitive tournaments

Some open problems

- Multiple round index coding:
  Recall that $\ell(G_1, G_2) \leq \ell(G_1) + \ell(G_2)$.
  How does $\lim_{k \to \infty} \ell(\underbrace{G, \ldots, G}_{k \text{ times}})/k$ behave?

Inequality may be strict

- What is the expected value of $\ell(\mathcal{G}_{n,\frac{1}{2}})$?
- Can $\text{minrk}_2(G)$ be exponentially larger than $\ell(G)$?
- Generalized setting: multiple receivers may be interested in the same bit.

Thank you.