Locally Testable Codes and ℓ_1 −Embeddings of Cayley Graphs

Parikshit Gopalan (MSR-SVC)
with
Salil Vadhan (Harvard), Yuan Zhou (CMU).

Locally Testable Codes

Local Tester for an $[n,k,d]_2$ linear code $\mathcal{C}$:

- Queries few co-ordinates.
- Accepts codewords.
- Rejects words far from the code with high probability.

[BenSasson-Harsha-Raskhodnikova]: A local tester is a distribution $\mathcal{D}$ on (low-weight) dual codewords.

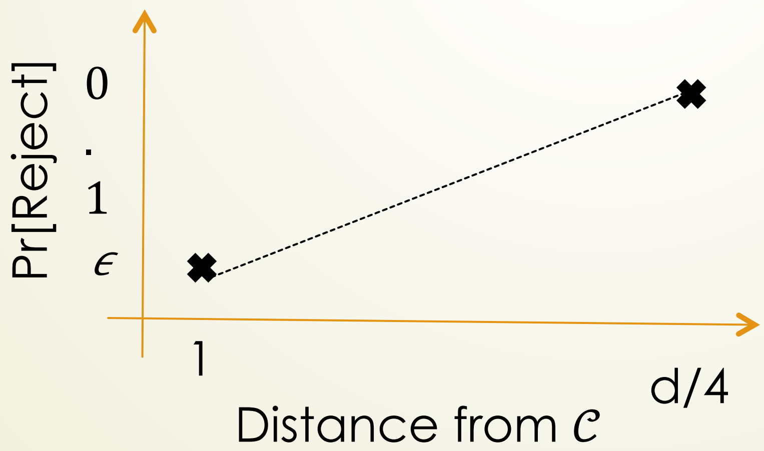
Locally Testable Codes

[Blum-Luby-Rubinfeld'90, Rubinfeld-Sudan'92, Freidl-Sudan'95]

Randomized Tester for an $[n,k,d]_2$ code:

- Queries coordinates according to $\mathcal{D}$ on $\mathcal{C}^{\perp}$.
- ϵ-smooth: queries each coordinate w.p. $\leq \epsilon$.
- Rejects words at distance d' with prob $\delta d'$.

Must have $\delta \leq \epsilon$, would like $\delta = \Omega(\epsilon)$.



The Price of Locality?

Asymptotically good regime: $r = \Omega(1), \delta = \Omega(1)$.

Are there asymptotically good 3-query LTCs?

- Existential question! [Goldreich-Sudan'02]
- LTCs with 3 queries, $n = k(\log k)^{c}$, $d = \Omega(n)$. [Dinur'05, …,Viderman'13]

Rate 1 regime: Let d be a (large) constant and $n \to \infty$.

How large can k be for an $[n,k,d]_2$ LTC?

- Fix smoothness $\epsilon = \Theta(1/d)$.

The Price of Locality?

Asymptotically good regime: $r=\Omega(1)$, $\delta=\Omega(1)$.

Are there asymptotically good 3-query LTCs?

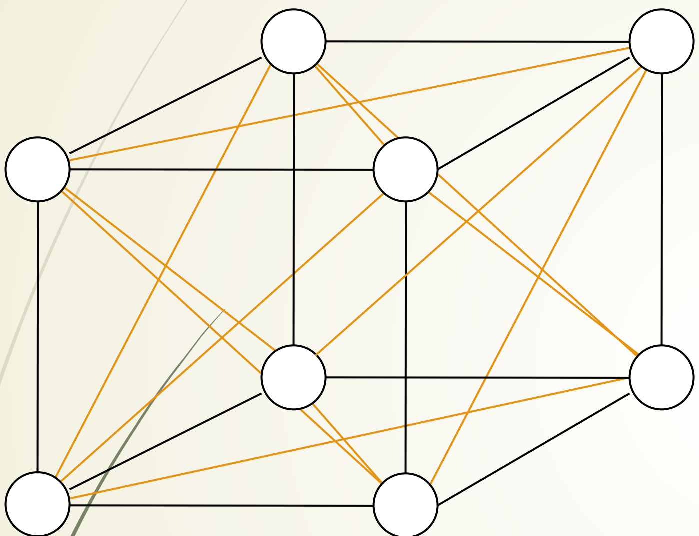
- Existential question! [Goldreich-Sudan'02]
- LTCs with 3 queries, $n=k(\log k)^{c}$, $d=\Omega(n)$. [Dinur'05, …,Viderman'13]

Rate 1 regime: Let d be a (large) constant and $n\to\infty$.

How large can k be for an $[n,k,d]_2$ LTC?

- Fix smoothness $\epsilon=\Theta(1/d)$.
- BCH gives $n-k=d/2\log(n)$. But not locally testable.
- [BKSSZ'08]:$[n, n-(\log n)^{\log(d)}, d]$ LTC from Reed-Muller.
- Can we have $n-k=O_d(\log(n))$?

Cayley Graphs on $\mathbb{F}_2^h$



Graph $\mathcal{G}(\mathbb{F}_2^h, S)$

$S = \{s_1, \dots, s_n\} \subseteq \mathbb{F}_2^h$

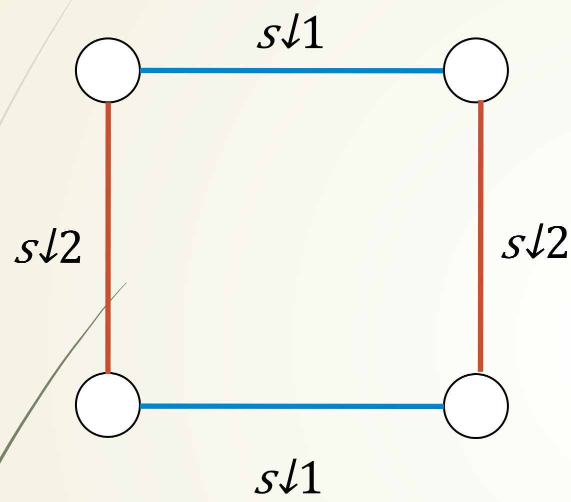
Vertices: $\mathbb{F}_2^h$

Edges: $\{(x, x+s_i) : x \in \mathbb{F}_2^h,\ i \in [n]\}$.

Hypercube: $S = (e_1, \dots, e_h)$ so $h = n$.
We are interested in $n > h$.

Def: S is d-wise independent if every $T \subseteq S$ where $|T| < d$ is linearly independent.

Cayley Graphs on $\mathbb{F}_2^h$



Graph $\mathcal{G}(\mathbb{F}_2^h, S)$

$S=\{s_1,\dots,s_n\}\subseteq\mathbb{F}_2^h$ is d-wise independent.

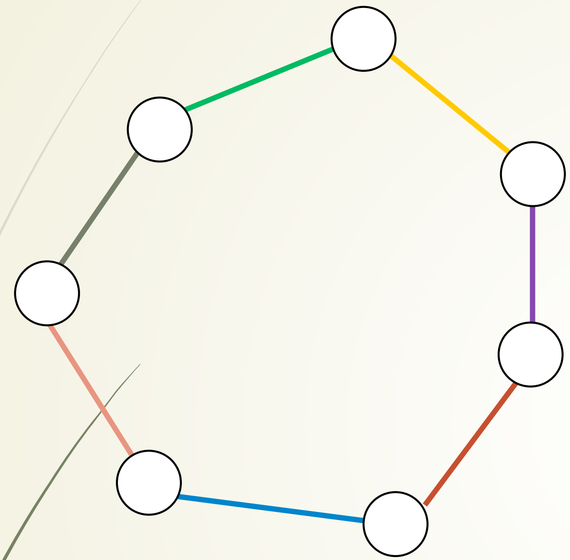
Vertices: $\mathbb{F}_2^h$

Edges: $\{(x, x+s_i) : x\in\mathbb{F}_2^h,\ i\in[n]\}$.

d-wise independence: Abelian analogue of large girth.

- Cycles occur when edge labels sum to 0.
- $\mathcal{G}(\mathbb{F}_2^h, S)$ will have 4 cycles.

Cayley Graphs on $\mathbb{F}_2^h$



Graph $\mathcal{G}(\mathbb{F}_2^h, S)$

$S=\{s_1, \ldots, s_n\} \subseteq \mathbb{F}_2^h$ is d-wise independent.

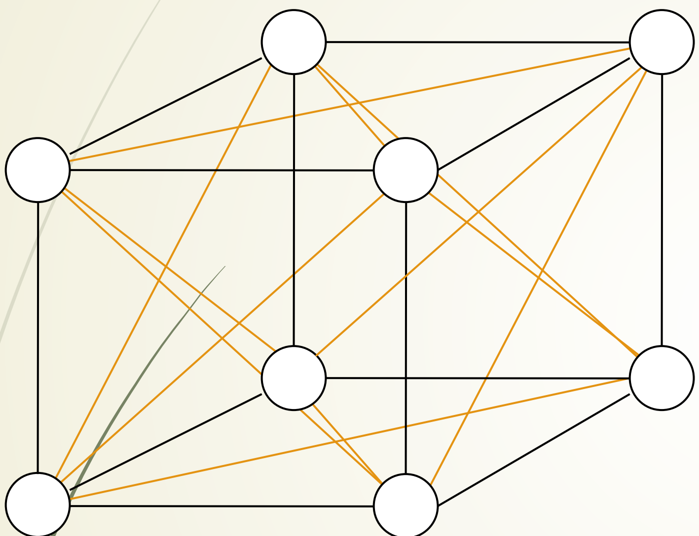
Vertices: $\mathbb{F}_2^h$

Edges: $\{(x, x+s_i) : x \in \mathbb{F}_2^h,\ i \in [n]\}$.

d-wise independence: Abelian analogue of large girth.

- Cycles occur when edge labels sum to 0.
- $\mathcal{G}(\mathbb{F}_2^h, S)$ will have 4 cycles.
- Non-trivial cycles have length at least d.

Cayley Graphs on $\mathbb{F}_2^h$



Graph $\mathcal{G}(\mathbb{F}_2^h, S)$

$S=\{s_1,\ldots,s_n\} \subseteq \mathbb{F}_2^h$ is d-wise independent.

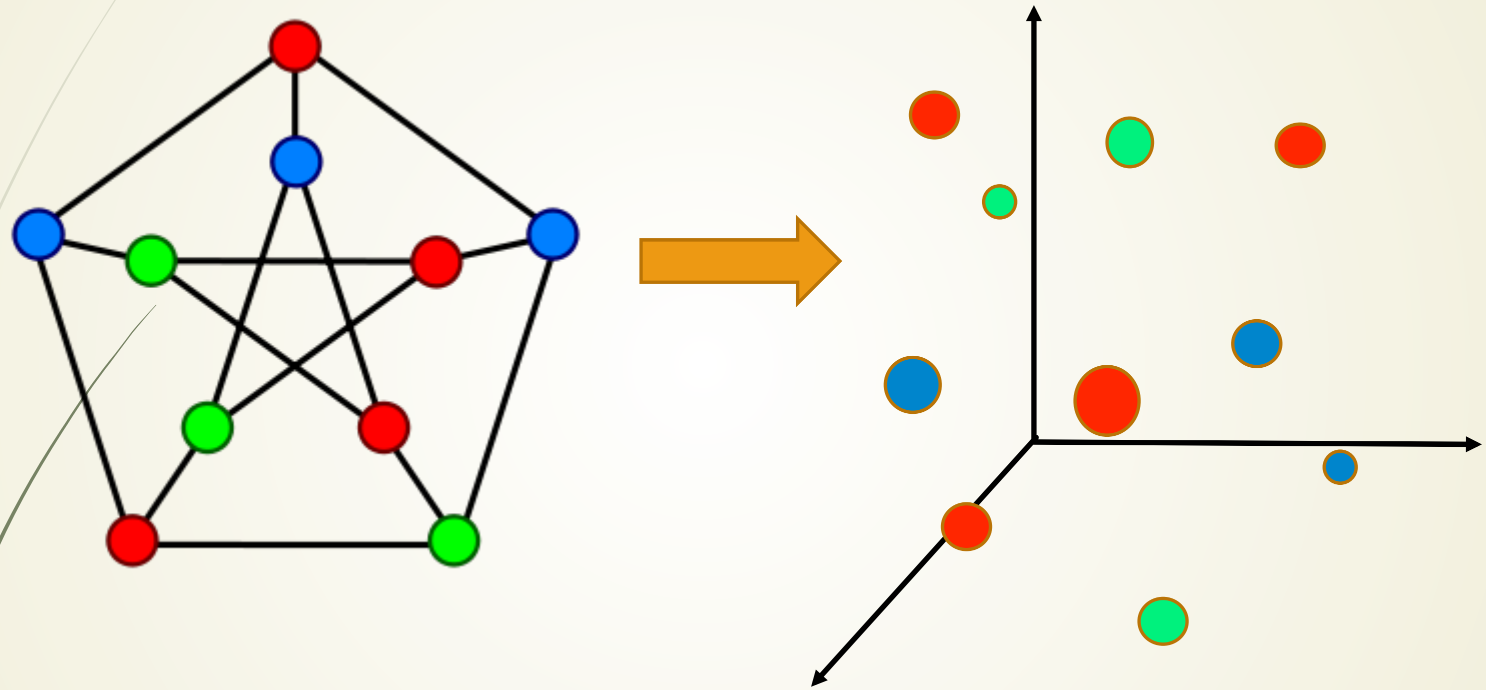
Vertices: $\mathbb{F}_2^h$

Edges: $\{(x, x+s_i) : x \in \mathbb{F}_2^h,\ i \in [n]\}$.

d-wise independence: Abelian analogue of large girth.

- Cycles occur when edge labels sum to 0.
- $\mathcal{G}(\mathbb{F}_2^h, S)$ will have 4 cycles.
- Non-trivial cycles have length at least d.
- $d/2$-neighborhood of any vertex is isomorphic to $B(n, d/2)$, but the vertex set has dimension $h \ll n$.

ℓ_1 – Embeddings of graphs



Embedding $f: V(\mathcal{G}) \to \mathbb{R}^d$ has distortion c if
$|f(x) - f(y)|_1 \leq d_{\mathcal{G}}(x,y) \leq c|f(x) - f(y)|_1$

$c_1(\mathcal{G})$= minimum distortion over all embeddings.

The Equivalence

Main Theorem: The following are equivalent:

- An $[n,k,d]_2$ code $\mathcal{C}$ with a tester of smoothness ϵ and soundness δ.
- A Cayley graph $\mathcal{G}(\mathbb{F}_2^{n-k},S)$ where $|S|=n$, S is d-wise independent with an embedding of distortion ϵ/δ.

[Khot-Naor'06]: Codes with large dual distance give Cayley graphs where $c_1(\mathcal{G})=\omega(1)$.

The Equivalence

Main Theorem: The following are equivalent:

- An $[n,k,d]_2$ code $\mathcal{C}$ with a tester of smoothness ϵ and soundness δ.
- A Cayley graph $\mathcal{G}(\mathbb{F}_2^{n-k},S)$ where $|S|=n$, S is d-wise independent with an embedding of distortion ϵ/δ.

Corollary: There exist asymptotically good strong LTCs iff there exist Cayley graphs $\mathcal{G}(\mathbb{F}_2^{h},S)$ where

- $|S|=(1+\Omega(1))h$,
- S is $\Omega(h)$-wise independent,
- $c_1(\mathcal{G})=O(1)$.

The Equivalence

Main Theorem: The following are equivalent:

- An $[n,k,d]_2$ code $\mathcal{C}$ with a tester of smoothness ϵ and soundness δ.
- A Cayley graph $\mathcal{G}(\mathbb{F}_2^{n-k},S)$ where $|S|=n$, S is d-wise independent with an embedding of distortion ϵ/δ.

Corollary: There exist $[n,n-O_d(\log(n)),d]_2$ strong LTCs iff there exist Cayley graphs $\mathcal{G}(\mathbb{F}_2^h,S)$ where

- $|S|=2^{\Omega_d(h)}$,
- S is d-wise independent,
- $c_1(\mathcal{G})=O(1)$.

The Equivalence

Main Theorem: The following are equivalent:

- An $[n,k,d]_2$ code $\mathcal{C}$ with a tester of smoothness ϵ and soundness δ.
- A Cayley graph $\mathcal{G}(\mathbb{F}_2^{n-k},S)$ where $|S|=n$, S is d-wise independent with an embedding of distortion ϵ/δ.

Proof Sketch:

- Codes from Graphs (and vice versa).
- Testers from Embeddings (and vice versa).

Some Applications.

Codes and Cayley Graphs

Graph $\mathcal{G}(\mathbb{F}_2^h, S)$: $S=\{s_1, \ldots, s_n\} \subseteq \mathbb{F}_2^h$ is d-wise independent.

$[n, n-h, d]_2$ Code $\mathcal{C}$: $h \times n$ Parity check matrix: $[s_1, \ldots, s_n]$

Codewords: $x \in \mathbb{F}_2^n$ such that $\sum_i x_i s_i = 0$.

What does the shortest path metric in $\mathcal{G}$ correspond to?

The (quotiented) Hamming metric on $\mathbb{F}_2^n / \mathcal{C}$.

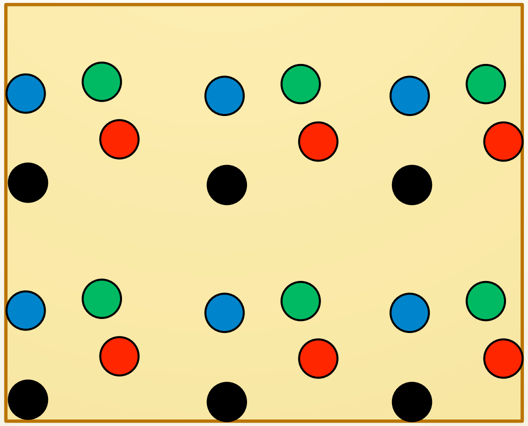
The Quotiented Hamming metric

Let $x = \{x + \mathcal{C}\}$ be a coset of $\mathcal{C}$.

Let $wt(x) = \min_{c \in \mathcal{C}} wt(x+c)$ and $d(x, y) = wt(x + y)$.

View cosets as received words grouped by error vector.

$wt(x)$ is the number of errors.



Codes and Cayley Graphs

Graph $\mathcal{G}(\mathbb{F}_2^h, S)$: $S = \{s_1, \ldots, s_n\} \subseteq \mathbb{F}_2^h$ is d-wise independent.

$[n, n-h, d]_2$ code $\mathcal{C}: x \in \mathbb{F}_2^n$ such that $\sum_i x_i s_i = 0$.

Shortest path metric ≡ Quotiented Hamming metric.

- Each vertex in $\mathcal{G}$ corresponds to a coset of $\mathcal{C}$.

  Start at 0, take a walk according to $x \in \{0,1\}^n$.

  For i∈$[n]$, if $x_i = 1$, take the edge labelled s_i.

  Reach the vertex $\sum_i x_i s_i \in \mathbb{F}_2^h$.

  Set of x leading to any vertex is a coset of $\mathcal{C}$.

Codes and Cayley Graphs

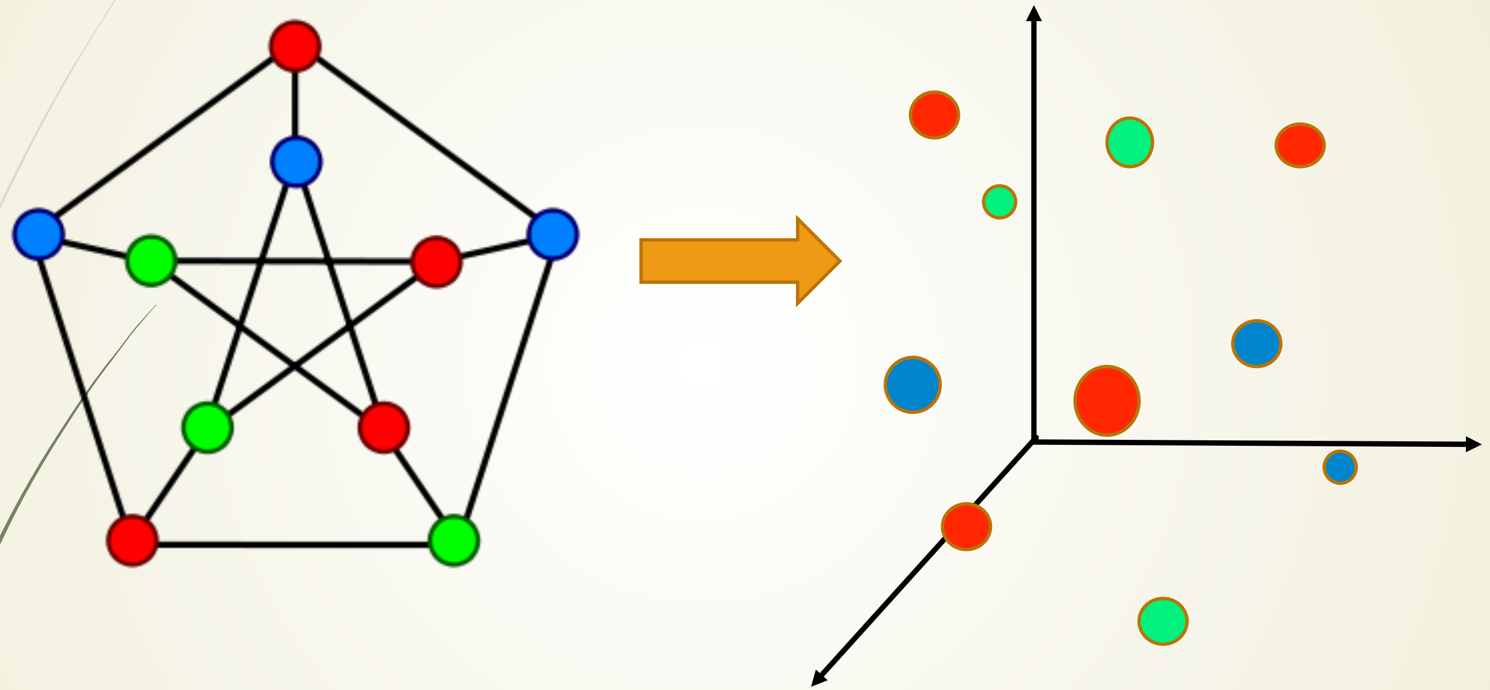
Graph $\mathcal{G}(\mathbb{F}_2^h, S)$: $S=\{s_1,\dots,s_n\}\subseteq\mathbb{F}_2^h$ is d-wise independent.

$[n, n-h, d]_2$ code $\mathcal{C}$: $x\in\mathbb{F}_2^n$ such that $\sum_i x_i s_i = 0$.

Shortest path metric ≡ Quotiented Hamming metric.

- Each vertex in $\mathcal{G}$ corresponds to a coset of $\mathcal{C}$.
- Shortest path to x corresponds to smallest weight $x\in x$.

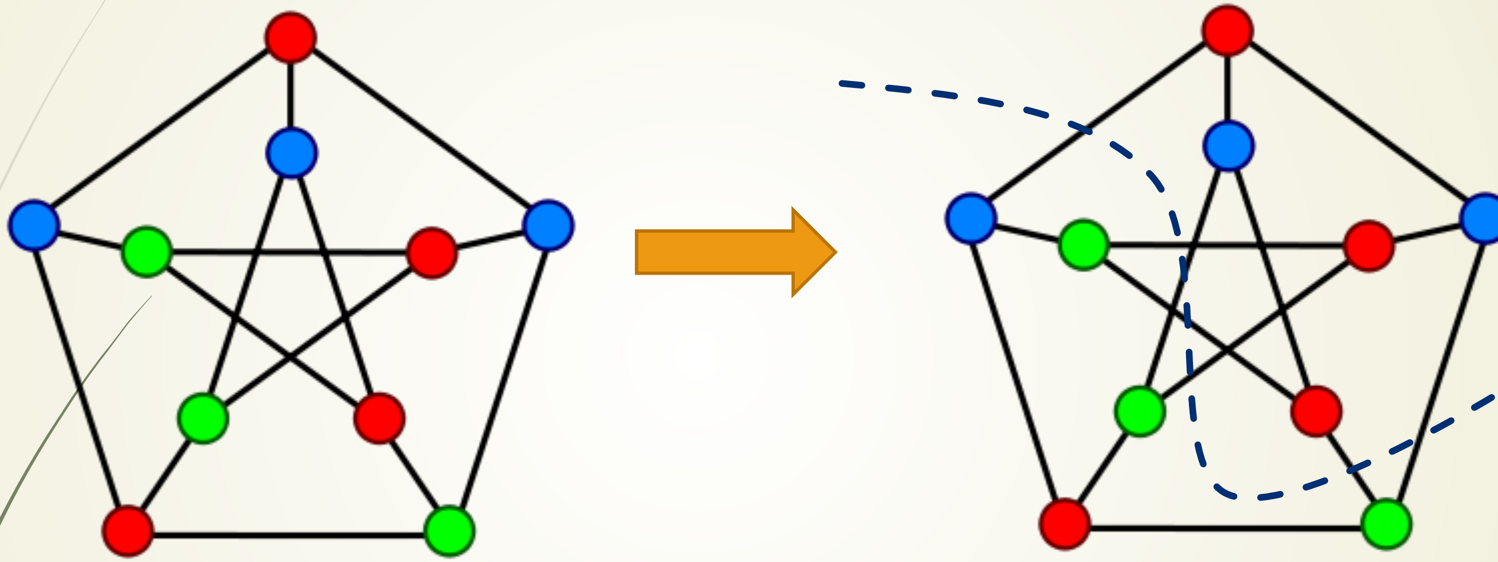
ℓ_1 –Embeddings of graphs



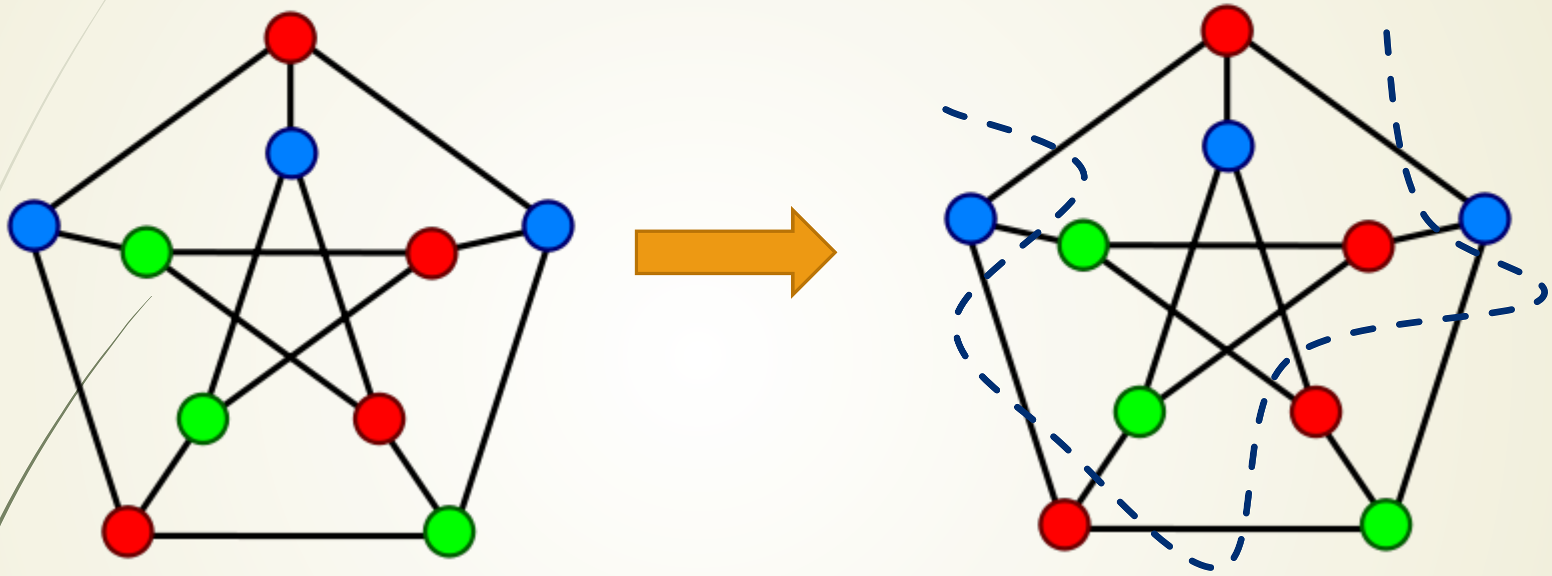
Embedding $f\colon \mathcal{G} \to \mathbb{R}^d$ has distortion c if
$|f(x)-f(y)|_1 \le d_{\mathcal{G}}(x,y) \le c|f(x)-f(y)|_1$

$c_1(\mathcal{G})$= minimum distortion over all embeddings.

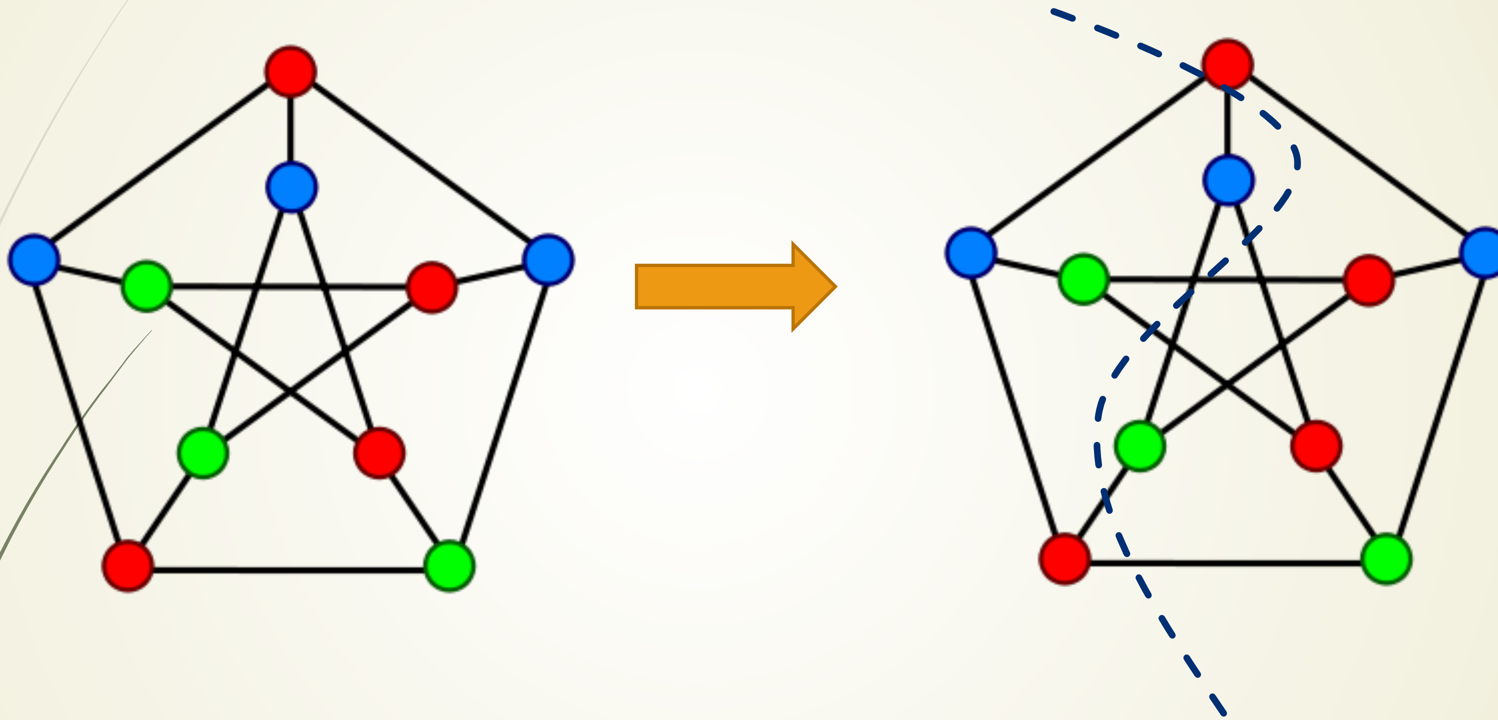
Cut-cone Characterization of ℓ_1



Cut-cone Characterization of $\ell\downarrow 1$



Cut-cone Characterization of ℓ_1



Distribution $\mathcal{D}$ on cuts $f: V(G) \to \{-1,1\}$.
$\rho(x,y) = \Pr_{f\sim\mathcal{D}}[f(x) \neq f(y)]$.
Embedding $\mathcal{D}$ has distortion c if there exists $\alpha \in \mathbb{R}$ such that
$\alpha d_G(x,y) \leq \rho(x,y) \leq c \cdot \alpha d_G(x,y)$

Embeddings from Testers

Given tester $\mathcal{D}$ distribution on $\mathcal{C}^{\perp}$.

Each $\alpha \in \mathcal{C}^{\perp}$ defines a cut on $V(\mathcal{G}) = \mathbb{F}_2^n / \mathcal{C}$.

Claim: The embedding $\mathcal{D}$ has distortion ϵ/δ.

Proof: Suffices to consider $(x, 0)$ by linearity.

$$d_{\mathcal{G}}(x, 0) = wt(x).$$

$$\delta \cdot wt(x) \leq \rho(x, 0) = \Pr[\mathcal{D} \text{ rejects } x] \qquad \leq \epsilon \cdot wt(x)$$

Testers from Embeddings

Distribution $\mathcal{D}$ on $f: \mathbb{F}_2^n/\mathcal{C} \to \{-1,1\}$ giving distortion c.

If $\mathcal{D}$ was supported on linear functions, we'd be (essentially) done.

Claim: There is a distribution $\mathcal{D}'$ on linear functions with distortion c.

Proof Outline:

- Extend f to all of $\mathbb{F}_2^n$.
- Its Fourier expansion is supported on $\mathcal{C}^\perp$:

$$f(x)=\sum_{\alpha\in\mathcal{C}^\perp} \hat{f}(\alpha)\chi_\alpha(x).$$

- If $\mathcal{D}$ samples f, $\mathcal{D}'$ samples α with probability $|\hat{f}(\alpha)|^2$.

Why does this work?

$\Pr_{\mathcal{F}ar}[f(x)\neq f(y)] / \Pr_{\mathcal{N}ear}[f(x)\neq f(y)]$

Distributions $\mathcal{F}ar, \mathcal{N}ear$ on $V\times V$.

$f: V\to\{-1,1\}$

Distributions of the form $(\mathcal{U}, \mathcal{U}+\mathcal{A})$.

$\chi_\alpha : \mathbb{F}_2^n/\mathcal{C} \to \{-1,1\}$

$$\mathbb{E}_{x\in\mathcal{U},\, a\in\mathcal{A}}[f(x,x+a)] = \sum_{\alpha\in\mathcal{C}^\perp} \hat{f}(\alpha)^2\, \mathbb{E}_{a\in\mathcal{A}}[\chi_\alpha(a)]$$

Applications …

[Khot-Naor'06]: If $\mathcal{C}^{\perp}$ is asymptotically good, then $c_1(\mathcal{G})=\Omega(n)$.

Proof: Suffices to lower bound ϵ/δ.

- Since $d^{\perp}=\Omega(n)$, $\epsilon=\Omega(1)$.
- Let t be the covering radius of $\mathcal{C}$. Then $\delta\leq 1/t$.

  We have $t=\Omega(n)$, since $\mathcal{C}^{\perp}$ has rate $\Omega(n)$.
- So $\delta=O(1/n)$ and $\epsilon/\delta=\Omega(n)$.

Analogue of [BenSasson-Harsha-Raskhodnikova'03]:
Small dual distance necessary for Local testing.

[BHR'03]: Codes where $d^{\perp}=O(1)$, but not locally testable.

A spectral view of LTCs

[G-Vadhan-Zhou]: $[n,k,d]_2$ LTCs are equivalent to Cayley graphs on $\mathbb{F}_2^{n-k}$ whose eigenvalue spectrum resembles the n-dimensional ϵ-noisy hypercube for $\epsilon=1/d$.

Gives a converse to a result of Barak-G.-Hastad-Meka-Raghavendra-Steurer'2012.

Conclusions

- Many known connections between codes and graphs:
  - Relate pseudorandom objects.
  - This work relates objects whose existence is unclear!
- Can it be used for better constructions?
- Or better lower bounds?