Storage Capacity as an Information-Theoretic Analogue of Vertex Cover

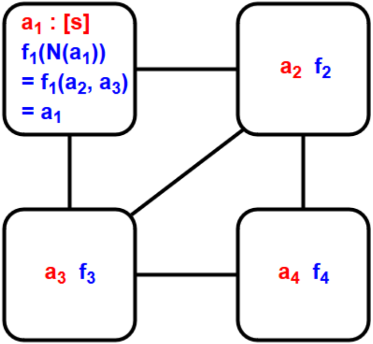
Sofya Vorotnikova

University of Massachusetts Amherst

Joint work with Andrew McGregor and Arya Mazumdar

Problem Description

Motivated by applications in distributed storage. Related to index coding, locally repairable codes, and generalized guessing games.



Input: graph.
Objective: come up with recovery functions that allow us to restore info at any vertex from info at its neighbors.
Question: How much information can we store in a given graph?

Results

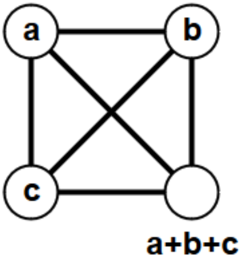
Lower Bounds

- Fractional clique partition:
  - 3/2-approximation of Cap in planar graphs
  - 4/3-approximation in triangle-free planar graphs

Upper Bounds

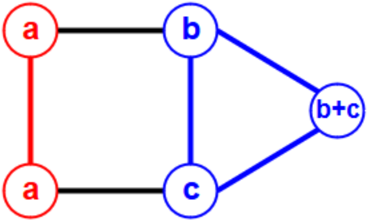
- Information-theoretic LP
- Cover by gadgets
- Vertex partition

Clique



Store any symbols at $k-1$ vertices of a k-clique.
Store their sum (mod s) at the last vertex.
$\text{Cap}(C_k) = k-1$

Lower Bound: Clique Partition

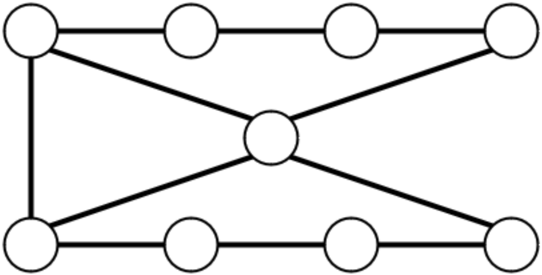


Define a variable $x_C \in \{0, 1\}$ for every clique C.

$$\text{maximize} \sum_C x_C(|C| - 1)$$

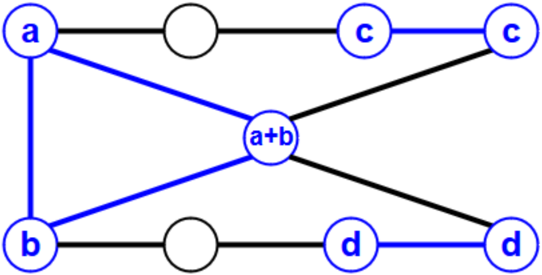
$$\text{s.t.} \sum_{C:v \in C} x_C = 1 \quad \forall v \in V$$

Lower Bound: Fractional Clique Partition



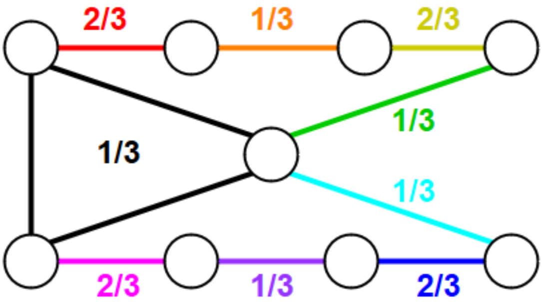
Can relax the previous LP to obtain a fractional solution.

Fractional Clique Partition



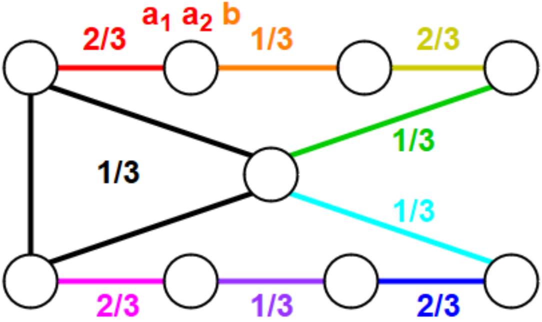
Clique partition: $\mathrm{Cap}(G) \geq 4$

Fractional Clique Partition



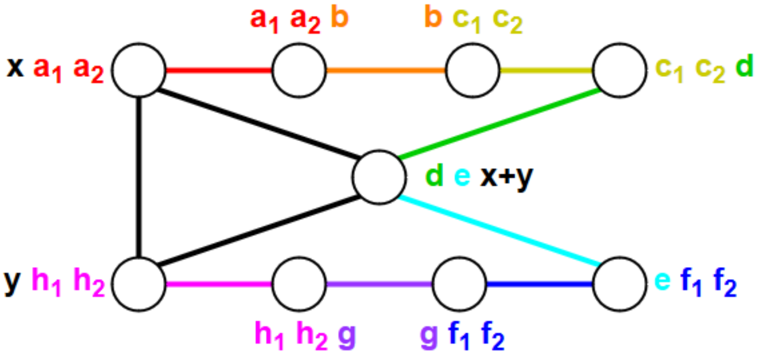
Fractional clique partition.

Fractional Clique Partition



Store a vector at each vertex = symbol of larger alphabet

Fractional Clique Partition



Store a vector at each vertex = symbol of larger alphabet
Ex: alphabet of size 8
$\mathrm{Cap}(G) \geq \log_8(2^{14}) = 14/3$

Fractional Clique Partition: Limitations

Vertex cover is an upper bound on Cap [Mazumdar 2015].

Since fractional clique partition in some graphs is the same as maximum matching, in general, this approach gives **2-approx**.

We concentrate on particular families of graphs, where approximation factor can be improved.

Fractional Clique Partition in Planar Graphs

Optimal fractional clique partition gives

- $3/2$-approx in **planar** graphs
- $4/3$-approx in **triangle-free planar** graphs.

Fractional Clique Partition in Planar Graphs

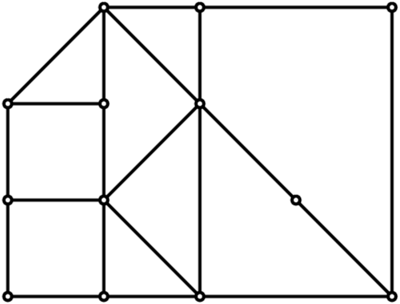
Optimal fractional clique partition gives

- 3/2-approx in **planar** graphs
- 4/3-approx in **triangle-free planar** graphs.

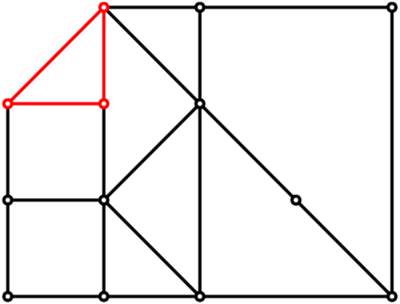
Proof idea: construct a fractional CP (not necessarily optimal) in the following manner:

- Remove triangles one by one in a greedy manner
- In the resultant triangle-free planar graph find optimal fractional matching

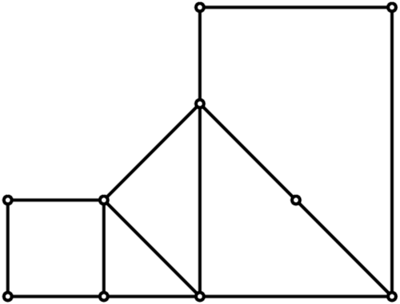
Fractional Clique Partition in Planar Graphs



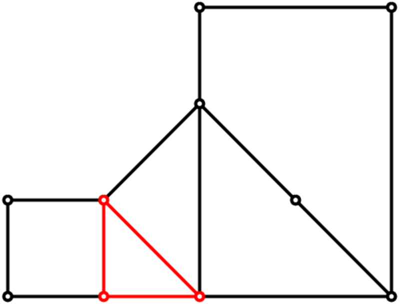
Fractional Clique Partition in Planar Graphs



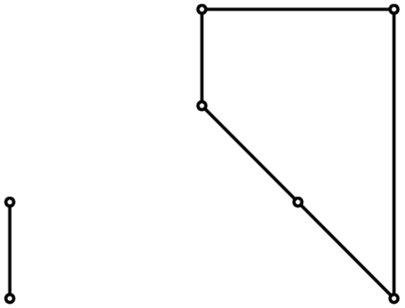
Fractional Clique Partition in Planar Graphs



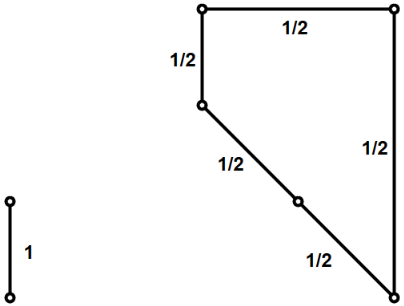
Fractional Clique Partition in Planar Graphs



Fractional Clique Partition in Planar Graphs



Fractional Clique Partition in Planar Graphs



Fractional Clique Partition in Planar Graphs

Triangle elimination
In each triangle store 2 units of information. Note, that at most 3 units are possible.

Fractional matching in planar graph
opt fractional matching $=$ opt fractional VC $\geq 2/3$ opt integral VC

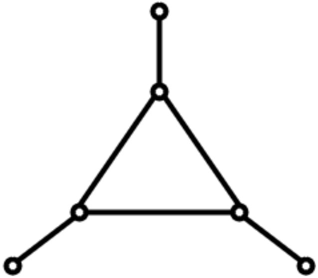
Fractional matching in triangle-free planar graph
opt fractional matching $=$ opt fractional VC $\geq 3/4$ opt integral VC

Thus, we get

- $3/2$-approx for planar graphs with triangles
- $4/3$-approx for triangle-free planar graphs

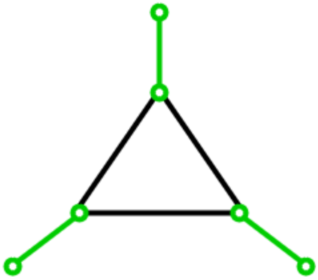
Fractional Clique Partition in Planar Graphs

Note, that this analysis is not optimal.



Fractional Clique Partition in Planar Graphs

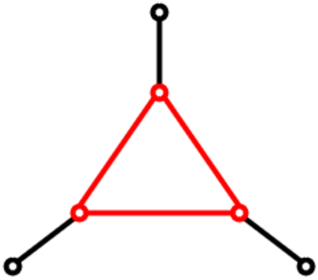
Note, that this analysis is not optimal.



$\text{Cap}(G) = 3$

Fractional Clique Partition in Planar Graphs

Note, that this analysis is not optimal.



Our analysis only shows $\text{Cap}(G) \geq 2$

Index Coding

Suppose, the stored symbols are arbitrary and no vertex knows which symbol it is storing. We want to broadcast one message to all vertices, such that each vertex can compute its symbol from the info at its neighbors and the broadcast.
Objective: minimize the length of the broadcast message.

Mazumdar 2015:

$$\mathrm{Cap}(G) = n - \mathrm{Ind}(G)$$

Using fraction clique partition we show

- 2-approx in planar graphs
- 3/2-approx in triangle-free planar graphs

Information Theoretic Formulation of Storage Capacity

$$\begin{aligned} \text{maximize} \quad & H(X_1, X_2, \ldots, X_n) \\ \text{s.t.} \quad & X_1, X_2, \ldots, X_n \text{ are random variables} \\ & H(X_i | N_G(X_i)) = 0 \quad \forall i \end{aligned}$$

Optimal solution to this optimization problem is the storage capacity of graph G.

Upper Bound: Information Theoretic LP

By relaxing the constraints of the optimization problem, we obtain the following LP (adapted from the index coding LP due to Blasiak, Kleinberg, and Lubetzky).

Define a variable z_S for every $S \subseteq V$.
Let $\mathrm{cl}(S) = S \cup \{v : N(v) \subseteq S\}$ denote the *closure* of the set S consisting of vertices in S and vertices with all neighbors in S.

$$\begin{aligned} \text{maximize} \quad & z_V \\ \text{s.t.} \quad & z_\emptyset = 0 \\ & z_T - z_S \leq |T \setminus \mathrm{cl}(S)| \quad \forall S \subseteq T \\ & z_S + z_T \geq z_{S \cap T} + z_{S \cup T} \quad \forall S, T \end{aligned}$$

Optimal solution to the above LP is an upper bound on $\mathrm{Cap}(G)$.

Example: Odd Cycle

We don't always need to solve the LP.
Consider the following subset of constraints:

$$
\begin{aligned}
2 &\geq z_{\{1,3\}} \\
2 &\geq z_{\{2,4\}} \\
1 &\geq z_{\{i\}} \quad \forall i \in \{5, 6, \ldots, n\} \\
0 &\geq z_{\{1,2,3\}} - z_{\{1,3\}} \\
0 &\geq z_{\{2,3,4\}} - z_{\{2,4\}} \\
z_{\{1,2,3\}} + z_{\{2,3,4\}} &\geq z_{\{2,3\}} + z_{\{1,2,3,4\}} \\
z_{\{2,3\}} + z_{\{5\}} + z_{\{7\}} + \cdots + z_{\{n\}} &\geq z_{\{2,3,5,7,\ldots,n\}} \\
z_{\{1,2,3,4\}} + z_{\{6\}} + z_{\{8\}} + \ldots z_{\{n-1\}} &\geq z_{\{1,2,3,4,6,8,\ldots,n-1\}} \\
0 &\geq z_V - z_{\{2,3,5,7,\ldots,n\}} \\
0 &\geq z_V - z_{\{1,2,3,4,6,8,\ldots,n-1\}}
\end{aligned}
$$

Constraints sum up to $n \geq 2z_V$ and thus $\mathrm{Cap}(C_n) \leq n/2$.

Upper Bound: Gadget Cover

Gadget cover lets us come up with such sets of constraints in a more systematic way.

A **gadget** $g(A, B)$ is created the following way:

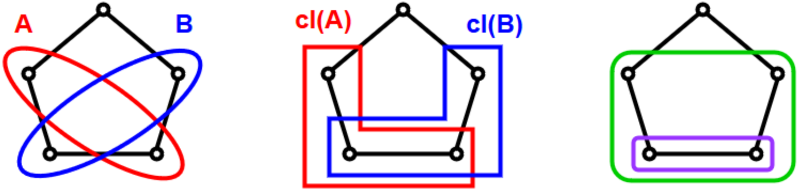
- take two sets of vertices A and B
- take their closures $\text{cl}(A)$ and $\text{cl}(B)$
- find $S = \text{cl}(A) \cup \text{cl}(B)$ and $T = \text{cl}(A) \cap \text{cl}(B)$

Then S and T form a gadget with weight $|A| + |B|$.

Color every gadget set with one of k colors, such that for every color c, vertices colored c form a vertex cover.
Then the total weight of gadgets is an upper bound on $k\,\text{Cap}(G)$.

Gadget Cover: Proof Idea

Bound follows from information theoretic LP. Note which constraints correspond to the steps of forming a gadget.



$$z_A \leq |A| \qquad z_{\text{cl}(A)} - z_A \leq 0 \qquad z_S + z_T$$

$$z_B \leq |B| \qquad z_{\text{cl}(B)} - z_B \leq 0 \qquad \leq z_{\text{cl}(A)} + z_{\text{cl}(B)}$$

If we sum these constraints, we obtain $z_S + z_T \leq |A| + |B|$.

Gadget Cover: Proof Idea

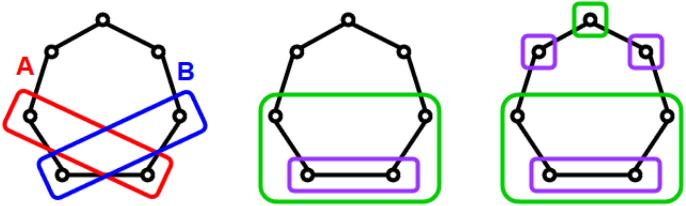
Find all constraints corresponding to building a gadget cover and coloring it k colors. If we sum these constraints, we obtain:

$$k\,\mathrm{Cap} \leq kz_V \leq |A_1| + |B_1| + |A_2| + |B_2| + \dots$$

where $|A_1| + |B_1| + |A_2| + |B_2| + \dots$ is the total weight of gadgets in the cover.

Example: Odd Cycle

Note: gadget $g(\{v\}, \emptyset)$ has one set that is empty.



Total weight is n. Since we used 2 colors,

$$\mathrm{Cap}(C_n) \leq n/2$$

This gadget cover corresponds to the set of constraints of Information Theoretic LP that provides the same bound.

Advantages of Gadget Cover

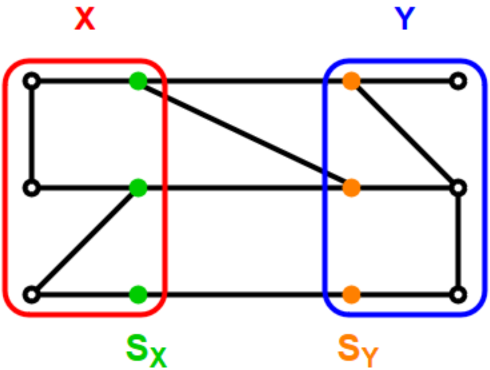
Info Theoretic LP	**Gadget Cover LP**
maximization	minimization
only optimal solution provides a bound	any feasible solution provides a bound
requires exponential time to solve	can consider small gadgets only and obtain a feasible solution in polynomial time

Upper Bound: Vertex Partition

If the vertices of G can be partitioned into sets X and Y s.t.:

- $G[X]$ and $G[Y]$ are both bipartite
- vertices in X with neighbors in Y form an independent set S_X
- vertices in Y with neighbors in X form an independent set S_Y

Then $\mathrm{Cap}(G) \leq n/2$.

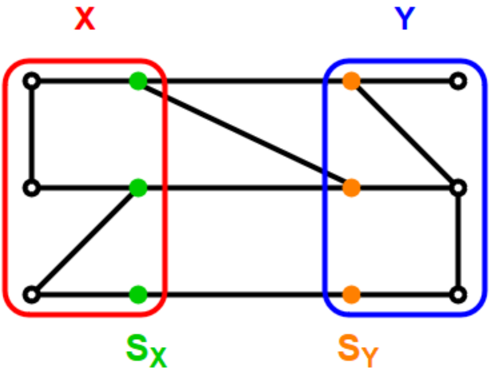


Upper Bound: Vertex Partition

If the vertices of G can be partitioned into sets X and Y s.t.:

- $G[X]$ and $G[Y]$ are both bipartite
- vertices in X with neighbors in Y form an independent set S_X
- vertices in Y with neighbors in X form an independent set S_Y

Then $\text{Cap}(G) \leq n/2$.



Idea: show that G has a gadget cover with 2 colors and weight n.

Other Results

- **Graph Products**: Cartesian product of a cycle of length at least 4 and a bipartite graph has Cap $= n/2$.
- **Cycles With Chords:** A cycle with chords, whose endpoints are at least distance 4 apart on the cycle, has Cap $= n/2$.
- **Partial Recovery:** Suppose every vertex stores a vector of length m. If up to any $\delta \in [0,1]$ fraction of the m coordinates are erased, they can be recovered by using the remaining content of the vertex and the contents in its neighbors. Find maximum total amount of information that can be stored in the graph.

Full version on arXiv: `https://arxiv.org/abs/1706.09197`

Summary

Upper and lower bounds on storage capacity and their applications in special families of graphs.

Lower bound via fractional clique partition.

Upper bounds via

- information theoretic LP
- gadget cover
- vertex partition

Summary

Upper and lower bounds on storage capacity and their applications in special families of graphs.

Lower bound via fractional clique partition.

Upper bounds via

- information theoretic LP
- gadget cover
- vertex partition

Thank you for your attention!