Storage Capacity as an Information-Theoretic Analogue of Vertex Cover[*]

Arya Mazumdar Andrew McGregor Sofya Vorotnikova

Abstract

Motivated by applications in distributed storage, the storage capacity of a graph was recently defined to be the maximum amount of information that can be stored across the vertices of a graph such that the information at any vertex can be recovered from the information stored at the neighboring vertices. Computing the storage capacity is a fundamental problem in network coding and is related, or equivalent, to some well-studied problems such as index coding with side information and generalized guessing games. In this paper, we consider storage capacity as a natural information-theoretic analogue of the minimum vertex cover of a graph. Indeed, while it was known that storage capacity is upper bounded by minimum vertex cover, we show that by treating it as such we can get a $3/2$ approximation for planar graphs, and a $4/3$ approximation for triangle-free planar graphs. Since the storage capacity is intimately related to the index coding rate, we get a 2 approximation of index coding rate for planar graphs[1] and $3/2$ approximation for triangle-free planar graphs. Previously only a trivial 4 approximation of the index coding rate was known for planar graphs. We then develop a general method of "gadget covering" to upper bound the storage capacity in terms of the average of a set of vertex covers. This method is intuitive and leads to the exact characterization of storage capacity for various families of graphs. As an illustrative example, we use this approach to derive the *exact* storage capacity of cycles-with-chords, a family of graphs related to outerplanar graphs. Finally, we generalize the storage capacity notion to include recovery from partial failures in distributed storage. We show tight upper and lower bounds on this partial recovery capacity that scales nicely with the fraction of failure in a vertex.

1 Introduction

The Shannon capacity of a graph [18] is a well studied parameter that quantifies the zero-error capacity of a noisy communication channel. There are also several other notions of graph capacities or graph entropies that model different communication/compression scenarios (for example, see [1]). In this paper, we are interested in a very recent definition of graph capacity, called the *storage capacity*, that we consider to be a natural information-theoretic analogue of the minimum vertex cover of a graph.

[*] University of Massachusetts, Amherst. {arya,mcgregor,svorotni}@cs.umass.edu. This work was supported by NSF Awards CCF-0953754, IIS-1251110, CCF-1320719, CCF-1642658, CCF-BSF-1618512, and a Google Research Award.

[1] A shorter version of this paper in the proceedings of the IEEE International Symposium on Information Theory, 2017 contains an error. The approximation factor for index coding rate for planar graphs was wrongly claimed to be 1.923. The correct approximation factor of our method is 2, and we have corrected Theorem 3 in this version.

Suppose, every vertex of a graph can store a symbol (from any alphabet) with the criterion that the content of any vertex can be uniquely recovered from the contents of its neighborhood in the graph. Then the maximum amount of information that can be stored in the graph is called the storage capacity of that graph [17]. This formulation is mainly motivated by applications in distributed storage, and generalizes the popular definition of *locally repairable codes* [13]. In a distributed storage system, each symbol (or coordinate) of a codeword vector is stored at a different server or storage vertex. In the case of a single server failure, it is desirable to be able to recover the data of that server by accessing a small number of other servers. Given the topology of the storage network as a graph, it is quite natural to model the local repair problem as a *neighborhood repair* problem as above.

Formally, suppose we are given an n-vertex graph $G(V,E)$, where $V=[n]\equiv\{1,2,\dots,n\}$. Also, given a positive integer $q\geq 2$, let $H(X)$ be the Shannon entropy of the random variable X in q-ary units (for example, when $q=2$, the entropy is in bits). Let $\{X_i\}_{i\in V}$, be random variables each with a finite sample space of size q. For any $I\subseteq[n]$, let $X_I\equiv\{X_i : i\in I\}$. Consider the solution of the following optimization problem:

$$\max H(X_1,\dots,X_n) \tag{1}$$

such that

$$H(X_i|X_{N(i)})=0,$$

for all $i\in V$ where $N(i)=\{j\in V:(i,j)\in E\}$ is the set of neighbors of vertex i. This is the storage capacity of the graph G and we denote it by $\mathrm{Cap}_q(G)$. Note that, although we hide the unit of entropy in the notation $H(\cdot)$, the unit should be clear from context, and the storage capacity should depend on it, as reflected in the subscript in the notation $\mathrm{Cap}_q(G)$. The absolute storage capacity is defined to be,

$$\mathrm{Cap}(G)=\lim_{q\to\infty}\mathrm{Cap}_q(G). \tag{2}$$

In [17], it was observed that the storage capacity is upper bounded by the size of the minimum vertex cover $\mathrm{VC}(G)$ of the graph G.

$$\mathrm{Cap}(G)\leq|\,\mathrm{VC}(G)|. \tag{3}$$

The proof of this fact is quite simple. Since all the neighbors of $V\setminus\mathrm{VC}$ belong to VC,

$$H(X_V)=H(X_{\mathrm{VC}},X_{V\setminus\mathrm{VC}(G)})=H(X_{\mathrm{VC}(G)})+H(X_{V\setminus\mathrm{VC}(G)}|H(X_{\mathrm{VC}(G)}))=H(X_{\mathrm{VC}(G)})\leq|\,\mathrm{VC}(G)|.$$

Indeed, this proof shows that $H(X_V)=H(X_{\mathrm{VC}(G)})$. Because of this, we think it is natural to view storage capacity as an *information theoretic analogue of vertex cover*. It was also shown in [17] that the storage capacity is at least equal to the size $\mathrm{MM}(G)$ of the maximum matching of the graph G:

$$\mathrm{MM}(G)\leq\mathrm{Cap}_2(G)\leq\mathrm{Cap}(G). \tag{4}$$

Since maximum matching and minimum vertex cover are two quantities within a factor of two of each other and maximum matching can be found in polynomial time, this fact gives a 2-approximation of the storage capacity[2]. Improvement of this approximation factor is unlikely to be achieved by simple means, since that would imply a better-than-2 approximation ratio for the minimum vertex cover problem violating the unique games conjecture [15].

[2] Indeed, finding a *maximal matching* is sufficient for this purpose.

This motivates us to look for natural families of graphs where minimum vertex cover has a better approximation. For example, for bipartite graphs maximum matching is equal to minimum vertex cover and hence storage capacity is exactly equal to the minimum vertex cover. Other obvious class, and our focus in Section 3, is the family of planar graphs for which a PTAS is known [4, 5]. Another motivation for studying storage capacity on planar graphs is that they represent common network topologies for distributed systems. For example, see [10] to note how a surprising number of data networks are actually planar. To minimize interference, it is natural for a distributed storage system to be arranged as a planar network. Moreover it is useful to have wireless networks, video-on-demand networks etc. that are planar or almost planar.

Video-on-demand also motivates a related broadcast problem called *index coding* [6] for which planar topologies are of interest, and outerplanar topologies have already been studied [7]. It was shown in [17] that storage capacity is, in a coding-theoretic sense, dual to index coding and is equivalent to the *guessing game* problem of [12]. The index coding rate for a graph G is defined to be the optimum value of the following minimization problem:

$$\min H(Y) \tag{5}$$

where Y is such that

$$H(X_i|Y, X_{N(i)}) = 0,$$

for all $i \in V$. This is called the optimum index coding rate for the graph G, and we denote it as $\mathrm{Ind}_q(G)$. We can also define,

$$\mathrm{Ind}(G) = \lim_{q \to \infty} \mathrm{Ind}_q(G). \tag{6}$$

The index coding problem is the hardest of all network coding problems and has been the subject of much recent attention, see, e.g., [16]. In particular it can be shown that any network coding problem can be reduced to an index coding problem [11]. It has been shown [17] that,

$$\mathrm{Cap}(G) = n - \mathrm{Ind}(G). \tag{7}$$

and hence exact computation of $\mathrm{Ind}(G)$ and $\mathrm{Cap}(G)$ is equivalent although the approximation hardness could obviously differ. Note that,

$$\mathrm{Ind}(G) \geq \alpha(G),$$

where $\alpha(G)$ is the independence number of G. Since, for planar graphs $\alpha(G) \geq n/4$, taking Y to be $X_{[n]}$ already gives a 4-approximation for index coding rate for planar graphs (since $H(Y) \leq n$). In this paper, we give a significantly better approximation algorithm for index coding rate of planar graphs. Not only that, due to the relation between index coding rate and storage capacity, we can obtain an approximation factor significantly better than 2 for storage capacity. Note that, even to approximate the optimal index coding rate within a factor of $n^{1-\epsilon}$ seems to be a challenge [9].

To go beyond the realm of planar graphs, and to obtain better approximation ratios, we then develop several upper bounding tools for storage capacities. In particular by using these tools, we are able to exactly characterize storage capacities of various families of graphs. Our approach revisits a linear program proposed by Blasiak, Kleinberg, and Lubetzky [8] that can be used to lower bound the optimum index coding rate or upper bound the storage capacity. We transform the problem of bounding this LP into the problem of constructing a family of vertex covers for the input graph. This in turn allows us to upper bound the storage capacity of any graph that admits a specific type of vertex partition. We then identify various graphs for which this upper bound is tight.

Since, the storage capacity, or the vertex cover, act as absolute upper bounds on the rate of information storage in the graph, a natural question to ask is, if we store above the limit of minimum vertex cover in the graph, will any of the repair property be left? This is similar in philosophy to the rate-distortion theory of data compression, where one compresses beyond entropy limit and still can recover the data with some distortion. This question gives rise to the notion of recovery from partial failure, as defined below.

We define the *partial repair capacity* also keeping the application of distributed storage in mind. This is a direct generalization in the context of distributed storage application to handle partial failure of vertices. In particular, suppose we lose $0 \leq \delta \leq 1$ proportion of the bits stored in a vertex. We still want to recover these bits by accessing the remaining $(1-\delta)$-fraction of the bits in the vertex plus the contents of the neighborhood. What is the maximum amount of information that can be stored in the network with such restriction? Intuitively, the storage capacity should increase. We characterize the trade-off between δ and this increase in storage capacity from both sides (i.e., upper and lower bounds on the capacity). A surprising fact that we observe is that, if we want to recover from more than half of the bits being lost, then there is no increase in storage capacity.

In summary, we made progress on the study of storage capacity on three fronts:

- *Planar graphs.* We prove a $3/2$ approximation of storage capacity and 2 approximation for index coding rate for planar graphs. We provide an approximation guarantee that depends on the number of triangle in the graph and, in the special case of for triangle-free graphs, we get a $4/3$ approximation for storage capacity, and $3/2$ approximation for index coding rate.
- *Tools for finding storage capacity upper bounds.* We develop an approach for bounding storage capacity in terms of a small number of vertex covers. We first illustrate this approach by finding the exact storage capacity of some simple graphs. We then use the approach to show a bound on any graph that admits a specific type of vertex partition. With this we prove exact bounds on a family of Cartesian product graphs and a family closely related to outerplanar graphs.
- *Partial failure recovery.* We show that if recovery from neighbors is possible for up to δ-proportion failure of the bits stored in a server, then the capacity is upper bounded by the optimum value of a linear program; in particular this implies when $\delta \geq \frac{1}{2}$, then the partial recovery capacity is same as the storage capacity. For an odd cycle, the upper bound on partial recovery capacity is given by $\frac{n}{2}(1+R_2(\delta))$, where $R_2(\delta)$ is the maximum achievable rate of a binary error-correcting code with relative minimum Hamming distance at least δn. On the other hand, we also obtain general lower bounds on the partial recovery capacity of a graph. For an odd cycle, our results imply that a partial failure recovery capacity of $\frac{n}{2}(2-h_2(\delta))$ is polynomial time achievable, where $h_2(\delta)$ denotes the binary entropy function. Our bounds are very tight, since it is a widely believed conjecture that $R_2(\delta) = 1-h_2(\delta)$.

2 Preliminaries

Let $\mathrm{CP}(G)$ denote the fractional clique packing of a graph defined as follows: Let $\mathcal{C}$ be the set of all cliques in G. For every $C \in \mathcal{C}$ define a variable $0 \leq x_C \leq 1$. Then $\mathrm{CP}(G)$ is the maximum values of $\sum_{C \in \mathcal{C}} x_C(|C|-1)$ subject to the constraint that

$$\sum_{C \in \mathcal{C}: u \in C} x_C \leq 1 \qquad \forall u \in V$$

Note that $\mathrm{CP}(G)$ can be computed in polynomial time in graphs, such as planar graphs, where all cliques have constant size. Furthermore, $\mathrm{CP}(G)$ is at least the size of the maximum fractional matching and they are obviously equal in triangle-free graphs since the only cliques are edges.

The following preliminary lemma shows that $\mathrm{Cap}_q(G) \geq \mathrm{CP}(G)$ for sufficiently large q. An equivalent result is known in the context of index coding but we include a proof here for completeness. The basic idea is that we can store $k-1$ information on a clique of size k by assigning $k-1$ independent uniform random variables to $k-1$ of the vertices and setting the final random variable to the sum (modulo q) of the first $k-1$ random variables.

Lemma 1. $\mathrm{Cap}_q(G) \geq \mathrm{CP}(G)$ *for sufficiently large q.*

Proof. Let $\{x_C\}_{C\in\mathcal{C}}$ achieve $\mathrm{CP}(G)$. Let q be sufficiently large such that q^{x_C} is integral for every C. For each clique $C = \{u_1, \ldots, u_{|C|}\}$ in the graph, define a family of random variables $X^C_{u_1}, X^C_{u_2}, \ldots, X^C_{u_{|C|}}$ where $X^C_{u_1}, X^C_{u_2}, \ldots, X^C_{u_{|C|-1}}$ are independent and uniform over $\{0, 1, \ldots, q^{x_C} - 1\}$ and

$$X^C_{u_{|C|}} = \sum_{i=1}^{|C|-1} X^C_{u_i} \bmod q^{x_C} \ .$$

Note that each $X^C_{u_i}$ can be deduced from $\{X^C_{u_j}\}_{j\neq i}$ and the entropy of $\{X^C_{u_j}\}_{j\neq i}$ is $x_C(|C|-1)$. Finally, let X_u be an encoding of $\{X^C_u\}_{C\in\mathcal{C}:u\in C}$ as a symbol from a q-ary alphabet; the fact that this is possible followings because $\sum_{C\in\mathcal{C}:u\in C} x_C \leq 1$. Then the entropy of X_V is $\mathrm{CP}(G)$ as required. □

Notation. Let $G[S]$ denote the subgraph induced by $S \subseteq V$. Let $\alpha(G)$ be the size of the largest independent set and let $\mathrm{VC}(G)$ be the size of minimum vertex cover. Let $\mathrm{MM}(G)$ be the size of the largest matching and let $\mathrm{FM}(G)$ be the weight of the maximum fractional matching.

3 Approximation Algorithms for Planar Graphs via Vertex Cover

In this section, we present approximation results for the storage capacity and optimal index coding rate of planar graphs. Specifically we show that $\mathrm{CP}(G)$ can be used to achieve a $3/2$ approximation of the storage capacity and a 2 approximation of the optimal index coding rate.

In our storage capacity result we use ideas introduced by Bar-Yehuda and Even [5] for the purpose of $5/3$-approximating the vertex cover in planar graphs. Specifically, they first considered a maximal set of vertex-disjoint triangles, reasoned about the vertex cover amongst these triangles, and then reasoned about the triangle-free induced subgraph on the remaining vertices. We consider a similar decomposition and reason about the integrality gap of vertex cover in each component. We parameterize our result in terms of the number of triangles; this will be essential in the subsequent result on optimal index coding rate.

Theorem 2. *Assume G is planar and let T be a set of $3t$ vertices corresponding to maximal set of t vertex disjoint triangles. Then,*

$$1 \leq \frac{\mathrm{Cap}(G)}{\mathrm{CP}(G)} \leq \frac{3t+k}{2t+3k/4}$$

where k is the size of the minimum vertex cover of $G[V \setminus T]$. Hence $\mathrm{CP}(G)$ is a $3/2$ approximation for $\mathrm{Cap}(G)$ and $4/3$ approximation if G is triangle-free.

Proof. Let $G' = G[V \setminus T]$. Partition the set of vertices into $T \cup C \cup I$ where C is the minimum vertex cover of G' and $I \subset V \setminus T$ is therefore an independent set. Let X_V be the set of variables that achieve storage capacity. Therefore,

$$\mathrm{Cap}(G) = H(X_V) = H(X_T) + H(X_C|X_T) + H(X_I|X_C, X_T) \leq H(X_T) + H(X_C|X_T) \leq 3t + k$$

since for each $v \in I$, $H(X_v|X_C, X_T) = 0$ since $N(v) \subset C \cup T$.

Consider the fractional clique packing in which each of the t vertex-disjoint triangles in T receive weight 1. Then, $\mathrm{CP}(G) \geq 2t + \mathrm{CP}(G')$. Then it remains to show that $\mathrm{CP}(G') \geq 3k/4$. Note that since G' is triangle-free planar graph, it is 3-colorable by Grötzsch's theorem [14]. Furthermore, $\mathrm{CP}(G')$ is the maximum fractional matching which, by duality, is the minimum fractional vertex cover. Hence it suffices to show that the size of the minimum fractional vertex cover of 3-colorable graph is at least $3/4$ of the size of the minimum (integral) vertex cover, i.e., $3k/4$. This can be shown as follows. Let $x_1, \ldots, x_n$ be an optimal fractional vertex cover, i.e., for all edges $uv \in G'$, $x_u + x_v \geq 1$. Since fractional vertex cover is $1/2$-integral, we may assume each $x_u \in \{0, 1/2, 1\}$. Let I_1, I_2, I_3 be a partitioning of $\{u \in [n] : x_u = 1/2\}$ corresponding to a 3-coloring where

$$\sum_{v \in I_1} x_v \geq \sum_{v \in I_2} x_v \geq \sum_{v \in I_3} x_v \ .$$

Then consider $y_1, \ldots, y_n$ where $y_u = 1$ iff $u \in I_2 \cup I_3$ or $x_u = 1$. Then

$$\sum_{u \in [n]} y_u \leq \sum_{u \in I_2 \cup I_3} y_u + \sum_{u \in [n] : x_u = 1} y_u \leq 2/3 \cdot 2 \cdot \sum_{u : x_u = 1/2} x_u + \sum_{u : x_u = 1} x_u \leq 4/3 \cdot \mathrm{CP}(G') \ ,$$

and $y_1, \ldots, y_n$ is a vertex cover because for every edge uv, at least one endpoint one of $\{x_u, x_v\}$ is 1 or at least one of u and v is in $I_2 \cup I_3$. □

We next use the result of the previous theorem, together with the chromatic number of planar and triangle-free planar graphs to achieve a 2 approximation for $\mathrm{Ind}(G)$.

Theorem 3. *Assume G is planar and let T be a set of $3t$ vertices corresponding to t vertex disjoint triangles. Then,*

$$1 \leq \frac{n - \mathrm{CP}(G)}{\mathrm{Ind}(G)} \leq \begin{cases} \frac{3n+3t}{4n-12t} + \frac{3}{4} & \textit{for } t \leq \frac{n}{12} \\ \frac{t}{n} + \frac{7}{4} & \textit{for } \frac{n}{12} \leq t \leq \frac{n}{4} \\ 4 - \frac{8t}{n} & \textit{for } t \geq \frac{n}{4} \end{cases} .$$

Maximizing over t implies that $n - \mathrm{CP}(G)$ is a 2 approximation for $\mathrm{Ind}(G)$ and a $3/2$ approximation if G is triangle-free.

Proof. From Theorem 2, we know that $\mathrm{CP}(G) \geq 2t + 3/4\,\mathrm{VC}(G')$ where $G' = G[V \setminus T]$. Therefore, we can bound $n - \mathrm{CP}(G)$ as follows:

$$\begin{aligned} n - \mathrm{CP}(G) &\leq n - (2t + 3/4\,\mathrm{VC}(G')) \\ &= n - (2t + 3/4(n - 3t - \alpha(G'))) \\ &= (n+t)/4 + 3/4\,\alpha(G') \end{aligned}$$

On the other hand,

$$\mathrm{Ind}(G) \geq \alpha(G) \geq \alpha(G')$$

where α denotes the size of the maximum independent set of the graph. Note that $\alpha(G) \geq n/4$ since G is planar and thus 4-colorable [2, 3]. Since G' has $n - 3t$ vertices and is triangle-free and planar and thus

3-colorable [14], $n - 3t \geq \alpha(G') \geq (n-3t)/3$. By combining inequalities above we get

$$\begin{aligned}
\frac{n - \mathrm{CP}(G)}{\mathrm{Ind}(G)} &\leq \min\left(\frac{(n+t)/4 + 3/4\,\alpha(G')}{\alpha(G)}, \frac{(n+t)/4 + 3/4\,\alpha(G')}{\alpha(G')}, \frac{(n+t)/4 + 3/4\,\alpha(G)}{\alpha(G)}\right) \\
&\leq \min\left(\frac{(n+t)/4 + 3/4(n-3t)}{n/4}, \frac{(n+t)/4}{(n-3t)/3} + 3/4, \frac{(n+t)/4}{n/4} + 3/4\right) \\
&= \min\left(4 - \frac{8t}{n}, \frac{3n+3t}{4n-3t} + \frac{3}{4}, \frac{t}{n} + \frac{7}{4}\right) .
\end{aligned}$$

□

4 Upper Bounds on Storage Capacity via Multiple Vertex Covers

In this section, we start by considering a linear program proposed by Blasiak, Kleinberg, and Lubetzky [8] that can be used to lower bound the optimum index coding rate or upper bound the storage capacity.[3] Unfortunately there are $\Omega(2^n)$ constraints but by carefully selecting a subset of constraints we can prove upper bounds on the storage capacity for a specific graph without solving the LP.

Our main goal in this section is to relate this linear program to a finding a suitable family of vertex covers of the graph. In doing so, we propose a most combinatorial "gadget" based approach to constructing good upper bounds that we think makes the process of proving strong upper bounds more intuitive. This allows us to prove a more general theorem that gives an upper bound on the storage capacity for a relatively large family of graphs. As an application of this theorem we show that a class of graphs closely related to the family of outerplanar graphs and another family of cartesian product graphs have capacity exactly $n/2$.

4.1 Upper Bound via the "Information Theoretic" LP

We first rewrite the index coding LP proposed by Blasiak, Kleinberg, and Lubetzky [8] for the purposes of upper-bounding storage capacity. We define a variable z_S for every $S \subseteq V$ that will correspond to an upper bound for $H(X_S)$. 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.

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

The second constraint corresponds to

$$H(X_T) - H(X_S) = H(X_T | X_S) = H(X_T | X_{\mathrm{cl}(S)}) \leq H(X_{T\setminus \mathrm{cl}(S)}) \leq |T \setminus \mathrm{cl}(S)| \, ,$$

whereas the last constraint follows from the sub-modularity of entropy. Hence, the optimal solution to the above LP is an upper bound on $\mathrm{Cap}(G)$. We henceforth refer to the above linear program as the *information theoretic* LP.

[3] Blasiak et al. only consider index coding but it is straightforward to adapt the LP in a natural way for storage capacity; see below.

4.2 Upper Bound via Gadgets

k-cover by gadgets is a technique for proving upper bounds on the storage capacity of graph. The core idea is to construct a set of k vertex covers for the graph via the construction of various gadgets which we now define.

A gadget $g(A, B)$ is created the following way: take two sets of vertices A and B, take their closures $\mathrm{cl}(A)$ and $\mathrm{cl}(B)$, find $S = \mathrm{cl}(A) \cup \mathrm{cl}(B)$ and $T = \mathrm{cl}(A) \cap \mathrm{cl}(B)$. Then S and T form a gadget. Call S the outside of the gadget and T the inside. We note that by taking $A = \{v\}$ and $B = \emptyset$ we obtain a gadget with the outside $\{v\}$ and empty inside (assuming v has no neighbors of degree one); call such gadget *trivial*. Define the weight of a gadget to be $|A| + |B|$. If we color every inside and outside gadget set with one of k colors such that the union of all sets of the same color forms a vertex cover, the total weight of gadgets in such coloring provides an upper bound on $k\,\mathrm{Cap}(G)$. Note that for $k = 1$ using gadgets with non-empty inside can only increase the total weight, so the only gadgets we need to consider are the trivial ones corresponding to individual vertices, and thus the construction is just a single vertex cover.

We can formulate the k-cover by gadgets (for fixed k) as the following linear program: Let $x_{S,c}$ be a variable where S is a set that is an outside or an inside of a gadget and c is one of k colors. Each variable has a corresponding weight $w_{S,c} = (|A| + |B|)/2$ where A and B are the 2 sets used to form the gadget that S is a part of. $x_{S,c} = 1$ if set S is colored with color c and 0 otherwise.

$$\begin{aligned}
\text{minimize} \quad & \frac{1}{k}\sum_{S,c} w_{S,c} x_{S,c} \\
\text{s.t.} \quad & \sum_{S: u\in S} x_{S,c} + \sum_{S: v \in S} x_{S,c} \geq 1 \quad \forall (u,v) \in E, \forall c \\
& \sum_c x_{H,c} = \sum_c x_{H',c} \quad \begin{array}{l}\text{for all gadgets, where } H \text{ is the outside of the gadget and} \\ H' \text{ the inside of the same gadget}\end{array}
\end{aligned}$$

The first condition states that every collection of sets of a certain color is a vertex cover and the second one states that the outside and inside of every gadget are used the same number of times.

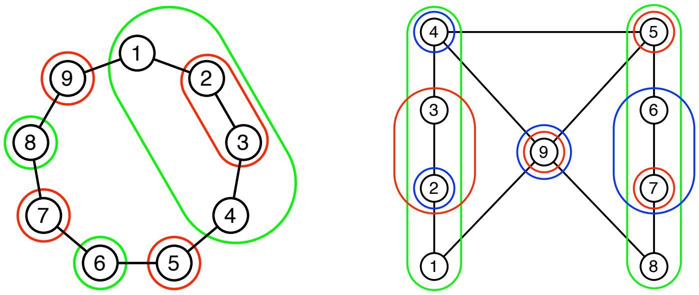
Theorem 4. *Any feasible integral solution to the above k-cover by gadgets LP is an upper bound on* $\mathrm{Cap}(G)$.

Proof. We prove this by showing that k-cover by gadgets follows from the fact that the optimum solution of the information theoretic LP is an upper bound. First, note which constraints correspond to the steps of forming a gadget:

$$\begin{aligned}
z_A &\leq |A| && \text{take set } A \\
z_B &\leq |B| && \text{take set } B \\
z_{\mathrm{cl}(A)} - z_A &\leq 0 && \text{find closure of } A \\
z_{\mathrm{cl}(B)} - z_B &\leq 0 && \text{find closure of } B \\
z_S + z_T &\leq z_{\mathrm{cl}(A)} + z_{\mathrm{cl}(B)} && \text{find } S = \mathrm{cl}(A) \cup \mathrm{cl}(B) \text{ and } T = \mathrm{cl}(A) \cap \mathrm{cl}(B)
\end{aligned}$$

If we sum all the constraints, we obtain $z_S + z_T \leq |A| + |B|$. Assume, that we used g gadgets in the cover. Then by summing all corresponding constraints, we get

$$z_{S_1} + z_{T_1} + z_{S_2} + z_{T_2} + \cdots + z_{S_g} + z_{T_g} \leq |A_1| + |B_1| + \cdots + |A_g| + |B_g|$$



(a) An Odd Cycle.

(b) An Outerplanar Graph.

Figure 1: Two examples of k-cover upper bounds. See text for details.

Group the sets into color classes $C_1, C_2, \ldots, C_k$. Let $U_i = \bigcup_{S \in C_i} S$. The corresponding constraints are then

$$z_{U_i} - \sum_{S \in C_i} z_S \leq 0 \quad \forall i \in \{1, 2, \ldots, k\}$$

$$z_{cl(U_i)} - z_{U_i} \leq 0 \quad \forall i \in \{1, 2, \ldots, k\}$$

Note that $z_V = z_{cl(U_i)}$ since U_i is a vertex cover. By summing these $2k$ constraints and the one obtained from building gadgets, we get $kz_V \leq |A_1| + |B_1| + \cdots + |A_g| + |B_g|$. □

4.2.1 Examples

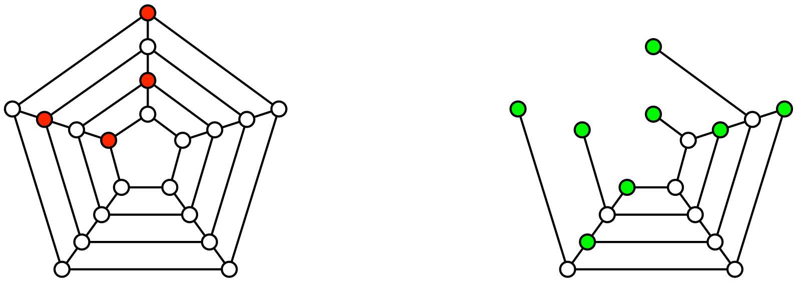
We next illustrate the use of the k-cover via gadgets approach with a couple of examples. First, we re-prove a result of Blasiak et al. [9] via a 2-cover by gadgets. Then we give an example of an outerplanar graph where it is necessary to consider a 3-cover by gadgets in order to establish a tight bound.

Odd Cycles. We prove that the storage capacity of an odd cycle of length n is $n/2$; see Figure 1(a) for an example where $n = 9$. $\mathrm{FM}(C_n) = n/2$, thus $\mathrm{Cap}(C_n) \geq n/2$. For the upper bound we form a gadget $g(A, B)$ by taking $A = \{v_1, v_3\}$, $B = \{v_2, v_4\}$ and obtaining outer set $S = \{v_1, v_2, v_3, v_4\}$ and inner set $T = \{v_2, v_3\}$. On the rest of the vertices we place trivial gadgets. Color S and trivial gadgets on $v_6, v_8, \ldots, v_{n-1}$ green, color T and trivial gadgets on $v_5, v_7, \ldots, v_n$ red. Green and red sets are then vertex covers and the total weight of all gadgets is n. Thus, $\mathrm{Cap}(C_n) \leq n/2$.

An Outerplanar Graph. We prove that the storage capacity of the graph in Figure 1(b) is $14/3$. This capacity is achieved by the fractional clique cover. Create a gadget $g_1(A_1, B_1)$ from $A_1 = \{v_1, v_3\}$ and $B_1 = \{v_2, v_4\}$ and another gadget $g_2(A_2, B_2)$ from $A_2 = \{v_5, v_7\}$ and $B_2 = \{v_6, v_8\}$. Place 1 trivial gadget on each of the vertices v_2, v_4, v_5, v_7 and 2 trivial gadgets on v_9. Color the sets as follows:

- Red: v_5, v_7, v_9 and the inside of gadget g_1
- Blue: v_2, v_4, v_9 and the inside of gadget g_2
- Green: the outside sets of both gadgets

Note that vertices of every color class form a vertex cover and the total weight of gadgets is 14.



(a) Graph G. Shaded vertices are $X = S_X$.

(b) Graph $G[Y]$. Shaded vertices are S_Y.

Figure 2: Example of storage capacity proof for a cartesian product graph G formed from a 5-cycle and a length 3 path. See text for details.

4.3 $n/2$ Upper Bound via Vertex Partition

The next theorem uses a 2-cover by gadgets to prove that a certain family of graphs have capacity at most $n/2$. Subsequently, we will use this theorem to exactly characterize the capacity of various graph families of interest.

Theorem 5. *Suppose that the vertices of a graph G can be partitioned into sets X and Y such that:*

1. *$G[X]$ and $G[Y]$ are both bipartite.*
2. *S_X is an independent set in $G[X]$ and S_Y is an independent set in $G[Y]$*

where $S_X \subseteq X$ consists of all vertices in X with a neighbor in Y and $S_Y \subseteq Y$ consists of all vertices in Y with a neighbor in X. Then $\mathrm{Cap}(G) \leq n/2$.

Proof. We prove this theorem by showing that G has a 2-cover by gadgets of total weight n. First, form a gadget $g_X(A, B)$ by letting (A, B) be a bipartition of the vertices of $G[X]$. Note that the vertices in X that are in the outside set of the gadget but not in the inside set, are exactly the vertices in S_X. This follows because for $v \in X$, $v \in \mathrm{cl}(A) \cap \mathrm{cl}(B)$ iff all of v's neighbors are in X. Similarly, form g_Y. Color the inside of g_X and the outside of g_Y red. Color the outside of g_X and the inside of g_Y blue. Observe that both color classes are vertex covers and the total weight of the 2 gadgets is $|X| + |Y| = n$. $\square$

4.3.1 Cartesian Product of a Cycle and a Bipartite Graph

The Cartesian product of graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ is denoted by $G_1 \Box G_2$ and defined as follows:

- The vertex set is the Cartesian set product $V_1 \times V_2$
- $(u, u')(v, v')$ is an edge iff $u = v$ and $u'v' \in E_2$ or $u' = v'$ and $uv \in E_1$

We next use Theorem 5 to show that any Cartesian Product of a cycle and a bipartite graph has storage capacity exactly $n/2$.

Theorem 6. *Let C_k be a cycle with $k > 3$, B a bipartite graph, and $G = C_k \Box B$. Then* $\mathrm{Cap}(G) = n/2$, *where n is the number of vertices in G.*

Proof. If k is even, G is a bipartite graph with $\mathrm{MM}(G) = \mathrm{VC}(G) = n/2$ and hence $\mathrm{Cap}(G) = n/2$. Assume for the rest of the proof that k is odd.

To show that $\mathrm{Cap}(G) \geq n/2$, consider the fractional matching where we assign weight $1/2$ to all edges of the form $(u, a)(v, a)$, i.e., edges that come from the cycle. Hence $\mathrm{Cap}(G) \geq n/2$.

To show that $\mathrm{Cap}(G) \leq n/2$ we proceed as follows. Consider the subgraph induced by vertices (u_i, v_1), (u_i, v_2), (u_i, v_3), etc., which is isomorphic to B. Color it using 2 colors and call the vertices of each color R_i and Q_i respectively. We now show that $X = R_1 \cup Q_2$ and $Y = V \setminus X$ satisfy the conditions of theorem 5. $G[X]$ has no edges and therefore is bipartite. $G[Y]$ is bipartite because it consists of $P_{k-3} \Box B$ which is bipartite (where P_{k-3} is a path of length $k - 3$ obtained by deleting edges $u_k u_1$, $u_1 u_2$, and $u_2 u_3$ from the cycle), edges between R_k and R_1, and edges between Q_2 and Q_3 which do not complete any cycles. $S_X = X$ is an independent set. $S_Y = R_k \cup Q_1 \cup R_2 \cup Q_3$ is also an independent set. □

4.3.2 Cycles With Chords That Are Not Too Close Together

We next apply Theorem 5 to prove that a family of graphs related to outerplanar graphs also has storage capacity $n/2$. Recall that any (connected) outerplanar graph without cut vertices is a cycle with non-overlapping chords. The family of graphs we consider is more general in the sense that we permit the chords to overlap but more restrictive in the sense that we require the endpoints of these chords to be at least a distance 4 apart on the cycle. A natural open question is to characterize $\mathrm{Cap}(G)$ for all outerplanar graphs. All that was previously known is that if we assume each X_i is a linear combination of $\{X_j\}_{j \in N(i)}$, then $\mathrm{Cap}(G)$ equals *integral* clique packing [7].

Theorem 7. *Let G be a cycle with a number of chords such that endpoints of chords are at least distance 4 apart on the cycle. Then* $\mathrm{Cap}(G) = n/2$.

Proof. To show that $\mathrm{Cap}(G) \geq n/2$, consider the fractional matching where we place weight $1/2$ on every edge of the cycle. To show that $\mathrm{Cap}(G) \leq n/2$ we proceed as follows. Label the vertices that are endpoints of chords $c_1, c_2, \ldots, c_k$ in the order they appear on the cycle. For every path between c_i and c_{i+1} (and between c_k and c_1) pick the middle vertex of the path to be included in X. If the path is of odd length, pick any of the 2 middle vertices. We now show that X and $Y = V \setminus X$ satisfy the conditions of theorem 5. $X = S_X$ is an independent set. $G[Y]$ is a forest and S_Y is an independent set due to the assumption on the distance between chord endpoints. □

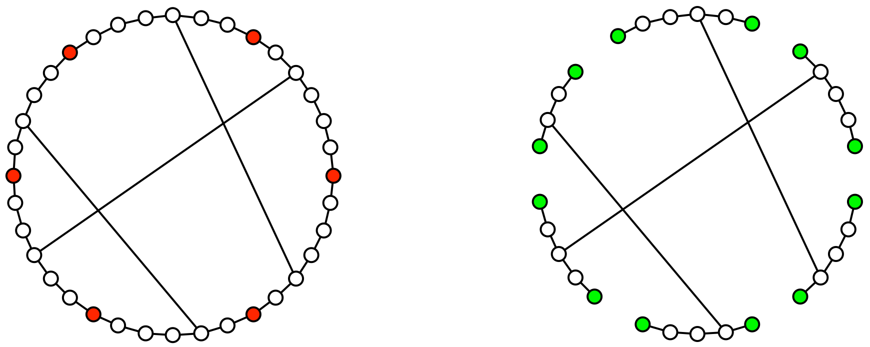
5 Partial Recovery

In this section, we extend the notion of storage capacity to cover for partial failures. This is a new generalization, that, as far as we understand, does not have a counterpart in index coding. As before, suppose we have a graph $G(V = [n], E)$ on n vertices. We assume here that vertex $i \in [n]$ stores $X_i \in \mathbb{F}_q^m$, a q-ary random vector of length m. We want the following repair criterion to be satisfied: if up to any $\delta, 0 \leq \delta \leq 1$, proportion of the m coordinates of $X_i, i \in [n]$ are erased, they can be recovered by using the remaining content of the vertex i and $X_{N(i)}$, the contents in the neighbors of the vertex.

We call the normalized asymptotic maximum total amount of information (in terms of q-ary unit)

$$\lim_{m \to \infty} \frac{H(X_1, X_2, \ldots, X_n)}{m},$$

that can be stored in the graph G, to be the *partial recovery capacity* of G. This is denoted by $\mathrm{Cap}_q(G, \delta)$.



(a) Graph G. Shaded vertices are $X = S_X$.

(b) Graph $G[Y]$. Shaded vertices are S_Y.

Figure 3: Example of a storage capacity proof for cycles with chords. See text for details.

We have the following simple facts.

Proposition 8. *For a graph G, $\mathrm{Cap}_q(G, 0) = n$ and $\mathrm{Cap}_q(G, 1) = \mathrm{Cap}(G)$.*

Proof. The first statement is quite evident. For the second, note that,

$$\mathrm{Cap}_q(G, 1) = \lim_{m\to\infty} \mathrm{Cap}_{q^m}(G) = \sup_m \mathrm{Cap}_{q^m}(G) = \mathrm{Cap}(G),$$

where we could use the $\lim$ and the $\sup$ interchangeably since $\mathrm{Cap}_{q^m}(G)$ is a nondecreasing and bounded sequence in m. □

In the remaining parts of this section, we will provide tight upper and lower bound on the quantity $\mathrm{Cap}_q(G, \delta)$.

5.1 Impossibility bound

Note that, the partial recovery capacity can be defined in terms of an entropy maximization problem, generalizing the storage capacity.

Theorem 9. *Let $H(X)$ be the entropy of X measured in q-ary units. Suppose, $X_i \in \mathbb{F}_q^m$, $i \in [n]$. For a graph $G([n], E)$, $\mathrm{Cap}_q(G, \delta)$ is upper bounded by the solution of the following optimization problem.*

$$\max \lim_{m\to\infty} \frac{H(X_1, \dots, X_n)}{m}, \tag{8}$$

such that,

$$H(X_i \mid X_{N(i)}) \le \log_q A_q(m, \delta m + 1),$$

where $A_q(m, d)$ is the maximum possible size of a q-ary m-length error-correcting code with minimum distance d.

Proof. Let $X_i \in \mathbb{F}_q^m$, $i \in [n]$ be the random variables that can be stored in the vertices of G satisfying the repair condition. Suppose we are given the values of $X_{N(i)}$. In this situation let $M \subseteq \mathbb{F}_q^m$ be the set of possible values of X_i ($P(X_i = a) > 0, \forall a \in M$). Let X_i^1, X_i^2 be any two different elements of M. We claim that, the Hamming distance between X_i^1, X_i^2 is at least $\delta m + 1$, or

$$d(X_i^1, X_i^2) \geq \delta m + 1.$$

Suppose this is not true. Then there exist $X_i^1, X_i^2 \in M$ such that $d(X_i^1, X_i^2) \leq \delta m$. Let $J \subset \{1, \dots, m\}$ be the coordinates where X_i^1 and X_i^2 differ. Therefore, $|J| \leq \delta m$. Suppose X_i^1 was stored in vertex i and the coordinates in J are erased. Now, there will not be any way to uniquely identify X_i: it can be either of X_i^1 or X_i^2. Hence the repair condition will not be satisfied which is a contradiction.

Therefore, $M \subseteq \mathbb{F}_q^m$ is a set of vectors such that any two elements of M is Hamming distance at least $\delta m + 1$ apart. Hence M is an error-correcting code with minimum distance $\delta m + 1$. And therefore, $|M| \leq A_q(m, d)$. This implies, $H(X_i \mid X_{N(i)}) \leq \log_q A_q(m, \delta m + 1)$, which proves the theorem. □

Let us define

$$R_q(\delta) \equiv \lim_{m \to \infty} \frac{A_q(m, \delta m + 1)}{m},$$

assuming the limit exists.

Corollary 10. *We must have, for any graph G, $\mathrm{Cap}_q(G, \delta) = \mathrm{Cap}(G)$ for $\delta \geq 1 - \frac{1}{q}$. In particular, $\mathrm{Cap}_2(G, \delta) = \mathrm{Cap}(G)$ for $\delta \geq \frac{1}{2}$.*

The proof of this fact follows since $R_q(\delta) = 0$ for $\delta \geq 1 - \frac{1}{q}$.

Generalizing the technique of upper bounding the storage capacity via an information theoretic linear program, we can obtain an upper bound on $\mathrm{Cap}_q(G, \delta)$. We define a variable z_S for every $S \subseteq V$ and let $\mathrm{bo}(S, T) = (\mathrm{cl}(S) \setminus S) \cap T$ denote the boundary of the set S consisting of vertices in T with all neighbors in S. Our main upper bound is the following.

Theorem 11. *Consider the LP below.*

$$\begin{aligned} \textit{maximize} \quad & z_V \\ \textit{s.t.} \quad & z_\emptyset = 0 \\ & z_T - z_S \leq |T \setminus S| - (1 - R_q(\delta))\,\mathrm{bo}(S, T) \quad \forall S \subseteq T \\ & z_S + z_T \geq z_{S \cap T} + z_{S \cup T} \quad \forall S, T \end{aligned}$$

The optimal solution to the above LP is an upper bound on $\mathrm{Cap}_q(G, \delta)$.

We omit the proof here since it is exactly same as the proof of the bound via information theoretic LP of Sec. 4.1.

5.1.1 Odd cycle

Consider an odd cycle with n vertices (n is odd). Below we show an example to illustrate the above bound on partial recovery capacity.

Consider the following subset of constraints:

$$\begin{aligned}
2 &\geq z_{\{1,3\}} - z_{\emptyset} \\
2 &\geq z_{\{2,4\}} - z_{\emptyset} \\
1 &\geq z_{\{i\}} \quad \forall i \in \{5, 6, \dots, n\} \\
R_q(\delta) &\geq z_{\{1,2,3\}} - z_{\{1,3\}} \\
R_q(\delta) &\geq z_{\{2,3,4\}} - z_{\{2,4\}} \\
z_{\{1,2,3\}} + z_{\{2,3,4\}} &\geq z_{\{2,3\}} + z_{\{1,2,3,4\}} \\
z_{\{2,3\}} + z_{\{5\}} + z_{\{7\}} + \cdots + z_{\{n\}} &\geq z_{\{2,3,5,7,\dots,n\}} + \frac{n-3}{2} z_{\emptyset} && \text{(a)} \\
z_{\{1,2,3,4\}} + z_{\{6\}} + z_{\{8\}} + \dots z_{\{n-1\}} &\geq z_{\{1,2,3,4,6,8,\dots,n-1\}} + \frac{n-5}{2} z_{\emptyset} && \text{(b)} \\
(n - \frac{n+1}{2}) - (1 - R_q(\delta))\frac{n-1}{2} &\geq z_V - z_{\{2,3,5,7,\dots,n\}} \\
(n - \frac{n+3}{2}) - (1 - R_q(\delta))\frac{n-3}{2} &\geq z_V - z_{\{1,2,3,4,6,8,\dots,n-1\}}
\end{aligned}$$

Equations (a) and (b) above are repeated applications of the inequality: $z_S + z_T \geq z_{S \cup T} + z_{\emptyset}$ if $S \cap T = \emptyset$. By summing up those constraints we get

$$n + 2R_q(\delta) + R_q(\delta)(n-2) \geq 2z_V - 2z_{\emptyset}$$

and thus

$$\mathrm{Cap}_q(G, \delta) \leq z_V \leq \frac{n}{2}(1 + R_q(\delta)),$$

whenever G is an odd cycle.

5.2 Achievability bound

A naive achievability bound on $\mathrm{Cap}_q(G, \delta)$ is given by,

$$\mathrm{Cap}_q(G, \delta) \geq n(1 - h_q(\delta)), \quad \delta \leq \frac{1}{2},$$

where $h_q(x) \equiv x \log_q(q-1) - x \log_q x - (1-x)\log_q(1-x)$. This amount of storage can be achieved by just using an error-correcting code of length m, distance $\delta m + 1$, and rate $1 - h_q(\delta)$ in each of the vertices. Such codes exist, by the Gilbert-Varshamov bound. Also,

$$\mathrm{Cap}_q(G, \delta) \geq 0,$$

for $0 \leq \delta \leq 1 - \frac{1}{q}$.

This simple bound can be improved by more carefully designing a code. Our main result of this section is the following.

Theorem 12. *Given a graph G, let $\mathcal{C}$ be the set of all cliques of G. The generalized clique packing number $\mathrm{CP}_\delta(G)$ is defined to be the optimum of the following linear program. For $0 \leq x_C \leq 1, \forall C \in \mathcal{C}$,*

$$\max \sum_{C \in \mathcal{C}} x_C(|C| - h_q(\delta)),$$

such that,

$$\sum_{C \in \mathcal{C}: u \in C} x_C \leq 1.$$

Then,

$$\mathrm{Cap}_q(G, \delta) \geq \mathrm{CP}_\delta(G), \quad \delta \leq 1 - \frac{1}{q},$$

and,

$$\mathrm{Cap}_q(G, \delta) \geq \mathrm{CP}(G), \quad \delta > 1 - \frac{1}{q}.$$

Proof. We illustrate the proof of this theorem by constructing a sequence of error-correcting codes that serves our purpose.

Let us first find a maximum matching $M \subseteq E$ of the graph. Now for each edge $(u, v) \in M$ use a q-ary error-correcting code of length $2m$ that can correct any δm erasures in the first m coordinates and any δm erasures in the second m coordinates.

We claim that such error-correcting code of length $2m$ and dimension $2m - mh_q(\delta)$ exists.

Randomly and uniformly choose a q-ary parity check matrix of size $(2m - k) \times 2m$ (that is, each coordinate of the matrix is chosen from $\{0, 1, \dots, q-1\}$ with uniform probability). The probability that a vector of weight δm is a codeword is $q^{-(2m-k)}$. Now the probability that there exists such a codeword that is an uncorrectable erasure pattern of the above type is

$$\leq 2\binom{m}{\delta m}(q-1)^{\delta m} q^{-(2m-k)} \sim q^{-(2m-k-mh_q(\delta))}, \delta \leq 1 - \frac{1}{q}.$$

Hence there exists such a code with dimension = $2m - mh_q(\delta)$. Therefore the total number of q-ary numbers that can be stored is

$$|M| \cdot (2m - mh_q(\delta)) = m \cdot \mathrm{MM}(G)(2 - h_q(\delta)),$$

where $\mathrm{MM}(G)$ is a maximum matching in G. Hence,

$$\mathrm{Cap}_q(G, \delta) \geq \mathrm{MM}(G)(2 - h_q(\delta)), \quad \delta \leq 1 - \frac{1}{q}.$$

For $\delta \geq 1 - \frac{1}{q}$, $\mathrm{Cap}_q(G, \delta) \geq \mathrm{MM}(G)$.

Following an argument similar to Lemma 1, it is possible to improve the maximum matching argument here to the fractional maximum matching. Therefore, we must have,

$$\mathrm{Cap}_q(G, \delta) \geq \mathrm{FM}(G)(2 - h_q(\delta)), \quad \delta \leq 1 - \frac{1}{q},$$

and, or $\delta \geq 1 - \frac{1}{q}$, $\mathrm{Cap}_q(G, \delta) \geq \mathrm{FM}(G)$, where $\mathrm{FM}(G)$ denote the size of the maximum fractional matching of G.

The above argument can easily be extended towards clique packing instead of maximum matching. If we have a clique of size L, then we need to choose a code of length Lm. We again, randomly and uniformly choose a q-ary parity check matrix of size $(Lm - k) \times Lm$. The probability that a vector of weight δm is a codeword is $q^{-(Lm-k)}$. Now the probability that there exists such a codeword that is an uncorrectable erasure pattern of the above type is

$$\leq L\binom{m}{\delta m}(q-1)^{\delta m} q^{-(Lm-k)} \sim q^{-(Lm-k-mh_q(\delta))}, \delta \leq 1 - \frac{1}{q}.$$

Therefore there exists a good code of dimension $Lm - mh_q(\delta)$. Using this code, total number of q-ary integers that can be stored is $m\,\mathrm{CP}_\delta(G)$, by following the argument of Lemma 1. This proves the claim. □

Example - Odd Cycle: Let us consider the example of n-cycle again where n is an odd number. Since the size of a fractional matching is $\frac{n}{2}$, we have

$$\mathrm{Cap}_q(G,\delta) \geq \frac{n}{2}(2 - h_q(\delta)), \quad \delta \leq 1 - \frac{1}{q},$$

and $\mathrm{Cap}_q(G,\delta) \geq \frac{n}{2}$ when $\delta > 1 - \frac{1}{q}$. Compare this with the impossibility bound that we have,

$$\mathrm{Cap}_q(G,\delta) \leq \frac{n}{2}(1 + R_q(\delta)).$$

It is widely conjectured that the optimal rate of an error-correcting code is given by

$$R_q(\delta) = 1 - h_q(\delta),$$

for small q, which is also known as the Gilbert-Varshamov conjecture. If this conjecture is true, then our upper and lower bounds match exactly. In particular, for large q (i.e., $q \to \infty$), we have $h_q(\delta) \to \delta$ and $R_q(\delta) \to 1 - \delta$. Hence, our bounds match definitively in the regime of large q.

References

[1] N. Alon and A. Orlitsky. Source coding and graph entropies. *IEEE Transactions on Information Theory*, 42(5):1329–1339, 1996.

[2] K. Appel and W. Haken. Every planar map is four colorable. part i: Discharging. *Illinois J. Math.*, 21(3):429–490, 09 1977.

[3] K. Appel, W. Haken, and J. Koch. Every planar map is four colorable. part ii: Reducibility. *Illinois J. Math.*, 21(3):491–567, 09 1977.

[4] B. S. Baker. Approximation algorithms for np-complete problems on planar graphs. *J. ACM*, 41(1):153–180, 1994.

[5] R. Bar-Yehuda and S. Even. On approximating a vertex cover for planar graphs. In *Proceedings of the 14th Annual ACM Symposium on Theory of Computing, May 5-7, 1982, San Francisco, California, USA*, pages 303–309, 1982.

[6] Z. Bar-Yossef, Y. Birk, T. Jayram, and T. Kol. Index coding with side information. *IEEE Transactions on Information Theory*, 57(3):1479–1494, 2011.

[7] Y. Berliner and M. Langberg. Index coding with outerplanar side information. In *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, pages 806–810. IEEE, 2011.

[8] A. Blasiak, R. Kleinberg, and E. Lubetzky. Lexicographic products and the power of non-linear network coding. In *IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS 2011, Palm Springs, CA, USA, October 22-25, 2011*, pages 609–618, 2011.

[9] A. Blasiak, R. D. Kleinberg, and E. Lubetzky. Index coding via linear programming. *CoRR*, abs/1004.1379, 2010.

[10] R. Bowden, H. X. Nguyen, N. Falkner, S. Knight, and M. Roughan. Planarity of data networks. In *Teletraffic Congress (ITC), 2011 23rd International*, pages 254–261. IEEE, 2011.

[11] M. Effros, S. El Rouayheb, and M. Langberg. An equivalence between network coding and index coding. *IEEE Transactions on Information Theory*, 61(5):2478–2487, 2015.

[12] M. Gadouleau, A. Richard, and S. Riis. Fixed points of boolean networks, guessing graphs, and coding theory. *SIAM Journal on Discrete Mathematics*, 29(4):2312–2335, 2015.

[13] P. Gopalan, C. Huang, H. Simitci, and S. Yekhanin. On the locality of codeword symbols. *IEEE Transactions on Information Theory*, 58(11):6925–6934, 2012.

[14] H. Grötzsch. Zur theorie der diskreten gebilde, vii: Ein dreifarbensatz fr dreikreisfreie netze auf der kugel. *Wiss. Z. Martin-Luther-U., Halle-Wittenberg, Math.-Nat. Reihe*, 8:109120, 1959.

[15] S. Khot and O. Regev. Vertex cover might be hard to approximate to within 2- ε. *Journal of Computer and System Sciences*, 74(3):335–349, 2008.

[16] M. Langberg and A. Sprintson. On the hardness of approximating the network coding capacity. *IEEE Transactions on Information Theory*, 57(2):1008–1014, 2011.

[17] A. Mazumdar. Storage capacity of repairable networks. *IEEE Transactions on Information Theory*, 61(11):5810–5821, 2015.

[18] C. Shannon. The zero error capacity of a noisy channel. *IRE Transactions on Information Theory*, 2(3):8–19, 1956.