Intersection between Information Theory and Combinatorics

Stoyan Dimitrov
University of Illinois at Chicago, sdimit6@uic.edu

Abstract

This short survey considers some examples of information-theoretic tools and reasoning applied to problems in Combinatorics, as well as some combinatorial problems arising in Information Theory. The concrete areas that are object of investigation are *Zero-error communication*, *Information Theory in counting problems* and *Information Theory in Additive Combinatorics*.

I. Introduction

Information Theory(IT) studies the ways to exploit patterns in a given data. Thus many of its most important problems have combinatorial nature, eventhough the classical channel coding problem is not attacked by using purely combinatorial observations. The main reason is that while considering the aforementioned problem, we tolerate a small probability of error. However, error-free version of the same problems can be considered and this is substantially motivated, because in many applications an eventual error might be either very expensive or inadmissible. In these new settings the problem is purely combinatorial.

In Section II, we consider the error-free communication problem over a noisy channel and define the so-called Shannon capacity of a channel and its corresponding graph. At the end of the section, we discuss some open problems and already known bounds for the Shannon capacity of certain graphs and families of graphs. Section III demonstrates the usability of IT in various counting problems and contains some nice proofs for some of them. In the last section, we briefly mention a few recent results in Additive Combinatorics concerning extremal values of certain expressions of sums and differences of entropies. The whole content of this work is an evidence that the intersection between IT and Combinatorics is huge and contains very diverse examples.

II. Shannon Capacity and zero-error communication

This section deals with zero-error capacity of communication channel, i.e. the maximum number of bits per channel use when a zero probability of error is required. The foundations of zero-error IT were given again by Shannon around 8 years after his seminal paper [5]. A good survey of the topic is given in [1].

Let us have a discrete memoryless channel $W : \mathcal{X} \to \mathcal{Y}$ with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$(ussualy, $\mathcal{X} = \mathcal{Y}$) and transitional matrix $p(y \mid x)$. This would mean that $p(y \mid x) = \mathbb{P}(\{x - transmitted, y - received\})$ and by the memoryless property, the chance to receive $Y = y_1 y_2 \dots y_n$, when transmitting $X = x_1 x_2 \dots x_n$ is simply

$$p^n(Y \mid X) = \prod_{i=1}^{n} p(y_i \mid x_i) \tag{1}$$

We will call the sequences X' and X'' indistinguishable for the channel W if there exists a sequence Y that can be received after transmitting both X' or X''. If two sequences are not indistinguishable we will call them distinguishable. Therefore, $X' = x'_1 x'_2 \dots x'_n$ and $X'' = x''_1 x''_2 \dots x''_n$ would be two distinguishable sequences if and only if for each sequence $Y = y_1 y_2 \dots y_n$, either $p(y_i \mid x'_i) = 0$ or $p(y_i \mid x''_i) = 0$ for at least one $i = 1, 2, \dots, n$. Note that we would be able to transmit without error any set of pairwise distinguishable sequences. Thus we would be interested in $N(W, n)$ - the maximum cardinality of a set of pairwise distinguishable sequences for a given channel W. Then, the zero-error capacity of a channel, as defined by Shannon([4]) is

$$C_0(W) = \limsup_{n\to\infty} \frac{1}{n} \log N(W,n) \tag{2}$$

In fact, C_0 is the least upper bound of all rates which can be achieved with zero probability of error. It may be seen that the value of C_0 depends only on which input letters are adjacent to each other. Here, $a \in \mathcal{X}$ and $b \in \mathcal{Y}$ are adjacent letters if $p(b \mid a) > 0$. This means that if one is interested only in zero-error transmission through a channel W, it suffices to consider the corresponding graph $G_W(V,E)$ with vertex set $V = \mathcal{X} \cup \mathcal{Y}$ and edge set $E = \{(a,b) \mid p(b \mid a) > 0\}$, i.e. there is an edge between two letters a and b iff they are adjacent. In order to mirror transmission of messages of length greater than 1, we will need the operation $G \boxtimes H$ denoting the strong product of the graphs G and H defined as follows. The vertices of $G \boxtimes H$ are all pairs (u,v), $u \in V(G)$, $v \in V(H)$, i.e the Cartesian product $V(G) \times V(H)$. As for the edge set, there is an edge betwen (v,v') and (w,w') if and only if one of the following holds - $(v,w) \in E(G)$ and $(v',w') \in E(H)$, or $v = w$ and $(v',w') \in E(H)$, or $(v,w) \in E(G)$ and $v' = w'$. We could note that the strong product $G \boxtimes H$ corresponds to the graph of a channel which is the direct product of the channels corresponding to G and H. Thus $G \boxtimes G$ would represents transmission of two symbols independently over the same channel W having corresponding graph G. To denote $G \boxtimes G \boxtimes \ldots \boxtimes G$, where G appears p times, we will write $G^{\boxtimes p}$. One may easily see that $N(W,n) = \alpha(G^{\boxtimes n})$, the independence number of $G^{\boxtimes n}$, which is the maximum number of vertices of the graph with no edge between them. Now, in these graph-theoretical settings, we can talk about the equivalent of the zero-error capacity $C_0(W)$ - the so-called Shannon capacity $\Theta(G) = 2^{C_0}$ of the graph G:

$$\Theta(G) = \lim_{n\to\infty} \sqrt[n]{\alpha(G^{\boxtimes n})} \tag{3}$$

One might expect that $\alpha(G^{\boxtimes n}) = [\alpha(G)]^n$ for all G and n, that is, if we choose the largest possible set of non-adjacent letters and form all sequences of these of length n, then this would be the best error free code of length n. However, we know that in IT often things become more efficient when the communication becomes longer. Indeed, let's consider the example of the graph C_5, also considered by Shannon(see Fig 1).

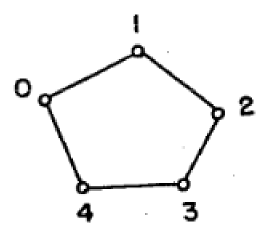

Fig. 1. The scheme of the cycle C_5 as given by Shannon in [4].

In the corresponding channel each transmitted symbol(number) x is received as either $x+1(\mod 5)$ or $x-1(\mod 5)$ with some positive probabilities. As you can see, for this channel the maximum cardinality of an independent set(set of pairwise non-confusable symbols) is 2. Indeed, we are not able to select 3 vertices of the graph without having an edge between at least 2 of these vertices. For instance, let's choose the two symbols 0 and 2. The set of sequences 00, 02, 20, 22 is of cardinality 4 and the length of each sequence is 2. However, it is possible to select 5 such non-confusable with each other words of length 2. Take, for example, 00, 12, 24, 31 and 43. It is readily verified that no 2 of these are adjacent in $(C_5)^{\boxtimes 2}$. Thus, $\Theta(C_5) \geq \sqrt{5}$. Surprisingly, when we have less than 5 input letter, the described strategy of selecting all the words with symbols from the largest set of non-adjacent letters indeed gives the independent sets of biggest cardinality(this was proved by Shannon).

The question of the exact value of $\Theta(C_5)$ has been opened for nearly 20 years, until in 1979, Lovász([6]) proved by a relatively simple geometric technique that $\Theta(C_5) \leq \sqrt{5}$. The latter means that the $\Theta(C_5) = \sqrt{5}$. More details on the proof and some recent facts related to Shannon capacities are contained in [10]. The paper of Lovász ([6]) generalized his technique and introduced his Lovász number $\theta(G)$ which turns out to be an upper bound for the Shannon capacity:

$$\theta(G) = \inf_{A \in \mathcal{A}_G} \lambda_1(A)$$

In the above definition, $\mathcal{A}_G$ is the set of all real symmetric matrices indexed by $V(G)$ that satisfy $A_{vw} = 1$ if $v = w$ or if v and w are non-adjacent, whereas $\lambda_1(A)$ stands for the largest eigenvalue of the matrix A. Here is the general theorem obtained by Lovász:

Theorem 1. *(Lovász [6]) For every graph G we have*

$$\Theta(G) \leq \theta(G)$$

It turns out that determining of Shannon capacity for arbitrary graph is very difficult. Around 60 years after Shannon defined his measure, the exact value of $\Theta(C_7)$ is still not known, eventhough the known bounds from above and below are quite close to each other. The current values are

$$3.2271 \approx \sqrt[4]{108} \leq \Theta(C_7) \leq \theta(C_7) \approx 3.3176$$

The latter lower bound was obtained very recently by deducing an existance of an independent set of size 350 in $C_7^{\boxtimes 5}$ by Mathew and Östergard([12]) and the upper bound follows from Theorem 1.

A. Known bounds on Shannon capacities

Eventhough determining the values of Shannon capacity for arbitrary graph could be extremely hard, some results were obtained for certain classes of graphs that have special structure and a lot of symmetries. The results cited below are reported in [7].

- **Vertex-Transitive Graphs**

  Regular is a graph for which all of its vertices have the same degree. A graph is vertex-transitive if for each pair of vertices (i, j) there exists an automorphism π such that $\pi(i) = j$. Not all regular graphs are vertex-transitive. The smallest example proving the latter statement is the union of the cycles C_3 and C_4.

  Theorem 2 (Lovász [6])**.** *Let G be a vertex-transitive graph on n vertices. Then*

  $$\theta(G)\theta(\bar{G}) = n$$

  Now, Theorem 1 implies that

  $$\Theta(G)\Theta(\bar{G}) \leq n$$

  It turns out that the inequality becomes equality if $G = \bar{G}$, that is, when G is self-complementary vertex-transitive graph. In this cases the value of the Shannon capacity is known:

  Theorem 3 (Lovász [6])**.** *Let G be a self-complementary vertex-transitive graph on n vertices. Then*

  $$\Theta(G) = \theta(\bar{G}) = \sqrt{n}$$

- **Circulant Graphs**

  Circulant graphs are a large subclass of the class of vertex-transitive graphs. Let $D = \{d_1, \ldots, d_l\} \subseteq [n]$, $n \geq 3$. A circulant graph $C_n(D)$ is a graph on n vertices such that $\{i, j\} \in E$ iff $j = i \pm d_k$ for some $k \in [l]$. All cycles are circulant graphs. We already considered the cases C_5 and C_7. Below is the table containing the values of the known bounds for $\alpha((C_n)^{\boxtimes p})$ for odd n. The Table is almost

n/p	**1**	**2**	**3**	**4**	**5**	**6**	p'	Lower bound on C_0
5	2	5	10	25	50	125	2	2.2361
7	3	10	33	115 108	350	1101	4	3.2271
9	4	18	81	363 324	1458	6561	3	4.3267
11	5	27	148	748	3996	21904	3	5.2896
13	6	39	247	1534	9633	61009	3	6.2743
15	7	52	381	2770	19864	145924	3	7.2495
17	8	68	578	4913	39304	334084	4	8.3721
19	9	85	807	7666	68994	651610	4	9.3571
21	10	105	1092	11441	114660	1201305	4	10.3423
23	11	126	1437	16466	181126	2074716	4	11.3278
25	12	150	1875	23125	281250	3515625	4	12.3316
27	13	175	2362	31522	413350	5579044	4	13.3246
29	14	203	2929	42017	594587	8579041	4	14.3171

Fig. 2. Known lower bounds for $\alpha((C_n)^{\boxtimes p})$, almost as reported in [8].

an exact copy of the one given in [8] and the difference is that the new bounds obtained by Mathew and Östergard in [12] were added.
The column p' contains the values of p which give the largest lower bound. The upper bounds received are known due to the following fact:

Theorem 4 (Lovász)**.** *For odd n,*

$$\theta(C_n) = \frac{n \cos(\pi/n)}{1 + \cos(\pi/n)}$$

Note that only the odd values of n are considered. It is known that $\Theta(C_n) = \frac{n}{2}$, when n is even (see [6] again, as well as [10]).

- **Kneser Graphs**

Finally, let us mention the known result about the so-called Kneser graphs, which are also vertex-transitive. Let q and r are integers such that $q \geq 2r$. A Kneser graph K(q,r) has vertices - some subsets of the set $[q]$ and an edge exists between 2 vertices iff the coresponding sets are disjoint. They were named after Martin Kneser who first investigated them in 1955. Kneser graphs are vertex-transitive.

Theorem 5 (Erd́os, Ko, Rado [9])**.**

$$\alpha(K(q,r)) = \binom{q-1}{r-1}$$

In fact, the eigenvalues of the adjacency matrix of Kneser graphs are known to be equal to

$$(-1)^i \binom{q-r-i}{r-1}, i = 0, 1, \ldots, r$$

The last two results were used to determine the Lovász number $\theta(K(q,r))$ which turned out to be equal to the Shannon capacity.

Theorem 6 (Lovász [6])**.**

$$\Theta(K(q,r)) = \binom{q-1}{r-1}$$

One may look at [7],[8] and [10] for more details on these and other related results.

A few unsolved problems

At the end of the section we list a few unsolved problems related to the Shannon capacities and mentioned in the great talk of Noga Alon([2]): –

1) What is the maximum possible capacity of a disjoint union of two graphs, each having capacity at most k?
2) Is the maximum possible capacity of a graph G with independence number 2 bounded by an absolute constant?
3) Is the problem of deciding if the Shannon capacity of a given input graph is at least a given real x decidable ?
4) What is the expected value of $\Theta(G(n, \frac{1}{2}))$?

III. Counting problems and entropy

Except the combinatorial problems occuring in IT, several applications of IT in classical combinatorics have been found in the recent years. For instance, such applications exist in various enumeration and counting problems. These problems are among the oldest types of problems considered. One of the simplest relevant examples is the following informal argument that in order to sort an array with n elements, we have to perform at least $\log n!$ comparisons. Since there are $n!$ permutations, one needs at least $\log n!$ bits of information to describe a permutation. On the other hand, each comparison gives us at most one bit of information. After performing these comparisons, a permutation is uniquely determined, so every algorithm would need to perform at least $\log n!$ comparisons. Therefore, it does not exist a sorting algorithm with complexity better than $\mathcal{O}(n \log n)$, which is a well-known fact. This statement can be formalized as follows: Assume we have a random permutation X and t other random variables Y_1, Y_2,... ,Y_t representing t comparisons performed by a sorting algorithm. We have

$$\begin{aligned}\log n! = H(X) &= H(f(Y_1, Y_2, \ldots, Y_t)) \\ &\leq H(Y_1, Y_2, \ldots, Y_t) \\ &\leq \sum_{i=1}^{t} H(Y_i) \\ &\leq t\end{aligned}$$

Using only simply properties of entropy, one may also prove facts in graph theory like the theorem of Bregman([15]) related to perfect matchings in bipartite graphs. A bipartite graph is a graph where the vertex set can be partitioned into sets A and B, such that all edges go between A and B. A matching in a graph is a set of pairwise disjoint edges. If each vertex of a graph has an edge of a matching incident to it, we call this matching perfect. By $d(v)$ we denote the degree of v, that is, the number of edges of G incident on v. The theorem, also known as Minc's conjecture is given below:

Theorem 7 (Bregman)**.** *Let $G = (A, B, E)$ be a bipartite graph with $|A| = |B| = n$. Then the number of perfect matchings in G is at most*

$$\prod_{v \in A} (d(v)!)^{\frac{1}{d(v)}}$$

Speaking about IT and Graph Theory, we cannot skip a valuable tool like the Shearer's entropy inequality:

Lemma 1 (Shearer)**.** *Let $X = (X_1, X_2 \ldots X_n)$ be a random variable and $\mathcal{A} = \{A_i\}_{i \in I}$ be a collection of subsests of $[n]$,such that each element of $[n]$ appears in at least k members of $\mathcal{A}$. For $A \subseteq [n]$, let $X_A = (X_j : j \in A)$. Then,*

$$\sum_{i \in I} H(X_{A_i}) \geq kH(X)$$

The lemma is an apparent generalization of the inequality $H(X) + H(Y) \geq H(X, Y)$. It was first introduced by Chung et all.([16]). Let us demonstrate the lemma's usefulness with the following problem:

Suppose n distinct points in $\mathbb{R}^3$ have n_1 distinct projections on the XY-plane, n_2 distinct projections on the XZ-plane and n_3 distinct projections on the YZ-plane. Then,

$$n_1 n_2 n_3 \geq n^2$$

Proof: Let $P = (A, B, C)$ be a point picked out of the n given points at random, i.e. according to uniform distribution. Then, for the projections, we have $P_1 = (A, B)$, $P_2 = (A, C)$,$P_3 = (B, C)$. The Shearer's lemma gives us $H(P_1) + H(P_2) + H(P_3) \geq 2H(P)$, but we also have $H(P) = \log n$ and $H(P_i) \leq \log n_i$ and the result follows. ■

This strategy can be generalized and the usual steps of a counting proof using entropy are below:

- Consider a random object $X \in \mathcal{C}$ drawn according to discrete uniform distribution over the class of objects $\mathcal{C}$, with cardinality that is what we want to find.
- Find a way to represent X as a vector of other random variables $(X_1, X_2, \ldots, X_n$ and apply Shearer's lemma for certain subsets of indexes A_i.
- Apply other inequalities like Jensen's, etc. if needed, to obtain the desired result.

Important applications of the Shearer's lemma to more difficult problems can be found in the works of Kahn([17],[19],[20]), Bansal et all.([21]) and Furedi([18]). The lemma is applicable even to problems connected to graph homomorphisms known for their high difficulty. A Theorem proved in [22] by Madiman and Tetali extends previous observations and gives a bound for the number of graph homomorphisms between two arbitrary graphs. The same work is a good summary on results related to inequalities between entropies of random variables consisted of more than one component.

Graph Covering problems

Theorem 8. *One needs at least $t \geq \log n$ bipartite graphs $G_1, G_2, \ldots, G_t$ with vertex set $[n]$ to "cover" the complete graph K_n, i.e. to suffice $G_1 \cap G_2 \cap \ldots \cap G_t = K_n$, where K_n is called complete because it contains every possible edge $\{(i, j) \mid 1 \leq i, j \leq n\}$.*

Proof: Denote the bipartite graphs with $G_i(A_i, B_i, E_i)$ and pick a random vertex $v \in [n]$. Let χ_i be the indicators for the events $v \in A_i$. i.e.

$$\chi_i = \begin{cases} 0 & v \in A_i \\ 1 & v \notin A_i \end{cases}$$

Note that once the χ_i are known, v is completely determined. Thus we will have

$$\begin{aligned} 0 = H(v \mid (\chi_i : i \in [t])) \\ &\geq H(v) - H((\chi_i : i \in [t])) \\ &\geq H(v) - \sum_{i=1}^{t} H(\chi_i) \\ &\geq \log n - t \end{aligned}$$

■

In other words, the random variable v has n bits of information and each χ_i has at most 1 bit of information. Thus, t must be at least $\log n$. The work [13] contains a nice proof of a generalization of the latter fact due to Hansel([23]), which is given below:

Theorem 9. *One needs at least $t \geq \log n$ bipartite graphs $G_1, G_2, \ldots, G_t$ with vertex set $[n]$, such that $G_1 \cap G_2 \cap \ldots \cap G_t = K_n$. If $size(G_i)$ stands for the number of non-isolated vertices of G_i then*

$$\sum_{i=1}^{t} \frac{size(G_i)}{n} \geq \log n$$

Graph Entropy

Since graphs are such a central term in Combinatorics, we finish the section by giving the definition of graph's entropy in a sense of Körner. Around 1970 (see [25]), he considered the following problem. Assume that we have discrete, memoryless and stationary information source over a finite set of symbols V associated with a set of vertices of a graph G. The latter means that we have a probability distribution $\mathbb{P}$ on V and at any given time a symbol v is emmited with probability $\mathbb{P}(v)$. What is the role of the edges of G? The symbols in V are not all distinguishable and some of them may be confused with one another. The edges represent these dependancies. Two symbols correspond to two adjacent nodes of G if and only if they are distinguishable. In fact, we have the same settings here like in Section II, except that now we put an edge if the two symbols are not confusable, as opose to II. The task is again to come up with the most efficient encoding of the information emitted. However, here, we have source emitting symbols according to an underlying distribution $\mathbb{P}$. This encoding should be a mapping from the strings of some fixed length t, consisted of the emitted letters, to a finite number of other fixed length strings of some other alphabet called codewords. The performance of our encoding is measured by its rate defined by the ratio $\frac{\log M}{t}$, where M stands for the number of different codewords we are using. Smaller rate means better performance, unlike the usual channel coding problem. Recall that by $G^{\boxtimes t}$ we denote the composition of t strong products of the graph G with itself(see Section II). Notice that if the edges in G describe distinguishability of letters according to our source then the edges describe distinguishability of the t-length sequences of these letters by the same source. One may also observe that if we take any set $U \subseteq V(G^{\boxtimes t})$ then the most efficient encoding of all the sequences in U requires $\chi(G^{\boxtimes t}(U))$ codewords - the chromatic number of the induced subragraph of $G^{\boxtimes t}$ on U. This leads to the original definition of graph entropy:

Definition 1 (Graph Entropy)**.**

$$H(G, \mathbb{P}) = \lim_{t\to\infty} \min_{\substack{U \subseteq V(G^{\boxtimes t}), \\ \mathbb{P}^t(U) > 1-\epsilon}} \frac{1}{t} \log \chi(G^{\boxtimes t}(U))$$

In the definition above, $\mathbb{P}^t(U) = \sum_{x\in U} \mathbb{P}^t(x)$, where $\mathbb{P}^t(x) = \prod_{i=1}^{t} \mathbb{P}(x_i)$, $x = x_1 x_2 \dots x_t$. Körner also gave another possiblle definition of graph entropy, as complexity measures of a graph, and proved the equivalence of the two. A good quality survey on graph entropy was written by Simonyi([26]). Graph entropy has been applied to various problems occuring in diverse areas - lower bounds on the size of families of perfect hash functions([27]), lower bounds for Boolean formula size ([28],[29]), algorithms for sorting partially ordered sets([30]) and even biology, chemistry and sociology(see [24] for a survey on these applications).

Summary

Most of the listed examples and several others for Entropy usage in enumeration problems can be found in the survey of Radhakrishnan([13]).

IV. Entropy- related questions in Additive Combinatorics

The last group of results we report is related to a "child" of "Number Theory" called "Additive Combinatorics". This particular area deals with problems where one is interested in the properties of the additive structure of a certain group(i.e. $\mathbb{N}$, $\mathbb{R}$, the set of real r.v., etc.). A classical book in Additive Combinatorics is the one of Tao and Vu ([35]). Below are given a few of the recent and easily comprehensible results:

Theorem 10. *If X_i are independent $\mathbb{R}^d$-valued random vectors, then*

$$H(X_1 + X_2) + H(X_2 + X_3) \geq H(X_1 + X_2 + X_3) + H(X_2)$$

Surprisingly, this simply result was claimed to be proved for the first time in 2008 by Madiman in [32]. Note that this statement is equivalent to $I(X; X+Y) \geq I(X; X+Y+Z)$. Another interesting fact, but with far more complicated proof, is the one below:

Theorem 11. *Let X, Y are i.i.d. random variables. Then*

$$\frac{1}{2} \leq \frac{H(X+Y)-H(X)}{H(X-Y)-H(X)} \leq 2$$

The upper bound was proved by Madiman[32] and the lower bound was proved independently by Ruzsa([34]) and Terence Tao([36]). Except their theoretical significance, it turns out that the class of results that the forementioned belong has also some applications. For instance in polar coding([33], 16') and for obtaining bounds of determinants of positive semi-definite-matrices([32], 08'). The cited articles, especially [32], [33] are good sources if one wants to become familiar with the topic.

V. Conclusion

Interesting results part of the interplay between Information Theory and Combinatorics were presented. They have diverse nature and applications and are object of work of some extraordinary mathematicians (e.g. Terry Tao, Noga Alon) which is and evidence of their importance. One may be confident that even more suprising, but applicable and beautiful results in this intersection of fields will be obtained in the future!

References

[1] J. Korner and A. Orlitsky, *Zeo-error information theory*, IEEE Transactions on Information Theory, 1998.

[2] N. Alon, *Combinatorial Reasoning in Information theory*, Online, Available: https://www.youtube.com/watch?v=mi7IM7tQvVY&index=2&list=PLdK7ayLCJPvVx0HZ7-nxgwTn0vzIkUV_v

[3] N. Alon, *Combinatorial Reasoning in Information theory*, IEEE Information Theory Society Newsletter, 2009.

[4] C. Shannon, *The zero error capacity of a noisy channel*, IRE Transactions on Information Theory., 1956

[5] C. Shannon, *A mathematical theory of communication*, Bell System Technical Journal. 27 (3): 379423., 1948

[6] L. Lovász, *On the Shannon Capacity of a Graph*, IEEE Transactions on Information Theory, 1979.

[7] M. Jurkiewicz, *A survey on known values and bounds on the Shannon capacity*, Selected Topics in Modern Mathematics Edition, 2014.

[8] M. Jurkiewicz, M. Kubale, K. Turowski *Some lower bounds on the Shannon capacity*, Journal of Applied Computer Science, 2014.

[9] P. Erdös, P. Ko, R. Rado *Intersection theorems for systems of finite sets*, Quart. J. Math. Oxford Ser. (2), Vol. 12, 1961.

[10] F. Bekius, *The Shannon capacity of a graph*, Bachelor Thesis, 2011.

[11] A. Vesel, J. Zerovnik, *Improved lower bound on the Shannon capacity of $C)7$*, Inf. Process. Lett., 2002.

[12] K. Mathew, P. Östergard, *New lower bounds for the Shannon capacity of odd cycles*, Designs, Codes and Cryptography., 2015.

[13] J. Radhakrishnan, *Entropy and Counting*, Computational Mathematics,Modelling and Algorithms , Narosa Publishers, New Delhi, 2001.

[14] J. Radhakrishnan, *An entropy proof of Bregman's theorem*, Journal of Combinatorial Theory, 1997.

[15] L. Bregman, *Some properties of nonnegative matrices and their permanents*, Soviet Mathematics Doklady, 1973.

[16] F. Chung, R. Graham, P. Frankl, J. Shearer *Some intersection theorems for ordered sets and graphs*, Journal of Combinatorial Theory, 1986.

[17] E. Freidgut, J. Kahn *On the number of copies of one hypergraph in another*, Israel Journal of Mathematics, 1998.

[18] Z. Furedi, *Scrambling permutations and entropy of hypergraphs*, Random Structures and Hypergraphs, 1996.

[19] J. Kahn *An entropy approach to the hardcore model on bipartite graphs*, Probability and Computing, 2001.

[20] J. Kahn *Entropy, independent sets and antichains*, Proc.Amer.Math.Soc., 2001.

[21] N. Bansal, R. Pendavingh, J. Van dr Pol *An entropy argument for counting matroids*, Journal of Combinatorial Theory, 2012.

[22] M. Madiman, P. Tetali *Information inequalities for joint distributions, with Interpretations and Applications*, IEEE Transactions on Information Theory, 2007.

[23] G. Hansel, P. Tetali *Nombre minimal de contacts de fermature n ecessaires pour r ealiser une fonction bool eenne sym etrique de n variables.*, C.R.Acad.Sci. Paris, 1964.

[24] M. Dehmer, A. Mowshowitz *A history of graph entropy measures*, Information Sciences, 2011.

[25] J. Körner *Coding of an information source having ambiguous alphabet and the entropy of graphs*, Transactions of the 6th Prague Conference on Information Theory, 1971.

[26] G. Simonyi *Perfect graphs and graph entropy. An updated survey.*, Information Sciences, 2001.

[27] J. Körner, K. Marton *New bounds for perfect hashing via information theory*, European Journal of Combinatorics, 1988.

[28] J. Radhakrishnan, $\sum\prod$ *threshold formulas*, Combinatorica, 1994.

[29] J. Radhakrishnan, *Better lower bounds for monotone threshold formulas*, Journal of Computer and System Sciences, 1997.

[30] J. Kahn, J. Kim *Entropy and sorting*, Journal of Computer and System Sciences, 1995.

[31] M. Madiman, *Information Theory in Complexity Theory and Combinatorics*, Online Talk, 2015, Available: https://www.youtube.com/watch?v=aWNKLUZJ-yM

[32] M. Madiman, *On the Entropy of Sums*, Proc. IEEE Inform. Theory Workshop, 2008

[33] E. Abbe, J. Li, M. Madiman, *Entropies of weighted sums and applications to polar codes*, 2016, Available:https://arxiv.org/pdf/1512.00135v4.pdf
[34] I. Ruzsa, *Entropy and Sumsets*, Random Struct. Alg., 2009
[35] T. Tao, V. Vu, *Additive Combinatorics*, ser. Cambridge studies in Advanced Mathematics, Cambridge University Press, 2006
[36] T. Tao, *Sumset and inverse sumset theory for Shannon entopy*, Combin. Probab. Comput., 2010