Combinatorial Reasoning in Information Theory

Noga Alon*

Abstract

Combinatorial techniques play a crucial role in the investigation of problems in Information Theory. We describe a few representative examples, focusing on the tools applied, and mentioning several open problems.

1 Introduction

Combinatorial ideas play a prominent role in the study of problems in Information theory. Indeed, the whole theory can be developed using a combinatorial approach, as done, for example, in [12]. In this brief survey we discuss several examples in which tools from Combinatorics and Graph Theory are applied in the investigation of problems in Information Theory. The combinatorial approach seems especially powerful for tackling problems in zero-error information theory which deals with scenarios in which no positive probability of error is tolerated. Problems of this type are discussed in a significant number of papers starting with [23], and are also the focus of the present short paper. This is not meant to be a comprehensive treatment of the subject, but hopefully provides an interesting description of several intriguing information theoretic results obtained by combinatorial reasoning.

2 The Shannon Capacity of graphs

For an undirected graph $G = (V, E)$, let $G^{\wedge n}$ denote the graph whose vertex set is V^n in which two distinct vertices $(u_1, u_2, \ldots, u_n)$ and $(v_1, v_2, \ldots, v_n)$ are adjacent iff for all i between 1 and n either $u_i = v_i$ or $u_i v_i \in E$. The *Shannon capacity* $c(G)$ of G is the limit $\lim_{n\to\infty}(\alpha(G^{\wedge n}))^{1/n}$, where $\alpha(G^{\wedge n})$ is the maximum size of an independent set of vertices in $G^{\wedge n}$. This limit exists, by super-multiplicativity, and it is always at least $\alpha(G)$. (It is worth noting that it is sometimes customary to call $\log c(G)$ the Shannon capacity of G, but we prefer to use here the above definition, following Lovász [19].)

The study of this parameter was introduced by Shannon in [23], motivated by a question in Information Theory. Indeed, if V is the set of all possible letters a channel can transmit in one

*Sackler School of Mathematics and Blavatnik School of Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, 69978, Israel. Email: nogaa@tau.ac.il. Research supported in part by an ERC advanced grant, by a USA-Israeli BSF grant, by the Israel Science Foundation and by the Hermann Minkowski Minerva Center for Geometry at Tel Aviv University.

use, and two letters are adjacent if they may be confused, then $\alpha(G^n)$ is the maximum number of messages that can be transmitted in n uses of the channel with no danger of confusion. Thus $c(G)$ represents the number of distinct messages *per use* the channel can communicate with no error while used many times.

There are several known upper bounds for the Shannon capacity of a graph. The most effective one is a geometric bound proved in [19], which is called the Lovász θ-function. Other bounds appear in [23], [16], [1].

The *(disjoint) union* of two graphs G and H, denoted $G+H$, is the graph whose vertex set is the disjoint union of the vertex sets of G and of H and whose edge set is the (disjoint) union of the edge sets of G and H. If G and H are graphs of two channels, then their union represents the *sum* of the channels corresponding to the situation where either one of the two channels may be used, a new choice being made for each transmitted letter.

Shannon [23] proved that for every G and H, $c(G+H) \geq c(G)+c(H)$ and that equality holds if the vertex set of one of the graphs, say G, can be covered by $\alpha(G)$ cliques. He conjectured that in fact equality always holds. In [1] it is proved that this is false in the following strong sense.

Theorem 2.1. *For every k there is a graph G so that the Shannon capacity of the graph and that of its complement $\overline{G}$ satisfy $c(G) \leq k, c(\overline{G}) \leq k$, whereas $c(G+\overline{G}) \geq k^{(1+o(1))\frac{\log k}{8 \log \log k}}$ and the $o(1)$-term tends to zero as k tends to infinity.*

The proof, which contains an explicit description of G, is based on some of the ideas of Frankl and Wilson [14], together with the basic approach of [16] and [2]. The main idea is to prove an algebraic upper bound for the Shannon capacity of a graph in terms of the dimension of an appropriately defined space of multivariate polynomials, and use this bound with polynomials over a field of one characteristic for the graph, and over a field of another characteristic for its complement. As shown in [1], the idea of using different fields is crucial here, and one cannot deduce the result using other known bounds like the θ-function.

The above counter-intuitive example is extended in [7], where it is shown that for every $\mathcal{F}$, a family of subsets of $[t]$, it is possible to assign a channel $\mathcal{C}_i$ to each sender $i \in [t]$, such that the capacity of a group of senders $X \subset [t]$ is high iff X contains some $F \in \mathcal{F}$. This corresponds to a case where only privileged subsets of senders are allowed to transmit in a high rate. The basic approach in the proof is similar to the algebraic one in [1], but requires some additional combinatorial arguments.

The behavior of the Shannon capacity of graphs in general is far from being well understood. Even the capacity of small, simple graphs, like the cycle of length 7, is not known (see [11] and some of its references for the known estimates on the capacity of odd cycles of length exceeding 5). As shown in [6], the sequence $(\alpha(G^{\wedge n}))^{1/n}$, whose limit is the Shannon capacity $c(G)$, can be very complicated, exhibiting a large number of jumps. It is not known if the maximum possible value of the Shannon capacity of a graph whose independence number is 2 is bounded by an absolute constant. This is equivalent to a well studied Ramsey-theoretic question about the asymptotic behavior of the maximum possible number of vertices in a complete graph whose edges can be colored by k colors with no monochromatic triangle (see [13], [3]).

Another interesting open problem is whether for every $\varepsilon > 0$ and every $n > n_0(\varepsilon)$ there exists

a graph G on n vertices satisfying $\alpha(G) < n^{\varepsilon}$ and $c(G) > n^{1-\varepsilon}$. The asymptotic behavior of the expected value of the Shannon capacity of the random graph $G(n, 0.5)$, which is known to be at most $O(\sqrt{n})$, as shown in [17], and at least $\Omega(\log n)$, as this is the typical independence number, is also open, and it seems plausible to conjecture that it is $\Theta(\log n)$. The maximum possible value of the Shannon capacity of the disjoint union of two graphs, each of capacity k is also unknown. This maximum is at least $k^{\Omega(\log k/\log\log k)}$, by the results of [1] mentioned above, but it is not even known that it is bounded by any function of k. Finally, it is not known if the problem of deciding whether the Shannon capacity of a given input graph exceeds a given value is decidable.

3 Multiple instances

There are several examples of communication problems in which the number of bits that have to be transmitted per instance for multiple independent instances decreases dramatically as the number of instances increases. Several examples of this type are given in [3], based on properties of the so-called Witsenhausen rate of a graph, introduced in [25]. Here we describe a few more recent examples.

3.1 Mail order and multiple product lines

In this short subsection we describe a simple yet intriguing result from [22], where the author applies the existence of expanders, which are sparse pseudo-random graphs, to describe an interesting result in which communication can be saved when several independent tasks are combined. For simplicity we describe only a very special case which captures the essence of the argument.

The problem was first considered in [24], its worst case version is analyzed in [22]. Consider a mail-order firm that sells m-different shirts. Assume, further, that each potential customer is interested in ℓ of the m shirts, and wants to get one of them, having no preference between those ℓ. The firm is interested in the minimum number of bits it should get from a customer, in order to be able to mail him one of the shirts he likes. We assume here that every subset of ℓ of the shirts may be the desired set of some customer. It is not difficult to see that there is a valid protocol enabling the customer to transmit only $\lceil \log_2(m-\ell+1) \rceil$ bits. Indeed, he simply sends the number of the first shirt among the first $m-\ell+1$ ones that appears in his set of desired shirts. Moreover, this is optimal. This is because if there are at most $m-\ell$ distinct transmissions that the company may get from any customer, then there are at most $m-\ell$ shirts that it can mail in response. Thus, a customer interested only in ℓ of the shirts that are not being sent in any of these responses will not get a shirt from his desired list, which is impossible, establishing the required lower bound.

Suppose, now, that the mail order firm expands into two product lines, and starts to sell pants as well. There are m kinds of pants, and each customer likes ℓ of them, and wants to get one of those he likes. Thus, each customer now wants to get one of his favorite ℓ shirts, and one of his favorite ℓ pants, where we assume no relation between the two sets. How many bits should the customer send? To be specific, consider only one representative case, when, say, $\ell = n/4$. Obviously one can use the same protocol separately for each product, sending a total of $2\lceil \log_2(m-\ell+1) \rceil$ bits. It is also clear that $\lceil \log_2(m-\ell+1) \rceil$ is a lower bound, as the communication problem for one

of the products requires that many bits. Which of these two bounds is closer to the best possible solution? Somewhat surprisingly it turns out that there is a protocol whose performance is close to the lower bound. The crucial observation here is to use expanders. In the particular example given here, we need a sparse bipartite graph with m vertices in each of its two color classes S, P, where $|S| = |P| = m$, so that between any two subsets $X \subset S$ and $Y \subset P$, with $|X| = |Y| = \ell$ $(= n/4)$, there is at least one edge. It is known (c.f., e.g., [5], Chapter 9) that there are such graphs with less than $64m$ edges. Fix such a graph, and view S as the set of shirts and P as the set of pants. If the preferred set of ℓ shirts of a customer is X and the preferred set of pants is Y, with $|X| = |Y| = \ell$, then he can simply transmit the label of an edge connecting X and Y. The number of bits required is thus less than $\lceil \log_2(64m) \rceil$, which is only the number required for one product plus a small constant number of bits.

3.2 Broadcasting with side information

The following variant of source coding, called Informed Source Coding On Demand was proposed by Birk and Kol [10]. A sender S wishes to broadcast a word $x = x_1 x_2 \dots x_n$, where $x_i \in \{0,1\}^t$ for all i, to m receivers $R_1, \dots, R_m$. Each R_j has some prior side information, consisting of some of the blocks x_i, and is interested in a single block $x_{f(j)}$. The sender wishes to transmit a codeword that will enable each and every receiver R_j to reconstruct its missing block $x_{f(j)}$ from its prior information. Let β_t denote the minimum possible length of such a binary code. The objective is to study the possible behavior of the numbers β_t for various scenarios. For simplicity we consider here only the case $t = 1$, although the case of bigger values of t, treated in [4], is also interesting.

The motivation for informed source coding is in applications such as Video on Demand. In such applications, a network, or a satellite, has to broadcast information to a set of clients. During the first transmission, each receiver misses a part of the data. Hence, each client is now interested in a different (small) part of the data, and has a prior side information, consisting of the part of the data he received [26]. Note that the assumption that each receiver is interested only in a single block is not necessary. Indeed, one can simulate a receiver interested in r blocks by r receivers, each interested in one of these blocks, and all having the same side information.

The problem above generalizes the problem of Index Coding, which was first presented in [10], and later studied in [8] and [20]. Index Coding is equivalent to a special case of the problem above in which $m = n$, $f(j) = j$ for all $j \in [m] = \{1, \dots, m\}$ and the size of the data blocks is $t = 1$.

It is natural to describe the above source coding problems in terms of a certain hypergraph. Define a *directed hypergraph* $H = (V, E)$ on the set of vertices $V = [n]$. Each vertex i of H corresponds to an input block x_i. The set E of m edges corresponds to the receivers $R_1, \dots, R_m$. For the receiver R_j, E contains a directed edge $e_j = (f(j), N(j))$, where $N(j) \subset [n]$ denotes the set of blocks which are known to receiver R_j. Clearly the structure of H captures the definition of the broadcast setting. Let $\beta_1(H)$ denote the minimal number of bits required to broadcast the information to all the receivers when the block length is $t = 1$.

We are interested in the asymptotic behavior of the number of bits that have to be transmitted when we consider parallel instances. Let $k \cdot H$ denote the disjoint union of k copies of H. Define $\beta_t^*(H) := \beta_1(t \cdot H)$. In words, β_t^* represents the minimal number of bits required if the network

topology is replicated t independent times. Such a scenario can occur when the topology is standard (resulting, for example, from using a common application or operation system). Therefore it is identical across networks, albeit with different data. A simple sub-additivity argument shows that the limit

$$\beta^*(H) := \lim_{t\to\infty} \frac{\beta_t^*(H)}{t} = \inf_t \frac{\beta_t^*(H)}{t}$$

exists.

Let $H = ([n], E)$ be a directed hypergraph for a broadcast network, and set $t = 1$. It is convenient to address the more precise notion of the *number of codewords* in a broadcast code which satisfies H. We say that $\mathcal{C}$, a broadcast code for H, is *optimal*, if it contains the minimum possible number of codewords (in which case, $\beta_1(H) = \lceil \log_2 |\mathcal{C}| \rceil$). We say that two input-strings $x, y \in \{0,1\}^n$ are confusable if there exists a receiver $e = (i, J) \in E$ such that $x_i \neq y_i$ but $x_j = y_j$ for all $j \in J$. This implies that the input-strings x, y can not be encoded with the same codeword. Let γ denote the maximal cardinality of a set of input-strings which is pairwise unconfusable. The following result, proved in [4], relates β^* and γ.

Theorem 3.1. *Let H and γ be defined as above. The following holds for any integer k:*

$$\left(\frac{2^n}{\gamma}\right)^k \leq |\mathcal{C}| \leq \left\lceil \left(\frac{2^n}{\gamma}\right)^k kn \log 2 \right\rceil$$

where $\mathcal{C}$ is an optimal code for $k \cdot H$. In particular, $\beta^*(H) = \lim_{k\to\infty} \frac{\beta_1(k\cdot H)}{k} = n - \log_2 \gamma$.

A surprising corollary of the above theorem is that β^* may be strictly smaller (and in fact even much smaller) than β_1. Indeed, as β^* deals with the case of disjoint instances, it is not intuitively clear that this should be the case: one would think that there can be no room for improving upon $\beta_1(H)$ when replicating H into t disjoint copies, given the total independence between these copies (no knowledge on blocks from other copies, independently chosen inputs). Note that even in the somewhat related Information Theoretic notion of the Shannon capacity of graphs (corresponding to channel coding rather than source coding), though, as described in the previous section, the capacity of a disjoint union may exceed the sum of the individual capacities, it is easy to verify that disjoint unions of the *same graph* can never achieve this. The following theorem demonstrates that the possible gap between $\beta_1(t \cdot H)/t$ and $\beta_1(H)$ can be very large: in fact, β^* may be bounded while β_1 is arbitrarily large:

Theorem 3.2. *There exists an explicit infinite family of broadcast networks for which $\beta^*(H) < 3$ is bounded and yet $\beta_1(H)$ is unbounded.*

The proofs combine properties of graph powers (specifically, the results of [21] and [9] on the chromatic numbers of the so called OR powers of a graph) with results about integral and fractional colorings of Cayley graphs, and about the chromatic numbers of Kneser graphs (see [18], [15]). More details can be found in [4].

4 Conclusions

Combinatorics is a powerful tool for tackling problems in Information theory. We have seen a few examples that illustrate this phenomenon; the study of the Shannon capacity of a graph, the investigation of a broadcasting problem with side information, and that of the potential merits of encoding multiple independent messages in certain scenarios.

Tools and techniques from Discrete Mathematics appear in the study of numerous additional problems in Information theory, and in particular play a crucial role in the theory of Error Correcting Codes.

The reverse direction, that is, that of applying information theoretic tools in the derivation of combinatorial results, is also a fruitful, active direction, which is not discussed here, and can be the topic of a similar article.

References

[1] N. Alon, The Shannon capacity of a union, Combinatorica 18 (1998), 301-310.

[2] N. Alon, L. Babai and H. Suzuki, Multilinear polynomials and Frankl-Ray-Chaudhuri-Wilson type intersection theorems, J. Combinatorial Theory, Ser. A **58** (1991), 165–180.

[3] N. Alon and A. Orlitsky, Repeated communication and Ramsey graphs, IEEE Transactions on Information Theory 41 (1995), 1276-1289.

[4] N. Alon, A. Hassidim, E. Lubetzky, U. Stav and A. Weinstein, Broadcasting with side information, Proc. of the 49^{th} IEEE FOCS (2008), 823-832.

[5] N. Alon and J. H. Spencer, *The Probabilistic Method, Third Edition*, Wiley, 2008, xv+352 pp.

[6] N. Alon and E. Lubetzky, The Shannon capacity of a graph and the independence numbers of its powers, IEEE Transactions on Information Theory 52 (2006), 2172-2176.

[7] N. Alon and E. Lubetzky, Privileged users in zero-error transmission over a noisy channel, Combinatorica 27 (2007), 737–743.

[8] Z. Bar-Yossef, Y. Birk, T.S. Jayram and T. Kol, Index coding with side information, Proc. of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006), pp. 197-206.

[9] C. Berge and M. Simonovits, The coloring numbers of direct product of two hypergraphs, In: C. Berge and D. Ray-Chaudhuri, editors, Hypergraph Seminar, Lecture Notes on Mathematics, # 411. Springer Verlag, 1974.

[10] Y. Birk and T. Kol, Coding-on-demand by an informed source (ISCOD) for efficient broadcast of different supplemental data to caching clients, IEEE Transactions on Information Theory 52 (2006), 2825-2830. An earlier version appeared in INFOCOM '98.

[11] T. Bohman, A limit theorem for the Shannon capacities of odd cycles II, Proc. Amer. Math. Soc. 133 (2005), 537–543.

[12] I. Csiszár and J. Körner, *Information theory: Coding theorems for discrete memoryless systems, third edition*, Akadémiai Kiadó, Budapest, 199?, xii+452.

[13] P. Erdős, R.J. McEliece, and H. Taylor, Ramsey bounds for graph products, *Pacific Journal of Mathematics*, 37(1):45–46, 1971.

[14] P. Frankl and R. Wilson, Intersection theorems with geometric consequences, Combinatorica **1** (1981), 357–368.

[15] J. E. Greene, A new short proof of Kneser's conjecture, Amer. Math. Monthly 109 (2002), 918–920.

[16] W. Haemers, An upper bound for the Shannon capacity of a graph, Colloq. Math. Soc. János Bolyai **25**, Algebraic Methods in Graph Theory, Szeged, Hungary (1978), 267–272.

[17] F. Juhasz, The asymptotic behaviour of Lovasz' θ function for random graphs, Combinatorica 2 (1982), 153-155.

[18] L. Lovász, Kneser's conjecture, chromatic number and homotopy, *Journal of Combinatorial Theory*, 25 (1978), 319–324.

[19] L. Lovász, On the Shannon capacity of a graph, IEEE Trans. Inform. Theory **25** (1979), 1–7.

[20] E. Lubetzky and U. Stav, Non-linear index coding outperforming the linear optimum, Proc. of the 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2007), 161-167.

[21] R.J. McEliece and E.C. Posner, Hide and seek, data storage, and entropy, The Annals of Mathematical Statistics, 42(5):1706–1716, 1971.

[22] A. Orlitsky, Scalar versus vector quantization: worst case analysis, IEEE Trans. Inform. Theory 48 (2002), 1393-1409.

[23] C. E. Shannon, The zero-error capacity of a noisy channel, IRE Transactions on Information Theory, 2(3) (1956), 8-19.

[24] D. Slepian, A. D. Wyner and J. K. Wolf, A note on specifying one of k items from a list of n items, Technical report, Bell Laboratories, 1973.

[25] H. S. Witsenhausen, The zero-error side information problem and chromatic numbers, IEEE Transactions on Information Theory, 22(5) (1976), 592-593.

[26] R. W. Yeung and Z. Zhang, Distributed source coding for satellite communications, IEEE Trans. Inform. Theory, vol. IT-45, pp. 1111–1120, 1999.