Trifference

János Körner* **Gábor Simonyi**
Mathematical Institute of HAS, 1364 Budapest, POB 127, Hungary.

Abstract

To distinguish n objects, we can label them by n binary sequences of length $\lceil \log_2 n \rceil$ each. Shorter sequences would not do. How about *tristinguishing* n objects? In this problem we use ternary sequences for labeling and require that any three of these be different in one and the same coordinate. This is the simplest unsolved case of a problem known as perfect hashing. We give a non-existence bound for a similar problem on binary sequences. We also deal with related problems of edge–colorings in graphs. It is shown that the minimum number of tricolorings needed to give every triangle of K_n all the three colors in at least one coloring is at most $\lceil \log_2 n \rceil$.

*The manuscript was prepared while this author was visiting the IASI–CNR, Rome, Italy.

1 Introduction

To distinguish n objects, we can address them by n binary sequences of length $\lceil \log n \rceil$ each. Shorter sequences would not do. (Notice that here and in the sequel all log's and exp's are binary.) From this trivial observation a surprisingly short way takes us to a hard unsolved combinatorial problem that emerges in several important models of computer science. More importantly, we will try to show that this problem represents a stumbling block whose "removal" might lead to a spectacular extension of the information–theoretic approach to extremal set theory from the case of graphs to hypergraphs.

Subsets or bipartitions of an n–set can be represented by binary sequences of length n. Likewise, k-partitions of an n–set can be represented by sequences over a k–ary alphabet. It was shown in the papers [6] and [7] that many problems in combinatorics regarding subsets or partitions of a set can be reformulated within a common information–theoretic framework in which the key notion is for sequences to be *really different* in a particular way pertinent to the problem. Let us consider a graph having as vertex set our k–ary alphabet. Considering two k–ary sequences *really different* if in some coordinate they differ in two elements of the alphabet which are the two endpoints of an edge of G leads us to the problem of Shannon capacity of the graph, [14]. In this problem, we ask how long each of n sequences over a k–ary alphabet must be if they have to be "really different" in the previous sense. This minimum length is easily shown to be asymptotically of the order $c \log n$ where the constant c is a characteristic of the graph. In fact, if we denote the above minimum length of the sequences by $l(G, n)$, then we can define the (logarithmic) Shannon capacity of the graph G as the always existing limit

$$\lim_{n \to \infty} \frac{\log n}{l(G, n)}.$$

In [2] Cohen, Körner and Simonyi extend this definition to families of graphs. Rather than requiring the occurrence of an edge of a fixed graph between sequences, they require the existence of an edge of any graph from a fixed family $\mathcal{G}$. They call the corresponding notion of capacity the Shannon capacity of the family of graphs $\mathcal{G}$. In [5] L. Gargano, J. Körner and U. Vaccaro introduce a further extension of this definition. Instead of simple graphs they consider directed graphs. Then, in case of a single digraph, the notion of really different requires, between any pairs of sequences the existence of arcs of the graph with opposite orientations. The corresponding analogon of Shannon capacity is called Sperner capacity.

Our paper [10] gives a simple example of the relevance of this kind of notions to extremal set theory. Subsequently, Gargano, Körner and Vaccaro have shown ([6], [7]) that the concept of Sperner capacity of a family of graphs offers a formally information–theoretic framework to treat and solve many interesting and even some long–standing open problems in extremal combinatorics in an asymptotic sense. These include various generalizations of Sperner's classical theorem on the maximum number of subsets of an n–set without one containing the other and the solution of Rényi's 1970 problem on the maximum number of pairwise qualitatively 2–independent k–partitions of an n–set, [13]. Beyond the above papers the interested reader is advised to consult [11] and [1] where the problem of computing the Shannon capacity and the Sperner capacity of a single graph are addressed.

If it is true that the Sperner capacity framework encompasses a great many combinatorial problems, we soon have to add that much more problems are left outside its scope. In fact, many problems in extremal combinatorics require more structure than what is offered by the framework of pairwise comparison of sequences. Whatever complicated notion of being really different we might come up with, it would not help; to formulate more problems in our language, we have to invoke some comparison of three or more sequences. Formally, this amounts to extend the investigation of capacities from graphs to hypergraphs, [9]. Of all such problems one is standing out. This is the problem of *trifference* discussed in the next section. We dare say that it is the conceptually simplest and more natural of them all. In one way or the other, solving it would shed light on the rest.

In this paper we just want to present some results concerning trifferentiating objects in some restricted manner. Some of these related problems are defined on graphs. Instead of trifferentiating any triple of elements of an n-set we might want to trifferentiate just a particular subset of these. Problems of the latter kind bring us closer to the interesting topic of anti–Ramsey theorems in the sense of [15]. A typical problem in anti–Ramsey theorems is to ask how many colors are needed to color the edges of K_m so that any three edges that form a triangle obtain different colors. (In fact, the answer to this question is trivial, but problems of this kind soon get complicated, cf. [15].) Reversing this question, Vera T. Sós asked how many tricolorings of the edges of K_m are needed if the edges of every triangle have to get three different colors in at least one of them. Setting $n = \binom{m}{2}$ Vera Sós' question can be reformulated in our language as follows. What is the minimum length of ternary sequences we have to use for labeling in order to assign trifferent labels to any triple of edges forming a triangle in K_m? Similar questions can be asked about three edges forming other subgraphs.

Problems of tristinction are strongly connected to Rényi's still unsolved question of qualitatively 3–independent partitions of an n–set and some other problems in combinatorics which keep coming back under different disguises.

2 Perfect hashing, trifference and quasi-trifference

Perfect hashing is a purely combinatorial model for the hashing problem in computer science. Its history and importance can be best understood from the paper of A. Yao [16]. We shall adopt the terminology of Fredman and Komlós [4].

Definition 1 *A family of b–partitions of a set X is called a (b, k)–system of perfect hash functions if every k–element subset of X meets k different classes of at least one of the* *partitions in the family. We denote by $Y(b,k,n)$ the minimum number of partitions in any (b,k)–system for a set of n elements. For given b and k set*

$$F(b,k) = \liminf_{n\to\infty} \frac{Y(b,k,n)}{\log n}.$$

(Notice that $F(b,k)$ is the reciprocal of the capacity of a particular uniform hypergraph in the sense of [9]). The exact value of $F(b,k)$ is unknown for $b \geq k > 2$. The best available bounds are due to Fredman and Komlós [4] and Körner and Marton [8], cf. also [9]. In said papers rather sophisticated information–theoretic proof techniques are used to obtain lower bounds. In exchange, in [4] the upper bound is derived using plain random selection. In [8] and [3], independently, this upper bound was improved in the case $b = k = 3$, thus showing that random selection gives rather poor results for this problem. In this paper we concentrate on problems related to this particular case which we like to call the problem of trifference. We recall the corresponding bounds available in the literature.

Körner and Marton [8] have proved

$$\frac{1}{\log\frac{3}{2}} \leq F(3,3) \leq \frac{4}{\log\frac{9}{5}}. \tag{1}$$

Numerically, this means that

$$1.709 \leq F(3,3) \leq 4.717$$

The upper bound is implicit also in [3].

Now we consider the following related problem.

Definition 2 *We call the binary sequences $\mathbf{x} = x_1, x_2, \ldots, x_t$, $\mathbf{y} = y_1, y_2, \ldots, y_t$ and $\mathbf{z} = z_1, z_2, \ldots, z_t$ quasi–trifferent if there exists a coordinate $1 \leq i \leq t-1$ for which the ordered pairs (x_i, x_{i+1}), (y_i, y_{i+1}), (z_i, z_{i+1}) are all different.*

Let $Y_2(2,3,n)$ be the minimum number t for which there exist n binary sequences of length t such that every three of them are quasi–trifferent. Set

$$F_2(2,3) = \liminf_{n\to\infty} \frac{Y_2(2,3,n)}{\log n}.$$

(Note that any three pairwise different binary sequences are trifferent at *some* two coordinates, i.e., if there is no restriction on the (relative) location of these two coordinates.)

The result of this section is the following

Theorem 1

$$F_2(2,3) \geq 2$$

Proof.

Let D be a quasi-trifferent set of t-length binary sequences, i.e., any three sequences in D are quasi-trifferent. Let $B = \{0,1\}^{\lfloor t/2 \rfloor}$ and for each $y \in B$ let $A(y)$ denote the set of all t-length binary sequences the even numbered coordinates of which form y.

Now we use double counting for the pairs $(x, A(y))$ where $x \in D$, $y \in B$ and $x \in A(y)$. Since every $x \in D$ uniquely determines the corresponding $A(y)$, obviously $|\{(x, A(y)) : x \in A(y), x \in D\}| = |D|$. On the other hand, $|A(y) \cap D| \leq 2$ for any fixed $A(y)$ since three binary sequences that coincide at every even-numbered coordinate could not be quasi-trifferent. This implies

$$|\{(x, A(y)) : x \in D, x \in A(y)\}| \leq 2|\{A(y) : y \in \{0,1\}^{\lfloor t/2 \rfloor}\}| = 2 \cdot 2^{\lfloor t/2 \rfloor}.$$

Combining this inequality with the previous equality we get $|D| \leq 2 \cdot 2^{\lfloor t/2 \rfloor}$. This implies $n \leq 2 \cdot 2^{\lfloor Y_2(2,3,n)/2 \rfloor}$ and thus the theorem follows. □

Tedious calculations give an upper bound by random choice that we omit here because of its irrelevance.

3 Tricolored triangles

Let K_n denote the complete graph on n vertices. An edge–tricoloring of K_n is a partition of the edge set of K_n into three different classes. We refer to the members of these respective classes as red, blue and green edges, respectively. We call a triangle edge–tricolored (ET) if all its edges are colored differently. Let $t(n)$ denote the minimum number of edge–tricolorings needed to make every triangle ET in at least one of them. Write

$$T = \liminf_{n \to \infty} \frac{t(n)}{\log n}.$$

Determining T seems hard and our lower and upper bounds are far apart. The lower bound is trivial. The main interest of the next result is that the upper bound is obtained via an explicit construction for this does not seem to happen frequently with similar problems.

Theorem 2

$$\frac{1}{\log 3} \leq T \leq 1.$$

Proof.

The lower bound is trivial. In fact, fix any vertex and look at all the adjacent edges of which there are $n-1$. Clearly, any two of them must have differing colors in at least one tricoloring. This gives

$$\frac{\log(n-1)}{\log 3} \leq t(n).$$

To prove the upper bound, assign to every node of K_n a different binary sequence of length $\lceil \log n \rceil$. We will define $\lceil \log n \rceil$ edge–tricolorings of K_n through these sequences. Let us look at the edge having the different vertices a and b as endpoints. Let $\mathbf{x} = x_1, x_2, \dots, x_t$ and $\mathbf{y} = y_1, y_2, \dots, y_t$ be the corresponding binary sequences. Define the i'th tricoloring of (a, b) as follows.

Let (a, b) be blue if $x_i = y_i$.

Let (a, b) be green if $x_i \neq y_i$ but $x_j = y_j$ for all $j < i$.

Let (a, b) be red else.

Let us say that the i'th coordinate *cuts the edge* (a, b) if the i'th coordinates of the sequences assigned to a resp. b are different, i. e., if $x_i \neq y_i$. We claim that the edges of every triangle get 3 different colors in a coloring in some coordinate $i > 1$. To prove this, notice that every edge of K_n is cut in some coordinate and that a coordinate cutting any edge of a triangle will cut exactly two of them. From these two observations it follows that in every triangle at least two different pairs of edges are cut in some coordinate.

Now fix a triangle and consider the smallest coordinate i for which all the edges of the triangle are cut in some coordinate with $j \leq i$. This means by the foregoing that in this coordinate i there is an edge cut for the first time and therefore never cut in a coordinate $j < i$. Notice, however, that the triangle cannot have more than one edge with these properties, for some pair of edges had to be cut before. Furthermore, there is an edge that is not cut in the i'th coordinate. This proves that our triangle has tricolored edges in this coordinate. Thus

$$t(n) \leq \lceil \log n \rceil - 1.$$

□

It seems unlikely that the lower bound be tight. In this context, it is worth noticing what happens if we just want to bicolor every pair of adjacent edges. More precisely, let $u(n)$ be the minimum number of edge–tricolorings needed to make every pair of adjacent edges of K_n bicolored in at least one of the tricolorings. Write

$$U = \liminf_{n \to \infty} \frac{u(n)}{\log n}.$$

We have

Proposition 1

$$U = \frac{1}{\log 3}.$$

Proof.

The lower bound is true by the same argument as in Theorem 1. To prove the upper bound label every vertex of K_n by a different ternary sequence of length $\lceil \frac{\log n}{\log 3} \rceil$. Next label every edge by the modulo 3 sum of the ternary vectors assigned to its two endpoints. The i'th coordinates of all these vectors give rise to the i'th edge–coloring in the obvious way. It is immediate that this family of colorings satisfies our condition. □

4 Other tricolored subgraphs

It follows from our previous observations that substantially more tricolorations of the edges of K_n are needed to tristinguish any triple of edges of the complete graph on n vertices than to tristinguish just the three edges of a triangle. In fact, by definition, the number of tricolorings needed to tristinguish every triple of the edges is $Y(3,3,\binom{n}{2})$ which is about $2F(3,3)\log n \geq 3.4\log n$ while we have seen that for the tristinction of the three edges of any triangle we need not more than about $\log n$ tricolorings of the edges. It is therefore interesting to understand what happens if we want to tristinguish the three edges of some other 3–edge subgraphs of K_n by the minimum number of 3–colorings of the edges of this complete graph. In particular, it is interesting to see whether there is a single type of subgraph which in itself is responsible for the total number of colorings needed to tristinguish all the edge triples of K_n, in the asymptotic sense.

Let $s(n)$ denote the minimum number of tricolorings of the edges of K_n needed to make every tristar edge–tricolored in at least one of them. (Here a tristar is a graph on 4 points with 3 edges all of which have a common endpoint. Just as for triangles, we say that a tristar is edge–tricolored if all its edges are colored differently.) Write

$$S = \liminf_{n\to\infty} \frac{s(n)}{\log n}.$$

We have

Proposition 2

$$S = F(3,3)$$

Proof.

The lower bound $F(3,3) \leq S$ follows from the fact that if we want to tristinguish just the $n-1$ edges meeting in a single fixed vertex of K_n, then this is equivalent to the problem of trifference and thus $Y(3,3,n-1) \leq s(n)$.

To prove the upper bound label every vertex of K_n by a trifferent ternary sequence of length $Y(3,3,n)$. Then label every edge by the modulo 3 sum of the ternary vectors assigned to its two endpoints. □

We will call trident a graph on 6 vertices with three vertex–disjoint edges. Let $r(n)$ denote the minimum number of tricolorings of the edges of K_n needed to tristinguish the three edges of any trident in at least one of them. Write further

$$R = \liminf_{n\to\infty} \frac{r(n)}{\log n}.$$

We claim that

Proposition 3

$$R = F(3,3)$$

Proof.

The lower bound $F(3,3) \leq R$ is obvious. In fact, notice that a maximal matching of the graph consists of $\lfloor \frac{n}{2} \rfloor$ pairwise vertex–disjoint edges. If we only restrict ourselves to coloring these, we see that $Y(3,3,\lfloor \frac{n}{2} \rfloor) \leq r(n)$.

To prove the upper bound, let us assign trifferent ternary sequences of minimum length to each of the vertices of K_n. Next assign to every edge one of the ternary sequences assigned to its two endpoints in a completely arbitrary manner. The sequences will define the tricolorings in the obvious way establishing

$$r(n) \leq Y(3,3,n).$$

□

The last proposition is somewhat surprising for numberwise the configuration of three vertex–disjoint edges is dominant among all the configurations of three edges. The proof shows that for very general criteria for colorings of k vertex–disjoint edges the minimum number of colorings is asymptotically the same as for criteria on vertex–colorings involving arbitrary k–tuples of vertices. The same remark applies for stars with k edges.

All our above constructions share the feature that the edge–colorings are constructed from vertex–colorings in a straightforward manner. We wonder whether this is due to our lack of imagination or something more relevant to the subject.

We have failed to give non–trivial bounds for the minimum number of tricolorings needed to tristinguish the edges of the two missing subgraphs.

5 Acknowledgement

Warmfelt thanks are due to Vera T. Sós for useful discussions in a unique atmosphere. We thank Gábor Tardos for finding an error in an earlier version of our manuscript.

We are grateful to the Zentrum für interdisziplinäre Forschung of Bielefeld University for its generous hospitality.

References

[1] R. Calderbank, P. Frankl, R. L. Graham, W. Li, L. Shepp, The cyclic triangle problem, *J. Alg. Comb.*, submitted

[2] G. Cohen, J. Körner, G. Simonyi, Zero-error capacities and very different sequences, in: *Sequences. Combinatorics, Security and Transmission*, papers presented at the Advanced International Workshop on Sequences, Positano, Italy, June 1988, Springer, New York, 1990, R. M. Capocelli, ed., 144–155

[3] P. Elias, Zero Error Capacity Under List Decoding, *IEEE Trans. Inform. Theory*, 5(**34**), Part I., 1988, 1070–1074

[4] M. L. Fredman, J. Komlós, On the size of separating systems and families of perfect hash functions, *SIAM J. Alg. Disc. Meth.*, 1(**5**), 61–68

[5] L. Gargano, J. Körner, U. Vaccaro, Sperner theorems on directed graphs and qualitative independence, *J. Comb. Theory*, Ser. A, to appear

[6] L. Gargano, J. Körner, U. Vaccaro, Sperner capacities, *Graphs and Combinatorics*, to appear

[7] L. Gargano, J. Körner, U. Vaccaro, Capacities: from information theory to extremal set theory, *J. Amer. Math. Soc.*, to appear

[8] J. Körner, K. Marton, New bounds for perfect hashing via information theory, *European J. Combinatorics*, **9**(1988), 523–530

[9] J. Körner, K. Marton, On the capacity of uniform hypergraphs, *IEEE Trans. Inform. Theory*, **36**, Jan. 1990, 153–156

[10] J. Körner, G. Simonyi, A Sperner-type theorem and qualitative independence, *J. Comb. Theory*, Ser. A, **58**, 1991, 351–359

[11] L. Lovász, On the Shannon capacity of a graph, *IEEE Trans. Information Theory*, IT–**25**(1979), pp. 383–390

[12] F. J. MacWilliams, N. J. A. Sloane, The Theory of Error Correcting Codes, North Holland, Amsterdam, 1977 and New York, 1983

[13] A. Rényi, *Foundations of Probability*, Wiley, New York, 1971

[14] C. E. Shannon, The zero–error capacity of a noisy channel, *IRE Trans. on Information Theory*, **2**(1956), pp. 8–19

[15] M. Simonovits, V. T. Sós, On restricted colourings of K_n, *Combinatorica*, 1(**4**), 101–110

[16] A. C. Yao, Should tables be sorted, *J. ACM*, 28, 1981, 615–628.