Available online at www.sciencedirect.com

ScienceDirect

Electronic Notes in Discrete Mathematics 44 (2013) 23–29

Electronic Notes in DISCRETE MATHEMATICS

www.elsevier.com/locate/endm

More on Connector Families

Gérard Cohen [1]

ENST, FRANCE

Emanuela Fachini [2]

"La Sapienza" University of Rome, ITALY

János Körner [3,4]

"La Sapienza" University of Rome, ITALY

Abstract

Let $\mathcal{G}_{k,n}$ be the family of all graphs on the same n vertices each having at least k connected components. We are interested in the largest cardinality of a subfamily in which the union of any two of the member graphs has at most $k-2$ connected components, and determine its exponential asymptotics.

1 Introduction

The union of two graphs with the same vertex set is a graph on their common vertex set whose edge set is the union of the edge sets of the two. In this

[1] Email: cohen@telecom-paristech.fr
[2] Email: fachini@di.uniroma1.it
[3] Email: korner@di.uniroma1.it
[4] Department of Computer Science, University of Rome, La Sapienza, via Salaria 113, 00198 Rome, ITALY

http://dx.doi.org/10.1016/j.endm.2013.10.005

paper we continue our investigation [3] of problems from the following general framework. Let $\mathcal{F}$ and $\mathcal{D}$ be two disjoint families of graphs on the same vertex set $[n]$. We are interested in the largest cardinality $M(\mathcal{F}, \mathcal{D})$ of a subfamily $\mathcal{C} \subseteq \mathcal{F}$ for which the union of any two different member graphs of $\mathcal{C}$ is in $\mathcal{D}$.

The study of this class of problems was initiated by Messuti, Simonyi and the third author in [8]. For the most part of [8] the family $\mathcal{F} = \mathcal{F}_n$ consisted of all the Hamilton paths in K_n with various choices of $\mathcal{D}$. However, in a significant special case the resulting extremal problems have roots in nineteenth century mathematics. We will say that the graph family $\mathcal{D}$ is *monotone* if for every graph G in $\mathcal{D}$ the graph G' obtained from it by adding some new edges between its vertices, is still in $\mathcal{D}$. Now, if $\mathcal{D}$ is monotone and complementary to $\mathcal{F}$, i. e. $\mathcal{D} = \overline{\mathcal{F}}$, then the subfamily $\mathcal{G} \subseteq \mathcal{F}$ of maximum cardinality $M(\mathcal{F}, \mathcal{D})$ must consist of all the maximal graphs in $\mathcal{F}$. A graph is maximal in $\mathcal{F}$ if adding any edge to it the new graph is not in $\mathcal{F}$ anymore. This implies that the union of two different maximal graphs from $\mathcal{F}$ is already in $\overline{\mathcal{F}} = \mathcal{D}$. For the same reason, for any family $\mathcal{C} \subseteq \mathcal{F}$, we can replace its members with maximal graphs from $\mathcal{F}$ containing them. No two members of $\mathcal{C}$ will be contained in the same maximal graph because their union is outside of $\mathcal{C}$. Thus $M(\mathcal{F}, \mathcal{D}) = M(\mathcal{F}, \overline{\mathcal{F}})$ equals the number of maximal elements in $\mathcal{F}$. The study of this number, in certain special cases, goes back to Dedekind [4] and has a vast literature. More precisely, we can assign to $\mathcal{F}$ a graph $G_{\mathcal{F}}$ whose vertices are the member graphs of $\mathcal{F}$. Two vertices are adjacent in $G_{\mathcal{F}}$ if the union of the corresponding graphs is in $\overline{\mathcal{F}}$. Then $M(\mathcal{F}, \overline{\mathcal{F}})$ is the maximal order of a clique in the graph $G_{\mathcal{F}}$. For the literature of this problem, we refer the reader to [7] and the quite recent paper [5]. In these classical problems one is concerned with the asymptotic enumeration of extremal structures. If $\mathcal{F}$ is the family of all graphs of vertex set $[n]$ and $\mathcal{D}$ is monotone, then it is easily seen that

$$M(\mathcal{F}, \mathcal{D}) = |\mathcal{D}| + M(\overline{\mathcal{D}}, \mathcal{D}).$$

Our focus here is on problems of a different kind, in which the family $\mathcal{D}$ is but a small fraction of the complement of family $\mathcal{F}$. This class of problems is related to zero–error information theory, in the sense of Shannon [9].

The present paper continues our previous investigations [3] where the following problem was introduced. Let $k > l$ be two positive integers and let $\mathcal{G}_{k,n}$ be the family of all the graphs on $[n]$ having at least k connected components. Further, let $\mathcal{D}_l$ be the family of those graphs on the same vertex sets having at most l connected components. Hence $\mathcal{D}_l$ is monotone and disjoint from $\mathcal{G}_{k,n}$. We are interested in $M(\mathcal{G}_{k,n}, \mathcal{D}_l)$. It was shown in [3] that for any k and $l = 1$

one has

$$\lim_{n\to\infty} \log \sqrt[n]{M(\mathcal{G}_{k,n}, \mathcal{D}_1)} = h(1/k)$$

where $h(t) = -t\log t - (1-t)\log(1-t)$ is the binary entropy function. We will refer to this problem as that of connector families. The result just quoted deals with its "high end". We are now going to determine the exponential asymptotics of $M(\mathcal{G}_{k,n}, \mathcal{D}_l)$ for more values of k and l, concentrating on the "low end". If $l = k-1$, we have $\mathcal{D}_{k-1} = \overline{\mathcal{G}}_{k,n}$ and the easy problem of counting the number of maximal k–disconnected graphs already discussed in [3]. This is the number of k-partitions of an n-set. In conclusion, the first interesting case at the low end is

$$l = k-2,$$

the subject of the next section.

Note that logarithms and exponentials are to the base 2.

2 Results

Let $\mathcal{G}_{k,n}$ be the family of all the graphs on vertex set $[n]$ with at least k connected components and let $\mathcal{D}_l$ be that of at most l connected components. We set $l = k-2$.

Theorem 2.1 *For $k \geq 4$ we have*

$$\lim_{n\to\infty} \log \sqrt[n]{M(\mathcal{G}_{k,n}, \mathcal{D}_{k-2})} = \log k - \frac{2}{k}$$

Sketch of proof. The proof of the theorem is rather long and technical and thus it will not be included here. We limit ourselves to give some indications. As in [3], also the present problem can be reduced to one about the Shannon capacity of a family of simple graphs. First of all, one should notice that since $\mathcal{D}_{k-2}$ is a monotone graph family, there is a family $\mathcal{C} \subseteq \mathcal{G}_{k,n}$ among those satisfying our pairwise condition with maximum cardinality, whose member graphs are complements of complete k-partite graphs on $[n]$. These graphs can be represented by k-ary sequences of length n, with the whole family $\mathcal{C}$ forming a subset C of $[k]^n$. For technical reasons, we will only consider sequences having the string $12\dots k$ as a fixed prefix. Clearly, this restriction will not effect our asymptotic evaluations. In order to satisfy the pairwise union condition, any two of the elements of C should satisfy the condition that for at least 2 different unordered pairs of distinct elements of $[k]$ each of these unordered pairs is present among the n pairs of coordinates of the unordered pair of sequences. Furthermore, this condition on $C \subseteq [k]^n$ is equivalent to our original condition on $\mathcal{C}$. Thus we now have the problem of determining the exponential asymptotics of the maximum cardinality of a set $C \subseteq [k]^n$ such that for any unordered pair of its sequences there are at least 2 different unordered pairs of distinct elements of $[k]$ among their unordered pairs of coordinates. Such a problem can easily seen to be solvable in the framework of the Shannon capacity of a corresponding family of graphs [6]. To explain this framework, we need a few definitions.

Let G be a finite undirected graph with vertex set V. For every natural number n let $M(G, n)$ be the largest cardinality of a set $C \subseteq V^n$ with the property that for any unordered pair of distinct members $\{\mathbf{x}, \mathbf{y}\} \in \binom{C}{2}$ there is a coordinate $i \in [n]$ with $\{x_i, y_i\} \in E(G)$ where $E(G)$ is the set of edges of the graph. We define the always existing limit

$$Sh(G) = \lim_{n\to\infty} \log \sqrt[n]{M(G, n)}$$

and call it the Shannon capacity of the graph G.

The limit exists by Fekete's lemma [2]. A set of vertices $C \subseteq V(G)$ is said to induce a clique if any unordered pair of distinct members of C constitutes an edge in the graph. A central concept in our proof and implicit in this definition is the concept of co–normal power of a graph.

Definition 2.2 The n-th co–normal power G^n of the graph G is the graph with vertex set $[V(G)]^n$ in which two vertices, $\mathbf{x}$ and $\mathbf{y}$ are adjacent if there is a coordinate $i \in [n]$ with $\{x_i, y_i\} \in E(G)$.

Thus (the binary logarithm) of Shannon capacity is the asymptotic exponent of the growth of the largest clique in the powers of G. It should be recalled that this is the somewhat unusual formulation of the capacity of a simple graph, used in [6] and [2] for reasons explained there. In fact, this is the right formulation leading to the natural generalization of graph capacity to directed graphs. (Shannon's original problem statement–to which the present one is mathematically equivalent–is in complementary terms.)

The last basic definition we need is a generalization of Shannon capacity from single graphs to graph families.

Definition 2.3 Let $\mathcal{G}$ be a finite family of simple graphs with a common finite vertex set V. We denote by $M(\mathcal{G}, n)$ the largest cardinality of a set $C \subseteq V^n$ that induces a clique in G^n for every member graph $G \in \mathcal{G}$. We call the always existing limit

$$Sh(\mathcal{G}) = \lim_{n\to\infty} \log \sqrt[n]{M(G, n)}$$

the Shannon capacity of the graph family $\mathcal{G}$.

In order to prove our main result, we will use the main theorem of [6] and this in turn forces us to introduce more technical notions of graph capacity; capacity within a given type.

For an arbitrary sequence $\mathbf{x} \in V^n$ we shall denote by $P_{\mathbf{x}}$ the probability distribution on the elements of V defined by

$$P_{\mathbf{x}}(a) = \frac{1}{n}|\{i : \ x_i = a, \ i = 1, 2, \ldots, n\}|,$$

where $\mathbf{x} = x_1 \cdots x_n$. The distribution $P_{\mathbf{x}}$ is called the type of $\mathbf{x}$. Let $V^n(P, \epsilon)$ denote the set of those $\mathbf{x} \in V^n$ which satisfy

$$|P_{\mathbf{x}} - P| = \max_{a \in V} |P_{\mathbf{x}}(a) - P(a)| \leq \epsilon.$$

We write $V_P^n = V^n(P, 0)$. For an arbitrary directed graph G, let $M(G, P, \epsilon, n)$ be the largest cardinality of any set $C \subset V^n(P, \epsilon)$ that induces a clique in G^n. We write

$$Sh(G, P) = \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log M(G, P, \epsilon, n)$$

and call it the Shannon capacity of G within the type P. This quantity was introduced in [1]. It is clear that for every finite graph G

$$Sh(G) = \max_P Sh(G, P). \tag{1}$$

The main result of [6] is the following formula.

Theorem 2.4 [6] *For an arbitrary finite family of directed graphs $\mathcal{G}$ on the common vertex set V we have*

$$Sh(\mathcal{G}) = \max_P \min_{G \in \mathcal{G}} Sh(G, P)$$

where P is running over the probability distributions on V.

Consider now the family $\mathcal{G}'_{k,n}$ consisting of all those graphs with vertex set $[k]$ which contain at least $\binom{k}{2} - 1$ edges. It is not hard to establish by symmetry that the maximum in $\max_P \min_{G \in \mathcal{G}'_{k,n}} Sh(G, P)$ is attained for the uniform distribution U on $[k]$. Likewise, it is clear that the minimum of the value of $Sh(G, U)$ is achieved by the graphs having just one edge missing. These graphs are isomorphic and have capacity within type U equal to $\log k - \frac{2}{k}$. In conclusion, we have

$$Sh(\mathcal{G}'_{k,n}) = \min_{G \in \mathcal{G}'_{k,n}} Sh(G, U) = \log k - \frac{2}{k} \tag{2}$$

It is easy to see that $\lim_{n\to\infty} \log \sqrt[n]{M(\mathcal{G}_{k,n}, \mathcal{D}_{k-2})} = Sh(\mathcal{G}'_{k,n})$ to complete the sketch of the proof.

□

All the other problems where $l = k - c$ for some absolute constant c can be solved in a similar manner, even though already for $l = k - 3$ substantial complications arise.

3 Intersection problems

Our problem area seems close to intersection problems in the sense of [10]. Formally, an intersection problem, when formulated in complementary terms (complements of the graphs and pairwise unions instead of intersections), asks for the determination of $M(\mathcal{F}, \overline{\mathcal{F}})$. More on this relationship is explained in [3].

References

[1] I. Csiszár, J. Körner, On the capacity of the arbitrarily varying channel for maximum probability of error, *Z. Wahrscheinlichkeitstheorie verw. Geb.* **57**, (1981), pp. 87-101.

[2] I. Csiszár, J. Körner, *Information theory: coding theorems for discrete memoryless systems*, 2nd edition, Cambridge University Press, 2011.

[3] G. Cohen, E. Fachini, J. Körner, Connector families of graphs, submitted to *Graphs and Combinatorics.*

[4] R. Dedekind, Über Zerlegungen von Zahlen durch ihre grössten gemeinsamen Teiler, *Festschrift Hoch. Braunschweig u ges. Werke, II.* (1897), 103–148.

[5] R. Duffus, P. Frankl, V. Rödl, Maximal independent sets in the covering graph of the cube, *Discrete Applied Mathematics*, to appear.

[6] L. Gargano, J. Körner, U. Vaccaro, Capacities: from information theory to extremal set theory, *J. Comb. Theory* Ser. **A**, **68**(1994), no. 2, pp. 296–315.

[7] J. Kahn, Entropy, independent sets and antichains, *Proceedings of the AMS*, (**130**),2 (2001), 371–378.

[8] J. Körner, S. Messuti, G. Simonyi, Families of graph–different Hamilton paths, *SIAM J. Discrete Mathematics* **26**(2012), pp. 321-329,

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

[10] Simonovits, M. and V. T. Sós, *Intersection theorems on structures*, Annals of Discrete Mathematics, 6(1980), pp. 301–313.