Final report on the project OTKA NK 62321

The participants of the project were scientifically very active during the years 2006-2008. They did not only obtain many results, which are contained in their 158 papers appeared in strong journals, but effectively disseminated them in the scientific community. They participated and gave lectures in more than 100 conferences (with multiplicity), half of them were plenary or invited talks. Szemerédi received two important international prizes.

We are trying to give account on the complete work, but put more emphasis on the end of the period. It seems to be difficult to describe all the results, only the ones what can be shortly explained are shown. The results are grouped according to their subjects.

Traditional graph theory

The maximum number of edges in a C_6-free bipartite graph was determined in the case when one of the classes is much larger than the other one. (Győri.)

Interesting new results were obtained about connected graphs not containing paths of a certain length. They are improvements of the results of Kopylov and the old Gallai-Erdős tehorems. (Balister-Győri-Lehel-Schelp.)

It was shown that the size of a subgraph of the n-dimensional hypercube Q_n without a cycle of length 14 is of order $o(|E(Q_n)|)$. (Füredi-Özkahiya.)

A new graph parameter resembling the graph capacity is introduced. Its basic properties are studied and its value is determined for stars. We also show that this parameter is maximum for given number of edges if independent edges are chosen and minimum for the stars. (K-M-Simonyi.)

We studied the optimal representations of graphs by (small) subtrees (as vertices) of a basic tree. The problem was motivated by coding theory.

The Ramsey number is determined for 3 large odd cycles in a graph.

Continuing our earlier work using a topological approach to study the local chromatic number of several graph families including Kneser, Schrijver, generalized Mycielski, and Borsuk graphs we improved the previous estimates.

We proved that every connected, strongly regular graph contains a hamiltonian cycle if the number of vertices is large. (Pyber.)

The connection between the local chromatic number of a graph and its topological properties were studied in more papers. In the most recent one we solve the last, most difficult case. We prove that there is a graph which is topologically $2t$-chromatic, but its local chromatic number is $t+1$, but the local chromatic number of the strongly topologically $2t$-chromatic graphs is at least $t+2$. (Simonyi-Tardos-Vrećica.)

Color the vertices of a graph in such a way that the the sizes of the monochromatic connected subgraphs are the smallest possible. We prove that if the n-vertex graph is chosen from the minor-closed class of graphs, then the vertices can be colored with two colors in such a way that every connected monochromatic subgraph is of size at most $O(n^{2/3})$, and there are planar graphs in which there is a connected monochromatic subgraph with $O(n^{2/3})$ vertices for every 2-colorings. (Linial-Matoušek-Sheffet-Tardos.)

A sharpening of the famous "crossing number lemma" was found for the graphs which are very far from regularity. (Pach-Solymosi-Tardos.)

It is proved that if n points are chosen randomly (in a uniform way) and define a graph where two points are adjacent iff the minimal axis-parallel rectangle determined by them contains no other point, then its chromatic number is $\omega(\log n/\log^2 \log n)$. (Chen-Pach- B.Szegedy-Tardos.)

We determine the approximate structure of the nearly largest graphs on n vertices not containing certain "forbidden graphs" more precisely than the earlier works in this area. (Balogh-Bollobás-Simonovits.)

Let $\mathcal{G}$ be a hereditary class of graphs, in other words it is a hereditary property P. If $f(n,P)$ is the number of n-vertex non-isomorph graphs, then it is either polynomial or is at least $S(n) \approx exp(n^1/2)$, and the structure is well determined in the polynomial case. (Balogh-Bollobás-Saks-Sós.)

Random graphs, large graphs, regularity lemma

The "similarities" of large graphs were studied. We show that several different definitions of the metrics (and convergence) are equivalent. (These result are strongly related to statistical physics, regularity lemma, etc.) (Borgs-Chayes-Lovász-Sós-Vesztergombi.)

Let L be a given small graph. We proved that almost all L-free graphs are very similar to a random subgraph of an L-free extremal graph.

The following statement was proved. If a sequence of graphs has asymptotically the same distribution of small subgraphs as a generalized random graph modeled on a fixed weighted graph H, then these graphs have a structure that is asymptotically the same as the structure of H. Furthermore it suffices to require this for a finite number of subgraphs, whose number and size is bounded by a function of $|V(H)|$.

An r-partite graph has edges only among the parts. Suppose that r is fixed and n is large. The edges between two parts are chosen in a pseudorandom way, with different densities. Such a graph can be characterized by r^2 parameters what is small. If the parameters are given, n tends to infinity then the density of every ℓ-vertex F subgraph can be calculated. We proved the converse of this statement: if the densities are known for a few "test-graphs" then we can conclude that our graphs form a sequence of r-partite graphs with given densities among the parts. (Lovász-Sós.)

Lenz and Stollmann proved that for Euclidean graphs satisfying a strong ergodic property, there exists an integrated density of states in a uniform sense, too. We managed to extend this statement for the most general case what can be imagined. (Elek Func. Anal.)

Several new results like the Hereditary Property Testing, Inverse Counting Lemma and the Uniqueness of Hypergraph Limit were proved in connection with the regularity lemma. (Elek-B. Szegedy.)

We proved a conjecture of Bollobás and Riordan: Every unimodular tree can be approximated by large girth graphs. (Elek.) We also proved a regularity lemma for graphs with bounded degree. (Elek-Lippner.)

Hypergraphs, other combinatorics

Our previous work on a class of new Sperner type theorems was continued, asymptotically (first two terms) determining the maximum number of sets in a family of subsets of an n-element set where certain configurations are excluded. Beside the new results we have written a survey paper introducing the method. (Griggs-K, Carroll-K, K-survey.)

The exact maximum was determined for the number of hyperedges in a uniform hypergraph for two "excluded subhypergraphs". This is important, since very few exact results are known for excluded hypergraphs. (Füredi-Pikhurko-Simonovits.) There are some new results for the largest 3-hypergraph not containing the Fano-plane (as a 3-uniform subhypergraph).

(Füredi-Simonovits.)

The following general problem of extremal hypergraph theory has been solved. Let $a = (a_1, \ldots, a_p)$ be a sequence of positive integers, $p \geq 2$, $k = a_1 + \ldots + a_p$. An a-cluster is a family of k-sets $\{F_0, \ldots, F_p\}$ such that the sets $F_i \setminus F_0$ are pairwise disjoint $(1 \leq i \leq p)$, $|F_i \setminus F_0| = a_i$, and the sets $F_0 \setminus F_i$ are pairwise disjoint, too. Given a there is a unique a-cluster, and the sets $F_0 \setminus F_i$ form an a-partition of F_0. With an intensive use of the delta-system method we prove that for $k > p > 1$ and sufficiently large n, $(n > \ n_0(k))$, if $\mathcal{F}$ is an n-vertex k-uniform family with $|\mathcal{F}|$ exceeding the Erdős-Ko-Rado bound $\binom{n-1}{k-1}$, then $\mathcal{F}$ contains an a-cluster. The only extremal family consists of all the k-subsets containing a given element. (Füredi-Özkahiya.)

It was shown that there are excluded r-uniform hypergraphs when both the asymptotical Turán density and the Ramsey-Turán density are positive (that is the numbers are asymptotically cn^r with a positive c) but different. This is a new evidence of the fact that "excluded hypergraph" problems are very different for their "graph" counterparts. (Mubayi- Sós.)

Lovász has earlier answered a question of Toft, that the maximum number of edges of a k-uniform 3-critical hypergraph is at most $\binom{n}{k-1}$. We obtained an ordered variant of this theorem, moreover proved some generalizations (generalizing the concept of 3-criticality). The proofs of the estimates use algebraic methods. Constructions show that the estimates are asymptotically correct. (Füredi-Sali.)

We have written a survey of new and old results on subsums of a finite sum and extremal sets of vertices of the hypercube (around the Littlewood-Offord problem). (D. Miklós.)

A family of subsets is called (p, q)-chain intersecting if it contains no two chains of length p and q, respectively so that the top members are disjoint. The maximum number of members of a (p, q)-chain intersecting family was found. (Bernáth-Gerbner.) The extreme points of a very general class of families were also determined, this theorem contains most of the known cases when the extreme points were earlier determined. Moreover the extreme points of the family of non-trivially intersecting (they do not have one common point) families are also found, this is the first case when the extreme points of a non-hereditary class are determined. (Gerbner.)

The extreme points of other classes of objects were also found. These are the intersecting subspaces of a finite vector space and the intersecting families of chains of length ℓ of subsets. (Gerbner-Patkós.)

The k-Sperner systems originally were studied by Paul Erdős. A possible generalization is the notion of (k_1, k_2)-Sperner families. A set system forms such in the underlying set X with the partition $X_1 \cup X_2$ if there is no $(k_1 + 1)$-chain where all members are the same in X_2 and there is no $(k_2 + 1)$-chain where all members are the same in X_1. We determined the convex hull of this new class (which is not hard with standard methods) then we described all optimal systems for a wide range of the parameters. The main result is that all maximum systems are homogenous. (Aydinian-Czabarka-ELP-Székely.)

We continued our work in the theory of extremal matrices. The maximum number of 1's has to be determined if certain configurations (submatrices) are excluded. The problem has been solved for some new configurations. It is important to notice that there are two different types of problems. The exclusion is supposed only for the submatrix with a given order of rows and columns, or for any permutation of them. Both cases are heavily studied. Many new results were obtained in both directions. (Keszegh-Tardos, Anstee-Sali.)

Let us show one of the newest results. Suppose that the 0,1 matrix under investigation contains no identical columns. Let F_{abcd} be an $(a+b+c+d) \times 2$ matrix consisting of a rows [11], b rows [10], c rows [01], and d rows [00]. $\mathrm{forb}(m, F_{abcd})$ denotes the maximum number of columns of a $m \times n$ 0,1 matrix containing no copy of F_{abcd} where variants obtained by permuting the rows and columns are also excluded. $\mathrm{forb}(m, F_{abcd})$ has been determined for all cases, except for F_{2110}.(Anstee-Barekat-Sali.)

Embedding theorems are natural tools in constructions of designs. This is the crucial step of Wllson's celebrated result, too. We prove here that every system of edge-disjoint copies of C_4s on n vertices can be extended to a perfect C_4-packing with adding at most $\sqrt{n}$ vertices. (Füredi-Lehel.)

The work concerning the splitting properties for subsets was continued. (ELP)

A well-known theorem of Goldberg and West claims that two robbers can always equally divide the pearls of k sorts in an open necklace if the number of pearls od each sort is even and they can cut the necklace at at least k places. $k-1$ cuts are not sufficient. Now we proved that if they can cut only $k-1$ times, and this is not sufficient, then they can determine in advance who should get more pearls from which sort. Topological methods were used. (Simonyi.)

On the borderlines of combinatorics: algebra, geometry, stochastics

The following theorem was proved. Every simple, Lie-type group of characteristic p is a product of 5 p-Sylow subgroups, and 3 is not enough in general. (Pyber.)

The "cost" and the "L2-Betti number" are the most important invariants of the measurable equivalence relations. These were defined for sequences of random graphs together with some basic statements. (Elek.)

Using some algebraic methods of von Neumann we successed to prove global Cheeger inequalities for finite graphs. (Elek.)

A minimal non-amenable action was constructed which possesses both a hyper-finite and non-hyper-finite ergodic measure. This is a surprise because it was proved earlier that the ergodic measures on an amenable action are orbit equivalent. (Ceccherini- Elek.)

An equipartition lemma was proved for finite graphs containing balls of subexponential growth. A consequence is that the independence ratio can be tested within finite time with high reliability and high accuracy for such graphs. The same holds for the spectral distribution function and the edit-distribution from monotone properties. (Pyber.)

A new proof, using non-standard analysis, was given for the Rodl-Skokan-Nagle hypergraph regularity and removal lemma. This gives a relatively short proof of the famous Szemeredi theorem on arithmetic progressions.

We continued our research in convex geometry of random points. The maximum number of such points was investigated what can be ordered into a convex chain when the point are randomly chosen (independently and uniformly distributed) in a triangle. The limit form of the maximum convex chain was determined. This is a close analogue of the problem of the maximum increasing subsequence. (Bárány és trsai)

In another paper we proved that n points on the plane can be arranged along a broken line where every angle is at least 20 degrees. This problem was open for more than 10 years, no lower bound was known. (Bárány and al.)

We have written a survey paper, for invitation, about "Lattice points and random points in convex bodies". (Bárány)

New central limit theorems were proved for random polytops, for the solution we had to show some interesting properties of the so called "floating body." (Bárány-Reitzner.)

Given a set of random points. At most how many of them are in convex position. This problem is strongly related to the question of the longest monotone subsequence. We determined the most important properties like expectation, concentration, limit shape, etc. (Ambrus-Bárány.)

Some interesting problems in discrete geometry were solved by tools in algebraic geometry (Bárány -Hubard-Jeronimo, Bárány-Arocha- Bracho-Fabilla-Montajano, Bárány-BlagojevicA. Szucs.)

Take 3 curve-sets each with one parameter. Choose n curves from each of the three sets. We show that in "all normal cases" the number of points covered 3 times is at most $O(n^{2-c})$. An important special case is when one set contains the unite circles containg a fixed point.

A maximal antichain of a poset *splits* if it can be partitioned into two parts such that the downset of the first part and the upset of the second part together supply every element of the poset. We gave conditions ensuring the splitting property in infinite posets and developed a general method to construct non-splitting antichains in reach enough infinite posets. This latter statement gave a very strong generalization of the counterexample given by Ahlswede and Khachatrian. Perhaps the most interesting result is that a certain instance of the Splitting Property is equivalent to the Axiom of Choice. (PL Erdős-Soukup.)

We proved that in the posets of homomorphisms of finite graphs every finite, non-maximal antichain can be extended into so called *generalized duality* which is a maximal splitting infinite antichain. It is also true that the same antichain can be extended into an infinite, non-splitting one. (Duffus - PL. Erdős - Nešetřil - Soukup.) Further extensions are done in a series of papers.

A new, simple modern proof was found on a result of Erdős and Rogers, claiming that the n-dimensional space can be covered by tesselated copies of an arbitrary convex body with a density not more than $20n \log n$. The proof uses the Lovász Local lemma after discreditation. (Füredi-Kang.)

For all positive integers k, r a finite set of axis-parallel rectangles are constructed in the plane in a such away that coloring them by k colors in an arbitrary way, the plane has a point which is in exactly r rectangles and they have the same color. (Pach-Tardos.)

A result in number theory has been improved by improving the tool in extreme graph theory. (Gyarmati-Elsholz-Simonovits.)

Theoretical computer science

Let $Sym(n)$ be a group of permutations with minimum degree m. We give sharp estimate on the order of the group and the number of elements with a given support. As an application we show that certain fundamental quantum-computer algorithms cannot be used for the fast solution of the Hidden Subgroup Problem in $Sym(n)$. ("We prove that something cannot be done on a computer which does not exist.") (Pyber.)

A given triangle should be covered by given smaller homothetic triangles and their variants obtained by rotation with 180 degrees. We found an algorithm giving the most economical cover.

It was conjectured that the maximum number of independent functional dependencies for n attributes in a database (or equivalently the maximum number of independent closures in an n-element set) is n choose n/2. We found a rather complicated construction which is essentially bigger. Later we proved an upper estimate showing that out construction is nearly optimal. (Katona-Miklós.)

We axiomatized the weak functional dependencies (disjuctions of functional dependencies) in the case of different contructors with complex values.

Good estimates were given for a coding problem motivated by the relational database model.

A discrepancy type theorem was obtained for matrices, motivated by geografical databases. (Anstee-Sali.)

The following coding problem arose from an extremal problem for relational databases. Let $\mathcal{C}$ be a set of codewords of length n over the q-element alphabet. The minimum Hamming distance is $n-k+1$. Suppose that for every choice of $k-1$ positions there is a pair of codewords being equal exactly in these positions. Let $f(q,k)$ denote the smallest n for which such a code exists. We gave lower and upper estimates on $f(q,k)$ later we improved these estimates for both directions. (Katona-Sali-Schewe, Sali-Székely.)

Characterizational results were achieved for higher order data models. (Sali-Schewe.)

The following problem has its origin in bioinformtics. Every word is identified with its reverse complement: take the reverse of the word and exchange all letters with their complements $a \leftrightarrow \bar{a}$, etc. As it turned out every n-word is uniquely identified by its subwords of length $\leq 2/3\,n$. (Pl. Erdős-Ligeti-Sziklai-Torney.)

Given two alphabets, and a (long) text above one of the alphabets, while a (relatively short) pattern above the other alphabet. For each position of the text we are looking for that injection between the alphabets which minimizes the number of mismatches between the text and the image of the pattern. We gave a fast algorithm to solve this problem parallel for each position assuming that the text is given by a run-length description. (Apostolico - PL. Erdős - Lewenstein.)

New type of fingerprint codes were constructed asymptotically achieving the the largest possible density. (Amiri-Tardos.)

We have written a survey paper on "Multiple Access Channels" containing some new results.

In connection with some problem originated in computational neuroscience, we studied a modified version of the existence problem of a graph with given degree sequence. The problem belongs to Tutte's well-known f-factor theory, our version gives a very simple way to answer the following question: a graphical degree sequence, and a graph G as a partial representation of this degree sequence is given. We want to check whether a bunch of edges, rooted at a given vertex, can be added to the current graph keeping the newly resulted graph to be extendable to a full representation of the original degree sequence. The result can be applied to systematically generate all possible representation of a given graphical degree sequence. (Kim - Toroczkai - Miklós - PL. Erdős - Székely.)