Our research in Computer Science leverages on the synergies of several interrelated fields, including the theory of algorithms and databases with emphasis on new parallel hardware architectures, data mining, information retrieval, machine learning, and graph theory.

Business intelligence, e-science and Web mining are rapidly growing sources of extreme large scale information processing problems. Our goal was to provide efficient solutions to distill knowledge from "big data". Unique to the project was the strong collaboration of engineers and mathematicians: while their results are always experimentally validated over real data, due to the huge size of the problem, their algorithms are based on deep theory of algorithms, probability and algebra.

We published 48 journal and 43 conference papers in the areas of the research plan:

1. Distributed algorithms and databases
2. Algorithms and complexity
2.1 Explicit isomorphisms, algebra, number theory
2.2 Quantum algorithms
2.3 Geometric aspects of the discrete logarithm problem
2.4 Singular Value Decomposition and Data Analysis
2.5 Parametrized complexity
3. Applications

Our basic research led to research and development projects in cooperation with Ericsson, Bosch, AEGON Hungary, and OTP Bank.

Dissemination: Most important activities.

www.highlightsofalgorithms.org

Dániel Marx gave a 1-hour invited tutorial on "The optimality program in parameterized algorithms" at the first edition of the new conference series Highlights on Algorithms (June 2016, Paris, France).

https://simons.berkeley.edu/programs/complexity2015

Dániel Marx co-organized the fall 2015 semester program "Fine-Grained Complexity and Algorithm Design" at the Simons Institute for the Theory of Computing, Berkeley, CA, USA. The 4-month event brought together researchers from a wide range of areas and included four week-long workshops in its program.

www.highlightsofalgorithms.org

The highly successful RecSys Challenge 2016 is co-organized by XING, CrowdRec and MTA SZTAKI. XING is a social network for business. People use XING, for example, to find a job and recruiters use XING to find the right candidate for a job. At the moment, XING has more than 15 Million users and around 1 Million job postings on the platform. Given a user, the goal of the job recommendation system is to predict those job postings that are likely to be relevant to the user. In order to fulfill this task, various data sources can be exploited. Job recommendations are displayed on xing.com as well as in XING's mobile apps.

https://dms.sztaki.hu/ecml-pkkd-2016/

We organized the The ECML/PKDD Discovery Challenge 2016 on Bank Card Usage Analysis to predict the user behavior of the OTP Bank Hungary, a key bank in the CEE Region. The Bank wants to know which branch will be visited by each customer to be able to optimize a contact list distribution. The customer will be proactively involved in campaigns from the branch that will be visited with the highest probability. The bank expects higher conversion rates in branch campaigns if the call is made in the branch mostly preferred by the customer. Internationally, over 30 teams competed and the results were presented at the ECML/PKDD 2016 conference in Italy.

https://dms.sztaki.hu/bubi

We organized the MOL Bubi public bike-sharing system Analytics Challenge together with the Centre for Budapest Transport (BKK) at the end of 2015. In the challenge, 5 months of data was given to predict usage patterns in three tasks, two for predictive analytics and one for open research. The data contains information about every MOL Bubi travel during this period, such as the source and destination stations and the starting and ending time. More than 20 teams competed for the three tasks.

1. Distributed algorithms and databases

Distributed algorithmic frameworks, originally, Hadoop, then Spark, Flink and many others have appeared to address scalability problems. Our research was enabled by using and improving the technologies below.

Daróczy, B., Vaderna, P., & Benczúr, A. (2015). Machine Learning Based Session Drop Prediction in LTE Networks and its SON Aspects.

2015 IEEE 81st Vehicular Technology Conference (VTC Spring). IEEE, 2015.

https://dms.sztaki.hu/sites/dms.sztaki.hu/files/file/2015/pid1159810-4.pdf

Abnormal bearer session release (i.e. bearer session drop) in cellular telecommunication networks may seriously impact the quality of experience of mobile users. A novel, very large scale ML method is presented that is able to predict session drops with higher accuracy than using traditional models.

Andras Garzo, Istvan Petras, Csaba Istvan Sidlo, Andras A. Benczur. Real-time streaming mobility analytics. NetMob 2013 - Third conference on the Analysis of Mobile Phone Datasets. May 1-3, 2013, MIT, Boston, USA
http://perso.uclouvain.be/vincent.blondel/netmob/2013/D4D-book.pdf

Andras Garzo, Andras A. Benczur, Csaba Istvan Sidlo, Daniel Tahara, Erik Francis Wyatt. Real-time streaming mobility analytics. IEEE Big Data 2013

https://dms.sztaki.hu/sites/dms.sztaki.hu/files/file/2013/pid2922315.pdf

We deploy distributed streaming algorithms and infrastructures to process large scale mobility data for fast reaction time prediction. We evaluate our methods on a data set derived from the Orange D4D Challenge data representing sample traces of Ivory Coast mobile phone users.

In the second paper, we also addressed the problem of identifying the exact location and movement of an individual user by learning global patterns both on the user level (home, work location, daily routes) and the traffic (typical routes at time of the day).

As a key performance indicator of our applications, we measure the running time and the error of predictions in short range (5 minutes to 1 hour) and long range (daily, weekly) of the location of an individual user and the density in a given area. Over a cluster of a few old dual core servers, we are capable of processing tens of thousands of record in a second. Our results open the possibility for efficient real time mobility predictions of even large metropolitan areas as well.

We demonstrate our solution via a fast reaction visual dashboard application that can form the base of emergency or rescue services as well as provide grounds for ride sharing, traffic planning and optimization, thus saving natural resources.

Róbert Pálovics, András Benczúr. Temporal influence over the Last.fm social network. The 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining ASONAM 2013 Niagara Falls, Canada, August 25-28, 2013
https://dms.sztaki.hu/sites/dms.sztaki.hu/files/file/2013/lastfm-asonam.pdf

Several recent results show the influence of social contacts to spread certain properties over the network, but others question the methodology of these experiments by proposing that the measured effects may be due to homophily or a shared environment. In this paper we justify the existence of the social influence by considering the temporal behavior of Last.fm users. In order to clearly distinguish between friends sharing the same interest, especially since Last.fm recommends friends based on similarity of taste, we separated the timeless effect of similar taste from the temporal impulses of immediately listening to the same artist after a friend. We measured strong increase of listening to a completely new artist in a few hours' period after a friend compared to non-friends representing a simple trend or external influence. In our experiment to eliminate network independent elements of taste, we improved collaborative filtering and trend based methods by blending with simple time aware recommendations based on the influence of friends. Our experiments are carried over the two-year "scrobble" history of 70,000 Last.fm users.

L. Dudás, Zs. Fekete, J. Göbölös-Szabó, A. Radnai, A. Salánki, A. Szabó, G. Szûcs. OWLAP - using OLAP approach in anomaly detection (Award: Good Support for the Data Preparation, Analysis, and Presentation Process) IEEE Conference on Visual Analytics Science and Technology 2012 October 14 - 19, Seattle, WA, USA
http://www.cse.ohio-state.edu/~raghu/teaching/CSE5544/Visweek2012/vast/challenge/dudas.pdf

An award winner Big Data visualization research prototype.

B. Daróczy, D. Siklósi and A.A. Benczúr. SZTAKI @ ImageCLEF 2012 Photo Annotation. In Working Notes of the ImageCLEF 2011 Workshop at CLEF 2012 Conference, Rome, Italy
https://dms.sztaki.hu/sites/dms.sztaki.hu/files/file/2012/clef2012labs_submission_135.pdf

Our team made second place with tight margin at ImageCLEF 2012 Photo Flickr. We used our open-source GMM/Fisher vector toolkit based on our research for Gaussian Mixture Modeling and Fisher Kernel based learning methods implemented on graphics coprocessors.

2.1 Explicit isomorphisms, algebra, number theory

S. Kiss, Cs. Sándor: On the multiplicativity of the linear combination of additive representation functions, Ramanujan Journal 2016 1-15.
http://www.cs.elte.hu/~kisspest/multiplikativreprezentacio_20151217.pdf

In this paper we give a necessary and sufficient condition to the multiplicativity of the function formed by the linear combination of certain additive representation functions. Our proof is based on elementary ideas.

G. Ivanyos, Y. Qiao, K. V. Subrahmaniam, Non-commutative Edmonds' problem and matrix semi-invariants. Computational Complexity, in press.
http://arxiv.org/abs/1508.00690
DOI: 10.1007/s00037-016-0143-x

We developed an algorithm for computing the noncommutative rank of a matrix whose entries are linear polynomials in several variables. The noncommutative rank is the rank of the matrix, considered as a matrix over the free division algebra generated by the variables. The algorithm is deterministic and runs in polynomial time in special cases. It is also constructive in the sense that witnesses certifying matching lower and upper bounds for the rank are produced. The method is combinatorial, the main idea is a generalization of the augmenting path technique for finding maximum matchings in bipartite graphs. The ingredients of the algorithm appear to be strong enough for leading to a version that runs in polynomial time in arbitrary cases. We mention that, independently, Garg, Gurvits, Oliveira and Wigderson proposed a completely different, although polynomial time method for computing the noncommutative rank in the special case when the coefficients of the polynomial entries of the matrices are rational numbers. In contrast to our algorithm, their method is less constructive (it merely computes the rank) and is restricted to the rational base field.

G. Ivanyos, M. Karpinski, Y. Qiao and M. Santha: Generalized Wong sequences and their applications to Edmonds' problems. J. Comput. Syst. Sci. 81 (2015), 1373-1386.
http://arxiv.org/abs/1307.6429 ; http://dx.doi.org/10.1016/j.jcss.2015.04.006

We discovered deterministic polynomial time algorithms for finding maximum rank matrices in linear spaces of matrices in certain iteresting special cases including the case when the space is spanned by (unknown) rank one matrices and other cases where the matrix space has some hidden algebraic structure, e.g. the matrices in it are simultenously triangularizable. For the case of unkown rank one generators this is supplemented with finding appropriate subsitutes for the generators that, although they may have higher rank, behave from a certain point of view like rank one matrices.

T. Mészáros, L. Rónyai: Shattering-extremal set systems of VC dimension at most 2, The Electronic Journal of Combinatorics, Vol. 21/4, #P4.30, 2014.
http://www.combinatorics.org/ojs/index.php/eljc/article/view/v21i4p30

We say that a set system F over a (finite) ground set X shatters a given subset S of X, if every subset Y of S is the intersection of S and an element of F. A set system is called shattering-extremal if it shatters exactly $|F|$ sets. In this paper we characterize shattering-extremal set systems which shatter no 3 element sets (i.e their Vapnik-Chervonenkis dimension is at most 2) in terms of their inclusion graphs.

T. Mészáros, L. Rónyai: A note on Alon's combinatorial Nullstellensatz, Annales Univ. Sci. Budapest., Sect. Comp. Vol. 42 (2014), 249-260.
http://ac.inf.elte.hu/Vol_042_2014/249_42.pdf

Alon's Combinatorial Nullstellensatz and in particular the resulting nonvanishing criterion is a very useful algebraic tool in combinatorics. In the paper we characterize those subsets X for which the ideal I(x) of polynomials vanishing on X has a laxicographic Gröber basis which is similar (in a precisely defined sense) to the Gröber basis of I(S).

A. Nagy, L. Rónyai: Finite semigroups whose semigroup algebra over a field has a trivial right annihilator, International Journal of Contemporary Mathematical Sciences, 9(2014) 25--36.

http://www.m-hikari.com/ijcms/ijcms-2014/1-4-2014/nagyIJCMS1-4-2014.pdf

We consider an element x of a semigroup algebra F[S] over a field F is called a right annihilating element if xs=0 holds for all elements s from the semigroup algebra. We prove that there is no nontrivial right annihilating element in several important cases: if the finite semigroup S is a direct product, or semilattice, or right zero semigroup of semigroups S_i such that every algebra F[S_i] has the trivial right annihilator only.

Kiss, Sándor Z., Eszter Rozgonyi, and Csaba Sándor. "On Sidon sets which are asymptotic bases of order 4." Functiones et Approximatio Commentarii Mathematici 51.2 (2014): 393-413.

https://www.researchgate.net/profile/Csaba_Sandor/publication/236248264_On_Sidon_sets_which_are_asymptotic_bases_of_order_4/links/54207f3d0cf241a65a1e35cf.pdf

A set S of positive integer is said to be a Sidon set if all the two terms sums from S are distinct. We say a set S of positive integers is an asymptotic basis of order k if every sufficiently large positive integer can be represented as the sum of k terms from the set S. In this paper we prove the existence of Sidon sets which are asymptotic bases of order 4 by using probabilistic methods.

G. Ivanyos, Á. Lelkes, L. Rónyai: Improved algorithms for splitting full matrix algebras, JP Journal of Algebra, Number Theory and Applications
Volume 28, Number 2, 2013, Pages 141-156.
http://eprints.sztaki.hu/7533/1/Ivanyos_141_2474051_ny.pdf

Let K be an algebraic number field of degree d and discriminant D over the field of rational numbers. Let A be an associative algebra over K given by structure constants such that A is isomorphic to the full matrix algebra M_n(K) for some positive integer n. Suppose that d, n and D are bounded. In a previous paper a polynomial time ff-algorithm was given to construct explicitly an isomorphism from A to M_n(K).
Here we simplify and improve this algorithm in the cases n < 43, K = Q, and for two Euclidean imgainary quadratic fields in the case n = 2.
The improvements are based on work by Kitaoka and Coulangeon on tensor products of lattices.

G Kós. Representing the GCD as linear combination in non-PID rings
ACTA MATHEMATICA HUNGARICA 140:(3) pp. 243-247. (2013)
http://www.cs.elte.hu/~kosgeza/publications/005_10474_2013_314_printpdf.pdf

In this paper we prove the following fact: If finitely many elements p_1 , p_2 , . . . , p_n of a unique factorization domain are given such that the greatest common divisor of each pair (p_i , p_j) can be expressed as a linear combination of p_i and p_j, then the greatest common divisor of all the p_i-s can also be expressed as a linear combination of p_1 , . . . , p_n . We prove an analogous statement in general commutative rings.

P. Borwein, T. Erdélyi, G. Kós:
The multiplicity of the zero at 1 of polynomials with constrained coefficients
ACTA ARITHMETICA 159.4 (2013): 387-395.
http://www.cs.elte.hu/~kosgeza/publications/387-396.pdf

We sharpen our previous results concerning polynomials with restricted coefficients. We provide new upper and lower bounds on the maximal multiplicity of the root at point 1 of a polynomial, depending its constant term and the L_p norm of the vector of the other terms. We also give constructions for polynomials where the value at 0 is big with respect to the L_q norm of the values at 1, 2, ..., n is prescribed.

S. Kiss, Cs. Sándor and E. Rozgonyi: Groups, partitions and representation functions, Publicationes Math. Debrecen, 85 (2014), 425-433.
http://www.cs.elte.hu/~kisspest/trecikk.pdf

We give a necessary and sufficient condition such that for every element in an additive semigroup, the equation a + b = n have the same number of solutions where a in A, b in B as to where a in X-A, b in X-B.

S. Kiss, Cs: Sándor: On the maximum values of the additive representation functions, International Journal of Number Theory, 12.04 (2016): 1055-1075.
http://arxiv.org/abs/1504.07411v2

Let A be a set of nonnegative integers. For a positive integer n let R_A(n) denote the number of representations of n as the sum of two terms from A. One of the famous conjecture of Erdős and Turán asserts that if R_A(n) is positive from a certain point on, then it cannot be bounded. We improve a recent result of Haddad and Helou about the quantitative version of the Erdős-Turán conjecture.

2.2 Quantum algorithms

In the area of quantum computing, our most important result is the generalization of the famous hidden subgroup problem leads to novel connections to a seemingly unrelated algorithmic problem.

A. M. Childs, G. Ivanyos, Quantum computation of discrete logarithms in semigroups. J. Math. Cryptology 8 (2014) 405-416.
http://dx.doi.org/10.1515/jmc-2013-0038 http://arxiv.org/abs/1310.6238

We studied the quantum complexity of constructive subsemigroup membership in commutative semigroups. We showed that the semigroup discrete logarithm problem can be solved in polynomial time and that the case of k>=2 is difficult.

G. Ivanyos, M. Santha, On solving systems of diagonal polynomial equations over finite fields, Proc. FAW 2015, Springer LNCS, Vol. 9130 (2015), 125-137.
http://dx.doi.org/10.1007/978-3-319-19647-3_12
http://arxiv.org/abs/1503.09016

We studied the problem of finding nonzero solutions to systems of homogeneous diagonal polynomial equations over finite fields. We give a polynomial time algorithm for the case the degree is constant and the number of variables exceed a certain bound polynomial in the number equations.

T. Decker, G. Ivanyos, R. Kulkarni, Y. Qiao and M. Santha, An efficient quantum algorithm for finding hidden parabolic subgroups in the general linear group, Proc. MFCS 2014, Springer LNCS Vol. 8635 (2014), 226-238

http://eprints.sztaki.hu/7930/1/para_mfcs_nolncs.pdf

We developed a quantum algorithm for finding hidden parabolic subgroups in general linear groups and the closely related groups. In particular, we successfully solved the problem of finding hidden Borel subgroups over small base fields. The earlier method worked in polynomial time only over sufficiently large base fields.

G. Ivanyos, R. Kulkarni, Y. Qiao, M. Santha and A. Sundaram, On the complexity of trial and error for constraint satisfaction problems, Proc. ICALP 2014, Springer LNCS Vol. 8572 (2014), 663-675.

http://eccc.hpi-web.de/report/2014/034/download/

Using an approach that is independent of the known hidden subgroup techniques, we proposed a novel polynomial time quantum algorithm for the latter problem. This new method can be used to obtain polynomial time quantum algorithms for a broader class of hidden polynomials, including elliptic curves.

T. Decker, G. Ivanyos, M. Santha and P. Wocjan "Hidden symmetry subgroup problems." SIAM Journal on Computing 42.5 (2013): 1987-2007.
http://arxiv.org/abs/1107.2189
http://eprints.sztaki.hu/7281/1/hssp-authorcopy-86441.pdf

We investigated a generalization of the hidden subgroup problem (HSP). We showed that in certain cases the generalization can be efficiently reduced to the ordinary HSP. In this framework we found links between the HSP and certain other problems in which quantum computers have exponential advantage over classical randomized algorithms.

K. Friedl, G. Ivanyos, F. Magniez, M. Santha and P. Sen: "Hidden translation and translating coset in quantum computing." SIAM Journal on Computing 43.1 (2014): 1-24.
http://arxiv.org/pdf/quant-ph/0211091v3
http://www.sztaki.hu/~ivanyos/pp/fimss-authorcopy.pdf

We made substantial improvements on our techniques that appeared in 2003 in a conference paper on solving the HSP in a wide class of solvable groups on quantum computers in polynomial time. The new achievements include an elegant variant of a procedure which reduces the HSP of solvable groups to a special hidden shift problem in Abelian group as well as novel applications of the method, such as a subexpontial time hidden subgroup algorithm in general solvable groups which works in quasi-polynomial time in solvable groups of constant exponent.

G. Ivanyos, H. Klauck, T. Lee, M. Santha and R. de Wolf: New bounds on the classical and quantum communication complexity of some graph properties, Proc. FSTTCS 12 (Leibniz International Proceedings in Informatics Vol. 18), 148-159. http://drops.dagstuhl.de/opus/volltexte/2012/3852/

We gave lower and upper bounds on the communication complexity of certain graph properties in the deterministic, in the classical probabilistic as well as in the quantum setting. The investigated properties include connectivity, possessing perfect matchings and having Eulerian tours.

G. Ivanyos, M. Karpinski, Y. Qiao and M. Santha Generalized Wong sequences and their applications to Edmonds' problems, Proc. STACS 14 (Leibniz International Proceedings in Informatics Vol. 25 (2014), 397-408.

http://eprints.sztaki.hu/8532/1/Ivanyos_125_2982463_ny.pdf

We proposed deterministic polynomial time algorithms for finding maximum rank matrices in linear spaces of matrices in special cases, including the case when the space is spanned by (unknown) rank one matrices.

M. Arora, G. Ivanyos, M. Karpinski and N. Saxena, Deterministic polynomial factoring and association schemes, LMS J. of Computation and Mathematics 17 (2014), 123-140.

http://eprints.sztaki.hu/7927/1/sqrtsmooth-nolms.pdf

We investigated higher dimensional generalizations of association schemes and used our results to obtain speedup in deterministic algorithms for factoring polynomials of certain special degrees over finite fields.

T. Decker, P. Hoyer, G. Ivanyos and M. Santha, Polynomial time quantum algorithms for certain bivariate hidden polynomial problems, Quantum Information and Computation 14 (2014) 790-806.

http://eprints.sztaki.hu/8195/1/Decker_790_2707553_ny.pdf

We developed transfer theorems which translate the complexity of a constraint satisfaction problem (CSP) in various versions of a recently proposed trial and error computational model to the classical complexity of an associated CSP. The transfer theorems were applied to obtain polynomial time algorithms or hardness results for the complexity of several interesting problems in the trial and error model.

2.4 Singular Value Decomposition and Data Analysis

L. Csató, L. Rónyai: Incomplete pairwise comparison matrices and weighting methods. FUNDAMENTA INFORMATICAE 144:(3-4) pp. 309-320. (2016)

http://arxiv.org/abs/1507.00461 DOI 10.3233/FI-2016-1337

A special class of preferences, given by a directed acyclic graph, is considered. They are represented by incomplete pairwise comparison matrices as only partial information is available: for some pairs no comparison is given in the graph. A weighting method satisfies the linear order preservation property if it always results in a ranking such that an alternative directly preferred to another does not have a

lower rank. We study whether two procedures, the Eigenvector Method and the Logarithmic Least Squares Method meet this axiom. Both weighting methods break linear order preservation, moreover, the ranking according to the Eigenvector Method depends on the incomplete pairwise comparison representation chosen.

P. L. Erdos, S. Kiss, I. Miklos, L. Soukup: Approximate Counting of Graphical Realizations, PloS ONE, 10(7), 2015.
http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0131300

In 1999 Kannan, Tetali and Vempala proposed a Markov Chain Monte Carlo method to uniformly sample all possible realizations of a given graphical degree sequence and conjectured its rapidly mixing nature. We show that for the general problem the derived Markov Chain Monte Carlo process is rapidly mixing, and it provides a fully polynomial randomized approximation scheme (a.k.a. FPRAS) for counting of all realizations.

Pálovics, R., Benczúr, A. A., Kocsis, L., Kiss, T., & Frigó, E. (2014, October). Exploiting temporal influence in online recommendation. Proc. 8th ACM Conference on Recommender systems (pp. 273-280). ACM.
https://dms.sztaki.hu/sites/dms.sztaki.hu/files/file/2014/recsys.pdf

We give models for time-aware music recommendation in a social media service with the potential of exploiting immediate temporal influences between users.

Daroczy, B., Siklosi, D., Palovics, R., & Benczur, A. A. Text Classification Kernels for Quality Prediction over the C3 Data Set. Proc. 24th WWW Companion 2015 (pp. 1441-1446).
https://dms.sztaki.hu/sites/dms.sztaki.hu/files/file/2015/webq01.pdf

We give text kernel methods for automatically assessing the credibility, presentation, knowledge, intention and completeness by extending the attributes in the C3 dataset by the page textual content.

I Hegedűs, Á Berta, L Kocsis, AA Benczúr, M Jelasity. Robust Decentralized Low-Rank Matrix Decomposition. ACM Transactions on Intelligent Systems and Technology (TIST) 7 (4), 62 2016

http://www.inf.u-szeged.hu/~jelasity/cikkek/tist15.pdf

Low-rank matrix approximation is an important tool in data mining with a wide range of applications including recommender systems, clustering, and identifying topics in documents. When the matrix to be approximated originates from a large distributed system—such as a network of mobile phones or smart meters—this is a very challenging problem due to the strongly conflicting, yet essential requirements of efficiency, robustness, and privacy preservation. We argue that while collecting sensitive data in a centralized fashion may be efficient, it is not an option when considering privacy and efficiency at the same time. Thus, we do not allow any sensitive data to leave the nodes of the

network. The local information at each node (personal attributes, documents, media ratings, etc.) defines one row in the matrix. This means that all computations have to be performed at the edge of the network. Known parallel methods that respect the locality constraint—suchassynchronizedparallel gradientsearch or distributed iterativemethods—require synchronized rounds or have inherent issues with load balancing, thus they are not robust to failure. Our distributed stochastic gradient descent algorithm overcomes these limitations. During the execution, any sensitive information remains local, while the global features (e.g., the factor model of movies) converge to the correct value at all nodes. We present a theoretical derivation, as well as a thorough experimental evaluation of our algorithm. We demonstrate that the convergence speed of our method is competitive while not relying on synchronization and being robust to extreme and realistic failure scenarios. To demonstrate the feasibility of our approach, we present trace-based simulations, real smartphone user behavior analysis, and tests over real movie recommender system data.

T. Mészáros, L. Rónyai: Shattering-Extremal Set Systems of Small VC-Dimension, ISRN Combinatorics, vol. 2013, Article ID 126214, 8 pages, 2013.
http://downloads.hindawi.com/journals/isrn.combinatorics/2013/126214.pdf

We say that a set system family F shatters a given set S if every subset X of S can be obtained as the intersection of some element H of F and S. The Sauer inequality states that in general, a set system F shatters at least |F| sets. Here, we concentrate on the case of equality. A set system F is called shattering-extremal if it shatters exactly |F| sets. We characterize shattering extremal set systems of Vapnik-Chervonenkis dimension 1 in terms of their inclusion graphs. Also, from the perspective of extremality, we relate set systems of bounded Vapnik-Chervonenkis dimension to their projections.

B. Daróczy, A. Benczúr, L. Rónyai: Fisher kernels for image descriptors: a theoretical overview and experimental results, pp. 1-15.
Annales Univ. Sci. Budapest., Sect. Comp. 40 (2013)
https://dms.sztaki.hu/sites/dms.sztaki.hu/files/file/2013/annales2013.pdf

Visualisual words have recently proved to be a key tool in image classification. Gaussian Mixture Modeling (GMM) with a Fisher information based distance over the mixtures often yields the most accurate classification results. Here we overview the theoretical foundations of the Fisher kernel method. We indicate that it yields a natural metric over images characterized by low level content descriptors generated from a Gaussian mixture. We justify the theoretical observations by reproducing standard measurements over the Pascal VOC 2007 data.

Miklos Erdelyi, Andras A. Benczur. Balint Daroczy, Andras Garzo, Tamas Kiss, David Siklosi. The classification power of Web features. Internet Mathematics, 10.3-4 (2014): 421-457.
https://dms.sztaki.hu/sites/dms.sztaki.hu/files/file/2013/internet-mathematics1.pdf

In this paper we give a comprehensive overview of features devised for Web spam detection and investigate how much various classes, some requiring very high computational effort, add to the classification accuracy. We collect and handle a large number of features based on recent advances in Web spam filtering, including temporal ones, in particular we analyze the strength and sensitivity of

linkage change. We show that machine learning techniques including ensemble selection, LogitBoost and Random Forest significantly improve accuracy.

Our result is a summary of the Web spam filtering best practice with a listing of various configurations depending on collection size, computational resources and quality needs. To foster research in the area, we make several feature sets and source codes public, including the temporal features of eight .uk crawl snapshots that include WEBSPAM-UK2007 as well as the Web Spam Challenge features for the labeled part of ClueWeb09.

Levente Kocsis, András György, Andrea N. Bán. BoostingTree: Parallel Selection of Weak Learners in Boosting, with Application to Ranking. Machine Learning, 2013.
https://dms.sztaki.hu/sites/dms.sztaki.hu/files/file/2013/mlj12-rev-r3.pdf

Boosting algorithms have been found successful in many areas of machine learning and, in particular, in ranking. For typical classes of weak learners used in boosting (such as decision stumps or trees), a large feature space can slow down the training, while a long sequence of weak hypotheses combined by boosting can result in a computationally expensive model. In this paper we propose a strategy that builds several sequences of weak hypotheses in parallel, and extends the ones that are likely to yield a good model. The weak hypothesis sequences are arranged in a **boosting tree**, and new weak hypotheses are added to promising nodes (both leaves and inner nodes) of the tree using some randomized method. Theoretical results show that the proposed algorithm asymptotically achieves the performance of the base boosting algorithm applied. Experiments are provided in ranking web documents and move ordering in chess, and the results indicate that the new strategy yields better performance when the length of the sequence is limited, and converges to similar performance as the original boosting algorithms otherwise.

András Garzó, Bálint Daróczy, Tamás Kiss, Dávid Siklósi, András A. Benczúr. Cross-lingual web spam classification. The 3rd Joint WICOW/AIRWeb Workshop on Web Quality Rio de Janeiro, Brasil. May 13, 2013. Proceedigs of the 22nd international conference on World Wide Web companion
https://dms.sztaki.hu/sites/dms.sztaki.hu/files/file/2013/crosslingual-short.pdf

While English language training data exists for several Web classification tasks, most notably for Web spam, we face an expensive human labeling procedure if we want to classify a Web domain in a language different from English. In this paper we overview how existing content and link based classification techniques work, how models can be ``translated'' from English into another language, and how language-dependent and independent methods combine. In particular, we show that simple bag-of-words translation works very well and in this procedure we may also rely on mixed language Web hosts, i.e. those that contain an English translation of part of the local language text. Our experiments are conducted on the ClueWeb09 corpus as the training English collection and a large Portuguese crawl of the Portuguese Web Archive. To foster further research, we provide labels and precomputed values of term frequencies, content and link based features for both ClueWeb09 and the Portuguese data.

B. Daróczy, D. Siklósi and A.A. Benczúr. SZTAKI @ ImageCLEF 2012 Photo Annotation. In Working Notes of the ImageCLEF 2011 Workshop at CLEF 2012 Conference, Rome, Italy
https://dms.sztaki.hu/sites/dms.sztaki.hu/files/file/2012/clef2012labs_submission_135.pdf

Our team made second place with tight margin at ImageCLEF 2012 Photo Flickr. We used our open-source GMM/Fisher vector toolkit based on our research for Gaussian Mixture Modeling and Fisher Kernel based learning methods implemented on graphics coprocessors.

Julianna Göbölös-Szabó, András Benczúr. Temporal Wikipedia search by edits and linkage. SIGIR 2013 Workshop on Time-aware Information Access, 28 July - 1 August 2013, Dublin, Ireland
https://dms.sztaki.hu/sites/dms.sztaki.hu/files/file/2013/temp-wimmut.pdf

In this paper we exploit the connectivity structure of edits in Wikipedia to identify recent events that happened at a given time via identifying bursty changes in linked articles around a specified date. Our key results include algorithms for node relevance ranking in temporal subgraph and neighborhood selection based on measurements for structural changes in time over the Wikipedia link graph. We measure our algorithms over manually annotated queries with relevant events in September and October 2011; we make the assessment publicly available. While our methods were tested over clean Wikipedia metadata, we believe the methods are applicable to general temporal Web collections as well.

2.5 Parametrized complexity

Our paper "List H-coloring a graph by removing few vertices" voted Best Paper at the European Symposium on Algorithms. Dániel Marx gave several tutorials and invited lectures.

D. Marx, M. Pilipczuk. Optimal parameterized algorithms for planar facility location problems using Voronoi diagrams. In Proceedings of the 23rd European Symposium on Algorithms (ESA 2015), Lecture Notes in Computer Science Volume 9294, Springer, 865-877, 2015.
http://arxiv.org/abs/1504.05476
http://dx.doi.org/10.1007/978-3-662-48350-3_72

We study a general family of facility location problems defined on planar graphs and on the 2-dimensional plane. In these problems, a subset of k objects has to be selected, satisfying certain packing (disjointness) and covering constraints. Our main result is showing that, for each of these problems, the n^O(k) time brute force algorithm of selecting k objects can be improved to n^O(\sqrt{k}) time. The algorithm is based on focusing on the Voronoi diagram of a hypothetical solution of k objects; this idea was introduced recently in the design of geometric QPTASs, but was not yet used for exact algorithms and for planar graphs.

D. Marx, T. Miltzow. Peeling and nibbling the cactus: subexponential-time algorithms for counting triangulations and related problems. In Proceedings of the 32nd International Symposium on Computational Geometry (SoCG 2016), 52:1-52:16, 2016.
http://dx.doi.org/10.4230/LIPIcs.SoCG.2016.52

Given a set of n points S in the plane, a triangulation T of S is a maximal set of non-crossing segments with endpoints in S. We present an algorithm that computes the number of triangulations on a given set of n points in time n^(11+ o(1))sqrt{n}, significantly improving the previous best running time of O(2^n n^2) by Alvarez and Seidel [SoCG 2013]. Our main tool is identifying separators of size O(\sqrt{n}) of a triangulation in a canonical way. The definition of the separators are based on the decomposition of the triangulation into nested layers ("cactus graphs"). Based on the above algorithm, we develop a simple and formal framework to count other non-crossing straight-line graphs in n^O(\sqrt{n}) time. We demonstrate the usefulness of the framework by applying it to counting non-crossing Hamilton cycles, spanning trees, perfect matchings, 3-colorable triangulations, connected graphs, cycle decompositions, quadrangulations, 3-regular graphs, and more.

R. Curticapean, D. Marx. Tight conditional lower bounds for counting perfect matchings on graphs of bounded treewidth, cliquewidth, and genus. In Proceedings of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2016), 1650-1669, 2016.

http://www.cs.bme.hu/~dmarx/papers/curticapean-marx-soda2016-permanent.pdf
http://dx.doi.org/10.1137/1.9781611974331.ch113

By now, we have a good understanding of how NP-hard problems become easier on graphs of bounded treewidth and bounded cliquewidth: for various problems, matching upper bounds and conditional lower bounds describe exactly how the running time has to depend on treewidth or cliquewidth. In particular, Fomin et al. (2009, 2010) have shown a significant difference between these two parameters: assuming the Exponential-Time Hypothesis (ETH), the optimal algorithms for problems such as Max Cut and Edge Dominating Set have running time 2^O(t) n^O(1) when parameterized by treewidth, but n^O(t) when parameterized by cliquewidth. In this paper, we show that a similar phenomenon occurs also for counting problems. Specifically, we prove that, assuming the counting version of the Strong Exponential-Time Hypothesis (#SETH), the problem of counting perfect matchings

- has no (2-epsilon)^k n^O(1) time algorithm for any epsilon>0 on graphs of treewidth k (but it is known to be solvable in time 2^k n^O(1) if a tree decomposition of width k is given), and
- has no O(n^(1-epsilon)k) time algorithm for any epsilon > 0 on graphs of cliquewidth k (but it can be solved in time O(n^(k+1)) if a k-expression is given).

A celebrated result of Fisher, Kasteleyn, and Temperley from the 1960s shows that counting perfect matchings in planar graphs is polynomial-time solvable. This was later extended by Gallucio and Loebl (1999), Tesler (2000) and Regge and Zechina (2000) who gave 4^k n^O(1) time algorithms for graphs of genus k. We show that the dependence on the genus k has to be exponential: assuming #ETH, the

counting version of ETH, there is no 2^O(k) n^O(1) time algorithm for the problem on graphs of genus k.

M. Bateni, E. Demaine, M. Hajiaghayi, D. Marx. A PTAS for planar group Steiner tree via spanner bootstrapping and prize collecting. In Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing (STOC 2016), 570-583, 2016.
http://materials.dagstuhl.de/files/16/16221/16221.MohammadHosseinBateni.Preprint.pdf
http://doi.acm.org/10.1145/2897518.2897549

We present the first polynomial-time approximation scheme (PTAS), i.e., (1+epsilon)-approximation algorithm for any constant epsilon> 0, for the planar group Steiner tree problem (in which each group lies on a boundary of a face). This result improves on the best previous approximation factor of O(logn (loglogn)^O(1)). We achieve this result via a novel and powerful technique called spanner bootstrapping, which allows one to bootstrap from a superconstant approximation factor (even superpolynomial in the input size) all the way down to a PTAS. This is in contrast with the popular existing approach for planar PTASs of constructing light-weight spanners in one iteration, which notably requires a constant-factor approximate solution to start from. Spanner bootstrapping removes one of the main barriers for designing PTASs for problems which have no known constant-factor approximation (even on planar graphs), and thus can be used to obtain PTASs for several difficult-to-approximate problems.

S.A. Amiri, S. Kreutzer, D. Marx, R. Rabinovich. Routing with Congestion in Acyclic Digraphs. MFCS 2016: 7:1-7:11
http://dx.doi.org/10.4230/LIPIcs.MFCS.2016.7

We study the version of the k-disjoint paths problem where k demand pairs (s_1,t_1), ..., (s_k,t_k) are specified in the input and the paths in the solution are allowed to intersect, but such that no vertex is on more than c paths. We show that on directed acyclic graphs the problem is solvable in time n^{O(d)} if we allow congestion k-d for k paths. Furthermore, we show that, under a suitable complexity theoretic assumption, the problem cannot be solved in time f(k)n^{o(d*log(d))} for any computable function f.

D. Marx, V. Mitsou. Double-Exponential and Triple-Exponential Bounds for Choosability Problems Parameterized by Treewidth. ICALP 2016: 28:1-28:15
http://dx.doi.org/10.4230/LIPIcs.ICALP.2016.28

Choosability, introduced by Erdös, Rubin, and Taylor [Congr. Number. 1979], is a well-studied concept in graph theory: we say that a graph is c-choosable if for any assignment of a list of c colors to each vertex, there is a proper coloring where each vertex uses a color from its list. We study the complexity of deciding choosability on graphs of bounded treewidth. It follows from earlier work that 3-

choosability can be decided in time 2^(2^(O(w)))*n^(O(1)) on graphs of treewidth w. We complement this result by a matching lower bound giving evidence that double-exponential dependence on treewidth may be necessary for the problem: we show that an algorithm with running time 2^(2^(o(w)))*n^(O(1)) would violate the Exponential-Time Hypothesis (ETH). We consider also the optimization problem where the task is to delete the minimum number of vertices to make the graph 4-choosable, and demonstrate that dependence on treewidth becomes tripleexponential for this problem: it can be solved in time 2^(2^(2^(O(w))))*n^(O(1)) on graphs of treewidth w, but an algorithm with running time 2^(2^(2^(o(w))))*n^(O(1)) would violate ETH.

A.E. Feldmann, D. Marx. The Complexity Landscape of Fixed-Parameter Directed Steiner Network Problems. ICALP 2016: 27:1-27:14
http://dx.doi.org/10.4230/LIPIcs.ICALP.2016.27

Given a directed graph G and a list (s_1, t_1), ..., (s_k, t_k) of terminal pairs, the Directed Steiner Network problem asks for a minimum-cost subgraph of G that contains a directed s_i -> t_i path for every 1 <= i <= k. The special case Directed Steiner Tree (when we ask for paths from a root r to terminals t_1, . . . , t_k) is known to be fixed-parameter tractable parameterized by the number of terminals, while the special case Strongly Connected Steiner Subgraph (when we ask for a path from every t_i to every other t_j) is known to be W[1]-hard parameterized by the number of terminals. We systematically explore the complexity landscape of directed Steiner problems to fully understand which other special cases are FPT or W[1]-hard. Formally, if H is a class of directed graphs, then we look at the special case of Directed Steiner Network where the list (s_1, t_1), ..., (s_k, t_k) of requests form a directed graph that is a member of H. Our main result is a complete characterization of the classes H resulting in fixed-parameter tractable special cases: we show that if every pattern in H has the combinatorial property of being "transitively equivalent to a bounded-length caterpillar with a bounded number of extra edges," then the problem is FPT, and it is W[1]-hard for every recursively enumerable H not having this property. This complete dichotomy unifies and generalizes the known results showing that Directed Steiner Tree is FPT [Dreyfus and Wagner, Networks 1971], Strongly Connected Steiner Subgraph is W[1]-hard [Guo et al., SIAM J. Discrete Math. 2011], and Directed Steiner Network is solvable in polynomial-time for constant number of terminals [Feldman and Ruhl, SIAM J. Comput. 2006], and moreover reveals a large continent of tractable cases that were not known before.

É. Bonnet, L. Egri, D. Marx. Fixed-Parameter Approximability of Boolean MinCSPs. ESA 2016: 18:1-18:18

http://dx.doi.org/10.4230/LIPIcs.ESA.2016.18

The minimum unsatisfiability version of a constraint satisfaction problem (CSP) asks for an assignment where the number of unsatisfied constraints is minimum possible, or equivalently, asks for a minimum-size set of constraints whose deletion makes the instance satisfiable. For a finite set Gamma of constraints, we denote by CSP(Gamma) the restriction of the problem where each constraint is from Gamma. The polynomial-time solvability and the polynomial-time approximability of CSP(Gamma) were fully characterized by [Khanna et al. SICOMP 2000]. Here we study the fixed-parameter (FP-)

approximability of the problem: given an instance and an integer k, one has to find a solution of size at most g(k) in time f(k)n^{O(1)} if a solution of size at most k exists. We especially focus on the case of constant-factor FP-approximability. Our main result classifies each finite constraint language Gamma into one of three classes: (1) CSP(Gamma) has a constant-factor FP-approximation; (2) CSP(Gamma) has a (constant-factor) FP-approximation if and only if Nearest Codeword has a (constant-factor) FP-approximation; (3) CSP(Gamma) has no FP-approximation, unless FPT=W[P]. We show that problems in the second class do not have constant-factor FP-approximations if both the Exponential-Time Hypothesis (ETH) and the Linear PCP Conjecture (LPC) hold. We also show that such an approximation would imply the existence of an FP-approximation for the k-Densest Subgraph problem with ratio 1-epsilon for any epsilon>0.

S. Kratsch, D. Marx, M. Wahlström. Parameterized Complexity and Kernelizability of Max Ones and Exact Ones Problems. Transactions on Computational Theory 8(1): 1 (2016)
https://pure.royalholloway.ac.uk/portal/files/25485741/main_journal_clean.pdf
http://doi.acm.org/10.1145/2858787

For a finite set Γ of Boolean relations, Max Ones SAT(Γ) and Exact Ones SAT(Γ) are generalized satisfiability problems where every constraint relation is from Γ, and the task is to find a satisfying assignment with at least/exactly k variables set to 1, respectively. We study the parameterized complexity of these problems, including the question whether they admit polynomial kernels. For Max Ones SAT(Γ), we give a classification into five different complexity levels: polynomial-time solvable, admits a polynomial kernel, fixed-parameter tractable, solvable in polynomial time for fixed k, and NP-hard already for k = 1. For Exact Ones SAT(Γ), we refine the classification obtained earlier by taking a closer look at the fixed-parameter tractable cases and classifying the sets Γ for which Exact Ones SAT(Γ) admits a polynomial kernel.

M. Cygan, H. Dell, D. Lokshtanov, D. Marx, J. Nederlof, Y. Okamoto, R. Paturi, S. Saurabh, M. Wahlström. On Problems as Hard as CNF-SAT. ACM Trans. Algorithms 12(3): 41 (2016)
http://doi.acm.org/10.1145/2925416

The field of exact exponential time algorithms for non-deterministic polynomial-time hard problems has thrived since the mid-2000s. While exhaustive search remains asymptotically the fastest known algorithm for some basic problems, non-trivial exponential time algorithms have been found for a myriad of problems, including Graph Coloring, Hamiltonian Path, Dominating Set, and 3-CNF-Sat. In some instances, improving these algorithms further seems to be out of reach. The CNF-Sat problem is the canonical example of a problem for which the trivial exhaustive search algorithm runs in time O(2^n), where n is the number of variables in the input formula. While there exist non-trivial algorithms for CNF-Sat that run in time o(2^n), no algorithm was able to improve the growth rate 2 to a smaller constant, and hence it is natural to conjecture that 2 is the optimal growth rate. The strong exponential time hypothesis (SETH) by Impagliazzo and Paturi [JCSS 2001] goes a little bit further and asserts that, for every epsilon< 1, there is a (large) integer k such that k-CNF-Sat cannot be computed in time 2epsilon*n.

In this article, we show that, for every epsilon < 1, the problems Hitting Set, Set Splitting, and NAE-Sat cannot be computed in time O(2^{epsilon*n)) unless SETH fails. Here n is the number of elements or variables in the input. For these problems, we actually get an equivalence to SETH in a certain sense. We conjecture that SETH implies a similar statement for Set Cover and prove that, under this assumption, the fastest known algorithms for Steiner Tree, Connected Vertex Cover, Set Partitioning, and the pseudo-polynomial time algorithm for Subset Sum cannot be significantly improved. Finally, we justify our assumption about the hardness of Set Cover by showing that the parity of the number of solutions to Set Cover cannot be computed in time O(2^{epsilon*n)) for any epsilon < 1 unless SETH fails.

Y. Cao, D. Marx. Chordal Editing is Fixed-Parameter Tractable. Algorithmica 75(1): 118-137 (2016)
http://dx.doi.org/10.4230/LIPIcs.STACS.2014.214
http://dx.doi.org/10.1007/s00453-015-0014-x

Graph modification problems typically ask for a small set of operations that transforms a given graph to have a certain property. The most commonly considered operations include vertex deletion, edge deletion, and edge addition; for the same property, one can define significantly different versions by allowing different operations. We study a very general graph modification problem that allows all three types of operations: given a graph G and integers k1, k2, j3, the chordal editing problem asks whether G can be transformed into a chordal graph by at most k1 vertex deletions, k2 edge deletions, and k3 edge additions.

M. Grohe, D. Marx. Structure Theorem and Isomorphism Test for Graphs with Excluded Topological Subgraphs. SIAM Journal on Computing, 44(1):114-159, 2015. http://dx.doi.org/10.1137/120892234
http://www.cs.bme.hu/~dmarx/papers/grohe-marx-topdec-sicomp.pdf

We generalize the structure theorem of Robertson and Seymour for graphs excluding a fixed graph H as a minor to graphs excluding H as a topological subgraph. We prove that for a fixed H, every graph excluding H as a topological subgraph has a tree decomposition where each part is either ``almost embeddable'' to a fixed surface or has bounded degree with the exception of a bounded number of vertices.

Y. Cao, D. Marx. Interval deletion is fixed-parameter tractable. ACM Transaction on Algorithms, 11(3):21, 2015.

http://doi.acm.org/10.1145/2629595
http://www.cs.bme.hu/~dmarx/papers/cao-marx-interval-talg.pdf

We study the minimum interval deletion problem, which asks for the removal of a set of at most k vertices to make a graph on n vertices into an interval graph. We present a parameterized algorithm of runtime 10^k*n^O(1) for this problem, thereby showing its fixed-parameter tractability.

L.A. Végh, D. Marx. Fixed-parameter algorithms for minimum cost edge-connectivity augmentation
ACM Transaction on Algorithms, 11(4):27, 2015.
http://dx.doi.org/10.1145/2700210

We consider connectivity-augmentation problems in a setting where each potential new edge has a nonnegative cost associated with it, and the task is to achieve a certain connectivity target with at most p new edges of minimum total cost. The minimum cost augmentation of edge-connectivity from k-1 to k with at most p new edges is fixed-parameter tractable parameterized by p and admits a polynomial kernel.

R. Chitnis, M. Cygan, M. Hajiaghayi, D. Marx. Directed Subset Feedback Vertex Set is Fixed-Parameter Tractable. ACM Transaction on Algorithms, 11(4):28, 2015.
http://dx.doi.org/10.1145/2700209
http://www.cs.bme.hu/~dmarx/papers/chitnis-subset-talg.pdf

Given a graph G and an integer k, the FEEDBACK VERTEX SET (FVS) problem asks if there is a vertex set T of size at most k that hits all cycles in the graph. We complete the picture for feedback vertex set problems and their subset versions in undirected and directed graphs.

D. Marx, P. Wollan. An exact characterization of tractable demand patterns for maximum disjoint path problems. Proc. 26th ACM-SIAM SODA, 642-661,2015.
http://dx.doi.org/10.1137/1.9781611973730.44
http://www.cs.bme.hu/~dmarx/papers/marx-wollan-soda2015-paths.pdf

We study the following general disjoint paths problem: given a supply graph G, terminals T, a demand graph H on the vertices T, and an integer k, we seek to find a set of k pairwise vertex-disjoint valid paths, where we say that a path of the supply graph G is valid if its endpoints are in T and adjacent in the demand graph H. We give a complete characterization of the fixed-parameter tractable cases.

B.M.P. Jansen, D.Marx. Characterizing the easy-to-find subgraphs from the viewpoint of polynomial-time algorithms, kernels, and Turing kernels. Proc. SODA, 616-629, 2015.
http://dx.doi.org/10.1137/1.9781611973730.42
http://www.cs.bme.hu/~dmarx/papers/jansen-marx-soda2015-packing.pdf

We study two fundamental problems related to finding subgraphs: given graphs G and H, test whether H is isomorphic to a subgraph of G, or determine the maximum number of vertex-disjoint H-subgraphs that can be packed in G.

S. Guillemot, D. Marx: A faster FPT algorithm for Bipartite Contraction. Information Processing Letters, 113(22-24): 906-912 (2013) http://dx.doi.org/10.1016/j.ipl.2013.09.004

The Bipartite Contraction problem is to decide, given a graph G and a parameter k, whether we can can obtain a bipartite graph from G by at most k edge contractions. We present a new randomized FPT algorithm for the problem, which is both conceptually simpler and achieves an improved running time.

D. Marx, I. Razgon Fixed-parameter tractability of multicut parameterized by the size of the cutset. SIAM Journal on Computing, 43(2): 355-388 (2014)

http://dx.doi.org/10.1137/110855247

Given an undirected graph G, a collection (s_1,t_1),..., (s_k,t_k) of pairs of vertices, and an integer p, the Edge Multicut problem ask if there is a set S of at most p edges such that the removal of S disconnects every s_i from the corresponding t_i. We show fixed-parameter tractability parameterized by the size p of the cutset in the solution.

Andrei A. Bulatov, Dániel Marx: Constraint Satisfaction Parameterized by Solution Size. SIAM J. Comput. 43(2): 573-616 (2014)
http://dx.doi.org/10.1137/120882160

In the constraint satisfaction problem (CSP) corresponding to a constraint language (i.e., a set of relations) Gamma, the goal is to find an assignment of values to variables so that a given set of constraints specified by relations from Gamma is satisfied. We study the fixed-parameter tractability of CSPs parameterized by the size of the solution.

Dániel Marx, Paul Wollan: Immersions in Highly Edge Connected Graphs. SIAM J. Discrete Math. 28(1): 503-520 (2014)

http://dx.doi.org/10.1137/130924056

We consider the problem of how much edge connectivity is necessary to force a graph G to contain a fixed graph H as an immersion. We give a structure theorem which roughly characterizes those highly edge connected graphs which do not contain K_t as a strong immersion.

Martin Grohe, Dániel Marx: Constraint Solving via Fractional Edge Covers. ACM Transactions on Algorithms 11(1): 4 (2014)

http://doi.acm.org/10.1145/2636918

Many important combinatorial problems can be modeled as constraint satisfaction problems. Hence, identifying polynomial-time solvable classes of constraint satisfaction problems has received a lot of attention. We describe the structural properties that can make the problem tractable.

Dániel Marx, Anastasios Sidiropoulos: The limited blessing of low dimensionality: when 1-1/d is the best possible exponent for -dimensional geometric problems. Symposium on Computational Geometry 2014: 67

http://doi.acm.org/10.1145/2582112.2582124

We are studying d-dimensional geometric problems that have algorithms with 1−1/d appearing in the exponent of the running time.

Sylvain Guillemot, Dániel Marx: Finding small patterns in permutations in linear time. SODA 2014: 82-101

http://dx.doi.org/10.1137/1.9781611973402.7

Given two permutations σ and π, the Permutation Pattern problem asks if σ is a subpattern of π. We show that the problem is fixed-parameter tractable parameterized by the size of the subpattern to be found.

Yixin Cao, Dániel Marx: Interval Deletion is Fixed-Parameter Tractable. SODA 2014: 122-141

http://dx.doi.org/10.1137/1.9781611973402.9

We study the minimum interval deletion problem, which asks for the removal of a set of at most k vertices to make a graph on n vertices into an interval graph. We present a parameterized algorithm for this problem, thereby showing its fixed-parameter tractability.

Rajesh Hemant Chitnis, MohammadTaghi Hajiaghayi, Dániel Marx: Tight Bounds for Planar Strongly Connected Steiner Subgraph with Fixed Number of Terminals (and Extensions). SODA 2014: 1782-1801

http://dx.doi.org/10.1137/1.9781611973402.129

Given a vertex-weighted directed graph and a set of k terminals, the objective of the Strongly Connected Steiner Subgraph (SCSS) problem is to find a vertex set of minimum weight such that G[H] contains a ti → tj path for each i = j. We explore how much easier the problem becomes on planar directed graphs.

Philip N. Klein, Dániel Marx: A subexponential parameterized algorithm for Subset TSP on planar graphs. SODA 2014: 1812-1830 http://dx.doi.org/10.1137/1.9781611973402.131

Given a graph G and a subset S of vertices, the Subset TSP problem asks for a shortest closed walk in G visiting all vertices of S. Our main result is showing that the problem can be solved in time $k^{O(\sqrt{k})+W)n^O(1)if G is a planar graph with weights that are integers no greater than W.

Yixin Cao, Dániel Marx: Chordal Editing is Fixed-Parameter Tractable. STACS 2014: 214-225

http://dx.doi.org/10.4230/LIPIcs.STACS.2014.214

Graph modification problems are typically asked as follows: is there a set of k operations that transforms a given graph to have a certain property. The most commonly considered operations include vertex deletion, edge deletion, and edge addition; for the same property, one can define significantly different versions by allowing different operations. We study a very general graph modification problem which allows all three types of operations.

Dániel Marx, Michal Pilipczuk: Everything you always wanted to know about the parameterized complexity of Subgraph Isomorphism (but were afraid to ask). STACS 2014: 542-553

http://dx.doi.org/10.4230/LIPIcs.STACS.2014.542

Given two graphs H and G, the Subgraph Isomorphism problem asks if H is isomorphic to a subgraph of G. We systematically investigate the possible parameterized algorithms that can exist for Subgraph Isomorphism.

Guillemot, Sylvain, and Dániel Marx. "A faster FPT algorithm for bipartite contraction." *Information Processing Letters* 113.22 (2013): 906-912.
http://arxiv.org/pdf/1305.2743

The Bipartite Contraction problem is to decide, given a graph G and a parameter k, whether we can can obtain a bipartite graph from G by at most k edge contractions. The fixed-parameter tractability of the problem was shown by Heggernes et al., with an algorithm whose running time has double-exponential dependence on k. We present a new randomized FPT algorithm for the problem, which is both conceptually simpler and achieves an improved 2^{O(k^2)}*nm running time, i.e., avoiding the double-exponential dependence on k. The algorithm can be derandomized using standard techniques.

Marx, Dániel, and Igor Razgon. "Fixed-parameter tractability of multicut parameterized by the size of the cutset." SIAM Journal on Computing 43.2 (2014): 355-388.

http://arxiv.org/pdf/1010.3633

Given an undirected graph G, a collection (s_1,t_1),..., (s_k,t_k) of pairs of vertices, and an integer p, the Edge Multicut problem ask if there is a set S of at most p edges such that the removal of S disconnects every s_i from the corresponding t_i. Vertex Multicut is the analogous problem where S is a set of at most p vertices. Our main result is that both problems are fixed-parameter tractable parameterized by the size p of the cutset in the solution.

3. Applications

Our results on error protection methods in optical networks appeared at leading international conferences and journals. The applied research is exploited in contracts with Ericsson, OTP Bank, AEGON Hungary and Bosch. By computational methods, we were able to confirm a conjecture of Littlewood.

S. Bozóki, T.-L. Lee, L. Rónyai: Seven mutually touching infinite cylinders, Computational Geometry 48.2 (2015): 87-93
http://arxiv.org/abs/1308.5164
http://www.sztaki.mta.hu/~bozoki/BozokiLeeRonyai-CGTA-2015-Manuscript.pdf

We confirm a conjecture of Littlewood: there exist seven infinite circular cylinders of unit radius which mutually touch each other. In fact, we exhibit two such sets of cylinders. Our approach is algebraic and uses symbolic and numerical computational techniques. We consider a system of polynomial equations describing the position of the axes of the cylinders in the 3 dimensional space. To have the same number of equations (namely 20) as the number of variables, the angle of the first two cylinders is fixed to 90 degrees. The homotopy continuation method provided approximate solutions. The existence of exact real solutions was verified with the alphaCertified method.

A. Pasic, J. Tapolcai, P. Babarczi, E. Bérczi-Kovács, Z. Király, and L. Rónyai: Survivable routing meets diversity coding, Proc. IFIP Networking Conference, Toulouse, France, 2015, pp. 1-9.
http://lendulet.tmit.bme.hu/lendulet_website/wp-content/papercite-data/pdf/pasic_2015_networking_survivable.pdf
http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7145330

Survivable routing methods in transport networks have been thoroughly investigated in the past decades. However, the proposed approaches suffered either from slow recovery time, poor bandwidth utilization, high computational or operational complexity. We propose an alogorithm with a modest polynomial-time complexity for solving this routing problem. We show that the same routing problem turns to be NP-hard as soon as we limit the forwarding capabilities of some nodes and the capacities of some links of the network.

J. Tapolcai, L. Rónyai, É. Hosszu, L. Gyimóthi, P-H. Ho, S. Subramaniam: Signaling Free Localization of Node Failures in All-Optical Networks. IEEE Transactions on Communications Volume: 64, Issue: 6, June 2016, 2527—2538

http://lendulet.tmit.bme.hu/lendulet_website/wp-content/papercite-data/pdf/tapolcai2016node_mtrail_tcom.pdf
DOI 10.1109/TCOMM.2016.2545653

Network-wide local unambiguous failure localization (NL-UFL) has been demonstrated as an interesting scenario of monitoring trails (m-trails). It attempts to enable every node in a communication network to autonomously localize any failure event in the network in a distributed

and all-optical manner by inspecting a set of m-trails traversing through the node. This paper investigates the m-trail allocation problem under the NL-UFL scenario by taking each link and node failure event into consideration. Bound analysis is performed using combinatorial group testing (CGT) theory and this is followed by the introduction of a novel heuristic on general topologies. Extensive simulation is also conducted.

J. Tapolcai, P-H. Ho, P. Babarczi, L. Rónyai: Neighborhood Failure Localization in All-Optical Networks via Monitoring Trails. IEEE-ACM Transactions on Networking 23:(6) pp. 1719-1728. (2015)

http://real.mtak.hu/33452/1/Tapolcai_1719_2731569_ny.pdf
DOI: 10.1109/TNET.2014.2342222

Shared protection is recognized for its outstanding efficiency in all-optical mesh networks, at the expense of lengthy restoration time. This paper investigates a novel monitoring trail (m-trail) scenario, called Global Neighborhood Failure Localization (G-NFL), that aims to enable any shared protection scheme, for achieving all-optical and ultra-fast failure restoration. We first define the neighborhood of a node. By assuming every node can obtain the on-off status of traversing m-trails and working lightpaths via lambda monitoring, the proposed G-NFL instance routes a set of m-trails such that each node can localize any failure in its neighborhood. Bound analysis is performed on the minimum bandwidth required.

Róbert Pálovics, András Benczúr. Temporal influence over the Last.fm social network. The 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining ASONAM 2013 Niagara Falls, Canada, August 25-28, 2013

Róbert Pálovics, András A. Benczúr. Temporal influence over the Last.fm social network. Social Network Analysis and Mining. Springer. 2015, 5:4

https://dms.sztaki.hu/sites/dms.sztaki.hu/files/file/2015/lastfm-snam-revision2_0.pdf

We successfully decomposed influence from effects of trends, global popularity, and homophily or shared environment of friends. We present our new experiments that use a mathematically sound formula for defining and measuring the influence in the network.

Pálovics, R., Ayala-Gómez, F., Csikota, B., Daróczy, B., Kocsis, L., Spadacene, D., & Benczúr, A. A. (2014, October). RecSys Challenge 2014: an ensemble of binary classifiers and matrix factorization. Proc. 2014 Recommender Systems Challenge (p. 13). ACM.

https://dms.sztaki.hu/sites/dms.sztaki.hu/files/file/2014/challenge.pdf

In our second place winner solution, we devise (1) a mix of binary classification methods for predicting nonzero engagement, including logistic regression and SVM; (2) regression methods for directly predicting the engagement, including linear regression and gradient boosted trees; (3) matrix factorization and factorization machines over the user-movie matrix, by using user and movie features as side information.

Location-aware online learning for top-k recommendation. R Pálovics, P Szalai, J Pap, E Frigó, L Kocsis, AA Benczúr. Pervasive and Mobile Computing (2016)

https://dms.sztaki.hu/sites/dms.sztaki.hu/files/file/2016/recsysjournal2015.pdf

We address the problem of recommending highly volatile items for users, both with potentially ambiguous location that may change in time. The three main ingredients of our method include (1) using online machine learning for the highly volatile items; (2) learning the personalized importance of hierarchical geolocation (for example, town, region, country, continent); finally (3) modeling temporal relevance by counting recent items with an exponential decay in recency. For (1), we consider a time-aware setting, where evaluation is cumbersome by traditional measures since we have different top recommendations at different times.

We describe a time-aware framework based on individual item discounted gain. For (2), we observe that trends and geolocation turns out to be more importantthanpersonalizeduserpreferences: user-itemmatrixfactorizationimproves in combination with our geo-trend learning methods, but in itself, it is greatly inferior to our location based models. In fact, since our best performing methods are based on spatiotemporal data, they are applicable in the user cold start setting as well. Finally for (3), we estimate the probability that the item will be viewed by its previous views to obtain a powerful model that combines item popularity and recency. To generate realistic data for measuring our new methods, we rely on Twitter users with known GPS location and consider hashtags as items that we recommend the users to be included in their next message.

Bálint Daróczy, Róbert Pálovics, Vilmos Wieszner, Richárd Farkas, András A. Benczúr. Predicting User-specific Temporal Retweet Count Based on Network and Content Information. INRA workshop, RecSys 2015

https://www.ntnu.no/wiki/download/attachments/86727107/INRA_2015_proceedings.pdf

Twitter generates a constant flow of quality news and mixed social content. While it is relative easy to separate large popularity news sources from personal messages, we address a more difficult question

to predict the success of a single message among all messages of the same user. We describe a temporal evaluation framework to analyze which messages of which users will be retweeted the most. It turns out that global popularity depend mostly on the network characteristics of the user, while for a given user, the retweet count of a single message can be predicted best by using a variety of features of the content, including linguistic characteristics.

J. Tapolcai, Pin-Han Ho, P. Babarczi, L. Rónyai, "Internet Optical Infrastructure - Issues on Monitoring and Failure Restoration", Springer, pp. 1-212, 2014. [ISBN 978-1-4614-7737-2]

The book is mostly based on the authors' work in the past decade on optical network survivability and fault localization/management.

J. Tapolcai, L. Rónyai, É. Hosszu, Pin-Han Ho, S. Subramaniam, "Signaling Free Localization of Node Failures in All-Optical Networks", In Proc. IEEE INFOCOM, Toronto, Canada, pp. 1860-1868, 2014

http://eprints.sztaki.hu/8034/1/Tapolcai_1860_2696098_ny.pdf

We investigate network-wide local unambiguous failure localization (NL-UFL), an importnant fault localization scenario in all-optical networks which employs monitoring trails (m-trails). It attempts to enable every node to autonomously localize any failure event in the network in a distributed and all-optical manner. Heuristic algorithms were introduced and analysed via simulations.

J. Tapolcai, Pin-Han Ho, P. Babarczi, L. Rónyai, "Neighborhood Failure Localization in All-Optical Networks via Monitoring Trails", IEEE/ACM Transactions on Networking, 23.6 (2015): 1719-1728.

http://eprints.sztaki.hu/8561/1/Tapolcai_1719_2731569_ny.pdf

We investigated a novel monitoring trail approach, called Global Neighborhood Failure Localization (G-NFL), that aims to enable any shared protection scheme to achieve all-optical and ultra-fast restoration after failure. We obtain a bound on the bandwidth requirement via a general lower bound for CGT coverlength.

Péter Babarczi, János Tapolcai, Lajos Rónyai, and Muriel Médard, Resilient Flow Decomposition of Unicast Connections with Network Coding, in Proceedings of the 2014 IEEE International Symposium on Information Theory (ISIT), pp. 116--120.

http://www.rle.mit.edu/ncrc/wp-content/uploads/2014/06/Resilient-Flow-Decomposition.pdf

In the paper we close the gap between end-to-end diversity coding and intra-session network coding for unicast connections resilient against single link failures. In particular, we show that coding

operations are sufficient to perform at the source and receiver if the user data can be split into at most two parts over the field GF(2). Our proof is purely combinatorial and based on standard graph theoretical and network flow techniques. It is a linear time construction that defines the route of subflows A, B and A+B between the source and destination nodes.

Miklós, István, Sandor Z. Kiss, and Eric Tannier. "On sampling SCJ rearrangement scenarios." arXiv preprint arXiv:1304.2170 (2013).

http://arxiv.org/pdf/1304.2170

We consider the Single Cut or Join (SCJ) operation on genomes, generalizing chromosome evolution by fusions and fissions, the computationally simplest known model of genome rearrangement. While most genome rearrangement problems are already hard when comparing three genomes, it is possible to compute in polynomial time a most parsimonious SCJ scenario for an arbitrary number of genomes related by a binary phylogenetic tree. Most importantly we prove that if a Fully Polynomial Randomized Approximation Scheme or a Fully Polynomial Almost Uniform Sampler exist for the most parsimonious SCJ scenario, then RP = NP.

É. Hosszú, J. Tapolcai, L. Rónyai, P. Babarczi, P. Soproni, P-H. Ho:
Fast failure Localization in all-optical networks with length-constrained monitoring trails, In: Proc. of the Workshop on Reliable Networks Design and Modeling (RNDM), St. Petersburg, Russia, 2012, pp. 677-683.
http://www.sztaki.hu/~ronyai/RNDM2012.pdf

The paper investigates theoretical bounds for allocating monitoring trails with at most k hops in optipcal networks, via an optimal combinatoryal group testing construction. Moreover, a novel meta-heuristic approach for solving the length-constrained m-trail allocation problem is introduced. Through extensive simulations the performance gap of the proposed algorithm to the lower bound is presented on a wide variety of network topologies.

J. Tapolcai, Pin-Han Ho, P. Babarczi, L. Rónyai: On Signaling-Free Failure Dependent Restoration in All-Optical Mesh Networks, IEEE/ACM Transactions on Networking, 2013. (accepted)
http://opti.tmit.bme.hu/~tapolcai/papers/tapolcai2013fdp-mtrail.pdf

Failure dependent protection (FDP) is known to achieve optimal capacity efficiency among all types of protection in all-optical mesh networks, at the expense of long recovery time and complicated signaling overhead. To avoid these, the paper investigates a new restoration framework that enables all-optical fault management and device configuration without relying on any control plane signaling. With the proposed restoration framework, a novel spare capacity allocation problem is defined, and analysed on circulant topologies.

J. Tapolcai, Pin-Han Ho, P. Babarczi, L. Rónyai, "On Achieving All-Optical Failure Restoration via Monitoring Trails", In Proc. IEEE INFOCOM Mini-Symposium, Turin, Italy, 2013.
http://opti.tmit.bme.hu/~tapolcai/papers/tapolcai2013nfl_mtrail_infocom.pdf

The paper investigates a novel monitoring trail (m-trail) scenario that can enable any shared protection scheme for achieving all-optical and ultra-fast failure restoration. A set of m-trails is routed such that each node can localize any failure in its neighborhood according to the ON-OFF status of the traversing m-trails. An analysis of the minimum bandwidth required for the m-trails is given. Extensive simulation is conducted to verify the proposed scheme.

J. Tapolcai, L. Rónyai, Pin-Han Ho: Link Fault Localization using Bi-directional M-Trails in All-Optical Mesh Networks, IEEE Transactions on Communications, 61.1 (2013): 291-300.
http://opti.tmit.bme.hu/~tapolcai/papers/tapolcai2012tcom_proofs.pdf

The paper considers the problem of single-link failure localization in all-optical mesh networks. Our study follows a generic monitoring approach in which a set of bi-directional monitoring trails (bm-trails) are defined, such that the network controller can achieve unambiguous failure localization (UFL) for any single link by collecting the alarms from the affected bm-trails. With the objective of minimizing the number of bm-trails (or the length of alarm codes) required for single-link UFL, the paper provides optimal (or essentially optimal) solutions to the bm-trail allocation problem on a number of well known topologies.

É. Hosszu, S. Kiss, L. Rónyai, J. Tapolcai: On a Parity Based Group Testing Algorithm. Acta Cybernetica-Szeged, 22 (2015), 423-433.
http://www.inf.u-szeged.hu/actacybernetica/edb/vol22n2/pdf/Kiss_2015_ActaCybernetica.pdf

The aim of the traditional Combinatorial Group Testing procedures is to identify up to d defective items from a set of n elements. In this paper we describe a new group testing scheme called parity group testing. The difference between the classical and the parity group testing method is that we check the parity of the defective items in a subset instead of the existence of defective items. The parity group testing algorithm has two possible outcomes as well. The test is negative if the number of defective items in the considered subset is even, otherwise it is positive. Our method is based on the Chinese Remainder Theorem similarly as an earlier result of Hirschberg et al.