Aims and Scope
The periodical *Dagstuhl Reports* documents the
program and the results of Dagstuhl Seminars and
Dagstuhl Perspectives Workshops.
In principal, for each Dagstuhl Seminar or Dagstuhl
Perspectives Workshop a report is published that
contains the following:

- an executive summary of the seminar program
  and the fundamental results,

- an overview of the talks given during the seminar
  (summarized as talk abstracts), and

- summaries from working groups (if applicable).

This basic framework can be extended by suitable
contributions that are related to the program of the
seminar, e. g. summaries from panel discussions or
open problem sessions.

Report from Dagstuhl Seminar 16451

Structure and Hardness in P

Edited by

Moshe Lewenstein[1], Seth Pettie[2], and Virginia Vassilevska Williams[3]

1   **Bar-Ilan University – Ramat Gan, IL**, `moshe@cs.biu.ac.il`
2   **University of Michigan – Ann Arbor, US**, `pettie@umich.edu`
3   **Stanford University, US**, `virgi@cs.stanford.edu`

──── **Abstract** ────

This document contains description of the talks at the Dagstuhl seminar 16451 "Structure and Hardness in P". The main goal of the seminar was to bring together researchers from several disciplines and connect those who work on proving conditional lower bounds with those who or may benefit from it. This resulted in an extensive list of open problems which is also provided.

1   Executive Summary

Moshe Lewenstein
Seth Pettie
Virginia Vassilevska Williams

The complexity class **P** (polynomial time) contains a vast variety of problems of practical interest and yet relatively little is known about the structure of P, or of the complexity of many individual problems in P. It is known that there exist contrived problems requiring $\Omega(n1.5)$ time or $\Omega(n2)$ time, and yet to date no unconditional nonlinear lower bounds have been proved for any problem of practical interest. However, the last few years have seen a new resurgence in conditional lower bounds, whose validity rests on the conjectured hardness of some archetypal computational problem. This work has imbued the class P with new structure and has valuable explanatory power.

To cite a small fraction of recent discoveries, it is now known that classic dynamic programming problems such as Edit Distance, LCS, and Fréchet distance require quadratic time (based on the conjectured hardness of k-CNF-SAT), that the best known triangle enumeration algorithms are optimal (based on the hardness of 3-SUM), that Valiant's context-free grammar parser is optimal (based on the hardness of k-CLIQUE), and that the best known approximate Nash equilibrium algorithm is optimal (based on the hardness of 3-SAT).

This Dagstuhl Seminar will bring together top researchers in diverse areas of theoretical computer science and include a mixture of both experts and non-experts in conditional lower bounds. Some specific goals of this seminar are listed below.

- Numerous important problems (such as Linear Programming) seem insoluble in linear time, and yet no conditional lower bounds are known to explain this fact. A goal is to discover conditional lower bounds for key problems for which little is currently known.

- Recent work has been based on both traditional hardness assumptions (such as the ETH, SETH, 3SUM, and APSP conjectures) and a variety of newly considered hardness assumptions (such as the OMv conjecture, the k-CLIQUE conjecture, and the Hitting Set conjecture). Almost nothing is known about the relative plausibility of these conjectures, or if multiple conjectures are, in fact, equivalent. A goal is to discover formal relationships between the traditional and newer hardness assumptions.

- A key goal of the seminar is to disseminate the techniques used to prove conditional lower bounds, particularly to researchers from areas of theoretical computer science that have yet to benefit from this theory. To this end the seminar will include a number of tutorials from top experts in the field.

2 Table of Contents

Open problems

3   Overview of Talks

3.1   Hardness for Graph Problems

Amir Abboud (Stanford University, US)

This is a survey of the landscape of Hardness in P results that we have for graph problems.

3.2   Optimal Hashing for High-Dimensional Spaces

Alexandr Andoni, Thijs Laarhoven, Ilya Razenshteyn, and Erik Waingarten

We survey recent advances in the approximate nearest neighbor search problem in high-dimensional Euclidean/Hamming spaces, which go beyond the classic Locality Sensitive Hashing technique for the problem. The culmination of these advances is a new optimal hashing algorithm that achieves the full trade-off between space vs query time. For example, we obtain the first algorithm with near-linear space and sub-linear query time for any approximation factor greater than 1, which is perhaps the most important regime in practice.

Our algorithm also unifies, simplifies, and improves upon the previous data structures for the problem, combining elements of data-dependent hashing and Locality Sensitive Filtering.

Finally, we discuss matching lower bounds for hashing algorithms, as well as for 1- and 2-cell probe algorithms. In particular, the 2-cell probe lower bound exploits a connection to locally-decodable codes, and yields the first space lower bound that is not polynomially smaller than the 1-probe bound (for any static data structure).

3.3   Permanents as hardness for problems in P?

Andreas Björklund (Lund University, SE)

We know that solving the orthogonal vectors problem with two sets of n vectors of dimension d, for d superlogarithmic in n in less than quadratic time in n would violate SETH, even if the vectors are from $\{0,1\}^d$. This is used in many of the SETH hardness results for problems in P.

However, SETH is a hypothesis about one specific problem, that there are no significantly faster algorithms to solve CNF Sat on n variables than testing all assignments. Unfortunately there are very few fine-grained reductions between hard exponential time problems. Until we find them, there are several other hard problems we could make similar hypotheses about.

In this talk we take a look at the matrix permanent. We give a proof sketch that shows that you cannot count the pairs of orthogonal vectors in two sets of n vectors each from

$\{-1, 0, 1\}^d$, for d polylogarithmic in n, in truly subquadratic time in n, unless you can compute the permanent of a 0/1 matrix with bounded number of ones faster than the best algorithms we know of to date.

3.4   Hardness for Polytime String Problems

Karl Bringmann (MPI für Informatik – Saarbrücken, DE)

In this tutorial we surveyed recent conditional lower bounds for polynomial time problems on strings. We focused on hardness based on the Strong Exponential Time Hypothesis, specifically we discussed hardness of longest common subsequence [1, 2] as well as for pattern matching of regular expressions [3, 4].

References

1   Karl Bringmann and Marvin Künnemann. "Quadratic conditional lower bounds for string problems and dynamic time warping." FOCS 2015.
2   Amir Abboud, Arturs Backurs, and Virginia Vassilevska Williams. "Tight hardness results for LCS and other sequence similarity measures." FOCS 2015.
3   Arturs Backurs and Piotr Indyk. "Which Regular Expression Patterns are Hard to Match?" FOCS 2016.
4   Karl Bringmann, Allan Grønlund, and Kasper Green Larsen. "A Dichotomy for Regular Expression Membership Testing." arXiv:1611.00918

3.5   Hardness of string problems with small alphabet size

Yi-Jun Chang (University of Michigan – Ann Arbor, US)

In FOCS 2015, Amir Abboud, Arturs Backurs, and Virginia Vassilevska Williams demonstrated conditional lower bounds for fundamental string problems such as RNA folding, Dyck edit distance, and $k - LCS$.

However, all these lower bound proofs require the alphabet size to be large enough to work. For RNA folding, the required alphabet size is 36, making the result biologically irrelevant. For k-LCS, the alphabet size needed is $O(k)$, and it is an open problem whether the same lower bound holds when the alphabet size is a constant independent of k.

In this talk, we show how we can lower the alphabet size requirement for the hardness proofs of RNA folding (from 36 to 4) and Dyck Edit distance (from 48 to 10). We will also discuss some open problems and future work directions.

3.6   Recent insights into counting small patterns

Radu Curticapean (Hungarian Academy of Sciences – Budapest, HU), Holger Dell (Universität des Saarlandes, DE), and Dániel Marx

We consider the problem of counting subgraphs. More specifically, we look at the following problems $\sharp Sub(C)$ for fixed graph classes C: Given as input a graph H from C (the pattern) and another graph G (the host), the task is to count the occurrences of H as a subgraph in G. Our goal is to understand which properties of the pattern class C make the problem $\sharp Sub(C)$ easy/hard. For instance, for the class of stars, we can solve this problem in linear time. For the class of paths however, it subsumes counting Hamiltonian paths and is hence $\sharp$ P-hard.

As it turns out, the notion of $\sharp$ P-hardness fails to give a sweeping dichotomy for the problems $\sharp Sub(C)$, since there exist classes C of intermediate complexity. However, adopting the framework of fixed-parameter tractability, and parameterizing by the size of the pattern, it was shown in 2014 how to classify the problems $\sharp Sub(C)$ as either polynomial-time solvable or $\sharp$ W[1]-hard: A class C lies on the polynomial-time side of this dichotomy iff the graphs appearing in C have vertex-covers of constant size.

In this talk, we introduce a new technique that allows us to view the subgraph counting problem from a new perspective. In particular, it allows for the following applications:

1. A greatly simplified proof of the 2014 dichotomy result, together with almost-tight lower bounds under ETH, which were not achievable before.
2. Faster algorithms for counting k-edge subgraphs, such as k-matchings, with running time n^{ck} for constants $c < 1$.

3.7   Tight Bounds for Subgraph Isomorphism and Graph Homomorphism

Marek Cygan (University of Warsaw, PL)

This tutorial will consists of two parts. First, we will show simple reductions that allow obtaining (slightly) sublinear reductions from CNF-SAT, leading to (slightly) superexponential lower bounds under the Exponential Time Hypothesis. The goal is to expose main ideas used in such reductions.

Second part will be about tight lower bounds for graph homomorphism and subgraph isomorphism (under the ETH), which is a joint work with Fedor Fomin, Alexander Kulikov, Ivan Mihajlin, Alexander Golovnev, Jakub Pachocki, Arkadiusz Socała.

3.8 Popular Conjectures as a Barrier for Dynamic Planar Graph Algorithms

Søren Dahlgaard (University of Copenhagen, DK)

We consider dynamic problems in planar graphs and present hardness results for dynamic shortest paths and related problems. This result is the first of its kind for planar graphs, and we believe that our techniques might be helpful in proving hardness for other problems in planar graphs. In particular we show that, based on the APSP-conjecture, no algorithm can perform dynamic shortest paths in planar graphs faster than $O(\sqrt{(n)})$ query and update time.

3.9 New upper bounds for some basic problems in P

Omer Gold (Tel Aviv University, IL)

I will provide an overview on our recent upper bounds for some basic (geometric) problems in P. Particularly, improved subquadratic bounds for 3-SUM, the first subquadratic-time algorithms for Dynamic Time Warping and Geometric Edit Distance, near-linear decision tree bounds for the discrete Fréchet distance under polyhedral metrics, and reduction relations between Dominance Products and high-dimensional Closest Pair problems. This overview is based on results that appear in [2, 1, 3, 4].

References

1   Omer Gold and Micha Sharir. *Improved Bounds for 3SUM, k-SUM, and Linear Degeneracy.* CoRR, abs/1512.05279, 2015
2   Omer Gold and Micha Sharir. *On the Complexity of the Discrete Fréchet Distance under L_1 and L_∞.* The 31st European Workshop on Computational Geometry (EuroCG), 2015
3   Omer Gold and Micha Sharir. *Dominance Products and Faster Algorithms for High-Dimensional Closest Pair under L_∞.* CoRR, abs/1605.08107, 2016
4   Omer Gold and Micha Sharir. *Dynamic Time Warping and Geometric Edit Distance: Breaking the Quadratic Barrier.* CoRR, abs/1607.05994, 2016

3.10   How Hard is it to Find (Honest) Witnesses?

Isaac Goldstein (Bar-Ilan University – Ramat Gan, IL)

In recent years much effort was put into developing polynomial-time conditional lower bounds for algorithms and data structures in both static and dynamic settings. Along these lines we suggest a framework for proving conditional lower bounds based on the well-known 3-SUM conjecture. Our framework creates a compact representation of an instance of the 3-SUM problem using hashing and domain specific encoding. This compact representation admits false solutions to the original 3-SUM problem instance which we reveal and eliminate until we find a true solution. In other words, from all witnesses (candidate solutions) we figure out if an honest one (a true solution) exists. This enumeration of witnesses is used to prove conditional lower bound on reporting problems that generate all witnesses. In turn, these reporting problems are reduced to various decision problems. These help to enumerate the witnesses by constructing appropriate search data structures. Hence, 3SUM-hardness of the decision problems is deduced.

We utilize this framework to show conditional lower bounds for several variants of convolutions, matrix multiplication and string problems. Our framework uses a strong connection between all of these problems and the ability to find witnesses.

Specifically, we prove conditional lower bounds for computing partial outputs of convolutions and matrix multiplication for sparse inputs. These problems are inspired by the open question raised by Muthukrishnan 20 years ago. The lower bounds we show rule out the possibility (unless the 3-SUM conjecture is false) that almost linear time solutions to sparse input-output convolutions or matrix multiplications exist. This is in contrast to standard convolutions and matrix multiplications that have, or assumed to have, almost linear solutions.

Moreover, we improve upon the conditional lower bounds of Amir et al. for histogram indexing, a problem that has been of much interest recently. The conditional lower bounds we show apply for both reporting and decision variants. For the well-studied decision variant, we show a full tradeoff between preprocessing and query time for every alphabet size > 2. At an extreme, this implies that no solution to this problem exists with subquadratic preprocessing time and $O(1)$ query time for every alphabet size > 2, unless the 3-SUM conjecture is false. This is in contrast to a recent result by Chan and Lewenstein for a binary alphabet. While these specific applications are used to demonstrate the techniques of our framework, we believe that this novel framework is useful for many other problems as well.

3.11 Parameterised graph distance problems

Thore Husfeldt (IT University of Copenhagen, DK)

We study the complexity of computing the diameter and other distance measures in an unweighted, undirected graph. We sketch the ideas behind a tree decomposition-based repeated traversal that computes the diameter in time $n \exp(t \log d)$, where t is the treewidth and d is the diameter ([Husfeldt, IPEC 2016]), which matches a lower bound under the Strong Exponential Time Hypothesis of Abboud, Vassilevska Williams, and Wang [SODA 2016] for constant diameter. We observe that simple arguments establish tight bounds under the same hypothesis when the problem is paraemterised by vertex cover number and (with some help from the audience) domination number.

3.12 Finding Even Cycles

Mathias Bæk Tejs Knudsen (University of Copenhagen, DK)

We study the problem of finding a $2k$-cycle in a graph, for constant values of k. Previous results showed that it is possible to do this in time $O(n^2)$, and we have improved this to $O(m^{(2k/(k+1))})$, where n and m is the number nodes and edges in the graph. Since any graph with $m \gg n^{(1+1/k)}$ edges contains a $2k$-cycle, this bound is at least as good as the $O(n^2)$ bound.

I will tell a little bit about the result and then focus on why it seems difficult to show a conditional lower bound of $n^{2-o(1)}$, since it implies solving a problem related to the Erdos Girth Conjecture. However, this does not rule out the possibility, that it is easy (for whatever definition of "easy") to show a lower bound of $m^{2k/(k+1)-o(1)}$.

3.13 Birthday Repetition: Tool for proving quasi-poly hardness

Young Kun Ko (Princeton University, US)

In this talk we give a broad introduction to Birthday Repetition, a technique introduced to prove quasi-polynomial hardness of "Free" game assuming the Exponential Time Hypothesis. The main observation of the technique is Birthday Paradox, in particular that aggregating the variables in 2-CSPs to a tuple of size $\tilde{O}(\sqrt{n})$ then choosing two tuples at random will have a challenge from original 2-CSP with high probability. No prior material is assumed.

3.14   Tight Bounds for Gomory-Hu-like Cut Counting

Robert Krauthgamer (Weizmann Institute – Rehovot, IL)

A classical result of Gomory and Hu from 1961 shows that in every edge-weighted undirected graph $G = (V, E, w)$, the minimum st-cut values, when ranging over all $s, t \in V$, take at most $|V| - 1$ distinct values. That is, these $\binom{|V|}{2}$ instances exhibit "redundancy" by factor $\Omega(|V|)$. They further showed how to construct a tree on V that stores all minimum st-cut values.

Motivated by this result, we obtain tight bounds for the redundancy factor of several generalizations of minimum st-cut, namely, Multiway-Cut, Multicut, and Group-Cut. A natural application of these bounds is to construct small data structures that store all the cut values for these problems, a la the Gomory-Hu tree. We initiate this direction by giving some upper and lower bounds.

3.15   Advances in fully dynamic algorithms with worst-case update time

Sebastian Krinninger (MPI für Informatik – Saarbrücken, DE)

A major goal in dynamic graph algorithms is to strengthen amortized update time guarantees to to hard worst-case guarantees. As we have learned from recent conditional lower bounds, this unfortunately might not always be possible in many cases. In other cases, it is on open problem how much better the worst-case update time guarantees can get. I will give a short overview and then present my recent contributions in this area.

3.16   Deterministic Time-Space Tradeoffs for k-SUM

Andrea Lincoln (Stanford University, US), Joshua R. Wang, Ryan Williams, and Virginia Vassilevska Williams (Stanford University, US)

Given a set of numbers, the k-SUM problem asks for a subset of k numbers that sums to zero. When the numbers are integers, the time and space complexity of k-SUM is generally studied in the word-RAM model; when the numbers are reals, the complexity is studied in the real-RAM model, and space is measured by the number of reals held in memory at any point.

We present a time and space efficient deterministic self-reduction for the k-SUM problem which holds for both models, and has many interesting consequences. To illustrate:

- 3-SUM is in deterministic time $O(n^2 \lg \lg(n) / \lg(n))$ and space $O\left(\sqrt{\frac{n \lg(n)}{\lg \lg(n)}}\right)$. In general, any polylogarithmic-time improvement over quadratic time for 3-SUM can be converted into an algorithm with an identical time improvement but low space complexity as well.
- 3-SUM is in deterministic time $O(n^2)$ and space $O(\sqrt{n})$, derandomizing an algorithm of Wang.
- A popular conjecture states that 3-SUM requires $n^{2-o(1)}$ time on the word-RAM. We show that the 3-SUM Conjecture is in fact equivalent to the (seemingly weaker) conjecture that every $O(n^{.51})$-space algorithm for 3-SUM requires at least $n^{2-o(1)}$ time on the word-RAM.
- For $k \geq 4$, k-SUM is in deterministic $O(n^{k-2+2/k})$ time and $O(\sqrt{n})$ space.

3.17 Continuous Optimization Based Maximum Flow Algorithms Make Sense

Aleksander Madry (MIT – Cambridge, US)

I will explain how to compute the maximum flow in a graph by iteratively routing electrical flows in the residual graph. The resulting algorithm provides the state of the art running time bounds for the unit capacity maximum flow problem.

3.18 Shortest cycle approximation

Liam Roditty (Bar-Ilan University – Ramat Gan, IL)

Joint work of Liam Roditty, Virginia Vassilevska Williams

We study the problem of determining the girth of an unweighted undirected graph.

In this talk I will survey efficient approximation algorithms with additive and multiplicative approximations from the paper [1].

References
1 Liam Roditty, Virginia Vassilevska Williams. *Subquadratic time approximation algorithms for the girth.* Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms. Society for Industrial and Applied Mathematics, 2012.

3.19   Towards Conditional Lower Bounds for Tree Edit Distance

Oren Weimann (University of Haifa, IL)

The tree edit distance (TED) between two labeled trees T and T' is the minimum cost of transforming one tree into the other by a sequence of elementary operations consisting of deleting and relabeling existing nodes, as well as inserting new nodes. The current fastest algorithm for TED requires $O(n^3)$ time. In terms of conditional lower bounds, the problem has a string edit distance (hence SETH) flavor, yet a cubic (hence APSP) complexity. In my talk, I presented a joint work with Paweł Gawrychowski and Shay Mozes showing a possible first step towards APSP-hardness. Namely, a reduction from APSP to TED under the assumption that in TED we seek the edit distance between every subtree of T and every subtree of T'. All existing TED algorithms actually compute this all subtree-to-subtree information.

Right after the talk, Karl Bringmann who was in the audience came up with a way to remove the subtree-to-subtree assumption in the APSP to TED reduction, thus achieving APSP-hardness for TED. The reduction still required $\Omega(n)$ different labels. In the following few days at Dagstuhl, together with Karl we extended the reduction to a reduction from Max-weighted k-Clique to TED that requires only $O(1)$ different labels.

We are currently writing-up these results and plan to submit them soon to a conference. We are grateful for Dagstuhl and feel that such outcome as the above can only happen in meetings like Dagstuhl.

3.20   Fine-Grained Complexity and Conditional Hardness for Sparse Graphs

Vijaya Ramachandran (University of Texas – Austin, US)

There is a large class of path and cycle problems on graphs that currently have $\tilde{O}(n^3)$[1] time algorithms. Graphs encountered in practice are typically sparse, with the number of edges m being close to linear in n, the number of vertices, or at least with $m << n^2$. When considering sparsity, the current time complexities of these problems split into two classes: the $\tilde{\Theta}(mn)$ class, which includes APSP, Betweenness Centrality, and Minimum-Weight-Cycle, among several other problems, and the $\Theta(m^{3/2})$ class, which includes all problems relating to enumerating and detecting triangles. Here n and m are the number of vertices and edges in the graph. We investigate the fine-grained complexity of these problems on sparse graphs, and our main results are the following:

---

[1] $\tilde{O}$ hides polylog factors. For APSP on dense graphs, we use it to also hide a larger, but sub-polynomial factor

1. **Reductions and Algorithms.** We define the notion of a sparse reduction that preserves graph sparsity, and we present several such reductions for graph problems in the $\tilde{O}(mn)$ class. This gives rise to a rich partial order on graph problems with $\tilde{O}(mn)$ time algorithms, with the Minimum-Weight-Cycle problem as a major source in this partial order, and APSP a major sink. Surprisingly, very few of the known subcubic results are sparse reductions (outside of a few reductions that place Centrality problems in the sub-cubic equivalence class). We develop new techniques in order to preserve sparsity in our reductions, many of which are nontrivial and intricate. Some of our reductions also lead to improved algorithms for various problems on finding simple cycles in undirected graphs.

2. **Conditional Hardness.** We establish a surprising conditional hardness result for sparse graphs: We show that if the Strong Exponential Time Hypothesis (SETH) holds, then several problems in the $\tilde{O}(mn)$ class, including certain problems that are also in the sub-cubic equivalence class such as Betweenness Centrality and Eccentricities, *cannot* have 'sub-mn' time algorithms, i.e., algorithms that run in $O(m^{\alpha} \cdot n^{2-\alpha-\epsilon})$ time, for constants $\alpha \geq 0$, $\epsilon > 0$. In particular, this result means that under SETH, the sub-cubic equivalence class is split into at least two classes when sparsity is taken into account, with triangle finding problems having faster algorithms than Eccentricities or Betweenness Centrality. This hardness result for the $\tilde{O}(mn)$ class is also surprising because a similar hardness result for the sub-cubic class is considered unlikely since this would falsify NSETH (Nondeterministic SETH).

3.21 Computing Min-Cut with truly subquadratic cut queries

Aviad Rubinstein (University of California – Berkeley, US)

I describe preliminary progress on the problem of computing an exact minimum cut of an unknown graph, when the graph is accessed via queries to a cut-value oracle.

3.22 On the oblivious adversary assumption in dynamic problems

Thatchaphol Saranurak (KTH Royal Institute of Technology – Stockholm, SE) and Danupon Nanongkai (KTH Royal Institute of Technology – Stockholm, SE)

Many dynamic randomized algorithms make the "oblivious adversary assumption" i.e. they assume that the adversary fixed all the updates before it sees any output of the dynamic algorithm. This is in contrast to an "adaptive adversary" that see all previous algorithm's outputs before it generates a new update. It is a fundamental question whether the true source of power of randomized dynamic algorithms is the randomness itself or in fact the oblivious adversary assumption.

For example, in dynamic spanning forest problem which plays a central role in the development of dynamic graph algorithms, there is a randomized algorithm with polylog worst-case update time against oblivious adversaries but the best-known algorithms against adaptive adversaries have $O(\sqrt{n})$ update time.

In this talk, I will try to motivate an approach for understanding the power of adaptive adversaries in dynamic spanning forest problem via a problem called "dynamic cut oracle". This problem is interesting for two reasons. First, if it is "hard", this would separate the two models of oblivious vs. adaptive adversaries. That is, it implies that there is no algorithm with polylog worst-case update time against adaptive adversaries for dynamic spanning forest. Second, the technique we used for studying dynamic cut oracle leads to a new exciting algorithm for dynamic spanning forest itself, which indicates that this problem might capture the hardness of dynamic spanning forest problem.

3.23 Subquadratic Algorithms for Succinct Stable Matching

Stefan Schneider (University of California – San Diego, US)

We consider the stable matching problem when the preference lists are not given explicitly but are represented in a succinct way and ask whether the problem becomes computationally easier and investigate other implications. We give subquadratic algorithms for finding a stable matching in special cases of natural succinct representations of the problem, the d-attribute, d-list, geometric, and single-peaked models. We also present algorithms for verifying a stable matching in the same models. We further show that for $d = \omega(log n)$ both finding and verifying a stable matching in the d-attribute and d-dimensional geometric models requires quadratic time assuming the Strong Exponential Time Hypothesis. This suggests that these succinct models are not significantly simpler computationally than the general case for sufficiently large d.

3.24 RNA-Folding: From Hardness to Algorithms

Virginia Vassilevska Williams (Stanford University, US)

A fundamental problem in computational biology is predicting the base-pairing of an RNA secondary structure. Most algorithms for this rely on an algorithm for a simplified version of this problem, RNA-folding, defined as follows: given a sequence S of letters over the alphabet $\{A, U, C, G\}$ where A can only be paired with U and C can only be paired with G, determine the best "folding" of S, i.e. a maximum size *nested* pairing of the symbols

of S. For instance, in the sequence ACUG the best pairing is either matching A with U, or matching C with G, but not both as that pairing wouldn't be nested.

A dynamic programming algorithm from 1980 by Nussinov and Jacobson solves the RNA-folding problem on an n letter sequence in $O(n^3)$ time. Despite many efforts, until recently, the best algorithms for RNA-folding only shaved small logarithmic factors over this cubic running time.

Recent work [1] explained why it has been so difficult to obtain faster algorithms: if one can solve RNA-folding on n length strings faster than one can currently multiply n by n matrices, then the Clique problem would have surprisingly fast algorithms. The current fastest algorithm to multiply n by n matrices runs in $O(n^{2.373})$ time and the fastest known Clique algorithms use this result. Obtaining an $O(n^{2.36})$ time algorithm for RNA-folding would thus be potentially difficult as it would imply a breakthrough for Clique algorithms and potentially also for matrix multiplication.

While this hardness result is appealing, it does not explain the seeming n^3 barrier. No better hardness seemed possible to us, and thus it became increasingly more plausible that RNA-folding should have a faster algorithm and in fact one using fast matrix multiplication. Indeed, this turned out to be true. In this talk I will strive to give some insight into the first truly subcubic time algorithm for the problem.

References

1   Abboud, Amir, Arturs Backurs, and Virginia Vassilevska Williams. *If the current clique algorithms are optimal, so is Valiant's parser.* Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on. IEEE, 2015.

3.25 Amortized Dynamic Cell-Probe Lower Bounds from Four-Party Communication

Huacheng Yu (Stanford University, US)

This paper develops a new technique for proving amortized, randomized cell-probe lower bounds on dynamic data structure problems. We introduce a new randomized nondeterministic four-party communication model that enables "accelerated", error-preserving simulations of dynamic data structures.

We use this technique to prove an $\Omega(n(\log n/\log\log n)^2)$ cell-probe lower bound for the dynamic 2D weighted orthogonal range counting problem (2D-ORC) with $n/poly\log n$ updates and n queries, that holds even for data structures with $\exp(-\tilde{\Omega}(n))$ success probability. This result not only proves the highest amortized lower bound to date, but is also tight in the strongest possible sense, as a matching upper bound can be obtained by a deterministic data structure with worst-case operational time. This is the first demonstration of a "sharp threshold" phenomenon for dynamic data structures.

Our broader motivation is that cell-probe lower bounds for exponentially small success facilitate reductions from dynamic to static data structures. As a proof-of-concept, we show that a slightly strengthened version of our lower bound would imply an $\Omega((\log n/\log\log n)^2)$

lower bound for the static 3D-ORC problem with $O(n \log^{O(1)} n)$ space. Such result would give a near quadratic improvement over the highest known static cell-probe lower bound, and break the long standing $\Omega(\log n)$ barrier for static data structures.

4   Open problems

The following open problems were contributed by the seminar attendees, and compiled and edited by the organizers.

4.1   Parameterizing problems in P by treewidth

Fedor V. Fomin (University of Bergen, NO)

Background.   Let t be the treewidth of an input graph. Many **NP**-hard problems, particularly those expressible in MSOL, are solvable in $f(t)n$ time and there are lower bounds on the (exponential) function f conditioned on the Strong Exponential Time Hypothesis (SETH) [26]. For problems in **P** the picture is less clear. Consider your favorite problem Π in **P** solvable in $T_\Pi(n)$ time on a graph wtih n vertices. Some problems Π admit algorithms running in $\text{poly}(t) \cdot o(T_\Pi(n))$ time whereas others do not. For example, Abboud et al. [5] proved that *Diameter* can be solved in $2^{O(t \log t)} n^{1+o(1)}$ time, yet a $2^{o(t)} n^{2-\epsilon}$ time algorithm would refute SETH. On the other hand, *maximum cardinality matching* can be solved in randomized $O(t^3 \cdot n \log n)$-time [31].

Question.   Classify graph problems in **P** according to their dependence on treewidth. Which problems admit $f(t) \cdot n^{1+o(1)}$-time algorithms with polynomial f, and which require exponential f? A specific goal is the determine whether *maximum weight perfect matching* has an $\tilde{O}(\text{poly}(t)n)$ algorithm, for integer weights from a polynomial range.

Main paper reference: Abboud et al. [5], Fomin et al. [31].

4.2   Approximate all-pairs shortest paths

Amir Abboud (Stanford University, US)

Background.   In unweighted, undirected graphs, we can compute All Pairs Shortest Paths (APSP) in $O(n^3)$ time with a fast "combinatorial" algorithm, or in $O(n^\omega)$ time, where $\omega < 2.373$ is the matrix multiplication exponent. It is conjectured that a truly subcubic combinatorial algorithm does not exist, which is equivalent to the combinatorial Boolean matrix multiplication conjecture.

What about approximation algorithms? The best kind of approximation is an additive $+2$, so that for all pairs u, v we return a value that is between $d(u, v)$ and $d(u, v) + 2$. Dor, Halperin, and Zwick [30] presented a *combinatorial* algorithm with runtime $\tilde{O}(n^{7/3})$. Note

that this runtime is currently even better that $O(n^\omega)$, and has the advantage of being practical.

Questions. Is there a conditional lower bound for +2-APSP? Can we show that a combinatorial algorithm must spend $n^{7/3-o(1)}$ time? Would a faster non-combinatorial algorithm require improvements to ω? Alternatively, is there an $\tilde{O}(n^2)$ time algorithm for +2-APSP?

Main paper reference: Dor, Halperin, and Zwick [30].

4.3 Approximate diameter

Amir Abboud (Stanford University, US)

Background. Computing the diameter of a sparse graph in truly subquadratic time refutes SETH: Roditty and Vassilevska Williams [62] showed that a $(3/2 - \varepsilon)$-approximation to the diameter requires $n^{2-o(1)}$ time, even on a sparse unweighted undirected graph under SETH. On the other hand, there are algorithms [62, 22] that give a (roughly) 3/2 approximation in $\tilde{O}(m\sqrt{n})$ time on unweighted graphs, or $\tilde{O}(\min\{m^{3/2}, mn^{2/3}\})$ time on weighted graphs. Extending these algorithms further, Cairo et al. [19] showed that for all integers $k \geq 1$, there is an $\tilde{O}(mn^{1/(k+1)})$ time algorithm that approximates the diameter of an undirected unweighted graph within a factor of (roughly) $2 - 1/2^k$.

Question. If we insist on near-linear runtime, what is the best approximation factor we can get? It is easy to see that a 2-approximation can be achieved in linear time, but what about an α-approximation, where $3/2 \leq \alpha < 2$?

Main paper reference: Roditty and Vassilevska W. [62].

4.4 Finding cycles and approximating the girth

Mathias Bæk Tejs Knudsen (University of Copenhagen, DK) and Liam Roditty (Bar-Ilan University – Ramat Gan, IL)

Background. Consider an unweighted undirected graph $G = (V, E)$. The *girth* of G is the length of the shortest cycle. The problem of detecting 3-cycles (and odd cycles of any length) is reducible to matrix multiplication and there are reductions in the reverse direction; see [66]. Yuster and Zwick [67] showed that detecting $2k$-cycles can be computed in $O(f(k)n^2)$ time, where f is exponential.

Question. For any fixed constant k, give a conditional lower bound, showing that there does not exist an algorithm deciding whether G contains a $2k$-cycle in time $O(f(k)n^{2-\varepsilon})$ for any $\varepsilon > 0$, or one running in $O(f(k)m^{2k/(k+1)-\varepsilon})$ time, where m is the number of edges.

Main paper reference: Yuster and Zwick [67].

Question. Prove or disprove the following conjecture: There exists a truly subquadratic algorithm for finding a 4-cycle in a graph if and only if there exists a truly subquadratic algorithm for finding a multiplicative $(2 - \varepsilon)$-approximation of the girth.

Question. Prove or disprove the following conjecture from [63]: the problem of detecting a 3-cycle in a graph G without 4- and 5-cycles requires $n^{2-o(1)}$ time. Note that if there exists a subquadratic $(2 - \epsilon)$-approximation for the girth, it must be able to detect 3-cycles in graphs without 4- and 5-cycles. See [63] for more details.

Main paper reference: Roditty and Vassilevska W. [63].

4.5 Minimum cycle problem in directed graphs

Virginia Vassilevska Williams (Stanford University, US)

Background. Given an unweighted *directed* graph $G = (V, E)$ on n vertices, the problem is to find a shortest cycle in G. The potentially simpler *Girth* problem asks to compute just the length of the shortest cycle.

The girth and the minimum cycle can be computed in $O(n^\omega)$ time exactly, as shown by Itai and Rodeh [49], where $\omega < 2.373$. It is easy to see that the minimum cycle problem is at least as hard as finding a triangle in a graph. In fact, even obtaining a $(2 - \delta)$-approximation for the girth for any constant $\delta > 0$ is at least as hard as triangle detection. The fastest algorithm for the Triangle problem in n node graphs runs in $O(n^\omega)$ time.

Question. Is there any $O(1)$-approximation algorithm for the girth that runs faster than $O(n^\omega)$ time? In recent work, Pachocki, Roditty, Sidford, Tov, and Vassilevska Williams [59] showed that for any integer k, there is an $\tilde{O}(mn^{1/k})$ time $O(k \log n)$ approximation algorithm for the Minimum Cycle problem. Thus, in nearly linear time, one can obtain an $O(\log^2 n)$-approximation. Can one improve the approximation factor further? Can one even obtain a constant factor approximation in linear time?

Main paper reference: Pachocki et al. [59].

4.6 Linear Programming

Aleksander Madry (MIT – Cambridge, US)

Background. Consider a linear program of the following form: minimize $c^T x$ subject to $Ax \geq b$, where A is an d-by-n constraint matrix. Suppose that we could solve any such LP in time

$$\widetilde{O}\left(\left(\mathrm{nnz}(A) + d^2\right) d^\delta \log L\right),$$

where $\mathrm{nnz}(A)$ is the number of non-zero entries of A, L is the bound on the bit complexity of the input entries, and δ is a positive constant.

Question. Is there some value of δ for which the above (hypothetical) running time bound would disprove any of the popular hardness conjectures?

In [57], it is shown that one can achieve the above running time bound for $\delta = \frac{1}{2}$.

Main paper reference: Lee and Sidford [57].

4.7 Fully dynamic APSP

Sebastian Krinninger (MPI für Informatik – Saarbrücken, DE)

Background. In the fully dynamic all-pairs shortest paths (APSP) problem we are interested in maintaining the distance matrix of a graph under insertions and deletions of nodes. Demetrescu and Italiano [28] showed that the distance matrix can be updated in *amortized* time $\tilde{O}(n^2)$ after each node update. The current fastest *worst case* algorithms have update times of $\tilde{O}(n^{2+2/3})$ (randomized Monte Carlo [7]) and $\tilde{O}(n^{2+3/4})$ (deterministic [65]).

Questions. Can the worst case update time $\tilde{O}(n^2)$ be achieved? A barrier for current algorithmic approaches is $n^{2.5}$. Is there a conditional lower bound showing this to be a true barrier?

Main paper reference: Abraham et al. [7].

4.8 Dynamic reachability in planar graphs

Søren Dahlgaard (University of Copenhagen, DK)

Background. *Dynamic reachability* in a planar graph G is the problem of maintaining a data structure supporting the following operations: (i) Insert a directed edge (u,v) into G, (ii) delete an edge from G, and (iii) query whether v is reachable from u in G.

An algorithm with update and query time $\tilde{O}(\sqrt{n})$ is known (Diks and Sankowski [29]) for dynamic *plane graphs*—that is, the graph is dynamic but the plane embedding is fixed.

Question. Does an $n^{1/2-\Omega(1)}$ algorithm exist or is there a conditional $n^{1/2-o(1)}$ hardness result? Any polynomial hardness result would be interesting. A good place to start for the latter part would be the recent paper by Abboud and Dahlgaard [3] about hardness for dynamic problems in planar graphs.

Main paper reference: Abboud and Dahlgaard [3].

4.9   Static hardness for planar graphs

Søren Dahlgaard (University of Copenhagen, DK)

Background.   An important direction is to show conditional hardness for important problems, even on restricted (easier) classes of graphs, e.g., planar graphs. Abboud and Dahlgaard [3] recently showed hardness for several *dynamic* problems in planar graphs, but nothing is known for static problems.

Question.   On planar graphs, many problems (such as shortest paths, multi-source multi-sink max-flow, etc.) run in near-linear time. Can we show that some problem does not? No hardness results are known for any static problem in **P** on planar graphs. Two candidate problems to consider are *diameter* and *sum of distances*. Both require subquadratic time (Cabello [18]), but it may still be possible to show a hardness result, e.g., $n^{3/2-o(1)}$ hardness.

Main paper reference: Cabello [18].

4.10   Sparse reductions for graph problems

Vijaya Ramachandran (University of Texas – Austin, US)

Background.   Many graph problems are known to be as hard as APSP on dense graphs [66, 4, 64], in the sense that a subcubic algorithm for any of them implies a subcubic algorithm for all of them. When the graph sparsity is taken into account, these problems currently are no longer in a single class: many have $\tilde{O}(mn)$-time algorithms whereas finding minimum weight triangle and related problems have $\tilde{O}(m^{3/2})$-time algorithms. Most known fine-grained reductions between graph problems do not preserve the graph sparsity. Until recently, the only examples of sparseness-preserving truly subcubic reductions appeared in [4]. Agarwal and Ramachandran [8] presented several more such reductions, strengthening the connections between problems with $\tilde{O}(mn)$-time algorithms. A reduction from CNF-SAT to Diameter was presented in [62] to give SETH-hardness results for Diameter and Eccentricities. The notion of a *sub-mn time bound* was formalized later, in [8], where it was observed that the reduction in [62] gives SETH-hardness for any sub-mn time bound for these problems.

Questions.   Is there a sparseness-preserving, $\tilde{O}(n^2)$ time reduction from undirected weighted All Nodes Shortest Cycles (ANSC) to APSP? Is there a sparseness-preserving, $\tilde{O}(m+n)$ time reduction from undirected Min-Wt-Cycle to either Radius or Eccentricities? Is it SETH-hard to find a sub-mn bound for Min-Wt-Cycle or an $O(n^2+$ sub-$mn)$ bound on APSP?

Main paper reference: Agarwal and Ramachandran [8].

4.11 Hardness for partially dynamic graph problems

Søren Dahlgaard (University of Copenhagen, DK)

Background. Many results show hardness for *fully*-dynamic problems in graphs, but the techniques do not seem to extend well to amortized lower bounds in the incremental and decremental cases. (See Abboud and Vassilevska Williams [2], Henzinger, Krinninger, Nanongkai, and Saranurak [46], Kopelowitz, Pettie and Porat [55], and Dahlgaard [27] for some initial results on incremental/decremental problems.)

Question. Develop general techniques for showing amortized hardness of partially dynamic problems in graphs. One candidate problem is decremental single-source reachability. A result of Chechik, Hansen, Italiano, Lacki, and Parotsidis [23] shows that $\tilde{O}(m\sqrt{n})$ total time is sufficient. Is it necessary?

4.12 Hardness of vertex connectivity

Veronika Loitzenbauer (Universität Wien, AT)

Background. A connected undirected graph is k-vertex (resp. edge) connected if it remains connected after any set of at most $k-1$ vertices (edges) is removed from the graph. A strongly connected directed graph is k-vertex (edge) connected if it remains strongly connected after any set of at most $k-1$ vertices (edges) is removed from the graph. The vertex (edge) connectivity of a graph is the maximum value of k such that the graph is k-vertex (edge) connected.

The edge-connectivity λ of an undirected graph can be determined in time $O(m \log^2 n \log^2 \log n)$ [47, 54], and for directed graphs in time $O(\lambda m \log(n^2/m))$ [37]. In contrast, the vertex-connectivity κ can only be computed in time $O((n + \min\{\kappa^{5/2}, \kappa n^{3/4}\})m)$ [38], where for undirected graphs m can be replaced by kn.

Question. To *check* k-vertex connectivity means to either confirm that $\kappa \geq k$ or to find a set of $k-1$ vertices that disconnects the graph. Even when k is constant, no $o(n^2)$ time (or $o(mn)$ time for directed graphs) algorithms are known for checking k-connectivity. Is there a conditional superlinear lower bound?

Main paper reference: Gabow [38].

4.13 Parity and mean-payoff games

Veronika Loitzenbauer (Universität Wien, AT)

Background. *Parity games*, and their generalization *mean-payoff games*, are among the rare "natural" problems in NP ∩ co-NP (and in UP ∩ co-UP [52]) for which no polynomial-time

algorithm is known. Both parity games and mean-payoff games are 2-player games played by taking an infinite walk on a directed graph; one of the vertices is designated the start vertex. In parity games each vertex is labeled by an integer in $[0, c]$; in mean payoff games each edge is labeled by an integer in $[-W, W]$. (See [53] for a description of the game.) The algorithmic question is to decide, for each start vertex, which of the two players wins the game and to construct a corresponding winning strategy. Parity games can be reduced to mean-payoff games with $W = n^c$. Very recently, quasi-polynomial $O(n^{\log c})$ time algorithms for parity games were discovered [20, 51]. The best known algorithms for mean-payoff games run in pseudo-polynomial time $O(mnW)$ [16] and randomized sub-exponential time $O(2^{\sqrt{n \log n}} \log W)$ [15].

Questions.   Is there a polynomial-time algorithm for parity or mean-payoff games? Are there conditional superlinear lower bounds on these problems?

4.14   Unknotting

Seth Pettie (University of Michigan – Ann Arbor, US)

Background.   A *knot* is a closed, non-self-intersecting polygonal chain in $\mathbb{R}^3$. Two knots are *equivalent* if one can be continuously deformed into the other without self-intersection. The *unknot* problem is to decide if a knot is equivalent to one that is embeddable in the plane.

Knots can be represented combinatorially, by projecting the polygonal chain onto $\mathbb{R}^2$, placing a vertex wherever two edges intersect. The result is a 4-regular planar graph (possibly with loops and parallel edges) where each vertex carries a bit indicating which pair of edges is "over" and which pair is "under." *Reidemeister moves* (a small set of transformations on the knot diagram) suffice to transform any knot diagram to one of its equivalent representations.

The complexity of *unknot* and related problems (e.g., are two knots equivalent?, can two knots simultaneously embedded in $\mathbb{R}^3$ be untangled?) are known to be in **NP** [45] and solvable in $2^{O(n)}$ time [45, 50].

Questions.   Given a plane knot diagram with n intersections, can *unknot* or *knot-equivalence* be solved in time near-linear in n? If not, are there conditional lower bounds that show even some polynomial hardness?

4.15   3-Collinearity (general position testing)

Omer Gold (Tel Aviv University, IL)

Background.   A set S of n points in $\mathbb{R}^2$ is said to be in *general position* if there do not exist three points in S that lie on a line. The *3-Collinearity* problem is to test whether S is in general position. The 3-Collinearity problem is known to be as hard as 3SUM, and an algorithm that runs in $O(n^2)$ time is known.

Questions. The question is whether the $O(n^2)$ algorithm is optimal or whether it can be solved in $o(n^2)$ time. Recent subquadratic algorithms for 3SUM [12, 44, 34, 41] indicate that polylogarithmic improvements should be possible. A related question is whether there is an $O(n^{2-\epsilon})$-depth decision tree for 3-Collinearity; see [44, 13].

Main paper reference: Gajentaan and Overmars [39].

4.16 Element uniqueness in $X + Y$

Omer Gold (Tel Aviv University, IL)

Background. Given two sets X and Y, each of n real numbers, determine whether all the elements of $X + Y = \{x + y \mid x \in X, y \in Y\}$ are distinct. A somewhat stronger variant of this problem is to sort $X + Y$.

The decision tree complexity of sorting $X + Y$ and Element Uniqueness in $X + Y$ was shown to be $O(n^2)$ by Fredman [33].

Question. Can these problems can be solved in $o(n^2 \log n)$ time, even for the special case $X = Y$?

4.17 Histogram indexing

Isaac Goldstein (Bar-Ilan University – Ramat Gan, IL)

Background. The *histogram* $\psi(T)$ of a string $T \in \Sigma^*$ is a $|\Sigma|$-length vector containing the number of occurrences of each letter in T. The *histogram indexing problem* (aka *jumbled indexing*) is to preprocess a string T to support the following query: given a histogram vector ψ, decide whether there is a substring T' of T such that $\psi(T') = \psi$.

The state-of-the-art algorithm for histogram indexing [21] preprocesses a binary text T in $O(n^{1.859})$ time and answers queries in $O(1)$ time. Over a d-letter alphabet the preprocessing and query times are $\tilde{O}(n^{2-\delta})$ and $\tilde{O}(n^{2/3+\delta(d+13)/6})$, for any $\delta \geq 0$. On the lower bound side [10, 42], the 3SUM conjecture implies that it is impossible to simultaneously improve $n^{2-\delta}$ preprocessing and $n^{\delta(d/2-1)}$ query time by polynomial factors, where $\delta \leq 2/(d-1)$ and $d \geq 3$.

Question. Are there any non-trivial lower bounds on histogram indexing when $d = 2$? Is it possible to close the gap between the lower and upper bounds in general, or to base the hardness off of a different conjecture than 3SUM?

Main paper reference: Chan and Lewenstein [21].

4.18   Integer programming

Fedor V. Fomin (University of Bergen, NO)

Background.   The objective of *Integer Programming* (IP) is to decide, for a given $m \times n$ matrix A and an m-vector $b = (b_1, \ldots, b_m)$, whether there is a non-negative integer n-vector x such that $Ax = b$. In 1981, Papadimitriou [61] showed that (IP) is solvable in pseudo-polynomial time on instances for which the number of constraints m is constant. The rough estimation of the running time of Papadimitriou's algorithm is $n^{O(m)} \cdot d^{O(m^2)}$, where d bounds the magnitude of any entry in A and b. The best known lower bound is $n^{o(\frac{m}{\log m})} d^{o(m)}$ [32], assuming the Exponential Time Hypothesis (ETH).

Question.   Is it possible to narrow the gap between algorithms for IP and the ETH-hardness of IP?

Main paper reference: Fomin et al. [32].

4.19   All-pairs min-cut and generalizations

Robert Krauthgamer (Weizmann Institute – Rehovot, IL)

Background.   The *all-pairs min-cut* problem is, given an edge-capacitated undirected graph $G = (V, E, c)$, to compute the minimum s-t cut over all pairs $s, t \in V$. Gomory and Hu [43] showed the problem is reducible to $n-1$ s-t min-cut instances, and moreover, all $\binom{n}{2}$ min-cuts can be represented by a capacitated tree T on the vertex set V. On unweighted graphs, the construction of T takes time $\tilde{O}(mn)$ [14, 60].

Generalizations of this problem include finding the min-cut separating every triple $(r, s, t) \in V^3$, which is **NP**-hard, and finding the min-cuts separating all pairs of k-sets $\{s_1, \ldots, s_k\}$ from $\{t_1, \ldots, t_k\}$. See [24].

Questions.   Are there superlinear conditional lower bounds for all-pairs min-cut/Gomory-Hu tree construction? (Refer to [6] for conditional lower bounds for variants of the problem on directed graphs.) Are there non-trivial conditional lower bounds for all-triplets approximate min-cut, or all-k-sets min-cut?

4.20   Parameterizing string algorithms by compressibility

Oren Weimann (University of Haifa, IL)

Background.   The broad idea can be illustrated with a lower bound for string edit distance: Given two strings of length N whose compressed length (say, using Lempel-Ziv compression) is n, it is known that their edit distance can be computed in $O(nN)$ time. Is it possible to prove an $\Omega(nN)$ conditional lower bound? The known conditional lower bound [11, 1, 17]

reduces CNF-SAT (with n variables) to string edit distance by creating two strings each consisting of $O(2^{n/2})$ blocks. To make such a reduction suitable for proving $\Omega(nN)$ lower bound, one needs to generate instead two strings whose length is much more than $2^{n/2}$ but that compress to much less than $2^{n/2}$.

4.21 Reductions from low complexity to high complexity

Amir Abboud (Stanford University, US)

Background. We know that improving the runtime of our 10-Clique algorithms improves the runtime of our 100-Clique algorithms. E.g., if 10-Clique can be solved in $O(n^5)$, then 100-Clique can be solved in $O(n^{50})$. In general, we have many examples of reductions showing that a faster algorithm for a problem with best known runtime $O(n^a)$, implies a faster algorithm for a problem with runtime $O(n^b)$, where $a \leq b$.

However, we have no interesting reductions in the other way, showing that improvements over n^b imply improvements over n^a, where $a < b$. In particular, we do not know how to use an algorithm that solved 100-Clique in $O(n^{50})$ or even $O(n^{11})$ time, to speed up the known algorithms for 10-Clique.

Could it be that such reductions, from low complexity to high complexity, do not exist? It is not hard to construct artificial problems where this can be done, but what about the natural problems we typically study: Clique, Orthogonal Vectors, k-SUM, APSP, LCS, etc. Can we show that a fine-grained reduction from 10-Clique to 100-Clique is unlikely due to some surprising consequences? Another candidate is 3SUM (for which the complexity is n^2) vs. APSP (for which the complexity is $N^{1.5}$, where N is the input size). We repeatedly ask if faster 3SUM implies faster APSP, but maybe proving such a result (via fine-grained reductions) has unexpected consequences?

On the other hand, it would be of great interest to find examples of such reductions between interesting and natural problems.

4.22 Stable matching in the two-list model

Stefan Schneider (University of California – San Diego, US)

Background. Gale and Shapley's stable matching [40] algorithm runs in $O(n^2)$ time (linear in the input size) and it is known that $\Omega(n^2)$ is optimal if the preference lists are arbitrary. Künnemann, Moeller, Paturi, and Schneider [56] studied the complexity of stable matching when the preference lists are constrained, and encoded in some succinct manner. Many succinct input models nonetheless require $n^{2-o(1)}$ time, conditioned on SETH.

Question. A problem left open by [56] is *two-list stable matching*. A matching market in the two-list model consists of two sets M and W, both of size n, and permutations π_1, π_2 on M and σ_1, σ_2 on W. The preference list of each agent $m \in M$ is either σ_1 or σ_2 and the preference list of each agent $w \in W$ is either π_1 or π_2. The input size is $O(n)$. The goal is

to find a stable matching in the resulting matching market. Can this problem be solved in linear time, or is there a superlinear conditional lower bound?

Main paper reference: Kuennemann et al. [56]

4.23   Boolean vs. real maximum inner product

Stefan Schneider (University of California – San Diego, US)

Background.   In the *maximum inner product* problem we are given two sets of d-dimensional vectors U and V of size n as well as a threshold l. The problem is to decide if there is a pair $u \in U, v \in V$ such that their inner product $u \cdot v$ is at least l. If the vectors are Boolean, then a randomized algorithm by Alman and Williams [9] solves the problem in time $n^{2-1/\Theta(c \log^2 c)}$ where $d = c \log n$. In contrast, if the vectors are real or integer, then using ray-shooting techniques [58] we can solve the problem in time $n^{2-1/\Theta(d)}$. This leaves a large gap between the two problems. In particular, the Boolean case is strongly subquadratic if $d = O(\log n)$, while the real case is only strongly subquadratic for constant d. The conditional lower bounds of [9] show that any $n^{2-\epsilon}$ algorithm when $d = \omega(\log n)$ refutes SETH.

Questions.   Can the gap between the boolean and integer/real case be closed, with a better maximum inner product algorithm? If the gap is natural, can it be explained with a stronger conditional lower bound on (real or integer) maximum inner product?

4.24   Hardness of Approximating NP-hard Problems

Seth Pettie (University of Michigan – Ann Arbor, US)

Background.   Many approximation algorithms for **NP**-hard problems run in polynomial time, but not linear time. This is often due to the use of general LP or SDP solvers, but not always. To take two examples, the *chromatic index* (edge coloring) and *minimum degree spanning tree* problems are **NP**-hard, but can both be approximated to within 1 of optimal in $\tilde{O}(m\sqrt{n})$ time [36] and $\tilde{O}(mn)$ time [35], respectively.

Question.   Prove superlinear conditional lower bounds on the time complexity of *any* approximation problem, whose exact version is **NP**-hard.

4.25 Chromatic index/edge coloring

Marek Cygan (University of Warsaw, PL)

Background. The *chromatic index* of a graph is the least number of colors needed for a proper edge-coloring. Vizing's theorem implies that the chromatic index is either Δ or $\Delta + 1$ (where Δ is the maximum degree), but determining which one is NP-hard. The NP-hardness reduction of Holyer [48] reduces 3SAT to a 3-regular graph on $O(n)$ vertices, so the ETH implies a $2^{\Omega(n)}$ lower bound. There is an $O^*(2^m)$ algorithm for chromatic index, by reduction to vertex coloring, so the hardness is well understood when $m = O(n)$.

Questions. Does the ETH rule out a $2^{o(m)}$ algorithm for chromatic index on dense graphs? Is there, for example, an $n^{O(n)}$ or $2^{n^{2-\epsilon}}$-time algorithm?

4.26 Communication Complexity of Approximate Hamming Distance

Raphaël Clifford (University of Bristol, GB)

Background. Consider strings P of length n and T of length $2n$. Alice has the whole of P and the first half of T. That is she has P and $T[0, \ldots, n-1]$. Bob has the second half of T, that is $T[n, \ldots, 2n-1]$. Alice sends one message to Bob and Bob has to output a $(1 + \epsilon)$ multiplicative approximation of $\mathrm{HD}(P, T[i, \ldots, i+n])$ for all $i \in [n]$ where HD is the Hamming Distance.

In [25] a $O(\sqrt{n} \log n / \epsilon^2)$ bit communication protocol was given.

Question. Is there a matching lower bound for the randomized one-way communication complexity of this problem?

Main paper reference. Clifford and Starikovskaya [25].

References

1 A. Abboud, A. Backurs, and V. Vassilevska Williams. Tight hardness results for LCS and other sequence similarity measures. In *Proceedings 56th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 59–78, 2015.

2 A. Abboud and V. Vassilevska Williams. Popular conjectures imply strong lower bounds for dynamic problems. In *Proceedings 55th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 434–443, 2014.

3 Amir Abboud and Søren Dahlgaard. Popular conjectures as a barrier for dynamic planar graph algorithms. In *Proceedings 57th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 477–486, 2016. `doi:10.1109/FOCS.2016.58`.

4 Amir Abboud, Fabrizio Grandoni, and Virginia Vassilevska Williams. Subcubic equivalences between graph centrality problems, APSP and diameter. In *Proceedings of the 26th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1681–1697, 2015. `doi:10.1137/1.9781611973730.112`.

5 Amir Abboud, Virginia Vassilevska Williams, and Joshua R. Wang. Approximation and fixed parameter subquadratic algorithms for radius and diameter in sparse graphs. In *Proceedings of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 377–391, 2016. `doi:10.1137/1.9781611974331.ch28`.

6 Amir Abboud, Virginia Vassilevska Williams, and Huacheng Yu. Matching triangles and basing hardness on an extremely popular conjecture. In *Proceedings 47th Annual ACM Symposium on Theory of Computing*, pages 41–50, 2015. `doi:10.1145/2746539.2746594`.

7 Ittai Abraham, Shiri Chechik, and Sebastian Krinninger. Fully dynamic all-pairs shortest paths with worst-case update-time revisited. In *Proceedings of the 28th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 440–452, 2017. `doi:10.1137/1.9781611974782.28`.

8 Udit Agarwal and Vijaya Ramachandran. Fine-grained complexity and conditional hardness for sparse graphs. *CoRR*, abs/1611.07008, 2016.

9 Josh Alman and Ryan Williams. Probabilistic polynomials and hamming nearest neighbors. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*. IEEE, 2015.

10 Amihood Amir, Timothy M. Chan, Moshe Lewenstein, and Noa Lewenstein. On hardness of jumbled indexing. In *International Colloquium on Automata, Languages and Programming (ICALP)*, pages 114–125, 2014.

11 A. Backurs and P. Indyk. Edit distance cannot be computed in strongly subquadratic time (unless SETH is false). In *Proceedings 47th Annual ACM Symposium on Theory of Computing (STOC)*, pages 51–58, 2015.

12 I. Baran, E. D. Demaine, and M. Pătraşcu. Subquadratic algorithms for 3SUM. *Algorithmica*, 50(4):584–596, 2008.

13 L. Barba, J. Cardinal, J. Iacono, S. Langerman, A. Ooms, and N. Solomon. Subquadratic algorithms for algebraic generalizations of 3SUM. *CoRR*, abs/1612.02384, 2016.

14 Anand Bhalgat, Ramesh Hariharan, Telikepalli Kavitha, and Debmalya Panigrahi. An $o(mn)$ Gomory-Hu tree construction algorithm for unweighted graphs. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing (STOC)*, pages 605–614, 2007. `doi:10.1145/1250790.1250879`.

15 Henrik Björklund and Sergei G. Vorobyov. A combinatorial strongly subexponential strategy improvement algorithm for mean payoff games. *Discrete Applied Mathematics*, 155(2):210–229, 2007. `doi:10.1016/j.dam.2006.04.029`.

16 Lubos Brim, Jakub Chaloupka, Laurent Doyen, Raffaella Gentilini, and Jean-François Raskin. Faster algorithms for mean-payoff games. *Formal Methods in System Design*, 38(2):97–118, 2011. `doi:10.1007/s10703-010-0105-x`.

17 K. Bringmann and M. Künnemann. Quadratic conditional lower bounds for string problems and dynamic time warping. In *Proceedings 56th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 79–97, 2015.

18 Sergio Cabello. Subquadratic algorithms for the diameter and the sum of pairwise distances in planar graphs. In *Proceedings of the 28th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2143–2152, 2017. `doi:10.1137/1.9781611974782.139`.

19 Massimo Cairo, Roberto Grossi, and Romeo Rizzi. New bounds for approximating extremal distances in undirected graphs. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 363–376, 2016.

20 C. S. Calude, S. Jain, B. Khoussainov, W. Li, and F. Stephan. Deciding parity games in quasipolynomial time. In *Proceedings 49th Annual ACM Symposium on Theory of Computing (STOC)*, 2017.

21 Timothy Chan and Moshe Lewenstein. Clustered integer 3SUM via additive combinatorics. In *Proceedings 47th ACM Symposium on Theory of Computing (STOC)*, pages 31–40, 2015.

22 S. Chechik, D. Larkin, L. Roditty, G. Schoenebeck, R. E. Tarjan, and V. Vassilevska Williams. Better approximation algorithms for the graph diameter. In *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1041–1052, 2014.

23 Shiri Chechik, Thomas Dueholm Hansen, Giuseppe F. Italiano, Jakub Lacki, and Nikos Parotsidis. Decremental single-source reachability and strongly connected components in $\tilde{O}(m\sqrt{n})$ total update time. In *Proceedings 57th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 315–324, 2016. `doi:10.1109/FOCS.2016.42`.

24 Rajesh Chitnis, Lior Kamma, and Robert Krauthgamer. Tight bounds for gomory-hu-like cut counting. In *Proceedings 42nd International Workshop on Graph-Theoretic Concepts in Computer Science (WG)*, pages 133–144, 2016. `doi:10.1007/978-3-662-53536-3_12`.

25 Raphaël Clifford and Tatiana A. Starikovskaya. Approximate hamming distance in a stream. In *43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, July 11-15, 2016, Rome, Italy*, pages 20:1–20:14, 2016.

26 Marek Cygan, Fedor V. Fomin, Lukasz Kowalik, Daniel Lokshtanov, Dániel Marx, Marcin Pilipczuk, Michal Pilipczuk, and Saket Saurabh. *Parameterized Algorithms*. Springer, 2015. `doi:10.1007/978-3-319-21275-3`.

27 Søren Dahlgaard. On the hardness of partially dynamic graph problems and connections to diameter. In *Proceedings 43rd International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 48:1–48:14, 2016. `doi:10.4230/LIPIcs.ICALP.2016.48`.

28 C. Demetrescu and G. F. Italiano. A new approach to dynamic all pairs shortest paths. *Journal of the ACM*, 51(6):968–992, 2004.

29 Krzysztof Diks and Piotr Sankowski. Dynamic plane transitive closure. In *Proceedings of 15th Annual European Symposium on Algorithms (ESA)*, pages 594–604, 2007. `doi:10.1007/978-3-540-75520-3_53`.

30 D. Dor, S. Halperin, and U. Zwick. All-pairs almost shortest paths. *SIAM Journal on Computing*, 29(5):1740–1759, 2000.

31 Fedor V. Fomin, Daniel Lokshtanov, Michal Pilipczuk, Saket Saurabh, and Marcin Wrochna. Fully polynomial-time parameterized computations for graphs and matrices of low treewidth. In *Proceedings 28th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1419–1432, 2017. `doi:10.1137/1.9781611974782.92`.

32 Fedor V. Fomin, Fahad Panolan, M. S. Ramanujan, and Saket Saurabh. On integer programming and the path-width of the constraint matrix. *CoRR*, abs/1607.05342v2, 2016.

33 M. L. Fredman. How good is the information theory bound in sorting? *Theoretical Computer Science*, 1(4):355–361, 1976.

34 A. Freund. Improved subquadratic 3SUM. *Algorithmica*, pages 1–19, 2015. `doi:10.1007/s00453-015-0079-6`.

35 M. Fürer and B. Raghavachari. Approximating the minimum-degree Steiner tree to within one of optimal. *J. Algorithms*, 17(3):409–423, 1994. `doi:10.1006/jagm.1994.1042`.

36 H. Gabow, T. Nishizeki, O. Kariv, D. Leven, and O. Terada. Algorithms for edge-coloring graphs. Technical Report 41/85, Tohoku University, 1985.

37 Harold N. Gabow. A matroid approach to finding edge connectivity and packing arborescences. *Journal of Computer and System Sciences*, 50(2):259–273, 1995. `doi:10.1006/jcss.1995.1022`.

38 Harold N. Gabow. Using expander graphs to find vertex connectivity. *J. ACM*, 53(5):800–844, 2006. `doi:10.1145/1183907.1183912`.

39 Anka Gajentaan and Mark H. Overmars. On a class of $o(n^2)$ problems in computational geometry. *Comput. Geom.*, 45(4):140–152, 2012.

40 David Gale and Lloyd S. Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.

41 O. Gold and M. Sharir. Improved bounds for 3SUM, *K*-SUM, and linear degeneracy. *CoRR*, abs/1512.05279, 2015.

42 Isaac Goldstein, Tsvi Kopelowitz, Moshe Lewenstein, and Ely Porat. How hard is it to find (honest) witnesses? In *Proceedings 24th Annual European Symposium on Algorithms (ESA)*, pages 45:1–45:16, 2016.

43 R. E. Gomory and T. C. Hu. Multi-terminal network flows. *Journal of the Society for Industrial and Applied Mathematics*, 9, 1961.

44 A. Grønlund and S. Pettie. Threesomes, degenerates, and love triangles. In *Proceedings 55th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 621–630, 2014.

45 Joel Hass, J. C. Lagarias, and Nicholas Pippenger. The computational complexity of knot and link problems. *J. ACM*, 46(2):185–211, 1999. `doi:10.1145/301970.301971`.

46 M. Henzinger, S. Krinninger, D. Nanongkai, and T. Saranurak. Unifying and strengthening hardness for dynamic problems via the online matrix-vector multiplication conjecture. In *Proceedings 47th Annual ACM Symposium on Theory of Computing (STOC)*, pages 21–30, 2015.

47 Monika Henzinger, Satish Rao, and Di Wang. Local flow partitioning for faster edge connectivity. In *Proceedings 28th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1919–1938, 2017.

48 Ian Holyer. The NP-completeness of edge-coloring. *SIAM J. Comput.*, 10(4):718–720, 1981. `doi:10.1137/0210055`.

49 A. Itai and M. Rodeh. Finding a minimum circuit in a graph. *SIAM J. Comput.*, 7(4):413–423, 1978.

50 W. Jaco and J. L. Tollefson. Algorithms for the complete decomposition of a closed 3-manifold. *Ill. J. Math.*, 39:358–406, 1995.

51 M. Jurdziński and R. Lazić. Succinct progress measures for solving parity games. *CoRR*, abs/1702.05051, 2017.

52 Marcin Jurdziński. Deciding the winner in parity games is in UP ∩ co-UP. *Information Processing Letters*, 68(3):119–124, 1998. `doi:10.1016/S0020-0190(98)00150-1`.

53 Marcin Jurdziński, Mike Paterson, and Uri Zwick. A Deterministic Subexponential Algorithm for Solving Parity Games. *SIAM Journal on Computing*, 38(4):1519–1532, 2008. `doi:10.1137/070686652`.

54 Ken-ichi Kawarabayashi and Mikkel Thorup. Deterministic global minimum cut of a simple graph in near-linear time. In *Proceedings 47th ACM Symposium on Theory of Computing (STOC)*, pages 665–674, 2015. `doi:10.1145/2746539.2746588`.

55 T. Kopelowitz, S. Pettie, and E. Porat. Higher lower bounds from the 3SUM conjecture. In *Proceedings 27th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1272–1287, 2016. `doi:10.1137/1.9781611974331.ch89`.

56 Marvin Künnemann, Daniel Moeller, Ramamohan Paturi, and Stefan Schneider. Subquadratic algorithms for succinct stable matching. *CoRR*, abs/1510.06452, 2015.

57 Yin Tat Lee and Aaron Sidford. Efficient inverse maintenance and faster algorithms for linear programming. In *Proceedings 56th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 230–249, 2015. `doi:10.1109/FOCS.2015.23`.

58 Jiří Matoušek. Efficient partition trees. *Discrete & Computational Geometry*, 8(1):315–334, 1992.

59 Jakub Pachocki, Liam Roditty, Aaron Sidford, Roei Tov, and Virginia Vassilevska Williams. Approximating cycles in directed graphs: Fast algorithms for girth and roundtrip spanners. *CoRR*, abs/1611.00721, 2016. URL: http://arxiv.org/abs/1611.00721.

60 Debmalya Panigrahi. Gomory-Hu trees. In Ming-Yang Kao, editor, *Encyclopedia of Algorithms*, pages 858–861. Springer New York, 2016. `doi:10.1007/978-1-4939-2864-4_168`.

61 Christos H. Papadimitriou. On the complexity of integer programming. *J. ACM*, 28(4):765–768, 1981. `doi:10.1145/322276.322287`.

62 L. Roditty and V. Vassilevska Williams. Fast approximation algorithms for the diameter and radius of sparse graphs. In *Proceedings 45th ACM Symposium on Theory of Computing (STOC)*, pages 515–524, 2013.

63 Liam Roditty and Virginia Vassilevska Williams. Subquadratic time approximation algorithms for the girth. In *Proceedings of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 833–845, 2012.

64 Barna Saha. Language edit distance and maximum likelihood parsing of stochastic grammars: Faster algorithms and connection to fundamental graph problems. In *Proceedings 56th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 118–135, 2015. `doi:10.1109/FOCS.2015.17`.

65 M. Thorup. Worst-case update times for fully-dynamic all-pairs shortest paths. In *Proceedings 37th ACM Symposium on Theory of Computing (STOC)*, pages 112–119, 2005.

66 V. Vassilevska Williams and R. Williams. Subcubic equivalences between path, matrix and triangle problems. In *Proceedings 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 645–654, 2010.

67 R. Yuster and U. Zwick. Finding even cycles even faster. *SIAM J. Discrete Mathematics*, 10(2):209–222, 1997.

Participants

Amir Abboud
Stanford University, US

Alexandr Andoni
Columbia University –
New York, US

Arturs Backurs
MIT – Cambridge, US

Andreas Björklund
Lund University, SE

Karl Bringmann
MPI für Informatik –
Saarbrücken, DE

Maike Buchin
Ruhr-Universität Bochum, DE

Yi-Jun Chang
University of Michigan –
Ann Arbor, US

Raphaël Clifford
University of Bristol, GB

Radu Curticapean
Hungarian Academy of Sciences –
Budapest, HU

Marek Cygan
University of Warsaw, PL

Søren Dahlgaard
University of Copenhagen, DK

Holger Dell
Universität des Saarlandes, DE

Fedor V. Fomin
University of Bergen, NO

Omer Gold
Tel Aviv University, IL

Isaac Goldstein
Bar-Ilan University –
Ramat Gan, IL

Thore Husfeldt
IT University of
Copenhagen, DK

Mathias Bæk Tejs Knudsen
University of Copenhagen, DK

Young Kun Ko
Princeton University, US

Tsvi Kopelowitz
University of Michigan –
Ann Arbor, US

Robert Krauthgamer
Weizmann Institute –
Rehovot, IL

Sebastian Krinninger
MPI für Informatik –
Saarbrücken, DE

Marvin Künnemann
University of California –
San Diego, US

Moshe Lewenstein
Bar-Ilan University –
Ramat Gan, IL

Andrea Lincoln
Stanford University, US

Veronika Loitzenbauer
Universität Wien, AT

Aleksander Madry
MIT – Cambridge, US

Kurt Mehlhorn
MPI für Informatik –
Saarbrücken, DE

Ivan Mikhailin
University of California –
San Diego, US

Danupon Nanongkai
KTH Royal Institute of
Technology – Stockholm, SE

Ramamohan Paturi
University of California –
San Diego, US

Seth Pettie
University of Michigan –
Ann Arbor, US

Ely Porat
Bar-Ilan University –
Ramat Gan, IL

Vijaya Ramachandran
University of Texas – Austin, US

Liam Roditty
Bar-Ilan University –
Ramat Gan, IL

Aviad Rubinstein
University of California –
Berkeley, US

Thatchaphol Saranurak
KTH Royal Institute of
Technology – Stockholm, SE

Stefan Schneider
University of California –
San Diego, US

Arkadiusz Socala
University of Warsaw, PL

Virginia Vassilevska Williams
Stanford University, US

Joshua R. Wang
Stanford University, US

Oren Weimann
University of Haifa, IL

Philip Wellnitz
MPI für Informatik –
Saarbrücken, DE

Huacheng Yu
Stanford University, US

Beyond-Planar Graphs: Algorithmics and Combinatorics

Edited by

Seok-Hee Hong[1], Michael Kaufmann[2], Stephen G. Kobourov[3], and János Pach[4]

1   **The University of Sydney, AU, `seokhee.hong@sydney.edu.au`**
2   **Universität Tübingen, DE, `mk@informatik.uni-tuebingen.de`**
3   **University of Arizona – Tucson, US, `kobourov@cs.arizona.edu`**
4   **EPFL – Lausanne, CH, `janos.pach@epfl.ch`**

─── **Abstract** ───

This report summarizes Dagstuhl Seminar 16452 "Beyond-Planar Graphs: Algorithmics and Combinatorics" and documents the talks and discussions. The seminar brought together 29 researchers in the areas of graph theory, combinatorics, computational geometry, and graph drawing. The common interest was in the exploration of structural properties and the development of algorithms for so-called *beyond-planar graphs*, i.e., non-planar graphs with topological constraints such as specific types of crossings, or with some forbidden crossing patterns. The seminar began with three introductory talks by experts in the different fields. Abstracts of these talks are collected in this report. Next we discussed and grouped together open research problems about beyond planar graphs, such as their combinatorial structures (e.g, thickness, crossing number, coloring), their topology (e.g., string graph representation), their geometric representations (e.g., straight-line drawing, visibility representation, contact representation), and applications (e.g., algorithms for real-world network visualization). Four working groups were formed and a report from each group is included here.

1   Executive Summary

Seok-Hee Hong
Michael Kaufmann
Stephen G. Kobourov
János Pach

Relational data sets, containing a set of objects and relations between them, are commonly modeled by graphs/networks, with the objects as the vertices and the relations as the edges. A great deal is known about the structure and properties of special types of graphs, in particular *planar graphs*. The class of planar graphs is fundamental for both Graph Theory and Graph Algorithms, and extensively studied. Many structural properties of planar graphs

Beyond-Planar Graphs: Algorithmics and Combinatorics, *Dagstuhl Reports*, Vol. 6, Issue 11, pp. 35–62
Editors: Seok-Hee Hong, Michael Kaufmann, Stephen G. Kobourov, and János Pach
        Dagstuhl Reports
        Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

are known and these properties can be used in the development of efficient algorithms for planar graphs, even where the more general problem is NP-hard.

Most real world graphs, however, are *non-planar*. In particular, many scale-free networks, which can be used to model web-graphs, social networks and biological networks, consists of sparse non-planar graphs. To analyze and visualize such real-world networks, we need to solve fundamental mathematical and algorithmic research questions on *sparse non-planar* graphs, which we call *beyond-planar graphs*. The notion of beyond-planar graphs has been established as non-planar graphs with topological constraints such as specific types of crossings or with some forbidden crossing patterns, although it has not been formally defined. Examples of beyond-planar graphs include:

- *k-planar* graphs: graphs which can be embedded with at most k crossings per edge.
- *k-quasi-planar* graphs: graphs which can be embedded without k mutually crossing edges.
- *bar k-visibility* graphs: graphs whose vertices are represented as horizontal segments (bars) and edges as vertical lines connecting bars, intersecting at most k other bars.
- *fan-crossing-free* graphs: graphs which can be embedded without fan-crossings.
- *fan-planar* graphs: graphs which can be embedded with crossings sharing the common vertices.
- *RAC (Right Angle Crossing)* graphs: a graph which has a straight-line drawing with right angle crossings.

The aim of the seminar was to bring together world-renowned researchers in graph algorithms, computational geometry and graph theory, and collaboratively develop a research agenda for the study of beyond-planar graphs. The plan was to work on specific open problems about the structure, topology, and geometry of beyond-planar graphs. One of the outcomes of the workshop might be an annotated bibliography of this new field of study.

On Sunday afternoon, 29 participants met at Dagstuhl for an informal get-together. Fortunately, there were no cancelations and everybody who registered was able to attend. On Monday morning, the workshop officially kicked off. After a round of introductions, where we discovered that eight participants were first-time Dagstuhl attendees, we enjoyed three overview talks about beyond-planar graphs from three different points of view. First, Géza Tóth from the Rényi Institute in Budapest talked about the combinatorics of beyond-planar graphs in connection to graph theory. Next, Giuseppe Liotta from the University of Perugia gave an overview about the connections between graph drawing and beyond-planar graphs and presented a taxonomy of related topics and questions. Finally, Alexander Wolff from the University of Würzburg discussed beyond-planar graphs in the context of geometry and geometric graph representations.

On Monday afternoon, we had lively open problem sessions, where we collected 20 problems covering the most relevant topics. The participants split into four groups based on common interest in subsets of the open problems. The last three days of the seminar were dedicated to working group efforts. Most of the groups kept their focus on the original problems as stated in the open problem session, while one group modified and expanded the problems; see Section 4. We had two progress reports sessions, including one on Friday morning, where group leaders were officially designated and plans for follow-up work were made. Work from one of the groups has been submitted to an international conference, and we expect further research publications to result directly from the seminar.

Arguably the best, and most-appreciated, feature of the seminar was the opportunity to engage in discussion and interactions with experts in various fields with shared passion about graphs, geometry and combinatorics. We received very positive feedback from the participants (e.g., scientific quality: 10.5/11, inspired new ideas: 23/25, inspired joint projects:

21/25) and it is our impression that the participants enjoyed the unique scientific atmosphere at the seminar and benefited from the scientific program. In summary, we regard the seminar as a success, and we are grateful for having had the opportunity to organize it and take this opportunity to thank the scientific, administrative, and technical staff at Schloss Dagstuhl.

2   Table of Contents

3 Overview of Talks

3.1 Graph Drawing Beyond Planarity

Giuseppe Liotta (University of Perugia, IT)

It is well known that drawings of graphs with many edge crossings are hard to read. On the other hand, edge crossings are simply unavoidable when the graph to be drawn is not planar. As a consequence a number of relaxations of the notion of graph planarity relaxation have been proposed in the literature. They allow edge crossings but forbid specific configurations which would affect the readability of the graph representation. For each such relaxation, different research questions can be asked having both algorithmic and combinatorial nature. Aim of the invited talk "Graph Drawing Beyond Planarity" was to briefly survey this rapidly growing research area by pointing out some of its most investigated questions and some of the most prominent open questions.

Graph Drawing Beyond Planarity

The classical literature on graph drawing showcases elegant algorithms and sophisticated data structures under the assumption that the input relational data set can be displayed as a network where no two edges cross (see, e.g., [5, 10, 11, 14]), i.e. the input is a planar graph. Unfortunately, almost every graph is non-planar in practice and the question on how to simplify the visual analysis of non-planar networks has become a central topic in the graph drawing research agenda.

We recall that planar graphs can be expressed in terms of forbidden subgraphs: A graph G is planar if and only if it does not contain a subdivision of K_5 or $K_{3,3}$. Then, a fundamental natural step towards understanding non-planar graphs is to consider realizations where some types of crossings are forbidden while some other types are allowed. For example, we recall a sequence of HCI experiments by Huang et al. [7, 8, 9] proving that crossing edges significantly affect human understanding if they form acute angle, while crossing that form angles from about $\frac{\pi}{3}$ to $\frac{\pi}{2}$ guarantee good readability properties. Hence it makes sense to explore complexity issues related to drawings of graphs where such "sharp angle crossings" are forbidden. In addition to these results, Purchase et al. [16, 17, 18]) prove that an edge is difficult to read if it is crossed by many other edges; hence, the current research agenda on graph drawing beyond planarity includes the study of representations where every edge is crossed by at most k other edges, for a given constant k.

Examples of "beyond planar" graphs and problems

A *drawing* of a graph G:
 (i) injectively maps each vertex u of G to a geometric body p_u in the plane;
 (ii) maps each edge (u, v) of G to a Jordan arc connecting p_u and p_v that does not pass through any other vertex;
(iii) is such that any two edges have at most one point in common.
A drawing of a graph is a *straight-line drawing* if every edge is a straight-line segment, it is a *poly-line drawing* if the edges are polygonal chains and may contain bends. Vertices are typically mapped to points in the plane, but in some cases they can be other geometric objects; for example, in a *rectangle visibility representation* each vertex is represented as

Forbidden configuration	Algorithmic question
no three mutually crossing edges	Complexity of the recognition problem
no crossing angle smaller than α	Straight–line drawability for graphs with bounded vertex degree
no fan crossing	Compute simultaneous emebddings of graph pairs
no edge with k crossings	Compute compact drawings of planar graphs with lmited number of crossings per edge
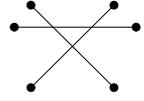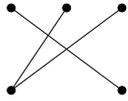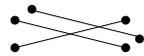
■ **Figure 1** A table with some forbidden crossing configurations and related computational questions.

a rectangle and each edge corresponds to a horizontal or vertical line of sight between its end-vertices. Rectangle visibility representations guarantees that edge crossings form $\frac{\pi}{2}$ angles and also that the edges are straight-line segments.

The "beyond planarity" research area could be briefly described as the (potentially uncountable) collection of problems of the type depicted in Figure 1, where the column "Forbidden" describes a forbidden crossing configuration in the drawing of a graph and the column "Question" describes a corresponding computational question of interest in graph drawing. We remark that both the forbidden configurations and the computational questions of Figure 1 are mere examples within a much larger research framework. The interested reader is referred, for example, to recent proceedings of the International Symposium on Graph Drawing [19] for more results and more open problems on the "beyond planarity" topic. (See also http://www.graphdrawing.org/symposia.html.)

For reasons of space we shall make just one concrete example in the next section about how the research on a specific problem has evolved.

A Research Stream Example: Edge Partitions of 1-planar Graphs

A *1-planar drawing* is a drawing where each edge is crossed at most once. A graph is *1-planar* if it admits a 1-planar drawing. A *1-planar embedding* is an embedding that represents an equivalence class of 1-planar drawings. A *1-plane graph* is a graph with a fixed 1-planar embedding. A 1-planar graph G with n vertices has at most $4n - 8$ edges [3, 15], which is a tight bound; namely those 1-planar graphs having n vertices and $4n - 8$ edges are called *optimal 1-planar graphs*.

An *edge partition* of a 1-planar graph G is an edge coloring of G with two colors, say *red* and *blue*, such that both the graph formed by the red edges, called the *red graph*, and the graph formed by the blue edges, called the *blue graph*, are planar. Note that, given a 1-planar embedding of G, an edge partition of G can be constructed by coloring red an edge for each pair of crossing edges, and by coloring blue the remaining edges. Czap and Hudák [4] proved that every optimal 1-planar graph admits an edge partition such that the red graph is a forest. This result has been later extended to all 1-planar graphs by Ackerman [1].

Motivated by visibility representations of 1-planar graphs (see, e.g., [2, 13, 13]), Lenhart et al. [12] and Di Giacomo et al. [6] studied edge partitions such that the red graph has maximum vertex degree that is bounded by a constant independent of the size of the graph. Namely, Lenhart et al. [12] proved that if G is an n-vertex optimal 1-plane graph with a an edge partition with the red graph G_R being a forest, then G_R has n vertices and it is composed of two trees. Based on this finding, they proved that for any constant c, there exists an optimal 1-plane graph such that in any edge partition with the red graph G_R being a forest, the maximum vertex degree of G_R is at least c. On the positive side, if we drop the acyclicity requirement, then every optimal 1-planar graph admits an edge partition such that the red graph has maximum vertex degree at most four, and degree four is sometimes needed [12]. Also, every 3-connected 1-planar graph admits an edge partition such that the red graph has maximum vertex degree at most six, and degree six is sometimes needed, as shown by Di Giacomo et al. [6]. Finally, for every $n > 0$ there exists an $O(n)$-vertex 2-connected 1-planar graph such that in any edge partition the red graph has maximum vertex degree $\Omega(n)$ [6].

References

1 Eyal Ackerman. A note on 1-planar graphs. *Discrete Applied Mathematics*, 175:104–108, 2014. `doi:10.1016/j.dam.2014.05.025`.

2 Therese C. Biedl, Giuseppe Liotta, and Fabrizio Montecchiani. On visibility representations of non-planar graphs. In Sándor P. Fekete and Anna Lubiw, editors, *32nd International Symposium on Computational Geometry, SoCG 2016, June 14-18, 2016, Boston, MA, USA*, volume 51 of *LIPIcs*, pages 19:1–19:16. Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2016. `doi:10.4230/LIPIcs.SoCG.2016.19`.

3 Rainer Bodendiek, Heinz Schumacher, and Klaus Wagner. Bemerkungen zu einem Sechsfarbenproblem von G. Ringel. *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, 53(1):41–52, 1983.

4 Július Czap and Dávid Hudák. On drawings and decompositions of 1-planar graphs. *Electr. J. Comb.*, 20(2):P54, 2013.

5 Giuseppe Di Battista, Peter Eades, Roberto Tamassia, and Ioannis G. Tollis. *Graph Drawing*. Prentice Hall, Upper Saddle River, NJ, 1999.

6 Emilio Di Giacomo, Walter Didimo, William S. Evans, Giuseppe Liotta, Henk Meijer, Fabrizio Montecchiani, and Stephen K. Wismath. Ortho-polygon visibility representations of embedded graphs. In *Graph Drawing and Network Visualization – 24th International Symposium, GD 2016, Athens, Greece, September 19-21, 2016, Revised Selected Papers*,
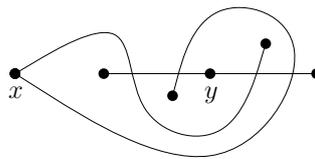
Figure 2 Edge $\{x, y\}$ cannot be added.

volume 9801 of *Lecture Notes in Computer Science*, pages 280–294. Springer, 2016. `doi: 10.1007/978-3-319-50106-2`.

7 Weidong Huang. Using eye tracking to investigate graph layout effects. In *APVIS*, pages 97–100, 2007.

8 Weidong Huang, Peter Eades, and Seok-Hee Hong. Larger crossing angles make graphs easier to read. *J. Vis. Lang. Comput.*, 25(4):452–465, 2014.

9 Weidong Huang, Seok-Hee Hong, and Peter Eades. Effects of crossing angles. In *PacificVis*, pages 41–46, 2008.

10 Michael Jünger and Petra Mutzel, editors. *Graph Drawing Software*. Springer, 2003.

11 Michael Kaufmann and Dorothea Wagner, editors. *Drawing Graphs*. Springer Verlag, 2001.

12 William J. Lenhart, Giuseppe Liotta, and Fabrizio Montecchiani. On partitioning the edges of 1-planar graphs. *Theoretical Computer Science*, to appear, 2017.

13 Giuseppe Liotta and Fabrizio Montecchiani. L-visibility drawings of ic-planar graphs. *Inf. Process. Lett.*, 116(3):217–222, 2016. `doi:10.1016/j.ipl.2015.11.011`.

14 Takao Nishizeki and Md.Saidur Rahman. *Planar Graph Drawing*. World Scientific, 2004.

15 János Pach and Géza Tóth. Graphs drawn with few crossings per edge. *Combinatorica*, 17(3):427–439, 1997.

16 Helen C. Purchase. Effective information visualisation: a study of graph drawing aesthetics and algorithms. *Interacting with Computers*, 13(2):147–162, 2000.

17 Helen C. Purchase, David A. Carrington, and Jo-Anne Allder. Empirical evaluation of aesthetics-based graph layout. *Empirical Software Engineering*, 7(3):233–255, 2002.

18 Colin Ware, Helen C. Purchase, Linda Colpoys, and Matthew McGill. Cognitive measurements of graph aesthetics. *Information Visualization*, 1(2):103–110, 2002.

19 Stephen K. Wismath and Alexander Wolff, editors. *Graph Drawing – 21st International Symposium, GD 2013, Bordeaux, France, September 23-25, 2013, Revised Selected Papers*, volume 8242 of *Lecture Notes in Computer Science*. Springer, 2013.

3.2 Saturated Topological Graphs

Géza Tóth (Alfréd Rényi Institute of Mathematics – Budapest, HU)

A *simple topological graph G* is a graph drawn in the plane so that any pair of edges have at most one point in common, which is either an endpoint or a proper crossing. *G* is called *saturated* if no further edge can be added so that it remains a simple topological graph. Obviously, if *G* is a *complete* simple topological graph, then it is saturated.

The simple topological graph G_1 on Figure 2, found by Kynčl, [7], has six vertices and if we connect x and y by any curve as an edge, two edges with a common endpoint will cross each other. So the resulting topological graph is not simple anymore. All other egdes can be added, so we obtain a saturated simple topological graph of 6 vertices and 14 edges. From
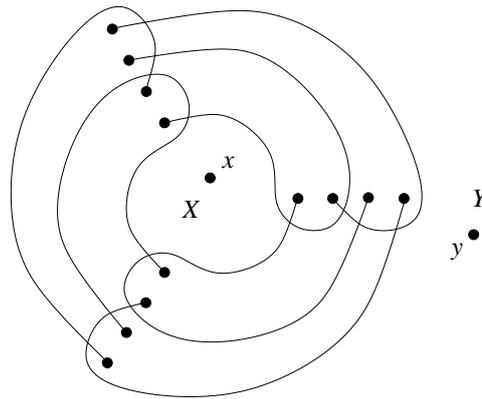
Figure 3 Edge $\{x, y\}$ cannot be added.

this we can construct a saturated simple topological graph of n vertices and $\binom{n}{2} - \lfloor n/6 \rfloor$ edges.

It is a natural question to ask, whether every saturated simple topological graph with n vertices must have $\Omega(n^2)$ edges. It turned out, that there are examples with only a linear number of edges.

▶ **Theorem 1** (Kynčl, Pach, Radoičić, Tóth, [6]). *For any $n \geq 4$, let $s(n)$ be the minimum number of edges that a saturated simple topological graph on n vertices can have. Then*

$$1.5n \leq s_1(n) \leq 17.5n.$$

The upper bound construction is an iterated version of the topological graph G_2 on Figure 3. It is a simple topological graph, but if we connect vertex x in region X, and vertex y in Y by a curve, it will cross one of the edges of G_2 at least twice.

For the lower bound, it is proved that in a saturated simple topological graph each vertex has degree at least three. Therefore, the number of edges is at least $1.5n$.

The upper bound has been improved recently by Hajnal, Igamberdiev, Rote, and Schulz [5]. For the lower bound, a natural way to improve it is to show that in a saturated simple topological graph each vertex has degree at least four, or five, or even more. In [5] it is also shown, that we can not expect too much improvement from this simple approach, there could be a vertex of degree four, or many vertices of degree five.

▶ **Theorem 2** (Hajnal, Igamberdiev, Rote, and Schulz [5]).
 (i) $s(n) \leq 7n$.
 (ii) *For every $n \geq 6$ there is a saturated simple topological graph on n vertices with a vertex of degree 4.*
(iii) *For every $m \geq 1$ there is a saturated simple topological graph on $10m$ vertices with m vertices of degree 5.*

▶ **Problem 3.** *Is there a saturated simple topological graph with a vertex of degree three?*

▶ **Problem 4.** *Construct a saturated simple topological graph with many vertices of degree four.*

▶ **Problem 5.** *Improve the bounds for $s(n)$.*

In general, for any positive integer k, a topological graph is called k-*simple* if any two edges have at most k points in common. We also assume that in a k-simple topological graph

■ **Table 1** Upper bounds on the minimum number of edges in saturated k-simple topological graphs.

k	1	2	3	4	5	6	7	8	9	10	≥ 11
upper bound [6]	$17.5n$	$16n$	$14.5n$	$13.5n$	$13n$	$9.5n$	$10n$	$9.5n$	$7n$	$9.5n$	$7n$
upper bound [5]	$7n$	$14.5n$									

no edge crosses itself. A 1-simple topological graph is exactly a simple topological graph. It is not obvious at all how to construct *non-complete* saturated k-simple topological graphs for $k > 1$.

▶ **Theorem 6** (Kynčl, Pach, Radoičić, Tóth, [6]). *For any positive integers k and $n \geq 4$, let $s_k(n)$ be the minimum number of edges that a saturated k-simple topological graph on n vertices can have. Then for $k > 1$ we have*

$$n \leq s_k(n) \leq 16n.$$

For $k = 2$, the upper bound was improved by Hajnal, Igamberdiev, Rote, and Schulz [5].

▶ **Theorem 7** (Hajnal, Igamberdiev, Rote, and Schulz [5]). $s_2(n) \leq 14.5n$.

For the best upper bounds see Table 1.

In a graph G, an *isolated triangle* is a triangle (K_3) which is not connected to any other vertices. In the proof of the lower bound $s(n) \geq 1.5n$, an essential step is that we prove that there is no isolated triangle in a saturated simple topological graph. The proof does not work for saturated k-simple topological graphs, for $k > 1$, therefore, in this case we can prove only that every vertex has degree at least 2, which implies $s_k(n) \geq n$.

▶ **Problem 8.** *For $k > 1$, can a saturated k-simple topological graph contain an isolated triangle?*

But unlike in the case of simple topological graphs, even if we knew that a saturated k-simple topological graph can not contain an isolated triangle, we still can not prove that in a saturated k-simple topological graph all vertices have degree at least 3.

▶ **Problem 9.** *Is there a saturated k-simple topological graph for some $k > 1$ with a vertex of degree two?*

Probably the most natural and exciting problem in this topic is the following.

▶ **Problem 10.** *Is it true that every saturated k-simple topological graph is connected?*

The answer might depend on the value of k, and we do not know the answer for any k.

We assumed that in a k-simple topological graph, no edge can cross itself. For any k, a graph drawn in the plane is called a *k-complicated topological graph* if any two edges have at most k points in common, and an edge is allowed to cross itself, at most k times. Somewhat surprisingly, for saturated k-complicated topological graphs we cannot even prove that every vertex has degree at least two! We can only prove that a saturated k-complicated topological graph does not have isolated vertices. Therefore, the best lower bound we have for the minimum number of edges of a saturated k-complicated topological graph is $c_k(n) \geq n/2$.

▶ **Problem 11.** *Is there a saturated k-complicated topological graph with a vertex of degree one, for every $k \geq 1$?*

Now we study a slightly different problem. It is an easy consequence of Euler's Formula, that every planar graph of n vertices has at most $3n - 6$ edges. If it has exactly $3n - 6$ edges, then it is a triangulation. If it has less edges and it is drawn in the plane without crossings, than we can extend it to a triangulation.

A topological graph is *1-plane*, if each edge is crossed at most once. A graph is *1-planar*, if it has a 1-plane drawing. It is known that the maximum number of edges of a 1-plane or 1-planar graph is $4n - 8$. Brandenburg et al. [3] and independently Eades et al. [4] observed a very interesting phenomenon. They noticed that maximal 1-plane or maximal 1-planar graphs can have much fewer edges.

▶ **Theorem 12** (Brandenburg, Eppstein, Gleissner, Goodrich, Hanauer, Reislhuber [3]). *Let $e_1(n)$ (resp. $e'_1(n)$) denote the minimum number of edges of a maximal 1-plane (resp. 1-planar) graph of n vertices. Then we have*

$$2.1n \le e_1(n) \le 2.33n, \quad 2.15n \le e'_1(n) \le 2.64n.$$

Both lower bounds were recently improved to $2.22n$ [2].

▶ **Problem 13.** *Improve the bounds for $e_1(n)$ and $e'_1(n)$.*

For any n, $e_1(n) \le e'_1(n)$ since any maximal 1-planar graph has a maximal 1-plane drawing. Now the best known lower bounds are the same.

▶ **Problem 14.** *Is it true that for every n $e_1(n) = e'_1(n)$?*

In general, for every $k \ge 1$, a topological graph is *k-plane*, if each edge is crossed at most k times. A graph is *k-planar*, if it has a k-plane drawing. Let $e_k(n)$ (resp. $e'_k(n)$) denote the minimum number of edges of a maximal k-plane (resp. k-planar) graph of n vertices.

Auer et al. [1] proved that $e_2(n) \le 1.33n$ and $e'_2(n) \le 2.63n$. It is not hard to see, that $e_k(n) \le cn/k$ for some $c > 0$.

▶ **Problem 15.** *Establish some nontrivial bounds for $e_k(n)$ and $e'_k(n)$.*

References

1 Christopher Auer, Franz J. Brandenburg, Andreas Gleissner, Kathrin Hanauer: On Sparse Maximal 2-Planar Graphs, *Graph Drawing 2012*, Lecture Notes in Computer Science 7704 (2013), 555–556.
2 János Barát, Géza Tóth: Improvements on the density of maximal 1-planar graphs, submitted.
3 Franz J. Brandenburg, David Eppstein, Andreas Gleissner, Michael T. Goodrich, Kathrin Hanauer, Josef Reislhuber: On the Density of Maximal 1-Planar Graphs, *Graph Drawing 2012*, Lecture Notes in Computer Science 7704 (2013), 327–338.
4 Peter Eades, Seok-Hee Hong, Naoki Katoh, Giuseppe Liotta, Pascal Schweitzer, Yusuke Suzuki: Testing Maximal 1-Planarity of Graphs with a Rotation System in Linear Time, *Graph Drawing 2012*, Lecture Notes in Computer Science 7704 (2013), 339–345.
5 Péter Hajnal, Alexander Igamberdiev, Günter Rote, André Schulz: Saturated simple and 2-simple topological graphs with few edges, *41st International Workshop on Graph-Theoretic Concepts in Computer Science, WG 2015*, Lecture Notes in Computer Science 9224 (2016), 391–405.
6 Jan Kynčl, János Pach, Rados Radoičić, Géza Tóth: Saturated simple and k-simple topological graphs, *Computational Geometry: Theory and Applications* 48 (2015), 295–310.
7 Jan Kynčl, Improved enumeration of simple topological graphs, *Discrete Comput. Geom.* 50 (2013), 727–770.

3.3 Drawing Graphs: Geometric Aspects Beyond Planarity

Alexander Wolff (Universität Würzburg, DE)

Long before Graph Drawing has been established as scientific field, researchers have been studying ways to draw *planar* graphs. A particularly intriguing question is whether any planar graph can be drawn straight-line, that is, by mapping the vertices to points and the edges to non-crossing line segments between their endpoints. This question was answered in the affirmative, independently by Wagner [25], Fáry [15], and Stein [22]. Koebe [18] showed an even stronger result by proving that any planar graph is a *coin graph*, that is, the vertices can be mapped to pairwise interior-disjoint disks such that two disks touch if and only if the corresponding vertices are adjacent. In a beautiful paper entitled "How to draw a graph", Tutte [23] gave a first constructive, efficient algorithm for drawing any (triconnected) planar graph with straight-line edges (and convex faces). It is easy, however, to give examples (such as the nested-triangles graph) where Tutte's algorithm produces a drawing with an exponential ratio between the lengths of the longest and the shortest edge. This led to the question whether planar graphs can always been drawn straight-line on a grid of polynomial size. Again, this was answered in the affirmative; independently by de Fraysseix et al. [5] and by Schnyder [20].

Few graphs, however, are actually planar. Therefore, researchers in graph drawing and related areas have recently become very interested in studying classes of close-to-planar graphs. For example, Huang et al. [16] did a user study that showed that the crossing angle has a strong influence on the readability of a graph drawing. Based on the results of this user study, Didimo et al. [7] introduced the class of *RAC graphs*, that is, graphs that can be drawn straight-line into the plane such that all crossings are at right angles.

In this abstract, I consider geometric aspects of the recent "beyond planarity" direction in graph drawing. I focus on drawing graphs with low *visual complexity*. By visual complexity I mean the number of geometric primitives needed to represent the graph; for example, slopes, line segments and circular arcs, or lines and, in 3-space, planes (see following sections).

Other obvious geometric aspects "beyond planarity" are ways of dealing with crossings; namely by making crossings nicer (such as insisting on large crossing angles [7] or using edge casing [14]) or by eliminating crossings (such as *confluent drawings* where edges are represented by the existence of locally-monotone curves between their endpoints [6] or *partial edge drawings* where only fractions of each edge are drawn [2]). The layout of graphs can also be subject to other geometric restrictions (such as tracks for the vertices [9]). In this short overview, I will not deal with any of these aspects nor with topological graphs, nor with graph representations beyond dot-link diagrams (such as intersection, contact, or visibility representations, map graphs etc.).

Slope Number

Why are metro maps often drawn with a restricted set of slopes, for example, orthogonal, hexagonal, or, most commonly, octilinear? Arguably, because such drawings have low *visual complexity*, that is, they appear simpler and clearer than drawings where the number of slopes is not restricted; especially if slopes are chosen such that the *angular resolution* – the smallest angle between two edges incident to the same vertex – is large. For example, in an *octilinear* drawing (where edge directions are multiples of 45°), the angular resolution is at least 45°.

Motivated by similar observations, Wade and Chu [24] introduced, for a given graph G, its *slope number*, slope(G), as the minimum number of distinct slopes in a straight-line drawing of G. Among others, they showed that, for $n \geq 3$, slope(K_n) = n. Later, Dujmović et al. [10] identified two simple lower bounds, namely slope(G) $\geq \Delta(G)/2$ and slope(G) $\geq \delta(G)$, where $\Delta(G)$ is the maximum degree and $\delta(G)$ is the minumum degree of G. They asked whether there is some universal function f such that, for any graph G, its slope number can be upperbounded by $f(\Delta(G))$, independently of the size of G. Their question was answered to the negative by Pach and Pálvögyi [19] and, independently, by Barát et al. [1]. Pach and Pálvögyi [19] showed that for any sufficiently large integer n and $\Delta \geq 5$, there is an n-vertex graph G of maximum degree Δ whose slope number is larger than $n^{1/2-O(1/\Delta)}$. This bound was later improved to $n^{1-O(1/\Delta)}$ by Dujmović et al. [11]. They also proved positive results for restricted graph classes; namely for interval, co-comparability, and AT-free graphs. Dujmović et al. also showed that, if every edge can have a bend, $\Delta(G)+1$ slopes suffice for any graph G. To name another positive result, Jelínek et al. [17] showed that, for any planar partial 3-tree, $(\Delta(G))^5$ slopes suffice.

Segment Number and Arc Number

Another way to keep the visual complexity of a graph drawing low is to use few line segments. This idea is captured by the *segment number* of a graph, that is, the smallest number of line segments that together constitute a straight-line drawing of the given graph. The *arc number* of a graph is defined analogously with respect to circular arcs. For a graph G, we denote its segment number by seg(G) and its arc number by arc(G). So far, both numbers have only been studied for planar graphs. Again, two obvious lower bounds for seg(G) are known [8]; the slope number of G and $\eta(G)/2$, that is, the number of odd-degree vertices of G over 2. Dujmović et al. [8], who introduced the segment number, showed that trees can be drawn such that optimum segment number and slope number are achieved simultaneously. In other words, any tree T admits a drawing with seg(T) = $\eta(T)/2$ and slope(T) = $\Delta(T)/2$. Unfortunately, these drawings need exponential area. Therefore, Schulz [21] suggested to study the arc number of planar graphs. He showed that any tree with m edges can be drawn on a polynomial-size grid ($O(n^{1.81}) \times n$) using at most $3m/4$ arcs. Upper bounds for segment number and arc number (as fractions of the number of edges, ignoring small additive terms) are known for series-parallel graphs (3/4 vs. 1/2), planar 3-trees (2/3 vs. 11/18), and triconnected planar graphs (5/6 vs. 2/3) [8, 21]. The upper bound on the segment number for triconnected planar graphs has been improved for the special cases of triangulations and 4-connected triangulations (from 5/6 to 7/9 and 3/4, respectively) by Durocher and Mondal [12]. Durocher et al. [13] showed that the segment number is NP-hard to compute, even in the special case of arrangement graphs.

Line Cover Number and Plane Cover Number

The *affine cover number* $\rho_\ell^d(G)$ of a graph G is a generalization of planarity. It asks how many ℓ-dimensional planes are needed to cover a straight-line, crossing-free drawing of G in d-dimensional space. Clearly, any graph can be drawn without crossings in 3-space, so only the cases $\ell \in \{1,2\}$ are interesting. As it turns out, host spaces of dimension greater than three don't help to reduce the affine cover number. Hence, only three combinations of ℓ and d are worth studying for a given graph G: the plane cover number $\rho_2^3(G)$, the line cover number $\rho_1^3(G)$ in 3-space, and, for any planar graph G, the line cover number $\rho_1^2(G)$ in the plane. Chaplick et al. [4] introduced the affine cover number and related it to many

known graph parameters. The also study the *weak affine cover number* where only the vertices (but not the edges) of the given graph need to be covered. Concerning computational complexity, Chaplick et al. [3] showed that deciding, for a given graph G and integer k, whether $\rho_3^2(G) \leq k$, $\rho_3^1(G) \leq k$, or $\rho_2^1(G) \leq k$ is (at least) NP-hard. On the positive side, they showed that the two versions of the line cover number are fixed-parameter tractable. For the plane cover number, however, the decision problem is NP-hard even for any *fixed k*; hence, this problem is *not* fixed-parameter tractable.

References

1 János Barát, Jiří Matoušek, and David R. Wood. Bounded-degree graphs have arbitrarily large geometric thickness. *Electr. J. Combin.*, 13:#R3, 2006.

2 Till Bruckdorfer and Michael Kaufmann. Mad at edge crossings? break the edges! In Evangelos Kranakis, Danny Krizanc, and Flaminia Luccio, editors, *Proc. 6th Int. Conf. Fun with Algorithms (FUN'12)*, volume 7288 of *Lect. Notes Comput. Sci.*, pages 40–50. Springer-Verlag, 2012.

3 Steven Chaplick, Krzysztof Fleszar, Fabian Lipp, Alexander Ravsky, Oleg Verbitsky, and Alexander Wolff. The complexity of drawing graphs on few lines and few planes. Arxiv report, June 2016. Available at http://arxiv.org/abs/1607.06444.

4 Steven Chaplick, Krzysztof Fleszar, Fabian Lipp, Alexander Ravsky, Oleg Verbitsky, and Alexander Wolff. Drawing graphs on few lines and few planes. In Yifan Hu and Martin Nöllenburg, editors, *Proc. 24th Int. Symp. Graph Drawing & Network Vis. (GD'16)*, volume 9801 of *Lect. Notes Comput. Sci.*, pages 166–180. Springer-Verlag, 2016.

5 Hubert de Fraysseix, János Pach, and Richard Pollack. How to draw a planar graph on a grid. *Combinatorica*, 10(1):41–51, 1990.

6 Matthew Dickerson, David Eppstein, Michael T. Goodrich, and Jeremy Yu Meng. Confluent drawings: Visualizing non-planar diagrams in a planar way. *J. Graph Algorithms Appl.*, 9(1):31–52, 2005.

7 Walter Didimo, Peter Eades, and Giuseppe Liotta. Drawing graphs with right angle crossings. *Theoret. Comput. Sci.*, 412(39):5156–5166, 2011.

8 Vida Dujmović, David Eppstein, Matthew Suderman, and David R. Wood. Drawings of planar graphs with few slopes and segments. *Comput. Geom. Theory Appl.*, 38(3):194–212, 2007.

9 Vida Dujmović, Attila Pór, and David R. Wood. Track layouts of graphs. *Discrete Math. & Theoret. Comput. Sci.*, 6(2):497–522, 2004.

10 Vida Dujmović, Matthew Suderman, and David R. Wood. Really straight graph drawings. In János Pach, editor, *Proc. 12th Int. Symp. Graph Drawing (GD'04)*, volume 3383 of *Lect. Notes Comput. Sci.*, pages 122–132. Springer-Verlag, 2005.

11 Vida Dujmović, Matthew Suderman, and David R. Wood. Graph drawings with few slopes. *Comput. Geom. Theory Appl.*, 38(3):181–193, 2007.

12 Stephane Durocher and Debajyoti Mondal. Drawing plane triangulations with few segments. In *Proc. Canad. Conf. Comput. Geom. (CCCG'14)*, pages 40–45, 2014.

13 Stephane Durocher, Debajyoti Mondal, Rahnuma Islam Nishat, and Sue Whitesides. A note on minimum-segment drawings of planar graphs. *J. Graph Algorithms Appl.*, 17:301–328, 2013.

14 David Eppstein, Marc van Kreveld, Elena Mumford, and Bettina Speckmann. Edges and switches, tunnels and bridges. *Comput. Geom. Theory Appl.*, 42(8):790–802, 2009. Special Issue on the 23rd European Workshop on Computational Geometry.

15 István Fáry. On straight-line representation of planar graphs. *Acta Sci. Math. (Szeged)*, 11:229–233, 1948.

16 Weidong Huang, Seok-Hee Hong, and Peter Eades. Effects of crossing angles. In *Proc. IEEE Pacific Visualization Symp. (PacificVis'08)*, pages 41–46, 2008.

17 Vít Jelínek, Eva Jelínková, Jan Kratochvíl, Bernard Lidický, Marek Tesař, and Tomáš Vyskočil. The planar slope number of planar partial 3-trees of bounded degree. In David Eppstein and Emden R. Gansner, editors, *Proc. 17th Int. Symp. Graph Drawing (GD'09)*, pages 304–315. Springer-Verlag, 2010.

18 Paul Koebe. Kontaktprobleme der konformen Abbildung. *Ber. Sächs. Akad. Wiss. Leipzig, Math.-Phys. Klasse*, 88:141–164, 1936.

19 János Pach and Dömötör Pálvölgyi. Bounded-degree graphs can have arbitrarily large slope numbers. *Electr. J. Combin.*, 13:#N1, 4 pages, 2006.

20 Walter Schnyder. Embedding planar graphs on the grid. In *Proc. 1st ACM-SIAM Symp. Discrete Algorithms (SODA'90)*, pages 138–148, 1990.

21 André Schulz. Drawing graphs with few arcs. *J. Graph Algorithms Appl.*, 19(1):393–412, 2015.

22 S. K. Stein. Convex maps. *Proc. Amer. Math. Soc.*, 2:464–466, 1951.

23 William T. Tutte. How to draw a graph. *Proc. London Math. Soc.*, 13(52):743–768, 1963.

24 Greg A. Wade and Jiang-Hsing Chu. Drawability of complete graphs using a minimal slope set. *Comput. J.*, 37:139–142, 1994.

25 Klaus Wagner. Bemerkungen zum Vierfarbenproblem. *Jahresbericht Deutsch. Math.-Verein.*, 46:26–32, 1936.

4   Working groups

4.1   Working group A: Generalization of the Crossing Lemma

Michael Kaufmann (Universität Tübingen, DE), János Pach (EPFL – Lausanne, CH), Vincenzo Roselli (Roma Tre University, IT), Géza Tóth (Alfréd Rényi Institute of Mathematics – Budapest, HU), Torsten Ueckerdt (KIT – Karlsruher Institut für Technologie, DE), and Pavel Valtr (Charles University – Prague, CZ)

The crossing number is a well-studied quantification on how far a given graph is from being planar. The Crossing Lemma yields a lower bound on the crossing number. The lemma has found many important applications to problems in incidence geometry, to sum-product estimates and other Erdős problems. The lemma was discovered by Ajtai, Chvátal, Newborn, Szemerédi [2], and, independently, by Leighton [1]. Twenty years ago an elegant proof based on randomized amplification was found.

The crossing number $\mathrm{cr}(G)$ of a graph $G = (V, E)$ is the smallest number of crossings over all drawings of G in the plane, that is, vertices are represented by points and edges are represented by curves connecting the corresponding points.

▶ **Lemma 1** (Crossing Lemma). *Let $G = (V, E)$ be a simple graph (no loops and multiple edges). If $|V| = n$, $|E| = m$ and $m \geq 4n$, then for the crossing number $\mathrm{cr}(G)$ we have $\mathrm{cr}(G) \geq cm^3/n^2$ for some constant $c > 0$.*

Since there are planar multi-graphs with a fixed number of vertices and an arbitrary number of edges, there is no analog of the Crossing Lemma for general multi-graphs. In the working group we were looking at a special case, where we only consider drawings which are

(i) *simple*, that is, edges are drawn as simple curves such that any two edges have at most one point in common, a crossing or a common endpoint, and

(ii) *non-homotopic*, that is, the drawing has no pair of *homotopic* parallel edges, i.e. , for any pair e_1, e_2 of parallel edges, both of the two regions bounded by $e_1 \cup e_2$ in the drawing contains at least one vertex of G.

Note that with these two assumptions, the maximum number of edges of a planar (multi)graph of n vertices is still $3n - 6$. We call such drawings of (multi-)graphs *simple non-homotopic*.

▶ **Question.** *Does the Crossing Lemma still hold for multigraphs with simple non-homotopic drawings? That is, are there constants $c', d > 0$ such that if $|V| = n$, $|E| = m$ and $m \geq dn$, then any simple non-homotopic drawing of G has at least $c'm^3/n^2$ crossings?*

It was observed that the restriction to simple drawings is necessary. The following is a description of a non-homotopic, but non-simple drawing where the Crossing Lemma fails: Let G be the vertex-disjoint union of an independent set $\{c_1, \ldots, c_n\}$ and a complete bipartite graph with bipartition classes $\{a_1, \ldots, a_n\}$ and $\{b_1, \ldots, b_n\}$ where each edge has multiplicity n. Hence G has $3n$ vertices and n^3 edges. The vertices of G are placed on three horizontal rows, with $\{a_1, \ldots, a_n\}$ on top, $\{c_1, \ldots, c_n\}$ in the middle, and $\{b_1, \ldots, b_n\}$ below. For each pair (a_i, b_j), $1 \leq i, j \leq n$, the k-th edge between a_i and b_j, $1 \leq k \leq n$, is routed using two straight segments: the first from a_i to a point between c_k and c_{k+1}, and the second from that point to b_j. This way every region defined by parallel edges contains at least one vertex c_k, i.e., no two parallel edges are homotopic. Clearly, each pair of independent edges crosses at most twice and any two adjacent edges cross at most once. Thus the number of crossings in the drawing of G is less than twice its number of edges squared, i.e., $\mathrm{cr} < 2n^6 = O(n^6)$. On the other hand the Crossing Lemma would predict a lower bound of $\Omega(m^3/n^2) = \Omega(n^9/n^2) = \Omega(n^7)$, which is not true for our example.

In this regime of topological non-homotopic drawings we can prove the following partial results.

- The Crossing Lemma still holds for any simple non-homotopic drawing of G in which each edge is an x-monotone curve.
- The Crossing Lemma still holds for any simple non-homotopic drawing of G with the additional property that parallel edges cross exactly the same edges.
- There exist constants $c', d > 0$ such that if $|V| = n$, $|E| = m$ and $m \geq dn$, then any simple non-homotopic drawing of G has at least $c'm^{2.5}/n^{1.5}$ crossings.

There is a lack of constructions for topological non-homotopic drawings with high edge multiplicities. Inspired by Moon's drawing of the complete graph [3] we could find drawings of multi-matchings with $|V| = 2n$, $|E| = 2n(n-1)$ and $\mathrm{cr} \approx n^4$.

References

1   Frank T. Leighton. Complexity Issues in VLSI. Foundations of Computing Series. MIT Press., 1983.
2   Miklós Ajtai, Vasek Chvátal, Monroe M. Newborn, and Endre Szemerédi. Crossing-free subgraphs. *In Theory and Practice of Combinatorics,* pages 9–12. North-Holland Mathematics Studies, 1982.
3   John W. Moon. On the Distribution of Crossings in Random Complete Graphs, *J. Soc. Indust. Appl. Math.* 13 (1965), 506–510.

(a) **(b)** **(c)** **(d)** **(e)** **(f)**
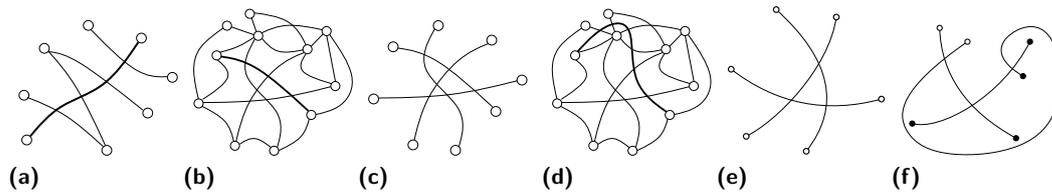
■ **Figure 4** (a) A crossing configuration that is forbidden in a 3-planar topological graph (the thick edge is crossed more than three times). (b) A 3-planar topological graph. (c) A crossing configuration that is forbidden in a 4-quasi planar topological graph. (d) A 4-quasi planar topological graph obtained from the one of Figure (c) by suitably rerouting the thick edge. (e) An untangled 3-crossing; all vertices belong to the same face of the arrangement (the outer face). (f) A tangled 3-crossing; the circled vertices and the solid vertices belong to distinct faces of the arrangement.

4.2 Working group B1: On the Relationship between k-Planar and k-Quasi Planar Graphs

Patrizio Angelini (Universität Tübingen, DE), Michael Bekos (Universität Tübingen, DE), Franz J. Brandenburg (Universität Passau, DE), Giordano Da Lozzo (University of California – Irvine, US), Giuseppe Di Battista (Roma Tre University, IT), Walter Didimo (University of Perugia, IT), Giuseppe Liotta (University of Perugia, IT), Fabrizio Montecchiani (University of Perugia, IT), and Ignaz Rutter (TU Eindhoven, NL)

Introduction and Preliminaries

Drawings of graphs are used in a variety of application domains. The aim of a graph visualization is to clearly convey the structure of the data and their relationships, in order to support users in their analysis tasks. In this respect, there is a general consensus that graph layouts with many edge crossings are hard to read, as also witnessed by several user studies on the subject (see e.g. [9]). An emerging research area, informally recognized as *beyond planarity*, concentrates on different models of graph planarity relaxations, which allow edge crossings but forbid specific configurations that would affect the readability too much.

Two of the most popular families in this context are the *k-planar* and the *k-quasi planar* graphs, which are usually defined in terms of *topological graphs*, i.e., graphs with a geometric representation in the plane with vertices as points and edges as Jordan arcs connecting their endpoints. Namely, a topological graph is *k-planar* ($k \geq 1$) if each edge is crossed at most k times, while it is *k-quasi planar* ($k \geq 2$) if it contains no k pairwise crossing edges. The 2-quasi planar graphs coincide with the planar graphs; also, 3-quasi planar graphs are simply called *quasi planar*. If G and G' are two isomorphic graphs, we write $G \simeq G'$. A graph G' is k-planar (k-quasi planar) if there exists a k-planar (k-quasi planar) topological graph G such that $G \simeq G'$. Figure 4a shows a crossing configuration that is forbidden in a 3-planar graph. Figure 4b depicts a 3-planar topological graph that is not 2-planar (the thick edge is crossed three times). Figure 4c shows a crossing configuration that is forbidden in a 4-quasi planar graph. Figure 4d depicts a 4-quasi planar topological graph that is not 3-quasi planar.

The k-planarity and k-quasi planarity hierarchies have been widely explored in graph theory, graph drawing, and computational geometry, mostly in terms of edge density. Pach and Tóth [7] proved that an n-vertex k-planar *simple* topological graph has at most $1.408\sqrt{k}n$

edges. For $k \leq 4$, they established a finer bound of $(k+3)(n-2)$, which is tight for $k \leq 2$. For $k = 3$, the best known upper bound is $5.5n - 11$, which is tight up to a constant [4, 5].

Concerning k-quasi planar graphs, a 20-year-old conjecture by Pach, Shahrokhi, and Szegedy [6] asserts that, for every fixed k, the maximum number of edges in a k-quasi-planar graph with n vertices is $O(n)$. However, so far, linear bounds have been proven only for $k \leq 4$; see the works of Agarwal et al. [3], Pach et al. [5], Ackerman and Tardos [2], and Ackerman [1]. For $k \geq 5$, several authors proved super-linear upper bounds; the most recent results are due to Suk and Walczak [8], who proved that any k-quasi planar simple topological graph on n vertices has at most $c_k n \log n$ edges, where c_k is a number depending only on k.

During this Dagstuhl Seminar, we studied inclusion relationships between the hierarchies of k-planar graphs and of k-quasi planar graphs. Note that some results on this relationship can be immediately derived from the definition of the two classes and from the previous results on their maximum edge density. For example, 3-quasi planar graphs can be denser than 3-planar graphs, and thus there are infinitely many 3-quasi planar graphs that are not 3-planar. On the other hand, for any $k \geq 1$, every k-planar graph is $(k+2)$-quasi planar, as in a set of $k+2$ mutually crossing edges every edge is crossed at least $k+1$ times. In this work we ask whether every k-planar simple graph is $(k+1)$-quasi planar, and prove that this is true for any $k \geq 3$. Note that for $k = 1$ the answer is trivially negative, since 2-quasi planar graphs are planar. We thus leave the question open for $k = 2$.

Basic Definitions

Two edges *cross* if they share one interior point and alternate around it. Two edges *intersect* if they either cross or share a common endpoint. A graph is *almost simple* if any two edges cross at most once, and *simple* if any two edges intersect at most once. A graph divides the plane into connected regions, called *faces*. The unbounded region is the *outer face*.

Given a subgraph X of a graph G, the *arrangement of X*, denoted by $\mathcal{A}_X$, is the arrangement of the curves corresponding to the edges of X. We denote the vertices and edges of X by $V(X)$ and $E(X)$. A *node* of $\mathcal{A}_X$ is either a vertex or a crossing point of X. A *segment* of $\mathcal{A}_X$ is a part of an edge of X that connects two nodes, i.e., a maximal uncrossed part of an edge of X. A *fan* is a set of edges that share a common endpoint. A set of k vertex-disjoint mutually crossing edges in a topological graph G is called a *k-crossing*. A k-crossing X is *untangled* if in the arrangement $\mathcal{A}_X$ of X all nodes corresponding to vertices in $V(X)$ are incident to a common face. Otherwise, it is *tangled*. For example, the 3-crossing in Figure 4e is untangled, whereas the one in Figure 4f is tangled. We observe the following.

▶ **Observation 1.** *Let $G = (V, E)$ be a k-planar simple topological graph and let X be a $(k+1)$-crossing in G. An edge in $E(X)$ cannot be crossed by any other edge in $E \setminus E(X)$. In particular, for any two distinct $(k+1)$-crossings X and Y in G, $E(X) \cap E(Y) = \emptyset$ holds.*

Edge Rerouting Operations and Proof Strategy

The strategy of our proof works as follows. Let G be any k-planar simple topological graph.

First, we show that it is possible to assume that every $(k+1)$-crossing in G is untangled. For this, we provide a technique to locally redraw the edges of any $(k+1)$-crossing without creating any other crossing, which may be of independent interest; see Figure 5c and 5d.

Second, we define an *edge rerouting* operation to redraw a single edge $e = \{u, v\} \in E(X)$ of an untangled $(k+1)$-crossing X of a k-planar simple topological graph G in order to remove this $(k+1)$-crossing while not introducing any new one. This operation is illustrated in Figure 5a and 5b and formally defined as follows. Consider a vertex $w \in V(X) \setminus \{u, v\}$.

(a)   **(b)**   **(c)**   **(d)**
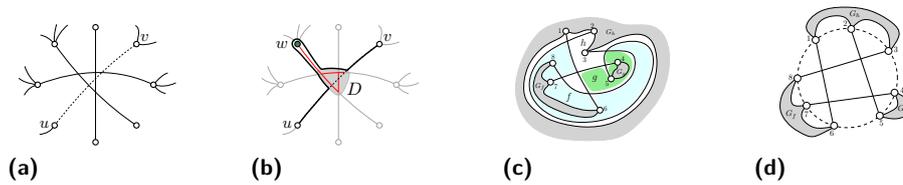
◼ **Figure 5** The rerouting operation for dissolving untangled k-crossings. (a) An untangled k-crossing X. (b) The rerouting of the dashed edge (u, v) around the marked vertex w. The arrangement $\mathcal{A}'_X$ is thin red, the removed nodes and segments are gray. Note that the dashed curve is part of $\mathcal{A}'_X$. Illustration of the untangling procedure. (c) A 3-planar simple topological graph with a 4-crossing X. (d) The topological graph resulting from the procedure that untangles X.

Denote by $\mathcal{A}'_X$ the arrangement obtained from the original arrangement $\mathcal{A}_X$ of X by removing all nodes corresponding to vertices in $V(X) \setminus \{u, v, w\}$, together with their incident segments, and by removing edge (u, v). The operation of *rerouting $e = \{u, v\}$ around w* consists of redrawing e sufficiently close to the boundary of the outer face of $\mathcal{A}'_X$, choosing the routing that passes close to w, in such a way that e does not cross any edge in $E \setminus E(X)$ except for the fan incident to w. More precisely, let D be a topological disk that encloses all crossing points of X and such each that edge in $E(X)$ crosses the boundary of D exactly twice. Then, the rerouted edge keeps unchanged the parts of e that go from u to the boundary of D and from v to the boundary of D. We call the unchanged parts of a rerouted edge its *tips* and the part that routes around w its *hook*.

We can prove the topological graph $G' \simeq G$ obtained by applying this operation has fewer $(k + 1)$-crossings than G and is almost-simple. However, G' may be not simple and may be not k-planar any longer. While simplicity can be later obtained by suitably redrawing some edges in G', as we will discuss later, the fact that G' is not k-planar does not allow us to use an algorithm in which the $(k + 1)$-crossings are removed one-by-one by repeatedly applying this operation.

We thus define a *global rerouting* operation, which picks an edge from each $(k+1)$-crossing in G and applies the rerouting operation *simultaneously* for all of these edges. Note that the global rerouting is well-defined since the $(k + 1)$-crossings are pairwise edge-disjoint by Observation 1. As discussed above, we proved that a single edge rerouting operation yields an almost-simple graph with fewer $(k + 1)$-crossings. We now discuss the analogous properties for a global rerouting. In terms of $(k+1)$-crossings, all those that were present in the original graph have been removed by the single operations composing the global rerouting. Further, we can prove that no new $(k + 1)$-crossing can be created by any global rerouting, and hence that the resulting topological graph $G' \simeq G$ is $(k + 1)$-quasi planar. We remark that this is the only argument that fails when $k = 2$, and is the reason why we can state the result only for $k \geq 3$. In order to ensure that G' is also almost-simple, on the other hand, we cannot perform any global rerouting, but we have to be careful in the choice of the edges to reroute and of the vertices around which these edges are rerouted. We discuss the conditions for G' to be almost simple in the following lemma.

▶ **Lemma 2.** *Graph G' is an almost-simple topological graph if and only if the following conditions hold.*

C.1 *No two edges are rerouted around the same vertex; see Figure 6a.*

C.2 *There is no pair of edges e, d such that e is rerouted around an endpoint of d and d is rerouted around an endpoint of e; see Figure 6b.*
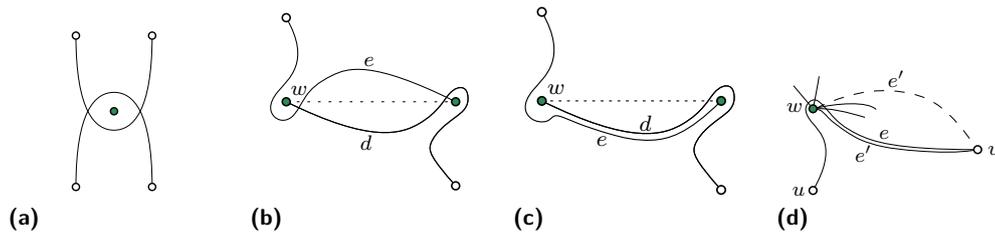
Figure 6 (a–b) Topological graphs that are not almost simple, arising from a global rerouting. (c) Avoiding the non-simplicity in (b) by redrawing one the two rerouted in edges. The vertices used for rerouting are filled green. Dotted parts of the drawing may consist of several vertices and edges. (d) Illustration for the redrawing technique to obtain simplicity.

In order to satisfy Condition C.2, we model the problem of choosing the vertices around which the edges have to be rerouted as a matching problem on a suitably defined bipartite graph, and prove that such a matching always exists by using the Hall's theorem.

Hence, after applying the global rerouting, if two edges cross more than once, then this is due to Condition C.2. In this case, we prove that it is possible to redraw one of these two edges, namely its portion between the two crossing points, without creating new $(k+1)$-crossings and without crossing any other edge more than once; see Figure 6b and 6c.

Once all the pairs of edges that cross more than once have been resolved, hence obtaining an almost-simple topological graph G' that is still $(k+1)$-quasi planar, it only remains to make G' simple, by resolving the possible pairs of adjacent edges that cross with each other. To do so, we again employ suitable redrawing techniques that do not break $(k+1)$-quasi planarity and do not introduce undesired crossings; see Figure 6d for an example.

As a final result of this work, we hence obtain the following theorem.

▶ **Theorem 3.** *For any $k \geq 3$, every k-planar graph is $(k+1)$-quasi planar.*

References

1   Eyal Ackerman. On the maximum number of edges in topological graphs with no four pairwise crossing edges. *Discrete & Computational Geometry*, 41(3):365–375, 2009.
2   Eyal Ackerman and Gábor Tardos. On the maximum number of edges in quasi-planar graphs. *J. Comb. Theory, Ser. A*, 114(3):563–571, 2007.
3   Pankaj K. Agarwal, Boris Aronov, János Pach, Richard Pollack, and Micha Sharir. Quasi-planar graphs have a linear number of edges. *Combinatorica*, 17(1):1–9, 1997.
4   János Pach, Rados Radoicic, Gábor Tardos, and Géza Tóth. Graphs drawn with at most 3 crossings per edge. Seminar on Combinatorial Computing, October 2002.
5   János Pach, Rados Radoicic, and Géza Tóth. Relaxing planarity for topological graphs. In *JCDCG*, volume 2866 of *LNCS*, pages 221–232. Springer, 2003.
6   János Pach, Farhad Shahrokhi, and Mario Szegedy. Applications of the crossing number. *Algorithmica*, 16(1):111–117, 1996.
7   János Pach and Géza Tóth. Graphs drawn with few crossings per edge. *Combinatorica*, 17(3):427–439, 1997.
8   Andrew Suk and Bartosz Walczak. New bounds on the maximum number of edges in k-quasi-planar graphs. *Comput. Geom.*, 50:24–33, 2015.
9   Colin Ware, Helen C. Purchase, Linda Colpoys, and Matthew McGill. Cognitive measurements of graph aesthetics. *Information Visualization*, 1(2):103–110, 2002.

4.3 Working group B2: Beyond-Planarity of Graphs with Bounded Degree

Muhammad Jawaherul Alam (University of California – Irvine, US), Kathrin Hanauer (Universität Passau, DE), Seok-Hee Hong (The University of Sydney, AU), Stephen G. Kobourov (University of Arizona – Tucson, US), Quan Nguyen (The University of Sydney), Sergey Pupyrev (Facebook – Menlo Park, US), and Md. Saidur Rahman (Bangladesh University of Eng.& Technology, BD)

We consider several problems related to beyond-planar graphs, that is, non-planar graphs with topological constraints such as specific types of crossings, or with some forbidden crossing patterns. In this context, we study graphs of bounded maximum degree.

Let $G = (V, E)$ be a graph with maximum vertex degree Δ. Furthermore, G is ...

cubic if every vertex of G has degree exactly three.

quasi-planar if G has an embedding so that there are no three pairwise crossing edges.

outer-quasi-planar if G is quasi-planar with the additional restriction that all vertices are incident to the outer face.

2-layer quasi-planar if G is outer-quasi-planar with the vertices being predefined to reside on exactly one of two layers.

1-sided 2-layer quasi-planar if G is 2-layer quasi-planar and the ordering on one layer is fixed.

right-angle crossing (RAC) if G has a straight-line drawing in the plane so that all crossings have a 90 degree angle.

fan-crossing-free if G has an embedding so that no edge crosses two or more other edges that have a common end vertex.

Known Related Results

If G has geometric thickness of at most two, it can be embedded as two stacked planar graphs, which cannot produce a pairwise crossing of more than two edges.

▶ **Proposition 1.** *Every graph with a geometric thickness of at most two is quasi-planar.*

▶ **Proposition 2** ([2])**.** *Leveled-planarity testing is $\mathcal{NP}$-hard.*

▶ **Proposition 3** ([1])**.** *Every hamiltonian cubic graph is RAC.*

Discussed Questions

▶ **Question 4.** *What is the computational complexity of recognizing quasi-planar graphs with maximum degree Δ?*

As the (geometric) thickness of graphs with $\Delta \leq 4$ is known to be two, all such graphs are quasi-planar by Proposition 1. An interesting case is $\Delta = 5$. A possibly useful observation is that the graphs of maximum degree five have linear arboricity three, that is, they can be decomposed into three sets of edge-disjoint linear forests (every forest is a set of paths).

▶ **Question 5.** *What is the complexity of recognizing outer-quasi-planar or 2-layer quasi-planar graphs with maximum degree Δ?*

As leveled-planarity testing is $\mathcal{NP}$-hard (see Proposition 2) and every leveled-planar graph is outer-quasi-planar, also testing 2-layer quasi-planarity might be $\mathcal{NP}$-hard. However, the other direction of the reduction is unclear.

The general outer-quasi-planar case remains open.

▶ **Question 6.** *For what values of Δ are the graphs of bounded maximum degree RAC? What is the complexity of recognizing RAC graphs of bounded maximum degree?*

Since K_6 does not admit a RAC drawing and graphs with $\Delta = 2$ are cycles, the above question is interesting for $\Delta = 3$ and $\Delta = 4$.

Bekos et al. show how to construct a RAC drawing for a hamiltonian cubic graph (Proposition 3), but the general case is open.

Note that every RAC drawing is quasi-planar as well as fan-crossing-free (graphs which can be embedded without fan-crossings). Hence, in order to answer Question 6, we first need to resolve the following one:

▶ **Question 7.** *Is it true that every graph of maximum degree $\Delta = 3$ ($\Delta = 4$) admits a (non-geometric, that is, with curves) both quasi-planar and fan-crossing-free drawing?*

Question 7 seems to be straightforward for $\Delta = 3$ (at least for 3-connected cubic graphs that admit a decomposition into three matchings) but less so for $\Delta = 4$.

Another set of questions is related to the density of non-planar graphs:

▶ **Question 8.** *What is the minimum density of maximal quasi-planar/outer-quasi-planar/RAC graphs?*

References

1 Evmorfia Argyriou, Michael Bekos, Michael Kaufmann, and Antonios Symvonis. *Geometric RAC Simultaneous Drawings of Graphs*, pages 287–298. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

2 Lenwood S. Heath and Arnold L. Rosenberg. Laying out graphs using queues. *SIAM Journal on Computing*, 21(5):927–958, 1992.

4.4   Working group C: Smooth Crossings

Evmorfia Argyriou (yWorks GmbH Tübingen, DE), Sabine Cornelsen (Universität Konstanz, DE), Martin Nöllenburg (TU Wien, AT), Yoshio Okamoto (The University of Electro-Communications – Tokyo, JP), Chrysanthi Raftopoulou (National Technical University of Athens, GR), and Alexander Wolff (Universität Würzburg, DE)

Orthogonal drawings date back to the 80's, with Valiant's [8], Leiserson's [6] and Leighton's [7] work on VLSI layouts and floor-planning applications and have been extensively studied over the years. The quality of an orthogonal drawing can be judged based on several aestetics criteria such as the required area, the total edge length, the total number of bends or maximum number of bends per edge. Towards the direction of "smoothening" an orthogonal drawing and improving its readability, Bekos et al. [4] introduced *smooth orthogonal drawings* that combine the clarity of orthogonal layouts and the artistic style of Lombardi drawings [5] by "replacing" the bends of the edges with smooth circular arcs.

(a) OC$_3$-layout   **(b)** SC$_1$-layout
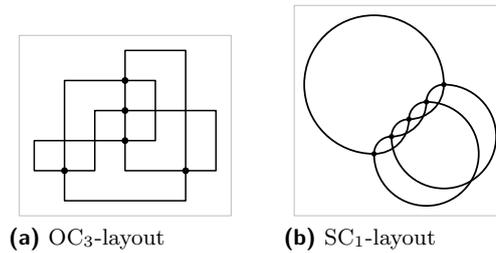
Figure 7 Different 2-planar drawings of K_5.

To the best of our knowledge, smooth orthogonal drawings have only been considered for planar graphs. We are interested in orthogonal and smooth orthogonal layouts of non-planar graphs such as 1-planar graphs that admit a drawing with at most one crossing per edge. Our goal is to study how typical aesthetics criteria for planar (smooth) orthogonal drawings, e.g., edge complexity, extend to non-planar drawings. We consider drawings, where vertices are mapped to points in $\mathbb{R}^2$ and edges are mapped to curves of the following two types.

Orthogonal Layout: Each edge is drawn as a sequence of vertical and horizontal line segments. Two segments meet in a bend.

Smooth Orthogonal Layout [4]: Each edge is drawn as a sequence of vertical and horizontal line segments as well as circular arcs: quarter circles, semicircles, and three-quarter circles. Where segments meet, they must have a common tangent.

The curve complexity of a drawing is the maximum number of segments used for an edge. An OC_k-*layout* is an orthogonal layout with curve complexity k, i.e., an orthogonal layout with at most $k-1$ bends per edge. A SC_k-*layout* is a smooth orthogonal layout with curve complexity k. We concentrate on graphs with a fixed embedding, i.e., a fixed rotation system and a fixed outer face.

As already mentioned, smooth orthogonal drawings have only been considered for planar graphs. Moreover, the vertex degree is usually restricted to four since every vertex has four available *ports* (North, South, East, West), where the edges enter and leave a vertex with horizontal or vertical tangents. In addition, the usual model insists that no two segments incident to the same vertex can use the same port. Note that using the same port necessarily causes overlaps only in the case of straight-line segments.

Biedl and Kant [3] presented a linear-time and -space algorithm that draws any connected graph of maximum degree four orthogonally on a grid of size $n \times n$ with at most $2n + 2$ bends, where each edge is bent at most twice. Note that their approach introduces crossings to the produced drawing. For the case that the given graph is planar, they describe how to obtain a planar orthogonal drawing with at most two bends per edge, except possibly for one edge on the outer face.

The smallest non-planar graph is K_5, which is 1-planar. Following the general algorithm of Biedl and Kant, we get an orthogonal drawing of K_5 with edge-complexity three (two bends per edge) as in Figure 7a. The resulting drawing is 2-planar. Bekos et al. [2] have shown that any biconnected graph of maximum degree four admits a (non-planar) SC$_1$-layout. For a (2-planar) SC$_1$-drawing of K_5, see Figure 7b.

We focus on 1-planar (and 1-plane) graphs. In the following, we examine the curve complexity of 1-plane drawings in the orthogonal and smooth orthogonal drawing style.

(a) OC_4-layout of K_5 **(b)** OC_4-layout of a 9-vertex graph
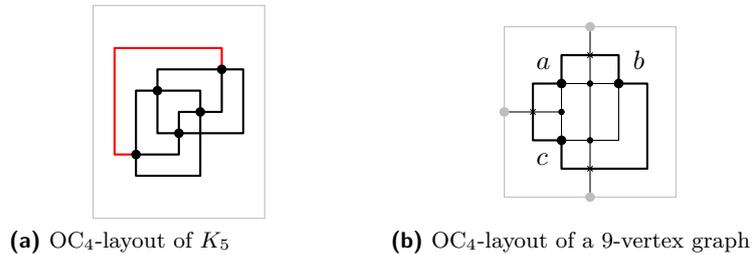
Figure 8 Biconnected 1-plane graphs that do not admit an OC_3-layout.

Orthogonal 1-Plane Drawings

In this section, we examine the case of orthogonal 1-plane drawings and we present one negative and one positive result. Namely, we show using a counterexample that not every biconnected graph of maximum degree four with a fixed embedding admits an OC_3-layout, whereas we prove that every biconnected 1-planar graph of maximum degree four admits an OC_5-layout.

▶ **Theorem 1.** *Not every biconnected graph of maximum degree four with a fixed embedding admits an OC_3-layout.*

Proof. The complete graph on five vertices has the above property (Figure 8a). For another example refer to Figure 8b: Vertices a, b and c create a triangle T and all vertices have their two remaining ports in the interior of T. Then, T has at least seven bends, and therefore at least one edge of T has at least three bends and edge-complexity four.   ◀

▶ **Corollary 2.** *There is a biconnected 1-planar graph of maximum degree four with n vertices and a given embedding that has $\mathcal{O}(n)$ edges with at least three bends in any OC_4-layout respecting the embedding.*

Proof. We use t copies of the graph of Figure 8b in a column by connecting the gray vertices. The graph has $n = 9t$ vertices and $\mathcal{O}(n)$ edges.   ◀

▶ **Theorem 3.** *Every biconnected 1-planar graph of maximum degree four admits an OC_5-layout.*

Proof. Our algorithm is a slight modification of the algorithm of Biedl and Kant [3]. First we planarize the given 1-planar embedding by introducing dummy vertices at crossings.

By the algorithm of Biedl and Kant, all edges have at most two bends, except possibly for one edge on the outer face that can have three bends. We only have to make sure that we do not introduce more than four bends in any edge adjacent to the outer face. All other crossing edges have at most four bends and therefore edge-complexity five.

The algorithm of Bield and Kant computes the drawing incrementally based on an st-numbering of the vertices of the graph i.e., each edge must have at least one predecessor (except s) and at least one successor (except t). Also, since our graph is biconnected, for any $s, t \in V$, there exists an st-numbering such that s is the sink vertex and t is the source vertex. We claim that we can choose s and t so that no edge has four bends. We consider the following three cases.
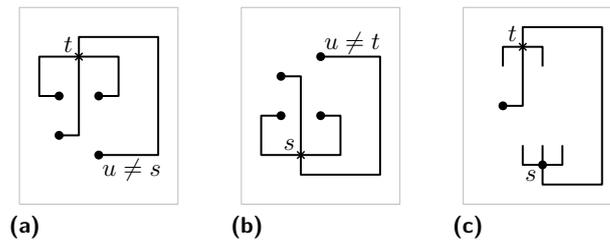
Figure 9 Different cases for the outer face.
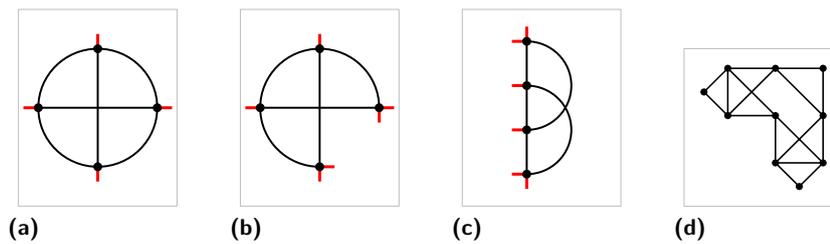
Figure 10 (a) SC_1-layouts for K_4 and (b)–(c) for $K_4 - e$ with restricted ports. (d) A biconnected outer-1 plane graph that does not have an SC_1-layout.

The outer face has at least two crossings. Let s and t be the dummy vertices at the crossings. Then the crossing edge that enters t from above (s from below) had at most three bends before entering and at most one after entering; see Figures 9a-9b respectively.

The outer face has one crossing. Let t be the dummy vertex at the crossing. For t we can argue as above. Let s be any vertex on the outer face. If s has degree less than four, we don't use the bottom port of s, and there is no problem. Otherwise s has degree four; there is at least one neighbor $s' \neq t$ on the outer face. We route the edge (s, s') through the bottom port of s. It has no crossing, hence it gets at most three bends.

The outer face has no crossings. No problem (see Figure 9c).

◀

Smooth Orthogonal 1-Plane Drawings

We focus on outerplane 1-planar graphs (in short: outer-1 plane graphs), and start with the following observation. The complete graph on four vertices with free ports towards its outer face has a unique SC_1-layout, shown in Figure 10a. Removing one edge, and restricting all ports towards its outer face, there exist two SC_1-layouts, shown in Figures 10b and 10c.

▶ **Theorem 4.** *Not every biconnected outer-1 plane graph has an SC_1-layout.*

Proof. Take the graph in Fig. 10d. It has two subgraphs isomorphic to $K_4 - e$ (with restricted ports) that share a vertex. It is not possible to combine any of the two possible SC_1-layouts for one copy of $K_4 - e$ with any SC_1-layout for the other copy. ◀

▶ **Theorem 5.** *Every biconnected outer-1 plane graph has an SC_4-layout.*

Proof. Planarize and apply the algorithm of Bekos et al. [1] that produces an SC_2-layout. ◀

▶ **Theorem 6.** *Every biconnected outer-1 plane graph where the endpoints of any two crossing edges induce a K_4 has an SC_1-layout.*
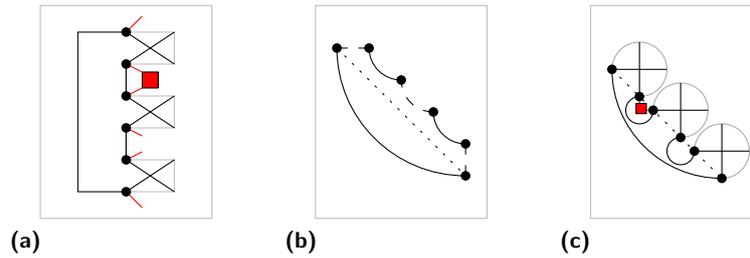
Figure 11 Example.

Proof. In this case all copies of K_4 are vertex-disjoint. We remove all pairs of crossing edges, producing a biconnected outerplane graph. This turns each copy of K_4 into a face of length four. We use the algorithm of Bekos et al. [1]. Whenever we want to add a "special" face of length four, we use the SC_1-layout of Figure 10a. We have to check and prove that invariants are preserved, that the algorithm can start, and that the diagonal stripes are well-defined. The area can be exponential. ◀

Future Work

Can this be extended to all outer-1 plane graphs where crossings are vertex-disjoint?

- In Figure 11a we have a face defined by vertices and crossing points, and in Figure 11c its SC_1-layout. The free ports, force components attached to the face with two edges, to be drawn before the diagonal.

References

1 Md. Jawaherul Alam, Michael A. Bekos, Michael Kaufmann, Philipp Kindermann, Stephen G. Kobourov, and Alexander Wolff. Smooth orthogonal drawings of planar graphs. In A. Pardo and A. Viola, editors, *LATIN*, volume 8392 of *LNCS*, pages 144–155. Springer, 2014.

2 Michael A. Bekos, Martin Gronemann, Sergey Pupyrev, and Chrysanthi N. Raftopoulou. Perfect smooth orthogonal drawings. In N. G. Bourbakis, G. A. Tsihrintzis, and M. Virvou, editors, *IISA*, pages 76–81. IEEE, 2014.

3 Therese C. Biedl and Goos Kant. A better heuristic for orthogonal graph drawings. *Computational Geometry*, 9(3):159–180, 1998.

4 Michael A. Bekos, Michael Kaufmann, Stephen G. Kobourov, and Antonios Symvonis. Smooth orthogonal layouts. *Journal of Graph Algorithms and Applications*, 17(5):575–595, 2013.

5 Christian A. Duncan, David Eppstein, Michael T. Goodrich, Stephen G. Kobourov, and Martin Nöllenburg. Lombardi drawings of graphs. *Journal of Graph Algorithms and Applications*, 16(1):85–108, 2012.

6 Charles E. Leiserson. Area-efficient graph layouts (for VLSI). In *FOCS*, pages 270–281. IEEE, 1980.

7 Frank T. Leighton. New lower bound techniques for VLSI. *Mathematical systems theory*, 17(1):47–70, 1984.

8 Leslie G. Valiant. Universality considerations in VLSI circuits. *IEEE Transactions on Computers*, 30(2):135–140, 1981.

5 List of Participants

We invited 40 researchers grouped in three main areas of research, representing 18 different countries; 29 of them attended the seminar. Annotation: 2⚒ (industry), 6♀ (female), 11★ (young researchers).

A) Graph Drawing

1.	**Alam, Md. Jawaherul**	Univ. of California – Irvine	U.S.A.	★
2.	**Angelini, Patrizio**	Univ. of Tübingen	Germany	★
3.	**Bekos, Michael**	Univ. of Tübingen	Germany	★
4.	**Brandenburg, Franz**	Univ. of Passau	Germany	
5.	**Cornelsen, Sabine**	Univ. of Konstanz	Germany	♀
6.	**Da Lozzo, Giordano**	Univ. of California – Irvine	U.S.A.	★
7.	**Di Battista, Giuseppe**	Univ. of Rome III	Italy	
8.	**Didimo, Walter**	Univ. of Perugia	Italy	
9.	**Hanauer, Kathrin**	Univ. of Passau	Germany	♀★
10.	**Hong, Seok-Hee**	Univ. of Sydney	Australia	♀
11.	**Kobourov, Stephen**	Univ. of Arizona	U.S.A.	
12.	**Liotta, Giuseppe**	Univ. of Perugia	Italy	
13.	**Montecchiani, Fabrizio**	Univ. of Perugia	Italy	
14.	**Pupyrev, Sergey**	Facebook	USA	★⚒
15.	**Roselli, Vincenzo**	Univ. of Rome III	Italy	★

B) Computational and Combinatorial Geometry

16.	**Argyriou, Evmorfia**	yWorks	Germany	♀★⚒
17.	**Felsner, Stefan**	TU Berlin	Germany	
18.	**Kaufmann, Michael**	Univ. of Tübingen	Germany	
19.	**Nöllenburg, Martin**	TU Vienna	Austria	★
20.	**Quan Nguyen**	Univ. of Sydney	Australia	★
21.	**Okamoto, Yoshio**	Tokyo Tech	Japan	
22.	**Pach, János**	EPFL	Switzerland	
23.	**Rutter, Ignaz**	TU Eindhoven	Netherlands	★
24.	**Wolff, Alexander**	Würzburg Univ.	Germany	

C) Graph Theory

25.	**Raftopoulou, Chrysanthi**	NTU of Athens	Greece	♀★
26.	**Rahman, Md. Saidur**	Bangladesh Univ.	Bangladesh	
27.	**Tóth, Géza**	Rényi Institute Budapest	Hungary	
28.	**Ueckerdt, Torsten**	KIT	Germany	★
29.	**Valtr, Pavel**	Charles Univ.	Czech Republic	

Participants

- Muhammad Jawaherul Alam
  University of California –
  Irvine, US
- Patrizio Angelini
  Universität Tübingen, DE
- Evmorfia Argyriou
  yWorks GmbH Tübingen, DE
- Michael Bekos
  Universität Tübingen, DE
- Franz J. Brandenburg
  Universität Passau, DE
- Sabine Cornelsen
  Universität Konstanz, DE
- Giordano Da Lozzo
  University of California –
  Irvine, US
- Giuseppe Di Battista
  Roma Tre University, IT
- Walter Didimo
  University of Perugia, IT
- Stefan Felsner
  TU Berlin, DE

- Kathrin Hanauer
  Universität Passau, DE
- Seok-Hee Hong
  The University of Sydney, AU
- Michael Kaufmann
  Universität Tübingen, DE
- Stephen G. Kobourov
  University of Arizona –
  Tucson, US
- Giuseppe Liotta
  University of Perugia, IT
- Fabrizio Montecchiani
  University of Perugia, IT
- Quan Nguyen
  The University of Sydney, AU
- Martin Nöllenburg
  TU Wien, AT
- Yoshio Okamoto
  The University of
  Electro-Communications –
  Tokyo, JP
- Janos Pach
  EPFL – Lausanne, CH

- Sergey Pupyrev
  Facebook – Menlo Park, US
- Chrysanthi Raftopoulou
  National Technical University of
  Athens, GR
- Md. Saidur Rahman
  Bangladesh University of Eng. &
  Technology, BD
- Vincenzo Roselli
  Roma Tre University, IT
- Ignaz Rutter
  TU Eindhoven, NL
- Géza Tóth
  Alfréd Rényi Institute of
  Mathematics – Budapest, HU
- Torsten Ueckerdt
  KIT – Karlsruher Institut für
  Technologie, DE
- Pavel Valtr
  Charles University – Prague, CZ
- Alexander Wolff
  Universität Würzburg, DE

Assessing ICT Security Risks in Socio-Technical Systems

Edited by

Tyler W. Moore[1], Christian W. Probst[2], Kai Rannenberg[3], and Michel van Eeten[4]

1   University of Tulsa, US, `tyler-moore@utulsa.edu`
2   Technical University of Denmark – Lyngby, DK, `cwpr@dtu.dk`
3   Goethe-Universität Frankfurt am Main, DE, `kai.rannenberg@m-chair.de`
4   TU Delft, NL, `m.j.g.vaneeten@tudelft.nl`

―――― **Abstract** ――

This report documents the program and the outcomes of Dagstuhl Seminar 16461 "Assessing ICT Security Risks in Socio-Technical Systems". As we progress from classic mechanical or electrical production systems, over ICT systems, to socio-technical systems, risk assessment becomes increasingly complex and difficult. Risk assessment for traditional engineering systems assumes the systems to be deterministic. In non-deterministic systems, standard procedure is to fix those factors that are not deterministic. These techniques do not scale to ICT systems where many risks are hard to trace due to the immaterial nature of information. Beyond ICT systems, socio-technical systems also contain human actors as integral parts of the system. In such socio-technical systems there may occur unforeseen interactions between the system, the environment, and the human actors, especially insiders. Assessing ICT security risks for socio-technical systems and their economic environment requires methods and tools that integrate relevant socio-technical security metrics. In this seminar we investigated systematic methods and tools to estimate those ICT security risks in socio-technical systems and their economic environment. In particular, we searched for novel security risk assessment methods that integrate different types of socio-technical security metrics.

1   Summary

Tyler W. Moore
Christian W. Probst
Kai Rannenberg
Michel van Eeten

The Dagstuhl Seminar 16461 "Assessing ICT Security Risks in Socio-Technical Systems" is part of a series of seminars that explore aspects of risk and security in socio-technical systems. After initial work on insider threats, the focus has turned towards understanding of relevant metrics and their application in novel security risk assessment methods.

Classical Risk Assessment

Assessing risk in classic mechanical or electrical production systems is difficult but possible, as experience shows. Historical knowledge provides us with approximation of likelihood of failures of machines and components. We can combine these likelihoods in modular ways, mapping the impact of the loss of a system component to the processes that component contributes to.

This approach works for traditional engineering systems, since the dependencies between components are expected to be known, and their behaviour is assumed to be deterministic. To reach a comparable level of predictability for risk assessment in areas that are less governed by machines, for example economics, standard procedure is to fix those factors that are not deterministic. The behaviour of buyers and sellers on a market, for example, is assumed to be gain-oriented and rational; dealing with irrational actors is less explored and hard to model.

Risk in ICT Systems and the Immaterial Nature of Information

The techniques that work well for assessing risk in classic mechanical or electrical production systems do not scale to ICT systems. A primary reason is that there is no clear connection between the usage of a system and the risk of failure. Many risks that we need to consider in ICT are hard to trace due to the immaterial nature of information [2]. Examples of such risks are unauthorized or illegitimate information flows. Information flows are more difficult to trace than material flows, as the flow is usually caused by a copying operation; the information is not missing at the source, which in the material world is an indicator of an illegitimate flow, *e.g.*, when goods are stolen. Moreover, the damage is barely related to the measurable amount of information [5, 9]. Several megabytes of, *e.g.*, white noise can be relatively harmless, while a single health data record or financial record of limited size can be a major problem, *e.g.*, in terms of loss of trust, reputation, or damage compensation payments.

Another reason why classic approaches have struggled to assess risk is that the threat from strategic adversaries is harder to model than random failures. Whereas events triggered by nature occur randomly, attackers can readily identify and target the weakest links present in systems, and adapt to evade defenses.

A final reason why assessing security risks is hard is that there is often an incentive to hide failures from public view, due to fears of reputational damage. This makes collecting data to empirically estimate loss probabilities very difficult.

Risk in Socio-technical Systems

Beyond ICT systems, socio-technical systems also contain human actors as integral parts of the system. In such socio-technical systems there may occur unforeseen interactions between the system, the environment, and the human actors, especially insiders [8].

Assessing the risk of the ICT system for human actors is difficult [4]; the assessment must take into account the effect of the ICT system on the environment, and it must quantify the likelihood for this risk to materialize. Assessing the risk of the human actor for the ICT system is difficult, too. As mentioned above, economics models human actors by assuming them to be gain-oriented and rational; dealing with irrational actors is less explored and hard to model. However, one of the biggest risks from human actors for an ICT system is irrational behaviour, or an unknown gain function.

Economics of Risk Assessment

The economic aspect both of the risk identified and the process of assessing risk often prohibits either risk mitigation or the assessment itself. Protection against irrational threats requires appropriate preventive measure, be it too restrictive policies or too intense surveillance. Neither the cost nor the effects of these measures are easily predictable [1].

Even worse, the cost for risk assessment itself can also be prohibitive. For example, trying to identify the actual risk for irrational behaviour or its impact on the system can be impossible or at least imply a too high price [4].

Security Metrics

As we concluded after the previous Dagstuhl Seminar 14491, well-defined data sources and clear instructions for use of the metrics are key assets "to understand security in today's complex socio-technical systems, and to provide decision support to those who can influence security".

Security metrics obviously cannot be applied on their own, but must be embedded in a well-defined framework of sources for metrics and computations on them [7]. Important topics include understanding the aspects surrounding metrics, such as sources, computations on metrics, relations to economics, and the analyses based on metrics.

Assessing ICT Security Risks in Socio-Technical Systems

Making risk in socio-technical systems assessable requires an understanding of how to address issues in these systems in a systematic way [3, 6]. In this seminar, we built upon the work in the predecessor seminars on insider threats and security metrics, and explores the embedding of human behaviour and security metrics into methods to support risk assessment.

Main findings

We established five working groups in the seminar, that discussed several times during the week and reported back in plenum. The results are presented in Section 4, and briefly summarized here.

Which data do we need to collect?

In a working group on "Collecting Data for the Security Investment Model", we considered the relationship between efforts to secure an organisation, the actual security level achieved through these efforts, and their effect on moderating attacks and the induced losses. We identified relevant, measurable indicators for the components in the model that relate metrics about components to the expected risk. Model outcomes could be used to guide security investments.

Which security risks should we consider?

To identify relevant risk assessment methods, we discussed the kind of risk relevant to measure in two working groups. On the one hand we explored "New Frontiers of Socio-Technical Security", where we considered disrupting new technologies and how they influence and change our perception of risk, or its limitations. The main example were orphaned devices in IoT systems, which often cannot be switched off, but pose a threat to the overall system

if they remain unmaintained. A similar problem space was explored in the working group on "Software Liability Regimes", which considered liability or lack thereof of producers to identify and fix problems.

Which attacker and user traits do we need to consider?

To understand relevant aspects of human actors involved in socio-technical systems, we established two more working groups. The group on "Unpacking the motivations of attackers" discussed how to understand attacker motivations in highly integrated socio-technical systems, where purpose and means play a fundamental role in the way and at which level(s) the cyberspace can be disrupted.

Conclusions

Assessing risk in socio-technical systems is and remains difficult, but can be supported by techniques and understanding of limitations and properties inherent to the system and the risk assessment methods applied. This seminar has explored how to identify these limitations and properties by exploring the different layers of socio-technical systems, their interactions, and their defining attributes.

A total of 36 researchers participated in the seminar across from different communities, which together span the range relevant to developing novel security risk assessment methods and to ensure the continuation from the previous seminars' results: cyber security, information security, data-driven security, security architecture, security economics, human factors, (security) risk management, crime science, formal methods, and social science.

References

1 Anderson, R., Barton, C., Böhme, R., Clayton, R., van Eeten, M. J. G., Levi, M., Moore, T., and Savage, S. (2012). Measuring the Cost of Cybercrime. Paper presented at the Workshop of Economics and Information Security (WEIS 2012), Berlin.

2 Blakley, B., McDermott, E., Morgan, J. P., and Geer, D. (2001). Information security is information risk management. In: NSPW'01 Proceedings of the 2001 workshop on New security paradigms (pp. 97–104). New York, NY: ACM.

3 Gollmann, D. (2012). From Insider Threats to Business Processes that are Secure-by-Design. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 3 (1/2), 4–12.

4 Hunker, J. and Probst, C. W. (2009). The Risk of Risk Analysis, And its relation to the Economics of Insider Threats. In: WEIS'09 Proceedings of the 2009 workshop on Economics of Information Security.

5 Johnson, M. E, Goetz, E., and Pfleeger, S. L. (2009). Security through Information Risk Management, IEEE Security and Privacy, 7(3):45–52.

6 Littlewood, B., Brocklehurst, S., Fenton, N., Mellor, P., Page, S., Wright, D., Dobson, J., McDermid, J., and Gollmann, D. (1993). Towards Operational Measures of Computer Security. Journal of Computer Security, 2(2–3):211–229.

7 Pieters, W., Van der Ven, S. H. G., and Probst, C. W. (2012). A move in the security measurement stalemate: Elo-style ratings to quantify vulnerability. In Proceedings of the 2012 New Security Paradigms Workshop (NSPW'12) (pp. 1–14). New York, NY: ACM. doi: `10.1145/2413296.2413298`.

8 Probst, C. W., Hunker, J., Bishop, M. and Gollmann, D. (2009). Countering Insider Threats. ENISA Quarterly Review, 5(2).

9 Tschersich, M., Kahl, C., Heim, S., Crane, S., Böttcher, K., Krontiris, I., Rannenberg, K. (2011). Towards Privacy-enhanced Mobile Communities – Architecture, Concepts and User Trials. In: Journal of Systems and Software, Volume 84, Issue 11, pp. 1947–1960, November 2011,

2   Table of Contents

Working groups

3   Overview of Talks

3.1   Complex socio-technical systems

Ross Anderson (University of Cambridge, GB)

When analysing the risks of complex socio-technical systems we have a lot of historical examples to use as a guide, from the Roman army through the Chinese civil service to the banking system as it emerged in the 19th century. What changes when you add software is that action is no longer local and personal; so the traditional ways of scaling action (ideologies, hierarchies, markets) are no longer all there is. Another factor that changes is the time constant. An important and neglected aspect is safety regulation. Traditionally, safety-critical equipment from cars through medical devices to electrotechnical equipment has been regulated by pre-market inspection, which has a time constant roughly equal to the product lifecycle, typically ten years. Once everything is online and hackable, your car will need monthly software updates just like your laptop and your phone. The current institutional arrangements are unable to cope, and major change is necessary. Richard Clayton, Eireann Leverett and I have done a big project on this for the European Commission; an academic paper should be out next year. The political vision is that just as Europe has become the world's privacy regulator (as Washington doesn't care and nobody else is big enough to matter), so also Europe should aim to be the world's safety regulator too.

3.2   Dealing with High-Consequence Low-Probability Events

Johannes M. Bauer (Michigan State University – East Lansing, US)

With increased connectivity and the multiplication of devices the ICT system increasingly resembles a complex adaptive system. In such systems, small changes may have large repercussions for overall system performance. Security incidents that have a low probability but potentially catastrophic consequences cannot easily be captured by statistical analyses and traditional forms of cost-benefit analysis. The alternative of preventing such events has potentially serious downsides, such as reducing the rate of innovation and possibly prohibitive costs. An alternative is to aim at designing resilient organizations, communities, nations, and global systems.

3.3   Usable Recovery

Zinaida Benenson (Universität Erlangen-Nürnberg, DE)

We propose usable recovery as a new paradigm to solve the "user involvement" dilemma in security. The security community expects that users should perform perimeter security tasks that will protect their devices from compromise. These tasks include installing antivirus,

choosing and managing strong passwords, making informed decisions when encountering security warnings, paying attention to deception cues in emails and on websites, updating all installed software, etc. The dilemma is that the effort required for those tasks does not match the security benefit that users would get by performing them: The cumulative effort required from non-expert users has become overwhelming and they may still not be able to protect their devices and data because of lack of time or skills, and the never-ending evolvement of the attack landscape. Our paradigm argues that because of this mismatch, home computer users should not be expected to perform these security tasks and, instead, be equipped with usable, nearly transparent (minimum and reasonable effort) recovery mechanisms. Thus, when a security incident is detected, the users can bring their machine to a functional state with most of their functionality, settings and data preserved. We systematically discuss the pros and cons of usable recovery and map the way forward to achieving low-effort usable security by means of usable recovery.

3.4 A security measurement model

Rainer Böhme (Universität Innsbruck, AT)

I reported on the main insight of Dagstuhl Seminar 14491, "Socio-Technical Security Metrics", a security measurement model. We tried to map out the simplest core, putting all limiting assumptions on the board. Using the notation of structural equations modelling, the core of the model describes how an attacker's efforts cause a loss, a relation that is influenced by how a defender's efforts result in security. Randomness comes in in several forms, chiefly through measurement error and (by assumption) attacker behaviour.

3.5 Human-Centered Security Requires Risk Communication

L. Jean Camp (Indiana University – Bloomington, US)

Safe, reliable, and secure computing requires risk-aware human behaviors. Specifically individuals must be empowered to distinguish not only between high risk and low risk choices when interacting with computers but also be able able to act of that knowledge. They need to be able to identify risk and have those risks identified in a manner that informs their behavior. The capacity of humans as security managers depends on the creation of technology that is built upon a well-founded understanding of the behavior of human users and the half-century of research on risk communication. Thus systems must not only be "trustworthy" but also must systematically communicate the risks associated with behaviors enabled by even well-designed systems. Currently, computers that are trusted are not trustworthy, but in truth there is rarely 0% or 100% certainty of harm. My work seeks to align the semantically similar but fundamentally discordant concepts of risky and trustworthy by combining computer science with risk communication. Network analysts and computer scientists can identify possible harm, but only users can determine the behavior and risk appropriate for the task at hand. Beyond usable security we seek to design to enable users to make informed decisions.

3.6   Systems Modelling and Behavioural Economics

Tristan Caulfield (University College London, GB)

Security managers in companies must design security policies that meet their organization's' security requirements. The many interactions between security policy, security controls and technology, business processes, and human behaviour make it difficult to evaluate the consequences of different choices. Systems models can create a representation of the system which can then be used to explore the effects of different policy choices. These models include human decision-making and recently we have been interested in seeing if models of decision-making from behavioural economics are applicable to security decisions and, if so, how they can be integrated into the modelling framework. We use a simple example of challenging behaviour and implement a model where decisions are based on social learning/herding.

3.7   The Security Behavior Observatory

Nicolas Christin (Carnegie Mellon University – Pittsburgh, US)

I introduced the Security Behavior Observatory (SBO). The SBO is a dedicated panel of participants who opt in to instrumentation of their home computers. This allows us to take a hybrid approach to computer security user experimentation: the control and depth of laboratory experiments performed in the user's natural environment over a prolonged period of time. The SBO allows us to perform longitudinal studies of how attacks proliferate, how users notice and respond to attacks, and the effectiveness of various security mitigations. This infrastructure allows us to combine tracking, experimental, and survey data that will result in a unique understanding of real-world user behavior that will benefit researchers in a variety of computer science disciplines. We have a working prototype of the SBO, with approximately 400 participants enrolled to date. I also reported on a couple of experiments we have been running and solicited feedback on them.

3.8   Improving Risk Decisions on Mobile Devices

Serge Egelman (ICSI – Berkeley, US)

In the beginning, Android smartphones showed users all the permissions third-party apps needed at install-time. My group performed several qualitative and quantitative studies and showed this was bad because benign requests (e.g., Internet access, requested by 90% of applications) habituated users to more dangerous requests. Things changed: both iOS and Android now prompt at runtime, the first time an app requests any of a much narrower set of "dangerous" permissions. This runtime prompting allows the user to consider context: what they were doing at the time the first request occurs. Of course, this ignores the context of

subsequent requests. As a result, we have built a classifier to manage when users are shown permission requests, thereby reducing errors by a factor of five.

We have also been exploring users' decisions to screen lock their devices. We observed bounded rationality: many users provide seemingly rational decisions for not employing security, though often underestimate the sensitivity of their data. Similarly, even users using screen locks believe it takes too long – an average of 2s or 60s/day. Thus, any new mobile authentication scheme that takes more than two seconds is a non-starter.

3.9 Value-sensitive design of Internet-based services (the commensurability challenge)

Hannes Hartenstein (KIT – Karlsruher Institut für Technologie, DE)

The main thesis of this pitch is that value-sensitive design is about the assessment of the risk that certain values are not fulfilled – not simply on costs, but on values – and, thus, can be used to assess ICT risks of socio-technical systems. There is a lot of activity in recent years in the area of 'responsible innovation' and 'value-sensitive design' (VSD). Van den Poel gives a definition and a linking of the notions of values, norm and technical requirements/designs. For our 'use case', access control, we compare VSD with OM-AM of Ravi Sandhu. The challenge is to 'meet in the middle', at the level of a norm. We (a recent paper by my group) approach this challenge and show that the approach clarifies the line between objectivity and subjectivity. Decision making under value conflicts and tradeoffs are pretty similar from a philosopher's point of view and a computer scientist. The key challenges are clarification of terms, reaching consensus on value impacts and scoring approaches ('value commensurability'). Our first steps to check whether VSD could be applied in the field of access control focusses on "who has the power/resources?" and on "self-direction" and, thus, what is known as the 'administrative model' or the identity and access governance. Our first findings show that categories can be determined and scoring looks possible. To conclude, value-sensitive design helps in linking technical design decision with the values they affect – and the risk assessment can link to those values as well, as long as the 'commensurability challenge' is something we can manage.

3.10 Modeling and Analysis of Security for Human Centric Systems

Florian Kammüller (Middlesex University – London, GB)

We propose the application of formal methods to model infrastructure, actors, and policies in order to analyse security policies. In our work, we started from invalidating global policies by a complete exploration of the state space using Modelchecking. However, to counter the state explosion problem we now moved on to Higher Order Logic using the interactive theorem prover Isabelle. The expressive power allows modeling the process of social explanation inspired by Max Weber into an Isabelle Insider Threat framework. Recent applications are to airplane safety and security, insider threats for the IoT and for auction

protocols. The CHIST-ERA project SUCCESS plans to use the framework in combination with attack trees, and the Behaviour Interaction Priority (BIP) component architecture model to develop security and privacy enhanced IoT solutions. A pilot from the health care sector, cost-effective IoT-based bio-marker monitoring for early Alzheimer's diagnosis, will investigate the feasibility of the approach. Critical needs as an input for this process ICT Risk assessment methods to understand what security and privacy questions we need to answer to the stakeholders including patients, nurses and doctors.

3.11 Socio-Technical Q&A

Stewart Kowalski (Norwegian University of Science & Technology – Gjøvik, NO)

<div align="center">

To be or not to be secure
enough.
That Is the question
of our abilities
and of our courage
to ask
the challenging
socio-technical questions.
So
that we can
Thinking, feeling, know
and act
to
Deter our enemies
Protect our friends
Respond to attacks and crisis's.
And
If necessary
recover through questioning
not
blaming
all stakeholders.

</div>

3.12 Attack Surface of Socio-Technical Systems

Kwok-Yan Lam (Nanyang TU – Singapore, SG)

Researchers in technical disciplines are familiar with concepts such as risks, attack surface, attack vectors, exposure, threats and vulnerabilities. In this talk, we first look at the key security pillars behind the attack surface of a distributed system. Security protection behind the attack surface is typically enforced by a combination of physical security, technical

mechanisms, and organizational policies in order to achieve the security objectives in an efficient and effective manner. The risk and security controls employed by a system will depend on the nature of the system such as Government (civilian vs national security), Enterprise (Critical Information Infrastructure vs non-CII) or Consumer application systems (e.g. free Internet/cloud application services).

Security has always been taking into consideration people's aspects, though have been mostly piecemeal, ad hoc, implicit and over-simplified. For example, in the study of practical security, we choose key length and password length that is manageable by people, cover time of crypto algorithm before sensitivity of data does not matter anymore, classification of data according to their impact on real world.

In real world situations, security requirements/objectives always start from people e.g. defining business objectives of a system by the business owner. When studying the cybersecurity of socio-technical systems, it is useful to identify the steps and processes where people are involved in the security consideration of the system development lifecycle. The people/social factors are typically one of the major source of vulnerabilities, and vary from system to system. In reality, it is not uncommon that the same person may view security differently when involved in different type of systems. Conversely, different people involved in the same system may have different views on security. It is therefore important to identify the various stakeholders of a system, and study the security concerns of these stakeholders. In a typical real world scenario, we usually can identify the following stakeholders:

- System owners: their security concerns are mainly related to reputation and compliance with laws and regulations.
- System sponsors: their main security concerns include operational resilience (robustness and availability) and efficiency.
- System designer/developers: their security interests are typically the traditional information security issues.
- System administrators/operators: their main security interests include implementation audit and control procedures, enforcement of the corresponding policies, and sometimes involve in the compliance of physical security,
- End users: their security interests are mainly stemmed from their concerns for their accountability and liability.

Security risks lead to uncertainty to the liabilities of the people. Socio-technical systems take into consideration the social/people factors in the risk assessment of information systems. Ultimately, people are most concerned about liabilities though technical people tend to look at the security issues as risks. For example, in an Internet banking system, most users are mainly concerned about their liabilities should fraudulent transactions be detected. System owners and business sponsors of a system tend to consider liabilities while system designers/developers tend to look at risks. It'll be useful to establish a framework and model to understand the relationship between risk and liability.

The Information Systems community developed the Enterprise Architecture as a tool for establishing a framework and methodology for enabling/facilitating communications among all stakeholders so as to make sure the business/people requirements are taken into consideration in the entire system development lifecycle. It'll be useful for the security research community to explore the use of the EA as a basis for developing risk assessment models for socio-technical systems.

Another area that is important to look into is the massive deployment of IoT devices in new generation smart city/nation systems. The attack surface and risk model is more complicated. As a kind of socio-technical system, IoT-based applications are very different

from existing socio-technical systems because a lot of IoT devices are operated in open environment, hence a lot of physical security and organization processes are not enforceable. It deserves much attention of the cybersecurity research community in that it needs new approach to address cybersecurity of IoT-based application systems. Key socio-technical problems include:

- Identify the applicable laws and regulations for IoT applications.
- Study the social aspects of such systems such as the role of people in the security considerations of the development lifecycle of such systems.
- Investigate security requirements for IoT-based socio-technical systems.
- Develop models for analyzing security requirements and risks of such systems.
- Develop architectural frameworks for devising security protection and control measures for such systems.

3.13   Cyber Risk Information Sharing

Stefan Laube (Universität Münster, DE)

Cyber risk management in modern days reduces to a race for information between defenders and attackers. This is why cyber risk information sharing is recognized by scholars and politicians alike as a measure for defenders to gain an edge over attackers. However, rational defenders' sharing incentives are guided by selfish reasons, moderated by the effects of cyber risk information sharing. In my talk, I introduce how to rigorously survey these effects. Specifically, I consider information sharing as an act of communication, that can be analyzed by Lasswell's model of communication: "Who says what to whom in which channel with what effect?".

3.14   Heavy-tailed cyber security incidents: Which organization design to manage extreme events?

Thomas Maillart (University of Geneva, CH)

There is broad evidence that cyber risks have extreme "heavy-tail" components, with quantitative evidence for personal data breaches, browser update delays, and cyber security incidents. With co-authors, analyzing the statistical properties of sixty thousand security events collected over six years at a large organization, we found that the distribution of costs induced by security incidents is in general highly skewed, following a power law tail distribution. However, this distribution of incident severity becomes less skewed over time, suggesting that the organization under scrutiny has managed to reduce the impact of large events. We illustrate this result with a case study focused on the empirical effects of full disk encryption on the severity of incidents involving lost or stolen devices. However, quantifying extreme risks at the aggregate level may not necessary provide insights at the fine-grained level. In this talk, I have therefore introduced a simple fine-grained cascading model, inspired from the St. Petersburg Paradox, which calibrates to all sorts of heavy-tailed distributions

(e.g., power-law, stretched exponential). This model may be used to statistically map real attack processes, involving overcoming layers of defenses, with incremental lower probability and higher potential damage.

3.15 How do you know that it works? The curses of empirical security analysis

Fabio Massacci (University of Trento, IT)

In this talk I discuss the difficulties behind empirical security analysis by illustrating some particular pitfalls through bad papers.

1. Partial data: some dataset might no contain the same type of information for different entities in the very same DB field and as a result we might infer completely wrong information (typical example is using NVD for comparing ProS vs FOSS). Some example bad papers are
   - Muhammad Shahzad et al., "A Large Scale Exploratory Analysis of Software Vulnerability Life Cycles," Proc. of ICSE'13.
   - R. Sen, G. R. Heim, "Managing Enterprise Risks of Technological Systems: An Exploratory Empirical Analysis of Vulnerability Characteristics as Drivers of Exploit Publication," Decision Sciences, 2016.
2. Too big to fail: several statistical tests were invented to deal with small samples. Statistical significance ($p < 0.05$) comes therefore for free with large audit trails. If we confuse it with practical significance we obtain a wrong conclusion as the paper below:
   - K. Seung Hyun and K. Bung Cho, "Differential Effects of Prior Experience on the Malware Resolution Process," MIS Quarterly 2014.
3. Who said it works: this is an ubiquitous problem in CS. We test our own system. While this is acceptable (at the beginning of one's research) to knwo whether a technology really works we need to put it into the ends of its final users.
4. How do you know? We often decide which are the important metrics based on industry suggestions. This can be structured with proper qualitative studies.

References

1 L. Allodi, M. Corradin, F. Massacci. Then and Now: On The Maturity of the Cybercrime Markets. The lesson black-hat marketeers learned. IEEE Transactions on Emerging Topics in Computing. 4(1):35–46, 2016.
2 L. Allodi, V. Kotov, F. Massacci. MalwareLab: Experimenting with Cybercrime Attack Tools. In: Proc. of Usenix Security CSET 2013, Washington D.C., USA.
3 V. Kotov and F. Massacci. Anatomy of Exploit Kits: Preliminary Analysis of Exploit Kits as Software Artefacts. Proc. of ESSoS 2013, pp. 181–196.
4 F. Massacci, F. Paci, L. M. S. Tran, A. Tedeschi: Assessing a requirements evolution approach: Empirical studies in the air traffic management domain. Journal of Systems and Software 95:70–88 (2014).

5   L. Allodi, F. Massacci. The Work-Averse Attacker Model. In the Proceedings of the 23rd European Conference on Information Systems (2015).

6   V. H. Nguyen, S. Dashevskyi, and F. Massacci. An Automatic Method for Assessing the Versions Affected by a Vulnerability, Empirical Software Engineering Journal. (2015). Short version in V.H.Nguyen and F.Massacci. The (Un)Reliability of Vulnerable Version Data of NVD: an Empirical Experiment on Chrome Vulnerabilities. In: Proceeding of the 8th ACM Symposium on Information, Computer and Communications Security (ASIACCS)'13, May 7-10, 2013.

7   L. Allodi, F. Massacci. Comparing vulnerability severity and exploits using case-control studies. In ACM Transactions on Information and System Security (TISSEC). Short version in L. Allodi, F. Massacci. How CVSS is DOSsing your patching policy (and wasting your money). Presentation at BlackHat USA 2013, Las Vegas, USA.

8   F. Massacci, V. H. Nguyen. An Empirical Methodology to Evaluate Vulnerability Discovery Models. In IEEE Transactions on Software Engineering (TSE), 40(12):1147–1162, 2014.

9   Woohyun Shim, Luca Allodi, Fabio Massacci. Crime Pays If You Are Just an Average Hacker. Proceedings of IEEE/ASE 2012 Cyber Security Conference. Complementary publication in ASE Journal 2012, Vol. 2.

10   Massacci F., and Paci F. How to Select a Security Requirements Method? A comparative study with students and practitioners. In Proceedings of the 17th Nordic Conference in Secure IT Systems (NordSec), 2012.

11   K. Labunets, F. Massacci, F. Paci, S. Marczak, F. Moreira de Oliveira. Model Comprehension for Security Risk Assessment: An Empirical Comparison of Tabular vs. Graphical Representations To appear in Empirical Software Engineering.

12   K. Labunets, F. Massacci and F. Paci. On the Equivalence Between Graphical and Tabular Representations for Security Risk Assessment. To appear in Proceedings of REFSQ'17.

13   K. Labunets, F. Paci, F. Massacci. Which Security Catalogue Is Better for Novices? In Proc. of EmpiRE Workshop at IEEE RE'15.

14   M. de Gramatica, K. Labunets, F. Massacci, F. Paci, A. Tedeschi. The Role of Catalogues of Threats and Security Controls in Security Risk Assessment: An Empirical Study with ATM Professionals In the Proceedings of REFSQ 2015.

15   K. Labunets, F. Paci, F. Massacci, and R. Ruprai. An Experiment on Comparing Textual vs. Visual Industrial Methods for Security Risk Assessment. In Proc. of EmpiRE Workshop at IEEE RE'14.

16   Labunets, K., Massacci, F., Paci, F., and Tran, L. M. S. An experimental comparison of two risk-based security methods. In Proceedings of the 7th ACM International Symposium on Empirical Software Engineering and Measurement (ESEM), 163–172, 2013.

3.16   Enhancing Researches in the ICT Security Field

Kanta Matsuura (University of Tokyo, JP)

Let us explore approaches for enhancing researches in this field.

Establishing nice theories – Yes, they would have great impacts. They can help empirical studies, new problem formulation, and so on. Young researchers can be inspired.

Providing datasets – Yes, many of us wish to have good data for research purposes. Considering the scientific requirement such as reproducibility of the results, common datasets would be helpful. However, there is an intrinsic limitation: as long as the data is observed

before the defense method under research is published, the attack included in the dataset assumes that the attacker does not know the defense method.

Having competitions – Yes, the use of a common data on-site at a competition event associated with a conference can have educational effects as well.

Finally, testbeds – Yes, this may solve the problems inherent to the dataset approach to some extent. However, this approach can be very expensive (and hence quite hard to do in reality) if you try to do everything by yourself. In my poster and short pitch talk, I introduce a new venture of testbed: bsafe.network. Bsafe is a test network for the research of blockchain technologies and their applications. Although there are a lot of challenges, we consider the use of the lessons from the Internet. Details are described at http://bsafe.network, particularly in the white paper there. A wide variety of researches may be designed in the network, ranging from distributed ledger protocols to socio-technical applications.

3.17 Security Challenges in Small Organisations

Simon Parkin (University College London, GB)

Here I provide a brief overview of a programme of user-centred security research involving micro- and small-sized organisations. Formal security training and in-house expertise may be limited or non-existent in these smaller organisations. Security vendors and providers of IT services and leased premises can play a role in the security posture of these organisations. The work involves collaboration with outsourced IT providers (and their clients) and other groups of organisations, such as charity associations. Small organisations can be diverse in their use of IT and daily interactions, where we are developing a range of SME archetypes as representative examples. I also discuss the Available Responsibility Budget (ARB) of small organisations, as a representation of organisation members' collective capacity to manage a range of security controls, as well as high-level findings of initial interviews with representatives of small organisations.

3.18 Agent-based modelling for cybersecurity

Wolter Pieters (TU Delft, NL)

Based on the expertise in our faculty, we are interested in developing further applications of agent-based models to study aspects of the security ecosystem, in particular in the critical infrastructure domain. In agent-based modelling, effects at system level (such as number of attacks) are studied based on modelling and simulating the behaviour of individual agents involved (such as attackers, defenders). This enables defenders to make better decisions, taking the expected responses of the other actors into account.

We have a successful proof-of-concept representing the vulnerability ecosystem, including behavioural models of attackers, discoverers, software vendors and software users, providing suggestions for the best patching strategies for vendors and users. We are also developing methods for investigating how these different actors make choices. We are interested in getting in touch with problem owners in the critical infrastructure domain who face security decision problems in multi-actor situations. By extending our models to new case studies, we can support security decisions as well as advance our research.

References
1   Breukers, Y. P. (2016). The Vulnerability Ecosystem: Exploring vulnerability discovery and the resulting cyberattacks through agent-based modelling. Master's thesis, TU Delft.
2   Meeuwisse, K. V. M. (2016). The usability-security trade-off: Exploring employees' perceptions and preferences for techncial security measures using choice modelling. Master's thesis, TU Delft.
3   Slangen, R. (2016). Understanding Cyber-risk by Investigating the Behaviour of Defender and Threat Agent Organisations: Why a Complex Adaptive Systems Perspective Contributes to Further Understanding Cyber-risk. Master's thesis, TU Delft.
4   van Wieren, M., Doerr, C., Jacobs, V., and Pieters, W. (2016). Understanding Bifurcation of Slow Versus Fast Cyber-Attackers. In: International Workshop on Data Privacy Management (pp. 19–33). Springer International Publishing.

3.19 Metrics for security-related behaviours and organisational economics and performance

Martina Angela Sasse (University College London, GB)

In this talk, I reflected on metrics we have used over the past 20 years when studying security-related behaviours and organisational economics and performance. In the first study in 1996, realised it the importance of understanding the mental and physical workload [1] that security policies create for users, and the impact it has on their performance on the primary task. Users have an acute sense of the amount of time that is diverted from their primary productive activity, and when they have reached that compliance budget, they will look for ways of diverting it [2]. We subsequently proposed that security managers must measure the effort employees spend on security, and work with business decision-makers to balance that effort with the risk mitigation achieved [3]. In later studies, we found that not only the time spent on the security task counts, but that disruption of the primary task means up to 40% of that effort has to be re-done [4]. Users' behaviour is also driven by their degree of risk understanding, and users' emotional stance towards the organisation [5]. More recent studies have shown that the degree of communication and collaboration between security specialists and other members of the organisation is a good indicator for how effective security is [6], and finally the degree of trust the organisation places in their staff [7].

References
1   Adams, A., Sasse, M. A. (1999). Users are not the enemy. Communications of the ACM, 42 (12):40–46.

2   Beautement, A., Sasse, M. A., Wonham, M. (2008). The compliance budget: managing security behaviour in organisations. Proceedings of New Security Paradigms Workshop.

3   Parkin, S., van Moorsel, A., Inglesant, P., Sasse, M. A. (2010). A stealth approach to usable security: helping IT security managers to identify workable security solutions. Proceedings of New Security Paradigms Workshop.

4   Steves, M., Chisnell, D., Sasse, A., Krol, K., Theofanos, M., Wald, H. (2014). Report: Authentication Diary Study. National Institute of Standards and Technology. http://dx.doi.org/10.6028/NIST.IR.7983

5   Beris, O., Beautement, A., Sasse, M. A. (2015). Employee Rule Breakers, Excuse Makers and Security Champions:: Mapping the Risk Perceptions and Emotions That Drive Security Behaviors. Procs NSPW.

6   Kirlappos, I., Parkin, S., Sasse, M. A. (2015). "Shadow security" as a tool for the learning organization. ACM SIGCAS Computers and Society, 45 (1):29–37.

7   Kirlappos, I., Sasse, M. A. (2015). Fixing Security Together: Leveraging trust relationships to improve security in organizations. Proceedings USEC.

3.20   Socio-Technical-Artefact Driven Security and Risk Analysis

Christian Sillaber (Universität Innsbruck, AT)

Joint work of Ruth Breu

We propose a structured quality assessment process to utilize models of a system and related artefacts created by stakeholders (system, business process, activity diagrams, etc.) to improve the identification of potential security risks. While research considers users as the most important resource in security and risk management processes, little attention is given to various artefacts created by them during the creation and operationalization of complex systems. The proposed methodology can unveil inherent misunderstandings and miscommunications between the stakeholders, by comparing the produced models against each other, that could have introduced security issues during the development of the system. We conducted several field studies to evaluate the viability of the approach using a variety of different modelling languages and frameworks.

References
1   Sillaber, C., and Breu, R. (2015). Identifying Blind Spots in IS Security Risk Management Processes Using Qualitative Model Analysis. In T. Tryfonas & I. Askoxylakis (Eds.), Human Aspects of Information Security, Privacy, and Trust (Vol. 19, pp. 252–259).

2   Sillaber, C., and Breu, R. (2015). Using Stakeholder Knowledge for Data Quality Assessment in IS Security Risk Management Processes. In Proceedings of the 2015 ACM SIGMIS Conference on Computers and People Research (pp. 153–159).

3   Sillaber, C., and Breu, R. (2015). Using Business Process Model Awareness to improve Stakeholder Participation in Information Systems Security Risk Management Processes. In Wirtschaftsinformatik 2015 Proceedings.

3.21   Assessing ICT security risks in socio-technical systems: Policy implications

Edgar A. Whitley (London School of Economics, GB)

My presentation had two components. It focussed on the socio-technical aspect of the workshop title more than the assessment of risk. It began by considering the expressions of socio-technical as presented in the "about me" posters produced by workshop participants. This was supplemented by oral descriptions of "socio-technical" in my working group. This indicated that the notion of socio-technical is very broadly understood and has many inconsistent interpretations. This can be significant because the different ways in which we conceptualise technology and its relationship to the social can have very different effects in terms of research design and real world policy implications.

Academically, the concept of socio-technical has a long, and specific, history, from its earliest uses in relation to the work of the Tavistock institute. More recently, there has been increased theorisation of how the social and technological might relate to each other, going beyond a simplistic "everything that is not the technological".

One particular way of considering the relationship between the social and the technological (or facts versus values) is presented by Bruno Latour, particularly his book on the politics of nature. I reviewed this model and highlighted the ways in which it could help clarify the different roles of technologists: working through perplexities (i.e. areas where there is no agreement on what is "true" (e.g. the best ways to assess ICT security risk") and institution (i.e. areas where there is agreement on what is true (or what is no longer suitable / acceptable as a practice in the field)).

3.22   Human Factors in Information Security

Sven Übelacker (TU Hamburg-Harburg, DE)

The working title of my PhD thesis covers the ample intertwined area of human interaction with information and communication systems regarding security. I focus on the targeted side of socio-technical attacks: what kind of enablers exist; why are some people (more) susceptible to social engineering attacks (than others). By doing so, I am facing two main challenges:

First of all, evidence of social engineering attacks needs to be discovered, analysed and understand. However, successful real-life attacks are rarely documented reliably and in detail. How can we draw generalisable conclusions for the real world? Experiments can hopefully shed a bit of light on a specific sample size in a small time frame, but can be as realistic

as they can be intrusive, ethically questionable or counterproductive for a security culture in the long run. If we look at documented cases, how reliable can we determine the source is reporting honestly? We are living in click-bait times of online "news"; hear-saying and interviews – of victims or (former) criminals – on real events can not be re-evaluated by fellow researchers and can only be qualitative. Law enforcement agencies publish their statistics, but most of the time their databases with more detailed information remain confidential and re-calculating or re-categorising their data is almost impossible. Consulting companies publish their reports and surveys with no public access to the data and may have their own interests like to acquire new customers. Expert opinion and guesstimates on possible attack scenarios can help to understand the knowledge of novel attack vectors, but how realistic and feasible are they without proper real events or experimentalisation? Finding reliable, useful and detailed evidence for holistic analyses should still be a valuable prerequisite before designing detection and mitigation strategies of attacks targeting humans. I am confident a good combination of various approaches and sources can lead to a better understanding, i.e., conducting experiments based on real events, re-conducting questionnaires and experiments internationally, or mapping incidents to expert opinion.

Secondly and complementing the anterior paragraph, it is crucial to understand the susceptibility of targeted persons. Many disciplines present ideas about human behaviour, biases, habits, irrationality, heuristic systems, usability, persuasion, personality, gullibility, risk perception/assessment, impulsiveness, stressors or cultural background to name a few. Insights can be very helpful for adapting the work environment and its culture, improving security mechanisms with respect to usability, organisational policies, interventions and trainings. Interdisciplinary research and a combination of the mentioned factors may leverage results.

4 Working groups

4.1 Collecting Data for the Security Investment Model

Tristan Caulfield (University College London, GB), Rainer Böhme (Universität Innsbruck, AT), Kanta Matsuura (University of Tokyo, JP), Tyler W. Moore (University of Tulsa, US), Martina Angela Sasse (University College London, GB), Michel van Eeten (TU Delft, NL), and Maarten van Wieren (Deloitte – Amsterdam, NL)

We considered a security investment model, proposed by Rainer Böhme, to estimate from data the relationships between efforts to secure an organization, the level of security actually achieved, and how well this level of security moderates efforts to attack the organization and the losses experienced. This model can be applied to many different types of organizations or systems to understand how investments in different security efforts affect security and prevent incidents and losses. While we briefly discussed examples such as hosting providers and bitcoin exchanges, we mainly explored how it could be applied to typical organizations, such as small-to-medium enterprises (SMEs).

We compiled a list of different indicators that could be measured for each component of the model. For indicators of effort towards security, the list includes organizational efforts, such as how much budget or staff-hours are allocated to security, as well as formal controls,

such as the existence of policy documents and continuity plans, and informal controls, such as whether employees share passwords, use of encryption, and security awareness training. For the level of security, the indicators must be independent of the attacker; it would be incorrect, for example, to use the number of machines infected with malware as an indicator here. Indicators for security include patch level and patch speed, results of brute force password cracking, penetration testing results, indicators of network and system hygiene, such as open mail relays and unused but running VM instances or servers, as well as human measures, including the Behavioural Security Grid to gauge employees' attitudes and behaviours towards security.

For attacker effort, possible measurable indicators could include network attack indicators, such as IDS events, ports scans, and malware encounters, as well as social engineering attempts, and types and frequencies of different attacks. The value for attacker effort could also be held constant across different organizations, if this information is too difficult to gather. Losses represent the impact and consequential effects of attacks. Indicators discussed include the number of incidents of ransomware and phishing, incidents of data theft, PCI incidents, DDoS attacks that interrupt business, theft of information or equipment, and incidents of different types of fraud. It is also possible to use various types of monetary loss as indicators, such as fines, investigation or recovery costs, lost revenue, and productivity losses resulting from an incident.

With data from a large number of companies, this model can be used to explore the effects of investments in different security efforts on organizations' security. This can then be used to help guide security investment by, for example, determining if a particular security control has an impact on the incidents or losses for organizations of a given size, or by finding the optimal investment in different security efforts for a given budget. Finally, we discussed what steps could be taken to collect this data, with a focus on working with IT service providers to gather data about and distribute surveys to their SME clients.

4.2   New frontiers of socio-technical security

Carlos H. Ganan (TU Delft, NL), Zinaida Benenson (Universität Erlangen-Nürnberg, DE), Hannes Hartenstein (KIT – Karlsruher Institut für Technologie, DE), Stewart Kowalski (Norwegian University of Science & Technology – Gjøvik, NO), Kwok-Yan Lam (Nanyang TU – Singapore, SG), Christian W. Probst (Technical University of Denmark – Lyngby, DK), Edgar A. Whitley (London School of Economics, GB), and Jeff Yan (Lancaster University, GB)

Socio-technical security paradigms continue emerging and new frontiers have to be identified to address these paradigms. New approaches and methods are required to investigate these frontiers where the human factor becomes a critical component of these socio-technical systems (STS). Analyzing these systems using multi-actor theories shed light to these systems and provides an essential tool to investigate new frontiers. The Internet of Things constitutes an example of these complex STS where new frontiers have to be defined. One of the main security challenges of such systems is that a lot of existing security architecture/models are not applicable to such scenarios. Another challenge is that a lot of IoT devices are developed and released to the market, a lot of organizations are making big promises for the

pervasive use of IoT devices in smart city/nation applications, but there is very little prior experience/knowledge for designing and analyzing the security of such systems. New solutions have to be envisioned to take over the security of IoT devices which become "orphaned" either because they are abandoned by their manufacturer or because the manufacturer goes out of business. Software escrow and graceful degradation are potential mechanisms that will help to mitigate the impact of ophaned IoT devices. The implementation and coexistence of these mechanisms with IoT devices already in the market arises new challenges that need to be addressed from a socio-technical perspective.

4.3 Unpacking the motivations of attackers

Thomas Maillart (University of Geneva, CH), Ross Anderson (University of Cambridge, GB), Johannes M. Bauer (Michigan State University – East Lansing, US), Barbara Kordy (IRISA – Rennes, FR), Gabriele Lenzini (University of Luxembourg, LU), and Sebastian Pape (Goethe-Universität Frankfurt am Main, DE)

The Internet was once thought to remain a conflict free space. Following technical innovation and its progressive integration in the social fabric, attacks and conflicts have appeared early on. They have since then steadily developed in sophistication, diversity and span. Analyzing the strategic motivations and modus operandi of both landmark and outstanding cyber-attacks and cyber-conflicts, we have found that they are aimed at disrupting the socio-technical system through the social sub-system (e.g., social engineering, deception, blackmail, mass media manipulation) through the technical sub-system (e.g., exploiting a vulnerability, manipulating social media), or through a combination of both.

Disruption may occur at various levels and time-horizons, which can be partitioned as proposed by John Groenewegen:

- level 0: resource allocation (i.e., continuous time operations),
- level 1: governance (1 to 10 years),
- level 2: institutional environment (10 to 100 years),
- level 3: embeddedness (100 to 1000 years).

Cyber attacks are aimed at gaining technical, social, economic, political or cultural power. On the contrary, they may be aimed at ensuring the status quo. We provide a few examples to give flesh:

- Crypto War 1 was the NSA trying to defend its level 2 institutional business model against disruptive changes at levels 0 and 1. The debate over DRM was the exact same, more or less; Hollywood and the music industry wanted to defend their business model against disruptive change and set out to bully the tech industry to redesign the world. The tech industry pretended to do so but ate their lunch.
- Crime Policing: The failure of policing as crime goes online may reflect a failure of the police to create the needed international institutions for fast cheap MLAT, but on the surface is much like crypto wars or DRM. As for the cybercriminals there's now a new institution, namely the international network of cybercrime operators and their suppliers, who appear to collude with some governments.

- Early hacktivism (such as the cypherpunks) was an attempt to use action at level 0 (writing software) to change ultimately all higher levels, by creating an ungovernable space in which people could transact with digital cash over anonymous remailers regardless of the wishes of established governments. It was perhaps more an attack on level 2, trying to establish Californian social norms worldwide at level 3 – or at least trying to make available to all a space in which these norms were the rule.
- Cyberbullying is mostly operational; it can be worse than real-world bullying as it doesn't stop when you get home, but better in that the teacher has a trail of who did it (provided she can get US-based providers to cooperate, which is level 1 intrusion).

Unpacking the motivations of attackers in the cyberspace requires to recognize the increasing integration of the socio-technical system. In this integrated system, the (strategic) purpose and (tactical) means play a fundamental role in the way and at which level(s) the cyberspace can be disrupted (resp. the status quo preserved). As a result, cyber attacks may have implications at corresponding time horizons, ranging from the disruption of immediate operations to the disruption of governance, institutions, and even long-established cultural norms. Our approach shall provide a sense of importance of (intent of) attacks in the cyber space-time, and hence help set more insightful and efficient agenda in order to prevent or mitigate attacks, or on the contrary, to promote such attempts to disruption if perceived as a potentially positive outcome.

4.4   Software Liability Regimes

Christian Sillaber (Universität Innsbruck, AT), Nicolas Christin (Carnegie Mellon University – Pittsburgh, US), Richard Clayton (University of Cambridge, GB), Stefan Laube (Universität Münster, DE), Fabio Massacci (University of Trento, IT), Tyler W. Moore (University of Tulsa, US), Kai Rannenberg (Goethe-Universität Frankfurt am Main, DE), and Michel van Eeten (TU Delft, NL)

New software based services are currently not covered through product liability regimes as software that is sold through licenses is not considered a product. This introduced a market where the parties that are in a position to fix problems (i.e. software developers and service providers) are not assigned any liability for resulting damages. This increases social costs and does not incentivise companies to invest in quality and security improvements for their software and services.

To address this imbalance, we call for the introduction of a software liability regime to cover negligent software development practice. To establish whether a software developer has been negligent or not, standards are required (to prove a violation).

A first logical step would be the introduction of mandatory certifications and standards for commercial software. However, standards are quickly out-of-date and the software industry has no incentives to buy in. Therefore, software security standards may be enforced by law. These should focus on the processes to be followed during implementations, as it is easier to certify development processes and the usage of state-of-the-art components that software properties. A baseline for such a standards should define testing processes (e.g. buffer overflow testing). The certifying body should not only audit that the software development processes

follow best-practice but also perform (randomized) sampling of the actual functioning of used software tests.

To steer public discussion into an appropriate direction, it would make sense to differentiate between commodities (e.g. household appliances) that are sold to and used by uneducated users and professional appliances. Software running on such devices should either be fully managed or certified by a licensed professional to be "safe for uneducated users" without them needing to do anything. If the manufacturer of commoditized software chooses an upgrade strategy over the Internet, then any upgrade should be safe for the uneducated user. Upgrades developed for such devices should be subject to the same rules as spare parts for physical products.Professional appliances should be explicitly labelled as such and there should be a process to determine whether a user has indeed the required professional competence (e.g. licence, expertise) to handle and configure a professional device.

In summary, policy makers should get active and introduce appropriate legislation.

The following questions have been raised during the working group discussions and will be investigated further:

- How can we test for negligence as an indicator for liability?
- What does strict liability and negligence mean?
- How can we attribute the faulty code? Forensics? Proofs?
- How is a customer able to provide enough probable cause?
- Is it necessary to start a discussion on mandatory certification of commercial software?
- What about software that becomes vulnerable over time (e.g. RC4/MD5)? How long are re-sellers/developers required to provide security updates?
- Which accountability functionalities for software should be required by law?
- What should the lawmaker do about software whose functionality is not 100% understandable? E.g. AI gone wrong.
- How can poor development be proven by experts? E.g. openssl, easy ex-post, but difficult ex-ante.
- How to handle installation v.s. maintenance?
- What is the safety and security level that can be expected from a device?
- What about free software developers? Would a too strict liability regime stifle innovation?
- How should a universal computer be treated differently from a (closed) consumer device?
- How to handle negligence of a consumer device was jailbroken?
- How should maintenance be handled? Any device should be updatable by a layman/uneducated user?
- What is the relationship between insurance and certification?

4.5 User-centred Security Models

Sven Übelacker (TU Hamburg-Harburg, DE), Vincent Koenig (University of Luxembourg, LU), and Simon Parkin (University College London, GB)

The working group began with discussions about how cognitive biases leave users behind in some security tasks, and that we will suffer to achieve an acceptable level of security. This then led to identification of factors – contextual, psychological, social – which influence security-relevant choices. A range of approaches were discussed by the working group:

- Identifying stakeholders and mapping them to decisions about security infrastructure. Are all stakeholders represented adequately in security-related decisions?
- When designing processes, security is not an add-on. 100% security may not be achievable, and security is a process. Is it feasible and desirable to bound security responsibilities, e.g., by being responsible for others? Can we define a minimum level of security effort for users?
- How much security is "good enough"? We have no universal definition of "good" security behaviour, and we postulate to stay away from normative approaches where the user is forced to comply. This also points to the issue of ethics in security that is "good enough". The group discussed this and also compared the problematic to that of the "Collingridge dilemma".
- Whether there is any security advice which will remain stable over time, for instance when teaching children about security. This affects the lifetime of models.

Overall, many researchers are already investigating these areas, and yet barriers remain which may be better understood through development of user-centred models.

Participants

Ross Anderson
University of Cambridge, GB

Johannes M. Bauer
Michigan State University –
East Lansing, US

Zinaida Benenson
Universität Erlangen-
Nürnberg, DE

Rainer Böhme
Universität Innsbruck, AT

L. Jean Camp
Indiana University –
Bloomington, US

Tristan Caulfield
University College London, GB

Nicolas Christin
Carnegie Mellon University –
Pittsburgh, US

Richard Clayton
University of Cambridge, GB

Serge Egelman
ICSI – Berkeley, US

Carlos H. Ganan
TU Delft, NL

Dieter Gollmann
TU Hamburg-Harburg, DE

Hannes Hartenstein
KIT – Karlsruher Institut für
Technologie, DE

Florian Kammüller
Middlesex University –
London, GB

Vincent Koenig
University of Luxembourg, LU

Barbara Kordy
IRISA – Rennes, FR

Stewart Kowalski
Norwegian Univ. of Science &
Technology – Gjøvik, NO

Kwok-Yan Lam
Nanyang TU – Singapore, SG

Stefan Laube
Universität Münster, DE

Gabriele Lenzini
University of Luxembourg, LU

Thomas Maillart
University of Geneva, CH

Fabio Massacci
University of Trento, IT

Kanta Matsuura
University of Tokyo, JP

Tyler W. Moore
University of Tulsa, US

Sebastian Pape
Goethe-Universität
Frankfurt am Main, DE

Simon Parkin
University College London, GB

Wolter Pieters
TU Delft, NL

Christian W. Probst
Technical University of Denmark
– Lyngby, DK

Kai Rannenberg
Goethe-Universität
Frankfurt am Main, DE

Martina Angela Sasse
University College London, GB

Christian Sillaber
Universität Innsbruck, AT

Sven Übelacker
TU Hamburg-Harburg, DE

Michel van Eeten
TU Delft, NL

Maarten van Wieren
Deloitte – Amsterdam, NL

Melanie Volkamer
Karlstad University, SE

Edgar A. Whitley
London School of Economics, GB

Jeff Yan
Lancaster University, GB

Inpainting-Based Image Compression

Edited by

Christine Guillemot[1], Gerlind Plonka-Hoch[2], Thomas Pock[3], and Joachim Weickert[4]

1   **INRIA – Rennes, FR,** `christine.guillemot@inria.fr`
2   **Universität Göttingen, DE,** `plonka@math.uni-goettingen.de`
3   **TU Graz, AT,** `pock@icg.tugraz.at`
4   **Universität des Saarlandes, DE,** `weickert@mia.uni-saarland.de`

―― **Abstract** ――――――――――――――――――――――――――――――――――――――

Inpainting-based image compression is an emerging paradigm for compressing visual data in a completely different way than popular transform-based methods such as JPEG. The underlying idea sounds very simple: One stores only a small, carefully selected subset of the data, which results in a substantial reduction of the file size. In the decoding phase, one interpolates the missing data by means of a suitable inpainting process. It propagates information from the known data into the areas where nothing has been stored, e.g. by solving a partial differential equation or by clever copy-and-paste mechanisms.

Inpainting-based codecs (coders and decoders) are more intuitive than transform-based ones, they are closer to biological mechanisms in our brain, and first results show that they may offer promising performance for high compression rates. However, before these ideas become practically viable, a number of difficult fundamental problems must be solved first. They involve e.g. the selection of the data and the inpainting operator, coding strategies, and the search for highly efficient numerical algorithms. This requires a collaborative effort of experts in data compression, inpainting, optimisation, approximation theory, numerical algorithms, and biological vision. In this Dagstuhl seminar we have brought together leading researcher from all these fields for the first time. It enabled a very fruitful and inspiring interaction which will form the basis for future progress.

1   Executive Summary

Joachim Weickert

Since the amount of visual data is rapidly increasing, there is a high demand for powerful methods for compressing digital images. A well-known example is the lossy JPEG standard that is based on the discrete cosine transform. Unfortunately its quality deteriorates substantially for high compression rates, such that better alternatives are needed.

The goal of this seminar was to pursue a completely different strategy than traditional, transform-based codecs (coders and decoders): We studied approaches that rely on so-called inpainting methods. They store only a small, carefully selected subset of the image data. In the decoding phase, the missing data is reconstructed by interpolation with partial differential equations (PDEs) or by copying information from patches in other image regions. Such codecs allow a very intuitive interpretation, and first experiments show their advantages for high compression rates where they can beat even advanced transform-based methods.

However, inpainting-based codecs are still in an early stage and require to solve a number of challenging fundamental problems, in particular:

1. Which data gives the best reconstruction?
2. What are the optimal inpainting operators?
3. How should the selected data be encoded and decoded?
4. What are the most efficient algorithms for real-time applications?

These problems are highly interrelated. Moreover, they require interdisciplinary expertise from various fields such as image inpainting, data compression and coding, approximation theory, and optimisation. To design these codecs in an optimal way, one must also understand their connections to related areas such as sparsity and compressed sensing, harmonic analysis, scattered data approximation with radial basis functions, and subdivision strategies.

Our seminar constituted the first symposium on this topic. It brought together 29 researchers from 11 countries, covering a broad range of expertise in the different fields mentioned above. Many of them have met for the first time, which resulted in a very fruitful interaction.

In order to have a good basis for joint discussions, first all participants introduced themselves and briefly described their background and interests. Then the seminar proceeded with six tutorial talks (45 minutes plus 15 minutes discussion), given by the four organisers as well as by Simon Masnou and Nira Dyn. In this way all participants could acquire a general overview on the achievements and challenges of inpainting-based image compression and its various aspects such as coding, inpainting, convex optimisation, subdivision, and computational harmonic analysis.

Afterwards we decided to cluster the talks thematically into six sessions, each consisting of 3–4 talks (ca. 30 minutes plus 15 minutes discussion) and lasting half a day:

1. Harmonic Analysis
   (talks by Gerlind Plonka-Hoch, Naoki Saito, and Hao-Min Zhou)
2. Approximation Theory
   (talks by Martin Buhmann, Armin Iske, Nira Dyn, and Tomas Sauer)
3. Inpainting
   (talks by Aurelien Bourquard, Carola-Bibiane Schönlieb, and Yann Gousseau)
4. Compression
   (talks by Gene Cheung, Joan Serra Sagrista, and Claire Mantel)
5. Optimisation of Data and Operators
   (talks by Zakaria Belhachmi, Laurent Hoeltgen, Peter Ochs, and Pascal Peter)
6. Algorithms, Biological Vision, and Benchmarking
   (talks by Jalal Fadili, Johannes Ballé, and Sarah Andris)

These sessions triggered interesting discussions during the talks, in the breaks, and in the evening, and they allowed the different communities to learn many new things from each other.

Our program featured also an evening panel discussion on open research questions on the interface between image inpainting and image compression. It was a lively interaction between the five panel members and the audience, involving also controversial statements and views about the future of inpainting-based codecs.

The participants had a very positive impression of this seminar as an inspiring forum to bring together different fields. As a consequence, this symposium also created several new collaborations, e.g. regarding interpolation with radial basis functions, subdivision-based coding, and diffusion-based coding. There was a general consensus that it would be desirable to have another seminar on this topic in 2–3 years. Moreover, it is planned to compile a related monograph which will be the first in its field.

2 Table of Contents

3　Overview of Talks

3.1　A Benchmark for Inpainting-Based Image Reconstruction and Compression

Sarah Andris (Universität des Saarlandes, DE)

Well-designed benchmarks have been able to push research forward in various fields by providing a basis for fair comparison. To the best of our knowledge, there is no benchmark for inpainting-based image compression that provides suitable sets of images and allows to compare and rank different methods. Our goal is to introduce such a benchmark. To this end, we present a webpage that contains preliminary test sets and allows users to test the ability of masks and corresponding grey or colour values to reconstruct images from these sets. Rather than introducing a complete benchmark, this talk intends to establish a foundation for discussion on the design of this webpage, offering different possibilities and extensions. We are convinced that an acknowledged benchmark in this area is able to foster fair comparisons and can promote reproducible research.

3.2　Nonlinear Image Representations with Cascaded Local Gain Control

Johannes Ballé (New York University, US)

Local gain control is found throughout biological sensory systems. Modeled as an operation known as divisive normalisation, it can be implemented as an invertible nonlinear transform, and has several interesting properties useful to image processing. It can be used to Gaussianise image densities, eliminating statistical dependencies and providing probabilistic image models that outperform sparse representations. It can mimic human perception of visual distortions, often manifesting as masking effects, and additionally can be used for image compression. The image degradations due to compression appear reminiscent of the results of edge-preserving smoothing methods, such as anisotropic diffusion.

3.3　What is a Right Operator in Inpainting?

Zakaria Belhachmi (University of Mulhouse, FR)

The choice of a "right" model/operator for the reconstruction in inpainting is still a challenging problem. We present some ideas on modeling with diffusion operators for geometry inpainting. As framework we choose variational methods based on partial differential equations (PDEs) and their discrete counterparts. The objective is to provide a closed loop continuous to

discrete model. The loop consists of an initial family of simple PDEs depending on some parameters – "corrected" at the discrete level – which yields (in the Gamma-convergence sense) a sophisticated limit model that captures the jump set of the restored image. We also discuss inpainting-based compression within this approach.

3.4 Variations on Edge-Enhancing Diffusion: Sparse Interpolation and Beyond

Aurélien Bourquard (MIT – Cambridge, US)

In this talk, we shall introduce an approach to interpolate images from sparse sets of samples. The proposed method aims at preserving image edges and is thus inspired from edge-enhancing diffusion (EED). Meanwhile, one important characteristic of our algorithm is that it follows variational principles. Specifically, it involves the minimisation of successive quadratic-cost functionals, which relates to iteratively-reweighted-least-squares (IRLS) methods. Considering several image examples, we shall illustrate how such an approach is able to produce high-quality results while maintaining computational efficiency, even when as little as 2% of the image samples are available. Finally, we shall discuss how one could leverage the enhanced modularity of the proposed algorithm's structure for adaptations to wider classes of problems, including compressed sensing.

3.5 Convergence Properties of Multiquadrics with Parameters

Martin Buhmann (Universität Gießen, DE)

Radial basis function methods are a useful approach especially to multivariable approximation, e.g., by interpolation, but also using the successful alternative called quasi-interpolation. This is all based on the idea that one can employ shifts of a single radially symmetric function to form spaces of approximating functions in any number of variables. This univariate function can be a decaying function such as an exponential, but unbounded radial basis functions such as multiquadrics are successful, too.

The initial approach to choosing which approximants to take, given possibly a very large number of scattered data in high dimensions, is usually by interpolation. Therefore one uses shifts of a radial basis function, such as the multiquadrics or Gauss- or Poisson-kernel, all of which are related to solving certain partial differential equations and to minimising some particular semi-norms, to interpolate the mentioned data. This of course raises the question of solvability of the problem and of the efficient computability.

We will discuss in this talk mainly multiquadric interpolation, its convergence properties when centres become dense in multiple dimensions, and their behaviour when parameters of the multiquadric function change.

3.6   Graph Signal Processing for Image Coding and Restoration

Gene Cheung (National Institute of Informatics – Tokyo, JP)

Graph signal processing (GSP) is the study of discrete signals that live on structured data kernels described by graphs. By allowing a more flexible graphical description of the underlying data kernel, GSP can be viewed as a generalisation of traditional signal processing techniques that target signals in regular kernels – e.g., an audio clip sampled periodically in time – while still providing a frequency domain interpretation of the observed signals. Though an image is a regularly sampled signal on a 2-D grid, one can nonetheless consider an image or an image patch as a graph-signal on a sparsely connected graph defined signal-dependently. Recent GSP works have shown that such an approach can lead to a compact signal representation in the graph Fourier domain, resulting in noticeable gain in image compression and restoration. Specifically, in this talk I will overview recent advances in GSP as applied to image processing. We will first describe how a Graph Fourier Transform (GFT) – a generalisation of known transforms like Discrete Cosine Transform (DCT) and Asymmetric Discrete Sine Transform (ADST) – can be defined in a signal-dependent manner and leads to compression gain over traditional DCT for piecewise smooth images. We will then describe how suitable graph-signal smoothness priors can be constructed for a graph-based image denoising algorithm for piecewise smooth images. Similar graph-signal smoothness priors can also be used for other image restoration problems, such as de-quantisation of compressed JPEG images.

3.7   Linear and Nonlinear Subdivision Schemes – An Overview

Nira Dyn (Tel Aviv University, IL)

Subdivision schemes are processes for the generation of curves/surfaces from discrete data by repeated refinements. The implementation of these schemes is simple, but their mathematical analysis is rather involved. This talk will present briefly the "classical" case of linear univariate schemes refining points for generating curves. In particular, two basic schemes will be considered. A short introduction to bivariate schemes generating surfaces, together with a specific example will also be discussed.

The talk will conclude with an extension of the 4-point linear scheme to a nonlinear scheme, capable of handling non-uniform geometries, together with a few examples of its performance.

3.8   IPOL: Image Processing On Line

Gabriele Facciolo (ENS – Cachan, FR)

It is clear that software is important in modern science. It is used to solve real scientific problems in a variety of fields ranging from physics, to medicine, and economics. However, research software is not like general purpose equipment (like telescopes or compilers). Research software is usually made by scientists for scientists, tailored to a concrete experimental process. So this code is not meant to be released nor published and even less reviewed. Nevertheless, the research code is very important, since it embodies the process itself. So the research software is key for attaining a truly reproducible research.

In this talk I am going to share our experience with IPOL: an on-line journal for image processing, a journal that publishes algorithm descriptions along with the code that implement them. IPOL follows the open access and reproducible research models. Since it is a journal, the code is also peer-reviewed to verify that it corresponds to the algorithm description.

The distinguishing characteristic of IPOL articles is that each article also includes an on-line demo that permits to easily test the code with new data and parameters from the web, and a public archive that stores all original test data used in the demo. Both the demo and the archive provide great insight into the algorithm inner workings and limitations, facilitating the experimentation and leading in the end to a stricter verification of the algorithm.

3.9   Variational Analysis Meets Stochastic Differential Equations: Application to Inverse Problems

Jalal Fadili (Caen University, FR)

Unlike regularised estimators which amount to solving an optimisation problem, computing the posterior conditional mean estimator corresponds to an integration problem which becomes very involved to solve analytically or even numerically in high-dimension. An alternative is to approximate it via a Markov chain Monte-Carlo (MCMC) method which consists in sampling from the target distribution by constructing an appropriate Markov chain, and to compute sample path averages based on the output of the Markov chain. In this work, we study an exponentially weighted aggregator/estimator. To compute this estimator, we propose a family of forward-backward-type proximal Langevin Monte-Carlo algorithms to efficiently sample from the target distribution (which is not smooth nor log-concave) and derive its guarantees. We illustrate it on various inverse problems including inpainting.

3.10   Patch-Based Video Inpainting

Yann Gousseau (Telecom ParisTech, FR)

Video inpainting is a challenging task for several reasons, including: our sensibility to temporal discontinuity or fluctuations, the volume of data involved, the presence of complex motions and moving textures, etc. To the best of our knowledge, the most efficient approaches to this problem rely on heuristics to minimise a global, patch-based energy, as initially introduced by Y. Wexler et al. about ten years ago.

In this talk, I will first review these approaches, before presenting some recent contributions that can yield efficient reconstructions for complex and high resolution videos. Eventually I will show failure examples and open problems.

3.11   Overview of Image Compression: Basics and Research Issues

Christine Guillemot (INRIA – Rennes, FR)

Image and video compression is now an old research topic, spanning three decades from the mid 80s. Research in the field has produced an innumerable amount of algorithms and techniques. This talk will start with a few source coding basics, and we will see how these basics have guided the design of key coding operators. Focusing on two operators, transforms and prediction, the talk will then present advances from isotropic to anisotropic transforms, to sparse and graph-based transforms. We will then move to the problem of prediction which is analogous to inpainting, and present a prediction approach based on sparse models. We will finish by presenting a novel coding architecture building upon concepts of epitomes and based on inpainting.

3.12   Optimising Data for PDE-Based Inpainting and Compression

Laurent Hoeltgen (BTU Cottbus, DE)

Finding good reconstruction data is a key problem for PDE-based inpainting within the context of lossy image compression. Not only the location of important pixels but also their corresponding colour values should be optimal. In this talk we discuss how the spatial and tonal optimisation are related and provide strategies to retrieve this data. Finally, we will discuss additional optimisation challenges that occur during the encoding of the data.

3.13 Geometrical Methods for Adaptive Approximation of Image and Video Data

Armin Iske (Universität Hamburg, DE)

We survey more recent and less recent results concerning adaptive approximation of images and videos. The utilised approximation methods rely on linear splines over anisotropic simplices and adaptive thinning, a greedy point removal scheme for multivariate scattered data. We will address both computational and theoretical aspects of adaptive thinning, including optimal N-term approximations.

3.14 Compression of Infrared Images

Claire Mantel (Technical University of Denmark – Lyngby, DK)

Advances in the manufacturing of thermal sensors (LWIR) have led to increased accessibility, improved precision, and higher resolution. IR images differ from visible light ones by their characteristics (due to the formation process of IR images) but also by their bitdepth (most IR sensors are 14b). This implies either tone mapping the images to 8b or high bitdepth compression. We will present a preliminary study on the compression of IR images.

3.15 Inpainting: An Overview

Simon Masnou (University Claude Bernard – Lyon, FR)

The purpose of this overview talk is to review various approaches which have been proposed to address the inpainting problem: geometric, patch-based, variational, or harmonic methods, and their combinations.

Geometric approaches consist in advecting or diffusing the information from outside to inside the inpainting domain, using variational or PDE models. Such methods are usually efficient for reconstructing geometric informations, but are not adapted to texture reconstruction. On the contrary, patch-based methods were initially designed for texture synthesis, and have been later adapted to image inpainting. Roughly speaking, they are based on copying and pasting (or averaging) similar "patches", i.e. pixel neighbourhoods. Modern approaches combine both geometric and patch-based principles, and rely crucially on a multiscale strategy. Excellent results can be obtained, although stability cannot be guaranteed: slightly different inpainting domains or parameters may yield very different results. A third category of methods performs very well for sparse inpainting domains: harmonic approaches, which use image decomposition on various dictionaries, for instance one adapted to geometry, another one adapted to texture, etc.

Numerous results for 2-D or 3-D images will be shown during the talk. The challenges of video inpainting will also be illustrated.

3.16 Optimising Data for Anisotropic Diffusion Inpainting with an Optimal Control Approach

Peter Ochs (Universität Freiburg, DE)

A key problem for inpainting processes based on partial differential equations (PDEs) in image compression applications is to find optimal locations and values of the stored grey values. Even for homogeneous diffusion models this is a challenging task. An optimal control approach is one of the best strategies to find optimal locations. Advanced anisotropic inpainting operators such as edge-enhancing diffusion (EED) yield a high reconstruction quality. This talk discusses an optimal control approach with the EED inpainting operator.

3.17 Gradients versus Grey Values for Sparse Image Reconstruction and Inpainting-Based Compression

Pascal Peter (Universität des Saarlandes, DE)

Inpainting-based compression codecs store images in terms of a small fraction of pixels and interpolate the missing image parts for decoding. Recently, Brinkmann et al. (2015) have suggested an alternative sparse representation: They propose to store gradient data instead of grey values. However, the potential of this idea for PDE-based reconstruction and compression still needs to be evaluated.

The goal of this talk is to provide a fair comparison of gradient and grey value data for homogeneous diffusion inpainting. To this end, we establish a framework for optimising and encoding the known data. It allows a fair comparison of the grey value and the gradient approach, as well as combinations of both kinds of known data. In this talk, we distinguish between two separate assessments of quality: On one hand, pure reconstruction quality can be analysed by inpainting images from comparable amounts of known data for each method. On the other hand, compression quality is obtained by also taking storage costs into account by specifying a full codec for each approach.

Our evaluation shows that gradient-based reconstructions can improve the visual quality in a pure reconstruction setting: They avoid singularities involved in the reconstructions from grey values. Surprisingly, this advantage does not carry over to compression due to high sensitivity of gradient data to quantisation. Combining both kinds of data can attenuate this drawback of the gradients, but overall, pure grey-value reconstructions still remain the better choice for compression. This reveals an important general principle for inpainting-based compression: Good reconstruction quality from sparse data only leads to good compression performance, if the approach is also robust w.r.t. lossy compression steps.

3.18   Computational Harmonic Analysis Tools for Image Compression and Inpainting

Gerlind Plonka-Hoch (Universität Göttingen, DE)

In this talk we will give an overview on basic ideas for image compression and inpainting using transforms from computational harmonic analysis.

Standard compression schemes like JPEG are essentially based on data decorellation using the cosine transform or tensor-product wavelet transforms. In recent years, many further adaptive and non-adaptive transforms have been proposed that can be used for sparse data representation and image compression.

The talk shows some examples of adaptive wavelet transforms as well as non-adaptive directional wavelet frames. Adaptive transforms can be directly used for image compression. They need a careful design to ensure stability of the inverse transform or a precise analysis of occurring adaptivity costs. Nonadaptive redundant frames, as e.g. curvelets or shearlets, cannot be directly used for compression but may serve as a tool to construct suitable regularisation terms if we consider solving optimisation problems for inpainting.

3.19   Convex Optimisation and Inpainting: A Tutorial

Thomas Pock (TU Graz, AT)

In this talk, we will discuss recent trends in convex optimisation for image processing and in particular for image inpainting. First, we will discuss popular non-smooth convex models such as the total variation, higher-order extensions, and also curvature depending models. Then, we will see how these models can be applied for image inpainting problems and finally we will discuss efficient first-order algorithms to minimise these models.

3.20   Polyharmonic Local Cosine Transforms for Improving JPEG-Compressed Images

Naoki Saito (University of California – Davis, US)

I will discuss two image compression-decompression schemes that reproduce images with better visual fidelity, less blocking artifacts, and better PSNR, particularly in low bit rates, than those processed by the JPEG Baseline method at the same bit rates. We got the patents on these algorithms both in Japan (2009) and the US (2011). The additional computational cost is small, i.e., linearly proportional to the number of pixels in an input image. The first method, the "full mode" polyharmonic local cosine transform (PHLCT), modifies the encoder and decoder parts of the JPEG Baseline method. The goal of the full mode PHLCT is to reduce the code size in the encoding part and reduce the blocking artifacts

in the decoder part. The second one, the "partial mode" PHLCT (or PPHLCT for short), modifies only the decoder part, and consequently, accepts JPEG files, yet decompresses them with higher quality with less blocking artifacts. The key idea behind these algorithms is a decomposition of each image block into a polyharmonic component and a residual. The polyharmonic component in this talk is an approximate solution to Poisson's equation with Neumann boundary conditions, which means that it is a smooth predictor of the original image block only using the image gradient information across the block boundary. Thus the residual – obtained by removing the polyharmonic component from the original image block – has approximately zero gradient across the block boundary which gives rise to fast-decaying DCT coefficients, which in turn lead to more efficient compression-decompression algorithms for the same bitrates. Our numerical experiments objectively and subjectively demonstrate the superiority of PHLCT over the JPEG Baseline method and the improvement of JPEG-compressed images when decompressed by PPHLCT.

If time permits, I will also discuss my idea of using the Helmholtz equation instead of the Poisson/Laplace equations for images containing oscillatory textures.

3.21   Reconstruction of Sparse Exponential Polynomials from Samples

Tomas Sauer (Universität Passau, DE)

Prony's method, in its various concrete algorithmic realisations, is concerned with the reconstruction of a sparse exponential sum from integer samples. In several variables, the reconstruction is based on finding the variety for a zero dimensional radical ideal. If one replaces the coefficients in the representation by polynomials, i.e., tries to recover sparse exponential polynomials, the zeros associated to the ideal have multiplicities attached to them. Indeed, there is a precise and fairly explicit relationship between the coefficients in the exponential polynomial and the multiplicity spaces of zeros.

3.22   Anisotropic Surface Interpolation

Carola-Bibiane Schönlieb (University of Cambridge, GB)

In this talk we discuss the use of anisotropic total variation regularisation for interpolating highly structured functions. We will motivate the modelling, discuss its numerical solution and show applications to digital elevation maps and limited angle tomography.

3.23 Remote Sensing Data Compression

Joan Serra Sagristà (Autonomus University of Barcelona, ES)

Remote sensing data compression is an active area of research and has seen some critical developments recently. We will first motivate the need for efficient data transmission – where data compression plays a significant role – in current and future Earth Observation missions. Then we will introduce some of our own recent developments in this field.

3.24 Inpainting-Based Image Compression: State-of-the-Art, Challenges, and Connections

Joachim Weickert (Universität des Saarlandes, DE)

Inpainting-based approaches for lossy image compression store only a small, carefully selected subset of the image data. In the decoding phase, the missing data are reconstructed by a suitable inpainting process. This can be achieved e.g. by interpolation with partial differential equations (PDEs) or by copying information from patches in other image regions. Such codecs allow a very intuitive interpretation, and first experiments show their advantages for high compression rates where they can beat even advanced transform-based methods. However, much more research is needed to turn them into viable codecs. In this talk we will discuss the state-of-the-art, identify challenging problems, and mention relations to other fields. We will see that these problems can only be solved, if different communities with expertise on inpainting, approximation theory, optimisation, numerical analysis, and coding collaborate in an interdisciplinary way.

3.25 An Approximated PDHG Method for Total Variation Wavelet Inpainting

Hao-Min Zhou (Georgia Institute of Technology – Atlanta, US)

The primal-dual hybrid gradient (PDHG) algorithm has been successfully applied to a number of total variation (TV) based image reconstruction problems for fast numerical solutions. We show that PDHG can also effectively solve the computational problem of image inpainting in wavelet domain, where high quality images are to be recovered from incomplete wavelet coefficients due to lossy data transmission. In particular, we propose an approximated PDHG algorithm to tackle the non-orthogonality of Daubechies 7-9 wavelet which is widely used in practice. We show that this approximated version essentially alters the gradient descent direction in the original PDHG algorithm, but eliminates its orthogonality restriction and retains low computational complexity.

4   Panel discussions

4.1   Panel Discussion on Inpainting-Based Image Compression

Thomas Pock (TU Graz, AT)

The panel discussion took place on Thursday, 17.11.2016, 20:00–21:00. Its goal was to discuss open research questions on the interface between image inpainting and image compression. The panel consisted of five panel members: Joachim Weickert, Jalal Fadili, Carola-Bibiane Schönlieb, Christine Guillemot, and Gerlind Plonka-Hoch. The moderator of the panel discussion was Thomas Pock.

The panel discussion started with a question to Joachim Weickert, why it should be a good idea to use image inpainting techniques for image and video compression algorithms. Joachim Weickert referred to the properties of the human visual system, which is mainly guided by edges and textures. Compressing this information in a suitable way, it might therefore be a good idea to reconstruct the missing image information be means of image inpainting. Furthermore, people working in image inpainting need to understand and solve the same kind of problems as people working in image compression. Gerlind Plonka-Hoch pointed out that this was actually always the goal of the wavelet transform or image adaptive variants that have been proposed in the past.

When asked if image inpainting based techniques will lead to new codecs, Christine Guillemot pointed out that it is very difficult to establish new standards for image and video codecs. First, people who have worked for more than 30 years on codecs will not easily accept a new technique, second, existing codecs might be "good enough", and third, companies who have already created dedicated hardware for established codecs will try to promote their existing algorithms simply to make more money. This is in particular true for the JPEG2000 standard which on the mass market was never able to become a substitute for its predecessor JPEG. Christine also expressed her feeling that inpainting-based compression algorithms will probably be relevant only for certain niche applications.

At this point, Jalal Fadili told the story of a very specialised imaging device that is used in the Herschel space probe. He was part of the team developing a compressed sensing technique for this device. Although the compressed sensing technique would have led to improved image quality, JPEG2000 was used in the end, mainly due to software reasons.

To account for that, Joachim Weickert remarked that inpainting-based image compression should be as simple as possible (e.g. linear, homogeneous diffusion) in order to allow for fast decoding. Joachim Weickert also pointed out that each neuron in the human brain is doing only very simple operations. To understand how the complexity of natural images can be realised by very simple and local operations will be a major future research questions.

Another topic that was discussed in the panel was video compression. During the week we saw very interesting and impressive results for video inpainting but almost no work has been done to use such techniques for video compression. One of the reasons is the high computational complexity of video inpainting algorithms that hinders those algorithms from real-time applications. Moreover, the human perception of videos is often quite counter-intuitive and hence it is difficult to define suitable quality measurements which take these aspects into account.

Finally, the panel discussed the role of quality measurements for image compression. It turns out that it is still completely unclear which performance measure is the most relevant.

In fact, image compression might be even task-dependent. Carola Schönlieb pointed out that when trying to find a lost object in an image or identifying a certain person in a video, one might be interested in certain details which could have been lost when compressing the image or video. Moreover, in medical applications, CT or MRI images are actually measuring certain physical quantities and medical doctors will in fact base their diagnoses on the image content. Hence one has to be even more careful when applying image compression to such images.

In general it turned out that the panel discussion was a very good idea and it led to vital discussions between the panel members and the audience. The discussion showed that there are much more open questions than solutions and that bringing together different communities in this seminar was actually a very good idea.

Participants

- Sarah Andris
Universität des Saarlandes, DE
- Johannes Ballé
New York University, US
- Zakaria Belhachmi
University of Mulhouse, FR
- Aurélien Bourquard
MIT – Cambridge, US
- Eva-Maria Brinkmann
Universität Münster, DE
- Dorin Bucur
University Savoie Mont Blanc –
Le Bourget-du-Lac, FR
- Martin Buhmann
Universität Gießen, DE
- Gene Cheung
National Institute of Informatics –
Tokyo, JP
- Nira Dyn
Tel Aviv University, IL
- Gabriele Facciolo
ENS – Cachan, FR

- Jalal Fadili
Caen University, FR
- Irena Galic
University of Osijek, HR
- Yann Gousseau
Telecom ParisTech, FR
- Christine Guillemot
INRIA – Rennes, FR
- Laurent Hoeltgen
BTU Cottbus, DE
- Armin Iske
Universität Hamburg, DE
- Claire Mantel
Technical University of Denmark
– Lyngby, DK
- Simon Masnou
University Claude Bernard –
Lyon, FR
- Peter Ochs
Universität Freiburg, DE
- Pascal Peter
Universität des Saarlandes, DE

- Gerlind Plonka-Hoch
Universität Göttingen, DE
- Thomas Pock
TU Graz, AT
- Daniela Rosca
TU of Cluj-Napoca, RO
- Naoki Saito
Univ. of California – Davis, US
- Tomas Sauer
Universität Passau, DE
- Carola-Bibiane Schönlieb
University of Cambridge, GB
- Joan Serra Sagristà
Autonomus University of
Barcelona, ES
- Joachim Weickert
Universität des Saarlandes, DE
- Hao-Min Zhou
Georgia Institute of Technology –
Atlanta, U

Concurrency with Weak Memory Models: Semantics, Languages, Compilation, Verification, Static Analysis, and Synthesis

Edited by

Jade Alglave[1], Patrick Cousot[2], and Caterina Urban[3]

1   University College London, GB, j.alglave@ucl.ac.uk
2   New York University, US, pcousot@cims.nyu.edu
3   ETH Zürich, CH, caterina.urban@inf.ethz.ch

──── **Abstract** ────

This report documents the program and the outcomes of Dagstuhl Seminar 16471 "Concurrency with Weak Memory Models: Semantics, Languages, Compilation, Verification, Static Analysis, and Synthesis". The aim of the seminar was to bring together people from various horizons, including theoreticians and verification practitioners as well as hardware vendors, in order to set up a long-term research program to design formal methods and develop tools ensuring the correctness of concurrent programs on modern multi-processor architectures.

1   Executive Summary

Jade Alglave
Patrick Cousot

In the last decade, research on weak memory has focussed on modeling accurately and precisely existing systems such as hardware chips. These laudable efforts have led to definitions of models such as IBM Power, Intel x86, Nvidia GPUs and others.

Now that we have faithful models, and know how to write others if need be, we can focus on how to use these models for verification, for example to assess the correctness of concurrent programs.

The goal of our seminar was to discuss how to get there. To do so, we gathered people from various horizons: hardware vendors, theoreticians, verification practitioners and hackers. We asked them what issues they are facing, and what tools they would need to help them tackle said issues.

The first day was dedicated to theory. We had overviews of classic semanticists tools such as event structures, message sequence charts, and pomsets. The remaining days were

mostly dedicated to models and verification practices, whether from a user point of view, or a designer point of view. We chose to close the days early, so that our guests would have ample time to come back to an interesting point they had heard during one of the talks, or engage in deep discussions. The feedback we got was quite positive, in that the seminar help spark discussions with, for example, a PhD student in concurrency theory, and a verification practitioner from ARM.

2   Table of Contents

3 Overview of Talks

3.1 Robustness against Consistency Models with Atomic Visibility

Giovanni Tito Bernardi (University Paris-Diderot, FR)

To achieve scalability, modern Internet services often rely on distributed databases with consistency models for transactions weaker than serializability. At present, application programmers often lack techniques to ensure that the weakness of these consistency models does not violate application correctness. In this talk I will present criteria to check whether applications that rely on a database providing only weak consistency are robust, i.e., behave as if they used a database providing serializability, and I will focus on a consistency model called Parallel Snapshot Isolation. The results I will outline handle systematically and uniformly several recently proposed weak consistency models, as well as a mechanism for strengthening consistency in parts of an application.

3.2 Transactions on Mergeable Objects in Shared-Memory

Annette Bieniusa (TU Kaiserslautern, DE)

Under high contention, serializability for transactions results in frequent aborts. This limits possible parallelism and results in performance degradation. In this talk, we introduce a new transactional semantics, Mergeable Transactions, which allows concurrent transactions on the same objects to execute in parallel. Instead of aborting and re-executing, the conflicting updates on shared objects are merged using type specific semantics. We show that mergeable transactions outperform serializable transactions under high contention workloads.

3.3 New Lace is a Program Logic for Weak Memory (Probably)

Richard Bornat (Middlesex University – London, GB)

It is possible to reason about weak-memory executions of litmus tests (but not yet about synchronized assignment) using a version of rely/guarantee. Constraints between commands and/or control expressions control order of elaboration (local execution) and propagation of writes. The logic is driven by the temporal modality 'since' and some specialized modalities based on it.

The logic has only a weak grasp of causality (treated by auxiliary variables, as usual in Owicki-Gries logics). It has some surprising rules dealing with stability: five or six different kinds of stability. It has a proof-checker (Arsenic, available on GitHub); the proof-checker is needed for even smallish proofs, which shows that proofs in the logic are far too complicated.

3.4 A Denotational Framework for Weak Memory Conucrrency

Stephen Brookes (Carnegie Mellon University – Pittsburgh, US)

We present a denotational semantic framework for compositional reasoning about shared-memory parallel programs, assuming a form of weak memory. Traditional denotational models of shared-memory programs are trace-based, assume sequential consistency, and use global states and interleaving, rendering them poorly suited for expressing weak memory behavior. Instead we abandon sequential consistency and embrace "true" concurrency: a program denotes a set of pomsets (partially ordered multi-sets) of actions. Rather than global states we use footprints, built from "local" states. This framework is intended to offer an alternative to execution graphs, widely used in operational/axiomatic formalizations of weak memory. An execution graph represents the behavior of a complete program, all threads known in advance, running without external interference. The axiomatic method requires construction of various auxiliary relational edges (happens-before, reads-from, etc) constrained to satisfy a battery of axioms. These graph-based methods are inherently non-compositional, relying on knowledge of the entire program structure and assuming that execution takes place with no interference. A denotational semantics is by its very nature compositional, allowing us to take account of interference in a natural manner; and we can derive analogues of the relevant auxiliary relations automatically from the structure of pomset executions.

3.5   Weak Memory using Event Structures

Simon Castellan (ENS – Lyon, FR)

In this talk, I will introduce a methodology to model weak memory using event structures. The model is compositional and neatly separate thread & storage semantics. Moreover it comes from recent game semantics advances using causal models. The game semantics aspects allows to define the model by simply defining a few key strategies with higher-order type. In this talk, we show how to interpret the TSO model using event structures in a compositional way.

3.6   Analysing Snapshot Isolation

Andrea Cerone (Imperial College London, GB)

 Joint work of Andrea Cerone, Alexey Gotsman
Main reference A. Cerone, A. Gotsman, "Analysing Snapshot Isolation", in Proc. of the 2016 ACM Symp. on
Principles of Distributed Computing (PODC'16), pp. 55–64, ACM, 2016.
       URL http://dx.doi.org/10.1145/2933057.2933096

Snapshot isolation (SI) is a widely used consistency model for transaction processing, implemented by most major databases and some of transactional memory systems. Unfortunately, its classical definition is given in a low-level operational way, by an idealized concurrency-control algorithm, and this complicates reasoning about the behavior of applications running under SI. We give an alternative specification to SI that characterizes it in terms of transactional dependency graphs of Adya et al., generalizing serialization graphs. Unlike previous work, our characterization does not require adding additional information to dependency graphs about start and commit points of transactions.

3.7   Game Semantics based on Event Structures

Pierre Clairambault (ENS – Lyon, FR)

 Joint work of Simon Castellan, Pierre Clairambault, Silvain Rideau, Glynn Winskel
Main reference S. Castellan, P. Clairambault, S. Rideau, G. Winskel, "Games and Strategies as Event structures",
arXiv:1604.04390v3 [math.LO], 2016.
       URL https://arxiv.org/abs/1604.04390v3

Games are common objects in theoretical computer science, used in particular to model open systems: indeed, the dynamics of an open system can be regarded as a game between two players, one playing for the system and the other for the environment. In program semantics, the same idea yields a general methodology ("Game Semantics") for giving syntax-free representations of the execution in a compositional manner, for various high-level programming languages with complex control structure. In this talk, I will give an introduction to recent work on game semantics based on event structures, with a focus on the semantics of shared state concurrency. This served as the basis and inspiration for Simon Castellan's event structure semantics for weak memory, presented in a separate talk.

3.8 Proof of Mutual-Exclusion and Non-Starvation of a Program: PostgreSQL

Patrick Cousot (New York University, US) and Jade Alglave (University College London, GB)

Proof of mutual-exclusion and non-starvation of a program: PostgreSQL, Jade Alglave and Patrick Cousot.

Using the parallel program invariance proof method of Alglave and Cousot (POPL 2017), we prove the mutual exclusion property of the PostgreSQL program. The weakest memory model necessary and sufficient for this mutual exclusion property to hold is extracted from the proof by calculational design.

The invariance proof method allows us to reason on any set of executions as defined by a set of read-from relations (each read-from relation uniquely determining a single execution trace, if any). Thanks to this property, we can use the inductive invariant to prove non-starvation. We prove that any execution that starves is impossible, either because this proof method cannot satisfy the verification conditions and so (by soundness of the proof method) is not a possible execution of the program, or by disallowing this execution thanks to labeled fences (which does not change the invariance proof and which effect is defined in cat), or thanks to properties of hardware architectures (such as no read of a future write beyond a cut) not expressible in the current version of cat.

3.9 Modeling and Analysis of Remote Memory Access Programming

Andrei Marian Dan (ETH Zürich, CH)

Recent advances in networking hardware have led to a new generation of Remote Memory Access (RMA) networks in which processors from different machines can communicate directly, bypassing the operating system and allowing higher performance. Researchers and practitioners have proposed libraries and programming models for RMA to enable the development of applications running on these networks, However, the memory models implied by these RMA libraries and languages are often loosely specified, poorly understood, and differ depending on the underlying network architecture and other factors. Hence, it is difficult to precisely reason about the semantics of RMA programs or how changes in the network architecture affect them. We address this problem with the following contributions: (i) a coreRMA language which serves as a common foundation, formalizing the essential characteristics of RMA programming; (ii) complete axiomatic semantics for that language; (iii) integration of our semantics with an existing constraint solver, enabling us to exhaustively

generate core- RMA programs (litmus tests) up to a specified bound and check whether the tests satisfy their specification; and (iv) extensive validation of our semantics on real-world RMA systems. We generated and ran 7,441 litmus tests using each of the low-level RMA network APIs: DMAPP, VPI Verbs, and Portals 4. Our results confirmed that our model success- fully captures behaviors exhibited by these networks. More- over, we found RMA programs that behave inconsistently with existing documentation, confirmed by network experts. Our work provides an important step towards understand- ing existing RMA networks, thus influencing the design of future RMA interfaces and hardware.

3.10   Formalising the ARM Memory Model ... Again

Will Deacon (ARM Ltd. – Cambridge, GB)

Recent work within the ARM architecture group has led to the development of a formalization of the ARMv8 weakly consistent memory model using herd and 'cat'. Whilst this model can act as an invaluable tool when considering concurrent applications in userspace, its interactions with the system architecture are unclear and appear to be inexpressible with the current litmus test methodology. This talk will introduce the new model and highlight some of the challenges faced when integrating it with the broader architecture.

3.11   A Plea for Industrial-Strength Formal Methods for Concurrent Software

David Delmas (Airbus S.A.S. – Toulouse, FR)

Verification activities are liable for more than half of the overall effort in the development of critical avionics software. Therefore, (semi-) automatic formal verification techniques are increasingly used to improve industrial efficiency, while preserving safety. As required by the DO-333 formal method technical supplement to the DO-178C standard for avionics software development, such techniques must be sound, and associate tools have to undergo stringent qualification processes.

However, while most existing techniques and tools focus on sequential or synchronous software, an increasing share of embedded systems is being developed in asynchronous software, to save on cost, weight, and resources. AstréeA, an extension of the Astrée run-time error analyzer, is one of the very few sound static analyzers which may be used for such asynchronous software.

So far, asynchronous avionics software was mostly running on single-core platforms with real-time scheduling. Nonetheless, multi-core avionics architectures are currently being considered. There is interest for wait-free/lock-free message passing algorithms, to support ARINC 653 sampling and queuing schemes. At implementation level, there is a need for a sustainable way to ensure correctness via a minimal set of fences, without impairing hard real-time performance targets. At verification level, there is a need for sound techniques

guaranteeing functional correctness of source and compiled programs. An issue is that the CompCert compiler is only certified for sequential executions. Finally, there is a need for a sound approach to timing analysis with complex multi-core processors.

3.12 Embedding Transactions in Weak Memory Models

Stephan Diestelhorst (ARM Ltd. – Cambridge, GB)

Hardware Transactional Memory has been proposed as a higher-level memory primitive that can improve performance of and simplify parallel programming. The baseline premise of HTM is simple: transactions behave as if they are executing in isolation, despite them executing concurrently.

Recent implementations and architectures, however, need to embed this core principle into a memory model for non-transactional accesses.

Especially with a weak non-tx memory model such as ARM, embedding the strong TM semantics leaves ample room for "impedance matching" of the different strengths. Together with the combinatorial explosion of options to add transactions to well-known litmus tests, a mechanized model for experimenting with semantics is prudent.

In my talk, I will show our work in progress of using the CAT language for formalizing HTM, and also informally present some of the challenges associated with the interplay of transactional and non-transactional accesses.

3.13 A Discrete Model of Concurrent Program Execution

Charles Anthony Richard Hoare (Microsoft Research UK – Cambridge, GB)

A two-dimensional discrete (non-metric) geometry is proposed for recording the trace of execution of a concurrent program that has been expressed in an object-oriented high-level programming language. From this is derived an algebraic semantics for the programming language, and a logic for reasoning about correctness of its semantics, and an operational semantics for implementing it.

3.14 (A New Methodology for) Inductive Verification of Message-Passing Programs

Bernhard Kragl (IST Austria – Klosterneuburg, AT)

Joint work of Bernhard Kragl, Shaz Qadeer

Designing and implementing distributed systems is a hard and challenging problem. A major obstacle is to manage the complexity of the sheer number of behaviors of a system due to, e.g., nondeterministic scheduling and unreliable message delivery. This complexity transfers to the amount of annotations required in correctness proofs and significantly hinders the adoption of verification technology.

In this talk we present a methodology developed atop the CIVL verification system [CAV'15] (originally designed for shared-memory programs) to simplify the construction of correctness proofs for message-passing programs. The central theme is to establish conditions that allow message handlers to be inlined at message sends and thus enable sequential reasoning, which further eliminates complicated case distinctions in the necessary invariants. For example, in our proof of a two-phase commit protocol we do not need to state complex conditions on the history of the system or the current network state (i.e., messages in delivery).

3.15 A Promising Semantics for Relaxed-Memory Concurrency

Ori Lahav (MPI-SWS – Kaiserslautern, DE)

Joint work of Derek Dreyer, Chung-Kil Hur, Jeehoon Kang, Ori Lahav, Viktor Vafeiadis
Main reference J. Kang, C.-K. Hur, O. Lahav, V. Vafeiadis, D. Dreyer, "A Promising Semantics for
Relaxed-Memory Concurrency", in Proc. of the 44th ACM SIGPLAN Symp. on Principles of
Programming Languages (POPL'17), pp. 175–189, ACM, 2017.
URL http://dx.doi.org/10.1145/3009837.3009850

Despite many years of research, it has proven very difficult to develop a memory model for concurrent programming languages that adequately balances the conflicting desiderata of programmers, compilers, and hardware. In this talk, we present the first relaxed memory model that (1) accounts for a broad spectrum of features from the C++11 concurrency model, (2) is implementable, in the sense that it provably validates many standard compiler optimizations and reorderings, as well as standard compilation schemes to x86-TSO and Power, (3) justifies simple invariant-based reasoning, thus demonstrating the absence of bad "out-of-thin-air" behaviors, (4) supports "DRF" guarantees, ensuring that programmers who use sufficient synchronization need not understand the full complexities of relaxed-memory semantics, and (5) defines the semantics of racy programs without relying on undefined behaviors, which is a prerequisite for applicability to type-safe languages like Java.

The key novel idea behind our model is the notion of promises: a thread may promise to execute a write in the future, thus enabling other threads to read from that write out of order. Crucially, to prevent out-of-thin-air behaviors, a promise step requires a thread-local certification that it will be possible to execute the promised write even in the absence of the promise. To establish confidence in our model, we have formalized most of our key results in Coq.

3.16 Automatic Synthesis of Comprehensive Litmus Test Suites

Daniel Lustig (NVIDIA Corp. – Santa Clara, US)

Litmus tests are the basic units of testing weak memory models. Most memory model analysis infrastructures and testing suites make heavy use of litmus tests as the basic units of testing and understanding. The success of such techniques requires that the suites of litmus tests be comprehensive: that they cover every obvious and obscure corner of the memory model and/or of its implementation. However, most litmus test suites today are generated through some combination of manual effort and randomization, and this leaves them prone to human error and incompleteness.

We present a methodology for synthesizing comprehensive litmus test suites directly from the memory model specification. By construction, these suites contain all tests satisfying a minimality criterion: that no synchronization mechanism in the test can be weakened without causing new behaviors to become observable. We formalize this notion using the Alloy modeling language, and we apply it to a number of existing and newly-proposed memory models. Our results show not only that this synthesis technique can automatically reproduce all manually-generated tests from existing suites, but also that it discovers new tests that are not as well studied.

3.17 C11 Compiler Mappings: Exploration, Verification, and Counterexamples

Yatin Manerkar (Princeton University, US)

C and C++ atomic operations get mapped down to individual instructions or combinations of instructions by compilers, depending on the ordering guarantees and synchronization instructions provided by the underlying architecture. These compiler mappings must uphold the ordering guarantees provided by C/C++ atomics or the compiled program will not behave according to the C/C++ memory model. In this talk I discuss a counterexample we discovered to the well-known trailing-sync compiler mappings for the Power and ARMv7 architectures that were previously thought to be proven correct. I also discuss the loophole in the proof of the mappings that allowed the incorrect mappings to be proven correct, as well as a few optimization-related bugs that I discovered in the IBM XL C++ compiler's implementation of C++ atomics.

3.18   Taming CAT

Luc Maranget (INRIA – Paris, FR)

License ☺ Creative Commons BY 3.0 Unported license
               © Luc Maranget
Joint work of Jade Alglave, Luc Maranget
         URL http://diy.inria.fr

In this demo-talk, a Lamport style model of Sequential Consistency is created live in CAT.
CAT is the Domain Specific Language used by the memory model simulator herd7 to describe
and execute shared memory models.

See http://diy.inria.fr for software and documentation.

3.19   Linux-Kernel Memory Ordering: Help Arrives At Last!

Paul McKenney (IBM – Beaverton, US)

License ☺ Creative Commons BY 3.0 Unported license
               © Paul McKenney
Joint work of Jade Alglave, Luc Maraget, Paul McKenney, Andrea Parri, Alan Stern

It has been said that Documentation/memory-barriers.txt can be used to frighten small
children [1], and perhaps this is true. However, it is woefully inefficient. After all, there are
a very large number of children in this world, and it would take a huge amount of time and
effort to read it to all of them.

This situation clearly calls out for automation, which has been developed over the past
two years. An automated tool takes short fragments of C code as input, along with an
assertion, and carries out the axiomatic equivalent of a full state-space search to determine
whether the assertion always, sometimes, or never triggers. This talk will describe this tool
and give a short demonstration of its capabilities.

To the best of our knowledge, this is the first realistic Linux-kernel memory model, and
the first memory model of any kind incorporating a realistic model of RCU.

References
1   Mel Gorman. [PATCH 11/18] mm: fix TLB flush race between migration, and
    change_protection_range. Linux Kernel Mailing List, Hillsboro, OR, USA, 2013

3.20   Portability Analysis for Axiomatic Memory Models

Roland Meyer (TU Braunschweig, DE)

License ☺ Creative Commons BY 3.0 Unported license
               © Roland Meyer
Joint work of Florian Furbach, Keijo Heljanko, Roland Meyer, Hernán Ponce de León

We present PORTHOS, the first tool that discovers porting bugs in performance-critical
code. PORTHOS takes as input a program, the memory model of the source architecture for
which the program has been developed, and the memory model of the targeted architecture.
If the code is not portable, PORTHOS finds a porting bug in the form of an unexpected
execution – an execution that is consistent with the target but inconsistent with the source
memory model. Technically, PORTHOS implements a bounded model checking method that

reduces portability analysis to the satisfiability modulo theories (SMT) problem with integer difference logic. There are two problems in the reduction that are unique to portability and that we present novel and efficient solutions for. First, the formulation of portability contains a quantifier alternation (consistent + inconsistent). We encode inconsistency as an existential query. Second, the memory models may contain recursive definitions. We compute the corresponding least fixed points efficiently in SMT. Interestingly, we are able to prove that our execution-based notion of portability is the most liberal one that admits an efficient algorithmic analysis: for state-based portability, a polynomial SAT encoding cannot exist. Experimentally, we applied PORTHOS to a number of case studies. It is able to check portability of non-trivial programs between interesting architectures. Notably, we present the first algorithmic analysis of portability from TSO to Power.

3.21 Hazard Pointers: C++ Memory Ordering Issues

Maged M. Michael (Facebook – New York, US)

In this talk I review the hazard pointers method with focus on memory ordering issues for the main access patterns under the C++ memory consistency model. This is in the context of an ongoing effort at the C++ standard committee to add hazard pointers to the standard library.

One of the challenges in determining correct memory ordering for a hazard pointers library implementation is that main access patterns include user code that may use weak memory order specifiers. This makes using default sequentially consistent memory accesses insufficient.

I use the herd memory model simulator to find sufficient memory ordering options for preventing incorrect execution patterns. The conclusions from the experience are that support is needed for read-modify-write C++ atomic operations such as compare_exchange, fetch_add, and exchange; and that it would be very useful for complex access patterns if memory model simulation tools can generate a list of memory ordering options for a litmus test without requiring the user to specify memory ordering in the litmus test.

3.22 Static Analysis by Abstract Interpretation of Numeric Properties of Programs under Weak Memory Models

Antoine Miné (CNRS and University Pierre & Marie Curie – Paris, FR)

In this talk, we discuss the verification of concurrent programs running under weak memory models by sound and automatic static analysis based on abstract interpretation. We first recall the principles of abstract interpretation and the well-known result that abstracting

thread interference in a flow-insensitive way makes the analysis robust against reordering of independent reads and writes. Then, we focus on the TSO and PSO memory models, and propose more precise abstractions tailored for these models. Starting from an operational semantics, we leverage existing numeric abstract domains as well as array abstractions to model the store buffers in a sound way. Finally, we present an application to fence removal on small code examples.

3.23 Musketeer in Dagstuhl: Automated Fencing in Software?

Vincent Nimal (Microsoft Research UK – Cambridge, GB)

```
             |
            / \                    +
           / + \____              /\
          / ||| \   \\            ||
          |_____|____\\        /__\
____   _____|      |  ,  ||  _____|  |____
  /\ \/        | ||| |____||_/|   ,  |__| ,   |\
 _____\=====| === |       /_|_____|_\
 /\ /\ |      ||    ||_____| || /\       /\ || |
 || || | |+| |||||||| |+|   | || || /=\ || || |
 |       ||___||        | || || | | || || |
 || || | |+| || | || |+|   | ||     | |     || |
_____|_____||_|_||_____|_||___|===|___||_|

'Power, ARM!'                    'SC!'
     \                              /
      ~__   .                 __~
       |)  /                 (|
      /|__/              _____/
      | |                      \
      /|                       /\
     / |                      |  \

Musketeer in Dagstuhl: automated fencing in software?
----------------------------------------------------
```

Modern architectures rely on memory fences to prevent undesired weakenings of memory consistency. As the fences' semantics may be subtle, the automation of their placement is highly desirable, e.g., in the context of legacy code. But precise methods restoring consistency do not scale to deployed systems code. We choose to trade some precision for scalability: we present a technique suitable for larger code bases. This method is implemented in the

tool musketeer, that we experimented on more than 350 executables of packages found in a Debian Linux distribution, e.g. memcached (about 10000 LoC).

This talk recalls some results of our CAV 2014 paper, with updated results and insights from our TOPLAS paper. It then discusses some difficulties inherent to the fence insertion problem preventing good compositionality, which also apply to other approaches.

3.24 Verifying a Concurrent Garbage Collector

Gustavo Petri (University Paris-Diderot, FR) and Delphine Demange (IRISA – Rennes, FR)

We consider the problem of mechanically verifying a state-of-the-art, on-the-fly concurrent garbage collector. To facilitate this task, we present a compiler intermediate representation (IR) that subsumes a concurrent programming language and a proof methodology. Our IR provides strong type guarantees, abstract concurrent data structures, and intrinsic support for threads, roots management and object inspection via high-level iterators. Our IR is also accompanied with a rely-guarantee program logic which we use to prove the functional correctness of programs. In the implementation of our collector, data races are omnipresent. To argue about the correctness of our garbage collector under the TSO memory model, we plan to exploit the fact that the "safe publication idiom" under TSO provides a semantics equivalent to that under the SC memory model.

3.25 Mixed-Size Concurrency: ARM, POWER, C/C++11, and SC

Susmit Sarkar (University of St. Andrews, GB)

Previous work on the semantics of relaxed shared-memory concurrency has only considered the case in which each load reads the data of exactly one store. In practice, however, multiprocessors support mixed-size accesses, and these are used by programs in C/C++, and particularly systems software.

I will describe recent work on modeling mixed-size behavior of POWER and ARM architectures and implementations, showing new aspects of memory consistency models that arise in this setting. In particular, the abstract notion of coherence becomes more subtle, and adding a barrier between each instruction does not restore Sequential Consistency.

This work was published at POPL'17.

3.26 Reachability for Dynamic Parametric Processes

Helmut Seidl (TU München, DE)

In a dynamic parametric process every subprocess may spawn arbitrarily many, identical child processes, that may communicate either over global variables, or over local variables that are shared with their parent.

We show that reachability for dynamic parametric processes is decidable under mild assumptions. These assumptions are, e.g., met if individual processes are realized by pushdown systems, or even higher-order pushdown systems. We also discuss in how far these methods can also deal with weak memory models.

3.27 Data Consistency Check of Very Large Execution Traces

Suzanne Shoaraee (ARM France SAS – Sophia-Antipolis, FR)

Verifying ARM CPU implementations could be challenging especially when it deals with the memory system. In this talk, I present one of our current challenges: be able to check the data consistency of very large execution traces containing millions of memory accesses.

I introduce ARM specific requirements (use of temporal information, barriers parameters . . .) and the difficulty to adapt existing formalizations of the memory model. A limited but scalable checker has been developed and is presented as it is already of interest to our verification teams.

At last the remaining challenges to be solved such as the handling of single, multiple-copy atomicity, atomic instructions or barriers are discussed.

3.28 From Architecture to Implementation

Daryl Stewart (ARM Ltd. – Cambridge, GB)

An architecture specification represents a contract between hardware and software which defines the permitted behaviors of a system. For reasoning about software this contract should be weak so that programmers code defensively against undesirable outcomes. For hardware it should be strong, so that implementations exhibit no more behaviors than permitted (or expected.) We propose that a no-man's land between the two communities is safer than attempting perfection.

When verifying hardware during bottom-up development we seek a framework for ensuring that the local behavior of subunits is correct with respect to the specified global behaviors. I

will show some of the local behaviors of hardware which give rise to the surprising behaviors of weak memory systems, along with some verification properties which can be applied to them in order to highlight the semantic gap between architecture and implementation.

3.29 TriCheck: Memory Model Verification at the Trisection of Software, Hardware, and ISA

Caroline J. Trippel (Princeton University, US)

License	Creative Commons BY 3.0 Unported license
	© Caroline J. Trippel
Joint work of	Daniel Lustig, Yatin A. Manerkar, Margaret Martonosi, Michael Pellauer, Caroline J. Trippel
Main reference	C. Trippel, Y. A. Manerkar, D. Lustig, M. Pellauer, M. Martonosi, "TriCheck: Memory Model Verification at the Trisection of Software, Hardware, and ISA", to appear in Proc. of the 22nd Int'l Conf. Architectural Support for Programming Languages and Operating Systems (ASPLOS'17), 2017; pre-print available from author's webpage.
URL	http://mrmgroup.cs.princeton.edu/papers/ctrippel_ASPLOS17.pdf

The ISA is a multi-part specification of hardware behavior as seen by software. One significant, yet often under-appreciated aspect of this specification is the memory consistency model which governs inter-module interactions in a shared memory system. We make a case for full-stack memory model design and verification and provide a toolflow – TriCheck – to support it. We apply TriCheck to the open source RISC-V ISA, focusing on the goal of accurate, efficient, and legal compilations from C11/C++11. In doing so, we uncover under-specifications and potential inefficiencies in the current RISC-V ISA documentation and identify possible solutions for each. We also identify two counter-examples to previously "proven-correct" compiler mappings from C11/C++11 to POWER and ARMv7.

3.30 Explaining Relaxed Memory Models with Program Transformations

Viktor Vafeiadis (MPI-SWS – Kaiserslautern, DE)

License	Creative Commons BY 3.0 Unported license
	© Viktor Vafeiadis
Joint work of	Ori Lahav, Viktor Vafeiadis
Main reference	O. Lahav, V. Vafeiadis, "Explaining Relaxed Memory Models with Program Transformations", in Proc. of the 21st Int'l Symp. on Formal Methods (FM'16), LNCS, Vol. 9995, pp. 479–495, Springer, 2016.
URL	http://dx.doi.org/10.1007/978-3-319-48989-6_29

Weak memory models determine the behavior of concurrent programs. While they are often understood in terms of reorderings that the hardware or the compiler may perform, their formal definitions are typically given in a very different style – either axiomatic or operational. In the talk, we investigate to what extent weak behaviors of existing memory models can be fully explained in terms of reorderings and other program transformations. We prove that TSO is equivalent to a set of two local transformations over sequential consistency, but that non-multi-copy-atomic models (such as C11, Power and ARM) cannot be explained in terms of local transformations over sequential consistency. We then show that transformations over a basic non-multi-copy-atomic model account for the relaxed behaviors of (a large fragment of) Power, but that ARM's relaxed behaviors cannot be explained in a similar way. Our

positive results may be used to simplify correctness of compilation proofs from the promising semantics of Kang et al. [1] to TSO or Power. More details can be found in [2].

References
1   J. Kang, C.-K. Hur, O. Lahav, V. Vafeiadis, D. Dreyer. A promising semantics for relaxed-memory concurrency. POPL 2017, pp. 175–189, ACM, 2017.
2   O. Lahav, V. Vafeiadis. Explaining relaxed memory models with program transformations. FM 2016, LNCS, Vol. 9995, pp. 479–495, Springer, 2016.

3.31   Event Structures and Stable Families

Glynn Winskel (University of Cambridge, GB)

This talk revisits old work on Event Structures (1978) and Stable Families (1981) which are relevant or potentially relevant in the modeling of weak memory in hardware design. Some recent work, e.g. that of Alan Jeffrey or separately Simon Castellan, uses event structures in modeling weak memory. In particular, Castellan uses a product of event structures; his work also fits within concurrent games where the composition of strategies uses the pullback of event structures. Both product and pullback of event structures are difficult to define directly on event structures. Here stable families come to the rescue: a coreflection from a category of event structures to a category of stable families allows us to transport the constructions from the simpler construction of product and pullback in stable families.

3.32   Weak Memory Models: Balancing Definitional Simplicity and Implementation Flexibility

Sizhuo Zhang (MIT – Cambridge, US)

RISC-V, a newly developed open source ISA, has not finalized its memory model, and thus offers an opportunity to explore the design space of weak memory models. We propose two new weak memory models: WMM and WMM-S, which balance definitional simplicity and implementation flexibility differently. Both allow all instruction reorderings except overtaking of loads by a store. We show that this restriction has little impact on performance and it considerably simplifies operational definitions. It also rules out the out-of-thin-air problem that plagues many definitions. WMM is simple (it is similar to the Alpha memory model), but it disallows behaviors arising due to shared store buffers and shared write-through caches (which are seen in POWER processors). WMM-S, on the other hand, is more complex and allows these behaviors. We give the operational definitions of both models using

Instantaneous Instruction Execution (I2E), which has been used in the definitions of SC and TSO. We also show how both models can be implemented using conventional cache-coherent memory systems and out-of-order processors, and encompasses the behaviors of most known optimizations.

Participants

- Jade Alglave
University College London, GB
- Giovanni Tito Bernardi
University Paris-Diderot, FR
- Annette Bieniusa-Middelkoop
TU Kaiserslautern, DE
- Richard Bornat
Middlesex University –
London, GB
- Stephen Brookes
Carnegie Mellon University –
Pittsburgh, US
- Simon Castellan
ENS – Lyon, FR
- Andrea Cerone
Imperial College London, GB
- Pierre Clairambault
ENS – Lyon, FR
- Patrick Cousot
New York University, US
- Andrei Marian Dan
ETH Zürich, CH
- Will Deacon
ARM Ltd. – Cambridge, GB
- David Delmas
Airbus S.A.S. – Toulouse, FR
- Delphine Demange
IRISA – Rennes, FR
- Stephan Diestelhorst
ARM Ltd. – Cambridge, GB

- Charles Anthony Richard
Hoare
Microsoft Research UK –
Cambridge, GB
- Vincent Jacques
University College London, GB
- Bernhard Kragl
IST Austria –
Klosterneuburg, AT
- Ori Lahav
MPI-SWS – Kaiserslautern, DE
- Daniel Lustig
NVIDIA Corp. –
Santa Clara, US
- Yatin Manerkar
Princeton University, US
- Luc Maranget
INRIA – Paris, FR
- Paul McKenney
IBM – Beaverton, US
- Paul-Andre Mellies
University Paris-Diderot, FR
- Roland Meyer
TU Braunschweig, DE
- Maged M. Michael
Facebook – New York, US
- Antoine Miné
CNRS and University Pierre &
Marie Curie – Paris, FR

- Vincent Nimal
Microsoft Research UK –
Cambridge, GB
- Andrea Parri
INRIA – Paris, FR
- Gustavo Petri
University Paris-Diderot, FR
- Susmit Sarkar
University of St. Andrews, GB
- Helmut Seidl
TU München, DE
- Suzanne Shoaraee
ARM France SAS –
Sophia-Antipolis, FR
- Daryl Stewart
ARM Ltd. – Cambridge, GB
- Caroline J. Trippel
Princeton University, US
- Caterina Urban
ETH Zürich, CH
- Viktor Vafeiadis
MPI-SWS – Kaiserslautern, DE
- Derek Williams
IBM Research Lab. – Austin, US
- Glynn Winskel
University of Cambridge, GB
- Sizhuo Zhang
MIT – Cambridge, US

QoE Vadis?

Edited by

Markus Fiedler[1], Sebastian Möller[2], Peter Reichl[3], and Min Xie[4]

1   **Blekinge Institute of Technology – Karlskrona, SE,** `markus.fiedler@bth.se`
2   **TU Berlin, DE,** `sebastian.moeller@tu-berlin.de`
3   **Universität Wien, AT,** `peter.reichl@univie.ac.at`
4   **Telenor Research – Trondheim, NO,** `min.xie@telenor.com`

──── Abstract ──

This report documents the program and the outcomes of Dagstuhl Perspectives Workshop 16472
"QoE Vadis?", which was preceded by the three Dagstuhl Seminars 09192 "From Quality of Service to Quality of Experience" (2009), 12181 "Quality of Experience: From User Perception to Instrumental Metrics" (2012), and 15022 "Quality of Experience: From Assessment to Application" (2015). As suggested by the name, the Perspectives Workshop set out to jointly and critically reflect on future perspectives and directions of Quality of Experience (QoE) research. This report reflects upon the organization of the workshop. It also provides a set of personal statements and feedbacks (through the innovative "Advocatus Diaboli" approach), as well as a marriage proposal with the area of User Experience (UX). Finally, an overview of the recommendations in the upcoming Dagstuhl Manifesto is given.

1   Executive Summary

Markus Fiedler
Sebastian Möller
Peter Reichl
Min Xie

During the recent decade, the transition from the technology-oriented notion of QoS (Quality of Service) to the user-centric concept of QoE (Quality of Experience) has become an important paradigm change in communication networking research. Simultaneously, the field of QoE as such has significantly developed and matured. This is amongst others reflected in the series of three Dagstuhl Seminars 09192 "From Quality of Service to Quality of Experience" (2009), 12181 "Quality of Experience: From User Perception to Instrumental Metrics" (2012), and 15022 "Quality of Experience: From Assessment to Application" (2015).

The QoE-related Dagstuhl Seminars had a significant impact on the understanding, definition and application of the QoE notion and concepts in the QoE community, for instance with respect to redefining fundamental concepts of quality. That work was performed in close collaboration with the COST Action IC1003 Qualinet [1] that has been concentrating on QoE in multimedia systems and services, and is still actively convening experts from all over the world to regular meetings and exchanges. In particular, this collaboration has led to the widely regarded Qualinet White Paper on "Definitions of QoE and related concepts" [1] and to the launch of a new journal entitled "Quality and User Experience" [2], fostering the scientific exchange within and between QoE and User Experience (UX) communities.

Realising the urgent need of jointly and critically reflecting the future perspectives and directions of QoE research, the QoE-related Dagstuhl Seminars were complemented by the present Dagstuhl Perspectives Workshop 16472 "QoE Vadis?", whose output is compiled in a Dagstuhl Manifesto. Besides of having brought together the two communities much closer, and besides triggering new events such as special sessions at conferences, the main outcome of the workshop has been concretized in terms of 11 recommendations to be communicated to stakeholders in the QoE and UX domains.

The workshop was organised around the writing process of the Manifesto draft: Starting from personal statements instead of talks, two sets of group works were arranged, whose output was criticially reviewed by "Advocatii Diaboli" and then refined and extended. A final review round by one representative of each the QoE and the UX group was performed before the Manifesto draft was completed by the end of the week.

References
1   European Network on Quality of Experience in Multimedia Systems and Services (COST IC 1003 Qualinet), http://www.qualinet.eu (last seen 2017-02-24).
2   Quality and Experience (QUEX), a journal published by Springer, http://link.springer.com/journal/41233 (last seen 2017-02-24).

2 Table of Contents

3 Organization and Inputs

3.1 Organization

Sebastian Möller (TU Berlin, DE) and Markus Fiedler (Blekinge Institute of Technology – Karlskrona, SE)

The 5-day workshop was organized as follows:

Monday: In the morning, an introductory session briefly explained goals and structure of the seminar, as well as the main work items to be focused on during the coming days:

1. a state-of-the-art and SWOT analysis of the current research landscape for QoE;
2. a set of anticipations how the area of QoE might develop in the future;
3. a set of theses on how the areas of QoE research will lead to innovative and improved products and services; and
4. a set of recommendations for future funding of these areas.

Then, brief personal introductions and personal statements (see Section 3.2) were given by all participants on the basis of a 1-slide 3-minute presentation. On the afternoon of that same day, the first group work was laid out and started. Group work was organized in two parallel groups and circled around the expected topics of the manifesto. The first session addressed the first two chapters of the manifesto, with marginal overlap between the groups.

Tuesday: The group work continued on the first half of the second day, followed by a social event, involving a visit of the UNESCO World Cultural Heritage Völkinger Hütte and a dinner in Saarlouis with focus on local food.

Wednesday: The results of the first group work were presented on Wednesday morning, followed by an "Advocatus Diaboli" (AD) session where the main results of the group works were thoroughly criticised, and suggestions for improvement were given to the two groups by senior members of the community (see Section 4). In the afternoon, the second group work started, again in two parallel groups, this time addressing the last two chapters of the manifesto.

Thursday: The group work was again continued on the first half of Thursday, followed again by another AD session. Following this, the groups compiled their final output text. On Thursday evening, a concert titled "Tribute to Tosti" was given by two of the co-organisers (Peter Reichl, song/piano and Markus Fiedler, piano).

Friday: The output of the group work was presented and challenged by one member of the QoE community and one member of the UX community, respectively. After departure of most of the workshop participants on Friday at lunchtime, the organizers and a set of supporters started compiling a first draft of the manifesto, which was handed over for further editing and collaborative optimization to an editing platform.

3.2 Collection of personal statements

Markus Fiedler (Blekinge Institute of Technology – Karlskrona, SE)

License Creative Commons BY 3.0 Unported license
© Markus Fiedler
Joint work of All participants of this Perspectives Workshop

Before the workshop, the delegates were asking to provide personal statements regarding the main contributions to the Manifesto to be addressed by the seminar, with regards to the questions listed in Section 3. These personal statements were presented during the personal three-minute presentations on the first day, alternatively upon arrival of the participants. This section contains an excerpt of those statements (in anonymous form), grouped around three topics: (1) fundamentals; (2) applications; and (3) values.

3.2.1 Fundamentals

1. Get to know the basics: What is quality perception?
2. New models and standards focusing individuals!
3. Role of IoT and wearable devices in measuring QoE
4. Building on top of what has been achieved in QoE and reinvent a modern version of QoL [Quality of Life].
5. UX vs. QoE: Root causes of the divide? Why is UX mode ubiquitous? Can we go the same road?
6. Key proposals are to marry QoE [and] user behavior using machine learning, how data can be made available from real services and encouragement to tackle the challenge with increasing number of services.
7. Complementary concepts and their relations: QoE, engagement, acceptance.
8. QoE has to join forces with UX research more in the near future.
9. We should change the paradigm from passive media consumption to interactive services, in which the user plays a crucial role for defining QoE by behaving in a certain way. For this, we will need to understand, formalize and simulate user behavior.
10. When a management action about network services is being performed on certain part of our network (e.g., traffic management within an autonomous system), what will be its direct or indirect impact on those social interactions being delivered via the affected or, in some cases, other autonomous systems? If the impacts on social interactions can be quantified, can we then measure and compare the effectiveness and quality of the management decision options? Second, with our near real-time 24/7 information exchanging system, can a network operation and management system leverage this powerful platform to close the gap between network operations and the demands from their customers?
11. How to develop practical QoE solutions applicable in large-scale networks?
12. How to integrate QoE with adjacent areas such as data measurement and machine learning technologies to serve future networks?
13. How to model, quantify, and engage QoE for existing and future networks and services?

3.2.2 Applications

1. Which application (area) next?
2. How to get out of the media (consumption) corner?

3. Candidate application area: Quality of Work Experience. What does QoE contribute more than UX and / or work psychology?

4. Get out of the multimedia comfort zone! There are so many more application domains where QoE provisioning in the sense of "avoid annoyance, create delight" could make a difference.

5. We should address services which adapt themselves to the behavior of the user. In such services, neither the service nor the user are constant, they adapt and learn from each other. QoE then becomes a floating target, which needs to be taken into account when assessing or predicting it.

6. We should extend QoE to services which deviate from a classical human-computer interaction paradigm, such as location-based services, interactions with invisible interfaces in smart environments, and alike. Such services may be relevant only in specific situations of the real life, thus QoE can only be addressed in the field.

7. It's been just a first step: The role of QoE for the future "Internet of People" (IoP). The transition from QoS to QoE has been hailed as a paradigm change enabling quality models for network services that are aligned to real human needs rather than purely technical parameters. With the advent of the Internet of Things (IoT), however, this step forward will immediately be threatened again, along with a broad range of further issues of high subjective importance for the end users, like privacy and network security. As a consequence, we need a much broader move towards an "Internet of People" (IoP) where humans are put more and more into the center of the Internet design, based on the increasing pervasiveness of personal mobile devices and the decentralisation of functions. Hence, key decision like content management or service provisioning have to be taken locally at my own device which, consequently, becomes my "proxy" while interacting with other devices/proxies. This evolution will have major impact on the way network, data and service management protocols are designed, and at the same time generalize the multidisciplinary path QoE has taken in the last decade towards a much broader human-centric concept for future communication networks.

8. A key challenge with respect to optimizing and managing QoE is to be able to effectively monitor relevant KPIs and determine root causes of QoE degradation. While many solutions exist, questions still remain as to what extend existing solutions are effective, and whether there is a need for new QoE monitoring tools in the context of different service scenarios (e.g., network probes that estimate QoE for encrypted OTT services, conversational services, etc)?

9. Finally, going beyond today's audio/visual communication services, future communication (conversational) services will likely move towards further immersive and interactive services that utilize advances in technologies such as VR/AR/3D, multi-sensory devices, etc. What key challenges do we need to address both from a user perspective as well as a networking perspective, for such services to become a reality? In particular, for such emerging services we are missing models relating QoS to QoE. Need for a multidisciplinary approach!

3.2.3   Value

1. Determine the business value of QoE provisioning! This will be mandatory for the future success of the area, both in terms of funding and adoption.

2. What's next? As a next step, the QoE eco-system needs to be considered. A couple of questions emerge from the QoE eco-system. Can we utilize QoE for network & service management? Or are complementary concepts more appropriate like user engagement or acceptance? Or is context information more valuable in QoE management? How to

compare QoE across apps and for different user groups? Which metrics are appropriate when looking at QoE (quantiles vs. MOS) or QoE fairness? How to transform QoE into business models, SLAs, etc.? Or can we trade QoE?

3. Who needs QoE (and QoE research)? The different stakeholders involved in the delivery of services and apps need to be considered. The question arises whom of them will benefit from QoE (research and concepts) and how. To advance in the field of QoE, there is a need to tackle QoE models for any kind of app (e.g. cloud services, IoT, ?) or even QoE models for data. This may require to integrate aspects like privacy, but also accuracy, pricing, etc. Complementary concepts like engagement or acceptance may be more appropriate, e.g. for monitoring in an encrypted Internet.

4. QoE framework from QoS to well-being and business profit

5. QoE needs to take application reliability and economics into consideration.

6. What are the ways in which we can exploit the move from QoS to QoE management? This question needs to be addressed from different perspectives: (a) network operator: What metrics are important? Is there a need for new QoE-driven network resource allocation mechanisms, algorithms? Will this reduce customer churn? Operator role in the context of OTT service delivery? (b) service provider: how to design QoE-driven service adaptation strategies meeting heterogeneous end user capabilities, requirements, and expectations? (c) end-user: what quality gains do users care about? What are user expectations with respect to service quality? To what extent are users willing to pay for quality?

7. With new networking paradigms emerging, such as virtualized networks and everything in the cloud, what is the position and potential of QoE management? What business models will dominate when considering different players involved in service delivery, and how will QoE be addressed in the context of these business models?

8. There are two main types of challenges when thinking about QoE management, one related to the technical aspects of how to implement it (which I would guess is more interesting to the audience of this seminar), and another one related to the business and political aspects of getting the different stakeholders involved cooperating. Depending on which field of research one hails from, either type of challenge can be more interesting, but, if QoE management is to take off in practice, both problem areas need solving.

9. Regarding the technical aspects, the progression of networks and services towards the cloud dictates a set of (quite interesting) technical challenges. Among these we can identify the definition of suitable SDN applications (and maybe even data-plane functionality, using e.g., P4) that enable QoE management. In the second area, we find business challenges, such as finding incentives and mechanisms for OTTs and telcos to cooperate, but also legal ones, such as the tighter net neutrality regulations in Europe.

10. A non-scientific but nonetheless relevant challenge to research on this area is that of funding. At least in Finland, it has been very hard to get any traction with the public funding authority, and even companies that are interested in the concepts, seem reluctant to invest in it. A valid question is whether the topic will actually leave the academic environment at some point (one would hope so, but we'll have to wait and see)

11. What are the benefits of QoE for customers and industry?

4   "Advocati Diaboli" feedbacks

In this section, we present some selected feedback from the "Advocatus Diaboli" (AD) sessions, in order to give an impression of how the AD concept was realized:

4.1   "Advocatus Diaboli" feedback, first example

Henning Schulzrinne (Columbia University – New York, US)

The AD reviewed the outcome of the first group work of group 1.

The first group of question regards the State-of-the-Art (SOTA) analysis.

- What's the scope of QoE? The definitions include applications TripAdvisor and RateMyProfessors. Is it telecommunications? Entertainment?
- QoE seems to be on relatively simple information transmission channels, *i.e.* fidelity of transmission. Does this exclude more complex interactions? What is the temporal scope? It seems to imply single sessions.
- What about predictability of experience?
- Producer vs. consumer – does this cover the OTT world, *i.e.* software vendors, "channel" providers, equipment vendors (smartphones). This means controllable and external, uncontrollable factors, such as "feelings" (did I like this?), performance and "actions" (will I drop the service? how much am I willing to pay?)

From a broad(er) perspective, it was observed that:

- QoE shares the "QoS problem" , which is managing scarcity: make it just tolerable enough that people won't leave [Tragedy of the Commons].
- Where is this still relevant? In home networks? Or in mobile networks?
- Or is it about improving experiences? What about UX? Does it offer a more positive message?
- Is there something distinctive in approach, tools and methods?

The second set of feedbacks addressed the SWOT analysis.

From an internal perspective:

- What would a QoMEX technical program committee think about a new paper?

With regard to weaknesses:

- Is QoE treated as "free" goods? The price-performance trade-off not easily considered.
- Evaluation ("sucks") vs. iterative refinement.
- Which are 10 key insights? What about the "end-to-end principle"?

From an external perspective:

- Why should I (dean) hire a new faculty in that area as opposed to (say) cybersecurity?
- Why should I (funding manager) spend money on this as opposed to cancer research?

From an outcomes perspective:

- Who exactly benefits from more research, beyond the researchers?
- In the old days: easy – codec and network designers
- Is this something that every [fill in the blank] should know? Who? Why?
- Should this become something that's part of the system design process? Where? Cf. "security by design" equivalence or "test-driven development?"

From a cost perspective:

- "Might not be cost effective"? Why? What is the cost? Research? "You don't have a choice anyway, so suck it up"? signed, your friendly ISP.

Last but not least, some comments about new research topics:

- What makes IoT and e-health different from the existing topics?
- Why do we have differentiating experience and insights compared to people who understand work places?
- What is the channel, "noise", evaluation metric? How does this overlap or differ from HCI, usable security, user interface design, . . .
- What would be part of a theoretical framework?
- More Dagstuhl seminars on QoE definitions?

4.2 "Advocatus Diaboli" feedback, second example

Sebastian Möller (TU Berlin, DE)

The AD provided input to the group work of group 2 regarding 2 questions. The first question addressed was the one regarding how QoE research would develop in the future. The following comments were given:

1. It was stated that system instrumentation is needed, e.g. for measuring interaction. This point seems to be very valid, however new questions arise from it, namely:
   - For which purpose is this necessary? For system adaptation and/or optimization? For service priorization? Or for the prediction of QoE/UX/user state/business impact?
   - What type of data is to be collected? User interaction with the system? Should user feedback be captured with a feedback button?
   - How can privacy be maintained with all this data collection?
2. One of the guiding ideas of the Dagstuhl seminar was to bring the QoE and UX communities, as well as the research fields, together. To be effective in this aim, the following questions need to be clearly answered:
   - What is the difference between QoE and UX?
   - What can QoE research bring to UX, and vice versa?
   - Which tools or best practices are helpful for this aim? Can we automate user research by a "semantic layer"?
3. User modelling was mentioned as a necessary and helpful step. The following questions arise:
   - What aspects of the user should be modelled? QoE perception and judgment? User behavior, and (if yes) at which level?
   - What is the purpose of the modelling? Adaptation? And what is the outcome? MOS?

The second block of feedback from the AD to group 2 addressed the recommendations to be given as an outcome of the workshop. This related tio the following recommendations proposed by group 2:

1. Societal impact:
   - Are open data and privacy conflicting goals? If not, how can these be addressed?
   - Is a "coherent, functional" theory of QoE and UX a requirement? Why? Can such a theory be expected, given the diversity and complexity of the services?

2. Business impact:
   - "QoE and UX will be key aspects for the adoption of new technologies": What is the link between QoE/UX and acceptance?
3. Academic/scientific impact:
   - If the question is still "where are we?", is this then a good topic to educate Master students in?
   - Collaboration: UX has been approached by QoE for a while; what are the showstoppers?
4. Industry impact
   - Is it reasonable to think of a "QoE-collecting framework that could be used in a wide range of applications?"

5   Outputs

5.1   The QoE-UX wedding proposal

Marianna Obrist (University of Sussex – Brighton, GB)

Obviously, QoE and User Experience (UX) are very closely related. Thus, the idea came up to "marry" both areas. The question arises: What does each partner bring to the marriage?

1. *A set of unique features*: While QoE brings the quantification of users (audio-visual) quality, UX brings qualitative and quantitative methods. QoE contributes measurement tools and models technical efficiency, while UX contributes individual and social UX models. Main factor studied by QoE are latency, accuracy, synchronization, etc., in contrast to pragmatic and hedonic qualities by UX. QoE extends a.o. towards user perception and Mean Opinion Scores (MOS), while UX extends towards user acceptance, trust, safety, emotions, wow as well as engagement, fun, flow, immersion, presence. QoE contributes standards, and UX contributes design guidelines and principles. QoE brings technical expertise with a good ability to talk to businesses, while UX brings expertise on understanding users, their interaction, and needs.
2. *A set of aspirations*: QoE aims at adding more subjective user data, automation, and interactivity (interaction patterns), while UX would like to be able to measure the added value of UX and to have numbers. QoE intends to improve users quality of life, and to make users and businesses happy. Similarly, UX would like to improve users personal wellbeing (long term value), and make users (and society) happy.
3. *Different, complementary stakeholders*: For QoE, these are infrastructure providers and operators; content providers and creators, while for UX, these are consumers, designers, and artists.

Indeed, a combination of QoE and UX will make a difference in the fields of multimedia/entertainment/gaming; IoT/wearable interfaces; and multisensory interaction.

5.2 Special Session "QoE Vadis?" at QoMEX 2017

Markus Fiedler (Blekinge Institute of Technology – Karlskrona, SE) and Marianna Obrist (University of Sussex – Brighton, GB)

As an immediate follow-up of the Perspectives Workshop, we proposed and got accepted a Special Session at the predominant conference in the QoE area, QoMEX [1].

Motivation and objective of the Special Session

This session is based on the recent Dagstuhl Perspectives Workshop "QoE Vadis?", with a special focus on contributions on future directions of QoE research beyond the multimedia comfort zone. Of particular interest are amongst others the exploration of services beyond multimedia consumption, investigation of new emerging technologies such as VR and AR, and stimulation of discussions on the design of multisensory experiences through opening up partnerships with adjacent research areas such as User Experience (UX) and Human-Computer Interaction (HCI), which attract huge investments by major players such as Apple, Google, etc. Pressing questions are amongst others: (i) How cross- and multi-disciplinary design perspectives can improve the quality of service creation, delivery and perception? (ii) What new processes, methods and measures can be applied to ensure the role of QoE in innovation? (iii) What new business and societal values can be envisaged for QoE research? Overall, this session shall provide a forum for researchers and practitioners from industry to look into the future of QoE research, joining forces with experts from other communities such as UX, HCI, as well as Economics, Management, etc.

Session outline

- Keynote (candidate: David Geerts, University of Leuven, Belgium)
- 2 sessions with 5 papers in total
- Short panel on "The marriage of QoE and UX: what do both communities bring into the partnership?"

Specific topics of interest

- The Devil's Advocate?s view on different aspects of QoE research
- Value of QoE research beyond multimedia consumption
- New applications and services relevant to QoE research
- New approaches and tools relevant to QoE research (e.g. Big Data analytics)
- Joint efforts between QoE and User Experience research
- Relationship between qualitative and quantitative data, user behavior and churn
- QoE by design for future and emerging technologies (e.g., Virtual and Augmented Reality)
- Business and societal aspects of QoE
- Innovation through QoE research
- Industrial aspects of QoE
- Lifecycle aspects (e.g. requirements engineering and upgrade strategies) for services
- Application of QoE research results "in real life"

References

1 9th International Conference on Quality of Multimedia Experience (QoMEX 2017), May 31– June 2, 2017, Erfurt, Germany, http://www.qomex2017.org (last seen 2017-02-24).

5.3 A glance at the upcoming Dagstuhl Manifesto

Sebastian Möller (TU Berlin, DE)

The current manifesto draft is a 19-page document which has been started during the Dagstuhl workshop, compiled in several post-workshop sessions, and re-iterated with all participants of the workshop in order to ensure correctness and readability. It is structured as follows: Section 2 provides a state-of-the-art and SWOT analysis of the current research landscape for QoE. Section 3 contains projections of how the area of QoE might develop in the future, and how it will lead to innovative and improved products and services. Finally, Section 4 provides a set of recommendations for future funding of QoE-related activities.

The main outcome of the manifesto is a list of 11 recommendations to stakeholders in the QoE and UX communities. They may be summarized as follows:

- R0: Put the end user into the focus of all your considerations.
- R1: Promote interdisciplinary research.
- R2: Provide access to open data and tools.
- R3: Drive investigation beyond the comfort zone.
- R4: Turn QoE from reactive to proactive research.
- R5: Implement mechanisms for direct quality feedback.
- R6: Join forces within industry.
- R7: Support QoE research as scientific approach to a substantial and unsolved problem.
- R8: Respect QoE as key paradigm for the future digital society.
- R9: Create a cross-disciplinary and cross-institutional research community.
- R10: Support market diversity and sustainability.

Participants

- Jan-Niklas Antons
  TU Berlin, DE
- Luigi Atzori
  University of Cagliari, IT
- Katrien De Moor
  NTNU – Trondheim, NO
- Touradj Ebrahimi
  EPFL – Lausanne, CH
- Sebastian Egger-Lampl
  AIT Austrian Institute of
  Technology – Wien, AT
- Markus Fiedler
  Blekinge Institute of Technology –
  Karlskrona, SE
- Jörgen Gustafsson
  Ericsson Research – Luleå, SE
- Tobias Hoßfeld
  Universität Duisburg-Essen, DE
- Lucjan Janowski
  AGH Univ. of Science &
  Technology – Krakow, PL

- Kalevi Kilkki
  Aalto University, FI
- Udo Krieger
  Universität Bamberg, DE
- Effie Lai-Chong Law
  University of Leicester, GB
- Sebastian Möller
  TU Berlin, DE
- Marianna Obrist
  University of Sussex –
  Brighton, GB
- Peter Reichl
  Universität Wien, AT
- Virpi Hannele Roto
  Aalto University, FI
- Henning Schulzrinne
  Columbia University –
  New York, US
- Lea Skorin-Kapov
  University of Zagreb, HR

- Jan Van Looy
  Ghent University, BE
- Martín Varela
  VTT Technical Research Centre
  of Finland – Oulu, FI
- Katarzyna Wac
  University of Geneva, CH
- Felix Wu
  University of California –
  Davis, US
- Min Xie
  Telenor Research –
  Trondheim, NO
- Hans-Jürgen Zepernick
  Blekinge Institute of Technology –
  Karlskrona, SE

New Directions for Learning with Kernels and Gaussian Processes

Edited by

Arthur Gretton[1], Philipp Hennig[2], Carl Edward Rasmussen[3], and Bernhard Schölkopf[4]

1   **University College London, GB,** `arthur.gretton@gmail.com`
2   **MPI für Intelligente Systeme – Tübingen, DE,** `ph@tue.mpg.de`
3   **University of Cambridge, GB,** `cer54@cam.ac.uk`
4   **MPI für Intelligente Systeme – Tübingen, DE,** `bs@tue.mpg.de`

―――― **Abstract** ――――――――――――――――――――――――――――――――――

The Dagstuhl Seminar on 16481 "New Directions for Learning with Kernels and Gaussian Processes" brought together two principal theoretical camps of the machine learning community at a crucial time for the field. Kernel methods and Gaussian process models together form a significant part of the discipline's foundations, but their prominence is waning while more elaborate but poorly understood hierarchical models are ascendant. In a lively, amiable seminar, the participants re-discovered common conceptual ground (and some continued points of disagreement) and productively discussed how theoretical rigour can stay relevant during a hectic phase for the subject.

Seminar November 27 to December 2, 2016 – http://www.dagstuhl.de/16481
1998 ACM Subject Classification G.1.1 Interpolation, G.1.2 Approximation, G.3 Probability and Statistics, I.2.6 Learning
Keywords and phrases gaussian processes, kernel methods, machine learning, probabilistic numerics, probabilistic programming
Digital Object Identifier 10.4230/DagRep.6.11.142
Edited in cooperation with Michael Schober

1 Summary

Arthur Gretton
Philipp Hennig
Carl Edward Rasmussen
Bernhard Schölkopf

Machine learning is a young field that currently enjoys rapid, almost dizzying advancement both on the theoretical and the practical side. On account of either, the until quite recently obscure discipline is increasingly turning into a central area of computer science. Dagstuhl seminar 16481 on *"New Directions for Learning with Kernels and Gaussian Processes"* attempted to allow a key community within machine learning to gather its bearings at this crucial moment in time.

Positive definite kernels are a concept that dominated machine learning research in the first decade of the millennium. They provide infinite-dimensional hypothesis classes

that deliver expressive power in an elegant analytical framework. In their probabilistic interpretation as Gaussian process models, they are also a fundamental concept of Bayesian inference:

A *positive definite kernel* $k : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ on some input domain $\mathbb{X}$ is a function with the property that, for all finite sets $\{x_1, \dots, x_N\} \subset \mathbb{X}$, the matrix $K \in \mathbb{R}^{N \times N}$, with elements $k_{ij} = k(x_i, x_j)$, is positive semidefinite. According to a theorem by Mercer, given certain regularity assumptions, such kernels can be expressed as a potentially *infinite* expansion

$$k(x, x') = \sum_{i=1}^{\infty} \lambda_i \phi_i(x) \phi_i^*(x'), \qquad \text{with} \qquad \sum_{i=1}^{\infty} \lambda_i < \infty, \tag{1}$$

where $*$ is the conjugate transpose, $\lambda_i \in \mathbb{R}_+$ is a non-negative *eigenvalue* and ϕ_i is an *eigenfunction* with respect to some measure $\nu(x)$: a function satisfying

$$\int k(x, x') \phi_i(x) d\nu(x) = \lambda_i \phi_i(x'). \tag{2}$$

Random functions $f(x)$ drawn by independently sampling Gaussian weights for each eigenfunction,

$$f(x) = \sum_{j=1}^{\infty} f_j \phi_j(x) \qquad \text{where} \qquad f_j \sim \mathcal{N}(0, \lambda_i), \tag{3}$$

are draws from the centered *Gaussian process* (GP) $p(f) = \mathcal{GP}(f; 0, k)$ with *covariance function* k. The logarithm of this Gaussian process measure is, up to constants and some technicalities, the square of the norm $\|f\|_k^2$ associated with the *reproducing kernel Hilbert space* (RKHS) of functions reproduced by k.

Supervised machine learning methods that *infer* an unknown function f from a data set of input-output pairs $(X, Y) := \{(x_i, y_i)\}_{i=1,\dots,N}$ can be constructed by minimizing an empirical risk $\ell(f(X); Y)$ regularized by $\|\cdot\|_k^2$. Or, algorithmically equivalent but with different philosophical interpretation, by computing the *posterior* Gaussian process measure arising from conditioning $\mathcal{GP}(f; 0, k)$ on the observed data points under a likelihood proportional to the exponential of the empirical risk.

The prominence of kernel/GP models was founded on this conceptually and algorithmically compact yet statistically powerful description of inference and learning of nonlinear functions. In the past years, however, hierarchical ('deep') parametric models have bounced back and delivered a series of impressive empirical successes. In areas like speech recognition and image classification, deep networks now far surpass the predictive performance previously achieved with nonparametric models. One central goal of the seminar was to discuss how the superior adaptability of deep models can be transferred to the kernel framework while retaining at least some analytical clarity. Among the central lessons from the 'deep resurgence' identified by the seminar participants is that the kernel community has been too reliant on theoretical notions of universality. Instead, representations must be learned on a more general level than previously accepted. This process is often associated with an 'engineering' approach to machine learning, in contrast to the supposedly more 'scientific' air surrounding kernel methods. But its importance must not be dismissed. At the same time, participants also pointed out that deep learning is often misrepresented, in particular in popular expositions, as an almost magic kind of process; when in reality the concept is closely related to kernel methods, and can be understood to some degree through this connection: Deep models provide a hierarchical parametrization of the feature functions $\phi_i(x)$ in terms of a finite-dimensional family. The continued relevance of the established theory for kernel/GP models

hinges on how much of the power of deep models can be understood from within the RKHS view, and how much new concepts are required to understand the expressivity of a deep learning machine.

There is also unconditionally good news: In a separate but related development, kernels have had their own renaissance lately, in the young areas of probabilistic programming ('computing of probability measures') and probabilistic numerics ('probabilistic descriptions of computing'). In both areas, kernels and Gaussian processes have been used as a descriptive language. And, similar to the situation in general machine learning, only a handful of comparably simple kernels have so far been used. The central question here, too, is thus how kernels can be designed for challenging, in particular high-dimensional regression problems. In contrast to the wider situation in ML, though, kernel design here should take place at compile-time, and be a structured algebraic process mapping source code describing a graphical model into a kernel. This gives rise to new fundamental questions for the theoretical computer science of machine learning.

A third thread running through the seminar concerned the internal conceptual schism between the probabilistic (Gaussian process) view and the statistical learning theoretical (RKHS) view on the model class. Although the algorithms and algebraic ides used on both sides overlap *almost* to the point of equivalence, their philosophical interpretations, and thus also the required theoretical properties, differ strongly. Participants for the seminar were deliberately invited from both "denominations" in roughly equal number. Several informal discussions in the evenings, and in particular a lively break-out discussion on Thursday helped clear up the mathematical connections (while also airing key conceptual points of contention from either side). Thursday's group is planning to write a publication based on the results of the discussion; this would be a highly valuable concrete contribution arising from the seminar, that may help drawing this community closer together.

Despite the challenges to some of the long-standing paradigms of this community, the seminar was infused with an air of excitement. The participants seemed to share the sensation that machine learning is still only just beginning to show its full potential. The mathematical concepts and insights that have emerged from the study of kernel/GP models may have to evolve and be adapted to recent developments, but their fundamental nature means they are quite likely to stay relevant for the understanding of current and future model classes. Far from going out of fashion, mathematical analysis of the statistical and numerical properties of machine learning model classes seems slated for a revival in coming years. And much of it will be leveraging the notions discussed at the seminar.

2 Table of Contents

3   Overview of Talks

3.1   Random Fourier Features for Operator-Valued Kernels

Florence d'Alché-Buc (Telecom ParisTech, FR)

Devoted to multi-task learning and structured output learning, operator-valued kernels provide a flexible tool to build vector-valued functions in the context of Reproducing Kernel Hilbert Spaces. To scale up operator-valued kernel-based regression devoted to multi-task and structured output learning, we extend the celebrated Random Fourier Feature methodology to get an approximation of operator-valued kernels. We propose a general principle for Operator-valued Random Fourier Feature construction relying on a generalization of Bochner's theorem for shift-invariant operator-valued Mercer kernels. We prove the uniform convergence of the kernel approximation for bounded and unbounded operator random Fourier features using appropriate Bernstein matrix concentration inequality. Numerical experiments show the quality of the approximation and the efficiency of the corresponding linear models on multiclass and regression problems.

3.2   Practical Challenges of Gaussian Process Applications

Marc Deisenroth (Imperial College London, GB)

In many applications, we face practical challenges with Gaussian processes and kernel methods. For example, in robotics and personalized healthcare, data- efficient learning (i.e., learning from small data sets) is critical. We can achieve this in multiple ways, e.g., by carefully modeling uncertainty in the model and the inference, transfer learning or the incorporation of structural priors. Focusing on uncertainty representation, it is critical to propagate uncertainty through a (Gaussian process) system, which is computationally expensive (training may not be the computational bottleneck). Other applications include the optimization (or learning) of simulators of very expensive experiments (e.g., LHC, bioprocesses or neotissue engineering). Challenge we face are high-dimensional optimization problems and scalability in the number of data points. Generally, scalability seems to be a general problem, and we should think about scale-free model architectures, inference and the software that allows us to perform distributed computing.

3.3   Deep kernels and deep Gaussian processes

David Duvenaud (Toronto, CA)

To suggest better neural network architectures, we analyze the properties different priors on compositions of functions.

We showed how we can construct deep kernels by composing their implicit features, and examine the properties of such kernels as we increase their depth.

We then showed how such models are different from deep Gaussian processes, and by visualizing draws from deep GP priors examined their properties as a function of depth.

Finally, we show that you get additive covariance if you do dropout on Gaussian processes.

3.4   Finding Galaxies in the Shadows of Quasars with Gaussian Processes

Roman Garnett (Washington University – St. Louis, US)

I discussed recent application of Gaussian processes to a problem from astrophysics: detecting damped Lyman-α absorbers in lines of sight to quasars. DLAs represent proto-galaxies in the ancient universe and their distribution is of interest to cosmology. The state of the art for detecting DLAs is visual inspection; however we show we can construct an automated method via Bayesian model selection, with GPs as our models of spectroscopic data. We use a dataset of $\sim 50,000$ quasar observations from SDSS-III to derive a custom "quasar kernel". The learned kernel has structure markedly different from off-the-shelf kernels. Performance on the detection task relied critically on this structure. Finally, I pointed out I had to rely on quasi- Monte Carlo to estimate model evidence for the DLA model because the integrand had dynamic range on the order of ~ 6000 nats. No off-the-shelf model can handle such data.

3.5 Comparing samples from two distributions

Arthur Gretton (University College London, GB)

We provide an overview of kernel approaches to comparing distributions. The focus is on choosing the function class, and adapting the test statistc, so as to maximize the power of the associated tests.

We begin with an introduction to embeddings of probabilities to a reproducing kernel Hilbert space (RKHS), where an embedding is simply the expectation of the kernel function that defines the RKHS. We demonstrate that the difference in these embeddings can be interpreted as an integral probability metric, called the Maximum Mean Discrepancy (MMD). This statistic can be used in a test of homogeneity, where two samples are observed, and the null hypothesis is that both samples are drawn from the same distribution.

The power of a statistical test based on the MMD will depend on the particular RKHS used. We show that the asymptotic distribution of the statistic is Gaussian under the alternative, and an infinite sum of weighted chi squared variables under the null. Since the null distribution has faster shrinking variance, it is shown that the kernel maximizing the test power is the one which gives the largest ratio of the MMD to its variance (the optimization is performed on a held-out validation set). We demonstrate that this optimized kernel can distinguish between samples from a generative adversarial network, and samples drawn from a reference test set.

An alternative approach to homogeneity testing is to look for maximum of the witness function associated with the MMD, which is a smooth function with largest amplitude where the probability mass of P and Q is most different. We can therefore use the values of the witness function at a particular set of points to construct a test statistic. Our statistic involves normalizing these witness function values by their joint covariance. We may optimize a lower bound on the test power by maximizing the test statistic over the witness point locations on a held-out validation set. We use this test to distinguish positive and negative emotions on a facial expression database, showing that a distinguishing feature reveals the facial areas most relevant to emotion.

Finally, we address the problem of comparing a model to a sample, for instance in the context of statistical model criticism. In this case, the MMD witness function can be modified by a Stein operator, to have zero expectation under the model distribution. The resulting statistic is denoted the Maximum Stein Discrepancy (MSD). This Stein operator can be computed even when the distribution cannot be normalized. We use the MSD to demonstrate the inadequacy of fit of a simple regression model to data with heteroscedastic noise.

3.6 GPs and Kernels for Computation – new opportunities in probabilistic numerics

Philipp Hennig (MPI für Intelligente Systeme – Tübingen, DE)

The computational complexity of machine learning models (not just kernel/GP models) is dominated by numerical tasks: optimization, integration, linear algebra, and the solution of differential algebra. The algorithms we use for these tasks have mostly arrived in our community from other disciplines, such as computational physics, and simulation. Interestingly, these methods can actually be interpreted as active learning algorithms themselves, since they estimate latent/incomputable quantities (e.g. the value of an integral) from observable/computable quantities (e.g. values of the integrand at various, actively chosen nodes). Over recent years, this observation has given rise to a class of numerical methods known as probabilistic numerical algorithms: Methods that take in and return probability measures, rather than point estimates. A string of papers have revealed that many popular and foundational numerical methods can be written as least-squares regression, and thus interpreted as MAP estimators arising from Gaussian probabilistic models. Careful analysis shows that the associated posterior variances can be calibrated at low computational cost, meaning that they provide a meaningful notion of uncertainty in computation. Now, this new framework can be used to build new functionality sorely needed in machine learning: Increased performance through custom prior assumptions; stability of computations under stochastic computations, and new notions of algorithmic safety through statistical hypothesis testing.

3.7 GPy and GPFlow

James Hensman (Lancaster University, GB)

In this talk, I present a live demo of working with the Python-based frameworks GPy and GPFlow.

Some discussion has arisen surrounding the reasons for the success of Deep Learning, and one of the contributing factors is widely agreed to be the availability of Deep Learning software. In this talk I argue that Deep Learning software can easily be adapted to suit kernel methods.

I describe how reverse mode differentiation of the Cholesky algorithm has been added to TensorFlow by Alex Matthews and myself. I then describe GPy and GPflow, two frameworks for Gaussian process computation.

I also present a live demo designed to introduce the audience to the concepts needed to understand TensorFlow, and how to adapt it to their own needs and projects.

3.8 Approximate EP for Deep Gaussian Processes

Joseé Miguel Hernández-Lobato (Harvard University – Cambridge, US)

Main reference T. D. Bui, D. Hernández-Lobato, Y. Li, J. M. Hernández-Lobato, and R. E. Turner, "Deep Gaussian Processes for Regression using Approximate Expectation Propagation", in Proc. of the 33rd Int'l Conf. on Machine Learning (JMLR 2016), Vol. 48, pp. 1472–1481, 2016.
URL http://jmlr.org/proceedings/papers/v48/bui16.pdf

Deep Gaussian processes (DGPs) are multi-layer hierarchical generalisations of Gaussian processes (GPs) and are formally equivalent to neural networks with multiple, infinitely wide hidden layers. DGPs are nonparametric probabilistic models and as such are arguably more flexible, have a greater capacity to generalise, and provide better calibrated uncertainty estimates than alternative deep models. We develop a new approximate Bayesian learning scheme that enables DGPs to be applied to a range of medium to large scale regression problems for the first time. The new method uses an approximate Expectation Propagation procedure and a novel and efficient extension of the probabilistic backpropagation algorithm for learning. We evaluate the new method for non- linear regression on eleven real-world datasets, showing that it always outperforms GP regression and is almost always better than state-of-the-art deterministic and sampling-based approximate inference methods for Bayesian neural networks.

3.9 Modelling Challenges in AutoML

Frank Hutter (Universität Freiburg, DE)

Main reference M. Feurer, A. Klein, K. Eggensperger, J. Springenberg, M. Blum, F. Hutter, "Efficient and Robust Automated Machine Learning", in Proc. of Advances in Neural Information Processing Systems 28 (NIPS 2015), pp. 2962–2970, Neural Information Processing Systems Foundation Inc., 2015.
URL https://papers.nips.cc/paper/5872-efficient-and-robust-automated-machine-learning
URL http://www.ml4aad.org

In this talk, I briefly overviewed recent developments in the field of automated machine learning, which gives rise to very popular applications of Gaussian processes in Bayesian optimization. I then discussed some of the modelling challenges that occur in this field (such as high dimensionality, conditional spaces, large number of data points, heteroscedasticity, large noise, modelling across data sets, and modelling of learning curves) and initial solutions; some of these solutions were based on Gaussian processes, and some were based on random forests and Bayesian neural networks. We then discussed the challenges of treating all of these problems using Gaussian processes.

3.10 Convergence guarantees for kernel-based quadrature

Motonobu Kanagawa (Institute of Statistical Mathematics – Tokyo, JP)

In this talk, I present recent results on kernel-based quadrature. Kernel- based quadrature rules are becoming important in machine learning and statistics, as they achieve super-$\sqrt{n}$ convergence rates in numerical integration, and thus provide alternatives to Monte Carlo integration in challenging settings where integrands are expensive to evaluate or where integrands are high dimensional. These rules are based on the assumption that the integrand has a certain degree of smoothness, which is expressed as that the integrand belongs to a certain reproducing kernel Hilbert space (RKHS). However, this assumption can be violated in practice (e.g., when the integrand is a black box function), and no general theory has been established for the convergence of kernel quadratures in such misspecified settings. In this talk, I explain that it is actually possible to prove that kernel quadratures can be consistent even when the integrand does not belong to the assumed RKHS, i.e., when the integrand is less smooth than assumed. Specifically, I show that one can derive convergence rates that depend on the (unknown) lesser smoothness of the integrand, where the degree of smoothness is expressed via powers of RKHSs or via Sobolev spaces.

3.11 Don't Panic: Deep Learning Methods are Mostly Harmless

Neil D. Lawrence (University of Sheffield, GB)

With the success of deep learning software and a wide variety of succesfull applications deep learning methods seem to be making the transition to a domain of engineering. A challenge is that the potential pitfalls of the deployment of these ideas has not been characterised. Currently empirical results are leading our theoretical understanding. To be a robust engineering discipline deep learning pipelines need to be placed on stronger theoretical foundations. This presents an opportunity for better characterized methods to augment deep learning ideas and prevent us from sucumbing to the pitfalls. The greater interpretability of kernel and GP methods as well as their more elegant mathematical characterization present opportunities in

1. Meta learning and characterisation of the deep deep learning pipeline.
2. Privacy, fairness and transparency.
3. Integration of physical systems with data driven models.
4. Quantifying the value of data.

3.12   Horses

David Lopez-Paz (Facebook – AI Research, US)

I talked about "horses", which are "systems that do not address the problem that they seem to be addressing". In particular, our current machine learning solutions are horses, since they are vulnerable to slight mismatches between training and testing data (domain adaptation, adversarial perturbation, etc.).

Taming horses is the biggest challenge for machine learning. More specifically, machine learning has a predictive focus (minimize the loss $L(y, y')$ between the true targets $y = f(x)$ and our estimates $y' = f'(x)$). This contrasts the scientific method, which explains "why" things happen in terms of mechanisms (minimize the loss $L(f, f')$ between the true mechanism f and our estimate f').

Since correlation is to prediction what causation is to explanation, I propose to tame horses by developing machine learning algorithms that leverage only causal dependencies, and ignore confounding dependencies.

3.13   Kernel Learning with Convolutional Kernel Networks

Julien Mairal (INRIA – Grenoble, FR)

In this talk, we present a new image representation based on a multilayer kernel machine that performs end-to-end learning. Unlike traditional kernel methods, where the kernel is handcrafted or adapted to data in an unsupervised manner, we learn how to shape the kernel for a supervised prediction problem. We proceed by generalizing convolutional kernel networks, which originally provide unsupervised image representations, and we derive backpropagation rules to optimize model parameters. As a result, we obtain a new type of convolutional neural network with the following properties: (i) at each layer, learning filters is equivalent to optimizing a linear subspace in a reproducing kernel Hilbert space (RKHS), where we project data; (ii) the network may be learned with supervision or without; (iii) the model comes with a natural regularization function (the norm in the RKHS). We show that the method achieves reasonably competitive performance on some standard "deep learning" image classification datasets such as CIFAR-10 and SVHN, and also state-of-the-art results for image super-resolution, demonstrating the applicability of the approach to a large variety of image-related tasks.

3.14 Shrinkage Estimators

Krikamol Muandet (Mahidol University, TH)

A mean function in a reproducing kernel Hilbert space (RKHS), or a kernel mean, is central to kernel methods in that it is used by many classical algorithms such as kernel principle component analysis, and it also forms the core inference step of modern kernel methods that rely on embedding probability distributions in RKHSs. Given a finite sample, an empirical average has been used commonly as a standard estimator of the true kernel mean. Despite a widespread use of this estimator, we show that it can be improved thanks to the well-known Stein phenomenon. We propose a new family of estimators called kernel mean shrinkage estimators (KMSEs), which benefit from both theoretical justifications and good empirical performance. The results demonstrate that the proposed estimators outperform the standard one, especially in a "large d, small n" paradigm.

3.15 MMD/VAE/f-GAN: Methods for Estimating Probabilistic Models

Sebastian Nowozin (Microsoft Research UK – Cambridge, GB)

Estimating generative or discriminative probabilistic models is important in practical applications. To formalize estimation we can think of measures of discrepancy between two distributions: the model distribution and the unknown true distribution. The classes of discrepancy measures are: 1. integral probability metrics, 2. proper scoring rules, and 3. f-divergences. Integral probability metrics take the supremum of a difference of two expectations over a class of functions. Depending on the choice of function class this realizes metrics such as the total variation, maximum mean discrepancy, or the Wasserstein metric. If the function class is taken to be a RKHS the resulting metric is the kernel MMD. Proper scoring rules require a more detailed access to the model distribution through its density function or properties thereof. Typical examples are the likelihood, the Brier score, or Bernardo's quadratic scoring rule. In some cases, taking the integral of a scoring rule yields an f-divergence. f-divergences require access to the density function of both the model distribution and the true distribution, which is not available. Recently a variational lower bound on f-divergences allows to circumvent this requirement by introducing an additional variational function. Training a generative model with this variational approach yields a

saddle-point problem to solve, an approach known as the generative-adversarial network (GAN) approach. These new approaches to estimating models in the likelihood-free setting have allowed new levels of performance in fitting complicated distributions such as learning distributions of natural images.

3.16   Score matching and kernel based estimators for the drift of stochastic differential equations

Manfred Opper (TU Berlin, DE)

Score matching is a method for estimating the logarithms of a probability density (up to a constant) which is not based on a likelihood. Using a kernel method this approach has recently been generalised to nonparametric density estimation.

I show that this method relates to a drift estimation problem for certain classes of stochastic differential equations and can be generalised to treat interesting types of Langevin equations.

I also show that the kernel method can be understood as a proper Bayesian approach in the limit, where observations of the stochastic process are densely sampled in time.

3.17   Gaussian Processes – Past and Future?

Carl Edward Rasmussen (University of Cambridge, GB)

I attempt to give an overview of challenges of GPs in the past and what the situation might look like in the future. I discussed 5 central theorems: 1) Uses of GPs; although GPs are often used simply to model functions, their central advantage are situations where the predictive error bars are central: probabilistic numerics, decision making, RL and active learning among others. 2) Practical considerations: providing good code/toolboxes and automation, covariance functions and inducing points. These questions haven't really been addressed satisfactorily. 3) Computational considerations: these questions have largely been solved, especially inducing point methods are very good. 4) Covariance structures: we still don't have a clear idea how to implement more sophisticated covariance functions, or how practically to do inference when the number of hyperparameters (statistically) prohibit ML type 2 treatment. 5) Towards the future: can we construct little Lego brick GPs which can take probabilistic inputs and can be assembled as stacked into useful structures?

3.18 How to fit a simple model

Carl Edward Rasmussen (University of Cambridge, GB)

In this talk, I present a paradox of suboptimal learning when fitting parameters of a simple model class. We show how it can be beneficial to learn a complex model which is then projected onto the simple model class rather than directly map from data to parameters.

3.19 String Gaussian Processes & Generalized Spectral Kernels

Stephen Roberts (University of Oxford, GB)

Joint work of Yves-Laurent Kom Samo, Stephen Roberts
Main reference Y. L. Kom Samo, S. Roberts, "String and Membrane Gaussian Processes", Journal of Machine
Learning Research, 17(131):1–87, 2016.
URL http://jmlr.org/papers/v17/15-382.html

In this talk we introduce ways of invoking highly non-stationary kernels. String GPs allow for a domain to be broken into a series of conditionally independent Gaussian Processes, which merely ensure continuity in f and f' at the boundaries. We show how this allows for not just non-stationary modelling in the extreme, but also a competitive scaling to large data sets. Using Lebesgue's decomposition theorem, it is showed that the two major methodologies in spectral kernel learning represent the continuous and singular components of the measure and how this can be extended to more general cases using a bi-measure.

3.20 Kernels – Past and Future?

Bernhard Schölkopf (MPI für Intelligente Systeme – Tübingen, DE)

Main reference A. Scibior, C.-J. Simon-Gabriel, I. O. Tolstikhin, B. Schölkopf, "Consistent Kernel Mean
Estimation for Functions of Random Variables", in Proc. of Advances in Neural Information
Processing Systems 29 (NIPS 2016), pp. 1732–1740, Neural Information Processing Systems
Foundation Inc., 2016.
URL http://papers.nips.cc/paper/6545-consistent-kernel-mean-estimation-for-functions-of-random-
variables

The talk summarized the development of kernel methods for machine learning, focusing on the main ideas and some future possibilities: introduction of p.d. kernels within the theory of integral equations, their use in potential functions methods, in SVM, the general "kernel trick", the observation that kernels can be defined on arbitrary sets of objects, the link to GPs, and finally the idea to represent distributions by kernel means, underlying kernel tests such as MMD and kernel independence tests.

Kernel mean representations lend themselves well to the development of kernel methods for probabilistic programming, i.e., methods for lifting functional operations defined for data types to the same functional operations for distributions over these data types.

3.21   Kernel Embeddings and Bayesian Quadrature

Dino Sejdinovic (University of Oxford, GB)

The talk overviewed kernel embeddings as implicit representations of probability measures, leading to the framework allowing nonparametric hypothesis testing and learning on distributions as inputs. In addition, the theory of kernel embeddings allows an alternative interpretation of Bayesian Quadrature (BQ) which does not require invoking the Gaussian process model which puts prior measures on known integrands. This interpretation leads to a recipe for the method applicable where kernel embeddings are not analytically available, while still matching the convergence rates of BQ.

3.22   Kernel Mean Embeddings

Carl-Johann Simon-Gabriel (MPI für Intelligente Systeme – Tübingen, DE)

We started with a brief introduction to KMEs, that motivated the embedding function:

$$m : \mathcal{P} \to \mathcal{H}$$
$$P \mapsto \int k(\cdot, x) \, dP(x)$$

This embedding defines a distance d_k between probability measures, which metrizes the usual weak convergence if and only if k is continuous and m is injective. We then showed how to systematically link the following three frequently used concepts: universal, characteristic and strictly positive definite kernels. From these links, we concluded that KMEs can be extended so as to embed not only probability measures, but also generalised measuers, aka. Schwartz-distributions. Furthermore, these extensions can remain injective, if the original embedding is injective. The sets of Schwartz-distributions can be seen as sets of measures and of their (distributional) derivatives. Interestingly, the embedding of the derivative P' of P can be easily deduced from the embedding of P. We hope that these extended embeddings will find applications in numerical methods for differential equation solving.

3.23   Random Fourier Features and Beyond

Bharath Sriperumbudur (Pennsylvania State University – University Park, US)

In this talk, I will recall the primal and dual formulations of linear ridge regression and kernel ridge regression as a motivating example to introduce feature approximations in the primal setting. Kernel methods have traditionally focused on the dual setting as it does not require the knowledge of the feature maps and also scales only with the sample size. To improve the scalability of kernel methods, various approximations to the dual problem has been studied in terms of incomplete Cholesky factorization, Nyström methods etc. Recently, a Fourier feature based finite dimensional approximation has been introduced which enables to work with the primal setting. In this talk, I will discuss the quality of approximation of Fourier features and present results on the optimality of approximation rates. Then, I will discuss various generalizations and directions of random feature approximations, some of which include rates of approximation for derivatives of kernels, optimal approximation rates for operator-valued kernels and possibility of other approximations to improve the scalability of kernel methods in the primal setting.

3.24   Learning with Hierarchical Kernels

Ingo Steinwart (Universität Stuttgart, DE)

We investigate iterated compositions of weighted sums of Gaussian kernels and provide an interpretation of the construction that shows some similarities with the architectures of deep neural networks. On the theoretical side, we show that these kernels are universal and that SVMs using these kernels are universally consistent. We further describe a parameter optimization method for the kernel parameters and empirically compare this method to SVMs, random forests, a multiple kernel learning approach, and to some deep neural networks.

3.25 Distribution Regression

Zoltán Szabó (Ecole Polytechnique – Palaiseau, FR)

We focus on the distribution regression problem (DRP): we regress from probability measures to Hilbert-space valued outputs, where the input distributions are only available through samples (this is the 'two-stage sampled' setting). Several important statistical and machine learning problems can be phrased within this framework including point estimation tasks without analytical solution (such as entropy estimation), or multi-instance learning. However, due to the two-stage sampled nature of the problem, the theoretical analysis becomes quite challenging: to the best of our knowledge the only existing method with performance guarantees to solve the DRP task requires density estimation (which often performs poorly in practise) and the distributions to be defined on a compact Euclidean domain. We present a simple, analytically tractable alternative to solve the DRP task: we embed the distributions to a reproducing kernel Hilbert space and perform ridge regression from the embedded distributions to the outputs. We prove that this scheme is consistent under mild conditions, and construct explicit finite sample bounds on its excess risk as a function of the sample numbers and the problem difficulty, which hold with high probability. Specifically, we establish the consistency of set kernels in regression, which was a 17-year- old-open question, and also present new kernels on embedded distributions. The practical efficiency of the studied technique is illustrated in aerosol prediction using multispectral satellite images.

3.26 Stochastic (partial) differential equations and Gaussian processes

Simo Särkkä (Aalto University, FI)

Stochastic partial differential equations and stochastic differential equations can be seen as alternatives to kernels in representation of Gaussian processes. Linear operator equations give spatial kernels, temporal kernels are equivalent to linear Itô stochastic differential equations. The differential equation representations allow for the use of differential equation numerical methods on Gaussian processes. For example, finite-differences, finite elements, basis function methods, and Galerkin methods can be used. In temporal and spatio-temporal case we can use linear-time Kalman filter and smoother approaches.

3.27   Consistent Kernel Mean Estimation for Functions of Random Variables

Ilya Tolstikhin (MPI für Intelligente Systeme – Tübingen, DE)

Given a random variable X and a function defined over the same space, we consider a problem of constructing a flexible representation for a distribution of $f(X)$. Following the approach of [1], we propose to do so by using mean embeddings of probability distributions into corresponding Reproducing Kernel Hilbert Spaces. Our new results show that a consistent estimation of the mean embedding of X leads to a consistent estimation of the mean embedding of $f(X)$. In particular, this result shows the consistency of a new estimator proposed by [1]. Apart form asymptotic results we also provide a finite sample guarantees for Matern kernels and discuss possible applications, including probabilistic programming.

References
1   B. Schölkopf, K. Muandet, K. Fukumizu, S. Harmeling, and J. Peters: "Computing functions of random variables via reproducing kernel Hilbert space representations." In: Statistics and Computing 25(4), 755–766, 2015.

3.28   Uncertain inputs in Gaussian Processes

Mark van der Wilk (University of Cambridge, GB)

We introduced Sparse GP inference using Variational Bayesian inference. In this framework, we aim to minimise the KL divergence of some easily computable approximate posterior to the true posterior. The structure of the approximations lends itself really well to handling uncertain inputs as well. The ability to do so is essential for the idea of making GPs "building blocks" of Machine Learning, like neural network layers, if uncertainty is to be taken into account. Finally we contrast the goal of uncertain input GPs to distribution regression from the kernel literature. Could there be any way to combine the methods?

3.29   Frequentist properties of GP learning methods

Harry van Zanten (University of Amsterdam, NL)

GPs have been routinely used as priors in a variety nonparametric statistical problems, including regression and classification. In this talk I gave a short overview of theoretical

results that have been obtained during the last 10 years or so about the frequentist performance of such methods.

The first wave of convergence rate results for nonparametric Bayes with fixed GP priors showed that to achieve optimal rates, the regularity of the GP has to exactly match the regularity of the function that is being estimated. These results apply to all the popular kernels, including the Matern and the squared exponential, for instance. Since the true regularity of the function of interest is typically not known, it is therefore necessary to have adaptive procedures that automatically set tuning, or hyper parameters in an optimal way. A second wave of results showed that both hierarchical and empirical Bayes methods can do this properly, provided they are carefully constructed. The results show that matters are actually a bit delicate. It is important for instance which priors are placed on hyper parameters, or which hyper parameters are held fixed and which are tuned. The third class of results that I discussed deal with the frequentist interpretation of nonparametric credible sets. Ideally, we would like a 95% credible set to be a frequentist 95% confidence set as well and at the same time have minimal, optimal size. It turns out that in the adaptive setting in which you don't know the regularity of the truth and use for instance hierarchical Bayes or empirical Bayes, this is fundamentally impossible. Whatever priors you use, there are always ground truths for which the credible sets are completely misleading. This means that confidence statements in nonparametric settings are fundamentally conditional: in nonparametric problems you can only believe credible sets, or error bars, if you *really* believe that your prior reflects the fine properties of the truth.

Several fundamental issues concerning GP methods are not fully understood yet. One of the most interesting ones is perhaps the issue of the fundamental limitations and possibilities of distributed GP methods. Under which conditions can such methods achieve the same optimal, adaptive performance as a centralised methods?

References

1   A. W. van der Vaart, and J. H. van Zanten: "Rates of contraction of posterior distributions based on Gaussian process priors." In: The Annals of Statistics, 1435–1463, 2008.

2   A. W. van der Vaart, and J. H. van Zanten: "Adaptive Bayesian estimation using a Gaussian random field with inverse Gamma bandwidth." In: The Annals of Statistics, 2655–2675, 2009.

3   A. W. van der Vaart, and J. H. van Zanten: "Information rates of nonparametric Gaussian process methods." In: Journal of Machine Learning Research 12, 2095–2119, 2011.

4   B. Szabó, A. W. van der Vaart, and J. H. van Zanten: "Frequentist coverage of adaptive nonparametric Bayesian credible sets." In: The Annals of Statistics 43(4), 1391–1428, 2015.

3.30   The ML Invasion of ABC

Richard Wilkinson (University of Sheffield, GB)

There are classes of models for which we can easily sample realizations from the model, but where we cannot compute the likelihood function $\pi(x|\theta)$. These typically are in scientific problems where the model represents our physical knowledge about the system. ABC methods are a class of Monte Carlo algorithms for doing all inference using simulation from the model.

In particular, they target some form of posterior distribution $\tilde{\pi}(\theta|D) \propto \tilde{\pi}(D|\theta)\pi(\theta)$ where the likelihood function used, $\tilde{\pi}$, may be very different from the true likelihood π.

In the past few years, some of the most interesting ideas in ABC have arisen in machine learning groups. These include approaches for bypassing the need to choose a set of summary statistics by instead using a kernel embedding & using the MMD metric to determine whether the simulator output is comparable to the data, & the use of generative adversarial networks as alternatives to pseudo-likelihood based approaches. There has also been significant work using surrogate models for the likelihood function, for example, by using Gaussian processes to approximate the likelihood function & then using the GP in a MCMC inference scheme to find the posterior.

4   Working groups

4.1   Generative Models

David Duvenaud (Toronto, CA)

Sebastian Nowizin introduced variational autoencoders, amortized inference, and the reparameterization trick. He said that kernel MMD is a good way to introduce non-probabilistic practitioners to the probabilistic way of doing things, because all you need to do is simply add noise to the input of whatever generative procedure and a diversity term on pairs of outputs to the loss. He says it's a lot easier to implement and get to work than GAN training.

Arthur, Ilya and Bernhard had a long discussion about the relationships between kernel MMD, VAEs and GANs. Ilya suggests that there is an opportunity for theoreticians to clean up and organize all the tricks that are required for training GANs.

Sebastian makes the point that modeling natural images is mainly interesting of a proxy task for modeling high-dimensional densities, and it's a task where humans can evaluate the quality of samples.

We talked about how the blurry images of VAEs can be address by using more sophisticated likelihoods, as in https://arxiv.org/pdf/1611.05013v1.pdf.

Krikamol asks if we can incorporate MCMC into VAEs or GANs, and I mention that Max Welling has a paper that does this: https://arxiv.org/abs/1410.6460.

We also talked about a paper by Roger Grosse that attempts to properly evaluate the predictive probability of VAEs and GANs: https://openreview.net/pdf?id=B1M8JF9xx.

I also suggest there might be room for kernel people to help with the new "Operator Variational Inference" method: https://arxiv.org/pdf/1610.09033.pdf.

Then we broke for coffee.

4.2   Limitations of GPs / non-Gaussian-Processes

Stefan Harmeling (Universität Düsseldorf, DE)

Participants: Carl Rasmussen, Hannes Nickisch, Manfred Opper, Marc Deisenroth Maren Mahsereci, Mark van der Wilk, Michael Osborne, Philipp Hennig, Roman Garnett, Simo Särkkä, Stefan Harmeling, Stephen Roberts.

1. *Question:* Are stacked or deep GPs useful extensions of GPs?
   Answer: Yes, stacked or deep GPs are working and they are useful, since they bring new possibilities a GP has not. Their deep structure might also lead to fewer parameters. However, learning the parameters of stacked or deep GPs might require lots of data.
2. *Question:* How can we formulate non-GPs on functions spaces?
   Answer: (i) The concatenation of two function-valued random variables is usually not a GP distribution. Or even simpler, (ii) a function-valued random variable can be concatenated with a nonlinear function to obtain a function-valued random variable that follows no longer a GP distribution. With such tricks we can impose constraints on function distributions such as monotonicity, convexity, non-negativity, etc. However, for lots of data the posterior GP might fulfill the constraints with high probability.
   We can also apply nonlinear functionals to the GPs (as Hilbert space elements), e.g., we can use GPs as inputs to nonlinear ordinary or partial differential equations to create non-Gaussian process outputs. The scaled mixtures of probability measures is one way which leads to e.g. Student-t and related processes. Using stochastic processes as inputs to other processes (as in deep GPs) also leads to non-Gaussian processes.
3. *Question:* In practice a large number of hyper-parameters make inference harder. Can we tie parameters to the rescue?
   Answer: It might help as is suggested by deep learning, where even randomly tying parameters can improve performance. Another difficulty of inference might be a situation, in which the input data lies on a low-dimensional manifold. This could lead to uncertainty off the manifold which might be sometimes problematic. Deep learning suggests the intuition that increasing the number of parameters can actually make optimisation of those parameters *easier*. While this might hold for Gaussian process models with lots of hyperparameters, we don't have much assurance that this wouldn't result in overfitting.
4. *Question:* What is a simple example of a family of functions that a GP can not model properly?
   Answer: The set of step functions (i.e. with exactly one step at an unknown location) can not be the support of a GP distribution in function space. So using covariance functions (kernels) to specify distributions in function space is limited. It is either that a GP spreads its mass too thinly, or spreads it too wide.

One other point: Perhaps certain *prior* distributions are hard to model using Gaussian processes. However, as data comes in, the posterior of a non-Gaussian process might be easy to model using a GP. This has the flavour of variational inference and has more desirable properties than relying on the data to constrain a GP prior.

Summary

1. Stacked and deep GPs are useful!
2. Create non-GPs e.g. by concatenation.
3. Lots of parameters make inference harder.
4. There is no GP for step functions.

4.3 The separation between Kernels/GPs and Deep Learning

Sebastian Nowozin (Microsoft Research UK – Cambridge, GB)

Motivation: what are the known success cases and failure modes of kernels and deep learning methods? Which applications are they best suited for?

Practical Observations. Bayesian probabilistic models provide a clean framework for thinking about applications such as one-shot learning and transfer learning; increasingly deep learning methods attempt to achieve the same applications where uncertainty is important, often successfully.

Gaussian processes are well established for regression problems, in particular in the small data setting. For example, in sample-efficient reinforcement learning. Similarly, in spatio-temporal modeling they work well.

Building kernels for high-dimensional input data (d>10 or d>100) or heterogeneous data is difficult. For Gaussian processes, performing hyperparameter optimization is challenging when there are a large number of hyperparameters. Perhaps stochastic optimization methods or variational inference for the hyperparameters can improve this.

Deep learning offer a flexible framework of overparameterized functions. They place representation learning first in terms of an explicit feature map, which allows representation to be useful for different applications.

Kernel methods still dominate in testing, such as with maximum mean discrepancy (MMD), because guarantees are important in hypothesis testing. Kernel methods are also popular in structured input settings, where we handle strings or graph structures, for example in bioinformatics.

In Bayesian optimization Gaussian processes are successful but other probabilistic models are possible. Bayesian neural networks may be an alternative but there are computational issues as well.

The deep learning community also spends a good amount of engineering and efficient implementation; does the GP community spend the same amount of effort?

Kernel methods provide mathematical tools to potentially prove guarantees within control applications.

Kernel methods are also *easy* to automate and predictable both in runtime and in the influence of parameters.

Computational Issues. Deep learning methods are perhaps also really successful because they are scalable; therefore, computational issues may be an important aspect of practical success.

Fast GPs scale with a flexible parameters, the number of inducing points. The variational approach is one approach but there are perhaps more efficient approaches. Yet, they remain much less scalable than deep learning methods.

Kernel methods (e.g. SVM) have also achieved better scalability and now can be used for up to 10M data points with controlled and guaranteed approximation guarantees. Popular packages such as scikit-learn also scales to 500k points without practical difficulties.

Representation power of functions: deep learning can represent complicated functions, sparse GPs with inducing points cannot. Perhaps there are interesting extensions of the inducing point approach to enhance the representational power of GPs.

Prediction: neural network prediction is fast, but for GPs prediction is expensive.

Dataset size: small data is good for kernel methods, big data is good for deep learning.

Theoretical Limitations.   Standard Gaussian processes have some known limitations.

What guarantees can be provided for the error bars of a GP? They correspond to posterior credible intervals, but therefore are conditioned on the assumed model class. Non-parametric uncertainty quantification is not possible in general (results of Richard Nickl). Do we care about the prediction accuracy or about the calibration properties of the error bard? Are error bars sufficient in practice, even if we know the model assumptions to be wrong? From the Bayesian viewpoint error bars depend only on the modeling assumptions, so we need to question modeling assumptions if we are not satisfied with the quality of the error bars.

Why does deep learning work in very high dimensions? Are there fundamental assumptions about real world densities that we have not understood yet?

Classification is still difficult with Gaussian process models.

There are two additional observations regarding kernel methods versus deep learning:

1. Degree-of-freedom bottleneck: a GP has effective N parameters to determine a function, where N is the number of samples. A deep network has a potentially larger number.
2. Kernel-information-bottleneck: a kernel consists of N*N scalars. If every instance contains a large amount of information (e.g. a megapixel image), more information flows to a deep neural network system than to a kernel method.

Deep Kernel Methods versus Deep Learning.   Kernel methods typically use a handcrafted kernel, whereas deep learning methods learn the representation by data.

Practical advantages of non-parametric methods are most likely not existing; the advantage is in theory, being able to prove how to increase function class as the data grows in order to guarantee the right function is recovered.

Deep learning methods can also represent uncertainty either by directly fitting a variance or by using Bayesian neural networks.

Main difficulties in deep kernel methods is in scalability. Main disadvantage compared to deep neural networks is the difficulty of running them on GPU, for example for computing covariance matrices as part of performing a deep kernel computation, or for computing the posterior variance for a given data point. Do deep GPs scale to many layers? This is unclear.

Optimization problems arising in deep neural networks and in deep learning are the same, for example initialization and optimization.

4.4 Connections and Differences between Kernels and GPs

Dino Sejdinovic (University of Oxford, GB)

The discussion recognized that there are shared mathematical foundations of the frequentist kernel methods and Gaussian Processes (GP). These foundations are based on the theory of Gaussian Hilbert spaces and the fact that the notions of *orthogonality and independence coincide* on the L_2-spaces of gaussian random variables. There is a need for a dictionary translating different concepts within these two communities using this shared mathematical framework – there is potentially lots to be gained from these different interpretations. As an example, we discussed the frequentist interpretation of the standard GP posterior covariance. It turns out this can be viewed as an inner product between the component of features orthogonal to the data subspace, i.e. *similarity not explained by the data*. There is also a "worst-case error" over the reproducing kernel Hilbert space (a specified class of functions with the encoded regularity) interpretation, which can be used to quantify uncertainty. Similar connection exists between maximum mean discrepancy (MMD) and the GP posterior variance in Bayesian quadrature.

Main similarities arise in standard supervised learning settings (e.g. kernel ridge regression and GP regression are closely related) but the connections in unsupervised settings are less well understood. For example, are there Bayesian counterparts to kernel PCA or density estimation using infinite dimensional exponential families? How are they related?

Another point of discussion revolved around the result that *the samples from GP almost surely do not lie in the RKHS* with the corresponding covariance kernel, even though the posterior mean does. This has important implications on model specification within the two frameworks. What are useful ways to think about this?

It was also reiterated that the two frameworks have different philosophies, with the frequentist focus on risk and the GP framework focusing on describing posterior measures and being oblivious to the task that follows it. Thus, there are important differences in the decision making process and since the two frameworks generally do different things with the same mathematical objects and interpret them differently – a caution should be exercised when translating these mathematical objects and this process may in some cases be misleading.

Participants

- Florence d'Alché-Buc
  Telecom ParisTech, FR
- Marc Deisenroth
  Imperial College London, GB
- David Duvenaud
  Toronto, CA
- Roman Garnett
  Washington University –
  St. Louis, US
- Arthur Gretton
  University College London, GB
- Stefan Harmeling
  Universität Düsseldorf, DE
- Philipp Hennig
  MPI für Intelligente Systeme –
  Tübingen, DE
- James Hensman
  Lancaster University, GB
- Joseé Miguel
  Hernández-Lobato
  Harvard University –
  Cambridge, US
- Frank Hutter
  Universität Freiburg, DE
- Motonobu Kanagawa
  Institute of Statistical
  Mathematics – Tokyo, JP
- Andreas Krause
  ETH Zürich, CH
- Neil D. Lawrence
  University of Sheffield, GB
- David Lopez-Paz
  Facebook – AI Research, US

- Maren Mahsereci
  MPI für Intelligente Systeme –
  Tübingen, DE
- Julien Mairal
  INRIA – Grenoble, FR
- Krikamol Muandet
  Mahidol University, TH
- Hannes Nickisch
  Philips – Hamburg, DE
- Sebastian Nowozin
  Microsoft Research UK –
  Cambridge, GB
- Cheng Soon Ong
  Data61 – Canberra, AU
- Manfred Opper
  TU Berlin, DE
- Peter Orbanz
  Columbia University –
  New York, US
- Michael A. Osborne
  University of Oxford, GB
- Carl Edward Rasmussen
  University of Cambridge, GB
- Stephen Roberts
  University of Oxford, GB
- Volker Roth
  Universität Basel, CH
- Simo Särkkä
  Aalto University, FI
- Bernt Schiele
  MPI für Informatik –
  Saarbrücken, DE

- Michael Schober
  MPI für Intelligente Systeme –
  Tübingen, DE
- Bernhard Schölkopf
  MPI für Intelligente Systeme –
  Tübingen, DE
- Dino Sejdinovic
  University of Oxford, GB
- Carl-Johann Simon-Gabriel
  MPI für Intelligente Systeme –
  Tübingen, DE
- Bharath Sriperumbudur
  Pennsylvania State University –
  University Park, US
- Ingo Steinwart
  Universität Stuttgart, DE
- Zoltán Szabó
  Ecole Polytechnique –
  Palaiseau, FR
- Ilya Tolstikhin
  MPI für Intelligente Systeme –
  Tübingen, DE
- Raquel Urtasun
  University of Toronto, CA
- Mark van der Wilk
  University of Cambridge, GB
- Harry van Zanten
  University of Amsterdam, NL
- Richard Wilkinson
  University of Sheffield, GB

Report from Dagstuhl Seminar 16482

Algorithms and Effectivity in Tropical Mathematics and Beyond

Edited by

Stéphane Gaubert[1], Dima Grigoriev[2], Michael Joswig[3], and Thorsten Theobald[4]

1   INRIA Saclay – Île-de-France, FR, `stephane.gaubert@inria.fr`
2   Lille I University, FR, `dmitry.grigoryev@math.univ-lille1.fr`
3   TU Berlin, DE, `joswig@math.tu-berlin.de`
4   Goethe-Universität Frankfurt am Main, DE, `theobald@math.uni-frankfurt.de`

─────── **Abstract** ───────

This report documents the Dagstuhl Seminar on Algorithms and Effectivity in Tropical Mathematics and Beyond, which took place from November 27 – December 02, 2016. The report contains an executive summary as well as abstracts of the talks which reflect recent progress in the topic of the meeting.

1   Executive summary

Stéphane Gaubert
Dima Grigoriev
Michael Joswig
Thorsten Theobald

Tropical mathematics is a uniting name for different research directions which involve the semi-ring of real numbers endowed with the operations min, + (called the tropical semi-ring). It has emerged in several areas of computer science and of pure and applied mathematics. For the first time, this seminar brought together the computer science and the mathematics viewpoints. A focus was on effective methods, algorithms and complexity bounds in tropical mathematics, and on their relations with open questions in various areas of computer science, including optimization, game theory and circuit complexity.

One of the oldest open algorithmic challenges in tropical mathematics is the complexity of solving systems of tropical linear equalities and inequalities. It is known to be equivalent to solving mean payoff games. The solvability of these problems is among the few known problems which are contained in the intersection NP ∩ co-NP, but not currently known

to be in P. This leads to new approaches in linear programming or convex semialgebraic programming over nonarchimedean fields.

According to the organizers' points of view the seminar was quite successful. In addition to 28 talks there were many informal discussions and exchange of ideas in small groups. We expect several new common papers of the participants conceived during the seminar. An important feature was to bring together experts with different backgrounds who often knew other participants just by their publications. According to the opinions expressed, the participants learned a lot of new things. The seminar was especially useful for the young people.

Every talk, in addition to new results, also contained open problems. This created a lot of interaction in subsequent discussions. The audience was very active, many questions were posed to the speakers during the talks and the breaks.

The talks can be conditionally partitioned into the following groups, although there were many interrelations between different groups:

- Algorithmical problems of foundations of tropical mathematics (H. Markwig, D. Maclagan, F. Rincon, V. Podolskii);
- Complexity of games and of tropical linear and convex algebra (M. Bodirsky, S. Gaubert, T. Hansen, M. Joswig, G. Loho, M. MacCaig, B. Schröter, S. Sergeev, M. Skomra);
- Algorithms and complexity bounds on tropical varieties (F. Bihan, D. Grigoriev, S. Hampe, A. Jensen, T. Jörgens, L. Tabera, T. Theobald, T. de Wolff). We mention that S. Hampe has made a demonstration of the software Polymake for computations in tropical algebra;
- Algorithms in tropical differential algebra (F. Aroca, C. Garay);
- Interactions of tropical mathematics with algorithmic issues in classical mathematics (M. Akian, X. Allamigeon, Bo Lin, M. Rojas, C. Vinzant).

During the seminar a manuscript appeared (on the Internet) with the very strong claim of a quasi-polynomial complexity algorithm for parity games. It was a lucky coincidence that so many experts with various backgrounds were present. So a special evening session on Thursday was created to analyze this result and its ramifications. This was one more highlight.

2   Table of Contents

3   Overview of Talks

3.1   Majorization inequalities for valuations of eigenvalues using tropical algebra

Marianne Akian (INRIA Saclay – Île-de-France, FR)

We consider a matrix with entries over the field of Puiseux series, equipped with its non-archimedean valuation (the leading exponent). We establish majorization inequalities relating the sequence of the valuations of the eigenvalues of a matrix with the tropical eigenvalues of its valuation matrix (the latter is obtained by taking the valuation entrywise). We also show that, generically in the leading coefficients of the Puiseux series, the precise asymptotics of eigenvalues, eigenvectors and condition numbers can be determined. For this, we apply diagonal scalings constructed from the dual variables of a parametric optimal assignment constructed from the valuation matrix.

Next, we establish an archimedean analogue of the above inequalities, which applies to matrix polynomials with coefficients in the field of complex numbers, equipped with the modulus as its valuation.

This talk covers joint works with Ravindra Bapat, Stéphane Gaubert, Andrea Marchesini, Meisam Sharify, and Francoise Tisseur.

3.2   Log-barrier interior-point methods are not strongly polynomial

Xavier Allamigeon (INRIA Saclay – Île-de-France, FR)

We identify a class of path-following interior-point methods which are not strongly polynomial. This class corresponds to primal-dual interior-point methods which iterates in the so-called "wide" neighborhood of the central path arising from the logarithmic barrier. It includes short step, long step as well as predictor-corrector types of interior-point methods.

We establish a lower bound on the number of iterations of these methods when applied to some parametric families of linear programs. The lower bound is expressed in terms of the number of tropical segments in a piecewise linear curve arising from the tropicalization of the central path. In this way, we derive that the aforementioned interior point methods require $\Omega(2^d)$ iterations to solve a linear program in dimension $O(d)$ which we studied in a previous work.

3.3 The Groebner fan of a differential ideal in tag ring of differential polynomials.

Fuensanta Aroca (Universidad Nacional Autonoma – Mexico, MX)

We will start with the Newton polygon of an algebraic plane curve, the Newton polytope of an algebraic hypersurface, and the tropical variety of an ideal. Then we will see the extension of these concepts to the differential case.

3.4 Fewnomial polynomial systems in the complex, real and tropical settings

Frédéric Bihan (University Savoie Mont Blanc – Le Bourget-du-Lac, FR)

In this talk I will first give a short survey of what is known about the number of solutions of a square polynomial system. I will mainly recall classical upper bounds in the complex, real and tropical settings. Tropical polynomial systems can be used to construct real polynomial systems with many real solutions. In the transversal case, this is known as the classical patchworking theorem. The non transversal case is more delicate since we don't know which tropical solutions lift to real ones in general. I will present a new construction due to Boulos El Hilany which uses a non transversal intersection of two plane tropical curves to produce a system with a record number of real positive solutions.

3.5 Max-closed semilinear constraint satisfaction problems

Manuel Bodirsky (TU Dresden, DE)

A semilinear relation S is max-closed if it is preserved by taking the componentwise maximum. The constraint satisfaction problem for max-closed semilinear constraints is at least as hard as determining the winner in Mean Payoff Games, a notorious problem of open computational complexity. Mean Payoff Games are known to be in the intersection of NP and co-NP, which is not known for max-closed semilinear constraints. Semilinear relations that are max-closed and additionally closed under translations have been called tropically convex in the literature. One of our main results is a new duality for open tropically convex relations, which puts the CSP for tropically convex semilinear constraints in general into NP ∩ co-NP. This extends the corresponding complexity result for and-or precedence constraints aka the max-atoms problem. To this end, we present a characterization of max-closed semilinear relations in terms of syntactically restricted first-order logic, and another characterization in terms of a finite set of relations L that allow primitive positive definitions of all other relations in the class. We also present a subclass of max-closed constraints where the CSP is in P; this class generalizes the class of max-closed constraints over finite domains, and the feasibility

problem for max-closed linear inequalities. Finally, we show that the class of max-closed semilinear constraints is maximal in the sense that as soon as a single relation that is not max-closed is added to L, the CSP becomes NP-hard.

3.6 Lopsided approximation of amoebas

Timo de Wolff (Texas A&M University – College Station, US)

The amoeba of a Laurent polynomial is the image of the corresponding hypersurface under the coordinatewise log absolute value map. We demonstrate that a theoretical amoeba approximation method due to Purbhoo can be used effectively in practice. To do this, we resolve the main bottleneck in Purbhoo's method by exploiting relations between cyclic resultants, which allow us to reduce the number of arithmetic operations needed to compute the 2^k-th cyclic resultant of a univariate degree d polynomial from $O(d2^k)$ to $O(kd^2)$. We use the same approach to give an approximation of a semi-algebraic description of the Log preimage of the amoeba of a Laurent polynomial. We also provide a `SINGULAR/SAGE` implementation of these algorithms, which shows a significant speedup when our specialized cyclic resultant computation is used, versus a general purpose resultant algorithm.

3.7 The fundamental theorem of tropical differential algebraic geometry

Cristhian Garay (Fluminense Federal University – Rio de Janeiro, BR)

Joint work of Fuensanta Aroca, Cristhian Garay, Zeinab Toghani

Let I be an ideal of the ring of Laurent polynomials $K[x_1^{\pm 1}, \ldots, x_n^{\pm 1}]$ with coefficients in a real-valued field (K, v). The fundamental theorem of tropical algebraic geometry states the equality $\mathrm{trop}(V(I)) = V(\mathrm{trop}(I))$ between the tropicalization $\mathrm{trop}(V(I))$ of the closed subscheme $V(I) \subset (K^*)^n$ and the tropical variety $V(\mathrm{trop}(I))$ associated to the tropicalization of the ideal $\mathrm{trop}(I)$.

In this talk we discuss an analogous result for a differential ideal G of the ring of differential polynomials $K[[t]]\{x_1, \ldots, x_n\}$, where K is an uncountable algebraically closed field of characteristic zero. We define the tropicalization $\mathrm{trop}(\mathrm{Sol}(G))$ of the set of solutions $\mathrm{Sol}(G) \subset K[[t]]^n$ of G, and the set of solutions $\mathrm{Sol}(\mathrm{trop}(G)) \subset \mathcal{P}(\mathbb{Z}_{\geq 0})^n$ associated to the tropicalization of the ideal $\mathrm{trop}(G)$. These two sets are linked by a tropicalization morphism $\mathrm{trop} : \mathrm{Sol}(G) \longrightarrow \mathrm{Sol}(\mathrm{trop}(G))$.

We show the equality $\mathrm{trop}(\mathrm{Sol}(G)) = \mathrm{Sol}(\mathrm{trop}(G))$, answering a question recently raised by D. Grigoriev.

3.8 The equivalence between zero-sum games and tropical convexity: non-linear Perron-Frobenius and metric fixed point methods

Stéphane Gaubert (INRIA Saclay – Île-de-France, FR)

Shapley operators represent the one day evolution of the value of a zero-sum two player game. They can be characterized as being non-expansive in a weak metric (Funk's metric) which arises in the Perron-Frobenius theory of non-linear maps over cones. The sets of sub or super fixed points of Shapley operators are precisely the (closed) tropical convex sets. In this way, metric fixed point techniques can be applied to tropical convexity. We shall survey some tools, like the existence of the mean payoff for nonexpansive mappings definable in an o-minimal structure, ergodicity conditions for zero-sum games, Kohlberg's invariant half-lines, non-linear spectral radii and Collatz-Wielandt type theorems. Then, we will discuss the applications of these results to tropical convexity and zero-sum games, starting with the equivalence between zero-sum games and tropical linear feasibility problems, and ending with extensions of this equivalence to different classes of non-expansive mappings, including ones arising in risk sensitive control and entropy games.

This talk is based on the following references:

References
1 Marianne Akian, Stéphane Gaubert, and Alexande Guterman. Tropical polyhedra are equivalent to mean payoff games. *International Journal of Algebra and Computation*, 22(1):1250001, 2012. `doi:10.1142/S0218196711006674`.
2 Jérôme Bolte, Stéphane Gaubert, and Guillaume Vigeral. Definable zero-sum stochastic games. *Mathematics of Operations Research*, 40(1):171–191, 2014. `doi:10.1287/moor.2014.0666`.
3 Marianne Akian, Stéphane Gaubert, and Antoine Hochart. Ergodicity conditions for zero-sum games. *Discrete and Continuous Dynamical Systems, series A*, 35(9):3901–3931, 2015. `doi:10.3934/dcds.2015.35.3901`.
4 Marianne Akian, Stéphane Gaubert, Jérémie Guillaud and Julien Grand-Clément. The operator approach to entropy games. Preprint. 2016.
5 Marianne Akian, Stéphane Gaubert, and Roger Nussbaum. A Collatz-Wielandt characterization of the spectral radius of order-preserving homogeneous maps on cones. 2011. `arXiv:1112.5968`.

3.9 Bounds on the number of connected components of tropical prevarieties

Dima Grigoriev (Lille I University, FR)

For a tropical prevariety in $\mathbb{R}^n$ given by a system of k tropical polynomials in n variables with degrees at most d, we prove that its number of the connected components is less than $\binom{k+7n-1}{3n} \cdot \frac{d^{3n}}{k+n+1}$. A similar bound (moreover, on the Betti numbers) for semi-algebraic sets was established in the papers by Oleinik-Petrovskii, Milnor, Thom, Basu-Pollack-Roy.

On a number of 0-dimensional connected components a better bound $\binom{k}{n} \cdot \frac{d^n}{k-n+1}$ is obtained, which extends the Bezout bound due to B. Sturmfels from the case $k = n$ to an arbitrary $k \geq n$.

Also we show that the latter bound is close to sharp, in particular, the number of connected components can depend on k.

3.10 Tropical computations in `polymake`

Simon Hampe (TU Berlin, DE)

In this talk I'll showcase some of the features of polymake that deal with tropical things. Most of these are new features of the recent polymake release 3.0. Examples include: Tropical arithmetic (Min or Max – it's your choice!), tropical convex hull computations, tropical linear spaces, tropical polynomials and tropical hypersurfaces. I will also present my own extension `a-tint`, which mostly deals with tropical intersection theory (but has a much wider range of features than that) and which has been included as a bundled extension in polymake.

3.11 An improved algorithm for feasibility of tropical linear programs

Thomas Dueholm Hansen (Aarhus University, DK)

Deciding feasibility of a tropical linear program was recently shown to be equivalent to solving a mean payoff game by Akian, Gaubert, and Guterman (2012). Allamigeon, Benchimol, Gaubert, and Joswig (2014) showed that combinatorial simplex algorithms for linear programming can solve mean payoff games. In this talk I present an improved version of the best known combinatorial simplex algorithm for linear programming, the Random-Facet pivoting rule by Kalai (1992) and Matousek, Sharir, and Welzl (1992), which due to the above reductions gives the best known algorithm for deciding whether a tropical linear program is feasible. The same improved running time is obtained for the more general class of abstract LP-type problems.

The talk is based on joint work with Uri Zwick, and in particular the paper *An Improved Version of the Random-Facet Pivoting Rule for the Simplex Algorithm* (2015).

3.12 Homotopy continuation for tropical polynomial systems

Anders Jensen (Aarhus University, DK)

In this talk we solve systems of n tropical polynomial equations in n unknowns by homotopy continuation, meaning that solutions are tracked as the coefficients of the system vary generically. Symbolically perturbing the coefficients gives an algorithm which is memoryless and exact, while it at the same time has complexity bounds similar to those of a recent

algorithm by Malajovich. As in the numerical algebraic geometry case, all isolated tropical solutions are obtained. We report on running times for our implementation in the setting of mixed volume computation.

3.13 Linear programs over fields of Puiseux fractions

Michael Joswig (TU Berlin, DE)

The formal Puiseux series with complex coefficients form the field which is probably the most popular for tropicalization. We start out by giving a short survey on known variations. For applications to linear programming we are particularly interested in real variants, as these admit an ordering. One example is the field of univariate rational functions with real coefficients, which was already studied by Hilbert. This construction can be extended to rational or even real exponents, and we call the resulting expressions *Puiseux fractions*. They have a natural finite encoding, which makes them useful for computations. In fact, they have been implemented in the software system `polymake`. Further, they have natural primal and dual valuations, such that they are particularly well suited to situations where min and max occur together. We show that generalizations of the classical linear programs of Klee and Minty can be interpreted over Puiseux fractions. Moreover, via Puiseux fractions we investigate a family of linear programs whose central paths exhibit an unusually large total curvature.

3.14 Imaginary projections of polynomials

Thorsten Jörgens (Goethe-Universität – Frankfurt am Main, DE)

For a polynomial $f \in \mathbb{C}[z_1, \ldots, z_n]$, its amoeba $\mathcal{A}(f)$ is defined as the image of its variety in the complex torus $(\mathbb{C}^*)^n$ under the log-absolute map $\log |\cdot|$, respectively as projection to the real part of the complex logarithm, $\mathrm{Re} \circ \log_{\mathbb{C}}(\cdot)$. Its behavior at infinity described by the logarithmic limit set

$$\lim_{r \to \infty} \left(\frac{1}{r} \mathcal{A}(f) \cap \mathbb{S}^{n-1} \right)$$

is one way to define a tropical variety.

We replace the underlying map by the projection of the variety onto its imaginary part. It turns out that the occurring structure, its imaginary projection $\mathcal{I}(f)$, has many similarities to amoebas, such as the convexity of the complement components. But there are also some differences, for example the limit set

$$\lim_{r \to \infty} \left(\frac{1}{r} \mathcal{I}(f) \cap \mathbb{S}^{n-1} \right)$$

is not always polyhedral. In this talk, we consider this structure from several viewpoints and discuss the situation at infinity.

3.15   Computing linear systems on metric graphs

Bo Lin (University of California – Berkeley, US)

The linear system $|D|$ of a divisor D on a metric graph has the structure of a cell complex. And the set $R(D)$ of corresponding tropical rational functions has the structure of a tropical semimodule. We introduce the anchor divisors and anchor cells in it – they serve as the landmarks for us to compute the f-vector of the complex and find all cells in the complex.

Then we compute the minimal set of generators of $R(D)$ using the landmarks. We apply these methods to some examples – namely the canonical linear systems of some small trivalent graphs.

3.16   Abstract tropical linear programming

Georg Loho (TU Berlin, DE)

Analogously to classical oriented matroid programming, we give the first feasibility algorithm for signed tropical oriented matroids, which encode the structure of tropical linear inequality systems. The feasibility problem for these systems is of special interest as it is in NP $\cap$ co-NP but no polynomial time algorithm is known.

To encode the combinatorial structure of tropical linear inequality systems, we equip tropical oriented matroids, which were introduced to study the combinatorics of tropical point configurations, with an additional sign information. We generalize the feasibility problem for tropical linear inequality systems to signed tropical oriented matroids. This allows us to formulate pivoting between basic covectors which only depends on the combinatorial structure and not on the coefficients of the inequalities.

3.17   On integer points in tropical polytopes

Marie MacCaig (Ecole Polytechnique – Palaiseau, FR)

We provide a brief overview of algorithms and complexity relating to finding integer points in tropical polytopes, discussing results for both tropical polytopes describes by vertices, and tropical polytopes described by inequalities.

For the question of existence of an integer point, pseudopolynomial algorithms exist but the complexity is unresolved. We outline a special case which is polynomially solvable and additionally show that adding a simple restriction to the solution set makes the problem NP-hard.

Further we discuss the complexity of counting the number of integer points in a tropical polytope, and the related problem of calculating the volume. We prove that, for tropical polytopes described by inequalities, these problems are #P-hard and, for tropical polytopes described by vertices, even approximation is hard.

3.18 Tropical commutative algebra

Diane Maclagan (University of Warwick – Coventry, GB)

In joint work with Felipe Rincón we have introduced a special class of ideals, called tropical ideals, in the semiring of tropical polynomials, motivated by describing subschemes of tropical toric varieties. We show that these ideals have a well-behaved Groebner theory, with a finite Groebner complex. Consequences include a simple version of the weak Nullstellensatz. The definition involves the theory of valuated matroids.

3.19 Computing faithful tropicalizations of curves: a moduli space perspective

Hannah Markwig (Universität Tübingen, DE)

Joint work in progress with Maria Angelica Cueto, Ralph Morrison and Ilya Tyomkin.

The tropicalization of a subvariety of a torus (or, more generally, of a toric variety) in general depends on the chosen embedding. A faithful tropicalization can be viewed as an embedding for which the tropicalization captures important geometric properties. We demonstrate how one can use modifications to compute faithful tropicalizations for curves for the case of elliptic curves embedded as plane cubics. We also discuss the situation of curves of genus two, three and four, and the relation to moduli spaces of (plane) tropical curves.

3.20 Polynomials in tropical algebra and related algorithmic problems

Vladimir Podolskii (Steklov Institute – Moscow, RU)

In this talk we will review some recent results on tropical polynomials including studies of systems of tropical linear polynomials and a tropical analog of Hilbert's Nullstellensatz.

We will discuss solvability problem and equivalence problem for tropical linear systems. We will discuss their computational complexity. We will also consider the problem of computing the dimension of the solution set of a given tropical linear systems.

For general polynomial systems we will discuss several versions of Hilbert's Nullstellensatz. We will present effective versions of these theorems. The upper bounds on the degrees of polynomials in effective tropical Nullstellensätze have matching lower bounds.

3.21   Tropical ideals

Felipe Rincon (University of Oslo, NO)

We study a special class of ideals, called tropical ideals, in the semiring of tropical polynomials, with the goal of developing a useful and solid algebraic foundation for tropical geometry. The class of tropical ideals strictly includes the tropicalizations of classical ideals, and it satisfies many desirable properties that mimic the classical setup. In particular, every tropical ideal has an associated variety, which we prove is always a finite polyhedral complex. In addition we show that tropical ideals satisfy the ascending chain condition, even though they are typically not finitely generated, and also the weak Nullstellensatz.

3.22   Tropical intersection multiplicity and complexity

J. Maurice Rojas (Texas A&M University – College Station, US)

We recall the Shub-Smale τ-Conjecture, which is a statement on integer roots of polynomials that implies a variant of P$\neq$NP. This conjecture remains open but, via some p-adic tricks, we give two new statements that are potentially easier to prove and still imply new separations of complexity classes. Also, these new variants of the τ-Conjecture lead to interesting open questions on the notion of intersection multiplicity for tropical varieties over p-adic fields. A key subtlety is that current tropical intersection products have not so far focussed on the number of distinct valuations (without multiplicity) in the tropically degenerate case.

3.23   Tropical linear spaces and k-splits

Benjamin Schröter (TU Berlin, DE)

Tropical (uniform) linear spaces are polyhedral complexes, which are dual to a subdivision of the hypersimplex into matroid polytopes. We give explicit constructions for such subdivisions, in particular, for coarsest nontrivial subdivisions. Our focus will be on those tropical linear spaces, whose bounded cells form a k-simplex. These are the so-called k-splits of the hypersimplex. From those tropical linear spaces we derive a new class of matroids, called split matroids. As an application we obtain a dimension bound for the moduli space of those tropical linear spaces, the Dressian.

3.24 Tropical linear algebra with the Lukasiewicz T-norm

Sergei Sergeev (University of Birmingham, GB)

The max-T semiring is defined as the unit interval $[0, 1]$ equipped with the arithmetics $a + b = \max(a, b)$ and ab being a T-norm. Linear algebra and convex geometry over such semiring can be developed in the usual way [1]. We then focus on the case of Lukasiewicz T-norm $\max(0, a + b - 1)$. We observe that any problem of the max-Lukasiewicz linear algebra can be equivalently formulated as a problem of the tropical (max-plus) linear algebra.

Based on this equivalence, we develop a theory of the matrix powers and the eigenproblem over the max-Lukasiewicz semiring [2].

References
1 V. Nitica and S. Sergeev. Tropical convexity over max-min semiring. *Tropical and Idempotent Mathematics and Applications, Cont. Math., AMS*, 616:241–260, 2014.
2 M. Gavalec, Z. Nemcova, S. Sergeev. Tropical linear algebra with the Lukasiewicz T-norm, *Fuzzy Sets and Systems*, 276:131–148, 2015

3.25 Tropical spectrahedra and stochastic mean payoff games

Mateusz Skomra (Ecole Polytechnique – Palaiseau, FR)

Joint work of Xavier Allamigeon, Stéphane Gaubert, Mateusz Skomra

Semidefinite programming (SDP) is a fundamental tool in convex and polynomial optimization. It consists in minimizing the linear functions over the spectrahedra (sets defined by linear matrix inequalities). In particular, SDP is a generalization of linear programming. In practice, approximate solutions to SDPs are obtained using interior point methods, but solving those programs exactly is a challenging task. One of the basic algorithmic questions associated with SDP is to decide whether a spectrahedron is nonempty. It is unknown whether this problem belongs to NP in the Turing machine model, and the state-of-the-art algorithms that certify the (in)feasibility of spectrahedra are based on cylindrical decomposition or the critical points method. We study the nonarchimedean analogue of this problem, replacing the field of real numbers by the field of Puiseux series. We introduce the notion of *tropical spectrahedra*, and show that, under genericity conditions, these objects can be described explicitly by systems of polynomial inequalities in the tropical semiring. Furthermore, we demonstrate a subclass of tropical spectrahedra which encode Shapley operators associated with stochastic mean payoff games. As a result, we show that a large class of semidefinite feasibility problems defined over Puiseux series can be solved efficiently using combinatorial algorithms designed for stochastic games.

3.26 Tropical discriminants in positive characteristic

Luis Tabera (University of Cantabria, ES)

In this talk I will present some results on tropical A-discriminants in positive characteristic and the p-adic case. Unlike the resultant, the A-discriminant depends on the characteristic of the underlying algebraic field. I will also show a discussion of all the cells in a tropical Severi variety in characteristic zero and show that characteristic p discriminants helps us understand some cells of these tropical Severi varieties.

3.27 From combinatorial tropical intersections to the mixed Ehrhart polynomial

Thorsten Theobald (Goethe-Universität – Frankfurt am Main, DE)

For lattice polytopes $P_1, \ldots, P_k \subseteq \mathbb{R}^d$, Bihan [1] introduced the discrete mixed volume $\mathrm{DMV}(P_1, \ldots, P_k)$ in analogy to the classical mixed volume. Departing from combinatorial questions on the intersection of the tropical hypersurfaces defined by tropical polynomials $f_1, ..., f_k$ in d variables with Newton polytopes $P_1, \ldots, P_k$ (as studied in [3]), the associated mixed Ehrhart polynomial $\mathrm{ME}_{P_1,\ldots,P_k}(n) = \mathrm{DMV}(nP_1, \ldots, nP_k)$ arises.

In the talk we discuss the mixed Ehrhart polynomial as well as its occurrence in tropical geometry. We provide a characterization of all mixed Ehrhart coefficients in terms of the classical multivariate Ehrhart polynomial. We also show that for large enough dilates $rP_1, \ldots, rP_k$ the corresponding mixed h^*-polynomial has only real roots and as a consequence the mixed h^*-vector becomes non-negative.

References
1 F. Bihan. Irrational mixed decomposition and sharp fewnomial bounds for tropical polynomial systems. *Discrete & Comp. Geom.*, 55(4):907–933, 2016.
2 C. Haase, M. Juhnke-Kubitzke, R. Sanyal, T. Theobald. Mixed Ehrhart polynomials, Preprint, arXiv: 1509.02254
3 R. Steffens and T. Theobald. Combinatorics and genus of tropical intersections and Ehrhart theory. *SIAM J. Discrete Math.*, 24(1):17–32, 2010.

3.28   Tropical hyperbolic polynomials

Cynthia Vinzant (North Carolina State University – Raleigh, US)

Hyperbolic and stable polynomials are real multivariate polynomials with certain topological properties. Their hypersurfaces bound convex cones, and optimization over these cones generalizes linear and semidefinite programming. There have been several interesting developments relating stable polynomials to matroids and other tropical objects. I will give an introduction to these polynomials and their connections to optimization and tropical geometry and present some related open questions.

Participants

- Marianne Akian
INRIA Saclay –
Île-de-France, FR

- Xavier Allamigeon
INRIA Saclay –
Île-de-France, FR

- Fuensanta Aroca
Universidad Nacional Autonoma –
Mexico, MX

- Frédéric Bihan
University Savoie Mont Blanc –
Le Bourget-du-Lac, FR

- Manuel Bodirsky
TU Dresden, DE

- Timo de Wolff
Texas A&M University –
College Station, US

- Cristhian Garay
Fluminense Federal University –
Rio de Janeiro, BR

- Stéphane Gaubert
INRIA Saclay –
Île-de-France, FR

- Dima Grigoriev
Lille I University, FR

- Alexander Guterman
Moscow State University, RU

- Christian Haase
FU Berlin, DE

- Simon Hampe
TU Berlin, DE

- Thomas Dueholm Hansen
Aarhus University, DK

- Anders Jensen
Aarhus University, DK

- Thorsten Jörgens
Goethe-Universität –
Frankfurt am Main, DE

- Michael Joswig
TU Berlin, DE

- Bo Lin
University of California –
Berkeley, US

- Georg Loho
TU Berlin, DE

- Marie MacCaig
Ecole Polytechnique –
Palaiseau, FR

- Diane Maclagan
University of Warwick –
Coventry, GB

- Hannah Markwig
Universität Tübingen, DE

- Thomas Markwig
Universität Tübingen, DE

- Vladimir Podolskii
Steklov Institute – Moscow, RU

- Felipe Rincon
University of Oslo, NO

- J. Maurice Rojas
Texas A&M University –
College Station, US

- Benjamin Schröter
TU Berlin, DE

- Sergei Sergeev
University of Birmingham, GB

- Mateusz Skomra
Ecole Polytechnique –
Palaiseau, FR

- Luis Tabera
University of Cantabria, ES

- Thorsten Theobald
Goethe-Universität –
Frankfurt am Main, DE

- Cynthia Vinzant
North Carolina State University –
Raleigh, US