Program

Tuesday May 20, 2014
(Institut Henri Poincaré)

9:00 - 9:15 **Registration**

9:15 - 9:30 **Opening Remarks**

9:30 - 10:15 Invited Talk: **Noam Nisan** - *Economic Efficiency Requires Interaction*

We study the necessity of interaction between individuals for obtaining approximately efficient allocations. The role of interaction in markets has received significant attention in economic thinking, e.g. in Hayek's 1945 classic paper.

We consider this problem in the framework of simultaneous communication complexity. We analyze the amount of simultaneous communication required for achieving an approximately efficient allocation. In particular, we consider two settings: combinatorial auctions with unit demand bidders (bipartite matching) and combinatorial auctions with sub-additive bidders. For both settings we first show that non-interactive systems have enormous communication costs relative to interactive ones. On the other hand, we show that limited interaction enables us to find approximately efficient allocations.

Joint work with Shahar Dobzinski and Sigal Oren.

10:15 - 10:45 **Coffee break**

10:45 - 11:15 **Fabio Pardi -** *Phylogenetic networks: what can we reconstruct?*

Phylogenies are used to describe the history of evolutionarily related biological entities (e.g. genes, individuals, species) and are central in many biological applications, including functional genomics, epidemiology and biodiversity assessment. Many methods for reconstructing and studying phylogenies have been proposed, almost all of which use trees to represent them. Although in many cases this is reasonable, in many others phylogenies should be represented as networks (more precisely directed acyclic graphs). This is due to a number of biological phenomena collectively known as reticulation events, whereby a species or a gene inherits genetic material from more than one parent organism. This may be caused by events such as hybridization (e.g. in plants), horizontal gene transfer (e.g. in bacteria) or recombination (e.g. in viruses or in genomes of sexually reproducing species).

Network inference methods are in their infancy, but they are almost invariably based on the following idea: the goodness of a candidate network is evaluated on the basis of how well the subtrees it contains fit the data. This poses a problem: different networks may contain exactly the same set of subtrees (up to isomorphism), meaning that these networks will be considered "indistinguishable" by most network inference methods, no matter the input data. We propose a novel definition of what constitutes a "uniquely reconstructible" network: for each class of indistinguishable networks, we define a canonical form. Under mild assumptions, the canonical form is unique. Given data coming from any phylogenetic network, only its canonical equivalent can be uniquely reconstructed. This is a fundamental limitation that implies a drastic reduction of the solution space in phylogenetic network inference.

Co-author: C. Scornavacca.

11:15 - 11:45 **Gregory Kucherov -** *Cascading Bloom filters applied to genome reconstruction*

One of the popular approaches to assembling genomic sequences from overlapping short fragments (reads) is based on so-called De Brujin graphs. Even if De Brujin graphs have been proposed to make the assembly more feasible, they still occupy a very large space and several approaches have been proposed recently to represent de Bruijn graphs compactly. In this work, we first show how to reduce the memory required by the data structure of Chikhi and Rizk (WABI'12) that represents de Brujin graphs using Bloom filters. Our method requires 30% to 40% less memory with respect to their method, with insignificant impact on construction time. We also show how a Bloom filter representation of de Brujin graphs can be made incremental (online). Finally, we will point out some interesting applications of our technique.

Joint work with Kamil Salikhov et Gustavo Sacomoto.

11:45 - 12:15 **Alessandra Carbone -** *Conservation and co-evolution: from sequence analysis to protein-protein interactions*

12:15 - 12:45 **Benny Chor -** *New facets of conservation in biological sequences*

Short DNA and amino acid sequences (e.g. genes or proteins) have driven early bioinformatics research. Later, however, focus has shifted to more sophisticated objects, such as various kinds of networks, and pure sequences have taken a back seat.

The advent of various high throughput sequencing technologies (aka "new generation sequencing") has given rise to new (or renewed) algorithmic challenges. It has also produced hundreds of complete genome and proteome sequences (with thousands more in the making). In addition to algorithmic questions, we also face many data discovery questions that were inaccessible before. I will describe a number of approaches and results in this area.

The talk will be self contained, and no prior biological knowledge is required.

12:45 - 14:15 **Lunch (To be found in restaurants around. Please see our list of restaurants.)**

14:15 - 15:00 Invited Talk: **Haim Kaplan -** *Two fruitful interactions of theory and practice: max flow and biased search trees*

I will describe two recent cases in which practical applications triggered the development of new algorithms with provable worst case bounds, that also have the best performance in practice. The first is a maximum flow algorithm that is particularly fast on vision applications and has a good strongly polynomial bound on its running time. The second is a variation of splay trees that does fewer rotations and thereby suitable for concurrent scenarios.

15:00 - 15:30 **Iordanis Kerenidis -** *Random Access Codes and Non-Locality*

15:30 - 16:00 **Coffee break**

16:00 - 16:30 **Mikos Santha -** *Generalized Wong sequences and their applications to Edmonds' problems*

We design two deterministic polynomial-time algorithms for variants of a problem introduced by Edmonds in 1967: determine the rank of a matrix M whose entries are homogeneous linear polynomials over the integers. Given a linear subspace B of the $n \times n$ matrices over some field F, we consider the following problems: symbolic matrix rank (SMR) is the problem to determine the maximum rank among matrices in B, while symbolic determinant identity testing (SDIT) is the question to decide whether there exists a nonsingular matrix in B.

The constructive versions of these problems are asking to find a matrix of maximum rank, respectively a nonsingular matrix, if there exists one.

In this talk I present an algorithm that solves the constructive SMR when B is spanned by unknown rank one matrices, answering an open question of Gurvits. The algorithm work over fields of size at least $n+1$. Our framework is based on a generalization of Wong sequences, a classical method to deal with pairs of matrices, to the case of pairs of matrix spaces. I will briefly mention a second algorithm that solves the constructive SDIT over fields of size at least $n+1$ when B is spanned by triangularizable matrices, but the triangularization is not given explicitly.

16:30 - 17:00 **Shai Vardi -** *Local Computation Mechanism Design - Stable Matching*

We introduce the notion of Local Computation Mechanism Design - designing game theoretic mechanisms which run in polylogarithmic time and space. Local computation mechanisms reply to each query in polylogarithmic time and space, and the replies to different queries are consistent with the same global feasible solution. In addition, the computation of the payments is also done in polylogarithmic time and space. Furthermore, the mechanisms need to maintain incentive compatibility with respect to the allocation and payments.

We present local computation mechanisms for a variety of classical game-theoretical problems, focusing on stable matching.

Time permitting, we will show that our local results have general implications. Specifically, that when the men's preference lists are bounded, we can achieve an arbitrarily good approximation to the stable matching within a fixed number of iterations of the Gale-Shapley algorithm.

Joint work with Avinatan Hassidim and Yishay Mansour

Wednesday May 21, 2014
(Institut Henri Poincaré)

9:00 - 9:30 **Guy Even -** *Deterministic Rateless Codes for BSC*

A rateless code encodes a finite length information word into an infinitely long codeword such that longer prefixes of the codeword can tolerate a larger fraction of errors. A rateless code achieves capacity for a family of channels if, for every channel in the family, reliable communication is obtained by a prefix of the code whose rate is arbitrarily close to the channel's capacity. As a result, a universal encoder can communicate over all channels in the family while simultaneously achieving optimal communication overhead.

In this paper, we construct the first *deterministic* rateless code for the binary symmetric channel. Our code can be encoded and decoded in $O(\beta)$ time per bit and in almost logarithmic parallel time of $O(\beta \log n)$, where β is any (arbitrarily slow) super-constant function. Furthermore, the error probability of our code is almost exponentially small $\exp(-\Omega(n/\beta))$. Previous rateless codes are probabilistic (i.e., based on code ensembles), require polynomial time per bit for decoding, and have inferior asymptotic error probabilities.

Our main technical contribution is a constructive proof for the existence of an infinite generating matrix that each of its prefixes induce a weight distribution that approximates the expected weight distribution of a random linear code.

Joint work with Benny Applebaum and Liron David.

9:30 - 10:15 Invited Talk: **Claire Mathieu -** *On the Glass Ceiling Effect in Social Networks*

The glass ceiling may be defined as "the unseen, yet unbreakable barrier that keeps minorities and women from rising to the upper rungs of the corporate ladder, regardless of their qualifications or achievements". Although undesirable, it is well documented that many societies and organizations exhibit a glass ceiling. In this paper we formally define and study the glass ceiling effect in social networks and provide a natural mathematical model that (partially) explains it. We propose a biased preferential attachment model that has two type of nodes, and is based on three well known social phenomena: i) rich get richer (preferential attachment) ii) minority of females (or other group) in the network and iii) homophily (preference to bond with similar people). We prove that our model exhibits a strong glass ceiling effect and that all three conditions are necessary, i.e., removing any one of them, will cause the model not to exhibit a glass ceiling effect. Additionally we present empirical evidence of student-mentor networks of researchers that exhibits all the above properties: female minority, preferential attachment, homophily and a glass ceiling.

Joint work with Chen Avin, Barbara Keller, Zvi Lotker, David Peleg, and Yvonne-Anne Pignolet.

10:15 - 10:45 **Coffee break**

10:45 - 11:15 **Frédéric Magniez -** *Unidirectional Input/Output Streaming Complexity of Reversal and Sorting*

We consider unidirectional data streams with restricted access, such as read-only and write-only streams. For read-write streams, we also introduce a new complexity measure called expansion, the ratio between the space used on the stream and the input size.

We give tight bounds for the complexity of reversing a stream of length n in several of the possible models. In the read-only and write-only model, we show that p-pass algorithms need

memory space $\Theta(n/p)$. But if either the output stream or the input stream is read-write, then the complexity falls to $\Theta(n/p^2)$. It becomes polylog(n) if $p = O(\log n)$ and both streams are read-write.

We also study the complexity of sorting a stream and give two algorithms with small expansion. Our main sorting algorithm is randomized and has $O(1)$ expansion, $O(\log n)$ passes and $O(\log n)$ memory.

11:15 - 11:45 **Adi Rosén -** *Semi-Streaming Set Cover*

We consider the set cover problem under the semi-streaming model. The underlying set system is formalized in terms of a hypergraph $G = (V, E)$ whose edges arrive one-by-one and the goal is to construct an edge cover $F \subseteq E$ while minimizing the cardinality of F. We also consider a relaxation of this problem, where given some $0 \leq \epsilon < 1$, the goal is to construct an edge $(1-\epsilon)$-cover, namely, a subset of edges covering all but an ϵ-fraction of the vertices. The key limitation imposed on the algorithm is that its space is limited to (poly)logarithmically many bits per vertex.

We give an asymptotically tight trade-off between ϵ and the approximation ratio: we design a semi-streaming algorithm that constructs for every $0 \leq \epsilon < 1$, an edge $(1-\epsilon)$-cover that approximates the optimal edge (1-)cover within a factor of $f(\epsilon, n)$ where $f(\epsilon, n) = O(1/\epsilon)$ for $\epsilon > 1/\sqrt{n}$, and $f(\epsilon, n) = O(\sqrt{n})$ for $\epsilon \leq 1/\sqrt{n}$. In particular for the traditional set cover problem we obtain an $O(\sqrt{n})$-approximation. This algorithm is proved to be best possible by establishing a family (parameterized by ϵ) of matching lower bounds.

Joint work with Yuval Emek.

11:45 - 12:15 **Moti Medina -** *Best of Two Local Models: Local Centralized and Local Distributed Algorithms*

We consider two models of computation: local centralized algorithms and local distributed algorithms. Algorithms in one model are adapted to the other model to obtain improved algorithms.

Distributed vertex coloring is employed to design improved local centralized algorithms for: maximal independent set, maximal matching, and an approximation scheme for maximum matching over bounded degree graphs. The improvement is threefold: the algorithms are deterministic, stateless, and the number of probes is poly$(\log^* n)$.

The recursive local centralized improvement technique by Nguyen and Onak is employed to obtain an improved distributed approximation scheme for maximum matching. The number of rounds of the distributed algorithm is $O(\log^* n)$ for bounded degree graphs.

12:15 - 14:15 **Lunch (To be found in restaurants around. Please see our list of restaurants.)**

14:15 - 15:00 Invited Talk: **Irit Dinur -** *On the complexity of correlated inputs*

We study the complexity of computational problems when multiple instances are given, and the instances are *correlated*. Such instances arise for example when they are derived from one "underlying source". Does this make the problem easier? We study this question in the domain of constraint satisfaction problems (CSPs). For example: given several CSP instances with solutions that are 99% the same, and for which the constraints are 95% the same, is it easier to find the solutions? We investigate both the case where we are told which bits differ between the different solutions, and the possibly more interesting case where the differences are unknown. For several variants of CSPs, we show that significant computational advantage can be gained when solving correlated instances. We believe that correlation is a natural

phenomena in practice, and exploiting correlation has potential widespread algorithmic benefits beyond CSPs. Our work suggests exploring correlation as a new dimension for achieving algorithmic goals.

Joint work with Shafi Goldwasser and Rachel Huijia Lin

15:00 - 15:30 **Michal Feldman -** *Beyond Walrasian equilibrium*

15:30 - 16:00 **Coffee break**

16:00 - 16:30 **Tova Milo -** *Query directed crowd mining*

Crowd data sourcing is increasingly used to gather information from the crowd and to obtain recommendations. In this talk, we explore a novel approach that broadens crowd data sourcing by enabling users to pose general questions, to mine the crowd for potentially relevant data, and to receive concise, relevant answers that represent frequent, significant data patterns. Our approach is based on (1) a simple generic model that captures both ontological knowledge as well as the individual history or habits of crowd members from which frequent patterns are mined; (2) a query language in which users can declaratively specify their information needs and the data patterns of interest; (3) an efficient query evaluation algorithm, which enables mining semantically concise answers while minimizing the number of questions posed to the crowd; and (4) an implementation of these ideas that mines the crowd through an interactive user interface. Experimental results with both real-life crowd and synthetic data demonstrate the feasibility and effectiveness of the approach.

16:30 - 17:00 **Serge Abiteboul -** *Introducing Access Control in Webdamlog*

We survey recent work on the specification of an access control mechanism in a collaborative environment. The work is presented in the context of the webdamlog language, an extension of datalog to a distributed context. We discuss a fine-grained access control mechanism for intentional data based on provenance as well as a control mechanism for delegation, i.e., for deploying rules at remote peers.

17:00 - 17:30 **Allan Borodin -** *A myopic model for the non-monotone submodular maximization problem*

What is a conceptually simple algorithm and what is the power and limitations of such algorithms? In particular, what is a greedy algorithm or more generally a myopic algorithm for a combinatorial optimization problem? And to be even more specific, motivated by the Buchbinder et al “online greedy algorithm” for the unconstrained non-monotone submodular maximization problem, what are (if any) the limitations of algorithms "of this genre" for the general unconstrained problem and for specific instances of the problem, such as Max-Di-Cut?

Joint work with Norman Huang

18:30 - 21:00 **Reception (cocktail dînatoire)** - on the ground floor of the IHP

Thursday May 22, 2014
Bâtiment Sophie Germain (U. Paris Diderot)

9:00 - 9:30 **Yossi Azar -** *Single parameter mechanism for unrelated machine scheduling*

We show a randomized truthful mechanism for the restricted-related scheduling model with 3 approximation on the makespan.

Joint work with Idan Maor.

9:30 - 10:15 Invited Talk: **David Peleg -** *Elite and Periphery in Social Networks: An Axiomatic Approach*

Many societies exhibit a two-tier structure composed of a social elite, namely, a relatively small but well-connected and highly influential group of powerful individuals, and the rest of society (referred to hereafter as the periphery). The talk will concern with understanding the structure, size and properties of the elite and the powers that shape it, using an axiom-based model for the relationships between the elite and the periphery.

(Joint work with Chen Avin, Zvi Lotker, Yvonne-Anne Pignolet, and Itzik Turkel)

10:15 - 10:45 **Coffee break**

10:45 - 11:15 **Amos Korman -** *Breathe before Speaking: Efficient Information Dissemination Despite Noisy, Limited and Anonymous Communication*

Distributed computing models typically assume reliable communication between processors. While such assumptions often hold for engineered networks, e.g., due to underlying error correction protocols, their relevance to biological systems, wherein messages are often distorted before reaching their destination, is quite limited. In this study we aim at bridging this gap by analyzing a model of communication in large anonymous populations composed of simple agents which interact through short and highly unreliable messages. We focus on the rumor-spreading problem and the majority-consensus problem, and propose extremely simple algorithms that solve them efficiently despite the severely restricted, stochastic and noisy setting assumed. Our algorithms suggest balancing between silence and transmission, synchronization, and majority-based decisions as important ingredients towards understanding collective communication schemes in anonymous and noisy populations.

11:15 - 11:45 **Pierre Fraigniaud -** *On local distributed decision*

This talk will briefly survey the main results regarding local distributed decision, by putting emphasis on the differences between the deterministic, non-deterministic, and randomized settings. The talk will also question the connections between property testing, local distributed decision, and other forms of decisions with limited information.

11:45 - 12:15 **Amos Fiat -** *A Double Exponential improvement over the cost of Free Parking*

We devise a framework for dynamic pricing schemes such as SFPark in San Francisco with the goal of minimizing social cost. In particular, we consider the objective of minimizing the sum of distances between parking spots and destinations. With this objective, the social cost of free parking may be exponentially worse than the minimal (non-selfish) social cost. We study this problem in various settings, and devise a pricing scheme that gives a double exponential reduction in social cost over the social cost of uniform pricing for parking (and in particular, over the social cost of free parking).

Joint work with Alon Ardenboim, Ilan Cohan, and Riccardo Colini-Baldeschi

12:15 - 14:15 **Lunch (To be found in restaurants around. Please see our list of restaurants.)**

14:15 - 15:00 Invited Talk: **Pascal Koiran -** *A τ-conjecture for Newton polygons*

One can associate to any bivariate polynomial $P(X, Y)$ its Newton polygon. This is the convex hull of the points (i, j) such that the monomial X^iY^j appears in P with a nonzero coefficient. We conjecture that when P is expressed as a sum of products of sparse polynomials, the number of edges of its Newton polygon is polynomially bounded in the size of such an expression. We show that this "τ-conjecture for Newton polygons," even in a weak form, implies that the permanent polynomial is not computable by polynomial size arithmetic circuits. We make the same observation for a weak version of an earlier "real τ-conjecture." Finally, we make some progress toward the τ-conjecture for Newton polygons using recent results from combinatorial geometry.

This talk is based on joint work with Natacha Portier, Sébastien Tavenas and Stéphan Thomassé. I will present our results and conjectures, starting from the very basic properties of Newton polygons (and in particular the role of Minkowski sum).

15:00 - 15:30 **Uri Zwick -** *Adjacency labeling schemes and induced-universal graphs*

We describe a way of assigning *labels* to the vertices of any undirected graph on up to n vertices, each composed of $n/2 + O(1)$ bits, such that given the labels of two vertices, and no other information regarding the graph, it is possible to decide whether or not the vertices are adjacent in the graph. This is optimal, up to an additive constant, and constitutes the first improvement in almost 50 years of an $n/2 + O(\log n)$ bound of Moon. As a consequence, we obtain an *induced-universal* graph for n-vertex graphs containing only $O(2^{n/2})$ vertices, which is optimal up to a multiplicative constant, solving an open problem of Vizing from 1968. We obtain similar tight results for directed graphs, tournaments and bipartite graphs.

Joint work with Stephen Alstrup, Haim Kaplan and Mikkel Thorup

15:30 - 16:00 **Cakes and coffee**