Between 2011 January and 2013 December, we worked on a bunch of research projects, we list all of those that lead to a publication and/or a submitted manuscript.

- We gave an MCMC method to sample the Bayesian distribution of parameters of vegetation dynamics. Given a set of vegetation maps from different years, the state space of the Markov chain consist of missing data (vegetation maps from other years where data is not available) and parameters that describe the speed of changing vegetation. The Markov chain converges to the Bayesian distribution, which is proportional to the product of the likelihood (the probability of observing the data and missing data under the given parameters) and the prior probabilities of dynamics parameters. The method has been applied to real data and the software is freely available. It was a joint research with Imelda Somodi and Klára Virágh, and the method has been published in Landscape Ecology.
- We showed that there was a positive selection on the HD motifs in Alzheimer Precursor Proteins (APPs). This HD motif is very likely has a role in metal binding and can be found in a disordered part of the protein. Interestingly, this motif is not position-specifically conserved, namely, cannot be detected by a multiple sequence alignment. However, we developed a statistical test to show that there is an evolutionary selection of this motif. Indeed, this motif can be found in the majority of the APP homologs, and if there were no selection then the number of homologous proteins in which the HD motif was presented would be significantly less due to random drift. Our findings might open new avenues in studying intrinsically disordered proteins. This research was a joint project with Zoltán Zádori and it has been published in PLoS Computational Biology.
- We proved that the most parsimonious Double Cut and Join (DCJ) scenarios can be almost uniformly sampled in polynomial time and their number can be FPRAS approximated (Fully Polynomial Randomized Approximation Scheme). The DCJ model revolutionized the computational biology research on genome rearrangement. It is one of the simplest, but still usable model for genome rearrangement. Due to its simplicity, it is possible to count and sample uniformly DCJ scenarios between co-tailed genomes, as was proven in an earlier work of Anna Bergeron and Aida Ouangroua. Together with Eric Tannier, we conjectured that the general case was computationally hard (#P-complete) and we proved that it is FPRAS approximatable and can be almost uniformly sampled in polynomial time. These proves have been published in Theoretical Computer Science.
- We proved that the swap Markov chain mixed rapidly on realizations of bipartite half-regular degree sequences. Generating simple graphs with a given degree sequence (or 0-1 matrices with given row and column sums) from the (almost) uniform distribution is essential in some statistical hypothesis testing. A few samples are sufficient to generate a distribution of a null model, and calculate threshold values for some p-values for a statistics of interest. Applications include analysis of presence-absence ecological matrices and statistical inferring of networks (neural networks, social networks, biochemical networks, etc.). One possible way to generate such matrices or networks is to use Markov chains. The swap Markov chain is conjectured to be rapidly mixing (and therefore it is computationally efficient to use such Markov chain to generate samples), but before our work, it was proved only regular degree sequences. We made a step further, and proved that it is also true for half-regular degree sequences. Joint work with Péter Erdős and Lajos Soukup, and published in Electronic Journal of Combinatorics.

- We also proved that the swap Markov chain is rapidly mixing on such realizations of half-regular degree sequences that contain a forbidden one-factor and a star-tree. Furthermore, it is also rapidly mixing on the balanced realizations of JDMs. The results are written into manuscripts available on arXiv (http://arxiv.org/abs/1301.7523, http://arxiv.org/abs/1302.3548, http://arxiv.org/abs/1307.5295) and under review at the time writing this report.
- We gave a formula for the swap distance. This purely combinatorial result might be useful for further research on proving the rapid mixing of the swap Markov chain. Joint work with Zoltán Király and Péter Erdős, and published in Combinatorics, Probability and Computating.
- We gave an efficient algorithm for modulated string searching. This stingology problem arose as a Next Generation Sequencing (NGS) problem. Row NGS data consist of measuring luminescence data, from which an estimation of the length of the run of the same nucleotides can be obtained. This data can be transformed into a regular expression and more generally into a form what is called modulated string searching. Joint work with Alberto Apostolico, Péter Erdős, Johannes Simmeons and published in Theoretical Computer Science.
- We showed that the number of most parsimonious Single Cut or Join (SCJ) scenarios cannot be approximated in an FPRAS manner and they cannot be sampled almost uniformly in polynomial time unless RP = NP. The SCJ model is the computationally simplest genome rearrangement model for which the small parsimony problem can be solved in polynomial time. Therefore, it is reasonable to assume that at least an FPRAS approximation exists for the number of such scenarios. However, the opposite is true. This negative result is written into a manuscript that is available on arXiv (http://arxiv.org/abs/1304.2170) and under review. Joint work with Sándor Z. Kiss and Eric Tannier.
- We proved the 'four reversal conjecture' for linear graphs. The four reversal conjecture is the following: consider the graph, whose vertices are the most parsimonious reversal scenarios sorting a signed permutation, and two vertices are connected if they differ in at most four reversals (roughly speaking, see the precise definition in the indicated reference). The conjecture is that this graph will always be connected, namely, a Markov chain whose transition kernel is built up using such small perturbations is irreducible. The relevance of this conjecture is that it would open a new avenue to develop a rapidly mixing Markov chain inferring the most parsimonious rearrangement scenarios between two genomes under the Hannenhalli-Pevzner rearrangement model. The conjecture is open for 7 years now, and we reached a partial result: we proved that the conjecture is true for signed permutations whose overlap graphs are linear graphs. Joint work with Eliot Bixby and Toby Flint, manuscript available on arXiv (http://arxiv.org/abs/1304.2170) and under review.
- I wrote an electronic book for BSc maths students on bioinformatics, available at http://www.renyi.hu/~miklosi/SztochasztikusModellek.pdf

To sum up, in the last 3 years, we published 6 scientific peer-reviewed papers, wrote 1 electronic book and submitted further 5 manuscripts.