

Lecture series on the subject "The interplay between fundamental computer science and data science"

Chair: [Prof. Dr. Florin Manea](#)

Abstract and brief biography

3. Gregory Kucherov: Compact data structures for bioinformatics with application to k-mer tables

Abstract: We first give a brief overview of several techniques increasingly used in modern data-intensive bioinformatic applications within the general paradigm of alignment-free methods. Those techniques include compact data structures — compressed static functions, Bloom filters — and probabilistic sketching techniques — locality-sensitive hashing and, in particular, minimizers. In the second part of the talk, we illustrate how these techniques can be applied to efficiently representing k-mer count tables for large genomic datasets.

Brief Bio:

Gregory Kucherov is a CNRS research director in Gaspard Monge Lab for Computer Science in Gustave Eiffel University (until 2019, Paris-Est University) at Marne-la-Vallée, France. His research interests include algorithm design and engineering, algorithmic complexity, word combinatorics, data structures, sequence and graph algorithms with applications to bioinformatics and other big data applications. Gregory Kucherov (co)-authored about 50 journal publications and more than 50 conference papers. He regularly serves on program committees of international conferences, has been invited to present his work to various international events, has served as editor of LNCS volumes and journal special issues. He supervised several software projects in bioinformatics, including ProPhyle, mreps, Yass, or Norine.