

Integer Linear Programming in Computational and Systems Biology

Keynote Address

Dan Gusfield
Department of Computer Science
University of California, Davis
Davis, CA 95616, USA
gusfield@cs.ucdavis.edu

Abstract

Integer (Linear) Programming, abbreviated “ILP”, is a versatile modeling and optimization technique that was developed for complex planning and operational decision making. However, it has been increasingly used in computational biology in non-traditional ways, most importantly and inventively as a computational tool to model biological phenomena, to analyze biological data, and to extract biological insight from the models and the data. Integer programming is often very effective in solving instances of hard biological problems on realistic data of current importance, despite the fact that many of those problems lack general algorithmic solutions that are efficient (in a provable, worst-case sense), and that the problem of solving integer programs also lacks a provable worst-case efficient general solution.

Highly engineered, commercial ILP solvers are available (now free to academics and researchers) to solve ILP formulations. The improvement of the best solvers has been spectacular, with an estimate that (combined with faster computers) benchmark ILP problems can now be solved 200-billion times faster than twenty-five years ago. Exploiting ILP, some biological problems of importance can be modeled in a way that allows a solution in seconds on a laptop, while more common (statistically-based) models require days, weeks or months of computation on large clusters.

The effectiveness of the best ILP solvers on

problem instances of importance in biology opens huge opportunities. The impact of faster and easier-to-implement computation could be truly transformative in several parts of biology. However, there are challenges in effectively using these tools for biological problems, and educational and outreach issues that must be addressed. In this talk, I will discuss some of the successes, opportunities, and challenges in exploiting ILP for computational and systems biology.

Bibliography

Dr. Gusfield’s primary interests involve the efficiency of algorithms, particularly for problems in combinatorial optimization and graph theory. These algorithms have been applied to study data security, stable matching, network flow, matroid optimization, string/pattern matching problems, molecular sequence analysis, and optimization problems in population-scale genomics. Currently, he is focused on string and combinatorial problems that arise in computational biology and bioinformatics. Dr. Gusfield served as chair of the computer science department at UCD from July 2000 until August 2004, and was the founding Editor-in-Chief of The IEEE/ACM Transactions of Computational Biology and Bioinformatics until January 2009.

Dan Gusfield received his Ph.D. in 1980 from UC Berkeley, working with Richard Karp, and was an Assistant Professor at Yale University from 1980 to 1986. His dissertation concerned problems of

sensitivity analysis in graphs, network flow and Matroid theory. In January 1987 Dan moved to UC Davis. In July 2016, he was promoted to the rank of Distinguished Professor.