

Introduction to the Special Issue on Biological Networks

Natasa Przulj

In this special issue on biological networks, we aim to interest the readership of *Internet Mathematics* in network theory applied to bioinformatics. Network biology is a new and emerging research area that is fast-growing, spurred by the collection of biological data representing connections or interactions of molecules in the cell. As such, it has the potential to have at least as profound an impact on our understanding of the cell as sequence data has had. However, the datasets are large, noisy and many graph theoretic problems are formally intractable (impossible to solve exactly in any time less than the age of the universe), and so heuristic approximations must be developed in an attempt to find approximate solutions. Furthermore, the tools developed to solve these problems must be made accessible to biological practitioners. In this direction, this issue contains papers on the many databases available, theoretical and algorithmic advances in analyzing these data, as well as papers on some specific biomedical applications, and two papers introducing software tools.

This issue presents six papers from some of the leading research groups in the area. Three papers present significant theoretical advances in techniques. Two of them (Elberfeld et al.; Crofts and Higham) look at directed graphs. First, Elberfeld et al. attack the “maximum graph orientation problem”, in which, given a list of source-sink pairs of nodes, we attempt to add direction to an undirected graph in such a way as to maximize the number of pairs for which directed paths exist from the source to the sink. This has applications in the problem of learning biological pathways, but Elberfeld et al. show that the problem is NP-hard

and then proceed to develop several approximate algorithms, as well as bounds on the quality of the solutions produced. Our second paper on directed graphs (Crofts and Higham) attempts to uncover hidden hierarchy in a directed graph that is likely to have hierarchy that is obscured simply because the nodes are ordered in a way that obscures the hierarchy. They show how to re-order the nodes in a way that exposes the hierarchy as obvious. They also derive measures of how hierarchical a graph actually is in comparison to a novel model of random hierarchical graphs. Applications include an analogy to Google's PageRank algorithm as well as network neuroscience predictions that can be tested biologically. Finally, our third paper presenting theoretical advances (Wittkop et al.) extends previous work by the Baumbach group on "Transitivity Clustering" by presenting a theoretical analysis of the technique, and then comparing their tool to existing tools for clustering biomedical objects based on common traits.

Without data, there would be no bridge between theory and practice. In this issue, Wong provides a comprehensive survey of the large number of databases currently available for the storage and analysis of biological networks.

Two new software tool papers are presented in this issue. First, Alcaraz et al. note that biological networks and gene expression data are often studied separately from each other. They introduce KeyPathwayMiner, which brings the two types of data together by attempting to find case-specific pathways of interest to a given series of gene expression studies. Along with the algorithmic development, they also provide a theoretical justification for why the problem is hard, and why their heuristic (Ant Colony Optimization) is a good way to solve this particular problem.

It is a simple fact that humans are still often better than computers at recognizing patterns. In fact, many algorithms are developed only after a human has visualized the problem in a new way and used the insight to design a novel algorithm to solve the problem. Our second software tool is a graph visualization tool called NAViGaTOR (Djebbari et al.). They note that, given the complexity of graphs, there is rarely a single best way to visualize even a single graph. As such, NAViGaTOR allows any graph to be viewed using multiple "scenarios", also called "workflows". Such visualization has the potential to spur the creation of new algorithms and better data analysis by allowing users to simultaneously visualize a graph in multiple ways.

As illustrated by these papers, this area of research promises to make a substantial contribution to biomedical understanding. We hope that the readership of *Internet Mathematics* will enjoy this issue of the journal, and that these papers will spur a new wave of submissions of similar papers in network biology to this journal.