SHORTEST CONNECTIVITY

COMBINATORIAL OPTIMIZATION

VOLUME 17

Through monographs and contributed works the objective of the series is to publish state of the art expository research covering all topics in the field of combinatorial optimization. In addition, the series will include books, which are suitable for graduate level courses in computer science, engineering, business, applied mathematics, and operations research.

Combinatorial (or discrete) optimization problems arise in various applications, including communications network design, VLSI design, machine vision, airline crew scheduling, corporate planning, computer-aided design and manufacturing, database query design, cellular telephone frequency assignment, constraint directed reasoning, and computational biology. The topics of the books will cover complexity analysis and algorithm design (parallel and serial), computational experiments and application in science and engineering.

SHORTEST CONNECTIVITY
An Introduction with Applications in Phylogeny

by

DIETMAR CIESLIK
Ernst-Moritz-Arndt University, Greifswald, Germany
Massey University, Palmerston North, New Zealand

Springer

Library of Congress Cataloging-in-Publication Data

A C.I.P. Catalogue record for this book is available
from the Library of Congress.

ISBN 0-387-23538-8 e-ISBN 0-387-23539-6 Printed on acid-free paper.

Printed in the United States of America.

9 8 7 6 5 4 3 2 1 SPIN 11336228

springeronline.com

CONTENTS

PREFACE

The problem of "Shortest Connectivity" has a long and convoluted history. Usually, the problem is known as Steiner's Problem and it can be described more precisely in the following way: Given a finite set of points in a metric space, search for a network that connects these points with the shortest possible length. This shortest network must be a tree and is called a Steiner Minimal Tree (SMT). It may contain vertices different from the points which are to be connected. Such points are called Steiner points.

Steiner's Problem seems disarmingly simple, but it is rich with possibilities and difficulties, even in the simplest case, the Euclidean plane. This is one of the reasons that an enormous volume of literature has been published, starting in the seventeenth century and continuing today.

Over the years Steiner's Problem has taken on an increasingly important role. More and more real-life problems are given which use Steiner's Problem or one of its relatives as an application, as a subproblem or as a model.
We will discuss the problem of "Shortest Connectivity" as a general approach to investigate real structures in nature. We will see that this involves the identification of a combinatorial structure that requires the smallest number of changes. It is often said that this principle abides by Ockham's razor, according to which the best hypothesis is the one requiring the smallest number of assumptions.[1]

At first we will give an overview of Steiner's Problem and its relatives as one of the most interesting optimization problems in the intersection of combinatorics and geometry. In this sense, the present book is an introduction to the theory of "Shortest Connectivity". We will see that Steiner's Problem is the core of the so-called "Geometric Network Design Problems", where the general problem can be stated as follows: given a configuration of vertices and/or edges, find a network which contains these objects, satisfies some predetermined re-

[1] Roughly speaking: Do not increase the number of entities without unnecessarity.

quirements, and which minimizes a given objective function that depends on several distance measures.

Secondly, we will discuss a new challenge, namely to create trees which reflect the phylogeny, which is the evolutionary history of "living entities". For 3.5 billion years, since life on earth began, evolution has created a remarkable variety of organisms. Millions of different species are alive today, while countless have become extinct. To describe the evolution of these species is a fundamental problem that has been of interest at least since Charles Darwin first proposed the theory of evolution more exactly.
Trees are widely used to represent evolutionary relationships. In biology, for example, the dominant view of the evolution of life is that all existing organisms are derived from some common ancestor and that a new species arises by a splitting of one population into two or more populations that not do not cross-breed, rather than by a mixing of two populations into one. The principle of Maximum Parsimony involves the identification of a combinatorial structure that requires the smallest number of evolutionary changes. Note that here, minimizing the number of assumptions does not mean minimizing the steps of an evolution[2], it means that among all possible structures we seek one which satisfies only one, and moreover a natural, condition.
We will consider the problem of reconstruction of phylogenetic trees in our sense of shortest connectivity. To do this we introduce the so-called Phylogenetic spaces. These are metric spaces whose points are arbitrary words generated by characters from some alphabet, and the metric measuring "similarity" of the words is generated by a cost measure on the characters. The "central dogma" will be: A phylogenetic tree is an SMT in a desired chosen phylogenetic space.

In any case this topic contains many problems for further research.

The aim in this graduate-level text is to outline the key mathematical concepts that underpin the important questions in applied mathematics. These concepts involve discrete mathematics (particularly graph theory), optimization, computer science, and several ideas in biology.

Acknowledgements.

I thank all people who supported my research and gave me helpful advice on how to write this book: A. Dress (Bielefeld), W.M. Fitch (Irvine), P.Gardner (Palmerston North), R. Graham (La Jolla), M.D. Hendy (Palmerston North), K.Huber (Uppsala), A.v.Haeseler (Jülich/Düsseldorf), A.O. Ivanov (Moscow),

[2]Whatever that means!

A. Kemnitz (Braunschweig), V.Moulton (Uppsala), P. Pardalos (Gainesville), D.Penny (Palmerston North), H.J.Prömel (Berlin), J.MacGregor Smith (Amherst), M.Steel (Christchurch), A.A. Tuzhilin (Moscow), D.M. Warme (Alexandria) and J. Wills (Siegen).

I thank Tim White (Palmerston North) and my student K.Kruse for proof reading of the manuscript.

Heidrun G. Köhler (Greifswald) gave a lot of remarks regarding writing this book in a suitable style.

Moreover, I thank my colleagues H.-R. Frieling, W. Girbardt and W. Passauer for helpful technical support.

I thank the Institute of Fundamental Sciences, Massey University, New Zealand; the von Neumann Institut for Computing, Forschungszentrum Jülich, Germany; and the Allan Wilson Centre for Molecular Evolution and Ecology, Massey University, New Zealand for hosting me during the winter 2001/02, the spring of 2002, and the spring of 2003, respectively.